Geo-spatial Data Quality and Errors
(Uncertainty in GIS)

Dr. Harish Karnatak
Scientist and Head, GIT&DL Department
Indian Institute of Remote Sensing, ISRO Dehradun, India
harish@iirs.gov.in

Why to study errors?

- GIS database model is an approximation of real world;
- Geographical data collected, entered and processed are not sufficiently reliable and error free;
- The spatial database has problems because the real world phenomena are not always discrete but also fuzzy in nature.

Causes of errors in spatial data

- Measurement errors: accuracy (ex. altitude measurement or soil samples, usually related to instruments);
- Computational errors: precision (ex. to what decimal point the data is represented?);
- Human error: error in using instruments, selecting scale, location of samples;
- Data model representation errors;
- Errors in derived data.

Basic concepts

- Error
  - It encompasses both the imprecision of data and its inaccuracies.
- Accuracy
  - It is the degree to which information on a map or in a digital database matches true or accepted values.
- Precision
  - refers to the level of measurement and exactness of description in a GIS database.

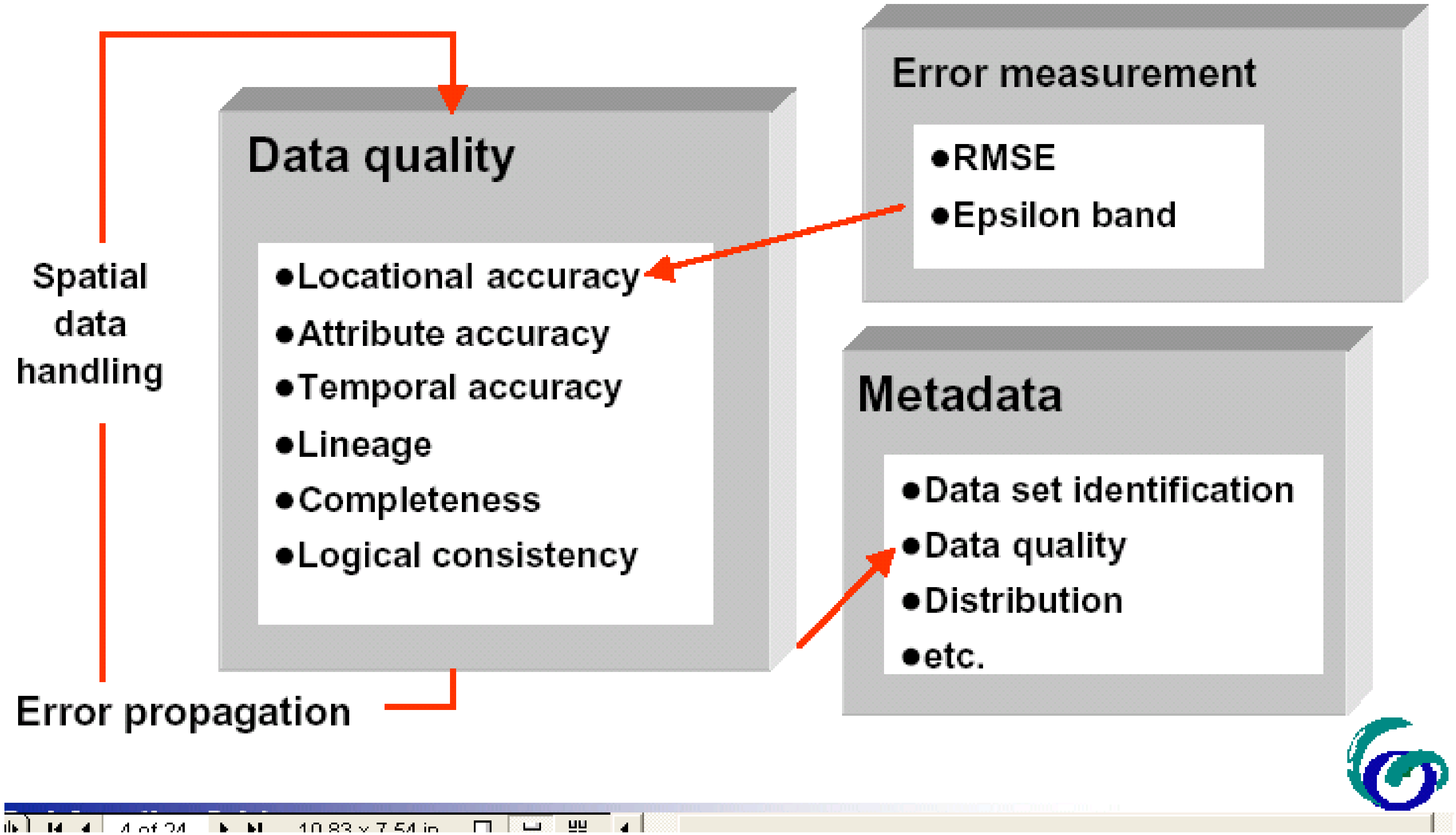
Basic concepts

- Data
  - A collection of facts from which conclusions may be drawn
- Quality
  - Data quality refers to the state of qualitative or quantitative pieces of information.
  - data is generally considered as high quality if it is "fit for its intended uses in operations, decision making and planning‟.

