FRAMEWORK FOR THE CONTROL OF QUALITY
IN
AUTOMATED MAPPING

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of the Ohio State University

By

Robert William Schmidley, B.S., M.S.

* * * * *

The Ohio State University
1996

Dissertation Committee:
John D. Bossler
J. Raul Ramirez
Alan Saalfeld

Approved by

Advisor
Department of Geodetic Science and Surveying
To My Family
ACKNOWLEDGEMENTS

I must acknowledge the help, support, advice and encouragement I received from many people in completing this work. Thanks to my advisor Dr. John Bossler of the Center for Mapping for the providing excellent opportunities to work and learn. A special debt of gratitude is owed Dr. Raul Ramirez for all his tireless efforts in my behalf. I am also grateful to Dr. Alan Saalfeld for his helpful feedback and to Dr. Gerald Newsom for his very careful reading of the draft. Dr. Joseph Loon, who introduced me to the subject of cartography, also deserves my appreciation.

I must also thank John Snowden and the staff at the Center for Mapping for their help, patience and support in this effort. Graduate student colleagues Joe Szakas, Eduardo Falcon and many others deserve my deep gratitude. Finally, to my wife Sally and son Mark, for their steadfast endurance throughout this project, a warm thanks.
VITA


1980 . . . . . . . . . . . B.S., Trinity College, Hartford, Connecticut


1988 . . . . . . . . . . M.S., The Ohio State University, Columbus, Ohio

1989-1994 . . . . . . Graduate Teaching Associate, Department of Geodetic Science and Surveying, Graduate Research Assistant, Center for Mapping, The University, Columbus, Ohio

PUBLICATIONS


Major Field: Geodetic Science

Studies in Cartography

Joseph C. Loon, Assistant Professor
J. Raul Ramirez, Adjunct Assistant Professor

Studies in Adjustment Computations

Burkhard Schaffrin, Associate Professor
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEDICATION</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>VITA</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>ix</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td><strong>I.  INTRODUCTION TO QUALITY</strong></td>
<td>1</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Purpose of Research</td>
<td>1</td>
</tr>
<tr>
<td>Scope and Limitations of Research</td>
<td>3</td>
</tr>
<tr>
<td>A Note on Terminology</td>
<td>5</td>
</tr>
<tr>
<td>Outline of Presentation</td>
<td>7</td>
</tr>
<tr>
<td>Primer on Quality</td>
<td>9</td>
</tr>
<tr>
<td>Brief History of Quality: The Rise of Japan</td>
<td>9</td>
</tr>
<tr>
<td>Factors Other than Competition</td>
<td>12</td>
</tr>
<tr>
<td>Quality/Efficiency Relationship</td>
<td>16</td>
</tr>
<tr>
<td>Conclusion</td>
<td>18</td>
</tr>
<tr>
<td>Quality and Quality Control</td>
<td>20</td>
</tr>
<tr>
<td>Quality Defined</td>
<td>20</td>
</tr>
<tr>
<td>Quality Control</td>
<td>24</td>
</tr>
<tr>
<td>Statistical Quality Control</td>
<td>30</td>
</tr>
<tr>
<td>Quality Measurement</td>
<td>35</td>
</tr>
<tr>
<td>Conclusion</td>
<td>39</td>
</tr>
<tr>
<td><strong>II. QUALITY ISSUES IN MAPPING</strong></td>
<td>42</td>
</tr>
<tr>
<td>Factors Affecting Mapping Quality</td>
<td>42</td>
</tr>
<tr>
<td>The Public Sector and the National</td>
<td></td>
</tr>
<tr>
<td>Spatial Data Infrastructure (NSDI)</td>
<td>47</td>
</tr>
<tr>
<td>Conclusion</td>
<td>54</td>
</tr>
</tbody>
</table>
### III. GENERAL CHARACTERIZATION OF CARTOGRAPHIC DATA QUALITY

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review of Spatial Data Standards</td>
<td>58</td>
</tr>
<tr>
<td>United States National Map Accuracy Standards</td>
<td>58</td>
</tr>
<tr>
<td>ASPRS Accuracy Standards for Large-Scale Maps</td>
<td>64</td>
</tr>
<tr>
<td>Draft National Cartographic Standards for Spatial Accuracy</td>
<td>68</td>
</tr>
<tr>
<td>USGS Standards for Digital Line Graphs</td>
<td>71</td>
</tr>
<tr>
<td>Spatial Data Transfer Standard</td>
<td>77</td>
</tr>
<tr>
<td>Definition of Cartographic Information Quality in SDTS</td>
<td>86</td>
</tr>
<tr>
<td>Conclusion</td>
<td>89</td>
</tr>
<tr>
<td>Defining Cartographic Quality from a Different Perspective</td>
<td>91</td>
</tr>
<tr>
<td>Cartographic Quality of Design</td>
<td>93</td>
</tr>
<tr>
<td>Cartographic Quality of Conformance</td>
<td>95</td>
</tr>
<tr>
<td>General Outline of Cartographic Quality</td>
<td>99</td>
</tr>
<tr>
<td>Conclusion</td>
<td>103</td>
</tr>
</tbody>
</table>

### IV. FRAMEWORK FOR THE CONTROL OF CARTOGRAPHIC DATA QUALITY

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Subjects for English Text</td>
<td>107</td>
</tr>
<tr>
<td>The Nature and Language of Cartographic Communication</td>
<td>112</td>
</tr>
<tr>
<td>Cartographic Alphabet</td>
<td>118</td>
</tr>
<tr>
<td>Cartographic Control Subjects</td>
<td>125</td>
</tr>
<tr>
<td>Applying Control Subjects in Cartographic Models</td>
<td>142</td>
</tr>
<tr>
<td>Problem of Higher Dimensions of Geographic Information</td>
<td>153</td>
</tr>
<tr>
<td>Conclusion</td>
<td>159</td>
</tr>
</tbody>
</table>

### V. MEASURING CARTOGRAPHIC DATA QUALITY

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Aspects of Quality Measurement in Cartography</td>
<td>172</td>
</tr>
<tr>
<td>Measurement in the Control of Quality in Cartography</td>
<td>172</td>
</tr>
<tr>
<td>Metric Concept</td>
<td>174</td>
</tr>
<tr>
<td>Measurement of Geometric Quality</td>
<td>177</td>
</tr>
<tr>
<td>Area Points</td>
<td>177</td>
</tr>
<tr>
<td>Point Discrepancy Measures</td>
<td>181</td>
</tr>
<tr>
<td>Summary and Statistical Measures for Geometric Point Accuracy</td>
<td>187</td>
</tr>
<tr>
<td>Conclusion</td>
<td>210</td>
</tr>
</tbody>
</table>
VI. LINEAR GEOMETRIC QUALITY ......................... 214

Introduction ........................................... 214
The Graphic Approach to Linear Quality ................. 216
Representing Lines ..................................... 218
Linear Measures and Problems .......................... 221
Maximum Separation .................................... 222
The Problem of Line Parametrization ................. 225
Continuous Distance .................................. 227
A Standardized Parametrization ....................... 229
L_p Metric ............................................ 236
Hausdorff Metric ..................................... 241
Area-Based Measures .................................. 244
Vector Component Measures ......................... 252
Discrete Linear Sampling .............................. 263
Conclusion ............................................ 271

VII. CONTROLLING QUALITY IN MAPPING ............... 274

Applying the Conceptual Framework to the Control of Quality in Mapping .... 274
Statistical Control in Mapping ....................... 289

VIII. FUTURE DIRECTIONS ................................. 301

LIST OF REFERENCES ................................. 308
LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Scale factors (k) for axes of the standard error ellipse with related probabilities (Mikhail and Ackerman, 1976, page 32)</td>
<td>201</td>
</tr>
<tr>
<td>2. Probabilities (P) of circular accuracy indices related to the circular standard error (σc) (Maling, 1989, page 153)</td>
<td>203</td>
</tr>
<tr>
<td>3. Comparison of approximations for the radius of the probability circle for case when σxy=0, σx=2.14596603 and σy/σx=0.25</td>
<td>208</td>
</tr>
<tr>
<td>FIGURES</td>
<td>PAGE</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>1. Deming's Quality Chain Reaction (Deming, 1986, page 3)</td>
<td>18</td>
</tr>
<tr>
<td>2. Control Feedback Loop (Juran and Gryna, 1993, page 99)</td>
<td>28</td>
</tr>
<tr>
<td>3. Change in DLG horizontal positional accuracy requirement</td>
<td>74</td>
</tr>
<tr>
<td>4. Cartographic Alphabet</td>
<td>121</td>
</tr>
<tr>
<td>5. Example of a skeletal level cartographic expression</td>
<td>121</td>
</tr>
<tr>
<td>6. Fragment from USGS 7.5 minute, 1:24,000 digital line graph file for Kassler, Colorado showing point information</td>
<td>127</td>
</tr>
<tr>
<td>7. Fragment from USGS 7.5 minute, 1:24,000 digital line graph file for Kassler, Colorado showing line information</td>
<td>130</td>
</tr>
<tr>
<td>8. Graphic example showing use of the area point to carry attribute information for an areal feature</td>
<td>135</td>
</tr>
<tr>
<td>9. Example showing how the area point can be used in digital information to carry attribute information for an areal feature</td>
<td>135</td>
</tr>
<tr>
<td>10. Fragment from USGS 7.5 minute, 1:24,000 digital line graph file for Kassler, Colorado showing areal attribute information</td>
<td>137</td>
</tr>
<tr>
<td>11. Example showing the indeterminacy of areal attributes carried on linear objects</td>
<td>139</td>
</tr>
<tr>
<td>12. Graphic example showing the indeterminacy of areal attributes carried on linear objects</td>
<td>140</td>
</tr>
</tbody>
</table>
13. Example of the raster data model .............. 144
14. Example of the objects in
   the vector data model .......................... 145
15. Euclidean distance metric for
   geometric point discrepancy .................. 182
16. Geometric point discrepancy
   measure in which x and y
   components are separate quantities ............ 184
17. Two-dimensional vector measure
   for geometric point discrepancy ............... 186
18. Horizontal error for a linear feature
   (Thompson, 1956, page 165) .................... 215
19. Maximum separation measure for lines
   and the problem of comparable points .......... 224
20. Example of asymmetry of
    the Hausdorff distance ....................... 242
21. Areal separation measure ....................... 244
22. Need for normalizing the
    areal separation measure ...................... 245
23. Example where areal separation measure
    fails to satisfy the triangle inequality .... 246
24. Ambiguity of areal separation measure
    for small closed linear features .............. 248
25. Illustration of the
    buffering/polygon overlay approach
    to linear discrepancy measurement ............ 249
26. Example where an area-based linear
    discrepancy measure fails to characterize
    the difference between two lines ............. 251
27. Example where average discrepancy
    vector is misleading .......................... 255
28. Example where the average discrepancy
    vector is a reasonable measure of linear
    difference .................................... 256
29. Example of the parametric linear error function associated with the difference between two lines .......... 265

30. Illustration of the difficulty of interpreting multiple occurrences of parametric linear error functions .......... 266

31. Nature of the error function in the case of polygonal data ............... 267

32. Operational steps in the creation of a product .......................... 275

33. Primary components of the framework for controlling quality in mapping .......... 277

34. Control Feedback Loop (Juran and Gryna, 1993, page 99) ............... 278

35. Translating user requirements into operational specifications .............. 281

36. Primary aspects of process evaluation, measurement of inputs, process and output ... 286

37. Quality control feedback loop in mapping based on the framework elements .......... 289

38. Example of a nonrandom distribution of error with respect to space .............. 297

39. Examples of rational subgrouping arrangements for geometric discrepancies based on spatial ordering ............... 299
CHAPTER I
INTRODUCTION TO QUALITY

Introduction

Purpose of Research

There is today a great deal of concern for quality. Evidence of the increased importance of product quality is found in the recent establishment of quality awards and standards. The International Standards Organization (ISO) published a set of quality standards, known as the ISO 9000 series, that has gained widespread acceptance in the European Community and in several other countries, including the United States. Quality issues are also being raised and considered in the information processing and mapping fields. For example, the Association for Computing Machinery (ACM) devoted a special section of the Communications of the ACM to software quality (November, 1993). The importance of quality in mapping and geographic information systems (GIS) has been noted as well (Nugent, 1995; Dominguez, 1994; Epner and Parmenter, 1993; Heideman and Angwin, 1992; and Bissex, Franks and Heitkamp, 1990). It is anticipated that quality will become, if it is not already, a major concern for members of the cartographic community.
However, most of the quality wisdom has evolved in the industrial and manufacturing domain and has not been widely applied in the context of data and information. Methods for controlling the quality of manufacturing processes are well developed but have not been extensively applied in the production of information in general and spatial data in particular. Therefore, much of the knowledge about quality needs to be transferred and adapted to the cartographic context. This research will investigate the general principles of quality and its control so they can be applied to the production of maps and spatial data. The focus will be on the development of a framework for controlling the quality of processes generating cartographic information.

One of the central issues in quality control is the problem of quality measurement. For this reason, the principal aim of this investigation will be to define what spatial data quality is and how it can be measured. By setting forth a framework for defining and measuring spatial data quality, it will then be possible to apply the methods of industrial quality control, particularly statistical quality control (SQC), to the production of cartographic information.

The application of quality control techniques to the production of geographic information will, in turn, set the stage for quality improvement. Most quality experts agree that improving quality leads to increased productivity, lower costs, and greater overall organizational
effectiveness (Feigenbaum, 1991, pages xvii, 45-47; Deming, 1986, pages 1-2 and Ishikawa, 1985, pages 54-55). The ultimate impact of this research, then, will be to establish the foundation for improving the quality and efficiency of the production of maps and spatial data.

Scope and Limitations of Research

In so far as measurement is the basis of the control of quality, this study will address the fundamental problems of what cartographic data quality is and how it can be measured. A satisfactory understanding of these issues will constitute the general framework for the control of spatial data quality. The research will concentrate on the individual unit of cartographic information and the problem of defining and evaluating its quality. In other words, the focus will be on the technical aspects of quality measurement in mapping. The broader problem of quality management--which addresses institutional, organizational, behavioral and economic issues--will not be considered in this research.

The research will concentrate on the kind of spatial data and information that is contained in maps and GIS databases. It is assumed that such information is limited to a static, planar model of the world. While several of the ideas developed may indeed be applicable in the context of three-dimensional spatial data, no deliberate attempt is made to be so inclusive. The results are intended to be
adequate for two-dimensional, planar cartographic information. In general, the class of cartographic information contemplated here is assumed to apply to general purpose, topographic maps rather than thematic maps or charts (Robinson, Sale, Morrison and Muehrcke, 1984, pages 7-10). Although no particular attempt will be made to avoid treatment of chart and thematic map data (and the results are probably applicable to these as well), the examples and arguments will be developed on the basis of general purpose, topographic map data. Since a static model is assumed, the problem of time is not addressed either. In other words, time is not treated as a continuous dimension in addition to the two (planar) spatial dimensions; it is considered only to the extent it is a descriptive attribute of a planar spatial object. Another consequence of the static view is the mapping process is regarded as a one-time operation rather than an on-going enterprise. Spatial information is generated only once in the mapping process; it is not treated as a database that must be continually kept up to date, as in a bank or insurance company. For this reason, the issue of data currency will not be significant in this research.

Since the emphasis is on controlling the quality of a process that generates spatial data, the issues typically associated with map design will not be covered in detail here. A complete description of cartographic data quality
would certainly have to take into account a host of design considerations. These range from the profound question of how phenomena are to be modeled in the data or represented on a physical map, to the more prosaic details of file format (binary or ASCII) or size of the sheet upon which the map will be printed. Although such items are of great importance in the final quality of the cartographic product, they are not within the primary scope of this research. Of principal interest here is the actual creation of the information (content) rather than its format and design.

A Note on Terminology

In this document various terms will be used rather imprecisely. Although fine distinctions can be, and often are, drawn between the terms data and information, here they will be treated as synonyms. Data and information shall denote collections of facts, observations, knowledge, ideas, signals, etc. that are commonly recognized as such. So, for example, though it may be rare to speak of a map or a newspaper as data, it is apparent that a map or newspaper has information. The information content in map or newspaper is universally recognized and acknowledged, if seldom precisely defined and identified in formal terms. No particular distinction is intended either, among data and information in computerized (digital), hardcopy (analog), or graphic form.
The primary subject of this research is obviously the data and information found in maps and GIS databases. And again, adjectives for this kind of information--geographic, spatial, map, and cartographic--will be used interchangeably. These terms certainly do not have identical dictionary meanings and each poses some problem with its use. Spatial data might be the most inclusive term, encompassing everything from a diagram of a DNA molecule to a chart of the entire universe, anything that pertains to space in some way. Although the present research may be applicable to more general forms of spatial information, it is intended to apply to the kind of data and information in conventional topographic maps. The term geographic information is not so general in scope as spatial data, but carries a strong earth-based connotation that may be unduly restricting. Why not include maps of the moon and other planets? Cartographic and map information may too narrowly imply traditional hardcopy map documents to the exclusion of the data now found in computerized geographic information systems. For these reasons then, all of these terms will be employed to denote the information typically contained in maps and digital GIS databases. Finally, the term mapping will be used to refer to any endeavor engaged in the production of either hardcopy documents or digital geographic information.
Outline of Presentation

The balance of the present chapter will set forth the basic ideas of quality and quality control. The brief primer on quality will provide a short history of the present quality movement, emphasizing the development of quality in Japan. Various other factors influencing quality are then discussed along with the major reason for embracing quality improvement: Quality makes bottom-line, economic sense. The final section of Chapter I sets forth a formal definition of product quality and presents the concepts of quality control, statistical quality control, and quality measurement.

The second chapter relates quality issues to the field of cartography. Its purpose is to demonstrate the importance of quality in the mapping domain. Many of the factors driving quality improvement in the overall economy are significant in mapping as well, but in the United States the public sector plays a major role in determining the quality of spatial data.

In the third chapter, a definition of cartographic data quality is developed. Several geographic information standards are reviewed from the perspective of how they define quality. This view of quality is then related to the definition accepted in the quality domain that was presented in Chapter I. The purpose of this chapter is to develop a general characterization of cartographic data quality.
A theoretical foundation for the control of spatial data quality is developed in Chapter IV. The nature of cartographic information is analyzed by means of the metaphor of natural language. A set of fundamental units of two-dimensional geographic information is presented and placed in the context of the raster and vector data models in use today. In establishing fundamental units of spatial data, the framework for controlling quality in a mapping process is largely complete. In essence, the question of what cartographic data quality is will be resolved.

The remaining part of the framework, namely measuring cartographic data quality, is tackled in Chapters V and VI. Here the problem of measuring and quantifying the geometric quality of geographic information is emphasized. An attempt is made to address inadequacies found in existing spatial data standards. The attribute component of geographic information quality is noted but not considered in depth.

The seventh chapter summarizes the significant aspects of the research by suggesting how the framework for controlling quality in mapping can be applied in the context of producing spatial data. In particular, an attempt will be made to relate the principles of statistical quality control to the geometric quality of geographic information.

A final chapter indicates aspects of the conceptual framework that were not resolved in the present research.
Other research issues that are important to the control of quality in mapping are indicated as well.

Primer on Quality

Brief History of Quality: The Rise of Japan

The roots of the present "quality movement" can be traced back at least as far as the late 1940's in postwar Japan. Indeed, the origins can be found in the pioneering work of Walter Shewhart who invented the control chart in the 1920's (Feigenbaum, 1991, page 394 and Grant and Leavenworth, 1988, page 1). Concern for "quality" in human endeavors has surely been present for a very long time: The pyramids and the Great Wall of China are, after all, still standing.

It is clear that quality is an important reason for Japanese industry’s great success (Ishikawa, 1985, page 5). Statistical quality control in Japan was inaugurated in 1946 when U.S. occupying forces ordered the Japanese telecommunications to begin using quality control (Ishikawa, 1985, page 15). In 1950, W. Edwards Deming went to Japan and presented seminars for the Union of Japanese Scientists and Engineers (JUSE) on the statistical control of quality in manufacturing industries. He taught Japanese engineers the methods and philosophy of Walter Shewhart (Deming, 1986, page 489). But, Deming did not teach the Japanese what quality was, for,
Before that time [1950], the quality of Japanese consumer goods had earned around the world a reputation for being shoddy and cheap. Yet anyone in our [the U.S.] Navy will testify that the Japanese knew what quality is. They simply had not yet bent their efforts toward quality in international trade.

(Deming, 1986, page 486)

For Japan, with no natural resources to speak of, national wealth was (and still is) a matter of successful manufacturing and export (Kondo, 1988, page 35F.2).

At the request of JUSE, another American quality expert, Joseph M. Juran, visited Japan in 1954 (Ishikawa, 1985, page 19 and Deming, 1986, page 489). Juran’s lectures expanded the scope of quality control from narrow applications in manufacturing and inspection to all phases and operations in the organization. The role of the worker in achieving quality was recognized early on in Japan and the quality control circle (QC circle) movement slowly emerged in the early 1960’s. QC circles are a means of focusing companywide quality efforts at the worker level and enhancing the effectiveness of workers in achieving the organization’s goals. A QC circle typically consists of a small group of workers that meets with a supervisor on a voluntary basis to study and tackle job-oriented issues and problems (Kondo, 1988, page 35F.4). By December 1983, there were 1,490,629 members of 173,953 QC circles formally registered with the QC Circle Headquarters (Ishikawa, 1985, page 139). Though participation is voluntary, it is believed that about one half of all workers in Japan
participate in these groups (Montgomery, 1991, page 15). QC circles in Japan have enhanced the creativity and willingness of workers, elevated morale and improved human relations, and reduced manufacturing defects.

According to Ishikawa (Ishikawa, 1985, pages 19-21), the Japanese quality movement evolved from an inspection-based approach (prevent the shipment of defective products) to an emphasis on control (prevent defects before they occur) and lately, to a point where all aspects of the organization, product design, marketing, manufacturing, purchasing, and so on, and all people from the president down, must engage in assuring and controlling quality (total quality control). What this brief history suggests is that the basic organizational and technical knowledge for achieving quality exists. Japan may be presently suffering economic setbacks but the overall economic and market success of the country since World War II is undeniable (particularly in the areas of consumer electronics and automobiles).

Japan’s remarkable quality success had been predicted. In 1967, Juran warned the European Organization for Quality Control that the Japanese would lead the world in quality by the end of the 1980’s because no other country was improving as quickly (Juran, 1988, page 35G.5). American products dominated in the period following World War II. However, by the late 1960’s competition from abroad could no longer be
ignored (Deming, 1986, page 27). The economic pressure exerted in the United States by the quality success of the Japanese and others was great. The "quality revolution" in Japan and the subsequent increase in imports to U.S. markets resulted in declining sales for U.S. industries, loss of jobs in export industries and an imbalance of trade that some experts believe jeopardizes the health of the overall economy (Juran, 1988, page 35G.6).

Of the possible responses to the market success of Japan (joint ventures, trade barriers, or quality improvement), it appears that businesses in the United States and around the world have learned the economic importance of high quality goods and services. Evidence of the increased attention being paid to product quality in the United States can be found in the Malcolm Baldrige National Quality Award. This award has been presented since 1988 under a program administered by the National Institute for Standards and Technology (NIST). It was established by the United States Congress to recognize U.S. companies with successful quality management programs and to raise overall awareness of quality.

Factors Other than Competition

In addition to the economic implications wrought by the quality revolution in Japan, there are other factors compelling quality improvement in the U.S. and around the world. Pressure for raising quality also comes from the
consumer movement, government regulation, the environmental movement, and societal concern for health and safety.

The rise of the consumer movement over the last thirty years has certainly focused attention on the quality of goods and services in the marketplace. As they become more sophisticated, consumers grow less tolerant of defects and demand higher quality in the products they acquire. Individuals such as Ralph Nader together with organizations such as the Consumer Federation of America and Consumers Union (which publishes the product testing magazine Consumer Reports) have been influential in passing consumer legislation and establishing regulatory policy in the United States (Mayer, 1987, pages 41-52, 101-102). It may be a common perception that products today are inferior to their predecessors. And although this common view is not correct (products today are generally a lot better than they were several years ago), consumers are becoming more vocal and demanding, and less forgiving of poor quality (Montgomery, 1991, pages 18-19 and Feigenbaum, 1991, page 30).

And it is not only individual consumers who are becoming more sophisticated in their buying habits. Industrial, commercial and government purchasing skill--focusing on total value--has increased as well (Feigenbaum, 1991, page 28).

Another trend (at least in the United States) has been the emergence of strict liability and growth in damage
awards. Product liability is not a new concept; manufacturers have generally been responsible to compensate for harm caused by their defective products. Relatively recently courts have expanded that responsibility through the application of strict liability (Feigenbaum, 1991, pages 34-37). Two principles comprise strict liability; they are product paternity and (what may be described as) truth in advertising. Under product paternity, the producer and seller assume responsibility for the ultimate use of the product and any harmful effects arising from its use (Montgomery, 1991, page 19). The truth in advertising aspect of strict liability requires that all promotional assertions concerning a product be supportable by valid data (Feigenbaum, 1991, pages 36-37). Strict liability places an expanded burden on producers for the products they put in the marketplace and has probably resulted in increasing financial penalties from damage awards. Improvement of the quality of products is an obvious response to the requirements of strict liability.

Another factor influencing quality arises from government regulation. Requirements for products and services imposed by government, in the form of legislation or regulation, are ubiquitous. In the United States, everything from insurance and banking to food, pharmaceuticals, and health care to child (automobile) safety restraints and the buildings we occupy are regulated
by government, at the local, state and federal levels. Such regulation places requirements (which are often minimal) on the characteristics a product must posses or how it may be used and distributed. Responding to and complying with government regulation is a major concern in many industries and certainly influences the quality of products in the marketplace.

The environmental movement has affected the nature of manufactured items and energy production. Consumers may not only be concerned with the environmental friendliness of the final product, they may be interested in the methods and procedures associated with its production as well. Buyers may be extending their sway from the showroom into the factory and onto the production line itself. Closely related to increasing societal attention to the environment are issues related to the production and use of energy. Energy consumption and environmental friendliness are now significant quality characteristics for many consumer goods (Feigenbaum, 1991, page 29). For instance, the amount of energy (a factor in the total cost) to burn a light bulb over its expected useful life span may be as important as the initial cost, durability and brightness of the bulb itself.

Health and safety concerns associated with product consumption and use are also a force in the marketplace. Warnings and announcements of the dangers associated with a
particular food or food additive, chemical, or pesticide are quite common. Producers must continually respond to consumer desires to avoid (as in the case of high fat content foods) items perceived to be dangerous or unhealthy. On the other hand, consumer demand may suddenly arise (say the recent oat bran craze) for a new product that is believed to be of some particular health advantage.

Examples of many of these factors are evident in the production and distribution of automobiles. Quality characteristics such as anti-lock brakes and airbags, gasoline mileage and emission controls, reliability, durability, and cost reflect most (if not all) of the societal, regulatory and economic forces driving quality improvement today.

**Quality/Efficiency Relationship**

In addition to external market and societal forces, the most compelling reason to improve quality may be that it makes sound economic sense. There is a general belief that quality and efficiency in production are contradictory goals, and that one must give up quality to enhance production efficiency or vice versa (Deming, 1986, page 1). But Deming and Feigenbaum (Deming, 1986, page 2 and Feigenbaum, 1991, pages 46-47) believe that such a view is deeply flawed. They argue that a great deal of the resources—material, human, energy, and machine—is wasted in production systems because it results in defective
output. Some of this output may have to be scrapped, resulting in virtually total loss. For that output which can be reworked or repaired, additional resources must be expended in the effort. Waste accrues as well when products have to be replaced or repaired under warranty. Resources are also spent in the process of testing and rechecking dubious product. Feigenbaum refers to this squandered proportion of productive capacity as a "hidden plant," and estimates that up to 40% of productive capacity is dissipated due to poor quality (Feigenbaum, 1991, page 46).

The simple logic suggested by the quality experts is to convert this hidden plant into productive use (Feigenbaum, 1991, page 47). Improving quality transfers the wasted material, human, machine, and energy resources into useful goods and services (Deming, 1986, page 2). By reducing the waste associated with poor quality production, efficiency is enhanced. According to Feigenbaum, quality improvement is the best return-on-investment opportunity available for manufacturing and service organizations (Feigenbaum, 1991, page xvii). Far from opposites, quality and efficiency in production are, in fact, mutually dependent and compatible objectives.

According to Deming, one reason for the success of Japanese companies was their understanding of the necessary relationship between quality and efficiency. The chain reaction shown in figure 1 became a way of life in Japan.
Figure 1. Deming’s Quality Chain Reaction (Deming, 1986, page 3).

Quality is at the very heart of success for Deming: Lower cost, increased productivity, and superior product. Therefore, improving quality will enhance an organization's efficiency and bottom line profitability.

Conclusion

The various market and social trends, competitive pressures, consumerism, government regulation and so on, are combining to make quality a central concern today in the United States and around the world. This general increase in the significance and awareness of quality issues may be referred to as "the quality movement." However, the various quality experts express skepticism whether the increased attention to quality will actually result in better goods and services. They argue that adopting Japanese techniques
(such as QC circles) or adhering to the European quality standard (the ISO 9000 series) or fulfilling the criteria for the Malcolm Baldrige National Quality Award will not necessarily lead to improved quality (Crosby, 1992, page xvii; Deming, 1986, page 129 and Feigenbaum, 1991, pages xvii-xviii). Actually achieving quality improvement is a difficult problem; it requires comprehensive organization-wide commitment and a great deal of hard work (see Deming’s 14 points, Deming, 1986, pages 23-96).

This section has considered the emergence of the quality movement and established the general significance of quality. Although quality issues often arise in the context of mass production manufacturing, they are no less significant in the nonmanufacturing sector. Organizations, both private and public, engaged in the delivery of services or "software" (for example computer programs or data and information), need to take quality seriously (Juran and Gryna, 1993, pages 6-7 and Deming, 1986, pages 183-247). It is also claimed that many quality techniques, especially statistical methods, are applicable in nonmanufacturing contexts (Grant and Leavenworth, 1991, page 27). If this claim is true, then quality issues should be central concerns in any mapping, surveying or geographic information endeavor. Before the issue of quality in mapping is more closely explored, the concepts of quality and quality control will be presented. In the next section, the
applicability of the ideas to nonmanufacturing situations should be kept in mind.

Quality and Quality Control

Quality Defined

Most quality experts emphasize that quality is ultimately determined by the customer (Deming, 1986, page 5 and Feigenbaum, 1991, page 7). The idea that quality efforts of an organization ought to be focused on the expectations and needs of the consumer leads to a simple characterization of quality: Quality is customer satisfaction (Juran and Gryna, 1993, page 3). In the context of this definition it is important to note that a customer is anyone affected by a product or process and that a product may be a tangible good (such as a car), software (computer program, data or report), or (as in transportation or banking) a service (Juran and Gryna, 1993, page 3). The broad conception of a customer suggests that quality is an issue for public and nonprofit organizations as well as private companies engaged in profit-making enterprises.

A major implication of defining quality as customer satisfaction is that quality is determined externally by the customer. Quality can not be defined (internally) by the producer of the good, software or service. It is important to note that the ultimate realization of a product's quality arises from the interaction of at least three factors. They are: (1) the product itself; (2) the customer and how the
product is used and cared for; and (3) the information and services available to support the customer (Deming, 1986, page 176). The relationships among these factors determine whether the product satisfies the customer. The fact that the producer and user are part of a coherent system suggests that actually achieving quality is a complicated proposition. While it may be helpful to stress the ultimate importance of the customer to an organization, the idea of "customer satisfaction" is rather vague.

A slightly more definite conception of quality has been adopted in a cartographic context.

Quality has various definitions in industrial engineering (Hayes and Romig, 1977), but one accepted definition is 'fitness for use' (Chrisman, 1983). Recently, the US National Committee Digital Cartographic Data Standards Task Force (DCDSTF 1988) has adopted this definition formally for inclusion in a US national standard for exchange of spatial data.

(Chrisman, 1991, page 165)

And indeed, while noting that it has not achieved universal acceptance, Joseph Juran adopts "fitness for use" as a kind of shorthand definition for quality in his *Quality Control Handbook* (Juran, 1988, page 2.8).

Customer satisfaction and fitness for use are useful starting points for considering quality. But, they are somewhat tautological in nature and do not provide a guide for action. A more functional characterization of customer satisfaction (and hence quality) is provided by Joseph Juran.
The word quality has multiple meanings. Two of those meanings dominate the use of the word:

1. Quality consists of those product features which meet the needs of customers and thereby provide product satisfaction.

2. Quality consists of freedom from deficiencies. (Juran, 1988, page 2.2)

Two particular components of quality, product features and freedom from deficiencies, are thus identified.

The product features aspect of quality concerns the characteristics of the product itself. For tangible goods, attributes such as reliability, durability, maintainability, attractiveness, performance, serviceability, ease of use, safety, size, weight (and many others) factor into the overall desirability and utility of the item to the customer. In the case of services and software, aspects such as timeliness, completeness, accuracy, friendliness, aesthetics, safety and so on comprise the product’s features (Feigenbaum, 1991, page 7; Juran and Gryna, 1993, page 4 and Montgomery, 1991, page 16). Whatever the particular context, product features constitute the principal motivation for its acquisition and use. The characteristics and features of a good or service are largely determined during the product’s planning and design phase. For example, the horsepower of an automobile engine, clarity of a TV picture, or nature of service and conveniences in a hotel are set in advance of the actual generation or delivery of the product. For this reason, the product
features component of quality is referred to as quality of design (Juran and Gryna, 1993, page 4).

The extent to which the final product (which may be a good, service, or software) actually meets the criteria and specifications called for in its design is the degree to which the product is free of deficiencies. The deficiency-free aspect of quality reflects the fact that a well designed product may be poorly realized in the processes by which it is created. In the manufacture of an automobile, for example, a "beautiful" design may be marred by a defective paint job. A surly employee may render poor service to the customer despite adequate training and resources. This second aspect of quality concerns the extent to which the product conforms to the (design) requirements. It is therefore referred to as quality of conformance (Juran and Gryna, 1993, page 5).

There are thus two fundamental components of customer satisfaction: Quality of design and quality of conformance. This characterization of quality is useful because it allows product quality to be analyzed from two different perspectives. A product may fail in the marketplace because it doesn’t possess the features desired by customers. On the other hand, a product may be rejected because it is defective or fails to conform to design specifications (as perceived by the customer). Distinctions among various product grades is a matter of quality of design and are
obviously intentional. A luxury model automobile, loaded with amenities, is often viewed as a high quality product. Because it normally includes a greater number of components, has tighter tolerances, requires more labor to assemble and so on it will, by design, cost more to produce than an economy model automobile. However, quality of conformance is significant in the production of both the luxury and low end models. A manufacturing process that turns out automobiles that comply with design requirements (whether high end or low end models) will cost less to produce and hence generate greater profit than a system that yields a large number of nonconforming units that have to be scrapped or reworked.

The two dimensions of quality--quality of design and quality of conformance--provide a basis for action. First, a product must be well designed and possess the proper features if it is to succeed. Second, its design must be well executed in the creation and delivery of the product; it must conform to requirements. In the next section, the concept of quality control will be presented.

Quality Control

It is interesting to note that the two-fold conception of quality is implicit in the work of Walter Shewhart, who wrote over sixty years ago.

Looked at broadly there are at a given time certain human wants to be fulfilled through the fabrication of raw materials into finished
products of different kinds. The first step of the engineer in trying to satisfy these wants is, therefore, that of translating as nearly as possible these wants into the physical characteristics of the thing manufactured to satisfy these wants. In taking this step intuition and judgment play an important role as well as the broad knowledge of the human element involved in the wants of individuals. The second step of the engineer is to set up ways and means of obtaining a product which will differ from the arbitrarily set standards for these quality characteristics by no more than may be left to chance.

(Shewhart, 1931, page 54)

While Shewhart’s description of the engineer’s task is presumably limited to the industrial or manufacturing context, the idea of the "ways and means" for obtaining the product can readily be expanded to include service and software commodities, such as banking and information, as well. Figuring out what the customer wants and designing a product to satisfy those wants is the problem of quality of design. Once the product (whether it is a tangible good, service, or software) is designed, it must be fabricated, produced, and delivered (to the customer) in conformance to design requirements. The engineer’s second step, according to Shewhart, concerns how the product is actually created.

A product-generating process, broadly conceived, consists of all the resources, human and material, and activities that go into creating the final product. In a traditional manufacturing context, the "process" typically consists of tools, machinery, human and managerial resources, source materials, chemicals, energy, procedures,
knowledge, skill and training, and so on. With services or software, the product-generating process may be more dependent on intangible (often human) factors, such as training, education, information, and procedures. In any case, the aggregate of these factors by which the product is created and delivered to the customer forms a production system. And, it is this production system that is regulated under a quality control program. Regulating the processes by which products (whether tangible goods, service or software) are created is the principal object of quality control.

As with the concept of quality there are several different ways to formally define quality control. Quality control may be broadly construed as "... the science of discovering and controlling variations" (Pyzdek, 1989, page vi). The idea is to carefully limit the amount of variation (from requirements) in the output of the product-generating process, thereby ensuring conformance to specifications. Joseph Juran defines the control process in terms of three basic steps: (1) Evaluate actual operating performance, (2) compare actual performance to goals, and (3) act on the difference (Juran, 1988, page 2.6). Whatever formal definition is adopted, the purpose behind quality control is to establish and maintain conformance of the product with design specifications and requirements. The control of quality concerns the conformance dimension of product
quality and focuses on the actual product-generating process. Quality control thus focuses on "how" the product is made and if it is being made correctly. "What" the product is--its performance, appearance, features and so on--is an issue of the design dimension of quality which would be a product planning and development activity.

At this point it will be noted that design and production--as well as other functions within an organization such as purchasing, accounting, customer service and so on--are activities that can not be operationally separated from each other. Indeed, bringing a successful product to the market will surely require cooperation of the marketing and engineering departments in developing a product that can be efficiently fabricated by the manufacturing department. In fact, there are management strategies, known as concurrent engineering or simultaneous engineering, that strongly emphasize cooperation with, and participation by, those affected by design decisions (e. g., purchasing, manufacturing, marketing) in the product development process.

While design-oriented and production-oriented activities can not be completely dissociated, the view of the control function as the process used to meet standards (Juran and Gryna, 1993, page 98) serves to focus attention on product generation and the conformance dimension of
quality. Controlling the quality of a product-generating process can be diagrammed as shown in figure 2. The

Figure 2. Control Feedback Loop (Juran and Gryna, 1993, page 99).

following universal sequence of steps needs to be accomplished in order to control process quality (Juran and Gryna, 1993, page 98):

1. Choosing the control subject: i.e., choosing what we intend to regulate
2. Choosing a unit of measure
3. Setting a goal for the control subject
4. Creating a sensor which can measure the control subject in terms of the unit of measure
5. Measuring actual performance
6. Interpreting the difference between actual performance and the goal
7. Taking action (if any) on the difference.

To control a process and the quality of its output, it is necessary to determine what aspects of the product or process that need to be regulated. In some cases, process parameters (say the temperature of an oven) may be critical,
while in others characteristics of the product itself (say dimension of a part) could be the major concern. In general, there may be many control subjects associated with a single product or process context. Once chosen, control subjects must be defined so they can be expressed and evaluated numerically (units of measure). Specifications (goals) the product or process is expected to meet should be well defined and expressed in the units of measure. A measurement system is required to carry out the comparison of process output (based on the units of measure) with the goal. Actual process output should be continuously monitored with "significant" variation or deviation from specifications triggering corrective action.

There are two fundamental problems in implementing the control sequence presented above. Perhaps the most obvious difficulty is associated with items 1 through 5; they all concern, in some way, the problem of quality measurement. Coming up with a quantitative unit of a product feature and then figuring out how to measure it can be challenging. For instance, consider the problem of characterizing and measuring the taste of a food product, or perhaps wine. The second difficulty arises in step 6, in which process performance is evaluated to see if there is anything going wrong. Interpreting and analyzing process performance is the focus of a body of techniques and approaches known as statistical quality control (SQC). The issue of quality
measurement will be addressed following the discussion of statistical quality control in the next section.

**Statistical Quality Control**

Statistical methods were first applied in the quality control process by Walter Shewhart in the 1920's (Shewhart, 1986 page 4). Recently, there has been a resurgence of attention in the United States to the use of statistical approaches to quality control. The "statistical process control" (or SPC) movement, which began in about 1980, relies on Shewhart's control chart (Juran, 1988, page 6.37). And, indeed, statistical methods play a central role in quality control. To see why requires an understanding of the concept of variation.

