University of Leicester

Semantic Uncertainty

by

Peter Fisher

Department of Geography, University of Leicester, UK

RSS/FERA workshop: Analaysing and Visualising Spatial Uncertainty

FERA York, 27th January 2010

Acknowledgement

- Many collaborators and colleagues over the years
- Particularly want to thank
  - Lucy Bastin
  - Lex Comber
  - Richard Wadsworth

Outline

1. Semantics
2. Examples of the problem of semantics
3. Solutions - Standards
4. Solutions - Discord-based
5. Solutions - Vagueness-based
6. Conclusions

Semantics

- Study of meaning
- Significance of signs
- Referring here to
  - The meaning in words
  - Vocabulary
- It is a very pressing problem and not a statistical one

Motivation

- To improve the use of GI …
- by …
- Enhancing the understanding of terms that might be used by
  1. Data Producers
  2. Data Users
  3. Anyone else who might be involved

Semantics in Geographical Information

- Semantics are fundamental in GI
- ALL classed information carries class labels
- Class labels tend to be words with specific meanings *and* often general meanings

- Data themes themselves carry semantic descriptions too

Class data descriptions can be

- Single words
  - Building, Woodland, Grassland, etc.
- Short phrases
  - Dense Dwarf Shrub Heath
- Extended descriptions
  - Species list with frequencies and/or necessity or otherwise of occurrence

