Geography 4203 / 5203

GIS Modeling

Class 12: Spatial Data Quality and Uncertainty

Some Updates - Progress Reports

- **Progress reports:** 11 & 14 April (instead of 14 & 16 April)
- Class on 16 April: Jeremy
- Class on 18 April: Babs comes for guest lecture
- **One-to-One meetings with project leaders:** 21 & 23 April (instead of 18 & 21 April)
- **Final presentations :** 23, 25, 28 & 30 April (as planned)

Last Lecture

- We finished with the very last part of spatial estimation
- We had a look at sampling issues and core area methods
- You understand why core area delineation is so important in different fields for extraction, modeling and evaluation
- You realize how to implement the ideas of transforming lower-order objects into higher-order objects
- You can explain how Kernel methods, convex hull and mean center approaches work

Today‘s Outline

- Coming to aspects of **spatial data quality**
- **Uncertainty** and spatial data quality - why are they **important**
- **See some examples of consequences of uncertainty in Spatial Data**
- Learning more about the **terminology** used and what these **terms mean**
- How to **describe** quality of spatial data and why **standards** have been evolved for transfer standard

Learning Objectives

- Understanding of **uncertainty** and **spatial data quality** as well as some of the **consequences**
- Terms and **terminology** such as **accuracy**, **error**, **precision**,…
- What the **SDTS** is and what stands behind the **famous five points**

Where to Start…?

- Maybe by asking: **What is quality**?

  - Reaching the **best result** possible…

  - Reaching the **requirements** defined by **standards** or **customers**…

  - Reaching the requirements for a specific **application** -- "**Fitness for use**"
- In summary this seems to be about **uncertainty**, how it **arises** and **propagates** throughout the process the spatial data are applied to

Quality <> Uncertainty???

- Uncertainty is considered to embrace the **unknown domain how good** our data are (too many types and subtypes to be listed here)
- However uncertainty **influences** our **spatial data quality** (question how this quality is specified)
- Uncertainty can be **introduced** at any stage of **GIS-based map production** and analysis (Reality observation, conceptual modeling, measurement, analysis steps, and … are we missing one here?)

Any Differences?

- Remember our readings discussion about **uncertainty and SDQ**
- Both fields are dedicated to related issues but went into **different directions** with regard to research foci…
- However we will explore the role of uncertainty and how it influences SDQ during this class set

Data Quality	Lineage	
	Accuracy	Positional
		Attribute
	Completeness	
	Logical Consistency	
	Semantic Accuracy	
	Currency	

Uncertainty	Error
	Vagueness
	Ambiguity
	Discord

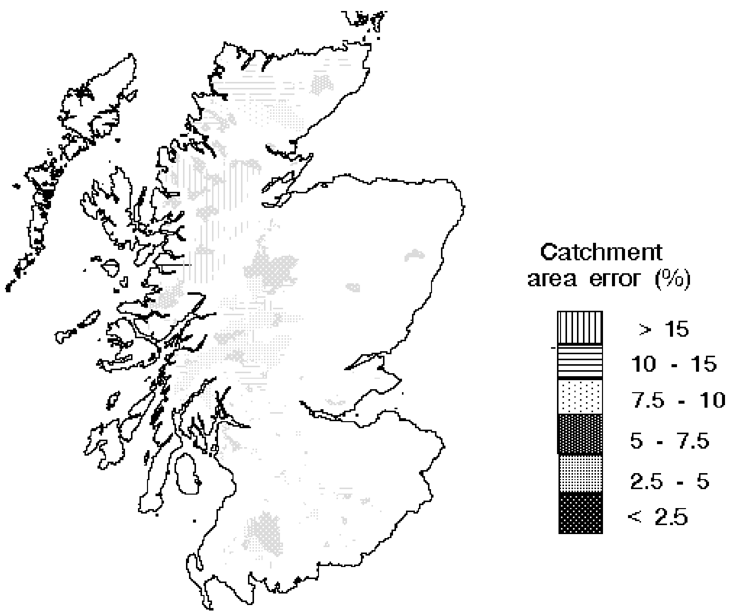
Uncertainty	Data Quality	
Error	Accuracy	Positional
		Attribute
	Completeness	
Vagueness, Discord and Ambiguity ?	Semantic accuracy	
Error, Discord, Vaguness and Ambiguity?	Currency	
Discord	Logical Consistency	
?	Lineage	