Data Quality

- Quality is a function of intangible properties such as completeness and consistency;
- Data are the result of a production process, and the manner in which this process is performed clearly affects data quality.

Data quality components in GIS



Factors affecting reliability of GIS data

- Age of data – collected at different times
- Areal coverage
- Map scale and resolution
- Density of observations
- Data formats and exchanges in formats
- Accessibility

Precision and Accuracy

- Precision: How exactly a location is specified (the exactness of the method used).
- Accuracy: How close it is to the true value (the exactness of the result)
- GIS data are capable of more precision 'double-precision'
- -> 6 decimal places 1234567.123456 (meters or square meters)
- But this is not meaningful without highly accurate data

Types and sources of errors in GIS

1. Types of error: spatial or attributes
2. Sources of error:
    1. instruments,
    2. human,
    3. change
3. The 'errors' that can occur during the four components of GIS:
    1. Input
    2. Database management
    3. Data Analysis
    4. Output

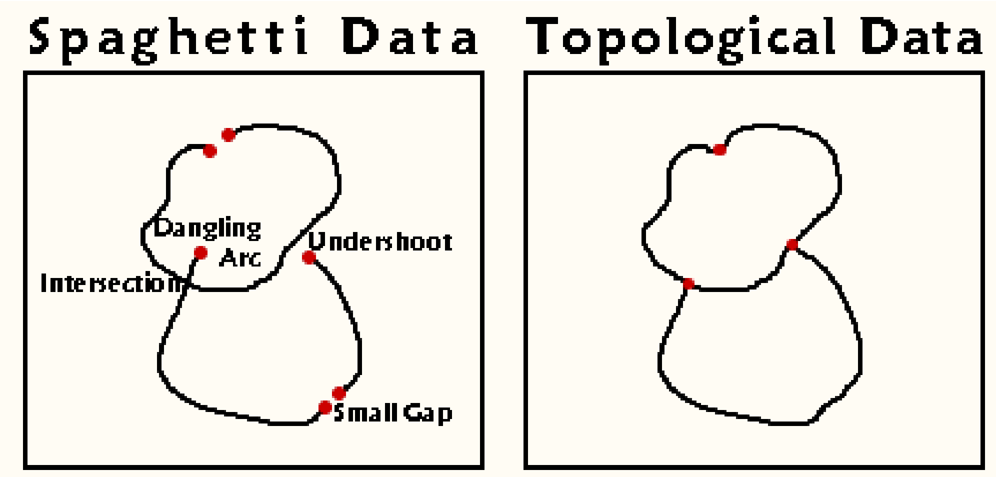
GIS processing errors

Input:

- Digitizing: human error and the width of a line
- Dangling nodes (connected to only one arc): permissible in arc themes

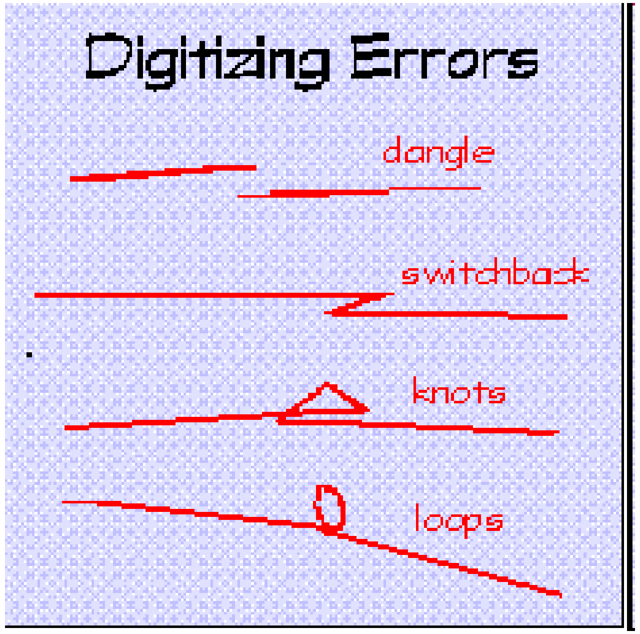
Topology is needed for GIS analysis

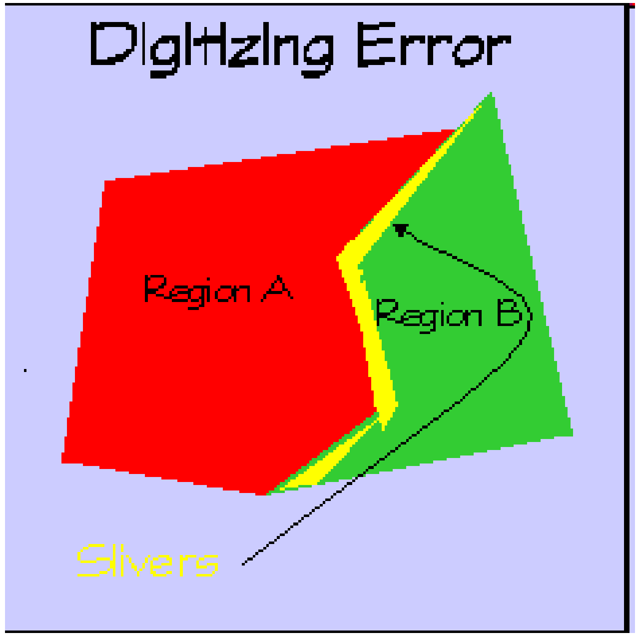
- [Topology = the spatial relationships between geographic features]