Semantics Examples

1. Landcover Mapping
2. Soil Mapping
3. Forest mapping

- Could be any number of other information themes

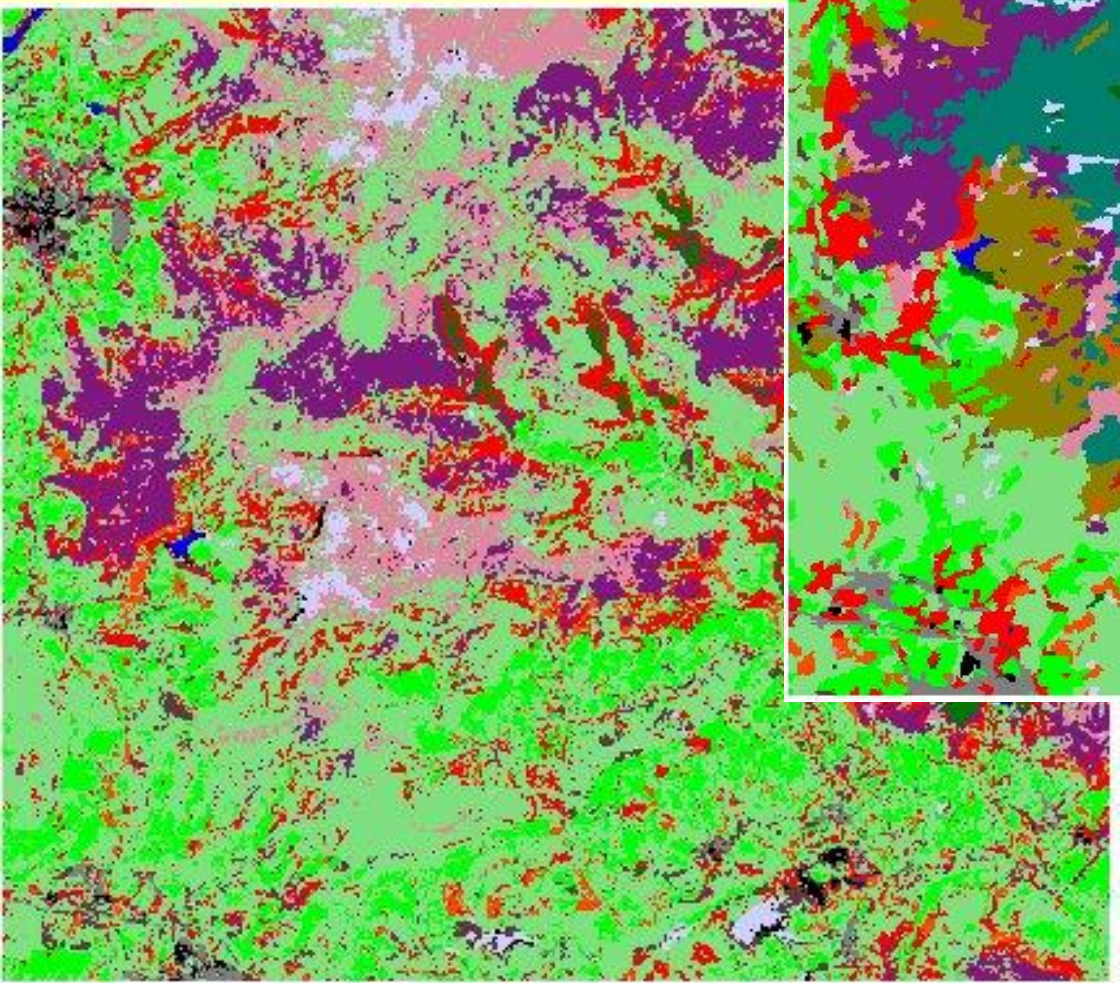
2. Examples A