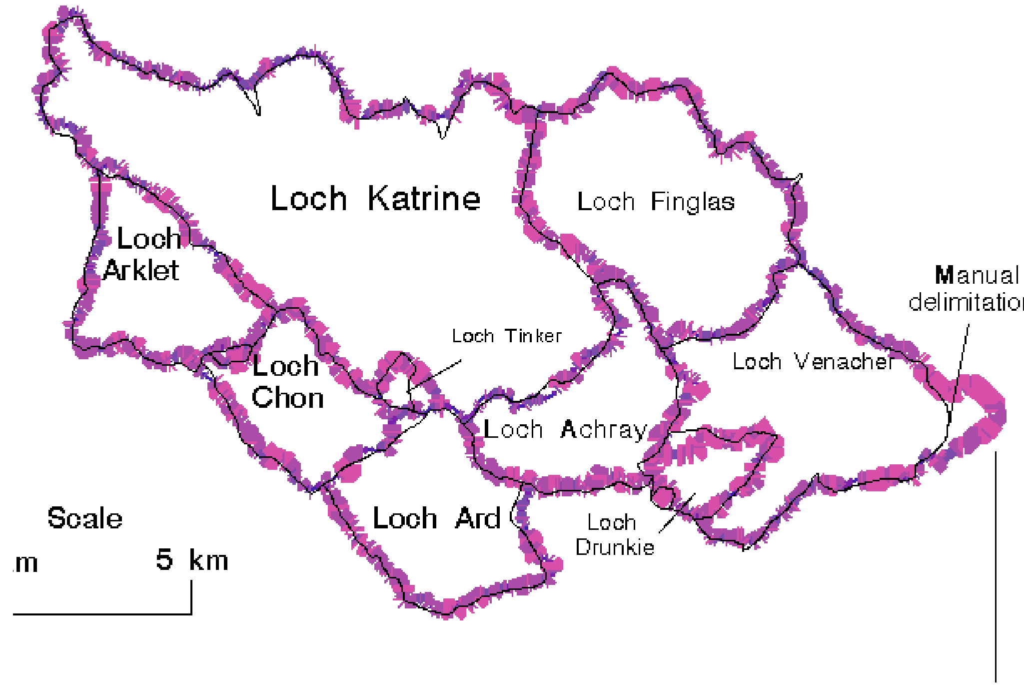
Some examples…

- Let’s look at some **examples** how important **uncertainty** can be
- What are the **potential** (or real) **consequences** of **decisions** based on **uncertain data** sources?
- Keep in mind how **SDQ** can be specified

How reliable is catchment delineation from topographic maps?

- What could be the consequences of unreliability in these boundaries??



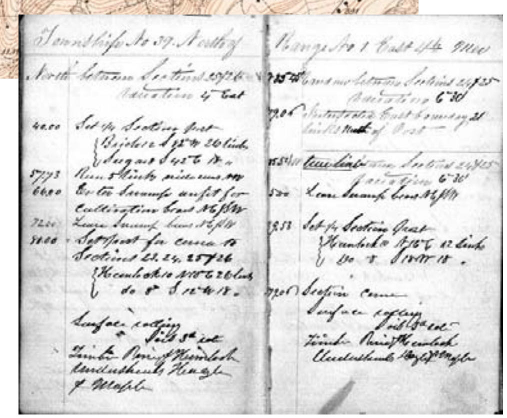


Boundary reliability: the broader the boundary line the lower the reliability

from http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/miller1_david/miller_paper1.html

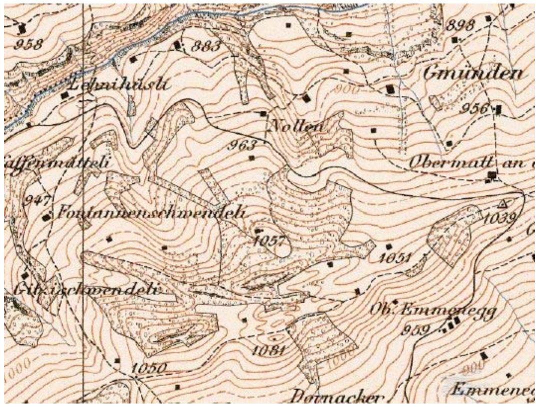
How reliable are estimations of land cover changes?

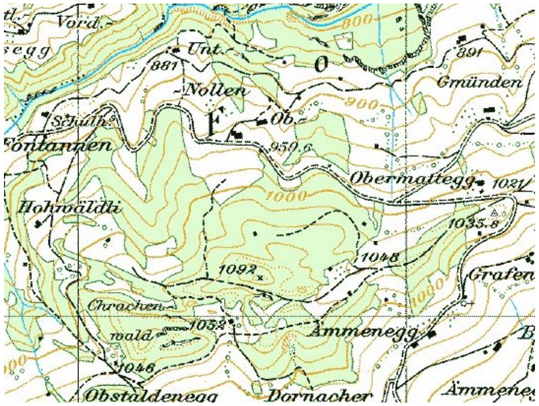
- Example Switzerland: “During the last 100 years the forest cover **increased by ca. 30%!!!”**
- Wow! But here’s the truth: We don’t know.



How reliable are estimations of land cover changes?

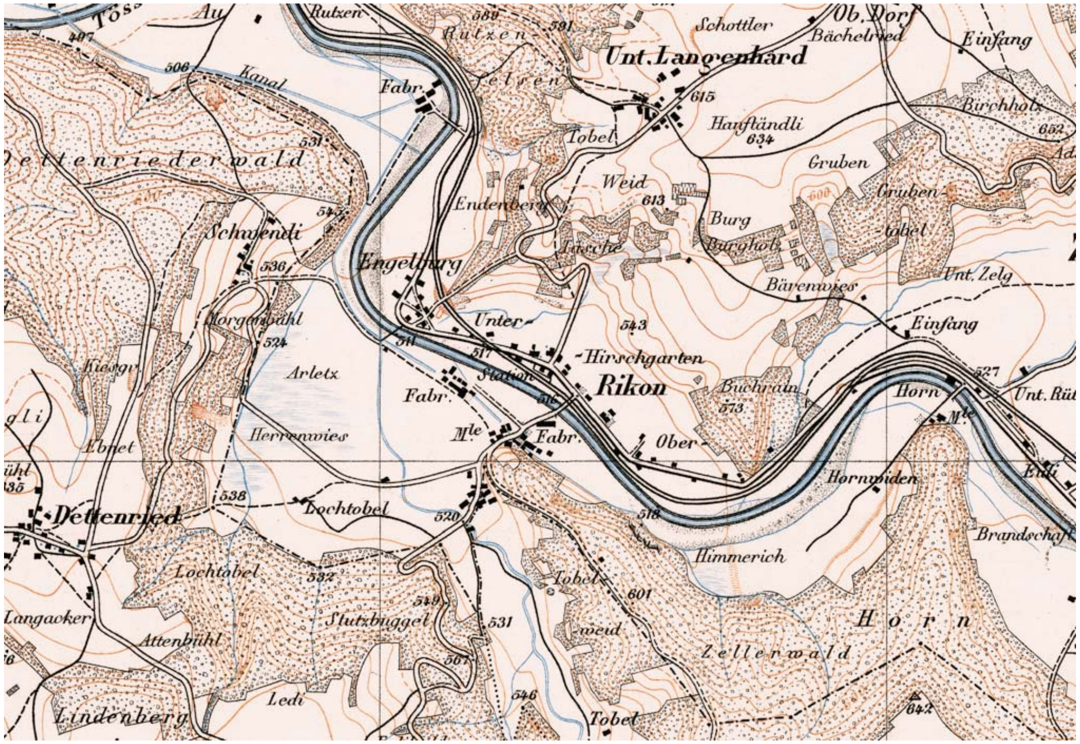
- **Forest definitions** in different time periods
- **Drivers** of forest mapping and **political interests** as well as **conflicts** (serious ones…)
- **Practice of mapping** in the 19th century
- What are the **materials available** to get evidence?