GIS processing errors

In-out :Digitizing errors

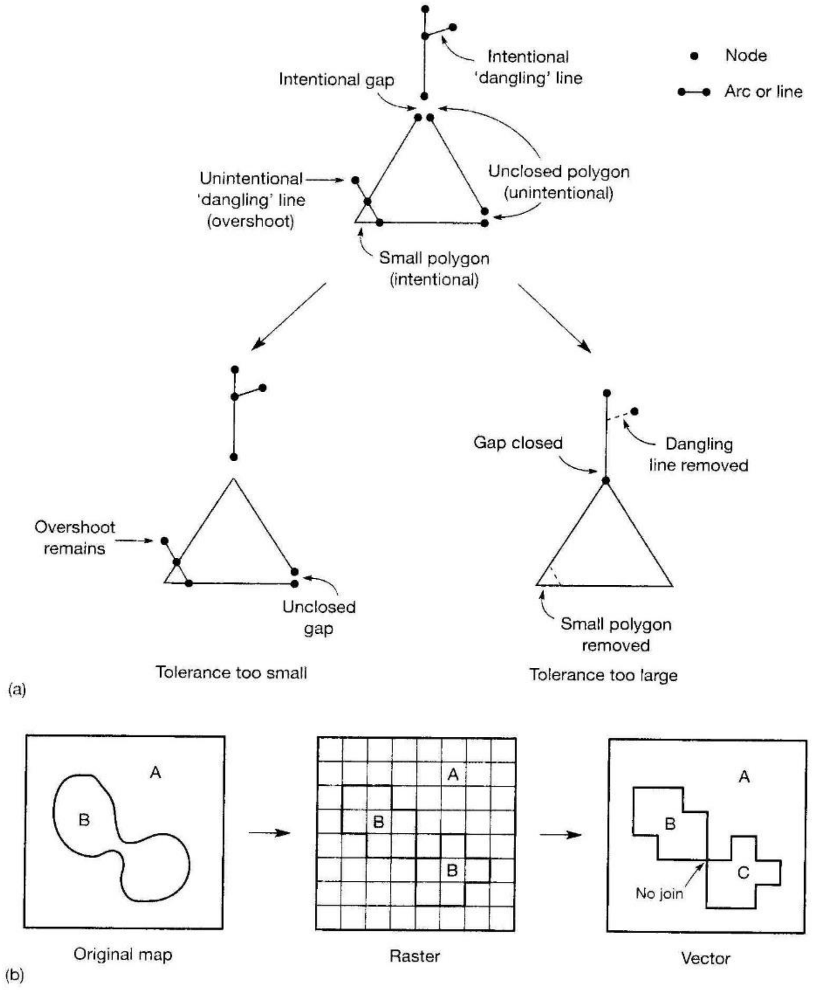




Topological errors in vector GIS

(a) Effects of tolerance on topological cleaning

(b) Topological ambiguities in raster to vector conversion



GIS processing errors

Database management

- Data precision: too little, too much
- Missing (null) values
- Metadata – how / when were data gathered etc..
- Units of measurement e.g. feet versus meters