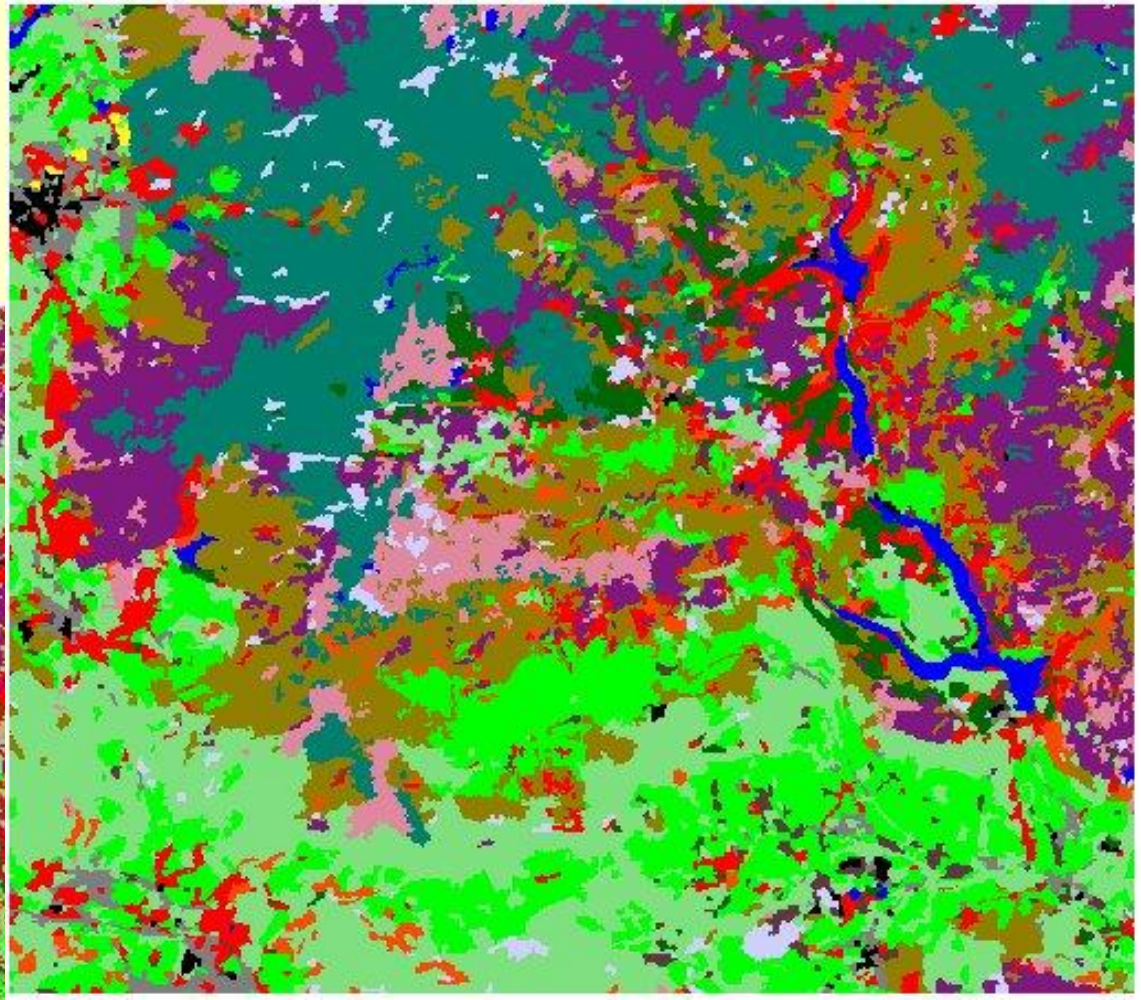
UK LCM2000

- Each phrase or word has a technical meaning defined in the supporting document
- AND
- Each has a general meaning

2. Examples A

Changing Landscape?