It is a fact that all manufacturing processes exhibit variation (Ott and Schilling, 1990, page 5). No matter how closely a production system is supervised, there will be differences in the output from unit to unit. Often reasons for the variation will be obvious as in the case of two or more workers or machines. Different people will generate slightly different output. In other situations, there is no obvious reason for variation in the output. Why a critical dimension of a single worker's output, say the length of a part, is slightly different from one unit to the next, is not readily apparent. Perhaps the pressure exerted by the operator's hand is not perfectly consistent, or the physical forces in the machine change over time due to changes in the
power supply. These examples illustrate the two general causes of variability found in production systems.

These causes were first identified by Shewhart (Deming, 1986, page 310) who recognized the distinction between a constant cause system and one affected by assignable causes (Shewhart, 1931, pages 12-14). A constant cause system is one in which the probability (of one of its characteristics falling within some given interval) is independent of time. A constant cause system is governed by chance; its output exhibits random variation, the magnitude of which (under ideal circumstances) amounts to "background noise." The unit to unit fluctuations in the dimension of a part or the number of errors per day in a data entry operation--to the extent they are stable--are examples of constant cause systems. Such variability is expected and reflects the fact that no process can be perfectly controlled. After all, the pressure of an operator's hand will be subject to some variability and the best clerk will make some mistakes!

Assignable causes of variation give rise to non-random, time-dependent disturbances in the process and its output. For instance, a machine may go "out of adjustment," causing the dimension of the produced part to be unacceptably small. A data entry clerk may get a headache or be faced with an increased workload and commit more than the usual number of mistakes. In these situations some specific reason or cause lies behind the change in the process.
Today the two sources of variation are often known as common (chance, constant) causes and special (assignable) causes (Deming, 1986, page 310 and Montgomery, 1991, page 102). In general, common causes are inherent in the product generating process and may be said to be faults of the system while special causes are due to specific and identifiable events or conditions (Deming, 1986, page 314). These causes are somewhat similar to the concepts from classical error theory to which they can be loosely related. Three kinds of error are usually identified: Random errors, blunders, and systematic errors (Mikhail and Ackerman, 1976, pages 60-71). Common causes are inherent in the production process and may be likened to random errors in so far as they manifest stable, predictable patterns and can not be totally eliminated from the system. Special causes are similar to blunders in their effect; they knock a stable system out of control. Since systematic errors normally have constant and variable components and arise from assignable as well as unassignable causes (Eisenhart, 1969, pages 30, 43), they are related to both assignable and common causes.

A production process that exhibits a stable and predictable pattern of variability is said to be in a state of statistical control (Shewhart, 1931, page 6). Such a system is subject only to common (chance) causes of variation; all special (assignable) causes have been
eliminated. Variability in the output from such a system (assuming it is acceptably small) must be tolerated; in other words, left to chance. The process should be left alone. Adjusting or tampering with a process that is in statistical control will at best increase variability and could cause the system to become unstable (Deming, 1986, pages 327-332). The effect on product quality would, of course, be devastating.

On the other hand, statistical control is an achievement, not a natural state (Deming, 1986, page 322). It results from the identification and removal of special (or assignable) causes from a production system. As production systems function over time, they will change. Sometimes the change will be gradual and occasionally it may be precipitous. Special (or assignable) causes will creep into the system and knock the process out of control. Their presence must be detected and then eliminated from the system in order to re-establish statistical control.

The primary goal of statistical quality control is therefore to maintain product generating processes in a state of statistical control. Special causes are eliminated from the system and common causes of variation are ignored. For this reason the major problem in statistical quality control is variation—recognizing and distinguishing special and common causes. In comparing process output to design specifications (step 6 of the quality control sequence and
the comparison box in the diagram) statistical methods help signal when to take action to fix a process that is, or soon will be, broken (that is, suffer special causes). Conversely, statistical quality control helps prevent unwarranted intervention in a process whose variability is due to common causes and is not otherwise broken. Statistical quality control would thus be an integral part of day to day quality control activities. In this sense then, statistical quality control is at the heart of the quality control process.

For Deming and Shewhart, it would hardly be possible to overstate the significance of statistical quality control (Deming, 1986, pages 340-341 and Shewhart, 1931, page 34). Perhaps the most important benefit is that stability, dependability, and predictability of the process, and hence product quality, flow from a state of statistical control. The number and quality of units a production system can produce (and thus the specifications it can meet) are known; the process has a definable capability (Deming, 1986, page 339). Statistical control is also a necessary precondition for quality improvement (Deming, 1986, page 338). Predictable costs, maximum productivity, just-in-time delivery (Kanban), and reduction in the cost of inspection are among some of the other advantages of statistical control (Shewhart, 1931, page 34 and Deming, 1986, page 341).
Stable, statistically controlled product generating processes are vitally important in achieving quality. The techniques of statistical quality control (SQC) or alternatively, statistical process control (SPC), aid in establishing and maintaining stable production systems and are constitutive in controlling quality. Statistical quality control is the primary tool for interpreting variation and its causes and thus forms the basis for taking corrective action.

An important issue that has been elided in the foregoing explication of statistical quality control is the problem of quality measurement. It is, of course, necessary to establish the control subjects and units of measure before statistical quality control can be contemplated; the product and process characteristics of interest need to be explicitly quantified in order to be amenable to mathematical analysis. However, as noted above, determining exactly what to measure and how to measure it can be daunting.

**Quality Measurement**

The problem of measuring quality is at the core of quality control and, by implication, statistical quality control. Recall that the first five steps in the quality control sequence set forth above involve measurement. Indeed,
Central to the process of quality control is the act of quality measurement: "What gets measured, gets done."

(Juran and Gryna, 1993, page 99)

Measuring quality requires the establishment of units of measure and a method of measurement for product and process goals. Units of measure constitute a numerical way to express the product or process feature (control subject) that is to be regulated. A method of measurement (the "sensor" in the control sequence set forth above) provides information on the control subjects in terms of the units of measure as the process functions. To control quality--to meet product and process specifications (goals)--it is necessary to figure out what to measure and how to measure it.

Reduction of a quality characteristic to a quantitative value establishes a unit of measure, a defined amount of a product or process feature that is expressed in numbers (Juran and Gryna, 1993, page 103). Figuring out what product and process characteristics to measure and transforming them to numbers is often straightforward. For instance, the length and diameter of a spindle may be important quality characteristics whose measurement is readily comprehended at the conceptual level.

(Notwithstanding, developing a method of measurement and actually obtaining values for these control subjects may not be a trivial task.) The mathematical concept of the "roundness" of the spindle is well defined. But, devising a
procedure for measuring the spindle's roundness, even conceptually, may be difficult (Deming, 1986, page 279). For some products and processes, it might be very difficult to conceive a meaningful numerical quantity for a control subject. The taste of a food product, aroma (some might say odor) of perfume, or "accuracy" of a computer program are elusive and hard to quantify.

A method of measurement involves the specification of instruments, materials, and equipment to be used, the procedures to be followed, and the conditions under which it is actually performed (Eisenhart, 1969, page 25). Closely following the work of Shewhart, Churchill Eisenhart suggests measurement (or more precisely the realization of a method of measurement) itself constitutes a production process (akin to mass production in industry) that has to be placed in a state of statistical control (Eisenhart, 1969, pages 22, 25-28). The basic idea is that "observations" which exhibit statistical control (or at least do not indicate of lack of statistical control) may be presumed to represent a probability distribution; that is, constitute a random sample from a population or universe.

At least three aspects of a method of measurement in the industrial quality control context can be identified (Grant and Leavenworth, 1988, pages 376-377). Precision relates to the variability (stability, repeatability) of values obtained under controlled conditions; reproducibility
refers to the variability of the method over time under operational (i.e., less strictly controlled) conditions; and accuracy relates to the absence of bias and conformity to the "true" value. In everyday parlance, the notion of measurement precision is presumed to imply something about reproducibility; that is, the system exhibits stability over time. Strictly speaking, the true value of any physical quantity is unknowable (Deming, 1986, 279-280 and Eisenhart, 1969, page 31). True values are those obtained from a particular measurement process (called a preferred procedure or exemplar method) deemed to be "best" for the purpose (Eisenhart, 1969, pages 20-21). Devising measurement processes (sensors in the control sequence) and placing them in statistical control can be technically challenging endeavors. And, it may be conceptually difficult to establish units of measure for the product and process goals for control subjects. The problems of measuring product and process quality arise from the need for well-defined, quantifiable control subjects and precise, unambiguous methods of measurement.

Deming summarizes these requirements as the need for operational definitions (Deming, 1986, pages 276-296). In order for quality specifications or goals to have precise, consistent and communicable meaning, they must be stated in terms of an operational definition. An operational definition sets forth a specific test, inspection and
sampling procedure, or criterion for deciding whether or not specifications have been achieved (Deming, 1986, pages 276-277). A quality specification without an operational definition is meaningless (Deming, 1986, page 277). Thus operational definitions are the mechanism for stating the goals for the control subjects (quality characteristics), standards, specifications, measurement systems and procedures for deciding conformance to specification in objective, repeatable, unambiguous, and above all meaningful terms.

Conclusion

This section has defined quality as customer satisfaction and identified its two components, quality of design and quality of conformance. Quality of design relates to the features of the product while quality of conformance is the degree to which the product actually meets design specifications. These two aspects of quality can be associated with the basic activities of planning and development (quality of design) and production (quality of conformance.) Quality control is the process of ensuring requirements are satisfied. The principal problems in the control of quality are measuring quality and dealing with variation.

Interpreting the variability of the product generating process and its output is the purpose of a body of approaches known as statistical quality control (SQC), or
statistical process control (SPC). Statistical quality control is at the heart of the control of quality and is the basis for maintaining stable and efficient stable production systems. Determining when process variation is acceptable and when it is not requires distinguishing special and common causes. Common causes (also called chance causes) are inherent in the production process and are random in nature. Special causes (also referred to as assignable causes) are non-random disturbances that can be attributed to specific events and conditions. The techniques of statistical quality control are used to identify and eliminate special causes from the production system. A production system from which special causes have been removed is said to be in statistical control; it is influenced only by common causes which must be ignored.

Along with deciding when the variation indicates a process that requires fixing, the actual measurement of quality is the other major issue in the control of quality. It involves the selection of product or process characteristics (control subjects) that are to be regulated and reducing them to numbers (units of measure). A method of measurement (sensor) that can continuously evaluate the process and product features in terms of the units of measure is also required. Measurement itself is similar to a production process that needs to be placed in a state of statistical control. Specifying units of measure and
establishing a quality measurement system result in operational definitions for product and process requirements.
CHAPTER II
QUALITY ISSUES IN MAPPING

The first chapter explored the emergence of the modern quality movement and set forth the basic principles of quality and quality control. Quality is a major concern today due to the influence of various social, economic and international forces, such as product liability, "consumerism" and foreign competition. Quality has become, in short, a matter of economic survival. This chapter will explore quality issues as they relate to cartographic data and information. Its purpose is to demonstrate the importance of quality in mapping and related fields. Many of the factors affecting quality improvement in general are also present in the cartographic context. There are other factors affecting quality in mapping that stem from the major role of the public sector in creating and using geographic information.

Factors Affecting Mapping Quality

While foreign competition in the actual production of geographic information may not yet be a major issue, it is certainly true that many of the forces influencing product quality in general are also affecting geographic data
quality. It can be argued that the emergence of computerized geographic information systems has increased the number of people who make use of digital spatial information. A basic aspect of the "market" for cartographic products has undergone a fundamental change. Map users now require data in digital rather than hardcopy form (Mapping Science Committee, 1990, page 21-23). As geographic information system (GIS) technology becomes available to a wider scope of users, individual consumers should begin to influence (by virtue of the marketplace) the content and quality of geographic data (Mapping Science Committee, 1994, page 15).

In addition to the new GIS technologies for using spatial data, there are new methods for acquiring it as well. New satellite surveying procedures (stemming from Global Positioning System or GPS technology), remote sensing satellites, and digital imaging and photogrammetric systems are capable of generating vast amounts of digital spatial data. Use of these new technologies for acquiring spatial data should, in principal, lower the cost of producing geographic information. Lower costs, in turn, should thereby encourage greater participation in the business of producing spatial information. Such enhancement in the potential for competition in its generation should result in increased pressure to improve the quality of cartographic information. And even if the new technology does not
dramatically lower the cost to collect spatial data (thereby bolstering competition), it must at least enhance its quality. Otherwise, adoption of new technology would be irrational!

Legal liability is as much a concern in the production and use of spatial information as it is for manufactured goods. Producers of spatial data and information are being held liable for the quality of their products and losses arising from their use (Aronoff, 1989, pages 134-135). It has been suggested that legal liability should be considered in the design of a municipal GIS database and is the primary motivation for a quality assurance program in the data collection phase (Foresman, Garza, Edwards, Kelley, Burgeson and Shalit, 1990, pages 132-133). In the development of standards for the exchange of digital charting data, the International Hydrographic Organization (IHO) has been strongly influenced by legal implications surrounding use of such information (Hogan, Wortman, Roswell and Winn, 1994, page 148). In the development of the Electronic Chart Display and Information Systems (ECDIS), S-57, IHO had to make digital data standards that closely parallel the legal basis of present hard-copy documents. For this reason, the IHO effort encompassed data display (user interface) standards, S-52, Specifications for Chart Content and Display of ECDIS, as well as transfer format standards. The purpose behind the additional standards is to assure users
that the electronic information is essentially the same (a result of the display standard) as that carried on the paper charts (Hogan, Wortman, Roswell and Winn, 1994, page 148).

It is difficult to imagine health, energy or environmental issues becoming a direct factor in the actual collection and distribution of geographic information. However, societal concerns over such issues will surely influence the content and use of spatial data, and hence its quality. For example, information resources concerning wetlands was the focus of a recent Mapping Science Committee study. The study indicated the lack of a consistent wetlands inventory mapping and analysis capability as being at the heart of the public debate on the issue (Mapping Science Committee, 1993, page 164). It is not hard to imagine the controversy over the disposition of toxic and nuclear waste, focusing concern on data quality issues. Public debate over the location of a toxic waste facility could call into question just about every spatial variable that could be mapped (transportation, hypsography and hydrography, demography, epidemiology, geology, soils, vegetation, meteorology, political boundaries, land ownership, and so on).

It is evident that producers of cartographic products are beginning to realize the bottom-line advantages of quality improvement. For instance, it has been noted that the cost of verifying (inspecting) GIS database quality is
small in comparison to the costs to repair it (Kelly and Merryman, 1993, page 66). The linkage between improving the quality of production processes and enhanced productivity has also been recognized (Epner and Parmenter, 1993, page 39); in fact, Deming’s chain reaction concerning quality and efficiency has been cited in an GIS data context (Hawkes, 1992, pages 123-124). By improving the quality of cartographic processes, an organization will actually become more efficient and productive. It is to be doubted that such a view is universally held in the mapping community, for one author notes the expensive and time-consuming aspects of implementing a quality control program in GIS database development (Nugent, 1995, page 527).

In any case, concern for quality and awareness of the modern quality movement is surely present in the cartographic domain. Total quality management (or TQM) is touted as the mechanism by which error free databases can be created without the need for rework in the GIS industry (Montgomery, 1992, page 70-72). To be "one of the first" data conversion companies in the world certified to the ISO 9000 quality standard is a major advertising point for one company (Baymont, 1994, page 55). And, the application of statistical quality control methods to GIS database development has been identified as a research need (Moellering, 1994, page 184). Whether all this attention
being paid to the problems of cartographic data quality will result in better geographic information remains to be seen.

The Public Sector and the National Spatial Data Infrastructure (NSDI)

In the United States, the public sector exercises considerable sway in the production of cartographic information. For instance, an estimate places federal level spending on geographic data activities at $4.4 billion (Mapping Science Committee, 1994, page 15). State level public agencies involved in activities such as environmental protection, agriculture, and transportation also make use of spatial information (Mapping Science Committee, 1994, page 43). Local level government, which includes approximately 80,000 municipal, town, township, school district and other bodies, are major producers of spatial data (Mapping Science Committee, 1994, page 50). The impact of such public sector activities in the mapping sector is similar to government regulation in other areas of the economy. Whether a public agency produces it, or contracts with private sector companies for its production, the format, content and quality of the information will be largely set by the requirements of the government organization.

The recently adopted Spatial Data Transfer Standard (SDTS) was formally approved by the Department of Commerce as Federal Information Processing Standard 173 (FIPS 173) on July 29, 1992 and became effective February 15, 1993
(Wortman, 1992, page 294 and Pegeas, Cascio and Lazaar, 1992, page 278). Federal Information Processing Standards are mandatory for U.S. federal agencies and are available for use by state and local governments and the private sector (Wortman, 1992, page 294). The profound influence of the U.S. SDTS in shaping (if not directly regulating) the transfer of geographic information can be seen in its formal adoption by the Commonwealth of Virginia and (with some changes) Australia (Musselman, 1994, page 175 and Miller and Hume, 1994, pages 176-179).

The major role played by the public sector (at least in the United States) in the creation and use of cartographic information affects quality in other ways. At least for the geographic data created by or for public agencies, there are essentially no free market mechanisms in place to effectively regulate their production. A public agency that generates spatial data is, of course, concerned about their quality, especially with respect to fulfilling its mandate. However, it does not have to respond to the needs of external customers the way an automobile manufacturer must in order to stay in business. A company that fails to provide customer satisfaction—that is, fails with respect to quality of design or quality of conformance—will be penalized in the marketplace. Quality is undoubtedly a major factor in the free market for consumer goods and services and competitive forces certainly influence product
quality. But, public agencies generally collect geographic data to fulfill their own mandates and are largely insulated from the need to satisfy an external customer in order to survive. In other words, the specter of competition in the production of this public information, say from Japan, does not exist.

It can be argued that the definition of cartographic data quality has been thus left largely to the public sector organizations responsible to collect data in support of their mandates. The data they produce may or may not be well suited to the needs of a broader community of map users due to their format, file structure, content, accessibility, accuracy and so on. Therefore, in the absence of a viable market, how can society decide how much quality should be built into its geographic information resources? What quality of design features (accuracy, format, timeliness, etc.) should geographic data possess? The basic problem is determining the "value" of spatial data to society. Or, stated another way, is it possible to realistically quantify the "loss" accruing to society due to inappropriate or deficient spatial information? These issues involve translating economic, "bottom line" factors into technical quality decisions without the benefit of market mechanisms.

A recent series of publications by the Mapping Science Committee of the National Resource Council (Mapping Science Committee, 1990, 1993, 1994) represents a deliberate attempt
to deal with precisely these concerns in the United States. These documents attempt to guide development of a coherent policy concerning geographic information resources, collectively known as the National Spatial Data Infrastructure, or NSDI. The NSDI initiative in the United States is a response to the changing technologies for producing and using spatial data. As a consequence of this technological revolution there have been profound changes in the nature and requirements associated with geographic information resources. As it has been remarked, "GIS is not simply a neutral technology change; it creates an imperative for and may indeed require social and institutional change as well" (Mapping Science Committee, 1990, page 20).

The NSDI is broadly defined as the sum total of available geographic information together with the human and technological resources for its production and distribution (Mapping Science Committee, 1994, page 7). The goal of the NSDI initiative is to purposefully manage the nation's spatial data resources to avoid (unnecessary) waste and realize the maximum benefit from public investment in geographic information resources (Mapping Science Committee, 1994, pages 5-7).

Of course, cartographic data quality is explicitly recognized as one of the cornerstones in the development of an enhanced NSDI (Mapping Science Committee, 1994, pages 20-21, 27). But, the NSDI model involves institutional changes
that will have a significant impact on geographic information quality. In the past public agencies in the U.S. have generally played a direct role in the production of cartographic information. But, this situation is likely to change. For example, regarding the National Mapping Division of the United States Geological Survey (USGS/NMD), it was noted,

... The most important function of the USGS/NMD in the future might be not to produce maps or even digital data, but to act as the interdepartmental administrator of the national geographic data infrastructure.

(Mapping Science Committee, 1990, page 23)

In other words, federal agencies (such as the USGS) would relinquish some of their responsibility and control over the generation of geographic information. Instead, they would administer and coordinate the collection and distribution of data under a partnership model. Under the data sharing program envisioned by the Mapping Science Committee, a federal agency may assume responsibility for data standards and quality control while state and local governments (and in all likelihood private companies too) will take over the actual data collection and maintenance functions (Mapping Science Committee, 1994, page 17 and 1993, pages 98-99). A major incentive to form data sharing partnerships is the idea that data producers would be assured their product meets national standards and had undergone independent QA/QC analysis by a responsible agency (Mapping Science Committee, 1993, page 99).
The idea that quality control be performed by an agency that neither owns nor controls the processes actually creating the spatial data presents a technical challenge. In so far as quality must be built into the product when it is created (it can not be subsequently inspected or "assured" in)--it would be difficult for a third party to actually control quality. This fact is acknowledged in the stewardship principle articulated by the Mapping Science Committee--the original producer is best able to maintain and assure quality (Mapping Science Committee, 1994, page 21). How responsibility for the quality of data in the NSDI ultimately gets apportioned between producing or coordinating agencies presents technical as well as institutional challenges.

One obvious technical issue confronting the "assuring agency" is developing techniques for evaluating the quality of conformance of geographic information (to requirements) solely on the basis of the end data product itself. In other words, how can the quality of the received cartographic information be determined by examining the data itself? To be sure, its logical consistency and conformance to format requirements can be established. But, the positional accuracy of a feature can not be checked without recourse to external (and ideally independent) sources of information. Another technical concern is the establishment of NSDI standards that are to be followed. According to the
Mapping Science Committee, the NSDI scheme depends on the establishment of standardized data structures, file formats, content specifications, accuracy requirements and metadata models (Mapping Science Committee, 1994, pages 22-23; 1993, pages 100-102, 115 and 1990, pages 46-47). While policing adherence to standards is a matter of quality of conformance, developing the standards (content, metadata, format and so on) is essentially a quality of design issue. Associating standards development and subsequent adherence as quality of design and quality of conformance issues underscores the technical quality issues in the NSDI initiative.

There are institutional quality issues in the NSDI to be resolved as well. Although establishing good data formats and accuracy requirements are challenging technical issues, getting the various producers and users of cartographic information to agree upon and then actually abide by such standards may be even more difficult. Achieving quality spatial data under the NSDI partnership model will depend on the establishment of appropriate vendor/vendee relationships between data producing and coordinating organizations. These relationships are crucial to the assurance of good geographic information because it is very hard for an agency to inspect (or police) quality "into" the data it receives.
Conclusion

This chapter has set forth some of the trends influencing quality in mapping. Market forces and competition should positively influence cartographic quality as technology works to simultaneously increase both the demand for geographic information and the capacity for its creation. Other social factors influencing quality in the economy as a whole, such as legal liability, and environmental and health issues are significant in the mapping context as well. It is also apparent that the mapping and GIS community is aware of the quality movement and the idea that improved quality can actually enhance productivity.

In the United States, the significance of the public sector in producing and distributing spatial information, together with the changes envisioned in the NSDI initiative, present both technical and institutional quality issues. The quality of cartographic information produced by or for a government agency is largely unregulated by market mechanisms and immune from the requirements of customer satisfaction. Establishment of NSDI format, content and accuracy standards are formidable quality of design issues. Similarly, developing techniques for ensuring quality of conformance to NSDI standards represents a difficult quality control problem. Resolution of these concerns may depend more on establishing the proper framework for cooperation
(in standards development) and appropriate vendor/vendee incentives (for quality control), than on technical solutions.
CHAPTER III
GENERAL CHARACTERIZATION OF CARTOGRAPHIC DATA QUALITY

In the previous chapter, various quality issues related to cartographic data were explored. Many of the forces driving quality improvement in general are at work in the mapping domain as well. The present chapter will address the problem of defining just what cartographic data quality is. Its purpose is to set forth a conception of quality that can be used in developing a framework for controlling mapping processes.

In the first section of the present chapter, several significant spatial data standards in the United States will be reviewed. All of the standards to be considered entail a particular view of cartographic quality. The group of standards examined includes the U.S. National Map Accuracy Standards of 1947 (Thompson, 1981, page 104), the ASPRS Accuracy Standards for Large-Scale Maps of 1990 (American Society for Photogrammetry and Remote Sensing, 1990, pages 1068-1070), draft United States National Cartographic Standards for Spatial Accuracy of 1994 (Federal Geographic Data Committee, 1994), the United States Geological Survey Standards for Digital Line Graphs (United States Department
of the Interior, U.S. Geological Survey, National Mapping Division), and finally the Spatial Data Transfer Standard (National Institute of Standards and Technology, 1992). The purpose of these standards and the kind of geographic information to which they apply differ in significant ways.

The U.S. National Map Accuracy Standards (NMAS 1947), the ASPRS Accuracy Standards for Large-Scale Maps (ASPRS 1990), and draft National Cartographic Standards for Spatial Accuracy (NCSSA 1994) set specific accuracy requirements and are largely applicable to graphic, two-dimensional hardcopy map documents. By contrast, the USGS Standards for Digital Line Graphs (USGS-DLG) pertain to the digital data that have been obtained from hardcopy map documents. But the USGS-DLG are similar to the first three standards in that they delineate requirements that are to be satisfied. These four standards are similar in the sense that they are all of a content-compliance type; they set forth specific conditions that are to be met by the geographic information product. The U.S. Spatial Data Transfer Standard (SDTS) is unlike the other four in that it is not compliance related. SDTS is a mechanism for the transfer of spatial data from one format to another; it does not impose requirements on the accuracy or content of the information. Like the USGS-DLG, SDTS is intended for use with digital spatial data rather than hardcopy map documents.
Despite these differences in application and purpose, the five geographic data standards all incorporate some definition of cartographic information quality. The standards will be analyzed from the perspective of the definitions of cartographic information quality they implicitly or explicitly embody. As they are reviewed, an attempt will also be made to present any particular difficulties or limitations that may be uncovered in the various standards.

In the second section of the chapter, the definitions of cartographic quality gleaned from the five standards will be related to the conception of product quality developed in Chapter I. In other words, the quality of a geographic information product will be examined from the point of view of quality of design and quality of conformance. The goal of this effort will be to develop a characterization of spatial data quality that the producer of a cartographic product can use in achieving customer satisfaction. Of particular interest is establishing the basic dimensions of cartographic information quality to use in controlling the quality of a mapping process.

Review of Spatial Data Standards

United States National Map Accuracy Standards

According to Morris Thompson, the notion of standard-accuracy topographic maps became practical in the 1930’s with the adoption of aerial photogrammetric techniques that
allowed large areas to be mapped in a uniform manner (Thompson, 1981, page 102). The process of developing uniform map accuracy standards was initialized in the United States in 1940 by the Bureau of the Budget. Thompson notes the formal standards that were first issued in 1941 contemplated three reference scales, 1:62,500, 1:24,000, and 1:12,000 with horizontal error tolerances of 1/50, 1/40, and 1/30th of an inch, respectively. The standards were amended in 1943 so that only two scale classes, 1:20,000 and smaller, and larger than 1:20,000, with respective horizontal tolerances of 1/50th and 1/30th of an inch, were considered. A third revision, which became effective in 1947, specified, among other things, that maps satisfying the standard shall bear the statement, "This map complies with the National Map Accuracy Standards" (Thompson, 1981, page 104).

The apparent motivation behind the U.S. National Map Accuracy Standards (NMAS 1947) is the "... utmost economy and expedition in producing maps" that satisfy the need for standard maps as well as the specific needs of government agencies. The standard consists of seven sections. The first two deal with horizontal and vertical accuracy and the third with procedures for determining compliance with the standard. Sections four, five and six set forth labeling requirements for maps that comply (section 4) and do not comply with the standard (section 5) and when a map involves
a considerable enlargement of an existing document (section 6). The final section calls for the publication of maps that conform to latitude and longitude boundaries of 3-3/4, 7.5, or 15 minutes extent.

The NMAS 1947 set forth two basic requirements for horizontal accuracy. For maps of publication scales 1:20,000 and larger, not more than 10 percent of test points shall be in error by more than 1/30 inch, measured at publication scale. For all other maps (that is, publication scales smaller than 1:20,000) no more than 10 percent of test points shall be in error by more 1/50th inch, measured at publication scale. The standard applies to well-defined points only, the definition of which constitutes the bulk of the first section (concerning horizontal accuracy). In general a well-defined point is plottable on the map within 1/100th inch and easily visible and recoverable on the ground--monuments or markers, intersections of roads, railroads, etc. that meet at right angles. The NMAS 1947 horizontal error tolerance can be thought of as a circular region of a given radius surrounding the true location of the test point on the map. To comply with the standard, the location obtained from the map should fall within the circle for at least 90 percent of the points tested.

Section 2 of the NMAS 1947 concerns vertical accuracy and applies to contour maps on all publication scales. It requires that not more than 10 percent of elevations checked
shall be in error more than one half the contour interval. When an elevation on the map is checked, it may be moved by as much as the allowable horizontal error tolerance for that scale in order to reduce the vertical error. The NMAS 1947 are silent with respect to exactly what (points, lines) are to be tested for vertical accuracy and how such tests are to be conducted. But, longstanding USGS practice is to compare elevations determined for well-defined points on the map by interpolating from the contours with elevations obtained from field-survey for the same points (Thompson, 1956, page 168). Because of the allowance for a horizontal shift in addition to the allowable elevation error, vertical accuracy can be expressed in the form of the Koppe formula. For example, if the slope angle of terrain at the test point is given by $\alpha$, scale is expressed as a representative fraction (say, 1:24,000), and the map's contour interval is expressed in inches, then the vertical accuracy standard for maps at scales 1:20,000 and smaller is given in equation (1).

$$vertical
tolerance = \frac{1}{2} CI + \frac{1}{scale} \cdot \frac{1}{50} \tan \alpha$$  \hspace{1cm} (1)

To obtain the appropriate Koppe formula for maps at scales 1:20,000 and larger, replace the fraction 1/50 by 1/30. For a 1:24,000 scale, 7.5 minute USGS topographic quadrangle map with a contour interval of 20 feet (240 inches) and a terrain slope of 10 degrees at the test point, the allowable vertical error would be about 17 feet. If the map were at
1:12,000 scale with a 10 foot (120 inch) contour interval, the allowable tolerance would be nearly 11 feet.

Of interest is the fact that horizontal accuracy under the NMAS 1947 is expressed in terms of inches on the map while vertical accuracy is stated in ground units. Also of significance in the standard are the procedures associated with compliance. Determining conformance to the standard is up to the producing agency that determines the extent of testing, and which maps are to be tested. Testing is based on comparing the positions of points on the map with positions determined by surveys of higher accuracy, but specific procedures to follow in checking conformance are lacking. The need to incorporate standardized testing procedures in the standard was noted by Thompson (Thompson, 1956, page 173). A final aspect of the NMAS 1947 to note is the prescription that map boundaries should follow the graticule (as opposed to a map projection grid system) and span 15, 7.5 or 3-3/4 minutes in either latitude or longitude. This requirement represents an early attempt to standardize the presentation of geographic information used and produced by the U.S. Federal government.

It is fairly clear that the definition of map quality implicit in the NMAS 1947 is limited to the horizontal accuracy of well-defined points; even contour elevation accuracy is based on the observations obtained for well-defined points. Very important omissions from the standard
are references to the horizontal (or positional) accuracy of features other than well-defined points and the accuracy of attribute information. Accuracy of linear features (roads, railways, power transmission lines, streams, and so on) and attributes (identifying names, road classifications, symbolization and the like) are not treated in the NMAS 1947. These particular shortcomings were discussed by Thompson nearly 40 years ago (Thompson, 1956, pages 164-166). He argued that map specifications should not contemplate these factors because factual errors (attribute accuracy) are negligible and the cost of determining linear accuracy would be too high.

A final problem with the NMAS 1947 arises when they are applied to the kind of spatial information products being produced today. The NMAS 1947 presuppose a physical map document of a given scale that can be measured. But, a good deal of the geographic information today is collected and used primarily in digital form. For instance, the continuous ground track obtained from a mobile mapping system may be used to develop a street centerline database. The information consists of a sequential list of three-dimensional coordinate locations with respect to an earth-based spatial reference system, say GRS 80 latitude, longitude and ellipsoid height. This kind of digital data is not bound to a particular scale in the way the information contained in a two-dimensional hardcopy map is.
Although such data possess a certain degree of accuracy—and therefore can not be represented on too large a scale map—the discontinuity in the NMAS 1947 requirement makes them difficult to apply. Which of the 1/50" or 1/30" criteria should be used on the ground track data residing in a GIS database is not immediately clear. In principal the horizontal accuracy of the data with respect to the ground could be either 56 feet (based on representation at 1:20,000) or 33 feet (assuming a 1:20,001 scale representation).

**ASPRS Accuracy Standards for Large-Scale Maps**

The American Society for Photogrammetry and Remote Sensing (ASPRS) published accuracy standards for large-scale maps in 1990 (American Society for Photogrammetry and Remote Sensing, 1990, pages 1068-1070). In general, the ASPRS 1990 standards are restricted to maps produced at scales of 1:20,000 and larger. The document asserts that one of its major features is that accuracy is indicated at ground scale as opposed to map scale. As in the NMAS 1947, the first two sections of the ASPRS 1990 contemplate horizontal and vertical accuracy. Allowance is made for various classes (or levels) of horizontal and vertical accuracy under the standard. Accuracy is defined in terms of the rms (root mean square) error for tests made on well-defined points. The horizontal accuracy specification for class 1 maps is presented in the form of a table showing the maximum
allowable rms errors (associated with the either the x or y planimetric coordinate) for various map publication scales. Two such tables, one for feet and one for meters, are provided and can easily be reproduced using simple relationships.

Though not explicitly mentioned in the standard, a consistent relationship between the limiting rms error criteria and map scale can be inferred. The limiting rms planimetric error in feet for a map published at a scale of 1:N (where N is the denominator of the representative fraction) is given in (2).

\[
\text{rms error (feet)} = \frac{0.01N}{12} \tag{2}
\]

The expression for the limiting rms planimetric error in meters is shown in (3).

\[
\text{rms error (meters)} = \frac{0.25N}{1000} \tag{3}
\]

For example, the limiting rms planimetric error for a 1:6,000 scale map for either the x or y coordinate is 5 feet while for a 1:12,000 scale map the maximum rms error is 10 feet. In other words, the horizontal portion of the standard boils down to rms error of 0.25 mm or 0.01 inch, in x and y separately, at map scale.

Provision is also made for map documents of lower accuracy. A class 2 map is one whose maximum rms error is twice that for a class 1 map; and a map whose limiting rms x
or y error is three times that for a class 1 map is designated a class 3 map. Presumably then, a 1:6,000 scale map with a limiting rms error of 10 feet would be considered a class 2 map under the standard. A major difference between the ASPRS 1990 and NMAS 1947 is the former express accuracy for the x and y coordinates separately while the NMAS 1947 treat error as one-dimensional distance quantity. Under the ASPRS 1990, accuracy is expressed in terms of the rms error for the collection test points. In contrast, the NMAS 1947 specify that not more than 10 percent of the test points shall be in error by more than the stated tolerance value.

The vertical accuracy section of the ASPRS 1990 is very similar to the NMAS 1947 requirement in so far as it is tied to the map’s contour interval. For class 1 maps, the maximum permissible rms error for the test points is one-third the contour interval; for spot elevations the limiting rms error is one-sixth the contour interval. Class 2 and 3 maps have respective limiting rms elevation errors twice and three times those for class 1 maps. Vertical accuracy is defined with respect to well-defined points only. But, the standards do not explicitly indicate how elevations of these points are determined from the map. Presumably, map elevations for points may be obtained by interpolation from the contours. In any case, as with vertical accuracy in the NMAS 1947, ASPRS 1990 allow the map position of a test point
to be moved in any direction up to twice the limiting rms error in position. The effect of permitting these shifts in map locations would be to minimize the apparent vertical error at each test point. But, unlike the NMAS 1947, the ASPRS 1990 vertical accuracy tolerance can not be reduced to a Koppe formula because the rms error is an aggregate statistic.

The ASPRS 1990 do differ substantially from the NMAS 1947 requirements with respect to testing and verification of conformance. Section 4 and the majority of the Appendix of the ASPRS 1990 (the bulk of the document) focus on the procedures for determining compliance with the standards. For instance the check surveys are to conform to the Federal Geodetic Control Committee (FGCC) specifications and procedures. A minimum of 20 check points per sheet is required. Their distribution over the map area is considered as well.

The definition of map quality implicit in the ASPRS 1990 is virtually identical to the one gleaned from the NMAS 1947: Map quality is positional accuracy of well-defined points. Shortcomings inherent in the NMAS 1947--failure to consider attribute accuracy and the positional accuracy of linear features--are not overcome in the ASPRS 1990. Though very much concerned with hardcopy map documents, the ASPRS 1990 state that digital spatial data of known ground-scale accuracy can be related to the appropriate map scale for
graphic presentation at a recognized standard. In this sense, the ASPRS 1990 would seem to acknowledge the fundamental difference between digital spatial data and their presentation on a two-dimensional graphic map. The separation of the data from their representation can be accomplished by means of the three accuracy classes established in the standard. For instance, planimetric geographic data may be collected and tested for conformance to a 10 foot rms specification. Under ASPRS 1990, the 10 foot rms data can be treated as class 2 if they are plotted on a map at a scale of 1:6,000 or, they can be plotted on a 1:12,000 scale map and thus be considered as class 1.

Draft National Cartographic Standards for Spatial Accuracy

In January of 1994, the Federal Geographic Data Committee (FGDC) sponsored a public review of the United States National Cartographic Standards for Spatial Accuracy (NCSSA 1994). At that time, it was anticipated the proposed standard would replace the NMAS 1947 altogether. This author is unaware of any further developments concerning revision or formal adoption of the NCSSA 1994. However, they will be considered here because they may represent a change in the definition of geographic information quality. The actual document considered was obtained from the FGDC Secretariat, USGS, 590 National Center, Reston, Virginia 22092.
The NCSSA 1990 requirements for horizontal and vertical accuracy are essentially identical to the ASPRS 1990 specifications. For class 1 maps, the allowable standard error for either the x or y coordinate is 0.25 mm at publication scale. Similarly, for class one maps, the limiting standard error for elevations determined from the contours is one-third the contour interval and one-sixth the contour interval for spot elevations. It should be noted that the NCSSA term, "standard error" is identical to the ASPRS 1990, "rms error." Class 2 maps under the NCSSA 1994 are those whose limiting standard errors are twice that for class 1 maps. Maps whose accuracy levels are neither class 1 nor class 2 do not conform to the standard. As in the NMAS 1947 and ASPRS 1990, accuracy checking is based on errors at well-defined points only. An interesting omission from the NCSSA 1994 is that no allowance is made for shifting the horizontal position of map points for the purpose of determining elevation accuracy. The effect--if indeed the failure to include planimetric displacement of vertical check points in the NCSSA 1994 is intentional--is a substantial tightening of the requirements with respect to both the NMAS 1947 and ASPRS 1990!

Other aspects of the ASPRS 1990 are incorporated in the draft NCSSA specifications. One is the stipulation that at least twenty but not more than fifty well-defined, well-spaced and evenly distributed test points be used for the
accuracy check for a map sheet. Where the draft NCSSA 1994 specifications differ from those of the ASPRS 1990 is in the rigor of testing conformance to the standards. Like the NMAS 1947 and unlike the ASPRS 1990, the NCSSA 1994 are rather vague in their prescription concerning check data; they require testing only against positions and elevations determined by surveys of a higher accuracy. Another feature of the NCSSA 1994 that distinguishes them from both the NMAS 1947 and ASPRS 1990 concerns the testing of a map series. A map series is at least one hundred maps produced using similar equipment, materials and procedures. Determining conformance of a series to the standard requires that at least three percent of the maps (but no fewer than 10 individual sheets) be tested. The class into which 90% of the standard errors of the individually tested maps fall will apply to the entire series. Attention to conformance of a series represents an attempt to say something about the quality of several maps, not all of which are checked. In this sense then, the NCSSA 1994 would seem to be suggesting the need for some sort of statistical quality or process control in the underlying production system.

It is clear that the definition of quality implicit in the draft U.S. National Cartographic Standards for Spatial Data is again: Map quality is positional accuracy of well-defined points. Linear and attribute accuracy are not treated under the draft standard and thus suffer the basic
limitations associated with both the NMAS 1947 and ASPRS 1990. The NCSSA 1994 seem as tied to and limited by the physical hardcopy map document as the NMAS 1947 document. Testing conformance under the NCSSA 1994 is about as "flexible" as under the NMAS 1947 with a few stipulations concerning the number and distribution of check points on the map sheet.