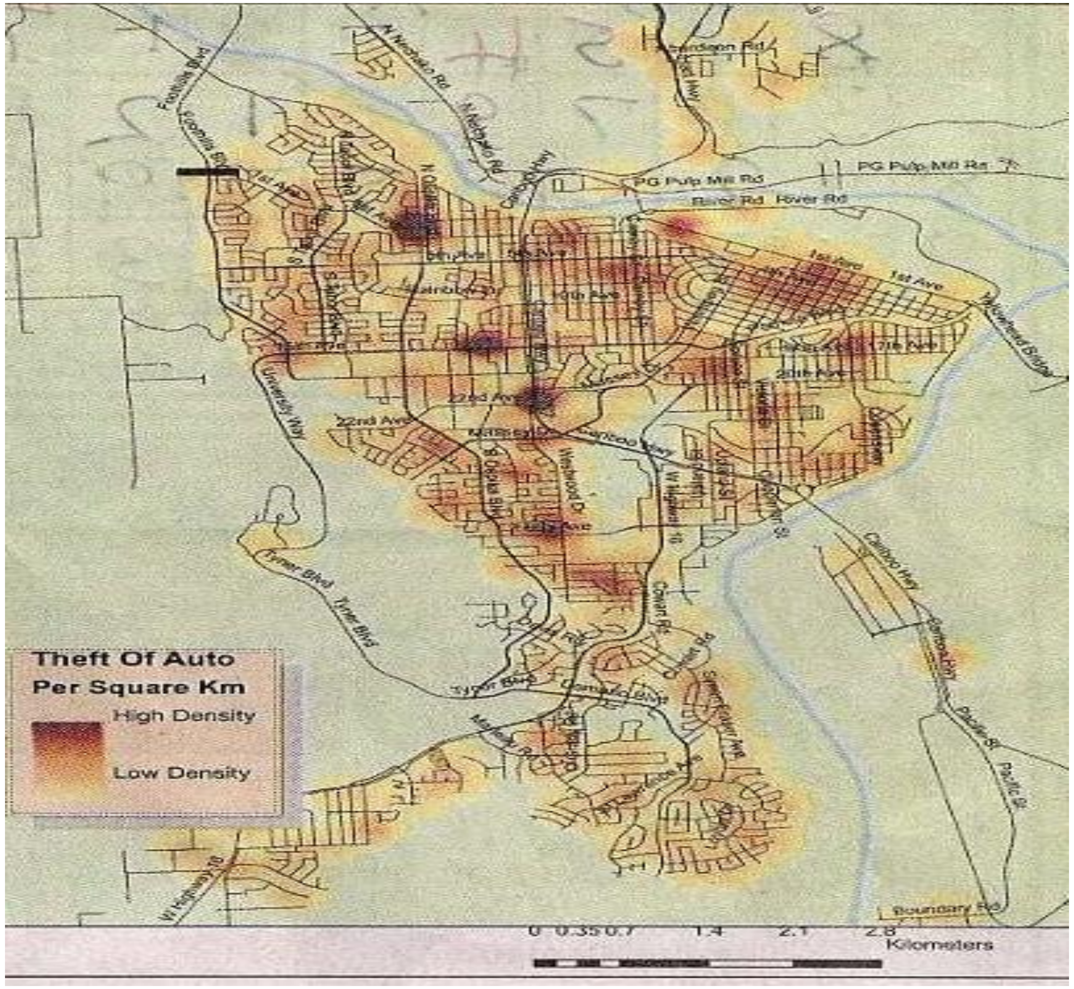
POLYGON	ESA_1	SPC1	PCT1	SPC2	PCT2	AGE_CL	HT_CL_IN	SITE_IDX	CRNCL_CL	SitePrep	Dist	YearDist	Regen	STTEND
67		HW	40	S	40	2	1	16.6	8	B	R	1985	1999	F
133			0		0	0	0	0	0			0	0	
199		HM	40	HW	30	9	3	7.2	5		L	1980	0	
353		HW	90	BA	10	9	4	11.6	1	B	L	1980	1999	F
229		HW	70	HM	20	9	3	9.5	5	B	L	1980	1999	F
264		HM	50	HW	30	9	3	7.5	5	H	L	1980	1999	F
162			0		0	0	0	0	0			0	0	
393		HW	60	HM	20	9	3	8.5	5	H	L	1980	1999	R
165		HM	80	BL	20	9	3	7	4	H	L	1980	1999	R

