**Title**
Unit 099 - Detecting and Evaluating Errors by Graphical Methods



**Authors**
099, CC in GIScience
Beard, Kate
Hunter, Gary

**Publication Date**
2000

Peer reviewed

# Unit 099 - Detecting and Evaluating Errors by Graphical Methods

by Kate Beard, Associate Professor, Department of Spatial Information Science and Engineering and NCGIA,
University of Maine, Orono, ME 04469

This section was edited by Gary Hunter, Department of Geomatics, University of Melbourne, Australia.



---

## Advanced Organizer

### Full Table of Contents

### Metadata and Revision History

---

# Detecting and Evaluating Errors by Graphical Methods

## 1. Rationale for Graphical Detection and Evaluation

- Graphical methods for error detection and evaluation are motivated by physiological, technical, and institutional factors.
- Physiological
  - human information processing system has strong acuity for visualization and

ability to recognize structure and relationships
  - spatial structure is more easily expressed and grasped through graphic or cartographic representation
  - graphical methods are a fast communication channel

- Technical
  - new initiatives e.g. digital libraries, National Spatial Data Infrastructure (NSDI) expand need to document spatial information reliability
  - more spatial data and geographic information processing resources are becoming accessible over the Internet and we need quick methods to process this larger volume

- Institutional
  - national and international standards efforts -(SDTS, 1992) Metadata Content Standard (FGDC, 1995) MEGRIN standards (Salge´ et al 1992) are requiring data quality assessment

### 1.1 Limitations of graphical methods

- Graphical methods are not always an effective solution nor a substitute for conventional numerical analytical tools.

- Graphical methods are open to misinterpretation.(Robinson et al 1985, Monmonier 1991)MacEachren (1994) suggests,

- data exploration tools allow identification of patterns we might otherwise miss, but do not guarantee that the pattern we see is real

---

## 2. Examples Of Graphical Methods

- Several disciplines have contributed including cartography, spatial statistics, statistical graphics, scientific visualization and spatial error modeling.

### 2.1 Graphical Methods in Statistics

- exploratory data analysis (EDA)) introduced graphical methods for exploring data. (Tukey 1977, Chambers et al 1983, Becker et al 1987, Cleveland 1993).

- highlight unusual values which may be errors

- a spatial methods do not consider spatial dependencies and do not detect values which may be unusual in a spatial context

- Cressie (1991) identifies EDA methods which overcome this limitation..

### 2.2 Graphical Methods in Cartography

- reliability diagrams were an early attempt to display variation in source documents used to compile maps (Wright 1942).
- theoretical treatments of projection distortion (Tissot 1881, Imhof 1964, Maling 1973).
- Bertin's (1983) graphical framework (visual variables).

### 2.3 Graphical Methods Related to GIS

- new visual variables including defocusing of features (MacEachren 1994, McGranaghan 1993) multivariate symbols (Hancock 1993).
- new visualization technologies (voxel-based 'true' 3-D displays, animation, hypermedia).

  **Specific examples:**

  - MacEachren et al (1993) developed a reliability visualization tool (RVIS) which supports several options for viewing data and metadata (reliability). Display options include side by side, overlay and merged displays.
  - Fisher developed error animation to view the reliability of classified imagery (1994a) and soil maps (1994b).
  - Goodchild et al (1994) use a fuzzy classifier to create multinomial probability fields. Display of realizations of the error model can inform users of the potential variation.
  - Paradis and Beard (1994) developed data quality filter that allows users to specify a data quality parameter (e.g. positional accuracy), a quality measure (e.g. RMSE) and a threshold value. The filter displays only data meeting thresholds.
  - Hunter and Goodchild (1995) describe a probability mapping approach for representing the uncertainty of the horizontal position of a nominated terrain elevation value.
  - Mitasova et al (1995) developed visualization tools for multidimensional interpolation and its accuracy based on cross validation.