**USGS Standards for Digital Line Graphs**

The United States Geological Survey Standards for Digital Line Graphs (USGS-DLG) consist of three parts. Part 1, General, sets forth the basic characteristics of DLG files, including their purpose, content, topological structuring, and the principles of graph theory upon which the data model is founded. The second section of the standard, Specifications, provides detailed descriptions of the coordinate systems, error characteristics, and data quality, as well as the optional and standard distribution formats. The final section, Part 3, Attribute Coding, specifies how descriptive feature codes are assigned to the node, line and area DLG objects. Except for Part 3, the bulk of the Standards for Digital Line Graphs considered here date from May, 1988; however there have been several changes in Parts 1 and 2 of the standards since then. Very recently, the Attribute Coding portion of the USGS-DLG was completely revised (United States Department of the Interior, Geological Survey, National Mapping Division, 1995).
Digital Line Graph (DLG) data are non-graphic in nature so these standards are genuinely applicable to digital spatial information. Production of DLG files primarily involves the conversion of existing USGS map documents into digital data files. It is noteworthy that Part 3 (concerning the attribute coding of features) represents the largest portion of the standard, amounting to 360 pages. Part 1 is 18 pages long while Part 2 spans 60 pages, 44 of which are appendices. If sheer volume is any indication, it would seem that the attribute component of spatial information is a more "significant" and complex problem than other aspects of the data, such as positional accuracy of spatial objects to name one.

The majority of Part 2 of the USGS-DLG is devoted to the nuts and bolts of the (standard and optional) data distribution formats. It includes sections on error definition as well as data quality. Errors are classified into three types, blunders, systematic, and accidental. Blunders are unacceptably large errors (often exceeding 3 times the standard error or 0.009 inch) and are easily identifiable. Systematic errors occur in a procedure-specific pattern and give rise to bias in the final product. Accidental errors are random in nature and result from the limitations associated with the measurement process. In accordance with general practice, the USGS-DLG assume that random errors conform to a normal distribution. The
quantity used in the USGS-DLG to describe horizontal error is referred to as a standard error statistic and is identical to the approach used in the ASPRS 1990 and NCSSA 1994 documents.

Data quality is rather broadly characterized in the USGS-DLG in terms of five components. These include lineage, positional accuracy, attribute accuracy, completeness and logical consistency. It should be noted that this definition of quality is essentially identical to the one contained in the recently adopted U. S. Spatial Data Transfer Standard (SDTS). For this reason, a detailed discussion of these five elements of data quality will be deferred until the next section when SDTS is considered. Of interest in the USGS-DLG are the actual specifications concerning DLG data.

Under the USGS-DLG, positional error shall be less than or equal to 0.003 inches standard error in both the x and y component directions, relative to the source map that was digitized. Since DLG data are derived from abutting maps, a provision for edge matching features at these boundaries limits the movement to no more 0.010 inches. It is of interest to note that the USGS-DLG positional accuracy specification has changed at least once (United States Department of the Interior, Geological Survey, National Mapping Division, 1992). At one time, the rms positional error of 0.003 inches applied to a single distance quantity.
In other words, the distance between the true location and the DLG position of a test point was considered. Under the present requirement, errors for the x and y coordinates are considered separately. This change is illustrated in figure 3; it parallels the different treatments of positional error

![Diagram](image.jpg)

Figure 3. Change in DLG horizontal positional accuracy requirement.

in the NMAS 1947 and the newer ASPRS 1990 and (proposed) NCSSA 1994 and represents an increase in the allowable tolerance.

Attribute accuracy is addressed in the USGS-DLG. The standards stipulate that all attribute codes of DLG data in the NDCDB (National Digital Cartographic Database) will agree within 98.5 percent to attribute codes set forth in Part 3, Attribute Coding. Just what "agreement within 98.5 percent" means is subject to some interpretation. Does it
mean 98.5 percent of the codes for a single mapping unit (map or separate DLG file) will be correct? Or, that 98.5 percent of all the codes in the entire database will be accurate, or perhaps that the coding associated with each feature will be 98.5 percent correct? However vague the requirement is, the standard does at least acknowledge the significance of something other than the positional accuracy of well-defined points.

Attention is paid in the USGS-DLG to the issue of edge matching, where features can be misaligned with respect to location or may have different attributes. For example, the location of a road at the boundary of map A might not be in agreement with its continuation on map B. On map A, the road may be symbolized as an unimproved dirt road while map B (produced some time afterward) may show the road as paved. Reporting on the status of the relationship of a DLG with its adjoining sheets is the primary function of a series of quality control flags in the file header. Since the DLG is implemented as a two-dimensional directed graph consisting of nodes, lines and areas, it is obvious that files must satisfy concomitant topological constraints. Finally, the USGS-DLG indicate that the DLG for a given category data will contain at least the same level of content and detail shown on the source graphic; in short, the DLG will be complete.
The USGS-DLG do address one of the primary shortcomings in the NMAS 1947, ASPRS 1990, and NCSSA 1994 in that they treat attribute accuracy. It may be assumed that positional accuracy in the standard is intended to apply to all DLG features. However, the computation of the standard error (that describes horizontal accuracy) is based on test points that are well distributed and well defined and have "true" coordinates of higher accuracy than the DLG criteria. So, in this sense, accuracy is only rigorously defined in terms of well-defined points. Mention is made of other aspects of cartographic data quality, such as lineage, completeness, and logical consistency. For these reasons the USGS-DLG provide an expanded definition of geographic information quality when compared to the other three standards reviewed so far.

However, the standards are less than complete in that they fail to explain the methods by which conformance is verified. For example, how many test points should be used in determining positional accuracy? How is completeness measured? And, with respect to attribute accuracy, the 98.5 percent agreement criterion is vague. Thus, a major shortcoming with the USGS-DLG would be the lack of precise operationally defined procedures for evaluating compliance with the standard. Presumably, these details would be set forth in a yet to be published section of the standard, Part 4, Quality Control and Verification.
**Spatial Data Transfer Standard**

Perhaps the most significant of the five standards considered, the U.S. Spatial Data Transfer Standard (SDTS) presents the broadest definition of cartographic data quality. It is, as noted at the opening of the chapter, unlike the other four standards in that it does not set specific quality requirements to be satisfied. The SDTS concerns data format rather than content. And, although quality information must be provided when the standard is used, compliance may be satisfied by reporting that nothing is known about the quality of the dataset (Fegeas, Cascio and Lazaar, 1992, page 281). Adhering to the letter (if not the spirit) of the standard with respect to spatial data quality would thus be easy.

The SDTS, which was formally approved by the Department of Commerce as Federal Information Processing Standard 173 (FIPS 173) on July 29, 1992 and became effective February 15, 1993, required twelve years of research and development (as well as review) and drew on the efforts of the academic, government, and private industry communities (Tom, 1994, pages 136-137; Fegeas, Cascio and Lazaar, 1992, page 279 and Wortman, 1992, page 294). Due to the length of time it took to craft the standard and the variety and number of individuals who participated in the process, it is reasonable (though perhaps not correct) to conclude that the SDTS should embody a consensus of the major participants in
the field of geographic information in the United States with respect to the quality of cartographic information.

The SDTS addresses the issue of quality by means of a required quality report that is a required component in a conforming SDTS transfer. It is interesting to note the specifications for the quality report changed very little from the time the proposed standard was published (Digital Cartographic Data Standards Task Force, 1988) and when it was formally adopted in 1992. Another point of interest (one that was indicated in Chapter I in the discussion concerning the definition of quality) is that the SDTS assumes an explicit definition of quality. It defines quality in terms of practical utility as an essential or distinguishing characteristic necessary for cartographic data to be fit for use (National Institute of Standards and Technology, 1992, Part 1, Section 1.4 and Part 1, Annex E). The primary purpose of the quality report is to provide a user (or customer) with enough information to decide if the data are satisfactory for a given application. Unlike the other four standards considered, SDTS quality reflects a "truth in labeling" approach to quality rather than setting specific requirements. The guiding philosophy in the SDTS is that quality is fitness for use.

The actual specifications for the data quality report are set forth in Part 1, Section 3 of the SDTS and consist of five sections:
Lineage--information about the sources and processing history of the data.

Positional accuracy--correctness of the spatial location of features.

Attribute accuracy--correctness of semantic (non-positional) information ascribed to spatial features.

Logical consistency--validity of relationships (especially topological ones) encoded in the data.

Completeness--mapping and selection rules and exhaustiveness of feature representation in the data.

Temporal information is to be included in each of the five sections. (As noted above, these five components also make up the definition of geographic information quality applicable to DLG data.) Under the SDTS, quality information should be presented in as rigorous and quantitative manner as possible to support the user in evaluating suitability of the data for a particular application. Such quality information may take the form of a description (textual report), quality attributes assigned to spatial features, or quality overlays.

The lineage component of the standard calls for a description of the source material and methods from which the data were derived. Dates are to be reported that reflect the situation on the ground (if available) in preference to the source's publication date. Although not a distinct component of the SDTS quality report (it is mentioned in the lineage section and subsumed in all of them), data currency is a factor that influences the use of
geographic information. Obviously, the age of a cartographic dataset can affect all the other aspects of data quality. For example, physical changes to a roadway (straightening or widening) may undermine the positional and attribute accuracy as well as the consistency (topological) of data that were formerly valid. Similarly, the creation of a new road altogether will affect the completeness dimension of quality. Particular attention is paid in the lineage report to the issue of mathematical and coordinate transformations involved in processing the source into a final data product. Although geodetic control is to be referenced, the standard is silent with respect to the reporting of details on the equipment, procedures, personnel and standards employed in a mapping process. The primary purpose of the lineage requirement is to provide the user with information on how and when the data were created.

Ideally, the lineage module would provide the user with an in-depth knowledge of the processes by which the spatial data were created. A potential user of geographic information might want to consider all sources and methods used in acquiring and processing the data in a transfer. Aerial and ground survey procedures, equipment (GPS, metric camera, photogrammetric instruments), calibration information, and quality assurance methods would, in principle, be helpful in evaluating geographic information generated by the mapping process. While inclusion of such
information in the context of the SDTS may be unrealistic, the goal of providing the user with information on the mapping process and how its quality was controlled is nonetheless valid.

It is interesting that only two of the five quality modules directly mention the concept of data accuracy. Accuracy is defined in the SDTS (Part 1, Section 1.4 and Part 1, Annex E) as the closeness of results of observations, computations or estimates to the true values or the values accepted as being true. Of course, obtaining a true value can often be difficult or even impossible. The SDTS quality report has distinct sections for positional and attribute accuracy and sets forth procedures for their evaluation.

The positional accuracy section of the quality report should reflect the quality of the final product after all transformations, rather than the accuracy of each step in the underlying process. Four methods for expressing positional accuracy are presented in the standard. They are deductive estimate, internal evidence, comparison to source, and (comparison with an) independent source of higher accuracy. A deductive estimate of the accuracy of location is developed using the principles of error propagation. Known errors in source materials and the limitations of each production step can be used to model the error in the final product. Internal evidence takes advantage of repeated or
redundant information to provide an indication of quality (for example, traverse misclosure or residuals from adjustment). Graphic overlay and visual inspection (by means of check plots) form the basis of the assessment based on comparison to source documents. The final and preferred method for testing positional accuracy is to compare the data to an independent source of higher accuracy and precision. Testing is to be done in accordance to the ASPRS 1990 (that is, ASPRS Accuracy Standards for Large Scale Maps). By subsuming the ASPRS 1990, it would seem that positional accuracy in the SDTS is restricted to well-defined points!

Evaluating the accuracy of attributes must reflect the four scales by which they can be measured. Nominal and ordinal scales represent categorical data while interval and ratio scale data can be expressed numerically. For example, a road name (nominal) or classification (ordinal) represent categorical attributes and Fahrenheit temperature (interval) and elevation above sea level (ratio) can be stated in numbers. With interval scales, values can be added and subtracted; values expressed in a ratio scale can be multiplied and divided. Testing the accuracy of interval and ratio scale attributes is accomplished using the techniques for evaluating positional accuracy. Some sort of numerical value for the "closeness" of the attributes to the truth can be provided.
However, there is no numerical measure of closeness to a true value for a categorical attribute (Chrisman, 1991, page 170). Therefore, nominal and ordinal scale attributes are evaluated using a deductive estimate, tests based on independent sampling, or tests based on polygon overlay. The deductive estimate can be no more rigorous than a guess based on experience. Sampling and overlay approaches are based on the idea of a misclassification matrix. In a misclassification matrix, the rows list values as they occur in the data for the selected attributes while the columns depict the "true" attributes. Diagonal elements of the matrix thus represent attributes that are correct. Off-diagonal elements by row are errors of commission--I called him Fred but he wasn’t really Fred. Off-diagonal elements by column are errors of omission--I should have called him Fred, but I didn’t. Simple counts of values are reported in the misclassification matrix when the sampling method is used. The cells of the misclassification matrix obtained from polygon overlay would involve a listing of total areas in each category.

Although the standard does not present the concept in such a general way, the basic idea behind the logical consistency component is satisfying a set of a priori constraints. As a simple example, a person’s age must agree with the attribute for birth date; otherwise, one or both of the values must be wrong. The logical consistency section
of the SDTS quality report sets forth three basic kinds of constraints: tests of valid values, general tests for graphic data, and specific topological tests. Testing valid values is a matter of checking whether a given datum can be found in the master list containing all the permissible values. For example, the string "horse" is not an allowable value for a contour's elevation. General graphic tests address such concerns as line overshoots and undershoots, duplicate line entries, and the presence of intersections where intended. Specific topological tests verify the integrity of the data by means of an automated procedure. In particular, they verify that all chains (lines) intersect at nodes, cycles of chains and nodes are consistent around polygons, and inner rings are correctly embedded within outer polygons.

As defined in the standard, the logical consistency component of the SDTS quality module may be too narrowly conceived. It would seem reasonable to allow and encourage a wider range of constraint checking be applied to the cartographic information. Why not encourage the imposition of more complex constraints that take into account the geographic nature of the data? That water can not flow up hill is an obvious cartographic rule that can be tested. Another extension is the fact that constraints can, in principle, be stochastic as well as deterministic. Deterministic edits may not be as powerful as probabilistic
tests in detecting problems in the data (Naus, 1975, pages 58-59). And indeed, logical consistency checking (well beyond that contemplated in the SDTS but far short of what is possible), together with topological tests, form the basis of the PROSYS quality validation system used by the USGS in the production Digital Line Graph (DLG) data (Bicking, 1994, pages 19-22). Cartographic information is so complex and contains so many redundancies that enhanced constraint checking may itself be an opportunity to improve quality. For this reason it is appropriate to treat logical consistency as the application of a priori constraints in general rather than just those associated with valid values, graphic inspection, and topological fidelity as set forth in the SDTS.

A very important warning concerning logical consistency is warranted at this point. It is perhaps likely, due to their potential for automation, that consistency checks might tend to become the primary (if not only) quality control method applied in the production of geographic information. Certainly, a failed constraint necessarily implies a problem with the data. If two (or more) values disagree, the test will generally not identify which one is wrong (and they may all be bad). A more important issue however, is that passing a set of a priori consistency checks does not guarantee good data, an obvious limitation that is often overlooked (Redman, 1992, page 64).
Therefore, logical consistency represents a necessary (but by no means sufficient) requirement of spatial data quality.

The final component of the SDTS quality report involves completeness. According to the standard, completeness refers to the relevant mapping rules, selection criteria and the relationship between the objects represented and the abstract universe of all such objects. Of specific interest is the exhaustiveness of the set of features. A distinction is made between spatial and taxonomic (attribute) completeness. Testing spatial completeness can be accomplished by means of formal topological consistency checking. Taxonomic completeness testing involves comparing a master list of geocodes to the codes present in the data.

The primary issue with respect to completeness is whether all features present in the terrain (that should be mapped) are in fact captured in the data representation. In short, do the data provide an exhaustive representation of the spatial features? Determining exhaustiveness of the representation—whether the spatial data embody all relevant terrain features—would almost surely require comparison against an independent data source believed to be reliable.

**Definition of Cartographic Information Quality in SDTS**

It is now possible to summarize the fundamental components of spatial data quality envisioned in the SDTS.
Spatial data quality is defined in terms of the following four properties:

- Accuracy of position (geometry) and attributes
- Consistency with respect to a priori (known) constraints
- Completeness of feature and attribute representations

These characteristics generally correspond to the five modules required in the SDTS Data Quality Report, the only difference being the omission of the lineage component. The reason lineage was dropped is because it is not a characteristic of the data; rather it describes the data. A detailed description of how and when a dataset was created can certainly help the user determine its fitness for a particular application. But, lineage information by itself does not constitute a quality characteristic of geographic information; rather it serves as an indicator of quality. Cartographic data are useful because they are accurate, consistent, and complete, not especially because of the manner in which they were created. (Of course, their accuracy, consistency and completeness are direct consequences of the methods by which they were produced.)

Another significant omission in the list is a reference to temporal information and currency. Whether spatial data are up to date is clearly an important factor in determining their fitness for use. But, as noted in Chapter I, a static mapping context has been assumed. Under this assumption,
the mapping process is conceived as a one-time operation, not as an on-going database enterprise. Spatial information is generated once in the mapping process and then, as far as the producer is concerned, abandoned. In this sense then, the issue of currency is largely beyond the control of producer of the information. For, once the date of the materials or ground observations and measurements are set, currency of the resulting data is established as well. There are situations where spatial data must be treated from a dynamic perspective. A tax assessor’s database, for example, must be continually updated as properties are transferred. But when the mapping is performed in a static context, with no intention to maintain it as an on-going database, currency is not a relevant quality characteristic of the data. The issue of data currency will, therefore, not be considered further in this research.

The SDTS definition of data quality is perhaps the most comprehensive one presented in the five spatial information standards considered. Because of its status as a neutral data transfer standard, the SDTS makes complete the break between the underlying data and its representation in a specific graphic or digital form. Attribute accuracy is acknowledged along with the other dimensions of logical consistency and completeness. Indeed, the components of accuracy, completeness and consistency are often listed as quality characteristics of information in general (Redman,
1992, pages 52-57 and Braithwaite, 1994, page 64). However, the SDTS does not present a complete definition of data quality. For instance, it suffers the limitation of the ASPRS 1990 in focusing solely on the positional accuracy of well-defined points. And, the treatment of lineage and logical consistency in the standard could, in principle, be expanded and clarified. Because it is not a prescriptive standard, the SDTS quality paradigm does not provide specific guidance in the actual measurement of quality. The quality report sets forth a reasonable definition of what cartographic data quality is and the things that should be evaluated but makes only the vaguest of suggestions as to how those characteristics can be rigorously quantified.

**Conclusion**

A dominant theme in the first three standards reviewed above (the ones that set forth definite requirements) is that quality is nearly synonymous with positional accuracy. After considering the USGS-DLG and SDTS standards, it is rather obvious that spatial data quality can not be equated with positional accuracy, though tendency to do so may remain a powerful influence in the cartographic community. (As an example, much of the new mapping and positioning technology is touted for its centimeter or millimeter accuracy.) A more comprehensive portrait of geographic information quality emerges from the last two standards considered, the USGS-DLG and especially, the SDTS 1992.
These last two standards are rather distinct in their scope and application. The USGS-DLG contain accuracy, content, and format criteria that are to be satisfied. On the other hand, the SDTS is a neutral data transfer mechanism that does not involve particular content and accuracy requirements.

However, even in the more inclusive standards, many limitations remain. Several problems have been identified in the foregoing discussion and will be summarized here. Perhaps most glaring is the lack of methods for rigorously quantifying the positional accuracy of anything other than well-defined points. If it is true that 80 percent of the information in a digital map consists of lines (McMaster and Shea, 1992, page 71), then the well-defined point approach would ignore the majority of the information on the map! Clearly, some suitable method for quantitatively evaluating the positional accuracy of poorly defined points and linear objects is needed if map quality is to achieve objective status. When attribute accuracy is considered in the standards, very little guidance is provided (other than use of a misclassification matrix) for actually measuring it. Operational definitions of the other dimensions of spatial data quality (logical consistency and completeness) for the purpose of rigorous quality measurement are also lacking. Of course, the latitude and flexibility in testing conformance to the quality (really point accuracy)
requirements diminishes their value. Numbers and distributions of points to be tested in a single dataset as well as the sampling of a few files from a multiple file project ought to be governed by the principle of obtaining useful information on the product's quality.

A series of questions suggested by Deming (Deming, 1986, page 286) can be posed at this point. How, for instance, is one to put a number on the amount of "completeness" in a digital dataset? How might a contract be written to specify that such and such a level of completeness is required? Is the method for testing compliance to the requirement unambiguous? Most important, do the requirements and procedures for testing conformance provide useful product quality information?

**Defining Cartographic Quality from a Different Perspective**

The SDTS quality report is not intended to set forth specific quality requirements. Instead, it is based on the idea of truth in labeling. Its goal is to allow the customer to make a good decision whether a dataset is suitable for a particular application. Under the SDTS approach to information quality, both the producer and the customer bear responsibility. The producer must clearly indicate what (if anything) is known about the accuracy, logical consistency, and completeness of the dataset and the user must decide if it is acceptable (Fegeas, Cascio and
Lazar, 1992, page 281 and Hunter and Williamson, 1990, page 122). A user can not anticipate that the data in the SDTS transfer will be appropriate for the task at hand. And, likewise, the producer is under no obligation (as far as the SDTS is concerned) to provide data of a certain level of quality.

The definition of quality assumed in the SDTS is fitness for use. In essence, every consumer must address the question of whether the product will be fit for the intended use. And, the basis of the SDTS quality module is to help the user make such a determination. In this sense, the perspective in the SDTS is that of the user or consumer of the geographic information product. In this section, however, the perspective will be shifted from the user’s point of view to that of the producer. In defining quality in Chapter I, it was argued that the idea of quality as fitness for use provided an insufficient basis for action to achieve quality in production. The two-fold conception of quality (product features and freedom from deficiencies) allows the producer to distinguish problems arising from the product’s design versus those that occur in its production. Such a distinction between design and production is not apparent in the SDTS or any of the other standards considered. But, treating the design of an information product separately from its production can be helpful to the producer of a geographic information product. From the
producer's point of view, fitness for use is encompassed in the design of the product (in the features it possesses) as well as in the extent to which it is free of defects.

**Cartographic Quality of Design**

Quality of design in spatial information encompasses those activities associated with the planning and design of a cartographic data product. Perhaps the most significant issue is deciding how the geographic data will model the real world. The process involves translating a conceptual view of the world into a more concrete data structure and then implementing it in a particular system-specific file structure. Peuquet identifies four levels of abstraction encountered in the design of a spatial database (Peuquet, 1990, pages 251-255). First there are phenomena as they exist in the real world. These phenomena must then be represented in the form of an abstract data model, which is a human conceptualization of the real world. A data structure is then developed that formally describes the objects and relationships in the (conceptual) data model. The final step involves implementing the data structure in a physical file structure. As an example, a two-dimensional, Euclidean point, line, polygon data model for the world might be translated into a relational (or perhaps an object-oriented or hierarchical) data structure and implemented in an ASCII file structure on a DOS-based computer platform.
Developing such a model of spatial phenomena may well be the most challenging aspect of producing geographic information. The degree to which the conceptual modeling fits the user’s application will affect the ultimate quality of the data.

For example, the decision to represent elevations with contour lines rather than a digital elevation model (DEM) will affect how the data are used. Contours are well-suited for display and visual analysis while DEM’s are efficiently handled in a computerized GIS for such tasks as earthwork computation and slope analysis. A curvilinear feature modeled with a series of connected straight line segments (polygonal line) will result in some loss of geometric fidelity, no matter how precisely or accurately the phenomenon is measured. Modelling fuzzy or imprecise phenomena (soils, for example) with sharply defined polygons might constitute too gross a simplification of reality, thus leading the user to inappropriate conclusions when used. Since no model can be a completely faithful model of reality, it is not possible to develop a single cartographic model that serves all purposes (Peuquet, 1990, page 255). Therefore, how the data model reality is an important component of fitness for use.

There are a host of other less profound issues to be addressed in the design of a cartographic data product. In determining the product’s features, the developer might ask
if the data will carry explicit topological information
and/or be checked for topological consistency? What spatial
reference frame (projection and datum) will be chosen? To
what precision and in what units (meters, feet) will
coordinate values be reported? What features (roads,
highways, trails, footpaths) will be included in the
database? What spatial objects (curves, lines, points,
polygons, etc.) will form the basis of the representation?
Will sections of well-defined curves or a polygonal
approximations be used for curved phenomena? Is the model
two-dimensional or three-dimensional? Is the file format to
be binary or ASCII? The point of this discussion is that
the design of a cartographic data product, from the
conceptual view and feature content, down to the physical
file structure, is an important and distinct facet of its
quality.

Cartographic Quality of Conformance

Once the cartographic data product has been designed,
it must then be "manufactured." In other words, spatial
information has to be collected, processed, and entered into
the data structure envisioned in the design of the product.
Quality of conformance addresses the possibility that a
well-designed cartographic database can be filled up with
bad data. The result of course will be a poor quality
gographic information product. The conformance aspect of
spatial data quality focuses on the extent to which the data
that actually get produced satisfy (design) specifications, that is, making sure defective (or nonconforming) data values are not created.

Of course, what constitutes non-conforming or defective data will depend on the particular mapping process and product being considered. As an example, the requirements may call for topologically "clean" linework (no undershoots, overshoots, or duplicate lines). A dataset containing linework with gaps and overshoots would be non-conforming. Evaluating this kind of defect would be a rather straightforward matter of verifying topological relationships with an automated edit check. But, other aspects of quality of conformance are not so easily tested, especially by means of an inspection step. For example, design specifications may stipulate that all hydrographic features be collected. Inspecting the final product to ensure the completeness requirement has been met may be tantamount to recreating the data. With a photogrammetrically compiled map, for instance, the checker would have to scrutinize the source imagery almost as closely as the original operator to verify that all the features had been captured. In other situations, it may be nearly impossible (short of recreating the data) to perform an independent check of the spatial data, as with the information contained in a fieldworker’s notebook. Achieving quality of conformance thus requires that the
processes for collecting and transforming spatial data into the final geographic information product be carefully monitored and controlled.

In the industrial context, the basic approach for ensuring that product-generating processes satisfy requirements is quality control. Similarly, in the mapping domain quality control can be used to establish and maintain conformance of the cartographic product to design specifications. While quality of design focuses on "what" the product is, quality of conformance addresses "how" the product is made and whether it is being made correctly.

As noted in the discussion on quality control in Chapter I, design and production activities can not be completely dissociated. For example, decisions concerning accuracy requirements can not be made without some idea of the capability of the production system and sources that will be used to generate data values. It would be somewhat frivolous to expect, say, millimeter positional accuracy (with respect to the ground) of data obtained from the conversion of 1:100,000 scale maps into digital form. The final product necessarily inherits the accuracy of the source document and that level has to be tolerated.

But, once the cartographic product has been designed, and the process for its generation engineered, quality is a matter of the extent to which the system conforms to requirements. Thus, it is in the context of ensuring
quality of conformance in mapping processes that the application of quality control can be contemplated.

At this point it is important to note that achieving quality of conformance in mapping cannot be equated with inspecting the final product. As the experts assert, quality cannot be inspected into the product, it must be built into it (Ishikawa, 1985, pages 77-80; Deming, 1986, pages 28-31 and Feigenbaum, page 77). In other words, well produced spatial data must be generated by processes that are not overly dependent on inspection and rework of non-conforming output.

And indeed, many have noted that implementation of quality control in mapping cannot rely on traditional approaches based on checking and rework, massive inspection and final product verification (Nugent, 1995, page 523; Epner and Parmenter, 1993, page 39; Hawkes, 1992, pages 125-126 and Jackson and Woodsford, 1991, page 248). However, the notion of geographic data quality control as a final inspection of the data at the end of the process may persist (Kelly and Merryman, 1993, page 66; Mapping Science Committee, 1993, pages 98-99 and Dominguez, 1994, pages 387-388).

Implementing quality control in a mapping process (and thereby achieving a high degree of quality of conformance) requires the establishment of control subjects (what is regulated) and methods for rigorously measuring quality.
The geographic data must be evaluated at critical stages of the production process to ensure they conform to requirements. Of course, quality control subjects and measurement systems (as well as the formal specifications) must be meaningfully related to the "fitness for use" of the final cartographic information product.

**General Outline of Cartographic Quality**

A general outline of the producer's view of cartographic information product quality can now be proposed. At the most basic level, the quality of a spatial data product consists of its quality of design and its quality of conformance. Poor quality data may result from an inadequate design or from a product generating process that fails to conform to requirements. These two causes of a bad geographic information product are quite different and call for differing approaches to their solution.

A dataset whose design fails to satisfy the customer may, nevertheless, be produced in conformance to requirements. In this case, a solution would require figuring out what the customer wants and redesigning the product. Of course, production systems may need reengineering as a result of changes in the product specifications. On the other hand, a well-conceived product may be unfit for use because it does not meet design requirements. In this situation, the map production system must be fixed so that output conforms to requirements.
Viewing a spatial information product from this two-fold quality perspective allows design and production issues to be unraveled and dealt with independently. Whether a given cartographic product is fit for a particular use depends on its design and if the processes by which it gets produced conform to the design requirements. It will be noted that this approach roughly corresponds to the division of spatial data error into a conceptual (or model based) component and a measurement (or operational) component (Veregin, 1989, pages 4-5 and Maffini, Arno and Bitterlich, 1989, page 56). Error can be attributed to a flawed model of reality or the necessarily imperfect processes for measuring spatial phenomena. Quality of design and quality of conformance are categories that easily encompass such treatments of error. For this reason they provide a more inclusive (and perhaps more useful) point of view from which to consider data quality.

Quality of design concentrates on the features of the geographic information product whereas quality of conformance focuses on the processes by which data values are obtained and transformed into the final product. While the conformance dimension of geographic information quality is the focus of the present research--especially the development of a framework for quality control in cartographic processes--the design aspect is considered for the sake of completeness. Although a general framework for
the design aspect of geographic information quality may be possible, no attempt will be made to develop one in this research.

Design of a geographic information product is clearly a complex proposition. It requires the development of a conceptual view of the world, its transformation into a formal data model and ultimately implementation in a physical file structure. Many specific decisions have to be made, ranging from issues of data format and feature representation to coordinate reference systems. Since the design will be largely determined by the particular application and use of the final product, generalizations about the nature of the process are difficult to make. For this reason, and the fact that the focus of the present research is on the conformance dimension of spatial data quality, the issue of cartographic quality of design will not be considered further.

Quality of conformance (freedom from deficiencies) in geographic information involves individual data values and the processes by which they are created; it does not entail issues normally associated with design, such as format, content, symbolization, and so on. For the most part, the five geographic information standards considered in the preceding section focus on individual data values, rather than product features. The NMAS 1947, ASPRS 1990, and NCSSA 1994 restrict attention to the positional accuracy of well-
defined points. Under these standards data quality is narrowly defined as conformance of location data values to very specific requirements. The USGS-DLG and SDTS standards present a broader conception of spatial data quality, but they too are also largely concerned with individual data values. Indeed, the USGS-DLG set forth requirements inherent in the design of the DLG format and is thus a conformance standard. Because it is a neutral, general purpose transfer vehicle, the SDTS can not address quality of design issues. It may be concluded that the product features (quality of design) component of quality is not a major concern of the standards considered above.

Of the five geographic information standards considered in the preceding section, the SDTS presented the broadest conception of quality. In particular, three spatial data quality characteristics were gleaned from the SDTS. They are accuracy (positional and attribute), logical consistency (with respect to a priori constraints), and completeness (of both feature and attribute representations). These three properties may be taken as the fundamental components that comprise quality of conformance in geographic information. Thus, currency, accuracy, consistency, and completeness represent a general characterization of the conformance dimension of cartographic data quality.
Conclusion

In this section, a characterization of spatial data quality—one that reflects the producer's perspective—has been developed. Cartographic information quality has been expressed in terms of its design (product features) and conformance (freedom from deficiencies) components. Cartographic information quality of conformance concerns the extent to which the individual data values produced in the mapping process conform to design requirements for geometric and attribute accuracy, logical consistency, and completeness.
CHAPTER IV
FRAMESWORK FOR THE CONTROL OF CARTOGRAPHIC DATA QUALITY

In the previous chapter the nature of cartographic data quality was explored. Several existing standards applicable in the United States were considered. Some significant limitations were found in several of these standards. In general, accuracy of descriptive attribute information for spatial features and the positional accuracy for features other than well-defined points are concerns that are poorly addressed in the standards considered. The broadest conception of cartographic information quality is set forth in the Spatial Data Transfer Standard (SDTS). SDTS presents a view of quality that can be essentially distilled into three fundamental components: Accuracy (position and attribute), consistency (with respect to a priori constraints), and completeness.

Cartographic information quality was also considered from the perspective of Juran's characterization of customer satisfaction, that is, as a function of product features (quality of design) and freedom from deficiencies (quality of conformance). The components of cartographic information quality obtained from the SDTS (some of which are present in
the other standards considered) basically encompass the quality of individual data values in terms of their accuracy, consistency, and completeness. In general, the SDTS does not take into account such quality of design issues as the conceptual model of phenomena, feature and content specifications, and the logical and physical formatting of the spatial information.

However, it is useful to differentiate design and production (conformance) aspects of the mapping product when characterizing its quality. Although design and production issues can not be wholly separated, viewing them separately enables the producer to analyze why a product fails to achieve customer satisfaction or fitness for use. A well-conceived product may not meet its intended level of quality (low quality of conformance) or it may be well-produced but not possess the right features (poor quality of design). Since the focus of this research is controlling the quality of spatial data, the quality of design aspect of spatial information will not be considered in any further depth. Quality control is primarily a matter of the conformance dimension of geographic information quality. Once the geographic information product is designed and the mapping process has been engineered, quality is a matter of whether or not the process has the requisite capability and can be maintained in a state of statistical control. Therefore, the focus will be on the problem of controlling the quality
of a mapping process and ensuring that output conforms to design specifications.

The present chapter will attempt to set forth a theoretical framework for the control of quality in mapping processes. Establishing stable, statistically controlled cartographic production systems requires the determination of the quantities that will be regulated in the quality control process. In essence, the issue to be addressed is determining a set of reasonable control subjects for processes that produce geographic information. These control subjects comprise a set of quality primitives for spatial data that can be used in evaluating process and product performance.

The investigation will argue that a linguistic perspective and the metaphor of natural language can form the basis for developing control subjects for cartographic data. It begins by considering the control subjects that might be used in the context of processing English text. An obvious set of primitive elements exists and forms basis for virtually all operations on such information. Control subjects for textual data are readily developed from the set of alphabet characters and can be used in measuring quality. After considering the role and significance of the alphabet characters in an English text processing context, the nature and language of cartographic communication will be explored. An argument is made that cartographic information comprises
a geometric (or positional) component together with descriptive (attribute) information and that maps can be broken down into individual units of information, much like the alphabet characters in text.

The development proceeds by considering the specific set of cartographic alphabet primitives proposed by Ramirez (Ramirez, 1988, pages 58-75). According to Ramirez, at the most primitive level a graphic map consists of four alphabet characters (and their combinations), point, line, curve, and blank space. How these primitive elements can be adapted and applied in the control of quality in mapping processes will then be explored. It will be argued that three spatial objects--the point, line, and space point--form a set of fundamental units of geographic information and constitute a reasonable set of control subjects for cartographic data. How these primitive elements carry the geometric and attribute facets of geographic information and their application to (primarily 2-dimensional) spatial data models are considered as well. As part of this effort, an attempt will be made to establish the applicability of the cartographic primitives to spatial data models in general and their extension to three-dimensional spatial models.

**Control Subjects for English Text**

Almost all would agree that the set of English language expressions represents a practically infinite body of information. The nature of what can be communicated in
English text is virtually unlimited and the semantic and syntactic structures are very complex. But as complex as English text may be, a fairly obvious set of fundamental units for such information exists. The basic elements of English data are the 26 alphabet characters (A-Z) together with the space character, punctuation marks (e.g., ";", "?") and the ten numerical characters (0-9). These building blocks of textual information constitute the basis for the control of processes operating on English text. Evaluation of the accuracy, consistency, or completeness of English text will be predicated on the alphabet characters.

Moreover, the "modeling" of this information in digital form is straightforward. At the conceptual level, English text is almost universally modeled by means of the standard ASCII coding scheme. The ASCII data structure involves a standard (8-bit) byte for each character in the alphabet. The implementation of the ASCII data structure in a physical file structure can be complex and depends on the particular hardware and software employed. It should be noted in any case that the physical file structure and formal data model are beyond the scope of the present research. Text data structures can be more or less complex, but a simple text file containing ASCII characters can adequately represent a page of written English.

To see the role of the alphabet characters in evaluating the conformance of a production system operating
on English text, consider the example of a typing operation. Assume the process involves a person using a computer word processor to "type" a page of handwritten English text. At the very basic level, the design requirement is to transform, in their proper sequence, the physical alphabet symbols (in graphite on the source document) into their ASCII equivalents in a physical storage medium on the machine. An obvious control subject in this production system is the ASCII byte, and the unit of measure a simple count of the number of incorrect values together with the total number of characters typed. Using such a simple scheme it would be quite possible to quantitatively evaluate the accuracy and completeness of the final data product.

A very important consequence of the ability to obtain such measurements is that the typing system can be placed in a state of statistical control. The capability and stability of the process can be known and communicated. Continued monitoring of the process will signal the presence of special causes of variation and the need to take corrective action.

Of course, the simple measurement scheme would leave a great deal to be desired. A fraction of incorrect to correct characters fails to differentiate the significance or meaning of a particular error. For instance, an error of a single byte in the number, "$900.00" that should be "$100.00" would have a precise value of eight hundred
dollars! On the other hand, one incorrect byte that transforms the word "digital" into the word "digetal" has no obvious metric value and may not distort the meaning at all. These examples suggest levels of data accuracy that reflect the meaning in the text as well as its representation in strings of ASCII bytes.

Constraint checking (that is, imposing logical consistency on the output) might be one approach for addressing the meaning of the English text data. For example, a design requirement could stipulate that every word in the file correspond to a word in a standard dictionary. Any non-standard words in the file must be considered as non-conformities, if not defects. Such a spelling checker would operate at the individual word level. With text information other consistency checks are surely possible. For instance, automated grammar checking is a desirable capability that would function at the level of phrases and sentences. A particularly attractive feature for a computerized word processor might be a "Did you really mean that?" function that actually contemplated the meaning of the text and its context. Such robust features would constitute a higher order of quality control subject in so far as they concern the syntactic (spelling and grammar) and semantic (meaning) dimensions of written text. To develop these features (without resorting to emulating human
intelligence) would necessitate a highly sophisticated and complex set of a priori constraints.

Nevertheless, any digital processing of written English text, from the level of the alphabet characters to the syntactic and semantic levels (and even with expert systems or artificial intelligence), is ultimately grounded in the processing of alphabet characters. They constitute the atomic (non-divisible) building blocks of both the written language and its representation in the ASCII data structure. Measurement of the quality of text at the syntactic or semantic level would involve operations on aggregations of the atomic units. A spell checker operates at a rather low syntactical (structural) level of individual words. But, a spell checker performs operations at the individual alphabet character level to ensure integrity at the word level. Grammar checking would function at a higher level, operating on aggregations of words in phrases and sentences. Again, such processing is based on words that are in turn made up of characters. It is reasonable to conclude then that the alphabet characters (or ASCII bytes) would form the basis of establishing control subjects and units of measure in the case of a process or system operating on written text.

The ability to identify a set of fundamental units of information in the case of English text greatly simplifies the development of control subjects and provides an immediate basis for measuring and controlling quality.
Cartographic information is unlike textual information in that it possesses a graphic (or non-textual) component and therefore has no such obvious fundamental units of information. Indeed, in Chrisman's view, a map can not be divided into completely independent pellets of information (Chrisman, 1991, page 165). He believes, in agreement with Robinson and Petchenik, that a map is an indivisible whole (Robinson and Petchenik, 1976, pages 37-40 and Chapter 4). When a graphic map is viewed as an image, it would appear to present a "whole" that can not be subdivided into atomic units of information the way English text can be reduced to individual alphabet characters.

But, to control a mapping process (that is, achieve quality of conformance) requires the establishment of just such individual, measurable units of geographic information, whether on a graphic map or in a digital database. In the next section, the relationship between the map and natural language will be explored in more detail, to see if cartographic information may be conceived in terms of a set of more fundamental units.

The Nature and Language of Cartographic Communication

If Robinson and Petchenik and Chrisman are correct in asserting the "indivisibility" of cartographic information, then the search for fundamental units of geographic data is doomed. In the case of written English text, the building
blocks are the alphabet characters, numerals and a few other symbols (space character and punctuation marks). Since maps and natural languages (such as English) are both means of communicating information, the fact that the English alphabet provides a handy set of control subjects in processing text suggests a similar set of primitive elements may exist in the cartographic domain.

Robinson and Petchenik strongly criticized the idea that maps can be conceived in terms of a discursive language such as English (Robinson and Petchenik, 1976, page 67). They concluded that any differences between discursive (such as English) and presentational (e.g., graphic map) communication forms were so profound that any attempt to apply linguistic concepts (such as grammar) is wasted effort. In contrast, however, others have considered the linguistic characteristics of map communication more favorably (Ramirez, 1988; Head, 1984; Morrison, 1976, pages 92-94 and Ratafski, 1971, pages 137-141). More recently, Andrews has argued that language provides a very useful metaphor for (if not entirely valid scientific hypothesis of) cartographic theory (Andrews, 1990, pages 1-2). Although conceiving a map in terms of such linguistic structures as noun, verb, adjective and so on may not be wholly correct, the map-language metaphor can nonetheless be of practical value. In the context of the present research and the problem of establishing control subjects for
geographic information, identifying the cartographic analogue for the natural language alphabet would be a valuable application of the metaphor.