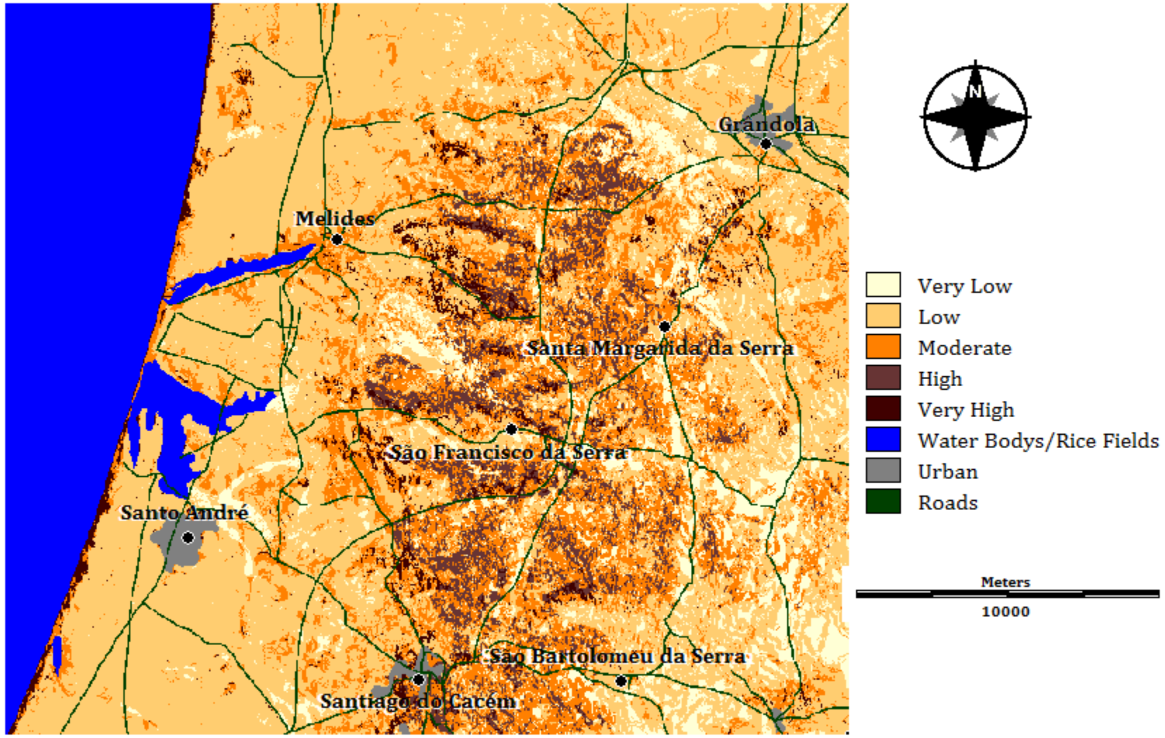
GIS processing errors

Data analysis

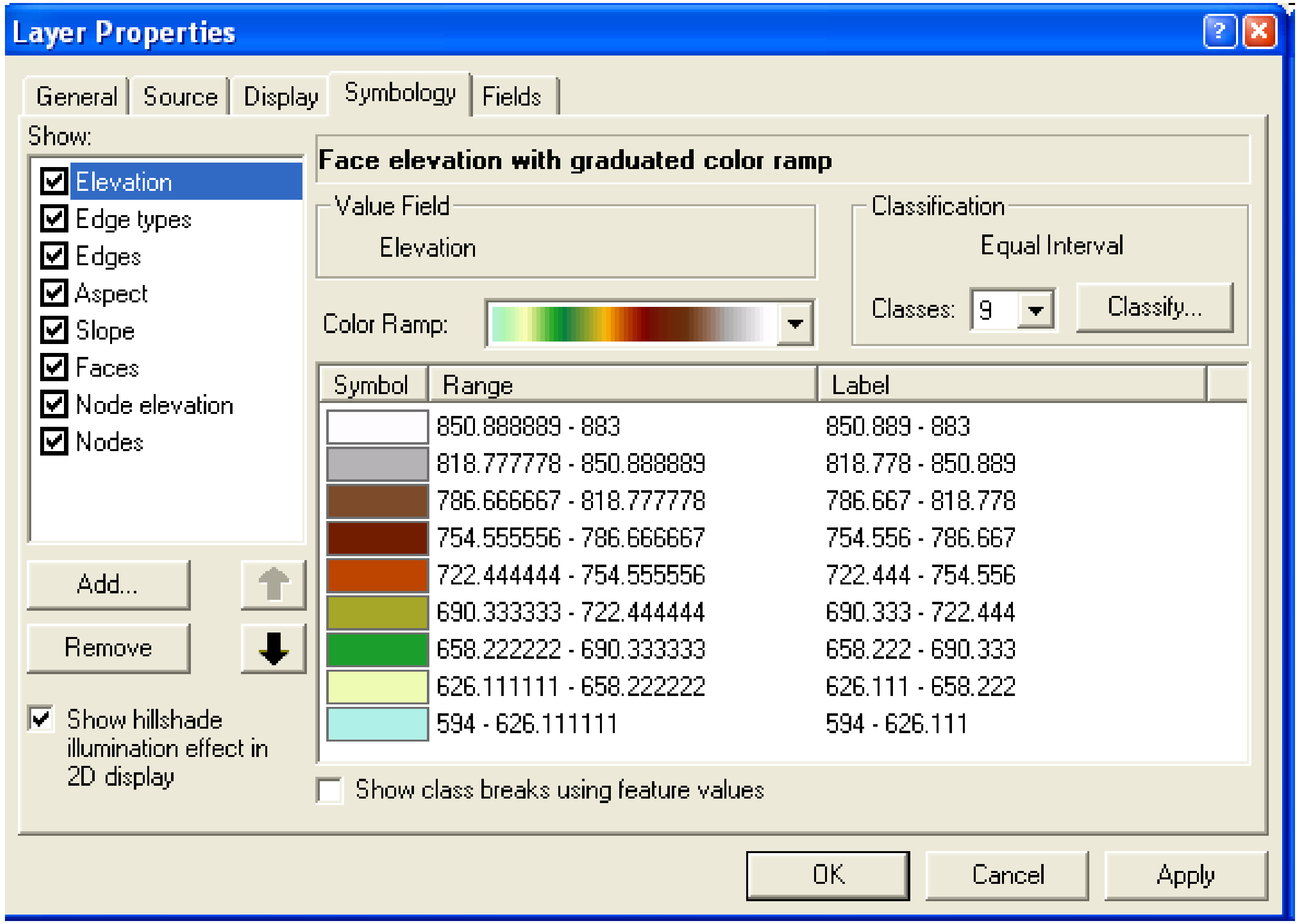
- Interpolation of point data into lines / surfaces e.g. TIN / contours.
- Overlay of layers, digitized separately from different sources or scales, e.g. soils and vegetation.
- They have common borders, but slight differences cause 'slivers'.
- The compounding effects of processing and analysis of multiple layers: for example, if two layers each have correctness of 90%, the accuracy of the resulting overlay is around 81%.
  - Inappropriate or inadequate inputs for models
  - Dubious classifications