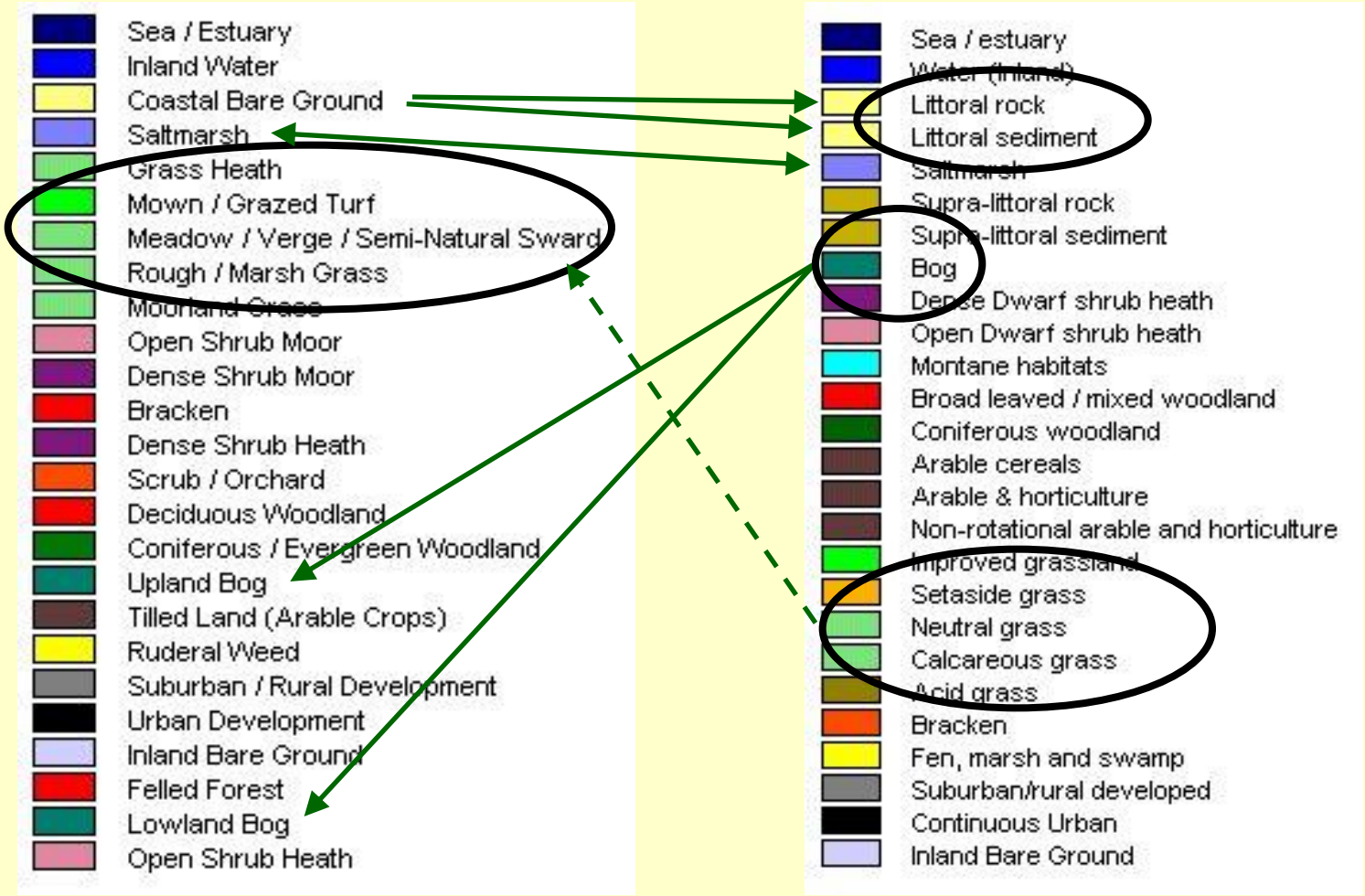
LCM 2000

LCMGB 1990

But the definition of "bog" has changed!

- LCMGB
  - 12 pixels (<1 ha) in SK tile
    - permanent waterlogging, resulting in depositions of acidic peat
    - mostly herbaceous communities of wetlands with permanent or temporary standing water
    - Lowland Bogs: carry most of the species of upland bogs, but in an obviously lowland context, with Myrica gale and Eriophorum spp. being highly characteristic.
    - Upland bogs: have many of the species of grass and dwarf shrub heaths
    - characterised by water-logging, perhaps with surface water, especially in winter. species such as bog myrtle *(Myrica gale)* and cotton grass (*Eriophorum* spp.) in addition to the species of grass and dwarf shrub moorlands.
- LCM2000
  - 120728 pixels (75 km^2) in SK tile
    - Bogs include ericaceous, herbaceous and mossy vegetation in areas with peat >0.5 m deep; ericaceous bogs are distinguished at subclass level. Inclusion of Ericaceous bogs contrasts with LCMGB 1990 where bogs were herbaceous or mossy in seasonal standing water
    - Peat depth determined from geological mapping !

Two Land Cover mappings of GB



LCMGB 1990

LCM 2000

Further example descriptions
LCM2000 Classes

1. Broad-leaved, mixed and yew woodland

   Broad-leaved, in stands > 5 m high with tree-cover > 20%; or scrub < 5 m and yew woodland with cover >30%. Mixed woodland is included if broadleaved trees in conifers cover > 20%. Stands ≥ 0.5 ha are mapped as separate blocks.

2. Coniferous woodland

   Coniferous woodland, semi-natural and plantations, with cover > 20%, and recently felled forestry. Once felled areas are colonised by rough grass, heath or scrub, they take that class.

3. Boundaries and linear features

   Larger linear features such as shelter belts or motorways; smaller linear features (hedges, walls, smaller roads) are only recorded by the field survey.

4. Arable and horticulture

   Annual crops, recent leys, freshly ploughed land, rotational setaside, and perennial horticulture crops such as berries and orchards. Once setaside is substantially vegetated with weeds or rough grass, it is included in the Improved grassland Habitat.