---

## 3. Challenges in Graphic Error Detection and Evaluation

- Challenges include 1) graphic design issues, 2) metadata issues, 3) error analysis issues, and 4) user satisfaction issues. A well-known case described by Blakemore (1985) provides a good example of the lack of understanding of geographic data accuracy requirements

### 3.1 Graphic Design Issues

- requires a representation of space or linkage of a spatial displays to a spatial representation (Monmonier 1989)

- Spatial displays provide users with information on whether errors are regular, random, or clustered in space.
- Two dimensional displays restrict views of full three dimensional space
- 3 dimensional displays add substantial cognitive and computational costs.
- need for both implicit and explicit displays of uncertainty.
- uncertainty conveyed implicitly with visual variables which suggest uncertainty (e.g. fog, unfocused displays, unsaturated colors) (McGranaghan 1993).
- explicit display requires quantification of the uncertainty arrived at through error analysis.
- graphic display should allow a data distribution and its reliability to be displayed independently or jointly
- three possibilities for joint display of data and reliability:
    - 1) side by side images,
    - 2) composite images, and
    - 3) sequenced images MacEachren (1994)
- side by side displays
    - viewer must interpret two images simultaneously.

      images should be comparable - same size, same coordinate scales, should be linked.
- composite images
    - requires overlay of contrasting visual variables, bivariate, or multivariate mapping. Bertin (1983) proposes different data variables with symbols of different dimensions (point, line, area). Examples Mitasova et al (1995) and MacEachren et al (1993) Brewer (1994) bivariate maps .
- images in sequence
    - need to interval of time between images, visual frame of reference must be constant between images
- linked displays and multiple version displays.
    - must be common visual cues for the same variable in different contexts- images (Monmonier 1989)
    - in multiple version displays need to display multiple realizations which by their differences indicate a range of uncertainty in the data.
    - these can be displayed as small multiples Tufte (1983), or sequenced using animation (Dibiase et al 1992). Uncertainty in this case is expressed implicitly.
- multiple views

- several iterations of a display can help to convey the uncertainty due to map design decisions MacEachren (1994)

### 3.2 Metadata Issues

- spatial data are frequently poorly documented.
- without information on data collection, sampling design, compilation or processing steps there is little basis on which to proceed.
- need to update metadata as data are updated

### 3.3 Error Analysis Issues

- Errors are often not detected simply by displaying the raw data (although examples of this are possible). Graphics gather their power from content and interpretation beyond the immediate display of numbers Tufte (1983.
- Good graphic design and by association effective detection and evaluation are highly dependent on effective error analysis.

### 3.4 Detection - determine presence of error

- All error detection requires some model or reference framework, either implicit or explicit, from which departures can be determined.
- These may include
    - a known or postulated distribution for a set of observations;
    - a hypothesized or assumed relationship;
    - an expected set or range of values, or
    - an independent (and more accurate) set of observations.
- These models and frameworks range from simple and inexpensive to complex and expensive.
- Plotting data works as a error detection device because we often have some expectation about the pattern we will see. Deviations from this pattern suggest errors.
- Statistics provide framework for detection by establishing an expected distribution for values.
- For spatial data we add departures from assumed stationary of mean or stationary of dependence as the basis for detection of possible errors. For example, we would be suspicious of observations when they are unusual with respect to their neighbors.

---

# 4. Techniques for raw data

- Exploratory techniques
  - identify outliers, detect blunders, and perform preliminary identification of data structure and statistical properties.
  - are most appropriate where observational data are not obtained by formal means, are not very accurate or at a high level of measurement or where real repetition is not feasible

    Cressie (1991) outlines some exploratory techniques for spatial data

- Many exploratory techniques require some processing and often "soft" models to generate interesting information for graphic display.

- Consistency rules
  - indicate ranges of expected values or expected relationships between values
  - For example topological rules such as the requirement that all chains begin and end with a node, or that all polygons must close are applied against the data and geometric configurations which deviate from these rules are flagged.
  - GIS editing packages support graphic highlighting of these inconsistencies to support easy visual detection as well as display of their spatial distribution..