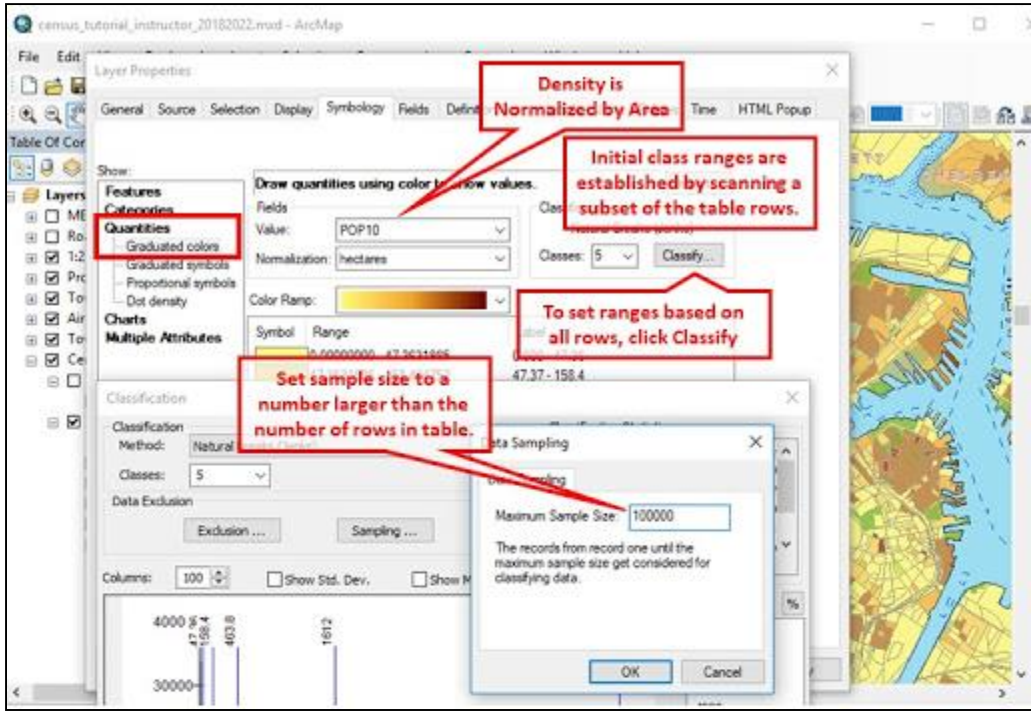
Dubious analysis



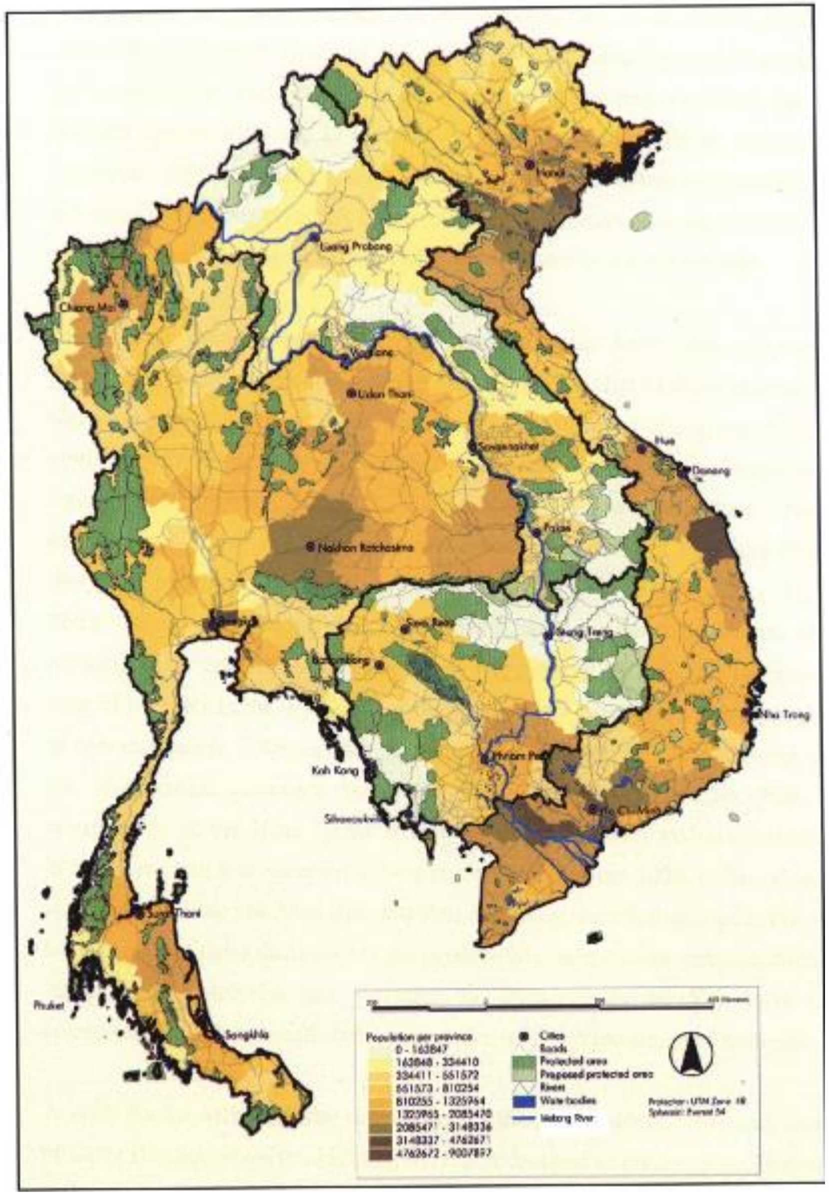


Default class boundaries

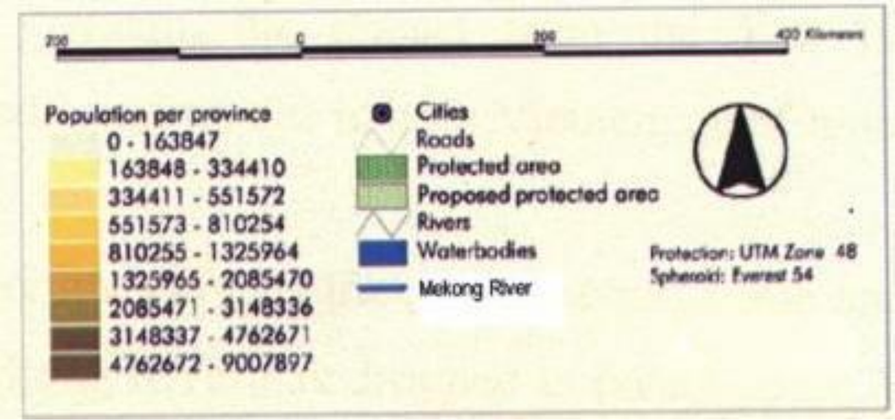


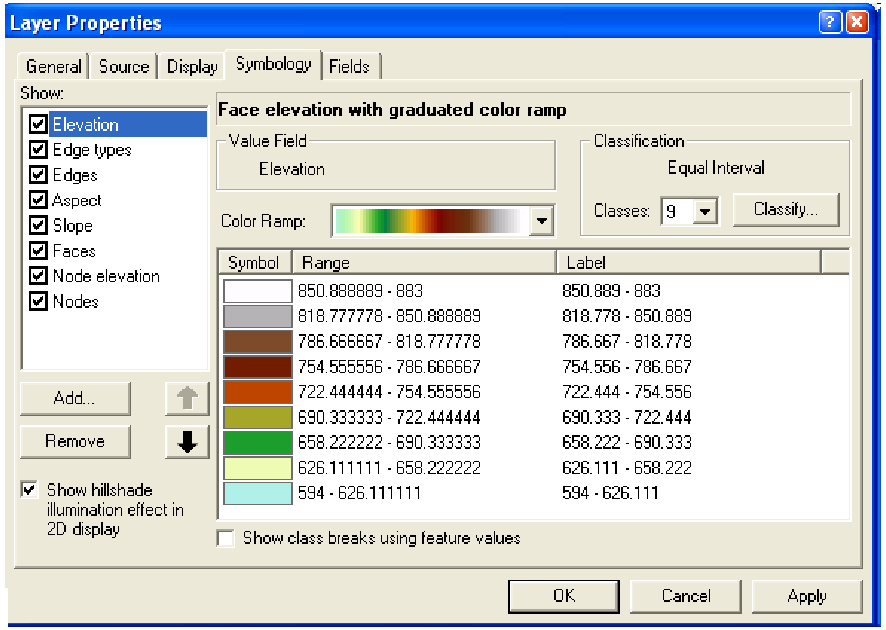


Data units and bad classes (density v numbers)



Map showing population density and protected area



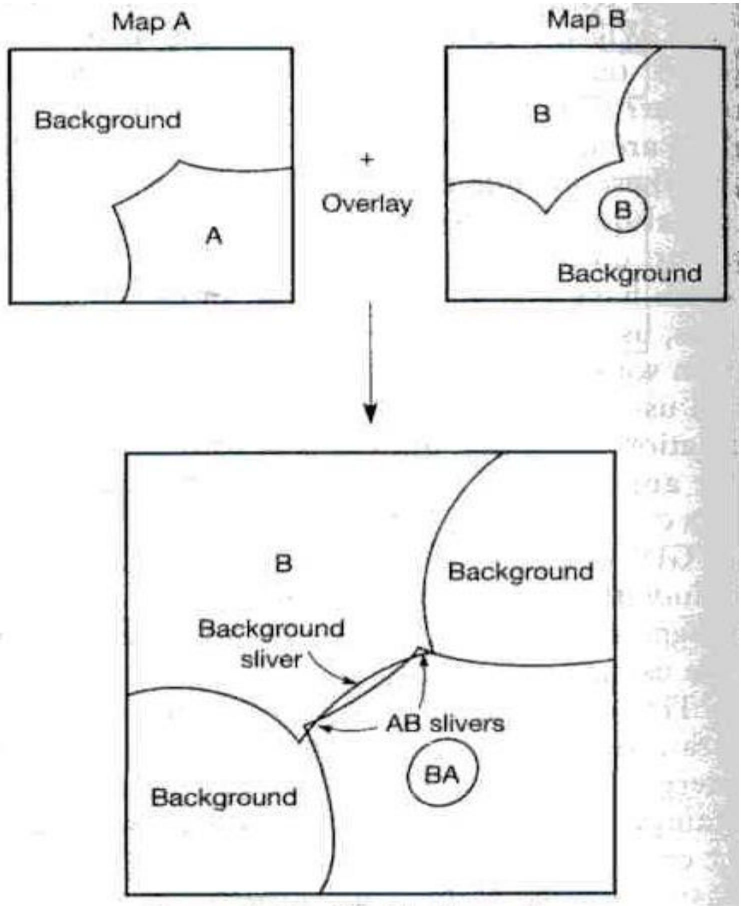


Errors in data processing and analysis

- GIS operations that can introduce errors include the classification of data, aggregation or disaggregation of area data and the integration of data using overlay technique.
- Where a certain level of spatial resolution or a certain set of polygon boundaries are required, data sets that are not mapped with these may need to be aggregated or disaggregated to the required level.