How to Approach & Apply Historical Spatial Data?

- … if you want to use them for GIS based change analysis…



Representing “on-the-ground” information

- How often is a value wrong (classified for **categorical** data)?
- **Numbers** (average error…) or error distributions…
- How do we “**conceptualize**” features?
- What is the **collection** method?
- What do we **misunderstand**?
- How much **up-to-date** are the data?

So where Comes Uncertainty from?

- Limitation in the data (raster cell's **resolution**)
- **Mixed categories** and homogenuous value of a raster cell ("loosing small fractions"): inclusion/generalization
- **Gradual changes** within **transition** zones
- **Collection**, map **production**, **processing**
- **Non-tracked changes** in land cover/ vegetation

Just Using Certain Data?

- There is no **perfect data**
- We have to be aware of **uncertainty** associated with the **data**, the underlying **meaning**, the **processing** steps to be carried out and the planned **application** (fit for use?)
- For this reason it is so fundamental to know more about uncertainty and its **influence to SDQ**
- Find some direction within the term clouds, first! We'll start with uncertainty and come to SDQ then
- Time for some definitions…

Definitions -- Uncertainty

- … **doubt** about the information which is recorded at a location (Fisher 2003)
- … a measure of the **difference** between the data and the **meaning** attached to the data by the current user Zhang and Goodchild (2002)
- … the result of **error**, **ambiguity**, **vagueness** or **lack of information** (Fisher 1999, Atkinson and Foody 2002) --> an **umbrella term** for these concepts

Uncertainty

- Fisher 1999

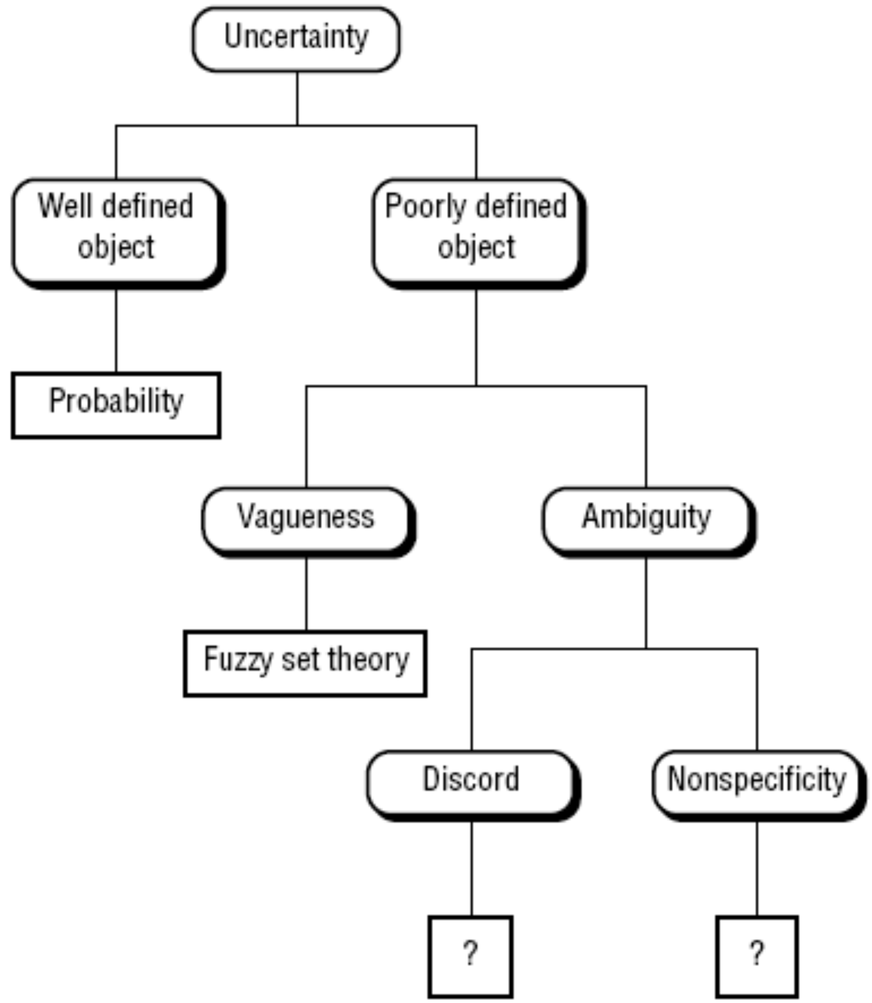


Fig 1. A conceptual model of uncertainty in spatial data (adapted from Klir and Yuan 1995: 268).

Definitions -- Uncertainty

- … **lack of knowledge** about:
- (1) objects of the real world due to **erroneous measurement**, **vague definitions** and **concepts** or unknown and **ambiguous meaning**;
- (2) effects of **transformations** performed on the data; and
- (3) the latter's suitability for the **intended application**
- Do you recognize three aspects here?