- Use of ground truth data or other sources of higher accuracy
  - example - root mean square error measures the error between a mapped point and a measured ground position. image classification uses ground interpretation
  - Comprehensive ground checks are expensive.

- Detection is an ongoing process
  - Error and uncertainty in spatial data are not static. New error and uncertainty occur as data are processed.

- Knowledge of lineage
  - Processes applied to the data should be known to utilize a specific graphic technique. For example to use Tissot's indicatrix to evaluate projection distortion we must first know what projection was applied.

- Where processes are unknown, simulations can be applied to generate information for graphic display.
  - realizations generated by simulation provide a distribution from which we can compute a variance and confidence limits.
  - simulations are computationally demanding.

---

# 5. Evaluation - determining magnitude/ significance of errors

- requirements for evaluation
  - the context of information use,

  - a model and
  - a hypothesis to determine significance.

- Evaluation techniques:
  - Cross validation
    - a common method used to assess statistical prediction
    - Observations are iteratively deleted and the remaining data are used to predict deleted observations.
    - Repeating this over many deleted subsets allows an assessment of the variability of prediction error.
  - Fuzzy classifiers
    - provide a means of describing uncertainty by associating pixels with a vector of class memberships (Goodchild, Sun and Yang (1992)
    - can create quite large processing or large storage overheads.

- Substantial costs and processing can be required to generate information for graphic display.The form and content of graphic displays is highly dependent on effectiveness of the error analysis. Implications are that GIS or other visualization software packages must either include error analysis tools or data producers must perform these analyses and store the results with their data.

### 5.1 User satisfaction issues

- User satisfaction issues relate to the packaging around the graphic and error analysis tools.
  - interface to these tools should be intuitive and easy to use.
  - ideal graphic displays are those which are simple, relevant, and unambiguous
  - users should be able to get the error information without losing sight of their original application goals.
  - uncertainty in the data should not be mapped to an uncertainty in the graphics such that a user has to search hard or spend a long time interpreting the results.
  - for most users the evaluation of uncertainty and error is a step on the path to some further goal rather than an end in itself..

---

# 6. Framework For Graphical Methods

- Framework as a two phase mapping
  - first between data, an application context, and a suite of appropriate error analysis methods.
  - second between the outcome of the error analysis and graphic display methods.

- The framework organizes information around three basic components: 1) the data, 2) the context of the analysis, and 3) error analysis/graphic methods..

- Data Characteristics
  - 1) status; whether the data are raw or processed and if processed what processes

    and parameters were applied,
  - 2) observed dimensions of the data: spatial, thematic, or temporal.

- range of possible dimensions includes the three spatial dimension X, Y and Z, several attribute dimensions A1... An and time, T. An observation could be a 2 or 3 dimensional spatial observation in which only geometry was observed (a survey measurement), a single or multivalued spatial observation or estimate in which geometry and attributes were observed or estimated (e.g. soil color and texture at location P), or a single or multivalued space time observation (e.g. observations on surface temperature and precipitation at the same station at the same time intervals)

---

# 7. Context description

- Indicates the environment in which the error analysis might be carried out.three components
  - 1) the task: error detection or evaluation,
  - 2) the desired dimensions of the error analysis: spatial, thematic, temporal, or combination, and
  - 3) the user types.

## 7.1 Detection is simplest

- Detection is simplest - may be accomplished by plotting the data and relying on the human eye to do the detection.

- Evaluation is more complex - methods may be exploratory or confirmatory and include tests for the significance of the errors.

## 7.2 Desired dimensions of analysis

- desired dimensions for error analysis can include spatial, thematic, temporal, or combined.
  - for example the only information of interest to a user may be the error or uncertainty in the location of an observation.
  - observed dimensions might restrict desired dimension e.g positional error analysis is limited if only two dimensions were observed rather than three.

## 7.3 User Types

- The user type influences the selection of error analysis and graphic methods--example users: data producer/distributor and the data browser in a digital library.