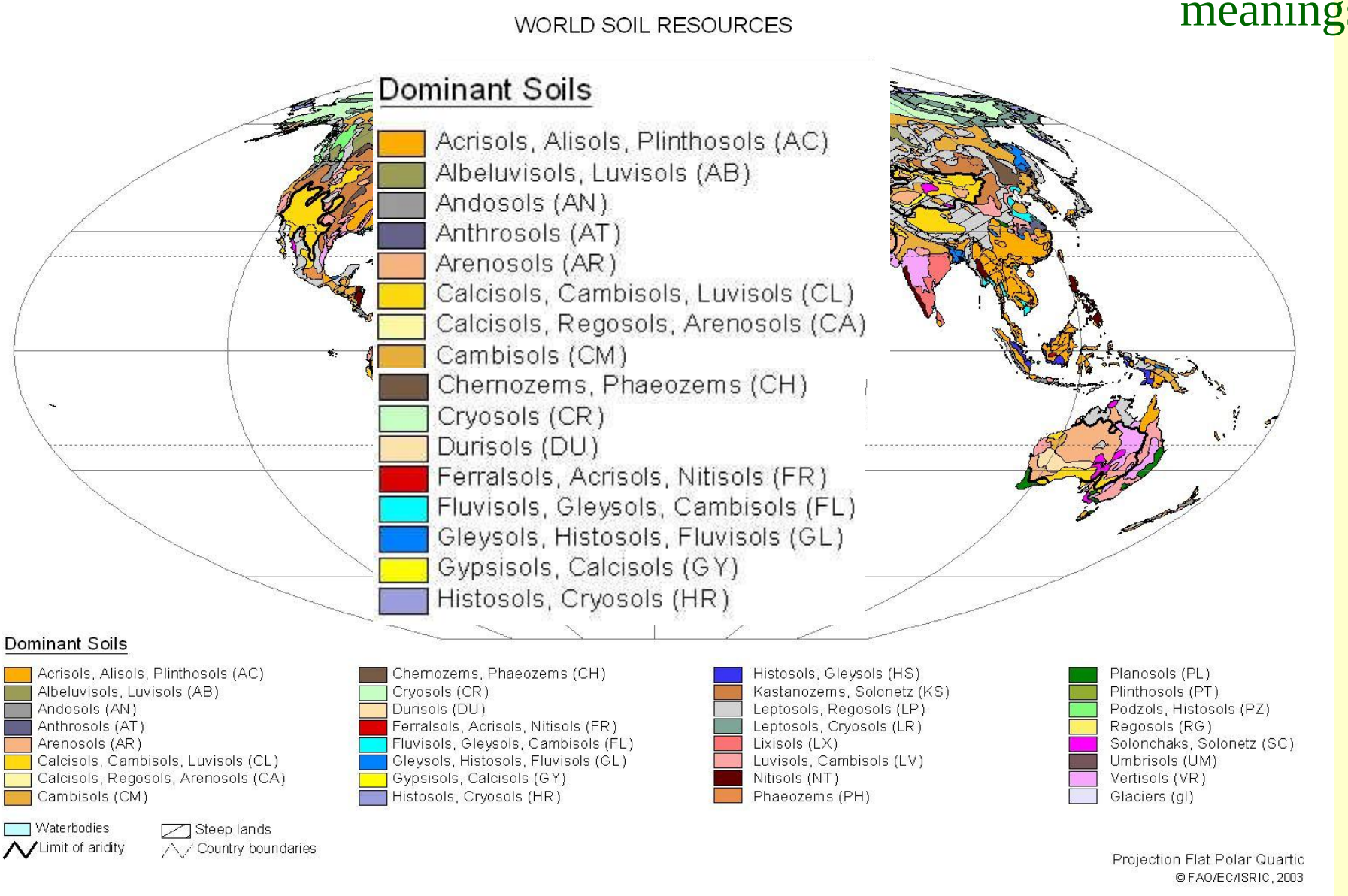
Definitions typified by

- Thresholds (trivial? Are they implemented literally?)
  - > 5 m Will 4.5 m be excluded if it is over a small area (0.5 ha)? Is it observable in a satellite image?
  - > 20% Will an area with only 19% be excluded?
- Fuzzy (vague) terms
  - Semi-natural
  - Recently
  - Smaller
  - Freshly
  - Substantially