Before continuing with the linguistic basis for cartographic control subjects, the nature of cartographic information should be more closely analyzed. It is worth noting that most cartographic theorists would agree that geographic information consists of two fundamental aspects. The two components relate to the fundamental elements of reality, specifically, locations and attributes at locations (Robinson, Sale, Morrison and Muehrcke, 1984, pages 4-5). At the most basic level geographic data address the two questions, "What is where?" And, "Where is what?" (Dale and McLaughlin, 1988, page 85). The "where" component refers to locations, and the "what" relates to attributes. These two aspects will be designated as the geometric and attribute dimensions of geographic data. Schlichtmann suggests ways of characterizing the spatial and non-spatial components of cartographic information (Schlichtmann, 1985, page 24). In particular, he indicates that the information on a map can be divided into plan (or geometric) and plan-free (non-geometric or attribute) information. The geometric information on a map is everything that is dependent on location. Attribute information, in contrast, is independent of spatial location. For instance, the geometric information for a lake feature is the location,
expressed in a spatial reference system (say, a closed string of latitude and longitude values), of its shoreline. The name of the lake would be attribute information because it is independent of the geometry (shape and location) of the shoreline.

It should be noted that attribute information is often expressed in a graphic map by geometric means (say, graduated circle symbols indicating city size or line weights representing road classifications). However, according to Schlichtmann, such information could easily be represented by other non-geometric visual variables (color, tone, e.g.) and does not constitute geometric information per se.

This discussion suggests that spatial information, whether graphic or digital, can be split into geometric and attribute components. In general, the geometric facet of geographic data comprises the positional (location) information and spatial character (shape or dispersion, for example) of features. The attribute component of geographic data carries information that explains and describes what the feature is. A significant implication of this depiction of cartographic data is that control subjects for mapping processes must reflect both the geometric and attribute components inherent in geographic information.

The connection between linguistic concepts (particularly the basic units of expression) and the search
for control subjects that reflect the geometric and attribute characteristics of geographic information will now be considered. It may be precisely those units that convey information that should comprise the control subjects for processes that generate cartographic information. Although he notes that both natural languages and map symbolism have minimal expressions, Schlichtmann is not entirely clear about what constitutes the smallest formal units which convey meaning in a map. The smallest "acts of message conveyance" in a map are referred to as localized signs by Schlichtmann and correspond to the sentence in a natural language (Schlichtmann, 1985, page 25). Unlike the map, a written language possesses a lower level of expression that serves to differentiate meanings but that does not by itself contain meaning. Such distinctive units in written natural languages comprise its alphabet. The letters e and a distinguish meaning in the words "bet" and "bat" but do not in themselves hold meaning. According to Schlichtmann, a map has no such distinctive signs (Schlichtmann, 1985, page 26).

According to Schlichtmann's view of map semiotics then, it is not possible to identify purely distinctive units of information analogous to the alphabet of a natural (written) language. The localized sign on the map, whatever it is, will not be so abstract as the alphabet characters of a written natural language. The building blocks of geographic
information, for Schlichtmann, will carry meanings as well as distinguish them. However, it would seem as if Robinson and Petchenik as well as Chrisman are mistaken in their belief that a graphic map constitutes an individualized whole that can not be divided into smaller units of information. Andrews, in agreement with Schlichtmann and Head, would argue that graphic maps do possess minimal units of information that are read sequentially (much like a written natural language) and that maps can be understood from a linguistic perspective (Andrews, 1990, page 2; Schlichtmann, 1985, page 25 and Head, 1984, pages 7-10).

As further indication that graphic maps can be analyzed into smaller units of information, consider the fact they can be generated from digital spatial data. It is clear that the computer data contain sufficient information to reproduce the graphic map. Since computers only operate on discrete units of data, a direct relationship between the units of information in the digital file and the information on the graphic map can be inferred. If a map were incapable of being divided into individual pellets of information, then formal computerized geographic data models could not exist. That the graphic map can be represented in a computer data model strongly suggests that the information it contains is capable of being subdivided into smaller units.
This section has argued that conceiving maps from a linguistic perspective is valid. It has been noted that cartographic data (whether graphic or digital) consist of a geometric component coupled with an attribute (descriptive) aspect. However, the most significant argument made in this section is that individual units of spatial information can be identified on the graphic map. Therefore, it is theoretically possible that control subjects for cartographic processes may be developed from such fundamental units of spatial information. The basic problem is to articulate the basic building blocks in which geometric and attribute aspects of spatial features are represented in cartographic information. An approach for setting forth a cartographic alphabet will be considered in the next section.

**Cartographic Alphabet**

The alphabet characters (A-Z, space character, punctuation, etc.) provide an obvious set of control subjects for information in the form of English text. But, in cartography, unlike language, no obvious alphabet presents itself as the basic units of information that can be regulated by the control process. Developing generic control subjects for use in controlling processes operating on cartographic data rests on the identification of a suitable set of such of objects in the mapping domain. Work by Ramirez suggests that maps do in fact possess fundamental
pellets of information, and in particular, a cartographic alphabet that can identified (Ramirez, 1988 and Ramirez 1993).

Ramirez approached the problem of identifying the minimal units of cartographic expression from the perspective of structural linguistics. (Of course, the compatibility of the linguistic paradigm to mapping was assumed.) Following the work of Noam Chomsky, Ramirez applied the concept of linguistic levels to the presentation of information on a map. The basis of Chomsky's structural language is the idea of linguistic levels. A language incorporates distinct levels that form a hierarchical structure (Ramirez, 1993, pages 98-99). Each particular level of the language has a fixed set of primes (or alphabet) that are concatenated (combined) to form more complex elements of expression at that level. These expressions then become the alphabet (or primes) for the next higher level of the language. To illustrate this concept in written language, letters (primes) are combined to form words. Individual words may then become the primes for more complex expressions (e.g., phrases and sentences) at the next higher level of the language. Thus a language has a hierarchy of levels ranging from the most primitive (lowest) level to the transformational (highest) level in which any grammatical expression of the language can be formed (Ramirez, 1993, page 99).
This structural approach was applied to cartographic representations in order to identify the basic units of information on a map. Attacking the problem from the highest level of cartographic expression (that is, the map document), Ramirez stripped away the composite layers of cartographic information to discover the map's most primitive level. At this fundamental level, the only type of information remaining,

... is the one originated by the TWO LOCAL PLANAR DIMENSIONS. This information cannot be removed, without destroying the two-dimensional display. This information allows the map reader to distinguish between different locations on the plane. This can be considered the most abstract representation of a map, and will be called the SKELETON of a map, the CARTOGRAPHIC SKELETON, or the SKELETON OF INFORMATION.

(Ramirez, 1988, page 58)

Ramirez then identified the actual primes or alphabet characters of the skeletal level of the map.

The PRIMES of CARTOGRAPHIC LEVEL 1 [i.e., the skeleton] are the alphabet signs POINT, LINE, CURVE, and BLANK. They are combined by CONCATENATION to form abstract shapes which represent the local position, path, or limit of the terrain features.

(Ramirez, 1988, page 71)

In this sense a graphic map does have the analogue of an alphabet of a natural language. The primes of cartographic expression are the point, line, curve, and blank space as shown in figure 4.
Expressions of the skeleton level of the map are obtained by combining the four basic primitives (alphabet characters). For example, an element of the skeleton level of a map can be constructed by combining line and curve primitives to form a more complex expression as shown in figure 5. It should be noted that the primes of the skeleton, as well as elements formed from their combination, are defined in a mathematical sense and should not be confused with their realization in a physical map document or graphic display. A point symbol on a graphic map, for example, must occupy a finite area. But, the point primitive has no length or area and is thus similar to a
geometric point (Ramirez, 1993, page 101). The line and curve characters have length but no width and do not occupy area.

In Ramirez' work a very clear distinction is made between the line and curve characters (Ramirez, 1988, page 67). The line character is limited to a straight line segment, defined by precisely two points. Generation of a more complex polygonal line, or string as it is often called, is accomplished by concatenating (combining) line segments. The curve is differentiated from the line (segment) largely because it requires a greater amount of information for its expression. For instance, a portion of the arc of a (planar) circle requires the specification of at least 5 quantities (center, radius, beginning angle, ending angle) where the line segment requires only four values (x and y coordinates at the two endpoints). Also, there are an infinite number of possible plane curves. But, as noted before, the number of curves actually implemented in a GIS is usually limited to a few types (e. g., piecewise polynomial, spline, or parametric function).

Expressing the blank space character, which is of critical importance in map communication, is accomplished by assuming a blank two-dimensional drawing space. Although this blank surface is empty in the sense that it does not contain geometric objects, the blank space primitive is inherent in the fact that the space is structured.
Locational information for the blank spaces is indicated by the particular coordinate reference system. By virtue of the structure of the coordinate system, the drawing space may be conceived as being filled with blank space characters. For example computer aided drafting systems such as Autodesk’s AutoCAD and Intergraph/Bentley MicroStation system begin with a drawing area that preserves blank spaces in places where other characters (points, lines, or curves) or features are not located (Ramirez, 1988, page 91-92). The systems generally have no explicit drafting operation or element for expressing blank space save for the plotting command, "pen up."

It is important to note that the skeleton level provides only geometric information concerning the mapped features. Attribute information is not present at the skeleton level. Attribute meaning is imparted to the map’s skeleton by modifying the appearance of its elements (formed by concatenating the alphabet characters, point, line, curve and blank). This operation, which converts the skeletal representation by means of the visual variables developed by Bertin (size, value, pattern, color, orientation, and shape) is referred to as image construction (Ramirez, 1993, page 103). For instance, the line character is given a certain weight (or thickness) and color to represent a particular type of road feature on the graphic map. In this manner, the abstract skeletal sign is transformed into a viewable
symbol that can carry attribute information as well as geometric information about spatial phenomena.

Other operations are required to generate the final map (which is the highest level of the linguistic structure). Specifically, the coordinate transformation and addition operations are needed as well as the concatenation and image construction functions mentioned above. The collection of these operations, together with a set of cartographic rules and a formal writing mechanism, form a complete cartographic language (Ramirez, 1993, page 102). The set of operations, rules and writing mechanism comprise a cartographic grammar.

While the details of Ramirez’ cartographic language, especially its grammar, are not of primary interest in the present research, the specification of the cartographic alphabet is significant. The cartographic primitives (point, line, curve, and blank) represent a collection of individual units of spatial information that may form the control subjects for mapping processes. It is also important that the alphabet characters derive from the consideration of the graphic map document—suggesting a general set of control subjects is applicable to both the digital and graphic contexts. In the next section, the primitives will be examined to see if they represent a suitable set of quality primitives that can be used in controlling the quality of processes that generate cartographic information.
Cartographic Control Subjects

A fairly obvious basis for specifying and measuring quality in the case of text data, whether graphic or digital, is the set of alphabet characters. The applicability of the ASCII data model to the control of quality in processes operating on English text was illustrated above. Preceding sections have also demonstrated that cartographic information incorporates the fundamental components of geometry and attribute and that it can be broken up into individual units. In particular, an alphabet for the expression of geometric (or positional) information has been presented. In graphic maps, attribute information (meaning) is imparted to the skeleton (constructed from combinations of the point, line, curve and text primitives) by means of Bertin’s visual variables. The basic proposition considered in this section is whether the four alphabet characters set forth by Ramirez can form the basis for control subjects in cartographic processes. It is fairly clear that the four alphabet primes (point, line, curve and blank) are applicable to the graphic map. But, to serve as a basis for cartographic data quality, the primitives must be adaptable to the digital context as well.

It is fairly obvious that the point object is the basis for virtually all direct spatial observation and measurement. Conventional surveying techniques rely on measuring distances, angles, and elevation differences
between points on or near the surface of the earth. Observations of these quantities are typically transformed into point locations for plotting on a map. Even satellite Global Positioning System (GPS) methods rely on the measurement of distances between points (the satellite and the ground receiver) to generate positional information. It is clear also that point objects are used in maps and GIS databases to represent a large class of geographic entities. Oil wells, survey monuments, man-made structures, utility poles, settlements, and even 500 sheep (in the case of a dot map) are just a few examples of punctual phenomena represented by the cartographic point object. The importance of the cartographic point object can be inferred from the spatial data standards considered earlier. As noted in Chapter III, the U.S. National Map Accuracy Standards (NMAS 1947) as well as the more recent ASPRS 1990 and NCSSA 1994 standards essentially define accuracy in terms of well-defined point features.

In graphic maps, point objects are endowed with attribute meaning by means of the visual variables (size, value, pattern, color, orientation, and shape). For example, the sizes of circles (on a graduated circle map) may indicate the population sizes of towns and cities. Shape is also often used to impart meaning to map points. On a USGS topographic map, for example, a triangle represents a horizontal control point, a small solid square
a structure, and an open circle can indicate a well. In the
digital domain, meaning may be assigned to the point object
by linking textual information to a spatial location. For
example, consider the fragment from the Kassler, Colorado
USGS hydrography digital line graph (DLG) file (7.5 Minute,
1:24,000 scale) shown in figure 6. In the DLG format points

```
L 521 190 110 6 6 499455.94 4370164.21 2 1 0
L 522 191 191 6 6 493393.28 4370350.79 2 1 0
L 523 192 192 23 23 491378.29 4370417.91 2 1 0
```

Figure 6. Fragment from USGS 7.5 minute, 1:24,000 digital
line graph file for Kassler, Colorado showing point
information.

are defined as degenerate lines; that is, as line segments
of zero length whose first and last coordinates are
identical. This fact explains the duplication of
coordinates and the appearance of the "L" (which stands for
line) at the beginning of the three records shown in figure
6. Locations of three points are given in eastings and
northings in the Universal Transverse Mercator reference
system for zone 13 (for example, the easting or x coordinate
for point #521 is 499,455.94 meters). All three points
represent nonflowing well features due to the attribute
code, 50 301, that appears on the last line of each record.
The code is "shorthand" for the descriptive textual
information found in the part 3 of the digital line graphs
standards (United States Department of the Interior, Geological Survey, National Mapping Division). This example shows how attribute and geometric information can be carried in digital cartographic data.

The point primitive forms the basis for a good deal of spatial measurement in surveying and related disciplines, especially with respect to its geometric characteristics. Associating the attribute information with the cartographic point, both in the graphic and digital contexts, is a straightforward matter of graphic symbolization or linking textual data to a geometric location. The point primitive is clearly a fundamental unit of cartographic information and thus constitutes an obvious control subject for mapping processes. And, in fact, it is already so employed in the NMAS 1947, ASPRS 1990, and NCSSA 1994 standards.

The distinction made by Ramirez between the line and curve characters is not too significant. In the mathematical sense, the difference between the straight line segment, the piecewise polygonal line, and more complex curves is one of degree rather than kind. The major differences among a segment, a string (polygonal line) and a curve derive from the amount of information needed in their expression. For this reason then, the distinction can be elided without losing general applicability. Henceforth, a line will be construed as any non-intersecting, non-branching planar line, a segment, polygonal line
(concatenation of segments), more complex curve, or a combination of curves and segments.

Lines are the basis for representing linear phenomena on maps and in digital spatial data as well. Power transmission lines, pipelines, railroads, roads, boundaries, and watercourses are just a few examples of spatial features represented by cartographic lines. Even non-linear phenomena can be represented by lines as in the use of contours or hachures to depict topographic relief. Lines also form the basis for representing areal phenomena as well because explicit definition of a region requires the specification of its bounding lines. Photogrammetric mapping methods primarily generate (in addition to point locations) information in the form of lines. For these reasons then, the line element clearly constitutes a fundamental unit of geometric (plan) map information.

Assigning attribute information to the cartographic line is done the same way as it is for points. On a graphic map, the mathematically defined line is modified by the visual variables (size, value, pattern, color, orientation, and shape) to endow it with specific meaning. For instance, a blue color on a USGS topographic map indicates a hydrographic feature, either a shoreline, river bank, or small stream. Pattern, for example dashes, dots and combinations, can be used to differentiate types of small streams (solid for perennial, interrupted for intermittent).
Attribute information is ascribed to the digital line by linking textual information to its geometric definition. Figure 7 is another fragment from the Kassler, Colorado hydrography DLG file and shows how attribute information can be carried in a digital format. This example contains records, each beginning with an "L" for two lines, numbered 26 and 27. In the DLG format lines are defined as a sequence of straight line segments and expressed as an ordered progression of x and y coordinates. In this example, the segments are expressed as a sequence of UTM zone 13 coordinate values. Line 26 in figure 7 consists of 5 points, the first of which has an easting of 492,587.01 meters and a northing of 4,358,795.63 meters. The last point (fifth) in line 26 has easting and northing values of 492,479.74 meters and 4,358,821.31 meters, respectively. Both line strings represent intermittent streams as indicated by the codes 50 412 and 50 610 that are defined in the DLG standards. This example demonstrates one method by which geometric and attribute information for a linear
cartographic feature can be carried in the nongraphic (digital) domain.

The geometric and attribute aspects of the line primitive are well understood and modeled, both in the graphic and digital mapping contexts. If, as suggested earlier, it is true that 80% of the information on a map (digital or analog) is in the form of lines (McMaster and Shea, 1992, page 71), then lines constitute perhaps the most important control subject for processes generating spatial data.

It is certainly not surprising that lines and points would comprise fundamental units of mapping information. And, indeed, with the possible exception of remote sensing devices such as film cameras and video and digital imaging sensor systems, virtually all spatial measurement and observation is in the form of punctual and linear information. While the assertion of the point and line as essential building blocks of cartographic information is hardly surprising or controversial, the role and significance of a blank space character is somewhat novel. Its importance as a fundamental unit of spatial information will be considered next.

The function of the blank space character in the cartographic language developed by Ramirez is largely that of positional place holder for the expression of cartographic features that are articulated by the line and
point primes. According to Ramirez, the blank space does not carry graphic signs, but the location and extent of an empty region is, nonetheless, essential positional information (Ramirez, 1988, pages 68-70). Thus the role of the blank space is largely geometric in nature; the blank space character does not carry representational information.

Many areas on a graphic map carry no explicit information; they are merely background upon which features are portrayed. A street centerline map pays no special attention to the regions between the lines. On a contour map, the areas between contours are not typically attributed, they exist as the implicit result of the contour lines. In these two examples areas are not formally identified, labeled or otherwise assigned explicit meaning and significance. On the other hand, regions may be of major importance on the map, in which case they are typically identified and labeled and meaning is ascribed to them. For example, the extent of a large waterbody and presence of water are indicated (often with blue tint) in a map portraying hydrography. Whether formally identified or not, the areas between the lines are of major cartographic significance. As noted by Robinson and Petchenik,

In cartography, a blank space can contain a great deal of information by experienced inference, such as the "empty" area shown between a meandering stream and its (floodplain!) edge shown by the sharp rise of the valley wall.

(Robinson and Petchenik, 1976, page 40)
The concept of the blank space is significant in cartography and rightly considered fundamental by Ramirez. For this reason then, it should be considered as a basic unit of cartographic information in both the graphic and non-graphic domains. But the issue of what function the blank space performs must be addressed.

Since all regions in a map are significant, it makes sense to explicitly account for them, even if they are not ascribed with specific attribute information. Formal identification of a blank area in a graphic map depends on whether or not the region had specific meaning assigned to it by the cartographer. The blank spaces between contours on a topographic map are truly empty in so far as they do not receive the explicit attention of the mapmaker. On the other hand, a green tint region on a USGS topographic map signifies the area as a woodland. In this situation, the blank area carries specific attribute information. Of course, the woodland area is geometrically defined at the skeletal level by its bounding lines. But, the expression of the particular significance of the empty space, its attribute information, is carried by the visual variable of color applied to the blank space. This discussion suggests that the blank space primitive can fulfill the dual purposes of geometric (positional) place holder and carrier of attribute information in the representation of areal phenomena.
Use of the blank space primitive as geometric (positional) place holder in the graphic map is essentially starting with a blank sheet of paper upon which the point and line elements are to be located. In order to impart attribute meaning to regions on a map, the cartographer modifies the empty space with one or more of Bertin’s visual variables (size, value, pattern, color, orientation, and shape). As suggested above, the use of green tint for woodlands and blue tint for lakes, ponds and wide rivers are specific instances where color and value are used to associate attributes with a blank region on a map. In the case of the swamp symbol on USGS topographic maps, pattern, shape and color are used to impart meaning to the blank space primitive.

In the digital domain, attribute information for areal phenomena (such as woodlands, political subdivisions and water features) can be accomplished by equating the blank space concept to the area point used in some digital spatial models. In the SDTS, the area point (NA) is a zero-dimensional spatial object within an area that usually carries attribute information about that area (National Institute of Standards and Technology, 1992, Part 1, Section 2.3). The idea behind this approach is to use a single point (interior to the area) and link to it all the attribute information associated with the area feature.
To illustrate how the area point is used, consider the following example shown in figure 8. Here, a lake (area A1) is bounded (and hence defined) by its shoreline (lines L2, L3, and L4). In addition, three roads (line L1, L5, and L6) approach the lake. In terms of the digital file, the information might be stored as shown in figure 9.

<table>
<thead>
<tr>
<th>Object ID</th>
<th>Point (X,Y)</th>
<th>Point (X,Y)</th>
<th>Attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>line L1</td>
<td>...</td>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td>line L2</td>
<td>1000</td>
<td>1000</td>
<td>3000</td>
</tr>
<tr>
<td>line L3</td>
<td>3000</td>
<td>4464</td>
<td>5000</td>
</tr>
<tr>
<td>line L4</td>
<td>1000</td>
<td>1000</td>
<td>5000</td>
</tr>
<tr>
<td>line L5</td>
<td>3000</td>
<td>4464</td>
<td>...</td>
</tr>
<tr>
<td>line L6</td>
<td>5000</td>
<td>1000</td>
<td>...</td>
</tr>
<tr>
<td>area A1</td>
<td>3000</td>
<td>2155</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 8. Graphic example showing use of the area point to carry attribute information for an areal feature.

Figure 9. Example showing how the area point can be used in digital information to carry attribute information for an areal feature.
case, the lake attribute is carried by a geometric point object (area A1) and "fills up" the region in which it is located (bounded by L2, L3, and L4). Geometric relationships alone are sufficient to determine the extent of the area to which the lake attribute (area point) object applies. Subsequent topological structuring of the data can be imposed on top of such a digital format and would, in all likelihood, facilitate computer analysis of the objects. However, strictly speaking topological structure is not necessary; the extent of the region affected by the area point is calculable. This example suggests how the area point primitive can function as an explicit surrogate in the digital domain for the blank space primitive (whether formally attributed or not) on the graphic map.

The area point approach is the basis for attributing areal features in the digital line graph format used by the USGS. In this format the record structure for nodes (i.e., point objects that possess topological structure) and area points is the same (United States Department of the Interior, Geological Survey, National Mapping Division, pages 2A-12, 5/88 and 2B-14, 5/88). The fragment from the Kassler, Colorado hydrography DLG in figure 10 is very similar to the example shown in figure 9. The only significant difference stems from the fact that the DLG incorporates formal topological structure. And indeed, the
<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>2</th>
<th>499566.87</th>
<th>4360448.92</th>
<th>24</th>
<th>0</th>
<th>2</th>
<th>0</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8</td>
<td>285</td>
<td>-7</td>
<td>-5</td>
<td>4</td>
<td>591</td>
<td>318</td>
<td>593</td>
<td>345</td>
</tr>
<tr>
<td></td>
<td>597</td>
<td>0</td>
<td>347</td>
<td>0</td>
<td>301</td>
<td>302</td>
<td>0</td>
<td>303</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>421</td>
<td>50</td>
<td>610</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>499549.54</td>
<td>4367106.99</td>
<td>53</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>257</td>
<td>255</td>
<td>254</td>
<td>-138</td>
<td>589</td>
<td>590</td>
<td>-4</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>-13</td>
<td>599</td>
<td>0</td>
<td>315</td>
<td>314</td>
<td>0</td>
<td>317</td>
<td>0</td>
<td>319</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>350</td>
<td>0</td>
<td>351</td>
<td>0</td>
<td>352</td>
<td>54</td>
<td>0</td>
<td>361</td>
</tr>
<tr>
<td></td>
<td>105</td>
<td>0</td>
<td>316</td>
<td>0</td>
<td>679</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>421</td>
<td>50</td>
<td>610</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>4</td>
<td>436492.62</td>
<td>4358630.77</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>287</td>
<td>-12</td>
<td>-10</td>
<td>-9</td>
<td>286</td>
<td>-6</td>
<td>598</td>
<td></td>
</tr>
</tbody>
</table>

Figure 10. Fragment from USGS 7.5 minute, 1:24,000 digital line graph file for Kassler, Colorado showing areal attribute information.

Information in this example is greatly complicated by the fact that the DLG format carries topological (connectivity) information in addition to the geometric and attribute information. Each area record is indicated by an "A" and contains a complete listing of the records for the lines that bound that area. So, for example, the line list for area 2 contains 24 items, 8, 285, 9, -7, -5, ..., 0, 304, 0, 346 (zeros in the listing precede lines that bound island polygons interior to the area). In other words, lines 8, 285, 9, and so on, bound area 2 while lines 304 and 346, among others, bound island polygons that are inside area 2. Conceptually, the attributes for the area record apply to the continuous, unbroken region bounded by the lines in the line list. Areas 2 and 3 represent intermittent lake or pond features due to the codes 50 412 (lake or pond) and 50 610 (intermittent). Area 4 is merely background because it is unattributed; it is cartographically empty. Formal topological structuring in this example is not necessary in
so far as each "area" is geometrically defined by a single UTM Zone 13 coordinate pair that is inside the area to which it applies. Area 2 has easting and northing values of 499566.87 and 4360448.92 meters.

It is interesting to note that the geometric location of the area point (which for area 2 in the example is \(x=499566.87, y=4360448.92\)) is not required in the DLG format; topological relationships alone are sufficient. In fact, according to the DLG format specifications, an arbitrary point, not necessarily geometrically interior to the area it represents, can be used to carry the attribute characteristics for the area (United States Department of the Interior, Geological Survey, 1990, page 4). The reason area points are not required to be interior to the areas they represent is due to the topological structure of the DLG format. Since each area "knows" the identity of its bounding lines and each line "knows" the areas to its left and to its right, there is no need to define the DLG area point with geometric precision.

Use of the area point (along with the geometric requirement that it be uniquely located interior to the region represented) is particularly convenient when areal objects (such as polygons) and topological structure are not included in the formal spatial model. It might be argued that the area point (blank space) primitive is no more fundamental than the polygon. The decision to include a
polygon (areal) object as one of the elements in a spatial data model is arbitrary. In the case of the DLG structure, polygons were not treated as primary elements. In the cartographic language developed by Ramirez, areas per se are not primitive elements. Instead, as in the DLG, they are formed from the skeleton of bounding lines. The fact that areas are not part of the alphabet in Ramirez' cartographic language, and are not necessary for map expression, suggests that the polygon (or other two-dimensional object) is not a fundamental unit or building block of cartographic information.

It might be argued that the blank space or area point primitive is not necessary at all. After all, lines could carry the attribute information for the areas they bound. However, to employ such an approach would almost surely require the dataset to include some formal topological structuring. Consider the example set forth in figure 11 where each of the three lines carry attributes red and green for adjacent areas. Without appeal to additional

<table>
<thead>
<tr>
<th>Object ID</th>
<th>Point (X,Y)</th>
<th>Point (X,Y)</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>line L2</td>
<td>1000 1000</td>
<td>3000 4464</td>
<td>red, green</td>
</tr>
<tr>
<td>line L3</td>
<td>3000 4464</td>
<td>5000 1000</td>
<td>red, green</td>
</tr>
<tr>
<td>line L4</td>
<td>1000 1000</td>
<td>5000 1000</td>
<td>red, green</td>
</tr>
</tbody>
</table>

Figure 11. Example showing the indeterminacy of areal attributes carried on linear objects.
information, it is difficult to decide which region the attribute red applies to. And, even when the information is portrayed graphically as in figure 12 it is not possible to judge whether the region interior to the triangle carries the attribute red or green. Of course, such information might be determined by means of an exhaustive analysis of

![Diagram](image)

Figure 12. Graphic example showing the indeterminacy of areal attributes carried on linear objects.

the entire data set and the use of topological relationships. Formal topological structuring, in particular line-node topology (for directed lines) and left-area, right-area relationships for lines, would clarify this situation.

However, it seems desirable to avoid topological structure when it comes to the development of cartographic control subjects. Topological information exists only implicitly in graphic maps and may or may not be present in
digital cartographic data. Restricting the cartographic primitives to the very simplest, most general objects that do not require formal topological structuring suggests that the area point (derived from Ramirez' blank space primitive) will be chosen over (more complex) methods requiring two-dimensional primitives such as polygons. Therefore, the blank space primitive will be taken as a fundamental building block of cartographic information. Its function, in the form of an area point, is to implicitly represent the geometry of spaces in a map and explicitly carry the attribute (plan-free) information for such areas.

In the foregoing discussion, three fundamental units of cartographic information have been considered. These are the point, line, and blank space primitives. According to Ramirez, these objects comprise the basic signs in the cartographic alphabet (Ramirez, 1993, page 100). Applicability of these primitives to spatial measurement and observation in general and cartographic information (digital and graphic) was also discussed. On the basis of this section, use of Ramirez' alphabet characters in slightly modified form (that is, point, line, and area point) can form the basis for cartographic control subjects both in the digital and analog mapping contexts. In other words, it is these objects, their geometry and attributes, which must be regulated in order to control the quality of a mapping process. The point, line, and area point objects comprise a
set of quality primitives for use in achieving quality of conformance in mapping.

**Applying Control Subjects in Cartographic Models**

Prior sections have emphasized the linguistic nature of cartographic communication and the idea that the information on a map can be divided into individual units that represent the geometric and attribute aspects of geographic phenomena. A particular set of quality primitives was presented as the basis for controlling cartographic data quality. Specifically, the point, line, and area point objects (together with their attribute information) are the fundamental things that need to be measured and regulated to ensure that a mapping process conforms to design requirements.

However, a careful reader will have noted that the development has been largely limited to the case of two-dimensional graphic maps and the general class of spatial data models known as vector models. It remains to be seen whether the cartographic information quality primitives (point, line, and area point) are applicable to the class of geographic data models referred to as raster models. Another potential difficulty related to the cartographic primitives is that they are inherently two-dimensional in nature. In general, true three-dimensional, solid representation is not possible using the point, line, and
area point constructs. Applicability of the quality primitives to cartographic data models will be considered in the present section. Problems associated with three or more dimensional geographic information are addressed in the following section.

The geometric and attribute characteristics of spatial phenomena are generally expressed in one of three basic cartographic models, the graphic map, the vector model, and the raster model. It will be noted here that some writers refer to the vector model as the object model (Haining, 1994, pages 47-48) and the raster model as the tessellation (Peuquet, 1990, page 259) or field model (Haining, 1994, pages 47-48). Although fine distinctions are often made between these terms and the models they represent, for the sake of simplicity and consistency they will be referred to here as the vector and raster models. The most prevalent and well-established method for presenting spatial information is the graphic (or viewable) map. It should be noted that portions of the theoretical work of Robinson and Petchenik, Schlichtmann, and Ramirez alluded to above are limited to viewable maps and map documents. However, computer-based representations of geographic phenomena are becoming important as spatial data are increasingly being collected, stored, and analyzed in digital form. Therefore, the control subjects presented above need to be related to digital methods for expressing cartographic information.
There are two basic ways of representing geographic phenomena in digital form, the vector-based approach and the raster model. These two paradigms in digital cartography are tied to how the world is conceptualized. In vector models, reality is viewed as empty space populated by individual objects while in the raster model the variation of a single geographic variable throughout the space is depicted (Haining, 1994, page 47 and Frank, 1992, page 411). The basic logical unit in the raster approach is the single cell or unit of space to which an attribute is attached (Peuquet, 1990, page 268). Space is typically represented as a two-dimensional matrix of uniform grid cells as shown in figure 13. Each cell represents a small (usually square)

```
4 4 5 6 4 3 2
5 4 4 5 3 2 1
3 2 3 4 3 2 1
4 3 5 6 5 4 3
5 5 7 8 7 6 4
```

Figure 13. Example of the raster data model.

portion of the earth’s surface to which some attribute (say soil class, elevation, or vegetation type) is assigned. Numbers (often integers) are used as codes for the actual attributes that are associated with each cell. Other
regular and irregular tiling (tessellation) schemes, based on triangular and hexagonal cells, are possible but not common.

In the vector model, space is (typically) represented as a two-dimensional surface or plane on which point, line and area (or polygon) objects (as shown in figure 14) are placed. Attribute information for geographic features (say, type of utility pole, road classification, and political subdivision) can be explicitly linked to the spatial objects. An approach for imparting attribute data to geometric entities is the one employed in the USGS DLG structure. This approach was described in the preceding section and illustrates how the vector model can be implemented in a particular data format.

There are, of course, several variations in these two spatial data models and there are numerous different ways in which they can be implemented in formal data and file
structures. For instance, a raster model can be implemented in a computer as a simple matrix or array structure, run-length encoded data, or quad-tree, to mention only a few options. The particular division of space into regular units (tessellation) in the raster approach might be based on square, rectangular, triangular, or hexagonal cells. Or, the tiling scheme might be based on a more irregular division of space. Specification of the attribute values in the raster domain could be expressed in integer or floating point values, or perhaps even ASCII text. Vector models, for example, can incorporate topological structuring so that connectedness and adjacency relationships among the objects are explicitly recorded. When vector models lack such structure they are often referred to as spaghetti models (Frank, 1992, page 414 and Peuquet, 1990, page 260) and are often employed in computer aided drafting (CAD) systems, such as Autodesk's AutoCAD and the Intergraph/Bentley MicroStation package. Another variation in vector models is the class of linear objects included. Lines might be collections of simple straight segments defined by an ordered set of (x,y) points, often referred to as polygonal lines or polylines. Or, a wider class of line based on piecewise circular, polynomial, spline, or parametric representations might be employed.

Now that the two methods for expressing cartographic information in digital form have been set forth, the
suitability of the quality primitives to these models will be considered. As Peuquet remarks, the simplest vector data structure, the spaghetti model, is a direct line-for-line translation of the paper map (Peuquet, 1990, page 260). In other words, the spaghetti vector model is essentially equivalent to the graphic map. Therefore, the quality primitives—the point, line, and area point—developed from Ramirez’ cartographic alphabet clearly apply to vector-based data models. Examples in the preceding section that illustrated how the geometric and attribute information for the basic units of spatial information can be implemented in the digital context were based on the vector model of geographic data. The only major distinction to be made between vector models presently in use and the set of quality primitives is the former are often (especially in GIS systems) topologically structured. Actual vector topological cartographic models tend to be more complex than the (point, line, and area point) primitives because they must encode spatial relationships among features as well as essential geometric and attribute information. In any case, two-dimensional digital vector geographic information (as well as the graphic map) are amenable to expression in terms of the quality primitives.

Raster data can be conceived as well-structured collections of areal units or cells. Each cell has an attribute, say soil or vegetation type, that applies
homogeneously throughout the small (and usually square) portion of the surface of the earth it represents. Although the cell value may represent information about a finite area of the surface, it is reduced to a single quantity or attribute. Variation of the attribute within an individual cell is not possible. In this sense, the raster model has an areal nature and might be seen as a collection of regular (albeit small) polygons. And, certain forms of raster spatial data readily lend themselves to such an interpretation. For example, the individual pixels in information derived from remotely sensed imagery are quite naturally perceived as areal units to which some attribute has been assigned. This fact is evident in the common expression for the ground area viewed by a pixel at a given point in time, instantaneous field of view or IFOV (Campbell, 1987, page 130).

However, the raster model can also be conceived as a discrete approximation to a continuously varying phenomenon. Each cell represents the value of the phenomenon that is recorded at an individual point location in space. The collection of cells is treated essentially as if it were a continuous representation of the phenomenon. In this sense, the raster model emulates a statistical surface (Robinson, Sale, Morrison and Muehrcke, 1984, page 281). Meteorological quantities such as air temperature and barometric pressure are examples of phenomena that are well
suited to modeling as field data. Gridded digital elevation model (DEM) information lends itself to such an interpretation as well. An elevation value in a DEM is conceived as valid only at a single location on the surface, certainly not over the entire extent of a grid cell. Values in the DEM represent locations of discrete geometric points in space and the individual cells carry no areal significance. (In fact, various methods of spatial interpolation have been developed to infer elevations for planimetric locations that are not contained in the DEM.) The continuous raster perspective is necessarily implicit whenever gridded raster data are differentiated to obtain gradient information, for instance. Therefore, the raster model can be seen as a structured collection of individual point objects that carry quantitative (ordinal, interval, or ratio scale) attributes. The discrete point observations are conceived as a finite approximation to the continuously varying characteristics of the phenomenon.

There are thus two different ways in which raster geographic information can be understood. On the one hand, as with the pixels derived from remotely sensed image data, it can be perceived as a collection of small areal units (or polygons). And on the other hand, it can be understood as a finite set of discrete point observations that emulates a continuous statistical surface. One possible reason for these differing interpretations of raster structured spatial
data may arise because the nature of the phenomenon and the characteristics of the raster model are confused.

Soil types, vegetation cover, landcover classifications, and roads are usually modeled as distinct objects in space. In other words, they are treated as if they are phenomena with very clear and abrupt boundaries separating them. Soils are often drawn on a polygon map where each sharply defined region represents a distinct object (soil classification). Roads are also rather sharply demarcated features identified by the pavement edges. When features are conceived as crisply defined objects, it is only natural to treat them as discrete objects with very well-defined boundaries when they are represented in a raster data structure. For instance, the road attribute in a raster model is usually defined in a binary fashion (road and not road), not by means of a continuously varying quantity (0-not road at all, 1-some road, . . . , 8-mostly road, 9-completely road). Since the road object must occupy a finite region of space, the individual grid cell (or pixel) must have an areal extent. So, although the raster model may be suited to interpretation in terms of a continuously varying quantity, the object model adopted for the mapped phenomenon persists.

Data that are essentially continuous in nature, such as temperatures or elevations, are generally not perceived as distinct spatial objects. Instead, they are seen in terms
of a statistical surface whose values vary continuously. The raster model is well suited to these kind of data; the mapped phenomenon and the model are congruent in so far as the raster model emulates a statistical surface. Raster data are properly understood as comprising a discrete approximation for a continuous phenomenon. As such, they consist of a collection of individual point objects.

The reason for considering these different interpretations of raster models and the phenomena represented in them is to suggest a conceptual link between the vector domain (point, line, and area point) and the raster model. If the raster model is conceived as a collection of point objects, then the point object would be the natural choice as the fundamental unit of raster information. This link between the vector and raster perspectives is perhaps most apparent for DEM data. The individual value in the raster DEM corresponds to a spot elevation in a graphic map or vector geographic data file. And, although a DEM can be viewed as a structured set of spot elevations, it is typically applied in the context of a raster model. For raster data that seem to possess an areal extent, as in image data, the point primitive would, at first glance, not seem appropriate. However, the point primitive could, at least in principle, possess the attribute of areal extent. Indeed, in hardcopy map documents, the cartographic point is necessarily so
portrayed. Any graphic point symbol must occupy a finite area. This duality between the geometric characteristics of the point primitive and its potential to "represent" an area presents no special difficulties.

Another reason to adopt the point primitive as the fundamental unit in raster data is the steady march toward higher and higher resolution data from remote sensing devices. The IFOV for the original LANDSAT multispectral scanner (MSS) was about 79x79 meters (Campbell, 1987, page 130). Spatial resolutions for sensors on commercial satellites scheduled for launch before the year 2000 are in the range of 1-5 meters (Brannon, Hill, Davis and Birk, 1994, page 329). A one meter pixel displayed on a 1:10,000 scale map is practically a point for it would be 0.1 mm square. The natural limit, assuming such higher resolution data are obtained, is obviously a point object. If one were to allow the resolution in the raster model to grow arbitrarily large (and thereby shrink the size of the individual cell or pixel) the result would be a dense collection of points. If, as cell sizes got sufficiently small, clusters of points emerged whose attributes were identical, it would be possible to identify clear boundaries between clusters. The cluster boundaries could then be transformed into line primitives and the cluster attributes transferred to area point primitives. In other words, the raster information would be expressed equivalently in the
vector domain. So, the point interpretation for raster data is appropriate in the limiting case (with respect to increasing resolution) in which case it would be compatible with expression by means of the vector primitives.