- Data producers
  - need robust error detection and correction tools that can operate quickly and effectively on large volumes of data.

deal primarily with raw data and objective is blunder detection and correction.
  - analysis applies to all dimensions of the data (space, theme, time)need review.
  - a goal could be to save the results of the error analysis and graphic displays as metadata for transfer with the data to end users.

- Digital library users
  - involved in searching for and evaluating data
  - both error detection and evaluation tasks apply
  - error analysis and graphic methods must be fast since users may be paying for connection time
  - error analysis and graphics will need to be simple and efficient to work over a range of client configurations

Table 1. shows mapping between data characteristics, context and error analysis methods..

# Table 1

Curly brackets under applicable dimensions indicate that the analysis method applies to combined dimensions rather than to dimensions individually. The underline indicates the dimension of primary interest. Computational complexity is by rank. As an example plotting is an error analysis technique that applies to raw data, can be applied to the analysis of all dimensions, serves the detection task and has low computational complexity

Each error analysis method produces an output which can be characterized. according to

  - the level of measurement of the result,
  - the spatial representation of the result (point, line, pixel, surface, etc.).

The graphic problem is one of representing k variables in an n dimensional field using a fixed set of spatial object representations (points, lines, pixels, surfaces). The range of possible variables which need to be displayed either separately or jointly includes:

  - 1) the observed data values;
  - 2) the errors in or reliability of the observed values;
  - 3) estimated data values; and
  - 4) the reliability of estimated values.

Any one of the four may be displayed independently or in some combination.

To combine displays of data and reliability we need to know the characteristics of both.

Table 2 links characteristics of the error analysis results and graphic display options. It

- identifies the level of measurement of the output and the spatial object representation to which the output may attach.

- can guide the choice of graphic display mode if the data and their reliability are to be displayed together

- graphic modes in the table refer specifically to the graphic techniques for combining data and reliability representations
  - side by side,
  - composite
  - sequenced images
  - small multiples (Tufte 1983).

**Table 2**

A composite map is the first choice since it is visually most efficient The efficiency of the composite image breaks down as the number of variables increases or the complexity of the spatial representation increases.
When this occurs two simple side by side images are preferable.

---

# 8. Future Research In Graphical Methods

- need enhancement and develop error models for spatial data, the development of error propagation techniques and enforcement or encouragement of better documentation of data sets.
- need evaluation of feature-oriented approaches to data quality representation,.
- need evaluation of how errors accrue differentially with specific GIS operations (buffering, overlay, coordinate conversion, etc.)
- need reduction in computational complexity of error detection and evaluation
- evaluation of error analysis on the fly versus storage of error analysis results
- improvements in data documentation - collection of metadata prior to data collection and parallel with data updates (Beard 1996).
- quality assessment of spatial data independent of GIS. .
- development of modular interoperable components which could be easily recombined.

---

# 9. References


- Anselin L 1997 (this volume)
- Beard M K 1996 A Structure for Organizing Metadata Collection Proceedings 3rd International Conference on GIS and Modeling Sante Fe, NM
- Becker R A, Cleveland W S and A Wilkes (1987) Dynamic Graphics for Data Analysis