Definitions -- Error

- **Error**:
- **Difference** between a computed, observed or measured value or condition and the **true**, specified or theoretically **correct** value or condition (Oxford Reference Online 1996)
- describes the **deviation** of a value from **truth** (Jones 1997)
- **Inaccuracy** in cases of **systematic** error and **Imprecision** in cases of **random** errors
- Describing cases of **measurable deviation** from the true state where no problems of definition occur.

Definitions -- Accuracy & Precision

Accuracy:

Difference between a recorded value and its true value (often divided into spatial, topological and attribute accuracy). In practice **truth** is a **reference value**, which is assumed to be more accurate

Systematic error

Precision:

Detail with which a measurement is reported - there is no point reporting a measurement to a higher precision than that with which it is measured

How repeatable is a process or measurement?

Random error

Precision & Accuracy

high average accuracy,
high precision

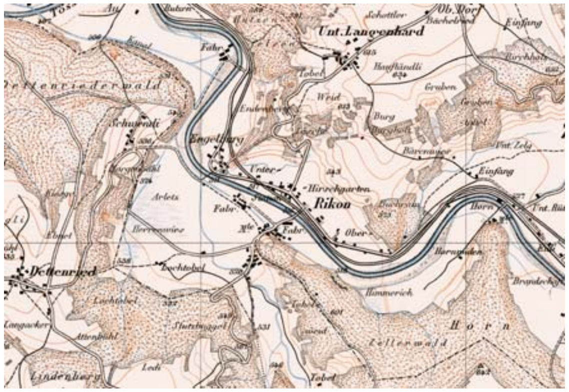
low average accuracy,
high precision

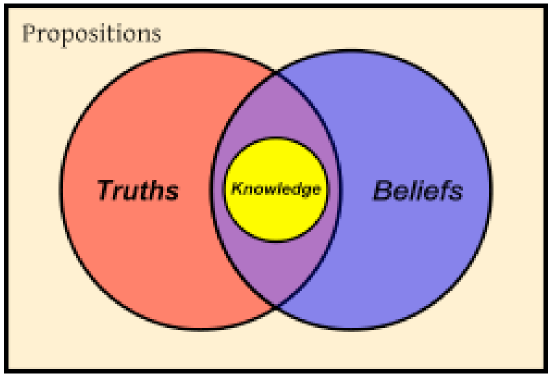
high average accuracy,
low precision

low average accuracy,
low precision

Vagueness

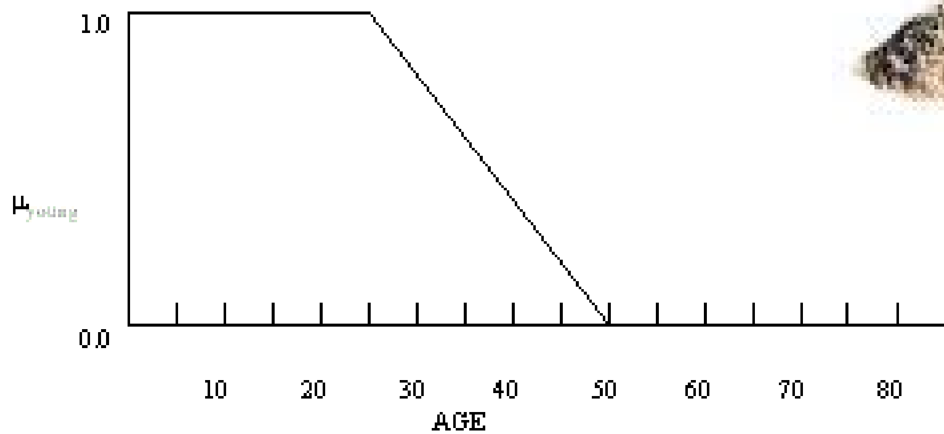
- **Indeterminacy** due to a lack of distinctness between **ill-defined** or fuzzy classes of objects or individual objects.
- In GIS vagueness in definition causes doubt over the **membership** of a considered location to one class or several classes
- Sorites paradox (Williamson 1994) & degree of truth
- (a) property of objects, (b) purely linguistic, (c) epistemic in nature or (d) purely semantic