For these reasons, it may be concluded that the point object comprises the fundamental unit of information in raster spatial data. This result is significant for it tends to provide another conceptual bridge between the vector and raster models. The fundamental units of information in the two approaches are drawn from the same set of primitive cartographic objects (point, line, and area point). Indeed, Peuquet noted that the two digital models are logical duals; entities are the primary units in vector data while space is the fundamental unit in raster data (Peuquet, 1990, page 268). An equally important conclusion is the set of quality primitives (point, line, area point) is adequate for the characterization of cartographic information in the graphic map as well as the raster and vector digital spatial data models in use today. A simple set of control subjects is applicable to the extant methods for modeling spatial data.

Problem of Higher Dimensions of Geographic Information

There is one final issue to consider before the quality primitives can be provisionally accepted. In short, what role do the point, line, and area point primitives play in
the context of true three-dimensional spatial information? Though never formally stated, most of the development so far has tacitly assumed a two-dimensional planar cartographic representation of the world. And although the surface of the earth—or perhaps its reduction to a mathematically defined horizontal datum surface (e.g., ellipsoid)—can be suitably represented (by means of a map projection) in a Euclidean planar framework (either graphic or digital), such a model does not reflect the three-dimensional nature of the world.

After asserting that all geographic data can be reduced to the point, line and area together with a label telling what it is, Burrough notes, in defining a map, there is no reason to exclude higher dimensions save for the problem of representation on a flat piece of paper (Burrough, 1986, page 13). A slightly broader view recognizes four categories of geographical phenomena, place (point), linear, areal, and volumetric (Robinson, Morrison, Sale and Muehrcke, 1984, page 107). An even more general conception of geographic reality would involve three spatial dimensions, a temporal dimension, and a practically unlimited number of attribute dimensions.

However, a general treatment of even true three-dimensional spatial information, encompassing solid or volumetric features, is problematic. Inclusion of the temporal dimension renders the problem even more difficult.
A general solution to the problem of handling topology in three dimensions has never been implemented; adding time only makes the problem harder (Greve, Kelmelis, Fegeas, Guptill and Mouat, 1993, page 1505). And, commenting on gaps in our present knowledge of spatial objects, Moellering notes,

A good understanding of the most efficient ways to specify primitives for 2½-dimensional objects (surfaces) and 3-dimensional objects (solids) remains a limitation of our knowledge. It is clear that much more work on 3-D topology is necessary to help solve the problem. (Moellering, 1994, page 182)

Not only are the topological relationships difficult to fathom in a three or four-dimensional model, but the nature of spatial objects becomes complex as well. It has been demonstrated in this chapter that the point, line and area point comprise a sufficient set of basic building blocks for planar cartographic information. Although the number of ways to define a linear object (and therefore to define a polygonal object as well) is infinite, specific data structures in use are generally limited to a small class of lines. The class of objects typically includes piecewise linear, quadratic, and cubic functions and piecewise circular arcs. In three-dimensional geographic modeling a consensus on the set of surfaces (faces) and solids to use has not been established.

Due to the outstanding problems associated with spaces of three or more dimensions and the tendency to limit
cartographic representations to planar models, GIS practice today is generally restricted to two considering problems with continuous geometric variables. Information concerning the third spatial dimension, and perhaps time, is typically treated as an attribute (Greve, Kelmelis, Fegeas, Guptill and Mouat, 1993, page 1505). Even the SDTS, which purports to be a general-purpose mechanism for the transfer of most forms of spatial data, is essentially limited to two-dimensional information. The basic object definitions in SDTS, though valid for nonplanar geometry since coordinate values may include z-values, are oriented to two-dimensional surface representation (Fegeas, Cascio, and Lazaar, 1992, page 279).

It can be concluded that significant unresolved problems remain in three-dimensional cartographic modelling of spatial phenomena. Since most existing forms of spatial data are limited to a two-dimensional representation, the point, line and area point quality primitives are an adequate foundation for the control of processes generating geographic information.

However, the quality primitives may yet prove to be a satisfactory set of control subjects even when the problems of three-dimensional spatial modelling are resolved. It was noted above that existing methods for measuring spatial phenomena are ultimately restricted to point and line quantities. Conventional surveying methods for capturing
geometric measurements of features are limited to observing distances, elevation differences, and angles among points. Similarly, photogrammetric techniques produce data in the form of points and lines. The basic operation in analytical photogrammetry involves the transformation between a point in the image space and a point in the object space (Slama, Ebner and Fritz, 1980, page 477). In essence then, photogrammetric measurement can be reduced to the point object. More recently developed methods for acquiring geographic information, such as the combination of Global Positioning System (GPS) and Inertial Navigation System (INS) technology in a mobile mapping context, are also limited to point and line objects. A mobile mapping system (based on GPS/INS) generates data in the form of a nearly continuous linear path through space. The INS provides a record of accelerations in three mutually orthogonal directions with respect to a reference frame (established and maintained by gyroscopes) that are integrated to provide the x, y, and z components of displacement. Inertial measurement units (IMU’s) rely on physical quantities—linear and angular accelerations—from which distances between locations are determined. GPS, which is used to control the INS, produces point positions based on distance measurements between the receiver and satellites. The GPS segment of a mobile mapping system is directly based on distances between points in space. And although the INS is
not based on point measurement per se, the data are integrated to provide linear displacements.

Thus, points and lines may be the only way in which the geometric component of spatial information (regardless how many dimensions are used in the data model) can be obtained through direct measurement. Even remote sensing devices record point summarizations (in the form of single-valued pixels) of reflectance characteristics of a surface. It is unlikely that methods of measurement exist that produce direct three-dimensional volumetric or surface data for spatial phenomena.

Clearly, the planar primitive objects, point, line and area point, can easily be applied to such three-dimensional measurements simply by adding a third coordinate for the z axis. Points are located in space as \((x,y,z)\) coordinate triplets, lines trace paths through space (possibly defined as ordered collections of point triplets), and area points can be redesignated as "volume points" that carry the attributes for the space in which they are located. Assuming methods employed to obtain geometric data that go into three-dimensional spatial models are predicated on point and line measurements--the quality primitives would constitute an adequate set of subjects for quality control. The only way to evaluate geometric accuracy of the geographic information (with respect to the ground) would rely on spatial measurement that is ultimately founded on
point and line observations. Although the information may reside in three-dimensional objects, for example, surfaces, faces and solids, the underlying measurements are still based on the point and line primitives.

This discussion suggests that the point, line, and area point primitives may be satisfactory control subjects even in the context of three-dimensional spatial data models. When suitably extended to three dimensions, these primitives are adequate when evaluating geographic information on the basis of geometric spatial measurements. The three quality primitives are suitable control subjects in the case of planar cartographic information. Since true three-dimensional spatial information is generally not produced at the present time, the quality primitives are sufficient for the geographic data being produced today.

Conclusion

This chapter has argued that cartographic information has both geometric and attribute components and can be broken down into individual pellets of information. A specific set of three primitive elements, the point, line, and area point objects, form the basic units of cartographic information in graphic map representations as well as two-dimensional digital raster and vector spatial data. Like the alphabet characters in the English text processing context, these primitives form the basis for measuring and evaluating the quality of cartographic information.
Therefore, the point, line, and area point objects, together with their attribute (descriptive) information, are the characteristics of a cartographic product to regulate in the control process. In this sense, they comprise a general purpose set of control subjects for mapping processes.

The concept of cartographic control subjects developed in the present chapter can now be combined with the definition of the conformance dimension of geographic information quality developed in Chapter III. Quality of conformance (freedom from deficiencies) in mapping processes is satisfying design specifications with respect to the accuracy (positional and attribute), logical consistency, and completeness (both feature and attribute) of the point, line, and area point objects that comprise the cartographic information product. In other words, statistical control focuses on regulating the point, line and area point objects generated by the mapping process to ensure their conformance to design specifications for accuracy, consistency, and completeness. By setting forth the fundamental units of geographic information, a general framework for quality control in cartography has been established.

This framework is quite general. Implementing statistical quality control in a particular mapping process will likely involve the measurement and regulation of other quantities in addition to the geometric and attribute characteristics of the point, line and area point objects.
For instance, the GISOM (Generating Information from Scanning Ohio's Maps) project at the Ohio State University Center for Mapping transforms information from existing 7.5 minute, 1:24,000 scale USGS topographic map information into computer Digital Line Graph (DLG-3) file data. One of the most significant steps in the GISOM process is transforming the analog source material (stable base map document) into a bi-level digital image by means of raster scanning. The raster scanning operation should be placed in a state of statistical control to ensure stability and reliability of the resulting digital image. It is an imaging step whose quality is defined by radiometric and geometric quantities and takes place prior to the generation of cartographic quality primitives. Calibrating and controlling the raster scanning process would involve measurements of pixels, which, as it has been argued above, are point objects. Similarly, conventional surveying methods require the calibration and control of instruments and procedures that measure distances, angles, and height differences. These processes occur before the geometric or attribute aspects of the cartographic point, line, and area point primitives are created.

As these examples suggest, there will typically be stages in cartographic processes that do not entail the generation or manipulation of the point, line, and area point objects. And, mapping processes will involve a wide
range of particular measurement and observation subsystems, ranging from GPS and mobile mapping to traditional ground surveying and photogrammetric methods. Controlling the quality of a mapping process will involve the regulation of a variety of subsystems and will be unique for each particular production context.

However, at some point in the production of geographic data, the fundamental units of cartographic information set forth in this chapter (point, line, and area point objects) will be created or transformed in some way. And, it is the extent to which these basic elements conform to design requirements (with respect to accuracy, consistency, and completeness) that will largely determine the fitness for use of the geographic information product. The user of a cartographic product is typically unconcerned with process requirements for the raster scanning device or theodolite but will take a keen interest in the quality of the point, line, and area point primitives. While the quality primitives do provide a general framework from which all mapping processes can and probably should be evaluated, it is impossible in the present research to describe the procedures for placing any mapping process in a state of statistical control. They are too numerous and too varied to be treated here.

But, in the basic and general sense, the principal goal in mapping quality of conformance will be ensuring that the
point, line, and area point objects comply with the requirements for accuracy, consistency and completeness called for in the product design.
CHAPTER V

MEASURING CARTOGRAPHIC DATA QUALITY

The previous chapter set forth a basic framework for the conformance dimension of cartographic information quality. A set of two-dimensional primitive elements, consisting of the point, line, and area point were identified as the basic building blocks of geographic information. These quality primitives are combined to produce the geometric (skeletal) expression of cartographic information. Attribute information is linked to the primitive elements to form a set of fundamental objects of geographic information. These objects comprise the basic control subjects that need to be regulated in the control of cartographic information quality.

In Chapter III, the basic dimensions of cartographic information quality were identified as accuracy, logical consistency, and completeness. Controlling quality in a mapping process is essentially a matter of ensuring that design specifications for accuracy, logical consistency, and completeness of the geometric and attribute aspects of the point, line, and area point objects are met. By setting
forth the things that need to be regulated, the framework for cartographic quality has been established.

What remains to be considered is the problem of how to go about evaluating and measuring cartographic data quality in a quantitative manner. In Chapter I, it was noted that quality measurement is at the heart of quality control; what gets measured, gets done. Measurement requires the establishment of quantitative methods by which the quality of conformance of production processes and their output (product) can be expressed and evaluated. Quality measurement thus forms the framework for articulating design requirements and operational criteria for mapping processes and their output. The basic issue yet to be resolved is the development of methods for the quantitative evaluation of the accuracy, consistency and completeness of the point, line, and area point objects with respect to their geometry and attributes.

The scope of this remaining issue is quite broad and encompasses the subproblems of establishing methods for measuring the accuracy of attributes, geometric accuracy, completeness, and logical consistency. Of these four particular concerns, logical consistency is fundamentally different from the other three. Geometric and attribute accuracy and completeness involve the relationship between the data and the phenomena they represent. Their evaluation will test the relationship between the data and "truth" that
is external to the data. Logical consistency, on the other hand, is based only on the data and does not appeal to "truth" beyond the data. Evaluating logical consistency is a matter of checking that the information satisfies a set of a priori constraints. In this way, accuracy and completeness are external quality characteristics while logical consistency is an internal quality characteristic.

In Chapter III, it was suggested that geographic information can be subjected to a vast array of a priori constraints. The reason is the complexity and redundancies inherent in cartographic data. There are indeed numerous rules that can be imposed on cartographic information (Ramirez, 1992, pages 187-195). Elevation attributes for adjacent contour lines may differ by no more than the contour interval, contour lines cannot cross a lake or pond, water cannot flow up hill nor can it reverse its flow, and road features should be connected to the wider network of roads, to suggest only a few possibilities. Constraints may also be based on the geometric nature of the phenomenon. Alignments for highways and railroads for instance, may be limited to a small class of geometrically defined curves whose radii of curvature are always greater than a given size. Other curves and curves that are too sharp would be deemed inconsistent.

And, if the spatial information covers a limited area (say a single state in the United States, for example)
additional rules might be developed based on geographic conditions within the region. For instance, a glacier feature would be impossible in Florida as would an elevation exceeding 345 feet. A simple constraint for a contour elevation attribute from a topographic map within Florida is to reject if greater than 345 feet.

Logical consistency in spatial data can also be based on stochastic constraints as well. Empirical evidence might be used to develop a set of probabilities concerning the likelihood of certain feature circumstances. For example, if a map contains an interstate highway feature, then it is very likely the highway will cross the map’s border.

Developing a set of consistency checks requires a thorough understanding of the various relationships among the features in the data. A rather comprehensive theory of spatial phenomena (as well as a good deal of empirical experience in the case of stochastic constraints) would be needed to establish a complete set of cartographic constraints for a geographic information product. That such knowledge presently exists is doubtful (Laurini and Thompson, 1992, page 591). In Chapter III, it was suggested that logical consistency checking could be extended well beyond that envisioned in the SDTS or embodied in the PROSYS quality validation system used by the USGS in the production of Digital Line Graph (DLG) data (Bicking, 1994, pages 19-22). Indeed, Bicking notes that PROSYS (and its Attribute
Verification Package or AVP module in particular) does not make use of the information encoded in map symbols nor does it incorporate constraints from the graph-theoretic and topological principles upon which the DLG is based (Bicking, 1994, page 22).

This discussion suggests that logical consistency for geographic data is a complicated issue that rests on an extensive understanding of the relationships and redundancies inherent in spatial data. A good understanding of logical consistency in geographic information is a subject for research in its own right. Logical consistency is, nonetheless, a major component of data quality that can and should be placed under statistical control (Redman, 1992, pages 248-249). The number, nature, and pattern of failed consistency checks would provide very useful information in the control of processes that generate cartographic information. A change in their overall number may signal the presence of assignable causes that require correction. Analysis of a mapping process that is in a state of statistical control with respect to logical consistency may also point to specific areas where the process can be improved.

To evaluate attribute information it is necessary to take into account the four measurement scales, nominal, ordinal, interval, and ratio, in which it can be expressed. As noted in Chapter III during the discussion of the SDTS,
there is no way to measure how close a categorical (nominal or ordinal scale) attribute value is to a true value. It is either correct or it is incorrect. For example, a road classification (interstate highway, state route, dirt trail) is either right or wrong. The name of a city may be misspelled and so be more or less correct (depending on how many letters are in error), but it is nonetheless wrong. Accuracy for nominally or ordinally scaled attributes is thus a binary quantity.

Summary statistics for categorical attributes are based on a count of non-conforming (defective or erroneous) values. It should be noted that counts of errors form the basis of the misclassification matrix that was briefly considered in Chapter III. The misclassification (confusion or error) matrix is the method of evaluating attribute accuracy suggested in SDTS.

For numerical (interval or ratio scale) attributes, it is possible to measure the degree of closeness between two values. An obvious situation where attributes are presented on a numerical scale is the digital elevation model or DEM where each elevation can be regarded as descriptive information associated with a correctly located point object. The discrepancy between the value of the elevation attribute in the data and its true value can be expressed in numbers, deviating so many meters or feet. Summarizing discrepancies in interval and ratio scale data can be
accomplished through the use of statistical quantities, average, standard error and root mean square error.

There may be circumstances too in which it may be entirely reasonable to treat ratio and interval attributes as categorical data. A good example would be the conversion of an existing analog map document into digital form. The aim of the process is to reproduce the analog information in a computer form. Whether the elevation attribute assigned to a given contour line or spot elevation is wrong by an inch or 1,000 feet is not as significant as the very fact that it is incorrect. In the context of analog to digital map conversion, it is likely that all attributes would be treated as categorical. And, in general, it may be desirable to reduce numerically scaled attributes to a nominal scale when attempting to control a mapping process. If a single attribute’s accuracy is specified in terms of an allowable range of values and out of range values are relatively rare, then a count of the number of defective attributes may prove more useful than an average numerical discrepancy.

Specifying and measuring quality for processes that generate cartographic attribute information depends on the measurement scale of the information being produced, which in turn depends on the particular mapping context. However measured, attribute accuracy requirements need to be stated in terms of operational definitions, in a clear and
meaningful way. As noted in Chapter III, there is no way to measure conformance to vague specifications such as "attribute codes shall agree within 98.5 percent to coding standards," contained in the USGS-DLG document.

Testing and evaluating geographic information for completeness is a matter of verifying the existence of a one-to-one correspondence between the objects in the data and the set of features to be mapped. It should be noted that establishing this correspondence between data and truth can be a nontrivial task. A terrain feature that has not been included in the spatial data is wrong (incomplete) with respect to both geometry and attributes. Since there is no spatial object in the data, detecting the incompleteness would require exhaustive search of the external source of truth. Assuming the number of incomplete or missing objects and attributes is relatively small in comparison to the number of objects in the data, the problem of discovering them is significant.

It is clear from this discussion that establishing quantitative methods for evaluating logical consistency, attribute accuracy and completeness of geographic information is nontrivial. Each subproblem is complex in its own right and will not be considered further in this research. In the balance of this document, the focus will be limited to the problem of evaluating the geometric accuracy of the point, line, and area point objects. Before
delving into the details of this problem, some general
observations will be made concerning the nature of
measurement in the mapping context.

General Aspects of Quality Measurement in Cartography

Measurement in the Control of Quality in Cartography

It is worth recalling that the principle aim of measurement in the control of quality is to determine whether the production system and its output conform to design specifications. How discrepancies between the process output and its goal are characterized should provide useful feedback on the functioning of the mapping process and the output generated. As noted in Chapter I, an important goal of quality control is establishing and maintaining stable production systems. In order to achieve stability, it is necessary to detect and remove special causes from the system. Measures of accuracy, logical consistency, and completeness should reveal as much as possible about the variation and causes present in the mapping process.

In so far as mapping is a measurement process, statistical control is itself a necessary precondition for quality. As a measurement operation, the mapping process must exhibit a stable, predictable pattern of variation before anything can be said of its accuracy, precision or
quality. Eisenhart makes this point rather forcefully, stressing,

... That a measurement operation to qualify as a measurement process must have attained a state of statistical control; and that until a measurement operation has been "debugged" to the extent that it has attained a state of statistical control, it cannot be regarded in any logical sense as measuring anything at all.

(Eisenhart, 1969, page 41)

In this sense, statistical control is not merely a desirable state to assure efficient production, it is constitutive in mapping. How the accuracy, logical consistency and completeness are evaluated must support the attainment of statistical control in the production of cartographic information.

Another aspect of measurement that should be recalled is the fact the goal or specifications to which the process output is compared can never be absolutely definite. The reason, as noted in Chapter I, is the true value of any measured property or datum can not be known in the strictest sense. Terms related to the "truth" or "true value" should probably be replaced by "standard," "reference," or "target value," as true values can not be known (Eisenhart, 1969, page 31). Specified target values must be obtained as the basis for evaluating quality. In general, target values are generated by the measurement procedure deemed to be best for the purpose at hand.

In most mapping situations it would be possible to obtain a standard or target value for a datum by means of an
independent measurement process of higher accuracy and precision than the original production system. Field survey, especially using GPS techniques, is one obvious method available for checking cartographic data. However, such independent verification in mapping processes is probably rare in actual practice. Nevertheless, some target or standard value must be assumed as the "truth" against which the product is compared for the purpose of evaluating quality. In the discussion that follows the term true value will be retained. However, it will denote the provisional target or reference value (by whatever means acquired) that serves as the basis for evaluating accuracy.

**Metric Concept**

The fundamental issue in quality measurement is characterizing the difference between the datum that gets generated by the mapping process and the true value (or value presumed to be true). Ideally, the discrepancy should be reduced to quantities amenable to mathematical and statistical analysis. In mathematics the notion of the distance between members of an abstract set leads to a very general way of describing differences. This abstract formulation of distance is formalized in the definition of a metric.

Let $X$ be a set. A metric on $X$ is a real function $d$ of ordered pairs of elements of $X$ which satisfies the following three conditions:

1. $d(x,y) \geq 0$, and $d(x,y) = 0 \iff x = y$;
(2) \( d(x,y) = d(y,x) \) (symmetry);
(3) \( d(x,y) \leq d(x,z) + d(z,y) \) (the triangle inequality).

(Simmons, 1963, page 51)

In the context of quality measurement, the set \( X \) would contain the control subjects (point, line, and area point objects) to be evaluated. That is, the set \( X \) contains both the quality objects that were "produced" in the mapping process and the standard or true ones. If \( x \) is the object that is produced and \( y \) is the true object, then the distance between \( x \) and \( y \) can be any real function that satisfies the three conditions of a metric.

The metric concept is quite general so it is possible to define several different measures on a set of objects. Moreover, the set of objects upon which the metric is defined is arbitrary. It can consist of numbers, functions, or any other well-defined collection of entities, including the cartographic quality objects. Any real-valued function that satisfies the conditions can fulfill the role of a metric. The value of the concept is that it reduces differences between objects to real numbers and thus enables quantitative comparisons to be made.

Perhaps the simplest example is the discrete metric shown in (4).

\[
d(x,y) = \begin{cases} 
0 & \text{if } x=y \\
1 & \text{if } x\neq y
\end{cases}
\] (4)
If the two objects, \( x \) and \( y \), are identical, then their difference is zero. If they are the slightest bit different, then the difference between them is unity. This metric does not discriminate between degrees of difference; it is binary and provides no indication how close two elements are.

A more familiar metric is the one induced by the Euclidean norm. A norm is a real number that characterizes the size of an element with respect to an origin or zero element of the set. The metric follows by defining distance as the norm of the difference between two elements of the set. For points in the plane, the Euclidean norm is defined in (5).

\[
\| (x, y) \| = (x^2 + y^2)^{1/2}
\]

(5)

For two points in the plane, \((x, y)\) and \((u, v)\), the metric induced by (5) is set forth in (6).

\[
d((x, y), (u, v)) = \| (x, y) - (u, v) \|
= \| (x-u, y-v) \|
= \sqrt{(x-u)^2 + (y-v)^2}
\]

(6)

This metric is the linear distance between the points and can easily be extended to points in three-dimensional space. It is the one that is usually adopted to describe the geometric difference between planar cartographic points. The three-dimensional version is the metric used in most surveying, geodetic, photogrammetric and geographic
applications. Unlike the discrete metric, this distance measure characterizes degrees of closeness; it is possible to analyze how close one point is to another.

Comparisons between the control subjects and the standards (or true values) may or may not fulfill the conditions of a metric. When a numerical difference is used, as opposed to its absolute value, the discrepancy can be either positive or negative. Or, perhaps differences might be expressed in terms of a vector rather than a scalar quantity. In such cases, the measure will not be a metric. Whether or not it constitutes a metric in the formal sense, a quality measure will often be apparent; the metric based on the Euclidean distance can be used in characterizing the geometric quality of cartographic point object. However, in the case of line objects, a measure does not so readily present itself for adoption. How can a single positive real number characterize the complex relationship between two lines? Quality measurement can be reduced to the problem of specifying measures (metric or otherwise) that can be used to quantify differences in the geometric and attribute aspects of the point, line, and area point objects.

**Measurement of Geometric Quality**

**Area Points**

It will be recalled from Chapter IV that the area point functions as an explicit surrogate for features that have an areal extent. It is defined geometrically as a point and
carries the attribute information for the region in which it is located. The actual extent of the areal feature is defined by its bounding lines. Evaluation of the accuracy of area points is a rather straightforward proposition for those data formats, like the DLG, that employ them. Their geometric accuracy is a matter of whether the area point is actually interior to the region to which its attributes apply.

Interactive graphic displays of the data will likely present the user with an explicit point symbol located interior to the areal feature it represents. However, geographic data structures in general need not possess the area point object. They may use explicit polygon objects, or topological structure together with attributes on lines to define areal features and endow them with descriptive information. In such cases, the measurement of area points will be essentially a virtual exercise, with the necessary transformations being accomplished by means of computer software.

The most natural way for a human operator to identify an areal feature in an interactive computer map display is to point the screen cursor somewhere within its interior. The machine then figures out the lines that comprise its boundary and retrieves the appropriate database item, such as a polygon. (When one of the area’s bounding lines must be identified by the operator, there may ambiguity as to
which area, left or right was intended.) Similarly, when an areal feature is identified by the computer as through a nonspatial query, it usually displays the region with some kind of crosshatching or point symbol the user can recognize. Such basic GIS and computer mapping functions can be used to emulate the existence of the area point object when it is not a part of the formal data structure. In most cases, the computer can be asked to act as if the area point were part of the underlying data structure.

Now precise geometric location of the area point object is not too significant. As long as an area point is clearly interior to the region to which its attributes apply, it is accurate. Strictly speaking, there are no degrees of accuracy in the location of an area point; its location is either correct or it is incorrect. Depending on the circumstances, particularly the scale, resolution, and complexity of the data, there may not be much room in which to locate the area point object. For example, placement of the area point may be difficult for very small or long and thin areas. In this case, resolution of the data can matter. But, actual location is not a matter that can be verified with respect to ground truth. Geometric accuracy of the area point primitive can thus be reduced to the basic issue of whether or not it is inside the proper region. In this way measurement of the geometric accuracy for the area point object is based on the discrete metric in (4).
However, most of the problems arising from the mislocation of this primitive (say duplicate area points, area points sitting on the region's bounding lines, missing area points, and so on) can probably be reduced to the level of logical consistency. Many potential geometric errors would also constitute inconsistencies that could be identified with a requirement that each unbroken closed region in a planar cartographic product possess no more than one area point. In this way, the majority of the geometric accuracy of the area point object is subsumed under logical consistency (internal quality). But, with those data formats where not every region is attributed or explicitly accounted for, the area point primitive must be positioned, even if imprecisely, inside the correct region. For these kinds of situations at least, the geometric accuracy can not be completely reduced to a matter of internal quality.

To get an idea of the overall level and variation of area point accuracy in a mapping process the total number of nonconformities might be divided by the total number generated in the file or map sheet. Or, perhaps, the number of nonconforming area points would be related to some other grouping such as operator, workstation, or unit time. In any case, because there is no numerical degree of geometric accuracy for area points, their aggregate behavior must be analyzed in terms of a count (or fraction derived from a count) of the number of nonconforming occurrences.
The various methods for evaluating the geometric accuracy of the point primitives will be reviewed in the remainder of the present chapter. In the next chapter, the problem of evaluating the geometric accuracy of the line primitive will considered.

**Point Discrepancy Measures**

If it is assumed the potentially distorting influences associated with the map projection and datum surface have been taken into account, then the basic problem of characterizing the geometric accuracy of the point primitive reduces to the problem of characterizing the discrepancy between two points in the plane. Given two points, \((a,b)\) and \((c,d)\) in the plane, how is their difference to be reduced to numbers? There are at least three common ways to characterize the spatial difference and these will be considered in turn.

Perhaps the most common numerical measure of the difference between two points in the plane is the Euclidean distance separating them. The distance, \(e\), is determined in accordance with (7)

\[
e((c,d),(a,b)) = \| (c,d) - (a,b) \| \\
= \| (c-a, d-b) \| \\
= \sqrt{(c-a)^2 + (d-b)^2}
\]  

and is shown graphically in figure 15. The difference, or positional error, between the true location and the one
depicted in the cartographic data is a one-dimensional quantity. Clearly, this measure will discriminate degrees of closeness between the two points and constitutes a metric in the formal sense. One consequence of this fact is that error will always be a nonnegative value (the first condition in the definition of a metric).

This metric is the approach employed in the United States National Map Accuracy Standards (NMAS 1947) that were discussed in Chapter III. Recall that expression of the horizontal accuracy requirement was of the form, no more than 10 percent of the points tested shall be in error by more than 1/50" or 1/30". In other words, horizontal map error is a one dimensional quantity consisting of the distance between the true position and the actual location of the point. As also noted in Chapter III, the NMAS 1947 horizontal accuracy requirement can be thought of as a circular region surrounding the test point on the map. The
distance metric corresponds to the radius of the circle and for this reason is often referred to as radial error.

The major drawback to (7) as a measure of point accuracy is that it collapses a difference that has both magnitude and direction to a single positive number. Maling notes that this approach to geometric point accuracy can be criticized because it can mask directional biases in the discrepancy (Maling, 1989, page 151). In reducing the planar discrepancy to a single real number, as with NMAS 1947, some potentially valuable information may be lost. For instance, consider a situation in which the error at all points on a map is attributable to differences in the direction of the y-axis; x coordinate values for the map and true locations are always identical. Such a situation may indicate some systematic bias in the mapping process. Or, in the language of quality control, it may signal the presence of special causes that need to be eliminated. The NMAS 1947 distance metric, by not reflecting direction in any way, would not provide an indication of the pattern in the discrepancies. Maling notes that the angle of the vectorial error could be measured quantitatively but that it is seldom done (Maling, 1989, page 151).

Another approach to geometric point accuracy treats the x and y components of the discrepancy as separate quantities. In other words given two points, (a,b) and
(c, d), their difference is characterized in terms of two
distinct values, \( e_x \) and \( e_y \) that are determined as in (8).

\[
e_x = c-a \quad \text{and} \quad e_y = d-b
\]  

(8)

Figure 16 displays the relationships involved in determining the separate components of the discrepancy. Instead of

![Diagram showing geometric point discrepancy measure in which x and y components are separate quantities.](image)

Figure 16. Geometric point discrepancy measure in which x and y components are separate quantities.

reducing the discrepancy to a one-dimensional metric quantity, this measure yields two scalar values, one each for the x and y dimensions. It should be noted that this approach is the basis of the USGS-DLG, ASPRS 1990, and NCSSA 1994 standards considered in Chapter III.

In principle, it is possible to detect some patterns or biases in point discrepancies using such a measure. Separate tallying of the errors, \( e_x \) and \( e_y \), would uncover the situation in the example suggested above, where the error associated with several points is due to differences only in the y-coordinate. However, a measure that treats \( e_x \)
and \(e_y\) as independent quantities would not detect the pattern where the point errors were all along the line \(y=x\) and had zero means in the two component directions.

Although the characterization of the geometric discrepancy of two points in terms of independent differences in \(x\) and \(y\) will shed some light on general patterns of errors in cartographic data, the approach may not be sensitive to all preferences in orientation. For this reason, the measure may not be particularly suited to the purpose of controlling quality where it is important to be able to detect special causes in the underlying process.

The fact that neither of the two methods for describing the difference between two planar cartographic points considered provides a complete picture of the geometric discrepancy suggests another method of evaluation be used. Directional bias (or patterns) in the error would be better detected if the discrepancies, \(e_x\) and \(e_y\), were treated as joint, rather than independent, quantities. This notion leads to the bivariate error model depicted in figure 17.
Figure 17. Two-dimensional vector measure for geometric point discrepancy.

Given the two points \((a, b)\) and \((c, d)\), the expression for the bivariate error is given in (9).

\[
(e_x, e_y) = (c-a, d-b)
\]  \hspace{1cm} (9)

Instead of reducing the geometric discrepancy to a pair of independent scalars, it characterizes difference in location as a two-dimensional vector whose coordinates are derived from the differences (or errors) \(e_x\) and \(e_y\) in the two component directions, \(x\) and \(y\). Since the bivariate error measure is a vector quantity and not a nonnegative real number, it does not constitute a metric.

In general, the bivariate vector error model preserves the entire character of the discrepancy in a two-dimensional form so that no information is lost. Both the magnitude and direction of the difference between two points can be inferred from the error vector. As a consequence, patterns or bias in the error should, in theory, be detectable using this measure. Another important aspect of the bivariate
model is that it subsumes the other two measures considered earlier. Both the metric distance measure and the separate discrepancy measure can be derived from the error vector.

The primary disadvantage of the vector-based model is its complexity. Evaluating a collection of metric distance measures (as used in NMAS 1947), or the scalar quantities $e_x$ and $e_y$ obtained from the separate discrepancy measure (as in USGS-DLG, ASPRS 1990, and NCSSA 1994), for patterns or biases is a rather straightforward matter of evaluating numbers. Summarizing the bivariate errors for several points involves a more complicated vector object.

Despite its complexity, the bivariate measure of point discrepancy will be adopted here. This measure is well-suited to the task of uncovering patterns or biases in the data that indicate the presence of special causes (and the lack of statistical control) in the underlying mapping process. And, in any case, the vector discrepancy measure can be reduced to either of the two simpler measures if circumstances warrant. For these reasons then, it is reasonable to adopt the bivariate model of measurement for geometric point accuracy.

**Summary and Statistical Measures for Geometric Point Accuracy**

To this point in the discussion, the stochastic model of differences in point location has not been considered. The focus has been on how the discrepancy should be
characterized in a strictly geometric sense. However, it seems that the principal concern in many treatments of map accuracy is how these point discrepancies are summarized and described in terms of a statistical model.

At the moment, there is a difference of opinion as to whether correct statistical theory has been used in the literature. Briefly, the nub of the debate is whether position error is a univariate (as previous literature assumes) or a bivariate meaning that it is composed of two variables, x and y).

(Thompson and Rosenfield, 1971, page 58)

The reference to previous literature is perhaps to the United States National Map Accuracy Standards (NMAS 1947). Here the debate seems to concern both the question whether map error is a bivariate or univariate quantity and how it should be described in terms of a stochastic model.

In the present section, an attempt will be made to unravel the relationship between the geometric discrepancy measure and the statistical methods used to summarize the aggregate behavior of a collection of such differences. In particular, the stochastic model that is usually adopted in the bivariate vector discrepancy model will be related to statistical measures typically applied to the metric distance and separate discrepancy measures.

The U.S. National Map Accuracy Standards of 1947 (NMAS 1947) do not assume a stochastic model for map error. Instead, they use a percentage criterion of the form, not more than 10 percent of points tested shall be in error by more than a given tolerance distance. In particular, for
the USGS 1:24,000 scale topographic map series, the
tolerance is 1/50" which corresponds to 40 feet on the
ground. For this map series then, the specification can be
restated as: 90% of test points shall be in error by 40
feet or less. Writing in 1956, Morris Thompson suggested
that the form of the specification could be improved if it
were stated in statistical terms that reflected size of the
tolerance on either side of the tolerance (Thompson, 1956, page
167). Rosenfield (Thompson and Rosenfield, 1971, page 60)
indicates that the 90% criterion can be related to the
equivalent standard error (which is presumably equivalent to
the root mean square error or RMSE), \(d_h\), for horizontal
position through (10).

\[
d_h = \frac{40}{1.66} = 24 \text{ feet}
\]  

(Both the standard error and RMSE or rms will be discussed
in more detail below.) The 1.66 divisor approximates the
(1.6449) conversion from the 90% (two-tailed) probability
level for the standard normal density function. The use of
1.66 in (10) by Thompson and Rosenfield is probably due to
the fact that it yields a value (24.09 feet) that is more
nearly an integer.

The apparent theoretical justification for (10) stems
from (11)

\[
\rho - N(0, \sigma^2) \text{ so } \frac{\rho}{\sigma} = z - N(0,1)
\]  

(11)
where horizontal errors, e, are assumed to follow a normal probability distribution with zero mean. Appealing to the 90% probability confidence interval for a standard normal random variable yields the final relationship in (12)

\[
Pr(-1.6449 \leq z \leq 1.6449) = 0.90
\]
\[
Pr(-1.6449 \leq \frac{e}{\sigma} \leq 1.6449) = 0.90
\]
\[
Pr(-1.6449\sigma \leq e \leq 1.6449\sigma) = 0.90
\]

after the appropriate substitution from (11) is made. Since the horizontal error e must be 40 feet or smaller with probability 0.90, the expression for \( \sigma \) can be developed from (12) as shown in (13).

\[
Pr(-40 \leq e \leq 40) = 0.90 \text{ so } -1.6449\sigma \leq 40 \leq 1.6449\sigma \text{ or }
\]
\[
\sigma = \frac{40}{1.6449} \approx 24
\]

Since the error is a distance quantity that must always be positive, the left hand inequalities in (13) are not meaningful. Since the true value for \( \sigma \) cannot be known, the standard error statistic, \( d_h \), is used instead. This relationship between the maximum allowable RMSE, \( d_h = 24 \) feet, is noted later by Thompson (Thompson, 1981, page 105).

This theoretical justification can be criticized on several related counts. At a very basic level, substitution of the observed standard error (RMSE or rms) for the population parameter \( \sigma \) might offend some. But, perhaps the most obvious problem in the development involves the fact
that one-dimensional map error can never be a negative quantity. The distance between the location of a point on the map and its true position must always be zero or greater. For this reason, it is theoretically impossible for the errors to be distributed about a zero mean. The mean value of a random variable based on the univariate distance quantity must be zero or greater. And, it will be zero only if all the errors are zero. Therefore, the assumption that the distribution of map errors has a zero mean is not appropriate.

Since the normal distribution is symmetric about its mean and has a long tail in both directions, it is not a good model to apply to positive distance errors—especially under the dubious assumption that the mean error (distance) is zero. To reasonably employ the normal distribution, it is necessary to assume at the very least the mean distance error is nonzero. It may be more theoretically correct to model the univariate horizontal map error with a one-tailed density function that vanishes for values less than zero. In either case, it is this writer's opinion that the 24 foot tolerance (or any other value) can not be related to a 90% normal confidence interval through the RMSE statistic as suggested by Thompson (Thompson, 1981, page 105).

A final problem concerns the use of the root mean square (rms) error (or RMSE) statistic to characterize the variation in horizontal map error. Of the five geographic
information standards reviewed in Chapter III, an RMSE is explicitly referenced in three of them (DLG 1988, ASPRS 1990, and NCSSA 1994) as well as implicitly in a fourth (SDTS by reference to ASPRS 1990). The RMSE, which is often also identified as the standard error, is based on the observed discrepancies, e₁, e₂, . . . , eₙ, between the map location and the true position at the n test points as set forth in (14).

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n}}
\]  

(14)

Use of the RMSE to approximate the standard deviation of the error σₑ is suspect whenever the expected (or mean) error μₑ is not zero. When μₑ ≠ 0 the RMSE may overstate the actual variance of the error because mean square error reflects the influence of bias as shown in (15).

\[
e^{-}(μₑ, σₑ^2), μₑ ≠ 0, \text{so } E(e^2) = σₑ^2 + μₑ^2 > σₑ^2
\]

(15)

And, since the square of the RMSE approximates the expected value of the squared error as shown in (16),

\[
RMSE^2 = \frac{\sum_{i=1}^{n} e_i^2}{n} = E(e^2) = σₑ^2 + μₑ^2 > σₑ^2
\]

(16)

it will be influenced by a nonzero mean value as well. In the case of the distance metric used in NMAS 1947, the expected error can be zero only when the probability of
error is zero. It is virtually certain then, that the RMSE value would tend to exaggerate the true standard deviation of the error. For this reason, the RMSE probably should not be used in approximating the variation of the error in NMAS 1947; assertions about the 90% confidence region based on the RMSE would certainly not be appropriate.

In any case, use of the RMSE as a method of estimating variation presumes that errors have zero mean. As noted, such an assumption can not reasonably be made in the case of the univariate distance metric. When errors for the x and y component directions, $e_x$ and $e_y$, are treated separately (as in DLG 1988, ASPRS 1990, and NCSSA 1994), such an assumption is at least theoretically valid. It is possible for discrepancies, say for the x dimension, to have zero mean, because they can be both positive and negative. However, even in this situation, the RMSE can be criticized because it requires deviations in x and y to have zero mean. If the errors, $e_x$ and $e_y$, have nonzero means, then there may be a preferred direction in the orientation of the discrepancies. Blind use of the $\text{RMSE}_x$ and $\text{RMSE}_y$ statistics would tend to mask this situation.

For this reason, Maling suggests the use of a different statistic to estimate the consistency of the map (Maling, 1989, pages 150-151). He believes the standard error is often confused with root mean square error and standard deviation and that the terms should be differentiated.
According to Maling, the standard error is not the same as the RMSE that was defined in (14). Instead he defines the standard error, which is also based on the observed discrepancies, $e_1, e_2, \ldots, e_n$, between the map location and the true position at the n test points, as in (17).

$$
\sigma_e = \sqrt{\frac{n \left( \sum_{i=1}^{n} e_i - \frac{\sum_{i=1}^{n} e_i}{n} \right)^2}{n}}
$$

(17)

Computation of this statistic differs from the RMSE in (14) in so far as an estimate of the mean error (18)

$$
\bar{e} = \frac{\sum_{i=1}^{n} e_i}{n}
$$

(18)

is removed from the estimate of the variance of the error.