Statistical Science 355-395

- Bertin J 1983 Semiology of Graphics: Diagrams, Networks, Maps Madison, Wisconsin, University of Wisconsin Press
- Bicking B and M K Beard 1995 Toward Implementing a Formal Approach to Automate Thematic Accuracy Checking for Digital Cartographic Datasets Proceedings Auto Carto 12 355-362
- Brewer C A 1994 Color Use Guidelines for mapping and visualization. In MacEachren A and D R F Taylor (eds) 1994 Visualization in Modern Cartography Oxford, Elsevier 123-148
- Burrough P 1989 Fuzzy mathematical methods for soil survey and land evaluation Journal of Soil Science 40: 477-492
- Chambers J M, Cleveland W S, Kleiner B, and Tukey P 1983 Graphical Methods for Data Analysis Boston, Duxbury Press
- Cleveland W S 1993 Visualizing Data Murray Hill NJ, AT&T Bell Laboratories
- Cleveland W and McGill R 1984 Graphical Perception: Theory, Experimentation and Application to the Development of Graphical Methods
- Journal of the American Statistical Association 79(387): 531-553
- Cox D R 1978 Some Remarks on the Role in Statistics of Graphical Methods Applied Statistics 27 9
- Cressie N 1991 Statistics for Spatial Data New York, John Wiley & Sons
- Dibiase D, MacEachren A M, Krygier J and Reeves C 1992 Animation and the role of map design in scientific visualization Cartography and GIS 19(4): 201-214 265-266
- Englund E 1993 Spatial Simulation: Environmental Applications In: Goodchild M F, Parks B O, Steyart L T (eds) Environmental Monitoring with GIS New York, Oxford University Press 432-437
- Federal Geographic Data Committee (FGDC) 1995 Content Standards for Geospatial Data
- Fisher P 1994a Visualization of the reliability in classified remotely sensed images Photogrammetric Engineering and Remote Sensing 60(7): 905-910
- Fisher P 1994b Visualizing the uncertainty of soil maps by animation Cartographica 30(2) 20 -27
- Fisher P 1997 (this volume)

Gershon N and Brown J R 1996 The Role of Computer Graphics and Visualization in the GII IEEE Computer Graphics and Applications. 61-62

- Goodchild M, Buttenfield B and Wood J 1994 Introduction to Visualizing Data Validity in Hearnshaw H and Unwin, D. (eds) Visualization in Geographic InformationSystems. Chichester, John Wiley & Sons 141-149
- Haining R 1990 Spatial Data Analysis in the Social and Environmental Sciences Cambridge, Cambridge University Press
- Hancock J R 1993 Multivariate regionalization: An approach using interactive statistical visualization Proceedings Auto Carto 11 Minneapolis MN 218-227
- Heuvelink G 1997 (this volume)
- Hunter G J and Goodchild M F 1995 Dealing with error in spatial databases: a simple case study Photogrammetric Engineering and Remote Sensing 61(5): 529-537
- Imhof E 1964 Beitrage zur Geshicte de topographischen Kartographie International Year Book of Cartography 4: 129-154
- Leung Y, Goodchild M F, Chih Chang L 1992 Visualization of fuzzy scenes and probability fields Proceedings 5th International Symposium on Spatial Data Handling Charleston, SC 480-490
- MacEachren A M 1994 Some Truth with Maps: A Primer on Symbolization and Design Washington, DC, American Association of Geographers
- MacEachren A M, Howard D. von Wyss M, Askov D and Taormino T 1993 Visualizing the health of Chesapeake Bay: An uncertain endeavor Proceedings GIS/LIS '93 Minneapolis, MN 449-45
- Mackinlay J 1986 Automating the design of graphical presentations of relational information. ACM Transactions on Graphics 5.(2): 110-141
- Maling D H 1973 Coordinate Systems and Map Projections London, George Philip and Son Limited
- McGranaghan M 1993 A cartographic view of data quality Cartographica 30 (2):8 19
- Mitasova H, Mitas L, Brown W, Gerdes D P, Kosinovsky I and Baker T 1995 Modeling spatially and temporally distributed phenomena: New methods and tools for GRASS GIS International Journal of GIS 9(4): 433-446
- Monmonier M 1991 How to Lie with Maps Chicago, IL, University of Chicago Press
- Monmonier M 1989 Geographic Brushing: Enhancing exploratory analysis of the scatterplot matrix Geographical Analysis 21(1): 81-84

Openshaw S, Charleston M and Carver S 1991 Error propagation: a Monte Carlo simulation. In Handling Geographical Information. Masser I and Blakemore M (eds) New York, John Wiley & Sons 78-101