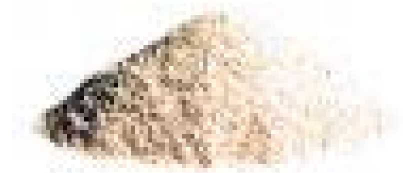


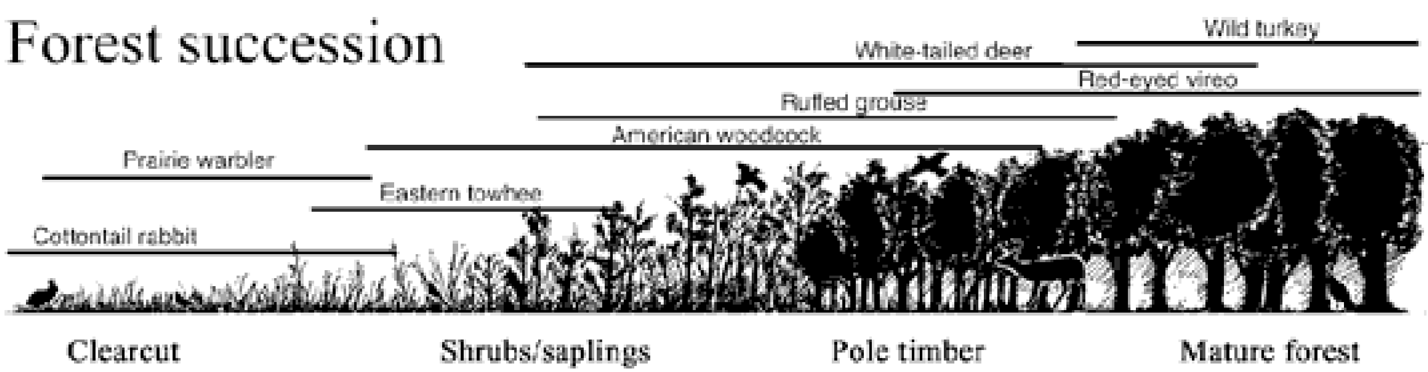
Vagueness

from www.nwhi.org/index/habdescriptions

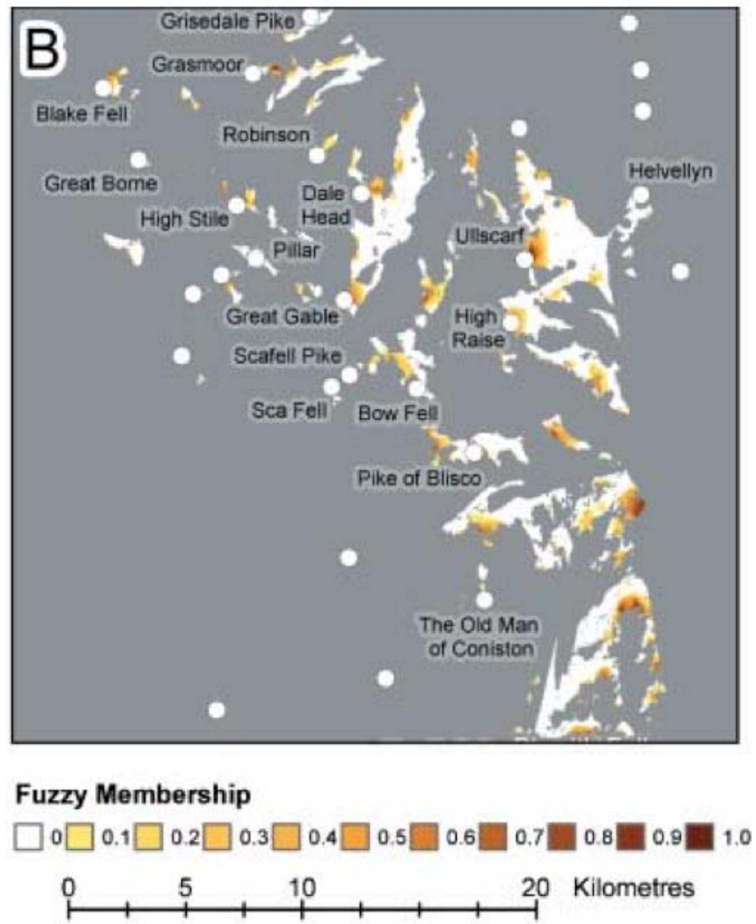


from http://www.iit.nrc.ca/IR_public/fuzzy/fuzzyJDocs/overview.html\





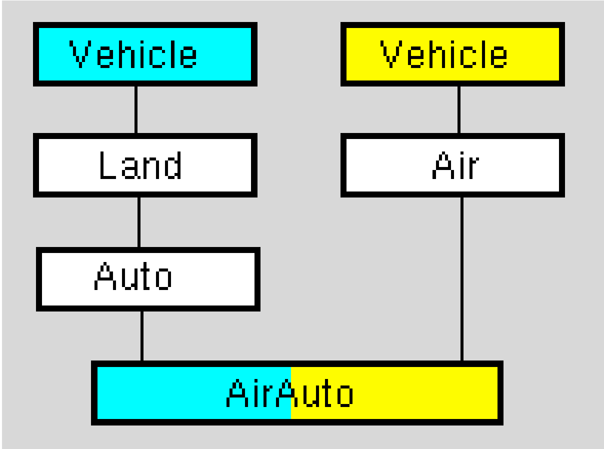
from hioline.osu.edu/b915/part_one.html

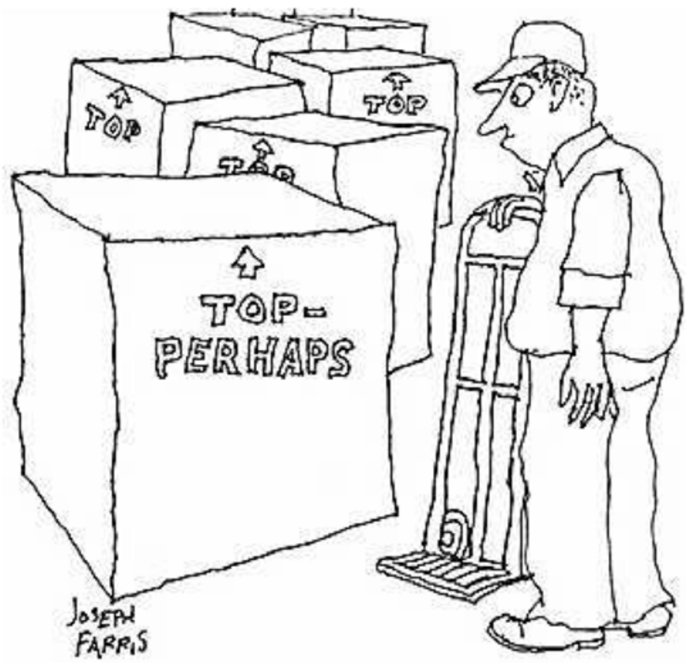


(from Fisher et al., 2004)