The standard error statistic (17) as defined by Maling should provide a better method of estimating the variance of the underlying distribution than the RMSE (14). Even when the RMSE is small, its use might obscure a problem with the underlying process if the expected values of the discrepancies $e_x$ and $e_y$ are not zero. It is more likely that nonzero mean errors in $x$ and $y$ (suggestive of a preferred orientation or trend in the discrepancies) will be revealed when the standard error (17) is determined in conjunction with the mean error (18).

It is worth noting at this point too that the normal distribution could be consistently applied to the one-
dimensional distance errors in the NMAS 1947 if the standard error (17) suggested by Maling is used in (10) rather than the RMSE (14). The reason is that the standard error of the one-dimensional distance metric relates discrepancies to an estimate for the mean error $\mu_e$. Differences from the mean error $\mu_e$, that is $e - \mu_e$, can be both positive and negative and can appear to be distributed symmetrically about the mean (in a two-tailed manner) in accordance to the normal distribution.

In the situation where the $x$ and $y$ components of the discrepancies are treated separately (as with USGS-DLG, ASPRS 1990, and NCSSA 1994), standard errors (17) and average errors (18) would be computed independently for the errors $e_x$ and $e_y$ in the two component dimensions. These quantities should be examined for evidence of systematic bias in the errors before the presumption of zero mean errors is made. Notwithstanding the possibility the standard error (17) and RMSE (14) are equal (due to zero mean errors), the discrepancies in the $x$ and $y$ dimensions are treated as completely independent quantities. It seems reasonable that the mapping process by which point geometry is created would result in the errors in the $x$ and $y$ dimensions, $e_x$ and $e_y$, to be closely related. At least intuitively, it would be surprising if $e_x$ and $e_y$ were truly independent.
To treat the discrepancies as related quantities, it is necessary to consider the bivariate model for geometric point accuracy. The balance of this section will explore the stochastic model that is typically used in the bivariate model. Statistical methods typically applied to the metric distance and separate discrepancy measures will be viewed as special cases of the bivariate model.

The stochastic model in the bivariate approach treats the discrepancy between true location and map position as a two-dimensional random error vector \((e_x, e_y)\). Although there may be less than ample evidence to do so (Thompson and Rosenfield, 1971, page 59) map error is usually assumed to be a normally distributed random quantity. This presumption is formally noted in USGS-DLG (United States Department of the Interior, U.S. Geological Survey, National Mapping Division, page 2-6, 10/92). Under this assumption, errors in \(x\) and \(y\) follow a bivariate normal distribution as in (19).

\[
e \sim N(\mu, \Sigma) \text{ where } \\
e = \begin{bmatrix} e_x \\ e_y \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \quad \text{and } \Sigma = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix} \tag{19}
\]

Perhaps the most significant aspect of treating point error from a bivariate perspective is that a consistent stochastic model for the univariate radial (distance) metric can be developed. To derive this relationship it is necessary to
transform the system in (19) to one in which the variables are uncorrelated.

It is always possible to transform the pair of dependent, correlated normal random variables represented by this distribution, $e_x$ and $e_y$, into a pair of uncorrelated, independent normal random variables $e_1'$ and $e_2'$ (Mikhail and Ackerman, 1976, page 31). The procedure for accomplishing this transformation is based on the normalized eigenvector diagonalization of the covariance matrix, $\Sigma$, which is described by Mikhail and Ackerman (Mikhail and Ackerman, 1976, pages 31ff).

The first step is to determine the eigenvalues, $\lambda_1$ and $\lambda_2$, and normalized eigenvectors $u_1$ and $u_2$ (where $u_1$ and $u_2$ have unit length) of the original covariance matrix, $\Sigma$. Then construct the matrix $U$ whose columns are the normalized eigenvectors, $u_1$ and $u_2$. Finally transform the original random vector $e$ to $e'$ by means of (20).

$$e' = U^T(e-\mu)$$

(20)

This transformation amounts to a rotation of the original system as the eigenvectors of a symmetric matrix are mutually orthogonal. The new random vector, $e'$, will be distributed as shown in (21)

$$e' \sim N(\overline{0}, \Sigma')$$

where

$$e' = \begin{bmatrix} e'_1 \\ e'_2 \end{bmatrix}, \quad \overline{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \text{and} \quad \Sigma' = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

(21)
and the variables \( e'_1 \) and \( e'_2 \) are independent, normally distributed random variables with zero means and respective variances, \( \lambda_1 \) and \( \lambda_2 \) (that are the eigenvalues of the original covariance matrix \( \Sigma \)).

In other words, the two random variables, \( e'_1 \) and \( e'_2 \), can be treated separately as in (22).

\[
e'_1 \sim N(0, \lambda_1) \quad \text{and} \quad e'_2 \sim N(0, \lambda_2) \quad \text{with} \quad \sigma_{e'_1 e'_2} = 0 \quad (22)
\]

It will be noted at this point that the treatment of map error underlying USGS-DLG, ASPRS 1990, and NCSSA 1994 implicitly assumes that individual errors in \( x \) and \( y \), \( e_x \) and \( e_y \) satisfy (or have been made to satisfy) the conditions of (22). Strictly speaking, it is only appropriate to consider rms errors in \( x \) and \( y \) separately when those errors fulfill (22). Otherwise, the assumption of zero mean, independent errors \( e_x \) and \( e_y \) may not be at all appropriate to the data. Significant information about the behavior of the mapping process may be obscured by the assumptions used in analyzing the measurements.

Continuing now with the derivation of the one-dimensional error, the random variables in (22) can be standardized by dividing them by their respective standard deviations (23).

\[
\frac{e'_1}{\sqrt{\lambda_1}} \sim N(0,1) \quad \text{and} \quad \frac{e'_2}{\sqrt{\lambda_2}} \sim N(0,1) \quad (23)
\]
Thus, independent, normally distributed random variables with zero means and unit variances are obtained. Since the square of a normally distributed random variable is a chi-square random variable with one degree of freedom (Hogg and Craig, 1978, pages 114-115), the squares of the transformed variables in (23) are independent chi-square random variables with one degree of freedom as shown in (24).

\[
\frac{e_1^2}{\lambda_1} \sim \chi_1^2 \quad \text{and} \quad \frac{e_2^2}{\lambda_2} \sim \chi_1^2
\]  \hspace{1cm} (24)

Since the sum of two independent chi-square random variables is a chi-square random variable with two degrees of freedom (Hogg and Craig, 1978, page 169), the expression in (25)

\[
\frac{e_1^2}{\lambda_1} + \frac{e_2^2}{\lambda_2} \sim \chi_2^2
\]  \hspace{1cm} (25)

is obtained. Finally, the relationship in (26)

\[
Pr\left\{ \frac{e_1^2}{\lambda_1} + \frac{e_2^2}{\lambda_2} \leq k^2 \right\} = Pr\{\chi_2^2 \leq k^2\}
\]  \hspace{1cm} (26)

is developed for the probability of the sum of the squares of the two independent, standardized normal random variables.

At this point, it is useful to recognize that (27)

\[
\frac{e_1^2}{\lambda_1} + \frac{e_2^2}{\lambda_2} = k^2
\]  \hspace{1cm} (27)
represents an equation for an ellipse with axes as indicated in (28).

\[ k_{\sqrt{\lambda_1}} \text{ and } k_{\sqrt{\lambda_2}} \]  

(28)

Therefore, the probability that the transformed random vector \( e' \) is interior to the given ellipse is given by the chi-square distribution in accordance with (29).

\[ Pr\left\{ \frac{e_1'^2}{\lambda_1} + \frac{e_2'^2}{\lambda_2} \leq k^2 \right\} = Pr\{\chi^2_s \leq k^2\} \]  

(29)

In other words, when map errors are treated as two-dimensional random variables, they can be related to an elliptical region in the plane with a given confidence probability.

The shape of the ellipse is determined from the eigenvalue diagonalization of the covariance matrix from the original bivariate normal error distribution. In particular, the map error will be interior to an elliptical region whose axes are \( k\lambda_1^{1/2} \) and \( k\lambda_2^{1/2} \) with the probability given in (29) from the chi-square distribution. (Note that \( k^2 \) is the value to be used in acquiring the relevant probability from the usual chi-square tables.) Some typical values for \( k \) are provided in table 1. An alternative
Table 1. Scale factors $(k)$ for axes of the standard error ellipse with related probabilities (Mikhail and Ackerman, 1976, page 32).

<table>
<thead>
<tr>
<th>Probability</th>
<th>0.394</th>
<th>0.500</th>
<th>0.900</th>
<th>0.950</th>
<th>0.990</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value of $k$</td>
<td>1.000</td>
<td>1.177</td>
<td>2.146</td>
<td>2.447</td>
<td>3.035</td>
</tr>
</tbody>
</table>

The probability value for $k=1$ might be 0.393 or 0.3935 while for the 0.950 probability, $k$ may be closer to 2.448. The particular ellipse whose axes are $\lambda_1^{1/2}$ and $\lambda_2^{1/2}$ (that is, for $k=1$) is called the standard ellipse (Mikhail and Ackerman, 1976, page 30) and represents 0.39 probability that the error is contained within its boundary. Such error ellipses, as they are often called, are widely used in surveying and photogrammetry to characterize the accuracy of point positions (Wong, 1980, pages 86-87 and Davis, Foote, Anderson and Mikhail, 1981, pages 33ff).

At this point, it is possible to relate the elliptical confidence region obtained from the bivariate normal distribution to the one-dimensional distance metric. The radial distance metric (as employed in NMAS 1947) is associated with a circular region centered on the point. So, the connection between the two-dimensional vector-based measure and the univariate distance metric is a matter of relating an elliptical region in the plane to a circular area. However, the elliptical confidence region can be
strictly compared to the circular one only under special conditions; otherwise, the relationship is approximate.

The correspondence between the one-dimensional radial distance metric and the bivariate error is exact only when the elliptical confidence region is a circle. Such a situation will occur (assuming the normal distribution applies) only when the errors in the x and y component directions, $e_x$ and $e_y$, are independent, normally distributed random variables with equal variances (that is $\sigma_x^2 = \sigma_y^2$). No eigenvector diagonalization of the covariance matrix $\Sigma$ is required in this case because it is a scaled identity matrix. Assuming the mean errors, $\mu_x$ have $\mu_y$, have been subtracted or accounted for as in (20), the confidence region for the point error in this case is given in (30).

$$\frac{e_x^2}{\sigma_x^2} + \frac{e_y^2}{\sigma_y^2} = k^2$$  \hspace{1cm} (30)

Since by assumption $\sigma_x = \sigma_y = \sigma_z$, the ellipse is really a circle with radius $k\sigma_z$, in accordance with (27) and (28). The probabilities associated with the interior of this circle are the same as those presented in the above table for the elliptical case. The quantity $\sigma_z$ is often called the circular standard error and corresponds roughly to 0.39 probability that the error is contained within a circle of radius $\sigma_z$. Other quantities related to $\sigma_z$ are defined in table 2. Multipliers (the last column in table 2) for $\sigma_z$
Table 2. Probabilities ($P_c$) of circular accuracy indices related to the circular standard error ($\sigma_c$) (Maling, 1989, page 153).

<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol</th>
<th>$P_c$</th>
<th>Derivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circular standard error</td>
<td>$\sigma_c$</td>
<td>0.3935</td>
<td>$\sigma_c$</td>
</tr>
<tr>
<td>Circular probable error</td>
<td>CPE or CEP</td>
<td>0.5</td>
<td>$1.1774\sigma_c$</td>
</tr>
<tr>
<td>Circular mean square positional error</td>
<td>MSPE</td>
<td>0.6321</td>
<td>$1.4142\sigma_c$</td>
</tr>
<tr>
<td>Circular map accuracy standard</td>
<td>CMAS</td>
<td>0.9</td>
<td>$2.1460\sigma_c$</td>
</tr>
<tr>
<td>Three-five sigma error</td>
<td>$3.5\sigma_c$</td>
<td>0.9978</td>
<td>$3.5\sigma_c$</td>
</tr>
</tbody>
</table>

come straight from the chi-square (2 degrees of freedom) density function and are directly related to the values presented above in table 1.

It should be noted that the various relationships set forth in table 2 are strictly applicable only in the circular error case; that is, when $\sigma_x=\sigma_c=\sigma_y$ and $\sigma_{xy}=0$. If the variances in the $x$ and $y$ dimensions differ (so that $\sigma_x\neq\sigma_y$), or correlation is introduced (so that $\sigma_{xy}\neq0$), then the confidence region will be an ellipse and not a circle. In this case, it is impossible to state a single circular standard error value $\sigma_c$ that can be linearly scaled by $k_\alpha$ to obtain circular confidence regions associated with the $100(1-\alpha)$% probabilities.

If it is truly the case that map errors in the $x$ and $y$ dimensions are independent ($\sigma_{xy}=0$) normally distributed random variables with $\sigma_x=\sigma_c=\sigma_y$, then the confidence region
will be a circle. In this situation either the radial distance metric from NMAS 1947, or the approach used in USGS-DLG, ASPRS 1990, and NCSSA 1994 where the errors in the \(x\) and \(y\) dimensions are treated separately, could be used to develop estimates of the proper value for \(\sigma_c\). According to Rosenfield (Thompson and Rosenfield, 1971, pages 61-62) normal USGS practice (at one time at least) was to develop approximations for \(\sigma_x\) and \(\sigma_y\) from the coordinate discrepancies at the test points in the two component dimensions \(e_x\) and \(e_y\) by means of (31).

\[
RMSE = \left[ \frac{\sum (e_{x_i}^2 + e_{y_i}^2)}{n} \right]^{1/2} = \left[ \frac{\sum e_{x_i}^2}{n} + \frac{\sum e_{y_i}^2}{n} \right]^{1/2} \approx (\sigma_x^2 + \sigma_y^2)^{1/2} (31)
\]

Implicit in (31) and this practice is the assumption that the mean discrepancies \(\mu_x\) and \(\mu_y\) are also zero. Otherwise, Maling's standard error (17) should be used. An approximation for the circular standard error \(\sigma_c\) is finally developed in accordance to (32).

\[
\sigma_x = \sigma_c = \sigma_y \quad \text{so} \quad \sigma_c = \left[ \frac{\sigma_x^2 + \sigma_y^2}{2} \right]^{1/2} \approx \frac{RMSE}{\sqrt{2}} (32)
\]

The RMSE that appears in equations (31) and (32) can be interpreted as the root mean square error from the one-dimensional distance metric or as a kind of "average" of the mean square errors in the two component dimensions \(x\) and \(y\). In this manner then, the relationship between the bivariate model and the other models has been established.
To illustrate the application of these ideas, the allowable RMSE for one-dimensional distance metric in the NMAS 1947 can be developed. For a 1:24,000 scale map the allowable horizontal error is 40 feet (based on 1/50" at publication scale). Since NMAS 1947 uses a 90% criterion the radius of the 90% circular confidence region will be determined. From table 2, this radius is obtained from the circular standard error as $2.146\sigma_c$. Combining this fact with equation (32) yields the result in (33)

$$\sigma_c \approx \frac{RMSE}{\sqrt{2}} \quad \text{and} \quad 40 \text{ feet} = 2.146\sigma_c \quad \text{so}$$

$$RMSE = \sqrt{2} \frac{40}{2.146} \approx 26.4 \text{ feet}$$

that is reported by Thompson (Thompson and Rosenfield, 1971, page 58).

Even under the assumption that errors in the $x$ and $y$ component dimensions $e_x$ and $e_y$ are independently distributed random variables, it is unlikely that empirically determined estimates for $\sigma_x$ and $\sigma_y$ would be equal. So, to reduce bivariate error to a single radial distance value will require some method of approximating the radius when $\sigma_x \neq \sigma_y$ or $\sigma_{xy} \neq 0$. The problem is one of determining the radius of a circle that contains the error with probability $100(1-\alpha)\%$ when the actual confidence region is elliptical in shape. In general, determination of the circular confidence region for a bivariate normal random variable requires the
integration of its density function over the circular shaped area of interest. Solution of the integral requires numerical methods (Leenhouts, 1985, pages 18-19).

Various methods for approximating the circular standard error when \( \sigma_x \neq \sigma_y \) are reported by Rosenfield (Thompson and Rosenfield, 1971, page 61) and elsewhere (Mikhail and Ackerman, 1976, pages 33-34). These are presented in (34).

\[
\sigma_{c1} = \left( \frac{\sigma_x^2 + \sigma_y^2}{2} \right)^{1/2} \quad \sigma_{c2} = (\sigma_x \sigma_y)^{1/2} \quad \sigma_{c3} = \left( \frac{\sigma_x + \sigma_y}{2} \right)
\]  \hspace{1cm} (34)

Citing unpublished work by Crombie (Crombie, 1967), Rosenfield suggests the \( \sigma_{c1} \) estimate in (34) for the circular standard error is the best approximation for the 90% probability circle (for all cases of \( \sigma_{\text{min}} / \sigma_{\text{max}} \) where \( \sigma_{\text{min}} \) is the smaller of \( \sigma_x \) or \( \sigma_y \) and \( \sigma_{\text{max}} \) is the larger of the two). This estimate for the circular standard error \( \sigma_{c1} \) must be scaled by \( k_4 = 2.146 \) to obtain the radius \( R = 2.146 \sigma_{c1} \) for the 90% confidence circle. If the confidence circle radius for some other probability level is desired, then one of the other methods of approximating \( \sigma_c \) might produce a better result.

To illustrate this fact consider the following case where \( \sigma_x = 2.14596603 \) and \( \sigma_y = \sigma_x / 4 \). Making the assumption of independent (\( \sigma_{xy} = 0 \)), zero mean, normally distributed errors in \( x \) and \( y \), the bivariate error \( e \) may be described as in (35).
\[ e \sim N(0, \Sigma) \] where
\[ e = \begin{bmatrix} e_x \\ e_y \end{bmatrix}, \quad 0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} 4.6051702 & 0 \\ 0 & 0.2878231 \end{bmatrix} \quad (35) \]

The ellipse whose axes are \( \sigma_x = 2.14596603 \) and \( \sigma_y = 0.53649151 \) represents the actual standard (elliptical) confidence region whose probability is about 0.39346903, or 39.35\%.

The 90\% confidence region is then an ellipse whose axes are \((\sigma_x \text{ and } \sigma_y \text{ scaled by } k_a = 2.146) \) \( k_a \sigma_x = 4.6051702 \) and \( k_a \sigma_y = 1.1512925 \). Approximations for \( \sigma_e \) were developed using the three approaches in (34). Radii of the 100(1-\(\alpha\))% confidence circles based on these three approximations were obtained by scaling \( \sigma_{c1} \) by \( k_a \) (obtained from table 2) so that \( R_{10} = k_a \sigma_{c1} \). For \( \alpha = 0.6065 \) and probability of 0.3935, \( k_a = 1 \), while for \( \alpha = 0.10 \) the value of \( k_a \) is about 2.146. The approximate radii, along with the true radius obtained from numerical integration, are provided in table 3. As claimed by Rosenfield, the \( \sigma_{c1} \) approximation is closest to the true
Table 3. Comparison of approximations for the radius of the probability circle for case when $\sigma_{xy}=0$, $\sigma_y=2.14596603$ and $\sigma_y/\sigma_x=0.25$.

<table>
<thead>
<tr>
<th>Approximation</th>
<th>$\sigma_{ci}$</th>
<th>PR=0.3935 ((\alpha=0.6065)) Radius, $R_{1\alpha}$</th>
<th>PR=0.90 ((\alpha=0.100)) Radius, $R_{1\alpha}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{c1}$</td>
<td>1.564</td>
<td>1.564</td>
<td>3.357</td>
</tr>
<tr>
<td>$\sigma_{c2}$</td>
<td>1.073</td>
<td>1.073</td>
<td>2.303</td>
</tr>
<tr>
<td>$\sigma_{c3}$</td>
<td>1.341</td>
<td>1.341</td>
<td>2.878</td>
</tr>
<tr>
<td>True Value</td>
<td>1.250</td>
<td>1.250</td>
<td>3.572</td>
</tr>
</tbody>
</table>

value of $R$ for the 90% confidence circle (which corresponds to $\alpha=0.10$). But, for the 39.35% probability associated with the circular standard error $\sigma_c$ (which corresponds to $\alpha=0.6065$) the $\sigma_{c3}$ method yields a better value. This example demonstrates that "best" approximations of the radius $R_\alpha$ of a circular confidence region depend on what $\alpha$ value and probability 100(1-$\alpha$)% are contemplated.

Leenhouts reviews other methods for approximating the radii of circular regions under the bivariate normal distribution for various probabilities and recommends an approach based on a third order polynomial (Leenhouts, 1985, pages 25-27). He provides coefficients, $q_i$, for several probability levels that are used to generate values for the radius, $R$, of the circular region. The equation is given in (36),
\[ \frac{R}{\sigma_{\text{max}}} = a_0 + a_1C + a_2C^2 + a_3C^3 \]  \hspace{1cm} (36)

where \( \sigma_{\text{max}} \) is the semimajor axis of the standard error ellipse (obtained after the covariance matrix has been diagonalized) and \( C = \sigma_{\text{min}} / \sigma_{\text{max}} \) is related to its eccentricity. For the particular case where the 90% probability circle is desired, the semimajor axis of the standard error ellipse (after diagonalization) is 2.14596603 and \( C = 0.25 \), the polynomial is given in (37).

\[ \frac{R}{\sigma_{\text{max}}} = 1.647743 - 0.062698C + 0.437189C^2 + 0.127667C^3 \]  \hspace{1cm} (37)

Solving (37) yields \( R/\sigma_{\text{max}} = 1.661387609 \) so that \( R = 3.565 \), which is close to the true value of \( R = 3.572 \).

The foregoing discussion has explored some of the statistical models applied to geometric point discrepancy measures. While the stochastic models associated with the univariate distance metric (used in the NMAS 1947) and separate discrepancy measure (that figure in USGS-DLG, ASPRS 1990, and NCSSA 1994) can be related to the bivariate normal approach applied in the vector discrepancy measure, the relationship is complex. There are no simple, exact transformations between the bivariate normal error model and the other one-dimensional measures. In general, the statistical behavior of the distance metric (7), separate discrepancy measures (8), and the vector approach (9) can be related only under special circumstances. Unless errors in
x and y are independent ($\sigma_x=0$) with equal variances ($\sigma_x=\sigma_y$), the relationships among the measures are approximate.

Another point raised in this discussion concerns the use of the rms error or RMSE (14) and the fact that it should not be used to summarize a set of geometric point discrepancies based on the one-dimensional distance metric. Though theoretically valid in the case where the discrepancy is treated as a pair of separate errors $e_x$ and $e_y$ in the component dimensions, the RMSE can overstate the variance of these errors if they do not have zero means. For this reason, Maling suggests that mean error (18) be evaluated and used in conjunction with the standard error (17) in order to obtain an indication of the existence of systematic effects or trends in the pattern of geometric point discrepancies.

**Conclusion**

The bulk of this chapter has been devoted to the problem of describing the discrepancy in point location as the basis for measuring the geometric accuracy of the point primitive. Measurement of other aspects of cartographic quality was acknowledged--such as attribute accuracy, logical consistency, and completeness--but not considered in depth. It was concluded that the geometric accuracy of the area point primitive can be largely (but not completely) reduced to a matter of logical consistency. Its geometric quality can only be evaluated on a binary scale as there are
no degrees of accuracy in its position; the location of an area point is either correct or it is wrong.

Three methods for characterizing the geometric discrepancy between two planar cartographic points were presented. These approaches included the univariate distance metric (7), the separate discrepancy measure (8) where differences for the x and y components are obtained, and the bivariate model (9). Which of these point discrepancy models to adopt depends on how helpful the measure is in detecting the lack of stability and the presence of special causes in the mapping process.

The univariate distance metric, which is the basis of the NMAS 1947, collapses the difference between two points into a single, nonnegative number. In the process some valuable information can be lost. Preferred orientations (directional biases), for example, could be masked by the distance metric. Use of the separate discrepancies, e_x and e_y, for the x and y components of the difference will enable certain patterns and preferred orientations to be detected. However, patterns that arise from correlations between the x and y components of the discrepancy may not be detected when they are treated as independent quantities. These particular limitations are not inherent in the bivariate vector model, which retains all the information about the geometric point discrepancy. For this reason, patterns, preferred orientations, and systematic effects are more
likely to be detected with the vector measure than with the other approaches.

Some difficulties were encountered in applying stochastic models to the univariate distance metric and separate discrepancy measures. The normal distribution cannot be applied to the one-dimensional distance metric. And, the differences in the x and y dimensions $e_x$ and $e_y$ must be independent when they are treated as separate quantities. In the vector discrepancy approach, a stochastic model based on the bivariate normal distribution can be consistently applied. This statistical model can accommodate correlations between the x and y components of the difference vector $(e_x, e_y)$.

In so far as the bivariate approach subsumes the distance metric and the separate discrepancy measure, retains more information than the other two measures and is amenable to a less restrictive stochastic model, it is reasonable to adopt it for the characterization of the geometric discrepancy between two planar cartographic points. Therefore, it may be concluded that geometric accuracy of the point primitive is best measured and described as a two-dimensional vector quantity.

This chapter has begun to address the ways in which cartographic objects can be measured with respect to geometric accuracy. In particular, methods for evaluating the geometric quality of the point primitive were treated in
some depth. For the most part, these approaches are well understood and used in the cartography. However, methods for measuring the geometric accuracy of line objects are not so apparent; none of the five spatial information standards reviewed in Chapter III even addressed the issue of linear quality. In the next chapter, an attempt will be made to set forth methods by which the geometric accuracy of the line primitive may be evaluated in a quantitative manner.
CHAPTER VI
LINEAR GEOMETRIC QUALITY

Introduction
From the discussion in the previous chapter it is clear that the treatment of map quality—i.e., in so far as it concerns the geometric accuracy of point data—is highly developed. Various methods were described for characterizing the discrepancy between the "true" location of the test point and its position in the cartographic data. Different stochastic models for summarizing these discrepancies were also considered. And, the various geographic information standards considered in Chapter III all present some method for describing positional accuracy of well-defined test points. But, having reduced cartographic data quality to a matter of points, lines and area points, it is quite obvious that a major control subject for cartographic data processes is the geometric accuracy of linear features. In fact, if 80% of the information on a map is in the lines, then describing geometric map accuracy solely in terms of points ignores most of the data. Clearly, some method of measuring linear geometric data quality in mapping is needed so
production processes and their output (product) can be assessed in a quantitative and objective manner.

As noted in Chapter III, this need was acknowledged by Thompson nearly 40 years ago.

It is also possible to state the horizontal accuracy in terms of the departure of "planimetric features" (rather than points) from their true positions. This makes the determination of error somewhat more complicated if the feature is an aligned feature such as a road or railway. (Thompson, 1956, page 165).

He then presents a figure similar to the one shown here in figure 18, where the dashed line indicates the true location

![Figure 18. Horizontal error for a linear feature (Thompson, 1956, page 165).](image)

and the solid line is the map location. He then sets forth three possibilities for describing the planimetric error of AB: The average distance between AB and A'B'; the maximum distance, AA'; or the angular value, A'O'A combined with the maximum difference, AA'. Thompson suggests that the last approach would be most complete but concludes that it would be too costly to determine the angular error A'O'A. In the end, he argues,
It can be assumed, therefore, that horizontal accuracy determination should invariably be based on the positions of identifiable points rather than on the location and orientation of features having length or area.  
(Thompson, 1956, page 166)

Writing in 1989, Maling noted that some work had been accomplished by the 1960’s in evaluating the accuracy of contour line data (for example, Blachut, 1964) that could be applied to other classes of linear cartographic data (Maling, 1989, page 156). "It seems, however, that little or no work has been done on these other applications" (Maling, 1989, page 156).

The Graphic Approach to Linear Quality

There is at least one context where the geometric quality of linear cartographic data is explicitly considered. Generally, whenever map overlay techniques are used, the positional accuracy of points and lines is evaluated. Map overlay approaches are often used in cases where analog maps are converted to digital form. However, in principle, they can be employed whenever an authoritative source, say an orthophoto, is available. A plot containing the linear data (generated by the mapping process) is placed on top of the source document on a light table and visually inspected. A common criterion used to judge error is the existence of areas (between the line on the check plot and the one on the source document) where light is visible (Campbell and Mortenson, 1989, page 1614). It should be
noted that this graphic overlay process and visual
inspection can be computerized. If authoritative source
information exists in computer form (say the result of
raster scanning existing hardcopy maps, digital orthophoto
production, or from a mobile mapping system using GPS/INS
technologies), they can be displayed on a video display
terminal together with the data generated by the mapping
process.

Whether based on a physical overlay or computer
display, this kind of linear data checking is based on
visual inspection by a human operator. As such, it is
heavily influenced by the subjective experience and skill of
the individual performing the evaluation. Linear error
criteria based on the appearance of light between two lines
(when overlay techniques are used) do not provide a
numerical measure of the discrepancy. Moreover, they do not
(apart from a count of the number of instances) provide a
good basis for summarization in terms of a stochastic model.
For these reasons, a more objective, quantitative method
needs to be developed for characterizing the discrepancy
between two planar lines.

This chapter will explore several approaches for
describing the discrepancy between two planar lines. Before
considering specific approaches for describing the error in
a linear feature, a method for representing lines
analytically will be set forth.
Representing Lines

There are several ways in which a straight line in the plane can be mathematically defined (Bowyer and Woodwark, 1983, pages 7-13). In general, the explicit form, as shown in (38),

\[ y = mx + b \]  \hspace{1cm} (38)

can not be used for cartographic lines. The reason is that lines on a map can not be expressed as functions of either axis alone. For instance, the vertical line from \((x,y)=(0.5,0)\) to \((x,y)=(0.5,1)\) can not be characterized as a function of \(x\). Another, more stable form is the implicit representation (Bowyer and Woodwark, 1983, page 8) shown in (39).

\[ ax + by + c = 0 \]  \hspace{1cm} (39)

For instance, the vertical line can be described as \(x=0.5\); that is, \(a=1\), \(b=0\), and \(c=-0.5\) for \(y\in[0,1]\). Implicit representations can be used for more complex lines. As an example, the graph of the function \(f(x)=x^2\) on the interval \(x\in[0,1]\) can be described in implicit form as \(x^2-y=0\), \(x\in[0,1]\). While this form may be satisfactory for some lines, it can be clumsy to use in the case of more complex lines, such as cubic curves. Implicit representations are generally not used for equations of degree higher than 2 due to the difficulty in solving them to obtain a \(y\) value for a given \(x\), and vice versa (Bowyer and Woodwark, 1983, page
76). For the general case, it is convenient to treat a line as a (real) vector-valued function of a single real parameter. The expression in (40)

\[ r(t) = (x(t), y(t)), \; t \in [a, b] \subseteq \mathbb{R} \] (40)

sets forth the general form of the parametric representation of a line. The domain of the function \( r \) is a closed interval of the real line. This form of representation provides a consistent method for defining simple (straight segments) and complex lines (sequences of segments and/or curves) that occur in mapping. For this reason, the parametric representation for defining lines will be adopted in this research.

It should be noted that in practice a linear cartographic feature is seldom so neatly represented by a simple vector-valued parametric function, especially one with a straightforward mathematical expression. Linear cartographic features are often complex in nature, consisting of straight segments and arbitrarily curved paths. The centerline of a meandering stream or topographic contour are examples of such general curves. In digital form these linear cartographic objects are modeled by a variety of geometric constructs, such as polygonal lines, circular arcs, and cubic spline functions, to mention only a few possibilities. (How linear phenomena are to be modeled in a GIS or digital spatial database is a matter of design.)
However, linear features are typically represented in a piecewise fashion; a few basic curves (circular arc, cubic spline function) are combined (concatenated) with straight line segments to depict a feature. As a notational convenience in the discussion that follows, \( r(t) \) will indicate a single simple function or the piecewise combination of several such functions.

Computerized representations of linear information (such as CAD systems as well as cartographic data) often use polygonal lines as the sole model for linear features. A polygonal line is a series of connected straight line segments. In this case, an array data structure is typically used to store the \((x,y)\) coordinates of the segment endpoints, the actual segments being implicit in the ordering of the points. A well-known application of the polygonal line format is the USGS Digital Line Graph (DLG) format where lines are exclusively defined as ordered sets of points (United States Department of the Interior, U.S. Geological Survey, page 1-10 5/88). An important property is that any reasonable cartographic line can be uniformly approximated by means of polygonal lines. And, of course, piecewise parametric representations of polygonal lines are easily developed.

A major disadvantage of a polygonal line representation is its dependence on scale. The arc of a circular curve, for example, can undergo scale changes without losing its
essential character. By contrast, the same arc approximated by a polygonal line (consisting of several small chords) can not be enlarged indefinitely before the "approximation" becomes obvious upon visual inspection. Although they may suffer disadvantages, polygonal linear structures are widely employed in cartographic applications and thus form an important class of linear information.

A significant issue that will be considered later is the choice of parametrization for the line. For example, the function suggested above, \(x^2 - y = 0, x \in [0,1]\), may be parametrized as \(r(t) = (x(t), y(t)), t \in [0,1]\) where \(x(t) = t\) and \(y(t) = t^2\). Another parametrization of this line could be based on the functions \(x(t) = t^{1/2}\) and \(y(t) = t\), with \(t \in [0,1]\). Although both representations generate precisely the same line (image) in the plane, in other respects they are not equal. The importance of a standardized parametrization will be considered below. For the time being, the two planar lines \(r_1\) and \(r_2\) to be compared will be represented in parametric form as shown in (41)

\[
\begin{align*}
\quad r_1(t) &= (x_1(t), y_1(t)) \quad \text{and} \\
\quad r_2(t) &= (x_2(t), y_2(t)), \quad \text{for } t \in [0,1]
\end{align*}
\]

where the parameter \(t\) ranges over the interval from 0 to 1.

Linear Measures and Problems

In this section several potential methods for characterizing the difference between two lines will be
investigated. The goal of this effort will be to develop a quantitative measure of the geometric quality of linear cartographic data to be used in controlling the quality of a mapping process.

**Maximum Separation**

The first measure to be considered here is based on the idea of the maximum separation or greatest distance separating two lines. In very general terms, the discrepancy between two lines may be defined as the greatest distance separating them. The idea for this measure is based on a metric of importance in topology that is induced by the following norm in (42).

\[
\|x\| = \sup\{\|r(t)\|_E: t \in [0,1]\}
\]  

(42)

Note that the E subscript in (42) is intended to denote the standard Euclidean norm for points in the plane. The "sup" (short for supremum) means the least upper bound of the set. For plane curves as defined in (41), this norm is expressed as in (43).

\[
\|x\| = \sup\{\|r(t)\|_E: t \in [0,1]\} \\
= \sup\{\|(x(t), y(t))\|_E: t \in [0,1]\} \\
= \sup\{\sqrt{x(t)^2 + y(t)^2}: t \in [0,1]\}
\]  

(43)

And, the metric induced by (43) for such lines reduces to the expression in (44).
\[ d(r_1, r_2) = \|r_1 - r_2\| \]
\[ = \sup\{\|r_1 - r_2\|_\infty : t \in [0, 1]\} \tag{44} \]
\[ = \sup\{\|(x_1(t) - x_2(t), y_1(t) - y_2(t))\|_\infty : t \in [0, 1]\} \]
\[ = \sup\{\sqrt{(x_1(t) - x_2(t))^2 + (y_1(t) - y_2(t))^2} : t \in [0, 1]\} \]

The final equality in (44) provides a formal description for the notion of the greatest distance separating two lines, \( r_1 \) and \( r_2 \).

Such a method of describing the discrepancy between two lines may hold potential for application in the context of statistical control. By signalling the maximum distance between two lines a single number for statistical summarization would be obtained. This kind of measure has been used in the evaluation of linear error.

To determine the maximum discrepancy between the two realizations of the digitizing process the vector representations of the lines were examined visually. Those parts of the sample sites where error were largest were greatly enlarged, and the maximum width between the two lines was measured.

(Dunn, Harrison and White, 1990, page 389).

This procedural description belies a significant problem with the maximum separation concept. Measuring the maximum distance between the two lines requires specification of which points on the lines are to be compared. While the intuitive sense of maximum separation may be reasonable, determining which points to compare is not as obvious as it might seem. Take for instance the simple example shown in figure 19, where the two lines are intended to be identical representations. Comparison of the lines \( r_1 \) and \( r_2 \) is
Figure 19. Maximum separation measure for lines and the problem of comparable points.

reasonably based on the distance between the points \((a_1,b_1)\) and \((c_1,d_1)\) or between \((a_2,b_2)\) and \((c_2,d_2)\). Certainly, two lines that are meant to be "close" to each other should have similar endpoints. But, someone might be inclined to suggest that the distance between the two lines should be based on the points, \((a_2,b_2)\) and \((c_1,d_1)\), or perhaps the smallest perpendicular distance separating them. The maximum separation obtained from these latter choices of points would yield a significantly smaller value than that obtained using the corresponding endpoints.

This example suggests a difficulty in applying the maximum separation measure. The problem arises from the uncertainty in specifying which pair of points from the two lines should form the basis of the maximum separation. Visual inspection and cartographic context may provide enough clues to specify the comparable points to used in a particular case. However, it is rather difficult to translate such judgment into a formal measurement criterion.
The choice of which points to compare is a significant issue associated with the maximum separation measure. This "comparable points" problem stems from the fact that a line can be parametrized in different ways. In the next section the issue of comparable points and dependence on parametrization will be explored in more detail.

The Problem of Line Parametrization

In order to demonstrate the role of the parametrization consider a single line defined in two distinct ways. The straight line segment, \( y = x \) for \( x, y \in [0,1] \), can be defined by means of the two parametrizations in (45).

\[
I_1(t) = (t, t) \text{ and } I_2(t) = (t^2, t^2), \ t \in [0,1] \tag{45}
\]

Both \( r_1 \) and \( r_2 \) in (45) represent the same line in the plane, namely \( y = x \) for \( x, y \in [0,1] \). Therefore, since the two lines are "identical" it is reasonable to expect that the maximum separation distance between them should be zero. That is, they should have a metric difference of zero. However, as evident in (46),

\[
d(r_1, r_2) = \sup\{|(x_1(t) - x_2(t), y_1(t) - y_2(t))| : t \in [0,1]| \\
= \sup\{\sqrt{(t-t^2)^2 + (t-t^2)^2} : t \in [0,1]| \\
= \sup\{\sqrt{2(t-t^2)} : t \in [0,1] | 
\tag{46}
\]

applying the maximum separation metric (44) to the "identical" lines in (45) yields a positive number for the difference. The maximum value of the function \( f(t) = 2^{1/2}(t-
for $t \in [0,1]$ is attained when $t = 1/2$, so that the value in (47)

$$d(r_1, r_2) = \sup \{\sqrt{2}(t-t^2) : t \in [0,1]\}$$

$$= \frac{\sqrt{2}}{4} \approx 0.354$$

is positive. It is of course true that $d(r_1, r_2) = 0$ when the two lines are identically parametrized, say $r_1(t) = (t, t)$ and $r_2(t) = (t, t)$, since $r_1 - r_2 = ((t-t), (t-t)) = (0, 0)$ for all $t \in [0,1]$.

That the single line segment in the plane, $y = x$ for $x, y \in [0,1]$, could possibly not be equal to itself is somewhat problematic. The reason is that the metric is comparing two functions $r_1$ and $r_2$ and not the two plane curves (image of the function). Different parametrizations of the same curve constitute different functions and therefore the metric distance between them can not be zero; $d(r_1, r_2) \neq 0$ precisely because $r_1 \neq r_2$. In general, metric measures may be sensitive to differences in the parametric representation of lines. The parametrization of the lines determines which points are compared under a given measure. In this sense, the problem of selecting comparable points—which is essentially the problem of choosing a method of parametrizing—is the cause of the problem with the maximum separation measure. Now, there are at least two ways to handle this problem. The first is suggested by Michael Godau who adopts a maximum separation type of measure, with some extra requirements.
Continuous Distance

Godau's continuous distance (Godau, 1991) metric is essentially the maximum separation idea with a further condition placed on the parametrizations of the lines. Specifically, the metric is based on those parametrizations (from all possible "reasonable" parametrizations) of the two lines that result in the smallest value (infimum, or greatest lower bound) of the maximum separation. The key point is that the metric is invariant under orientation (direction) preserving reparametrizations (Godau, 1991, page 128). The continuous distance is defined (with some alteration of Godau's original notation) as follows. Let \( r_1: [a, a'] \rightarrow V \) and \( r_2: [b, b'] \rightarrow V \), be curves in Euclidean vector space \( V \). The continuous distance \( d_c(r_1, r_2) \) is defined as

\[
d_c(r_1, r_2) = \inf_{\alpha \in [0,1]} \max_{t \in [0,1]} \| r_1(\alpha(t)) - r_2(\beta(t)) \| \quad (48)
\]

where \( \alpha \) and \( \beta \) are continuous, surjective, and increasing (that is, "reasonable") functions.