- Paradis J and Beard M K 1994 Visualization of Data Quality for the Decision Maker: A Data Quality Filter Journal of the Urban and Regional Information Systems Association 6(2): 25-34
- Robertson P K 1991 A Methodology for Choosing Data Representations IEEE Computer Graphics and Applications May 1991 56-67
- Robinson A H, Sale R D, Morrison J and Muehrcke P 1985 Elements of Cartography 5th ed. New York, John Wiley & Sons
- Salge´ F, Smith N and P. Ahonen 1992 Towards harmonized geographical data for Europe: MEGRIN and the needs for research. Proceedings 5th International Symposium on Spatial Data Handling Charleston, SC 294-302.
- Tissot A 1881 Memoire sur la representaton des surfaces et les projections des cartes geographiques Paris, Gauthier Villars.
- Tufte E R 1983 The Visual Display of Quantitative Information Cheshire CT, Graphics Press
- Tukey J W !977 Exploratory Data Analysis Reading Ma, Addison-Wesley.
- Wood, J. 1994. Visualizing Contour Interpolation Accuracy in Digital Elevation Models. In Hearnshaw, H. and Unwin, D. (eds) Visualization in Geographic Information Systems. Chichester: John Wiley& Sons. 168-180.
- Wright, J. K. 1942. Map makers are Human: Comments on the Subjective in Maps. The Geographical Review. 32(4): 527-544.

---

# Citation

To reference this material use the appropriate variation of the following format:

> Beard, Kate,(1998) Detecting and Evaluating Errors by Graphical Methods, *NCGIA Core Curriculum in GIScience*, http://www.ncgia.ucsb.edu/giscc/units/u099/u099, posted June 23, 1998.

---

Created: June 23, 1998
Last revised:June 23, 1998.

---

# Unit 099 - Detecting and Evaluating Errors by Graphical Methods

## Table of Contents



---

Back to the Unit

# Unit 099 - Detecting and Evaluating Errors by Graphical Methods

## Metadata and Revision History

### 1. About the main contributors


- Kate Beard, Associate Professor, Department of Spatial Information Science and Engineering and NCGIA,University of Maine, Orono, ME 04469


### 2. Details about the file

- unit title
  - Detecting and Evaluating Errors by Graphical Methods
- unit key number
  - 099

### 3. Key words

### 4. Index words

### 5. Prerequisite units

### 6. Subsequent units

### 7. Other contributors to this unit

### 8. Revision history

- Created: June 23, 1998

---

Back to the Unit.

| **Error analysis method** | **Data status** | **Applicable dimensions** | **Tasks** | **Computational complexity** |
|---|---|---|---|---|
| | | | | |
| Plots | raw | x,y,z,a,t | detection | low |
| Consistency checks | raw | x,y,z,a,t | detection | low |
| Ground truth checks | processed | x,y,z,a | detection, evaluation | low |
| Adjustment computation | raw | x,y,z | detection evaluation | low-moderate |
| Cross validation | processed | {x,y, z,a} | evaluation | moderate |
| Fuzzy classification | processed | {x,y,z,a} | evaluation | moderate |
| Simulation | processed | x,y,z,a,t | detection evaluation | high |

**Table 1. Provides a basis for associating error analysis methods with particular data set and context characteristics.**

| **Error analysis method** | **Level of measurement** | **Applicable spatial object** | **Spatial object evaluated** | **Graphic mode** |
|---|---|---|---|---|
| | | | | |
| Consistency check | nominal | point, line, area | point, line,<br>area | composite |
| Ground control check | real | point, pixel,<br>set of pixels | point, pixel,<br>set of pixels | composite |
| Adjustment computation | real | point | point | composite |
| Cross validation | real | point<br>surface | point<br>surface | composite<br>side by side |
| Fuzzy classification | real | pixel | pixel | small multiple<br>animation |
| Simulation | nominal-<br>real | surface | surface | small multiple, animation |

**Table 2 Basis for associating error analysis output with graphic display modes.**