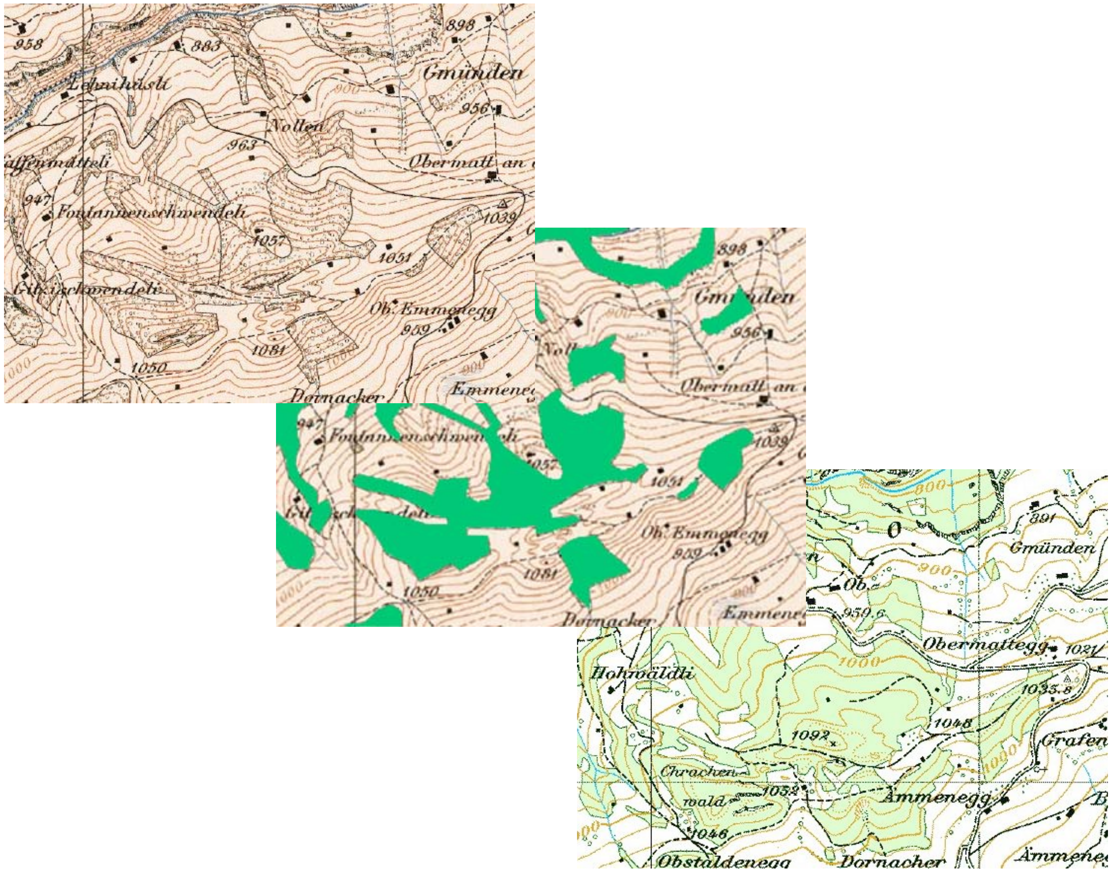
Ambiguity

- **Confusion** among concepts which have the same name, but more than one definition (Fisher 2000)
- **Discord**: lack of agreement if one object is clearly defined but is shown to be a member of different classes under differing classification schemes or interpretations
- **Non-specificity**: Occurrence of ambiguity if the assignment of an object to a class is unsettled at all (it is then a matter of interpretation and prone to subjectivity)





from http://www.cartoonstock.com/lowres/jfa0432l.jpg



A “Bringing-Together”

- **Error** is not due to problems of definition but due to measurable deviations from “truth”
- **Vagueness** arises due to overlapping definitions and is thus only considerable in the context of an environment with other objects or classes of objects
- **Ambiguity** is caused by definitions with different meaning under varying classification schemes (**discord**) or weak definitions (**non-specificity**) without consideration of its environment

Reporting Data Quality

- Data are passed around and manipulated by **many people**, within and between **organizations**, intentionally and unintentionally
- By reporting on data quality and understanding these concepts, we can **minimize uncertainty** or choose more appropriate products
- Burrough and McDonnell list factors affecting data quality...

Where we started…?

- … by asking: **What is quality**?