Now, in the case of the line \( y=x \), parametrized as \( r_1(t) = (t, t) \) and \( r_2(t) = (t^2, t^2) \), the continuous distance between \( r_1 \) and \( r_2 \) would be zero. In particular, the functions in (49)

\[
\alpha(t) = t \quad \text{and} \quad \beta(t) = t^2 \quad \text{for} \ t \in [0,1] \quad (49)
\]
would provide a greatest lower bound (infimum) for (48) because of (50).

\[
\max_{t \in [0,1]} \| r_1(\alpha(t)) - r_2(\beta(t)) \| = \\
\max_{t \in [0,1]} \| (t, t) - ((\sqrt{\varepsilon})^2, (\sqrt{\varepsilon})^2) \| = (50)
\]

\[
\max_{t \in [0,1]} \| (0, 0) \| = 0
\]

In this example, although \( r_1 \) and \( r_2 \) are not identical as functions, they are identical in the sense in which cartographers reckon equality; that is, they represent precisely the same planar line.

Godau does note that strictly speaking the continuous distance is not a metric. The reason is that it is possible for \( r_1 \neq r_2 \) and \( d_c(r_1, r_2) = 0 \), which violates the first condition of a metric.

By calling two curves equivalent iff their distance is zero then regarding the equivalence classes as the "true" curves and defining \( d_c \) appropriately we can realize \( d_c \) as a metric. (Godau, 1991, page 128)

Here the objects in the space are equivalence classes of parametrized curves that are "zero distance" apart under the continuous distance metric, \( d_c \). Thus, several distinct functions (namely those in a class where their mutual distances are all zero) become a single member of the space.

Rather than considering the class of all reasonable parametrizations of curves and the equivalence classes that might result, another approach is to simply adopt a
"standardized" parametrization. A standardized parametrization would impose uniformity in the way cartographic lines are represented as functions.

A Standardized Parametrization

In standardizing the parametrization of cartographic lines, the aim is to develop a set of objects whose members are meaningfully comparable under whatever particular measure of closeness is chosen. Imposing a particular parametrization on cartographic lines amounts to a formal structuring of the class of objects to be compared under the measure. In other words, the linear data to be compared are altered in such a way that parametrization is no longer an issue. In particular, the problem of comparable points is overcome by assuming a uniform method of line parametrization.

Perhaps the most conceptually direct approach to parametric representation is to tie the notion of "comparable" points to arc distance along a line. In other words, for purposes of evaluating differences among lines, all lines are standardized as vector-valued functions on the interval [0,1] in proportion to their arc length. Thus, a line is defined as in (41) but where the parameter $t$ in (51)

$$r:[0,1] \rightarrow \mathbb{R}^2 \text{ or } r(t) = (x(t), y(t)), \ t \in [0,1]$$

(51)

is the proportion of the length of the line traversed from its origin at $t=0$, $(x(0), y(0))$. The point $(x(1), y(1)),$
where \( t=1 \), is the end point of the line. The issue of determining the "comparable points" on two or more lines is solved once and for all under this approach. Comparable points on two or more lines under this standardized arc length parametrization are precisely those points for which the parameter values \( t \) are equal.

Before describing how the standardized arc length parametrization can actually be accomplished, it should be noted that cartographic lines encountered in practice will in general be rectifiable. That is, they will have finite length. If a curve \( r(t) \) has a continuous derivative on its domain, then it will be rectifiable (Kreyszig, 1988, page 469). Since cartographic features can have sharp corners (as with a rectangular parcel boundary), it may be necessary to view them in a piecewise fashion. The following development applies to the individual portions of the line over which it is well behaved (as between discontinuities or zeros in the derivative of its parametric representation).

Reparametrizing a curve in accordance with its arc length, \( s \), is relatively straightforward. The arc length for an arbitrary curve \( r(t) \) from a point \( r(a) \) on the curve is given in (52).

\[
s(t) = \int_a^t \sqrt{r'(u) \cdot r'(u)} \, du \tag{52}
\]
The total length $L$ of the curve from $r(a)$ to $r(b)$ is obtained by solving the definite integral in (52) for a particular value of $t$ as indicated in (53).

$$L = s(b) = \int_{a}^{b} \sqrt{r'(u) \cdot r'(u)} \, du$$  \hspace{1cm} (53)

Reparametrizing the curve in (40) with arbitrary parametrization (on the interval $[a, b]$ of the real line) for arc length is accomplished by the transformation in (54).

$$r_0(w) = r(\phi(w)) \quad \text{where}$$

$$\phi(w) = s^{-1}(Lw) \quad \text{and} \quad w \in [0, 1]$$

so that for $w \in [0, 1]$

$$r_0(w) = (x(\phi(w)), y(\phi(w)))$$

$$= (x[s^{-1}(Lw)], y[s^{-1}(Lw)])$$  \hspace{1cm} (54)

In equation (54) $s^{-1}(t)$ is the inverse of the arc length function in (52) and $L$ is the total length of the curve from $r(a)$ to $r(b)$ from (53). The arc length of the new function $r_0$ in (54) is proportional to parameter $w$. As $w$ varies from 0 to 1, the image point on the curve $r_0(w)$ varies from 0 to $L$. When $w = 1/4$, for example, $r_0(1/4)$ is one quarter of the distance along the curve from its beginning.

Consider again the example in (45) where two parametrizations for the line $y = x$, $x, y \in [0, 1]$ are given. These two functions would be reparametrized in accordance with their arc lengths, both of which are obviously $2^{1/2} = \cdots$
1.414. From (52), the arc length of $r_1(t) = (t, t)$ from 0 to $t$ is (55).

$$s_1(t) = \int_0^t \sqrt{r_1'(u) \cdot r_1'(u)} \, du = \int_0^t \sqrt{(1, 1) \cdot (1, 1)} \, du$$

$$= \int_0^t \sqrt{2} \, du = \sqrt{2}u \bigg|_0^t = \sqrt{2}t$$

(55)

The function needed for the reparametrization is easily obtained from (55) as shown in (56).

$$s_1(t) = \sqrt{2}t, \text{ so } s_1^{-1}(u) = \frac{u}{\sqrt{2}} \text{ and}$$

$$\phi_1(w) = s_1^{-1}(Lw) = \frac{Lw}{\sqrt{2}} \text{ for } w \in [0, 1]$$

(56)

In this way $r_1$ is transformed so that $w$ ranges from 0 to 1 in proportion to arc length $L = 2^{1/2}$ and the reparametrized line based on (54) becomes (57).

$$r_1(w) = (x[s^{-1}(Lw)], y[s^{-1}(Lw)])$$

$$= (x\left(\frac{Lw}{\sqrt{2}}\right), y\left(\frac{Lw}{\sqrt{2}}\right))$$

$$= (\sqrt{2}w, \sqrt{2}w)$$

$$= (w, w) \text{ for } w \in [0, 1]$$

(57)

Although the new line is identical to the original line, the sequence of steps helps illustrate how it is accomplished. A similar procedure achieves the reparametrization of the curve $r_2(t) = (t^2, t^2)$. Its arc length from 0 to $t$ based on (52) is developed (58).
\[
\begin{align*}
s_2(t) &= \int_0^t \sqrt{2u} \cdot \dot{r}_2(u) \, du = \int_0^t \sqrt{2u,2u} \cdot (2u,2u) \, du \\
&= \int_0^t 2\sqrt{2} \, u \, du = 2u^2 \Big|_0^t = \sqrt{2} t^2
\end{align*}
\]

(58)

The inverse function needed for the reparametrization is obtained from the arc length function as shown in (59).

\[
s_2(t) = \sqrt{2} t^2, \text{ so } s_2^{-1}(u) = \frac{\sqrt{u}}{\sqrt{2}} \text{ and }
\]

\[
\phi_2(w) = s_2^{-1}(Lw) = \frac{\sqrt{Lw}}{\sqrt{2}} \text{ for } w \in [0,1]
\]

(59)

Finally, noting that the total arc length of \(r_2\) from \(r_2(a)\) to \(r_2(b)\) is \(L = 2^{1/2}\), the reparametrized line is developed in (60).

\[
r_2(w) = (x[s^{-1}(Lw)], y[s^{-1}(Lw)])
\]

\[
= (x(\frac{\sqrt{Lw}}{\sqrt{2}}), y(\frac{\sqrt{Lw}}{\sqrt{2}}))
\]

\[
= (\frac{\sqrt{2}w}{\sqrt{2}}, \frac{\sqrt{2}w}{\sqrt{2}})
\]

\[
= (w, w) \text{ for } w \in [0,1]
\]

(60)

Unlike the situation for \(r_1\), the transformed version of \(r_2\) differs from its original functional representation. After they are standardized, the different functional representations \(r_1\) and \(r_2\) for the line \(y=x\) become identical, namely \(r(w) = (w, w), \ w \in [0,1]\). It is clear that the distance between the standardized representations will be zero under any metric because the functions are identical. These
examples show how the standardized parametrizations are obtained. In principal, any two lines that have the same paths in the plane will end up with identical standardized arc length representations (assuming the proper orientations are maintained). And, imposing this uniformity on cartographic lines is conceptually simpler than the continuous distance metric suggested by Godau.

The examples above demonstrate how the standardized arc length representation of a line is accomplished in an analytical sense. In order to use this method of standardized arc length parametrization as the basis for comparison, correspondence between lines must first be established. In other words, the matching endpoints of the lines need to be identified and the lines must be structured so their orientations (start to finish) are in agreement. A cartographic data file containing several lines must be segmented into individual line units as the basis of comparison. In the quality control context where the distance between the "produced" line is compared to its "goal" (or more accurately determined line accepted as the "truth"), establishing the relationship between test line and the "map" should be a fairly straightforward task for a human operator. In general, points on the two lines that correspond (or at least should correspond) are easily identified. For example, road intersections, confluences of streams, distinctive prominences, intersections with other
features (lines) and so on, could be used to unambiguously
determine the beginning and ending points on the lines to be
compared. If the cartographic database incorporates some
topological structure, then the identification of linear
units for comparison should be straightforward.

There are, however, situations where reasonable point
correspondence for parametrization might be difficult, if
not impossible to determine. Consider a smooth, closed line
feature, such as an elevation contour or the shoreline of a
lake. Assuming the feature lacks an obvious prominence or
other reasonable candidate (perhaps intersection with some
other feature), how does one establish a point of origin for
the parametrization? A simple rule, such as setting the
extreme northernmost point as the line's origin, might be
possible in some situations.

It is believed that the standardized arc length
parametrization of linear cartographic data is a reasonable
approach to take with respect to the problem of multiple
parametric representations and the definition of comparable
points. In the context of controlling the quality of a
mapping process, the two lines to be compared or measured
are intended to be essentially "equal". Any significant
discrepancy occurring between the two lines because of the
nature of the parametrization (especially the arc length)
would tend to cast such presumed "equality" into doubt!
Moreover, the length of cartographic lines is itself a
subject of significance in evaluating geometric data quality. Parametrization on the basis of arc length would provide rather obvious cues of any major discrepancy between lines. For these reasons, the standardized arc length parametrization will be used as the basis for representing linear cartographic features in terms of a vector-valued parametric function.

$L_p$ Metric

Once the standardized arc length method of curve representation is adopted, it is possible to consider possible metric comparisons among lines. The idea of maximum separation or greatest distance between two lines was presented above. A significant aspect of such a metric is that it reveals the "worst" case discrepancy between the lines. Although the maximum separation measure does not reveal much about the "average" discrepancy over the length of the whole line, it is a member of a class of metrics suggested by Saalfeld for comparing the closeness of two lines. The metrics are known as $L_p$ metrics and are defined as follows.

If two curves are expressed as functions $r_1(t)$ and $r_2(t)$ parametrized by the same normalized parameter $t$, then there are several fundamental measures for describing distance between the curves which are known in function theory as the $L_p$ measures [(61)]:

$$\|r_1 - r_2\|_p = \left[ \int \|r_1(t) - r_2(t)\|^p dt \right]^{1/p} \quad (61)$$
where the norm \( \| \| \) within the integral refers to a measure of distance between two points in the plane.

(Saalfeld, 1986, page 147, notation changed)

The expression (61) yields a valid metric when \( p \geq 1 \).

Assuming the standard Euclidean distance norm is used, the maximum separation measure is obtained from the \( L_p \) metric when \( p \) tends to infinity and is called \( L_\infty \). When \( p=1 \), the integral in (61) amounts to an average of the (positive) distances between comparable points on the two lines.

Of this potentially large class of measures, Saalfeld considers the case where \( p=2 \) and the norm is the standard Euclidean distance norm for points in the plane. He assumes also that the lines are parametrized on the closed interval, \([0,1]\). Under these conditions, given the two lines in (41), the \( L_2 \) measure for their difference is expressed by (62).

\[
d(r_1, r_2) = \left( \int_0^1 \| x_1(t) - x_2(t) \|^2 dt \right)^{1/2}
\]

(62)

\[
= \left( \int_0^1 [ (x_1(t) - x_2(t))^2 + (y_1(t) - y_2(t))^2 ] dt \right)^{1/2}
\]

The integration in (62) over the interval \( t \in [0,1] \) amounts to an average of the squared distances between the image points of the two functions (Saalfeld, 1986, page 148). As a consequence, it is reasonable to designate the \( L_2 \) metric as the average squared separation measure.

Of course, this average squared separation measure \( (L_2) \), as with the maximum separation measure \( (L_\infty) \), is
dependent upon the particular parametrization. That is, if the parametrizations are changed, then it is likely that a different value for the distance between two curves will be obtained. Of course, this situation is expected because distinct functions (arising from different parametrizations of the same line) are not identical when they are compared under a metric. This fact underscores the requirement for a standardized parametrization of lines (such as the arc length approach advocated in the previous section) if measures are based on metric comparisons.

The average squared separation metric has the appealing property that it provides an indication of the closeness of entire lines in an average sort of way. Unlike the maximum separation metric that signals the "worst" deviation of a line from a known standard, the average squared separation describes the nature of the discrepancy from the standard over the whole course of the line. As with the maximum separation metric, the average separation measure yields a single number that can be used as the basis for statistical summary. For these reasons then, this measure would seem to have some potential in the statistical control of quality for linear geometric data.

Actual use of this metric has been suggested by Saalfeld for quantifying the similarity between curves (Saalfeld, 1986, page 148). More importantly, Saalfeld provides a straightforward computational method for
computing the $L_2$ metric in the case of polygonal curves. Assume the two lines (41) to be compared are polygonal curves. That is, they are defined as a sequence of straight line segments and are parametrized for $t \in [0, 1]$. Although Saalfeld does not require that the lines have been standardized with respect to their arc lengths, it is nonetheless prudent to assume a standardized arc length parametrization. Finally, it is necessary to identify all those values of the parameter $t$ where either curve possesses a segment endpoint. In other words, the set of parameter values $\{t_0 = 0, t_1, t_2, \ldots t_n = 1\}$ correspond to the places where at least one of the two curves changes direction. The two lines can thus be expressed as sequences of coordinates as in (63).

$$I_1 = \{ (x_1(t_0), y_1(t_0)), (x_1(t_1), y_1(t_1)), \ldots, (x_1(t_n), y_1(t_n)) \}$$

$$I_2 = \{ (x_2(t_0), y_2(t_0)), (x_2(t_1), y_2(t_1)), \ldots, (x_2(t_n), y_2(t_n)) \}$$

Evaluation of the average squared separation metric (Saalfeld, 1986, page 152) for the lines in (63) reduces to (64).
\[ d^2(r_1, r_2) = \int_0^1 \| r_1(t) - r_2(t) \|^2 dt \]

\[ = \sum_{i=0}^{n-1} \frac{t_{i+1} - t_i}{3} \left\{ \left[ x_1(i+1) - x_2(i+1) \right]^2 + \left[ x_1(i+1) - x_2(i) \right] \left[ x_1(i) - x_2(i) \right] \right\} \]

\[ + \sum_{i=0}^{n-1} \frac{t_{i+1} - t_i}{3} \left\{ \left[ y_1(i+1) - y_2(i+1) \right]^2 + \left[ y_1(i+1) - y_2(i) \right] \left[ y_1(i) - y_2(i) \right] \right\} \]  

(64)

Note that in (64), \( x_j(k) \) and \( y_j(k) \) are shorthand expressions for \( x_j(t_k) \) and \( y_j(t_k) \), respectively. Derivation of this result hinges on the algebraic simplification, \( (a^3-b^3) = (a-b)(a^2+ab+b^2) \). Either \( d^2(r_1, r_2) \) or \( d(r_1, r_2) \) from (64) may be used as a numerical value of the closeness of the two polygonal lines. The average squared separation measure would be computationally expedient in the case of polygonal line data.

In concluding this section, it will be noted that the \( L_2 \) metric is analogous to the idea of the root mean square error. In Chapter V, the RMSE was considered in the evaluation of error in the point primitive. The RMSE amounts to an average of squared errors which is essentially equivalent to the average squared separation measure (or \( L_2 \) metric) presented here for linear data. The correspondence can be seen by replacing the summation in the RMSE in (14) with the integral in the definition of the average squared separation metric in (62). While the RMSE has a statistical
connotation in its use in summarizing point discrepancies, such an interpretation may not be appropriate with respect to the $L_2$ metric considered here.

**Hausdorff Metric**

All of the measures considered so far—maximum separation, continuous distance, and average squared separation—to the extent they are based on parametric representations, reflect the structure (namely order) in the collection of points that constitute the line. A more fundamental definition of a cartographic line would be simply a collection of points in the plane. Such a definition avoids the structure implicit in the ordering of the points. By neglecting this structure, problems associated with parametrization would be overcome.

One common measure of closeness for arbitrary point sets is the Hausdorff metric. Assume the two lines to be compared are conceived as point sets, $R_1$ and $R_2$ as in (65)

$$R_1 = \{(a, b) : (a, b) \in R_1\} \text{ and } R_2 = \{(c, d) : (c, d) \in R_2\} \quad (65)$$

rather than as parametrically defined functions in (41). The Hausdorff distance (Preparata and Shamos, 1985, page 217) from $R_1$ to $R_2$ is given in (66).

$$d_h(R_1, R_2) = \max_{(a, b) \in R_1} \min_{(c, d) \in R_2} \|(a, b) - (c, d)\| \quad (66)$$

To illustrate the application of this measure, suppose the two point sets, $R_1$ and $R_2$, have $n$ and $m$ elements,
respectively. That is, they are finite in size and can be listed as in (67).

\[ R_1 = \{(a_1, b_1), (a_2, b_2), \ldots, (a_n, b_n)\} \]
\[ R_2 = \{(c_1, d_1), (c_2, d_2), \ldots, (c_m, b_m)\} \]  

(67)

Evaluation of \( d_H(R_1, R_2) \) entails first finding the particular \((c_{10}, d_{10}) \in R_2\) for each \((a_i, b_i) \in R_1\) that minimizes the Euclidean distance between \((a_i, b_i)\) and \((c_{10}, d_{10})\). Then, from this set of \( n \) distances (one for each \((a_i, b_i) \in R_1\) is obtained) pick the largest.

This distance is not a metric for it fails to be symmetric. In figure 20, \( d_H(R_1, R_2) \) is the distance between

![Figure 20](image.png)

Figure 20. Example of asymmetry of the Hausdorff distance.

\((a_i, b_i)\) and \((c_i, d_i)\) but it does not equal the distance \( d_H(R_2, R_1) \) which is based on the points \((a_0, b_0)\) and \((c_2, d_2)\). The Hausdorff metric between the two point sets \( R_1 \) and \( R_2 \) is taken as the maximum value of the two distances, \( d_H(R_1, R_2) \)
and $d_n(R_2, R_1)$ (Preparata and Shamos, 1985, page 217). In more general terms, the Hausdorff metric is defined (Shephard and Webster, 1965, page 74) on an arbitrary metric space $(X, d)$ as (68).

$$d^H = \max\{ \sup_{x_2 \in X_2} \inf_{x_1 \in X_1} d(x_1, x_2), \sup_{x_1 \in X_1} \inf_{x_2 \in X_2} d(x_2, x_1) \}$$

(68)

In the example in figure 20, the Hausdorff metric $d^H(R_1, R_2)$ would be the distance between the points $(a_1, b_1)$ and $(c_1, d_1)$, as this distance would be maximum.

Godau suggests that the Hausdorff metric is a poor measure of the geometric similarity between lines, precisely because information in the ordering (or structuring) of points in lines is not reflected (Godau, 1991, page 127-128). Application of the Hausdorff metric—or any other abstract measures for distances between arbitrary point sets—in the control of the geometric quality of cartographic lines is not promising. The primary problem in applying this kind of measure is its failure to reflect the structure inherent in the ordering of the points in the line. The problem is that "closeness" between two linear features may not be based on a good choice of points from each line. To evaluate the geometric quality of a linear feature, it is necessary to reflect the structure (orientation and ordering) of the line and base the comparison on points that are "reasonably comparable."
While measures based on abstract point sets may not be applicable to controlling the geometric quality of cartographic linework, there are other measures that do not rely on parametric representations. The path of the line in the plane is the sole basis for these measures that will be considered in the next section.

**Area-Based Measures**

Like the Hausdorff metric just considered, area-based measures do not require the expression of the line in parametric form. For this reason, they are not sensitive to the nature of the parametrization. One area-based discrepancy approach is to determine the size of the area between the two lines to be compared. Consider the example shown in figure 21. If the corresponding terminal points on

![Diagram of Map Line and True Line](image)

*Figure 21. Areal separation measure.*
the two lines are connected by segments, then a chain of areas can be inferred. The size of the total of all these areas, $A_1+A_2+A_3$ gives an indication of the discrepancy between the true line and the map line. However, it is necessary to divide this total area by the length of the true line, $L_T$, so the areal separation for the lines in figure 21 becomes $(A_1+A_2+A_3)/L_T$.

Total area must be divided by length so the discrepancy measure is not influenced by how long the line is. If the area is not normalized in some way, the result might be misleading. As illustrated in figure 22, the areal separation measure can be influenced by line length. The lines in the right hand pair in figure 22 are obviously closer than the ones on the left despite the fact the latter enclose a smaller area. When these two areas are normalized for length, the discrepancy in the left hand pair is 1 (length=1) while the measure for the right hand pair is 0.5

![Figure 22. Need for normalizing the areal separation measure.](image)
(based on length = 4). This example shows that area must be normalized in order to yield a satisfactory measure of linear discrepancy. There may be circumstances in which it is difficult to decide which line to choose as the divisor. For example, two equally valid lines may be compared for similarity. In such cases, an arbitrary choice could be made or the average length of the lines might be used. In any case, area, without normalization, would be unsatisfactory.

It should be noted that this areal separation measure, in general, does not satisfy the requirements of a metric. The example in figure 23, though somewhat contrived,

![Figure 23](image)

Figure 23. Example where areal separation measure fails to satisfy the triangle inequality.

illustrates how it can fail to satisfy the triangle inequality. Assume the true length for normalization is 4
in this example and that endpoints are connected by lines to create the enclosed areas. When \(R_1\) and \(R_2\) are compared the result is \(d(R_1, R_2) = 16/4 = 4\). But, \(d(R_1, R_3) = d(R_3, R_2) = 6/4 = 1.5\) so that \(d(R_1, R_2) > d(R_1, R_3) + d(R_3, R_2)\), which is a violation of the requirement for a metric. While the example in figure 23 may be cartographically farfetched, it does demonstrate that the areal separation measure (in general) is not a metric.

Nevertheless, the areal separation measure outlined above does provide an intuitively meaningful way of characterizing the discrepancy between two lines. Lines that are close together will tend to enclose a small area while lines that are widely separated will have a large area between them. And, it has the advantage of being independent of parametric representation. The comparison of two lines is based solely on their geometric appearance in the plane and not on their functional form. This kind of measure has been used in an empirical study of discrepancies between lines forming polygon boundaries (Dunn, Harrison and White, 1990, page 390).

There are, however, some practical difficulties associated with its use. Some sort of rule is required concerning the handling of areas at the ends of the lines; the approach suggested here is to join corresponding endpoints with line segments to form the terminal areas. Another complication may arise in determining the true
linear length by which to scale total area, as in the case of lines whose accuracy is equally valid. The geometric configuration of the lines themselves may cause difficulty. For example, it would be convenient to restrict consideration of the area-based measure to "simple" curves, that is, curves that do not intersect themselves. Curves with multiple points (self-intersections) could lead to ambiguities in computing areas.

But even if curves with multiple points are not considered, ambiguity could still arise in determining how to evaluate areal separation between two lines. What area to reckon (A1, A2, or A3) in the case of small closed lines (such as occur in contour coverages at peaks) as shown in figure 24 may not be obvious and would almost certainly involve a heavy computational burden when implemented.

Developing general purpose computational rules to apply in all circumstances that might arise in practice could be difficult. Calculating the areas of the chain of polygonal
areas formed by two lines requires the determination of all points where they intersect. For two fairly similar polygonal lines, numerous intersections would arise and have to be computed and then administered to build the polygons the form the areas. Such an operation would surely be non-trivial to implement.

Another area-based approach to linear discrepancy measurement can be devised that does not suffer these particular limitations. Tveite and Langaas (Tveite and Langaas, 1995) describe measures based on line buffering and subsequent overlay analysis. Their method is illustrated in figure 25. It involves a buffering operation with radius $bs_i$ on both the true line $Q$ and the line in the data $X$. An overlay operation is then performed on the polygons, $Xbs_i$ and $Qbs_i$, that result from the buffering. Various measures
of discrepancy are developed based on the area \((A)\) and number of areas \((\#A)\) in the overlay data set \(XQbs_i\). In particular, Tveite and Langaas define average displacement \((DE)\) and oscillation \((O)\) as in (69).

\[
DE = bs_i \cdot \frac{A(Xbs_i \cap Qbs_i)}{A(Xbs_i)} \\
O = \frac{\#A(Xbs_i \cap Qbs_i)}{\text{Length}(X)}
\]

(69)

They indicate that \(DE\) is the lower bound of the average displacement of a line \((X)\) with respect to the true line \((Q)\) and that \(O\) can be used as an indication of bias.

Perhaps the greatest appeal of the method suggested by Tveite and Langaas is the ease with which it could be implemented in a vector GIS system. In general, however, area-based measures may not always provide good indications of the difference between two lines. Intuitively, if two curves are close, the area between them and the average displacement \((DE)\) will be small. But, as Saalfeld notes,

"The area between the curves may nevertheless be small if one or both of the curves have spikes; hence, the area measure is not always the best measure of closeness of curves."

(Saalfeld, 1986, page 147)

Consider the two lines shown in figure 26. Since they are very close, their difference, as measured by either of the
two area-based approaches considered in this section, would
be small. However, they are rather dissimilar lines. Even
though points A and B in figure 26 are not far apart in a
spatial sense (2 \epsilon \text{ where } \epsilon \text{ is small}), they are separated by a
significant distance along the line. This example
illustrates that an area-based measure can not be blindly
applied to linear data. The reason is neither the areal
separation nor the average displacement (DE) takes into
account the structure (order) of points along the line.

From this discussion, it may be concluded that the
area-based measures may not be good candidates for use in
controlling the geometric quality of cartographic lines.
Their chief advantages stem from conceptual simplicity,
independence of the functional (parametric) representation,
and (at least in the case of the average displacement) the
ease of implementation. These advantages are eclipsed by
the fact that they can be misleading.

One important characteristic of all the measures
considered so far is the reduction of linear discrepancy to
a single real number. Single values will reflect the relative magnitude of the difference among two lines but may not capture any information on the "direction" or spatial character of the discrepancy. A major aim in the control of cartographic processes would surely be detecting special causes that introduce locational bias in the output. It would be very important to identify small but systematic deviations in the locations of linework. For instance, all the linear features being mapped may be 2 meters to the northwest of their correct locations. Though such a deviation is so small that design specifications are not compromised, it represents an "uncentered" process suffering from special causes. For this reason then, a simple metric characterization or a single number may not be adequate. In the next section, an attempt will be made to set forth measures that can account for some degree of directional bias.

Vector Component Measures

As was the case in evaluating point discrepancies, a desirable property of a method for assessing linear geometric quality is that it be sensitive to biases or preference in directional error. In order to retain more information about the nature of the discrepancy in point data, it was necessary to consider differences in both the x and y dimensions. It was noted earlier that three of the geographic information standards reviewed in Chapter III,
USGS-DLG, ASPRS 1990, and NCSSA 1994, evaluate point error in terms of the separate discrepancies $e_x$ and $e_y$ obtained in the $x$ and $y$ dimensions. An analogous approach can be adopted in the case of lines.

Perhaps the simplest such method for comparing two planar lines would be the "average" deviations of their $x$ and $y$ components. To set this notion in more formal terms, suppose the two lines to be compared are expressed in parametric equations as in (41) and have been standardized with respect to arc length. The components of an average or mean (vector) discrepancy vector, $(e_x, e_y)$, are developed in accordance to (70).

$$
e_x = \int_{0}^{1} (x_1(t) - x_2(t)) \, dt$$

$$= \int_{0}^{1} x_1(t) \, dt - \int_{0}^{1} x_2(t) \, dt, \text{ and}$$

$$e_y = \int_{0}^{1} (y_1(t) - y_2(t)) \, dt$$

$$= \int_{0}^{1} y_1(t) \, dt - \int_{0}^{1} y_2(t) \, dt$$

As with other measures based on parametric representation, the vector $(e_x, e_y)$ that results from (70) is dependent on the particular functional form of the lines $r_1$ and $r_2$. The line $y=x$ for $x, y \in [0,1]$ defined and parametrized on $t \in [0,1]$ in the two ways in (45) demonstrates this fact. Performing the integrations in (70) for the lines in (45), one obtains
\((e_x, e_y) = (1/6, 1/6)\), when in fact the lines trace identical paths in the plane. It is because of this dependence that a standardized arc length parametrization has been assumed.

To illustrate application of this measure consider the two curves shown in figure 27 and defined in (71).

\[
I_1(t) = (x_1(t), y_1(t)) = (t, \frac{1}{2}), \text{ and}
\]

\[
I_2(t) = (x_2(t), y_2(t)) = (t, \frac{1+2t}{4}) \text{ with } t \in [0, 1]
\]

They are in standardized arc length parametric form.

Components of the mean discrepancy vector, \(e_x\) and \(e_y\) can be computed as in (72).

\[
e_x = \int_0^1 (x_1(t) - x_2(t)) \, dt = \int_0^1 (t - \frac{1+2t}{4}) \, dt = 0,
\]

\[
e_y = \int_0^1 (y_1(t) - y_2(t)) \, dt = \int_0^1 (\frac{1}{2} - \frac{1+2t}{4}) \, dt = \frac{t-t^2}{4} \bigg|_0^1 = 0
\]

The result is \((e_x, e_y) = (0, 0)\), suggesting the two lines are as similar as they could get under the average vector component separation measure. But, of course the lines are quite dissimilar, as indicated in figure 27. The averaging that
Figure 27. Example where average discrepancy vector is misleading.

obliterates the distinction between these two lines is reasonable in the sense that, overall, the lines are not too unlike. However, this example suggests the average discrepancy vector could be misleading in certain situations.

In the next example the average vector component separation measure would faithfully reflect the difference between the two lines. Consider the two lines $r_1$ and $r_2$ depicted in figure 28 and defined as in (73).

\[ r_1(t) = (x_1(t), y_1(t)) = (t, \frac{1}{2}), \text{ and} \]
\[ r_2(t) = (x_2(t), y_2(t)) = (t + \frac{1}{4}, \frac{1}{4}) \text{ with } t \in [0, 1] \]  

(73)

They are in standardized parametric form, and appear as two lines separated by constant offsets in $x$ and $y$. Computation
Figure 28. Example where the average discrepancy vector is a reasonable measure of linear difference.

of the components of the average vector separation is accomplished in (74).

\[
e_x = \int_0^1 (x_1(t) - x_2(t)) \, dt = \int_0^1 (t - (t + \frac{1}{4})) \, dt = -\frac{t}{4} \bigg|_0^1 = -\frac{1}{4},
\]

\[
e_y = \int_0^1 (y_1(t) - y_2(t)) \, dt = \int_0^1 (\frac{1}{2} - \frac{1}{4}) \, dt = \frac{t}{4} \bigg|_0^1 = \frac{1}{4}
\]

The result, \((e_x, e_y) = (-1/4, 1/4)\) faithfully reflects the difference between the two lines. The two preceding examples suggest that application of the average separation vector might be useful in detecting systematic patterns or trends in the relationship between two lines. For this reason, the average separation vector may serve as a good measure of how well the process is centered. However, this measure can be misleading in the same way any average measure can when extremes balance out. Care in its use would be necessary to avoid faulty conclusions about the geometric quality of a line.
Computation of the components of the average separation vector in accordance with the expressions in (70) for polygonal line data is a rather straightforward proposition. Assume the two lines (41) to be compared are polygonal curves and have been parametrized with respect to their arc lengths. Since the two lines are defined in terms of a chain of straight line segments, they can be expressed as sequences of coordinates as in (75).

\[ r_1 = \{(x_1(t_0), y_1(t_0)), (x_1(t_1), y_1(t_1)), \ldots, (x_1(t_n), y_1(t_n))\} \]
\[ r_2 = \{(x_2(t_0), y_2(t_0)), (x_2(t_1), y_2(t_1)), \ldots, (x_2(t_m), y_2(t_m))\} \] (75)

Evaluation of the average separation vector components \( e_x \) and \( e_y \) for the polygonal lines in (75) reduces to the expressions in (76).

\[ e_x = \int_0^1 (x_1(t) - x_2(t)) \, dt = \int_0^1 x_1(t) \, dt - \int_0^1 x_2(t) \, dt \]
\[ = \sum_{i=0}^{n-1} \frac{t_{i+1} - t_i}{2} [x_1(i+1) + x_1(i)] - \sum_{i=0}^{m-1} \frac{t_{i+1} - t_i}{2} [x_2(i+1) + x_2(i)] \] (76)

\[ e_y = \int_0^1 (y_1(t) - y_2(t)) \, dt = \int_0^1 y_1(t) \, dt - \int_0^1 y_2(t) \, dt \]
\[ = \sum_{i=0}^{n-1} \frac{t_{i+1} - t_i}{2} [y_1(i+1) + y_1(i)] - \sum_{i=0}^{m-1} \frac{t_{i+1} - t_i}{2} [y_2(i+1) + y_2(i)] \]

Note that in (76) \( x_j(k) \) and \( y_j(k) \) are shorthand expressions for \( x_j(t_k) \) and \( y_j(t_k) \), respectively. These formulas are
computationally convenient for comparing polygonal line data by means of the average separation vector.

The average vector separation measure indicates how well a line may be "centered" with respect to the true line. But, another measure is required in order to make a valid determination concerning the geometric quality of a line. One approach for overcoming the limitations in the average separation vector, \((e_x, e_y)\), would be to compute the average squared deviations of the vector components, \(e_x^2\) and \(e_y^2\). The formalization of this idea was already presented earlier in the discussion of the \(L_2\) metric.

The \(L_2\) metric was computed in accordance with (62) for the lines in (41) and by means of (64) for polygonal line data. The components of the average squared discrepancy vector \((e_x^2, e_y^2)\) for the lines in (41) is accomplished by splitting the integral in (62) into two integrals to obtain (77).

\[
e_x^2 = \int_0^1 (x_1(t) - x_2(t))^2 dt, \tag{77}
\]
\[
e_y^2 = \int_0^1 (y_1(t) - y_2(t))^2 dt
\]

The result from (64) can also be applied in the case of polygonal lines to obtain a computationally convenient formula as shown in (78).
\[ e_x^2 = \sum_{i=0}^{n-1} \frac{t_{i+1} - t_i}{3} \left\{ \left[ x_1(i+1) - x_2(i+1) \right]^2 + \left[ x_1(i+1) - x_2(i+1) \right] \left[ x_1(i) - x_2(i) \right] + \left[ x_1(i) - x_2(i) \right]^2 \right\} \]

\[ e_y^2 = \sum_{i=0}^{n-1} \frac{t_{i+1} - t_i}{3} \left\{ \left[ y_1(i+1) - y_2(i+1) \right]^2 + \left[ y_1(i+1) - y_2(i+1) \right] \left[ y_1(i) - y_2(i) \right] + \left[ y_1(i) - y_2(i) \right]^2 \right\} \]

(78)

The points \( t_0 = 0, t_1, t_2, \ldots, t_n = 1 \) are where either curve changes direction under the standardized arc length parametrization on \( t \in [0,1] \). As in (64), the terms \( x_j(k) \) and \( y_j(k) \) in (78) are abbreviations for \( x_j(t_k) \) and \( y_j(t_k) \), respectively.

To illustrate the application of the average squared separation vector, consider again the line pairs defined in (71) and (73) and depicted in figures 27 and 28, respectively. Computation of the components of the average squared separation vector for the lines in (71) is accomplished according to (79).

\[ e_x^2 = \int_0^1 (x_1(t) - x_2(t))^2 dt = \int_0^1 (t-t)^2 dt = 0, \]

\[ e_y^2 = \int_0^1 (y_1(t) - y_2(t))^2 dt = \int_0^1 (\frac{1}{2} \cdot \frac{1+2t}{4})^2 dt = \frac{-(1-2t)^3}{96} \bigg|_0^1 = \frac{1}{48} \]

(79)

The resulting vector is \( (e_x^2, e_y^2) = (0, \frac{1}{48}) \), which indicates a discrepancy in the \( y \) dimension only. Recall that the mean separation vector for these lines is \( (e_x, e_y) = (0,0) \). The components of the average squared deviation vector for the pair of lines (73) is determined in (80).
\[ \begin{align*}
e^2_x &= \int_0^1 (x_1(t) - x_2(t))^2 dt = \int_0^1 (t - (t + \frac{1}{4}))^2 dt = \frac{t}{16} \bigg|_0^1 = \frac{1}{16}, \\
e^2_y &= \int_0^1 (y_1(t) - y_2(t))^2 dt = \int_0^1 \left(\frac{1}{2} - \frac{1}{4}\right)^2 dt = \frac{t}{16} \bigg|_0^1 = \frac{1}{16}
\end{align*} \] 

Here the resulting vector, \((e_x^2, e_y^2) = (1/16, 1/16)\) suggests the discrepancy arises due to differences in both the \(x\) and \(y\) dimensions.

The two lines in figure 28 are probably closer than the ones in figure 27. However, the latter pair have zero mean separation and an average squared deviation that is smaller than the pair in figure 28. It appears as if the results of the two measures in these examples may not coincide with the intuitive judgment that the pair of lines in figure 28 is closer than the pair in figure 27. One way to reconcile these results with common sense would be to subtract the squares of the components of the average separation vector from the respective components of the average squared discrepancy vector. In this case, the pair of lines in figure 28 would have a zero \((0,0)\) difference which is smaller than the discrepancy of \((0,1/48)\) for the pair in figure 27. By considering this difference of the two measures, the discrepancy between the lines in figure 28 is smaller than it is for the lines in figure 27.

The approach outlined in the preceding paragraph can be loosely related to the standard error statistic (17) presented in Chapter V. Computation of the components of
the average squared separation vector (77) is analogous to computing the RMSE in (14) separately for the x and y discrepancies in the case of point data. And, the components of the average separation vector from (70) can be treated as average differences in the x and y dimensions. By subtracting the square of the average separation in x, \((e_x)^2\), from the average squared separation for x, \(e_x^2\), a quantity akin to the standard error for the x dimension is obtained. A similar result obviously applies for the y dimension. In this way, the vector component discrepancy measures presented in this section for lines correspond to the treatment of point differences as separate deviations in x and y that was outlined in Chapter V.

The two vector component measures, when taken together, do provide more information about the nature of the difference between two lines than either measure alone. In figure 28, the difference between the two lines is attributable to the mean separation since the components of the average squared separation vector and the square of the components of the mean separation vector are identical. This situation suggests the presence of a constant offset between the two lines. For the pair of lines in figure 27, the mean separation vector is zero, suggesting no simple offset between the lines.

Such measures might be useful in distinguishing the effects of bias from inherent variability with which the
particular mapping process is creating the linear data. A special cause in the mapping system may result in specific trends or bias in the cartographic product. The average vector separation measure would provide some insight into the presence of bias in the linear data, while the average squared vector separation measure could shed light on the overall size of the error in the curve data. In any case, the average vector component measures capture more information about the nature of the difference between two lines than any of the metric or single-valued measures considered in the preceding sections. Single-valued measures of closeness for lines may fail to provide a sufficiently complete characterization of differences to be useful.