Attribute error due to processing

- Attribute error result from positional error (such as the missing 'hole' feature in map A that is present as an island in map B). If one of the two maps that overlaid contains an error, then a classification error will result in the composite map (polygon BA).



Composite (A + B) with sliver polygons

Project management and human error

- Measuring error
- Typos/drawing errors
- Incorrect implementation error
- Planning/coordination error
- Incorrect use of devices error
- Erroneous methodology error
- Other human errors

Geometry related errors

- Rounding errors
- Processing errors
- Geometric coordinate transformation
- Map scanning, geometric approximations
- Vector to fine raster errors

Controlling and Dealing with Errors

How errors can be controlled?

- Estimating degree of an error is an interesting area of GIS and computational science.

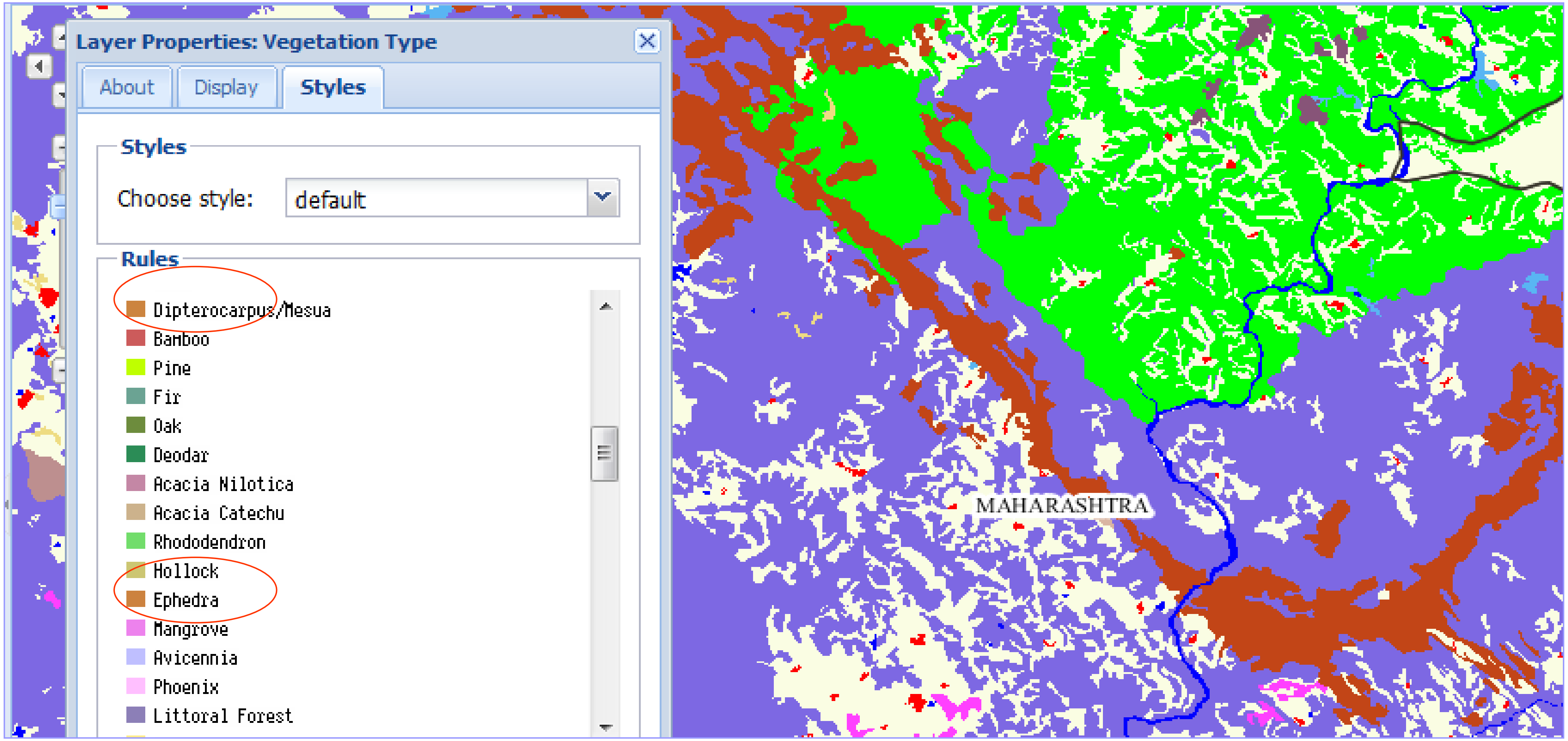
Methods to deal with errors:

- Initial data: control quality of measurement, develop standards, prevent human error.
- Data models: select correct data models based on experience or model appropriateness, reduce errors during conversion from one to another.

Finding and modeling errors in GIS

- Checking for errors
  - Probably the simplest means of checking for data errors is by visual inspection.
  - Various statistical methods can be employed to help pinpoint potential errors.
  - Estimating degree of an error helps in controlling and correcting errors

Output Error



Scale bar, north arrow, mix legend ???

Summary of GIS errors

- Computer data have as many or more errors than printed maps
- The difference between accuracy and precision
- The effects of scale and generalization
- Lack of documentation - the need for metadata
- Age and date of GIS data (relative to rate of change)
- Effect of area jurisdictions - e.g provincial differences
- The challenge of a large province and country