  - Reaching the **best result** possible…

  - Reaching the **requirements** defined by **standards** or **customers**…

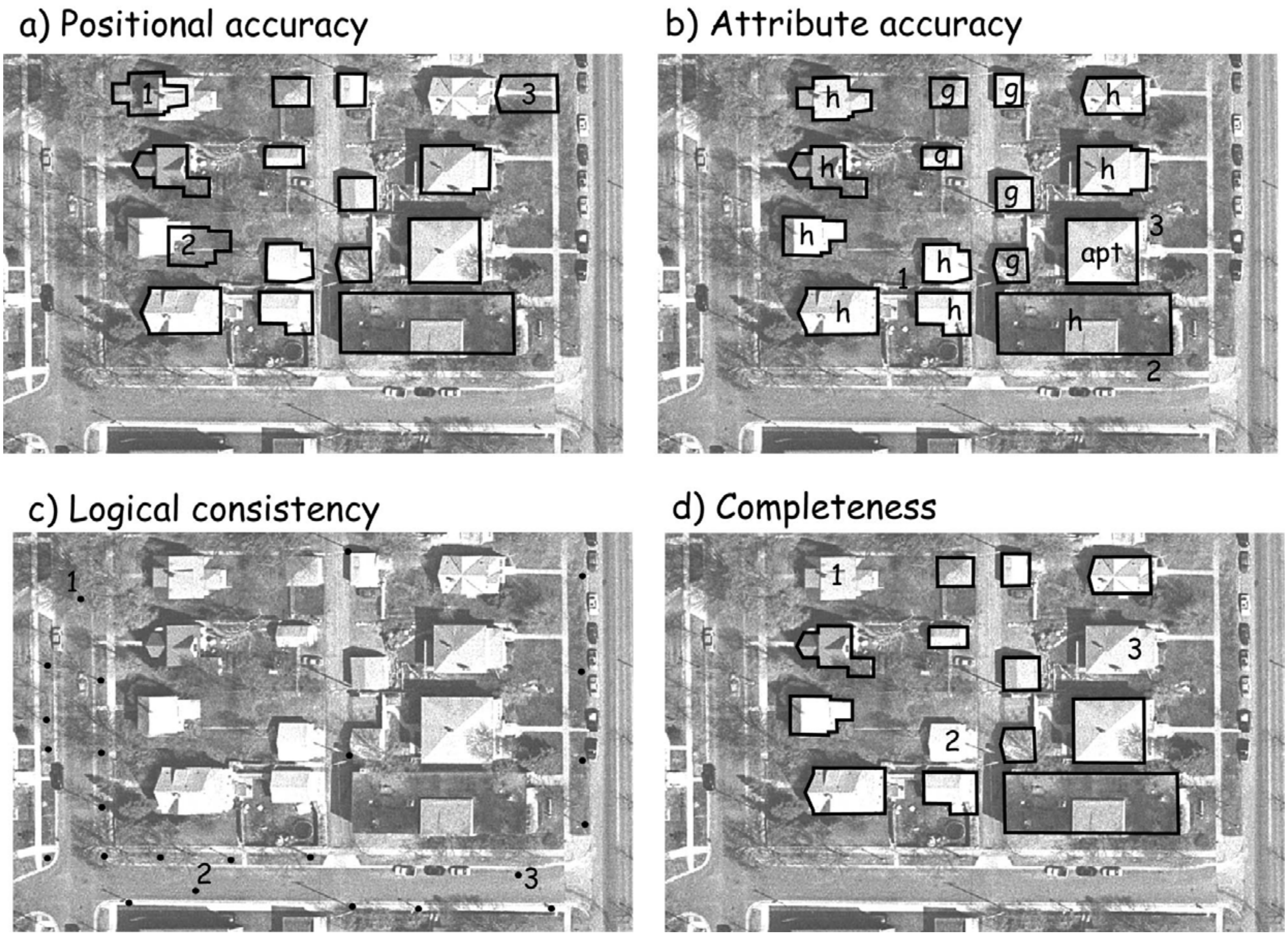
  - Reaching the requirements for a specific **application** -- "**Fitness for use**"
- In summary this seems to be about **uncertainty**, how it **arises** and **propagates** throughout the process the spatial data are applied to

SDTS

- Developed in the US to allow **transfer** of data between organizations using a defined and agreed **standard**
- **Spatial Data Transfer Standard** (SDTS 1992) is **obligatory** for US Federal Organisations to use the SDTS
- Included **compulsory** data quality fields (note these are only a small part of Burrough and McDonnell’s list of factors affecting data quality)

The SDTS famous Five

- **Positional accuracy**: e.g. the accuracy with which the positions of depth soundings were reported
- **Attribute accuracy**: e.g. the accuracy of the actual depth soundings; the classification of a pixel which is bog as urban area
- **Logical consistency**: e.g. are values valid (is there a class called "bag", are there topological problems (e.g. overlapping polygons with different classes)
- **Completeness**: e.g. Has a polygon not been digitized, what was the minimum size of object included, what was the density of observation?
- **Lineage**: Who produced the data, what methods did they use, why were the data produced, when?


a) Positional accuracy
1
3
2
b) Attribute accuracy
h
g
apt
1
2
3
c) Logical consistency
1
2
3
d) Completeness
1
2
3

The Famous Five???

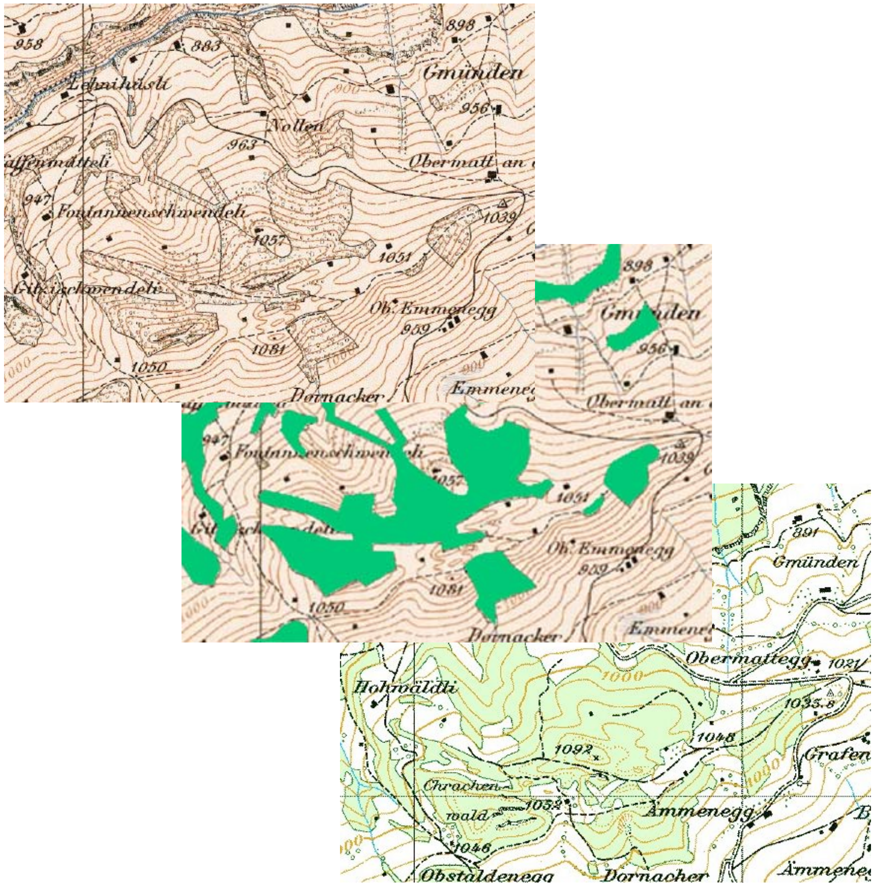
- Things change… from SDTS (1992) to…
- Guptill S C and C Morrison J L (eds) 1995 Elements of Spatial Data Quality.