A significant issue that has not been addressed to this point is how to summarize multiple occurrences of a linear discrepancy measure. Aggregating several instances of a metric or single-valued discrepancy measure is conceptually simpler than it would be for the vector measures presented in this section. In either case, however, some form of weighting will be necessary. It would be unreasonable to treat the discrepancy measure associated with a short linear feature on an equal footing with the discrepancy associated with a longer feature. A natural way in which the results of the measures from several lines could be combined would be to use each line's arc length as a weighting factor. In
this manner then, results for long lines would have a greater influence on the final statistic than the values obtained from shorter lines. In the next section, a method for comparing lines from the perspective of a set of discrete points will be presented. Under this approach, discrepancy measures for individual lines are not generated so that the problem of aggregating (weighting) the results for several lines does not arise.

**Discrete Linear Sampling**

In the quality control context, the principal concern when comparing two lines is the extent to which the "error" exhibits characteristics that warrant intervention in the mapping process. Determining whether the pattern of variability (or error) in the mapping process is stable (random) is a major goal of any method for evaluating the geometric quality of cartographic lines. The various methods considered to this point for characterizing the difference between two lines regard the individual line as the basic unit. Each of the measures collapses the nature of the discrepancy over the entire length of the linear feature into a metric value, single number, or vector quantity.

When a single value is created, as in the maximum separation, continuous distance, $L_2$ metric, Hausdorff metric, and areal separation measures, information about the nature of the difference between two lines can be lost. In
the vector component approach, patterns in the discrepancy between two lines can be detected. For instance, when the average discrepancy vector is not zero, \((e_x, e_y) \neq 0\), an offset (or translation) between the true line and the line on the map may be responsible. The vector component measures considered in the preceding section can be viewed more generally in terms of a bivariate error function.

Recall from Chapter V that when point discrepancies are regarded as bivariate quantities, error vectors in (9) are obtained. In the linear case, assuming that the standardized arc length parametrization is employed for the lines as defined in (41), the bivariate discrepancy is a parametrically defined error function (81).

\[
e(t) = r_1(t) - r_2(t) \\
= (x_1(t), y_1(t)) - (x_2(t), y_2(t)) \\
= (x_1(t) - x_2(t), y_1(t) - y_2(t)), \text{ for } t \in [0, 1]
\]  

Consider the two lines shown in Figure 29 and defined in (82).

\[
r_1(t) = (t, \frac{1}{2}) \text{ for } t \in [0, 1], \text{ and }
\]

\[
r_2(t) = \begin{cases} 
\left( \frac{3}{2} t, \frac{1-t}{2} \right), & t \in [0, \frac{1}{2}] \\
\left( \frac{t+1}{2}, \frac{3t-1}{2} \right), & t \in \left[ \frac{1}{2}, 1 \right]
\end{cases}
\]

They are in standardized arc length form so the error function can be developed as shown in (83).
\[ e(t) = \begin{cases} 
\left( -\frac{t}{2}, \frac{t}{2} \right), & t \in \left[ 0, \frac{1}{2} \right] \\
\left( \frac{t-1}{2}, \frac{2-3t}{2} \right), & t \in \left[ \frac{1}{2}, 1 \right]
\end{cases} \quad (83) \]

The two lines in (82) and the resulting error function are depicted in figure 29. As with point discrepancies, the bivariate error function, \( e(t) \), contains a complete characterization of the difference between the two lines.

![Figure 29. Example of the parametric linear error function associated with the difference between two lines.](image)

The vector component measures discussed in the previous section--average separation vector (70) and average squared separation vector (77)--are reasonable ways to summarize the information contained in the bivariate linear error function \( e(t) \). If several comparisons between lines on the map and the true lines are made, the result will be a collection of bivariate error functions, \( e_1(t), e_2(t), \ldots, e_n(t) \), as illustrated in figure 30.
Figure 30. Illustration of the difficulty of interpreting multiple occurrences of parametric linear error functions.

What the collection of lines indicates about the nature of the overall discrepancy in the data or the accuracy of the underlying mapping process is unclear upon inspection.

In principle, the average separation and average squared separation vectors, \((e_x, e_y)\) and \((e_x^2, e_y^2)\), for the \(n\) error functions can be combined on the basis of each line's arc length as the weighting factor. In other words, overall measures can be developed in accordance to (84),

\[
\begin{align*}
\overline{(e_x', e_y')} &= L_1(e_x, e_y)_1 + L_2(e_x, e_y)_2 + \ldots + L_n(e_x, e_y)_n \\
\overline{(e_x^2, e_y^2)} &= L_1(e_x^2, e_y^2)_1 + L_2(e_x^2, e_y^2)_2 + \ldots + L_n(e_x^2, e_y^2)_n
\end{align*}
\]

where the \(L_j\) are the arc lengths of the true lines, \(r_j(t)\). In this manner, summary statistics for several bivariate error functions can be obtained.

At this point it will be noted that, in the case of polygonal line data, a finite set of points can provide a complete description of the bivariate error function.
Recall that in evaluating $L_2$ metric and the average squared separation vector for polygonal data, it is convenient to identify all the values of the parameter $t$ where either curve, $r_1$ or $r_2$, changes direction. The parameter values chosen \( \{t_0=0, t_1, t_2, \ldots, t_n=1\} \) correspond to the places where there is a segment endpoint on at least one of the curves. For each interval, $t_i \leq t \leq t_{i+1}$, over which neither line changes direction, the error function is simply the line joining the points $r_1(t_i) - r_2(t_i)$ and $r_1(t_{i+1}) - r_2(t_{i+1})$. This situation is depicted in figure 31. In this way, the

![Diagram](image)

Figure 31. Nature of the error function in the case of polygonal data.

error function for polygonal data can be completely described by a suitably chosen discrete sample of point discrepancies.

The fact that all the information in the bivariate error function (in the case of polygonal line data) is
contained in a finite collection of points suggests that it might be useful to consider a discrete representation of the error function in other cases. Instead of dealing with the continuous error function $e(t)$ between two lines, $r_1$ and $r_2$ as a parametrically defined curve, it would be far simpler to consider a well-chosen sample of bivariate point discrepancies.

To make this idea more precise, assume the two lines $r_1$ and $r_2$ are defined as in (41) and are in standardized arc length form. As the parameter $t$ ranges from 0 to 1, proportional distances along each line are traversed. A discrete collection of point discrepancies between the two lines can be obtained for specific values of the parameter $t$, as shown in (85).

For $t \in \{t_0, t_1, t_2, \ldots, t_n\}$, obtain

$$\{r_1(t_0)-r_2(t_0), r_1(t_1)-r_2(t_1), \ldots, r_1(t_n)-r_2(t_n)\}$$

$$= \{(x_1(t_0)-x_2(t_0), y_1(t_0)-y_2(t_0)),$$

$$\ (x_1(t_1)-x_2(t_1), y_1(t_1)-y_2(t_1)),$$

$$\ \cdots, (x_1(t_n)-x_2(t_n), y_1(t_n)-y_2(t_n))\}$$

(85)

There is no ambiguity associated with choosing the comparable points on the two lines. Comparable points are determined on the basis of the image of each specific value of the standardized parameter $t$.

In implementing this discrete linear sample scheme, an appropriate set $\{t_0, t_1, t_2, \ldots, t_n\}$ of parameter values from $t \in [0,1]$ is required. Certainly it would be reasonable
to include the endpoints so that \( t_0 = 0 \) and \( t_n = 1 \) would be included. It would also be reasonable to link the sampling to the actual (rather than the standardized) arc length. For example, values of \( t \) for a given pair of lines would be chosen so that a point discrepancy is obtained at 20 meter intervals along each feature. Comparisons of long lines would thus tend to generate proportionately more sample points than comparisons of short lines. However obtained, the collection of bivariate point discrepancies for several lines in a cartographic dataset can be readily combined for analysis. In general, the point discrepancies can be subjected to the various statistical and summary measures (considered earlier) for evaluating punctual cartographic accuracy.

The chief advantage of this discrete linear sampling scheme is that it reduces information contained in a parametric error function to a collection of bivariate point discrepancies. Such discrete bivariate point errors will, in general, be far easier to handle than the continuous error functions \( e(t) \). To the extent that error functions are sampled in a uniform manner in accordance with arc length, the average separation and average squared separation vectors can be computed on the basis of simple point sums versus the integrations in (70) and (77). Aggregating statistics from comparisons of individual lines can be performed on a pointwise basis without taking into
account arc length weighting factors as in (84). Analyzing the bivariate behavior (correlations, for example) is computationally and conceptually simpler with a set of point discrepancies than it is with continuous error functions.

Another reason sampling the bivariate error function might be favored is the practical fact that complete knowledge of such a function may be quite hard to obtain. Use of any of the other methods presented in this chapter requires a fairly good approximation of the parametric function for the true alignment of the feature being evaluated. And, such information is indeed becoming increasingly available through new positioning techniques collectively known as mobile mapping. But, in the absence of such information, conventional surveying methods (say traverse along the centerline of a road) could be used to obtain a sample of positions along a linear feature. If the proportionate distances are known along the length of the aligned feature where positions are sampled, then these independently measured point locations can be related (by means of the arc length parametrization) to the feature’s representation on the map or in the data. Position samples from a conventional survey may not provide enough information about the line to evaluate the maximum separation metric or perform the integrations required to compute the $L_2$ or vector components measures.
Discrete sampling of the bivariate error function would seem to provide many advantages over most of the other measures considered in this chapter. In particular, it should detect systematic effects due to correlation between the x and y dimensions of the error. These relationships are not reflected in any of the single-valued measures (such as the maximum separation metric or L_2 metric) considered earlier. And, summarizing and analyzing individual point discrepancies is easier than manipulating a collection of continuous functions.

It should be noted at this point too that the linear sampling approach can be used in the practical implementation of the L_2 metric and the vector components measures presented earlier. The arc length sampling can also be used in the numerical evaluation of the various integrals associated with the L_2 metric (62) and the vector component measures, (70) and (77). For these reasons, the linear sample approach would seem to be the best one for evaluating the geometric accuracy of cartographic line objects.

**Conclusion**

This chapter has considered several potential measures for characterizing the difference between the line as it appears on the map or in the data and the true line. Ideally, truth means the line as determined by an independent source of higher accuracy and precision than the
original mapping process. In practice, it can mean any other definition of the line that is accepted as true.

With most of the measures of linear geometric quality considered in this chapter, the fundamental difficulty of determining which points on the two lines (true and map) to compare is usually present. Indeed, the comparable points issue may be the heart of problem in developing linear geometric measures. The fact that map accuracy standards generally restrict attention to well-defined points suggests that establishing correspondence between ground truth and the map—that is, the basis for comparison—is nontrivial. In this research, the comparable points problem has been solved by means of a standardized parametrization based on the arc length of the line. This approach would seem to be a reasonable one to adopt as the basis for evaluating linear geometric quality.

Of all the measures considered in this chapter, the linear sample approach is recommended here. Other measures reduce the difference between lines to a single number or result in unwieldy functions that are at best difficult to work with and at worst obscure the nature of the difference between two lines. A suitably chosen sample of points from the bivariate error function can, in principle, be obtained when complete knowledge of the true line is not available. By generating a set of individual point discrepancies, data acquired from the linear sample approach can be analyzed in
accordance with the stochastic model used to summarize the accuracy of point objects. If the sampling of the bivariate linear error function is appropriate, a useful portrait of the stability and accuracy of the underlying mapping process can be obtained.
CHAPTER VII
CONTROLLING QUALITY IN MAPPING

The principal goal of this research has been to set forth a conceptual framework for controlling quality in the production of maps and spatial information. Since measurement is the basis of quality control, defining cartographic data quality and specifying how it can be measured were the primary issues considered in this research. Specifying the things to measure (control subjects) and how to measure them sets the stage for the application of the methods of industrial quality control, particularly statistical quality control (SQC), in the production of cartographic information. The various chapters in this document have explored the components of the framework for controlling the quality of a mapping process. In the present chapter, these components will be related to the operational context of a mapping project.

Applying the Conceptual Framework to the Control of Quality in Mapping

In any production system, the basic operational steps are to design the product, figure out how it can be
"manufactured," and then actually generate the product. Figure 32 illustrates the sequence of these stages. The

![Diagram of the sequence of stages in product creation: Planning and Product Design lead to Process Design, which leads to Production and Product Generation. Quality of Design and Quality of Conformance are depicted as feedback loops between stages.]

Figure 32. Operational steps in the creation of a product. The first two steps occur before formal production and are part of the planning phase. The final step in Figure 32, the one in which the product is actually generated, corresponds to the production phase.

The principal components of product quality, that is quality of design and quality of conformance, can be related to the operational sequence set forth in figure 32. Quality of design is obviously set during the planning phase of the product life cycle. Quality of conformance, how well the product meets design specifications, is determined during production (rightmost box in figure 32). Design and conformance issues overlap when it comes to fashioning the process by which the product will be generated (middle box in figure 32). Design must take into account the process by
which the product can be efficiently fabricated. Similarly, whether a production system is capable of consistently meeting design specifications is essentially a question of conformance.

A mapping project can also be described in terms of the three stages depicted in figure 32. The geographic information product must be designed, methods for acquiring and transforming the spatial data established, and then the data have to be obtained and processed into final product. Deciding what the features of the cartographic product will be (product design) and determining how it can be produced (process design) are addressed in the planning and development phase of the project. The primary issue during the production phase is if the mapping process is generating information in conformance to specifications.

The conceptual framework developed in this research for controlling quality provides a coherent basis for tackling each of the three steps of a mapping project. Its primary components are summarized in figure 33. Spatial objects are
<table>
<thead>
<tr>
<th>Framework Elements</th>
<th>Quality Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Geometry</strong></td>
<td><strong>Attribute</strong></td>
</tr>
<tr>
<td>.</td>
<td>Point</td>
</tr>
<tr>
<td>✎</td>
<td>Line</td>
</tr>
<tr>
<td>✎</td>
<td>Area Point</td>
</tr>
</tbody>
</table>

Figure 33. Primary components of the framework for controlling quality in mapping.

developed from combinations of the fundamental units of cartographic information (point, line, and area point). Geographic meaning is imparted to the objects by attributes associated with the primitives or combinations of primitives. Quality in cartographic information can be reduced to the accuracy (geometric and attribute), logical consistency (a priori constraints), and completeness (exhaustiveness) of the point, line, and area point objects that comprise it.

The following discussion will consider how this framework is applied in each of the three stages of a mapping process. While the focus in this research has been on the conformance dimension of quality during production, what occurs during the planning phase is, nevertheless, important in the control of quality. A great deal of the work associated with controlling quality in mapping is
accomplished before the first bit of spatial data is created.

In particular, the principal elements of the control feedback mechanism presented in Chapter I need to be in place before production begins. At the culmination of the planning phase, precise operational specifications for the various boxes in the control feedback loop (originally presented in figure 2 and reproduced here as figure 34) must

![Control Feedback Loop Diagram](Figure 34. Control Feedback Loop (Juran and Gryna, 1993, page 99).

be established. Of course, product specifications (goals box in figure 34) are determined in conjunction with product design. And, the mapping system (process box) that will generate the final cartographic product is set up when the process is designed. Mechanisms by which the process and its output will be measured (sensor box) and evaluated (comparison box) with respect to requirements for signs of
trouble need to be established as part of the planning phase of the project. Finally, methods for maintaining process stability and strategies for handling out-of-control situations (actuator box) should be considered when the production system is being designed. The control of quality (ensuring conformance to requirements) is thus a major consideration during the planning phase of a mapping project.

As indicated above in figure 32, the planning phase encompasses the design of the product and the process for its generation. While the scope of quality of design in mapping was considered in Chapter III, it has not been dealt with extensively in the preceding investigation. Product design establishes the characteristics of the final cartographic product as well as the requirements by which conformance will be judged. How the cartographic data will model phenomena, choice of the raster or vector model, set of features to be included, spatial reference system, topological structuring, methods for representing feature geometry and attributes, data structure, file structure, and graphic appearance are but a few of the issues that are addressed during the course of the design. In essence, design establishes what the product will be and the requirements it is expected to meet.

In many cases, the features of the geographic information product may be determined by convention or
standard practice. Maps produced by a national mapping agency in a standard series are usually generated in accordance with well-defined guidelines for content, appearance, accuracy, and so on. The format and information content for subdivision plats in a given jurisdiction will probably be fairly standard as well. In such situations, both the customer and producer will know and agree on the particular features the cartographic information is to possess. The U.S. Geological Survey conducted a survey recently to determine user satisfaction with its base cartographic products (Snyder, 1995). Quality of design (product features) is even a concern in the context of the rather standardized set of spatial information products generated by USGS.

In other mapping situations, the product may have to be designed de novo. An example is the "automation" of mapping functions at the local government level. Although the design of municipal GIS databases in the U.S. may be a maturing field, it has not yet achieved conventional status. In such cases, an attempt may be made to analyze the users' requirements and develop an information system that satisfies their needs. Figuring out what a user wants and then devising the "ways and means" to fulfill them is a principal aim in a mapping enterprise. The customer may be able to state only the general purpose for the geographic information. Designers must spell out precisely what
spatial features will be included, how they will be modeled, and at what accuracy they will be collected.

Whether the cartographic product is of conventional or novel design, the basic problem in determining its features is translating the customers' sense of quality into a formal product specification. This process, which is diagrammed in figure 35, involves relating the users' concept of fitness for use (quality) to precise operational criteria; mapping rules, feature definitions, data formats, content specifications and so on. The geometry of the point, line, and area point primitives, together with their attributes, constitute the building blocks of cartographic information from which the product's design can be constructed in a formal manner. Of course, design of a cartographic information product will encompass issues not directly tied to the framework elements. For instance, the size of the map sheet or particular computer file structure are not
closely linked to the conceptual framework. Nevertheless, as noted in Chapter IV, the point, line, and area point elements are the basis for cartographic expression at a very primitive level. For this reason, they provide a good means for articulating the essential aspects of the design of a geographic information product.

The framework elements are also useful in the design step because they comprise the major control subjects of interest in a mapping process. In expressing user needs (fitness for use) in terms of spatial objects formed from the point, line, and area point primitives, precise requirements concerning their geometric and attribute accuracy, logical consistency, and completeness must be established as well. Operational definitions for product and process quality in the right hand side of figure 35 will thus include requirements that point, line and area point objects are to satisfy. A meaningful, unambiguous, and measurable definition of cartographic information quality is predicated on the control subjects formed from the framework elements.

Rigorous operational specifications for the cartographic control subjects must be expressed in terms of numbers. It is thus necessary to specify how the accuracy, completeness, and logical consistency of the geometric and attribute aspects spatial objects formed from point, line, and area point primitives will be defined and measured
quantitatively. In this research, an effort was made to address quantification of the geometric accuracy of point, line and area point objects. Although the notion of geometric quality (accuracy) of spatial data may be well-developed, the review of the various spatial information standards in Chapter III revealed significant limitations. In particular, a method for quantitatively specifying the geometric accuracy of linear objects does not exist.

As argued in Chapter V, geometric accuracy for cartographic points should be treated as a bivariate vector quantity. Under the bivariate approach it is possible (in principle at least) to detect relationships (correlations) between the x and y components of discrepancies between the true positions and "manufactured" locations of point features. Such relationships can be obscured when discrepancies are defined--as in the five geographic information standards reviewed in Chapter III--in terms of univariate distances or independent errors in the x and y dimensions.

A method for characterizing the geometric difference between the true and map line was proposed in Chapter VI. It is based on the notion that the difference between two lines can be reduced to the issue of determining comparable points upon which to base the evaluation of the discrepancy. The basic problem is deciding which pairs of points, one from each line, should be chosen for the sake of difference
measurements. This issue was resolved by assuming a standardized arc length parametrization as the basis for establishing comparable points. In other words, a point located 3/10 the total distance along the first line should be compared to the point on the second line that is precisely 3/10 the distance along its total length. A bivariate (vector valued) error function is obtained from the collection of point discrepancies between the lines.

Since a bivariate error function would be difficult to analyze, it was decided that a (potentially dense) discrete sample of individual point vector discrepancies from the error function be obtained instead. In this way, methods applicable to the evaluation of point discrepancies can be applied to linear features.

As far as the definition of geometric accuracy, the framework presented in this research is complete. It will enable the quantitative specification of operational criteria for the geometric component of geographic information quality. This research did not, however, address quantitative measurement of attribute accuracy and the problems of evaluating logical consistency and completeness. Although this research has not focused on the design component of geographic information quality, the conceptual framework for its control plays a central role in the product design step in a mapping project. It is the basis for expressing cartographic design at a very
fundamental level in a manner conducive to objective measurement.

Along with the design of the cartographic product, during the planning phase of the project, it is necessary to design the process for actually generating the geographic information. In mapping, the boundary between the design of the product and determining the "ways and means" for its generation may be hard to distinguish. The decision to use satellite imagery as the basis for a mapping project, for example, largely defines the kinds of features and positional accuracy that can go into the product. Similarly, the need for very high positional accuracy may dictate the use of GPS surveying techniques. In many situations, the choice of mapping technology and the nature of the end product are very closely linked. But, as noted in Chapter II, the technology for collecting and generating spatial data is changing so rapidly that an ever expanding range of production methods is becoming available. However these planning and design decisions are resolved, it is obvious that the cartographic process must be capable of delivering the desired spatial data product.

Engineering the mapping process is a matter of developing a system that will reliably produce geographic information in accordance with design requirements. Since both the cartographic information and its quality characteristics are ultimately defined in terms of point,
line and area point objects, they are instrumental in process design as well. However complex a mapping process may be, its evaluation involves the three components set forth in figure 36. Note that the measurement operations in

![Diagram](image)

Figure 36. Primary aspects of process evaluation, measurement of inputs, process and output.

this diagram represent an expanded conception of the sensor box in figure 34. Evaluation of the mapping process will in general require measurement of the inputs, the process itself, and the output.

In the case of spatial data, inputs might be geodetic control information obtained from a third party contractor, satellite imagery, or existing sources of geographic information. In so far as the quality of these source data will influence the final cartographic product, their quality should be known and controlled to the extent possible. The actual mapping process may be a complex sequence of operations (subprocesses) where data are created and
transformed into intermediate outputs. A particular production step may involve the direct manipulation of the framework elements as in the manual digitizing of a map document. In other cases, the controllable aspects of the process may not be related to the point, line, and area point objects. For instance, the temperature and humidity of the room in which hardcopy maps are generated on a plotter are not directly related to the framework elements. However, these process parameters may influence the dimensional accuracy of the final document and thus affect the geometric quality of the point, line, and area point objects represented on the map.

The major issue in designing the mapping process (and figuring out ways it should be evaluated during production) is determining how the characteristics of the input information and various process parameters affect the quality of the final cartographic information product. In ideal circumstances, the influence of these factors on the quality requirements for the output cartographic information (expressed in terms of the framework elements) would be well understood. This knowledge may be difficult to acquire. However, it is essential when trying to fix a broken process or secure an improvement in its capability. While the present research did not specifically consider the problem of process engineering, the conceptual framework provides a set of measurable control subjects to use in studying the
mapping system. By analyzing the geometric and attribute quality of the point, line, and area point objects, an understanding of how process factors affect the quality of the end geographic information product can be obtained.

Upon completion of the design phase of the mapping project, the features of the geographic information product have been determined, precise operational design specifications set and the production system developed. At this point, the primary concern in the sequence in figure 32 is the conformance of the production process to the design specifications. Formal production of the cartographic information commences and the control feedback loop depicted in figure 34 is activated. As discussed in Chapter I, measurements of the process and its output are compared to the specifications, and corrective action taken if requirements are not satisfied. The primary tool for judging (comparison box in figure 34) whether the mapping process is broken is statistical quality (or process) control (SQC or SPC). To the extent the framework developed in this research has set forth precise, measurable units of geographic information, implementation of formal SPC techniques for their control will be more or less straightforward. This subject is addressed briefly in the next section.

Before considering how SPC might be applied in mapping, it will be useful to consider the generic control feedback
loop in figure 34 from the perspective of the conceptual framework developed in this research. To the extent it is based on the framework elements, the control of quality in mapping can be diagrammed as in figure 37. Measurement of geographic information generated by the mapping process, as well as the specifications, will be in terms of the point, line, and area point objects. For geometric accuracy, the quantitative methods suggested Chapters V and VI provide a complete basis for specifying and evaluating quality. As long as the process exhibits statistical control, no action is taken and the process is left alone. If the system shows lack of stability, then it becomes necessary to search for special causes and take corrective action.

**Statistical Control in Mapping**

The actual steps involved in the control of quality in a mapping project were set forth in the preceding section.
At this point, it will be useful to consider how the concepts of SPC might be applied in practice. To that end, this section will consider the implementation of statistical control with respect to geometric quality. In essence, the question in figure 37: "Is the process in statistical control?" will be addressed. The discussion is not intended to be conclusive; rather it will suggest how the fundamental units of spatial information might be controlled.

Recall from Chapter I that the aim of statistical control is ensuring stability of the mapping process. The presence of special (or assignable) causes of variation must be discovered and eliminated so the process may be controlled within limits, that is, reduced to a system in which only common causes are present. Such a system is said to be in a state of statistical control; it is a random process (Deming, 1986, page 321). Determining if a mapping process is in statistical control is essentially a matter of answering the question: "Does it represent a random process?"

The principal tool for determining whether measurements of control subjects show evidence of nonrandom variability is the control chart. Control charts were developed by Walter Shewhart (Shewhart, 1931) and provide an operational means for deciding whether data constitute a random sample from some statistical distribution or universe (Grant and Leavenworth, 1988, page 74 and Eisenhart, 1969, page 27).
The key point in establishing statistical control is not the particular form or shape of a probability distribution associated with a quality characteristic, but rather that there is a universe at all (Eisenhart, 1969, pages 27-28 and Shewhart, 1931, page 301).

To decide if a set of n data values come from a constant system of chance causes (that is, constitute a random sample from a statistical universe), Shewhart proposed Criterion I, consisting of the following steps (Shewhart, 1931, page 304). First divide the data into m rational subgroups with \( n_1, n_2, \ldots, n_i, \ldots, n_m \) values. Then choose and compute statistics (typically estimates of the mean and dispersion of the measured phenomenon). For each statistic (\( \Theta \)) computed, estimates for its mean and standard deviation are obtained. Shewhart uses two very important criteria to guide the determination of these estimates (Shewhart, 1931, pages 301-302). The first is that estimates of the means and standard deviations of the statistics should converge in probability to the means and standard deviations of those statistics for the population (if indeed a population exists). In other words, for a given statistic \( \Theta \), estimates are chosen that fulfill (86).

\[
P(|\bar{\mu}_\Theta(n) - \mu_\Theta| > \epsilon) \rightarrow 0, \text{ and}
\]
\[
P(|\bar{\sigma}_\Theta(n) - \sigma_\Theta| > \epsilon) \rightarrow 0, \text{ as } n \rightarrow \infty 
\]
Secondly, the estimates for the means and standard deviations of the chosen statistics should be those most likely to indicate lack of stability in the cause system. Fulfilling these two conditions can at best be approximate because the nature of the underlying distribution, if it exists at all, can not be known. An important consequence of the second requirement is that estimates for the standard deviation will be based on an average (or possibly median) of the m subgroup dispersion statistics, rather than a single estimate derived from the whole sample (Wheeler and Chambers, 1992, pages 56-60).

The final step is to construct a control chart for each statistic θ whose limits are given by (87).

\[ \hat{\theta} \pm 3\hat{\theta} \]  

(87)

Lack of control is indicated if an observed value of the statistic for a subgroup falls outside the limits determined in (87). The two values in (87) constitute the upper and lower control limits, often referred to as 3-sigma limits (Grant and Leavenworth, 1988, page 77). This method is known as the Shewhart control chart. Statistics θ for which control charts are commonly used are the mean and variability (dispersion) of the process. Process variability is often characterized by the range rather than a standard deviation. When these characteristics are monitored with control charts, the aim is to detect the lack
of a single statistical universe due to changes in either the process average or its dispersion, or both average and dispersion.

At this point it should be noted that, while motivated by statistical theory, the control chart approach based on the three-sigma limits is robust with respect to the actual nature of the underlying distribution (Shewhart, 1931, pages 315-318). Some writers relate Shewhart's technique sketched above and the resulting control limits to actual probability values (Montgomery, 1992, pages 104-105). In particular, they suggest the control chart is essentially a sequential testing of the hypothesis that the process is in a state of statistical control. As an example,

For successive random samples of size n, this control chart [after Shewhart] can be viewed as repeated tests of significance of the form $H_0: \mu=\mu_0$ vs. $H_1: \mu\neq\mu_0$. The regions above the UCL and below the LCL thus correspond to the likelihood ratio test rejection region.

(Alt, 1985, page 111)

Others suggest that the Shewhart control chart can not be interpreted in terms of formal statistical theory.

According to Deming, couching statistical control in such a formal manner is theoretically incorrect and can be misleading.

The calculations that show where to place the control limits on a chart have their basis in the theory of probability. It would nevertheless be wrong to attach any particular figure to the probability that a statistical signal for detection of a special cause could be wrong, or that the chart could fail to send a signal when a special cause exists. The reason is that no
process, except in artificial demonstrations by use of random numbers, is steady, unwavering.

It is true that some books on the statistical control of quality and many training manuals for teaching control charts show a graph of the normal curve and proportions or area thereunder. Such tables and charts are misleading and derail effective study and use of control charts.

Rules for detection of special causes and for action on them are not tests of a hypothesis that the system is in a stable state.  
(Deming, 1986, pages 334-335)

One possible trap stemming from reliance on a formal statistical interpretation is the conversion of a process capability number (derived from control chart analysis) into a fraction of nonconforming product. Such results, "... come in two flavors: Ordinary fantasies and outright hallucinations" (Wheeler and Chambers, 1992, page 129). The problem arises because a probability distribution (usually normal) must be assumed in order to relate conformance specifications to process dispersion limits derived from the control chart. Wheeler and Chambers note that three-sigma limits on the Shewhart control chart have been proven in over 60 years of practice; they do not rely on any of the restrictive assumptions of formal deductive statistical models (Wheeler and Chambers, 1992, pages 56-84).

Controlling the geometric quality of a mapping process by means of the control chart approach would seem to be a fairly straightforward application of the approach outlined above. The process would be rather simple if a univariate measure (such as the distance metric used in the NMAS 1947,
for example) were the control subject. Distances between the target (provisionally true) locations of points and their respective locations as generated in the mapping process would be observed for appropriate subgroupings of the data. Estimates for the mean distance error and its variability (based on subgroup ranges or sample standard deviations) would be used to develop three-sigma control limits for average distance discrepancies and distance discrepancy variation (either range or standard deviation). In the case of linear discrepancies, a univariate measure, such as the \( L_2 \) metric or maximum separation metric, could serve as the control subject.

The approach would work equally well if discrepancies between the true and the process geometry were regarded as distinct quantities in the \( x \) and \( y \) dimensions. For instance, the average discrepancy and variability for the \( x \) coordinates of points would be controlled separately from \( y \) coordinates. In the linear case, the vector components measures could be similarly employed as control subjects.

However, the strategy outlined in the preceding paragraphs is probably inadequate for at least two reasons. The first is due to arguments in Chapters V and VI that geometric quality ought to be regarded as bivariate quantity. Univariate measures may obscure relationships in the nature of the discrepancies (preferred orientations or correlations between their \( x \) and \( y \) components) that may be
significant. Therefore, it is reasonable to require that discrepancies in \( x \) and \( y \) be controlled jointly. This first objection is not too significant because control charts for two or more related variables are available. It should be noted that the explicit theory behind these multivariate control charts, in contrast to Shewhart's Criterion I, seems to rest heavily on the assumption that the process control subjects are normally distributed (Tracy, Young and Mason, 1992, page 89; Montgomery, 1991, pages 322-333 and Alt, 1985, pages 111).

A second shortcoming of the method, even assuming a multivariate control chart is employed, arises from the fact that bivariate cartographic errors are themselves a spatially distributed phenomenon. Control charts depend on the rational subgrouping of the data as the basis of detecting nonrandomness. The idea is to choose subgroups so that the chance of variability within any one group is small while the chance of variability from group to group is high (Grant and Leavenworth, 1988, page 118). Detecting lack of stability (nonrandomness) in the process (shifts in the mean or variability of the statistical universe) by the control chart is linked to the particular choice of rational subgrouping. Often rational subgroups are based on the order of production or time. In mapping, for example, if the rational subgrouping for the control chart is based on
time (order of production), shifts in the process average, say, from day to day can be detected.

However, in the case of spatial data, it is equally important to detect nonrandomness in the process that can arise on the basis of spatial ordering. The need to reflect spatial as well as temporal ordering of the discrepancies for the sake of determining statistical control is illustrated in figure 38. Here the discrepancies, though

![Diagram of map region with error concentrated and negligible areas]

Figure 38. Example of a nonrandom distribution of error with respect to space.

randomly distributed (with respect to magnitude and orientation) and well behaved, are all concentrated in the northwest corner of the region. Clearly, a mapping process with this kind of discrepancy distribution over space would give rise to suspicions of nonrandom variation and special causes. In other words, what went wrong in the northwest
corner? However, such variation would generally not be detectable unless the subgrouping reflected spatial ordering of the error in some way. Statistical control in a mapping process is perhaps more complicated than industrial applications because the process must be random with respect to time and space.

In order to determine statistical control when working with spatial data, some means of detecting nonrandomness with respect to spatial order is clearly necessary. One approach may be to consider statistical control from the perspective of multiple rational subgroupings of the data. Of course, the temporal order of production would be a fairly obvious basis for choosing rational subgroups for the purpose of developing the control chart. And, indeed, there may be a clear connection between the temporal order in which the data were acquired and its spatial distribution. But, to be more deliberate in the detection of nonrandomness with respect to space, it may be useful to consider rational subgrouping on a geographic basis.

A few suggestions for rational subgrouping in the mapping context are provided in figure 39. For instance,
grouping by intervals along the x axis (or along other alignments) might uncover lack of stability in the process with respect to information gathered along that axis. In attempting to analyze stability with respect to space (or any other significant parameter), it is necessary to take into account the nature of the particular mapping process. The various spatial orderings suggested in figure 39 may make sense only in certain situations. For instance, the circular intervals radiating from the center of the region might be useful in controlling radially symmetric lens distortion in a metric camera. In the case of data obtained by means of conventional ground survey or GPS there may be little reason to suspect lack of stability would behave in a radially symmetric fashion; some other spatial hypothesis is
required. A key determinant of variation in geometric accuracy may be the nature of the terrain itself. For example, even though the process for its acquisition may be quite stable, the accuracy of elevation information may vary significantly in accordance to the roughness of the ground. It is conceivable then that influences of the mapped phenomenon on quality may be difficult to distinguish from problems with the underlying data collection system.

While multiple subgroupings may not be an elegant solution to statistical control in spatial processes, the technique would at least provide some basis for asserting that the mapping process is random with respect to space. The issue of the spatial stability of the mapping process is a complex problem that will require further research.
CHAPTER VIII

FUTURE DIRECTIONS

The purpose of this chapter is to identify some of the issues and problems that remain to be solved in the application of quality control to mapping. It is hoped this research has, at the very least, securely established the relevance of the quality movement (especially the ideas of its major exponents, Deming, Juran, and Shewhart) to the mapping domain. Methods developed to enhance the quality of manufactured goods can and should be applied to the generation of cartographic information products.

Before considering the areas that require further investigation, it is worth noting what has been accomplished in this research. At the conceptual level, this effort identified the point, line, and area point primitives as the fundamental building blocks of two-dimensional analog, vector and raster cartographic representations. These atomic units of geographic information are the basis for objective definition and measurement of quality in mapping. The three framework elements, together with their attribute (descriptive) information, are the essential characteristics of a cartographic product that should be regulated in the
quality control process. In this sense, the point, line, and area point objects comprise a general purpose set of control subjects for mapping processes. Very broadly, cartographic information quality can be defined in terms of the accuracy, logical consistency, and completeness of the attribute and geometric aspects of the point, line, and area point objects.

In addition to developing a general definition of spatial data quality, this research explored the more specific problem of characterizing geometric accuracy. Methods for the quantitative evaluation of the geometric accuracy of the point, line, and area point primitives were suggested. In Chapter VI, a method for measuring the geometric accuracy of linear features was developed. The proposed measure for line objects attempts to remedy the fact that the spatial data standards reviewed in Chapter III are limited to positional accuracy for well-defined points.

By identifying the control subjects to evaluate, as well as the methods for their quantitative evaluation, the groundwork for controlling quality mapping is established. Therefore, as far as the control of geometric accuracy in mapping, the framework developed in this research is complete. However, many aspects of the practical measurement and evaluation of cartographic quality primitives were not addressed.
While formal methods for characterizing the geometric accuracy of the quality primitives were considered in Chapters V and VI, quantitative methods for evaluating the accuracy of attribute information, completeness of the representation, and logical consistency were not developed. These issues were acknowledged in Chapter V and each one could be the focus of an extensive investigation in its own right. Logical consistency in geographic information is one area where a great deal of work may yet be required. Rigorous methods for characterizing the logical consistency, completeness and attribute accuracy of a geographic information product need to be developed in order to fill in the conceptual gaps in the framework that were not resolved in the present effort.

Other issues were deliberately avoided altogether in this research. Though it was suggested in Chapter IV that the point, line and area point elements may be sufficient for handling three-dimensional cartographic information, this fact remains to be demonstrated. It is possible that genuine three-dimensional GIS capability will be developed, requiring an expanded set of primitive objects. In either case, the results presented above are applicable to planar cartographic information. Time is another complicating factor that was not considered. New technologies for collecting and processing spatial data may destroy the notion of a static spatial database. Cartographic
information may come to be seen as a dynamic resource whose quality needs to be maintained over time. In such a situation, temporal accuracy of the data becomes an important quality characteristic to be continuously evaluated and controlled.

The problem of design is another major aspect of spatial data quality that was largely avoided in the present research. It was suggested in Chapter VII that the framework elements will be useful both in developing the features of the geographic information product and in figuring out how it can be efficiently produced. But, in some situations it may be difficult to translate a user's vague requirement into precise operational specifications for spatial objects formed from the point, line, and area point primitives. For instance, in designing a mapping system for interstate highways, what is the geometric accuracy required for emergency call box (point feature) locations? Is it 2, 10, or 50 meters? Why? Establishing a rational justification for a decision in this simple case may not be as trivial as first appears. And, rational justification may be impossible in the case of spatial data that undergo analysis in complex GIS models. In general, the problem of propagating "quality" from the cartographic information to its ultimate effect on users' decisions will need to be solved before quality of design can be treated in an objective manner.
A problem closely related to design is engineering the processes for actually producing the spatial information. While the framework elements are the principal control subjects of interest, unrelated parameters of the mapping process may have a profound impact on quality. Levels of training, equipment used and its calibration, ergonomics, quality of vendor-supplied inputs, and management competence are a few examples of these system variables. Establishing the relationship between such process parameters and the quality of the point, line and area point objects may be difficult. Although each mapping process is unique, applied research in the area of process engineering could shed light on the relative significance of certain general classes of parameters. This kind of information would be of great value in troubleshooting a mapping process that has fallen out of control.

A final technical question concerns the application of statistical process control to mapping. It was noted in Chapter VII that the system generating the geographic data should behave as a random process. So, in addition to being random with respect to the (time) order of production, a mapping process should also exhibit randomness with respect to space. That is, special causes should not enter the system due to shifts in location. This situation would seem to be unique to spatial data and may require statistical
quality control methods that are uncommon in the manufacture of physical commodities.

The outstanding problems discussed so far all relate to the technical aspects of controlling quality. However, the creation of geographic information takes place within a wider institutional context. Various economic, social, and political factors will determine the kind of maps and spatial information that get created, by whom they are generated, and how they are produced. As noted in Chapter II, changes in the United States associated with the NSDI initiative may jumble responsibilities for maintaining and ensuring the quality of the data that are produced. No quality control program in mapping can be considered optimal unless it factors in the bottom line social and economic costs and benefits associated with the geographic information that gets produced and used. By the same token, the full benefits to be obtained from quality improvement will become apparent only when these institutional issues are considered.

There are, of course, many institutional matters to be studied, ranging from the management concerns of a private mapping company to the geographic information requirements of society as a whole. One important research need is to determine the economic value of geographic information, and the extent to which varying levels of quality will influence its value. This information is necessary if rational
choices are to be made concerning the level of cartographic information quality that should be achieved.

The last and perhaps most important work that is required will be to actually test the ideas set forth in this research in practice. With respect to the reason for choosing three sigma as the basis for setting control limits, Shewhart noted,

The fact that the criterion which we happen to use has a fine ancestry of highbrow statistical theorems does not justify its use. Such justification must come from empirical evidence that it works.

(Shewhart, 1931, pages 17-18)

The same comment could apply to the conceptual framework for controlling quality in mapping. Quality control programs predicated on the objective evaluation of the point, line, and area point objects will ultimately decide the validity of the theory. In particular, the approach set forth in Chapter VI for characterizing the geometric accuracy of linear features needs to be applied in practical mapping contexts to determine if it is a valid quality measure.
LIST OF REFERENCES


Maffini, Giulio, Michael Arno and Wolfgang Bitterlich. 1939. "Observations and Comments on the Generation and