Data Quality	Lineage	
	Accuracy	Positional
		Attribute
	Completeness	
	Logical Consistency	
	Semantic Accuracy	
	Currency	

- So, how to bring them (SDQ, Uncertainty types, Uncertainty domains,…) together now?

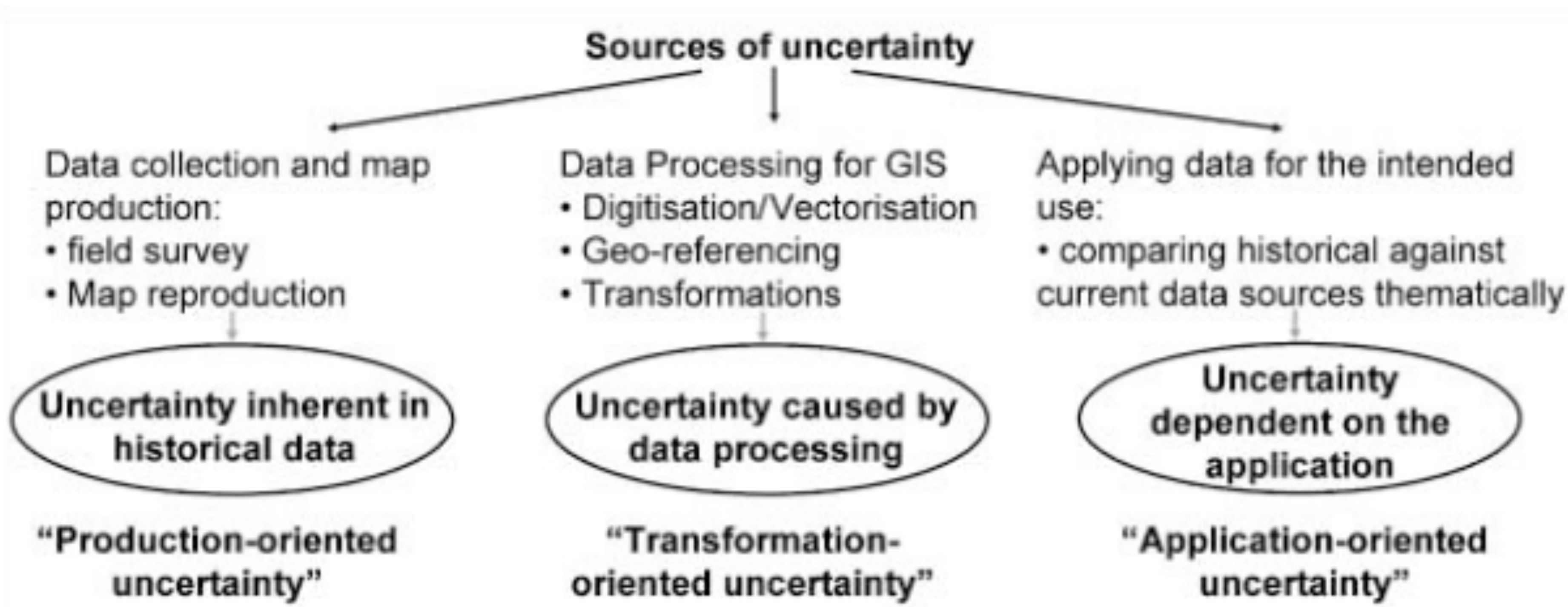
Remember…

- Uncertainty is considered to embrace the **unknown domain how good** our data are
- However uncertainty **influences** our **spatial data quality (is our data good enough in relation to standards?)**
- Uncertainty can be **introduced** at any stage of **GIS-based map production** and analysis (Reality observation, conceptual modeling, measurement, analysis steps, and … use of the data)



… where Uncertainty Comes from…

- The case of historical spatial data:



Metadata…

Summary

- **Uncertainty** and **Spatial Data Quality** have much in common, they consider **similar** topics and have related categories or domains
- **Uncertainty** embraces the **unknown domain** to give the basis for judgments, **how good** the data are
- **SDQ** rather allows to ask if our data are **good enough** with regard to **standards** or expectations
- Uncertainty in Spatial information is still an open research environment due to the many **unresolved questions** existing
- New **concepts** and methods such as from **fuzzy logic** have been introduced into **uncertainty research**
- As you can expect, **fitness for use** and **use error** are a very important part for industrial developments of spatial data technology

References

- Burrough, P.A. and McDonnell, R.A. (1998):Principles of Geographical Information Systems. Second Edition. Oxford University Press.
- Jones, C.B. (1997): Geographical Information Systems and Computer Cartography. Longman.
- Longley et al. 2001. Geographic Information Systems and Science. Wiley.
- Fisher P 1999 Models of uncertainty in spatial data. In Longley P, Goodchild M F, Maguire D J, and Rhind D W (eds) Geographical Information Systems: Principles, Techniques, Management and Applications (Volume 1). New York, John Wiley and Sons: 191–205
- Fisher P 2003 Data quality and uncertainty: Ships passing in the night! In Shi W, Goodchild M F, and Fisher P (eds) Proceedings of the Second International Symposium on Spatial Data Quality. Hong Kong, Hong Kong Polytechnic University: 17–22
- Guptill S C and C Morrison J L (eds) 1995 Elements of Spatial Data Quality. Oxford, Pergamon
- … if you like endless reference lists: Leyk et al., 2005 in TGIS