2. Examples B

FAO Soil Map of the World

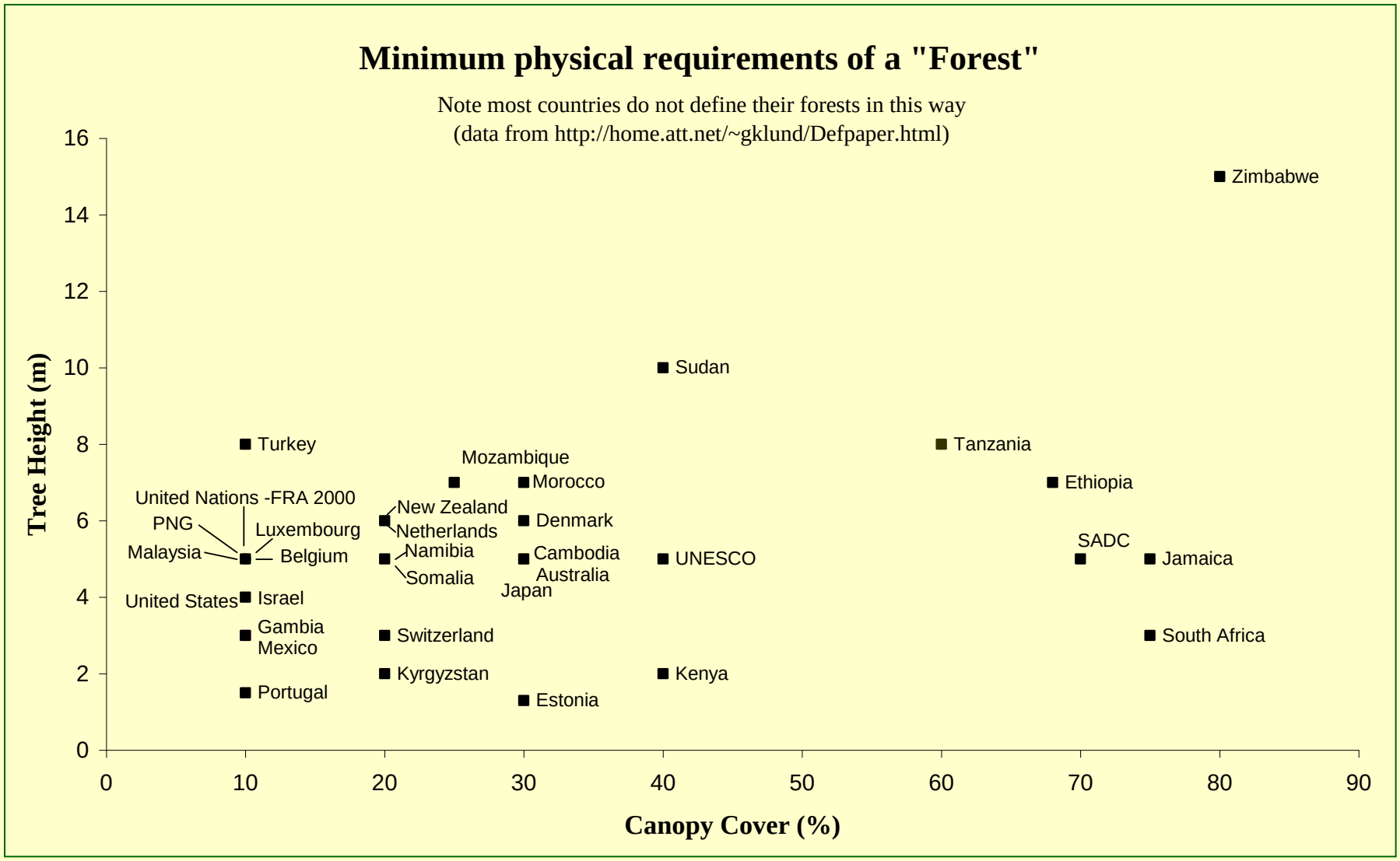
Uses special terms with well-defined meanings



But is it universal?

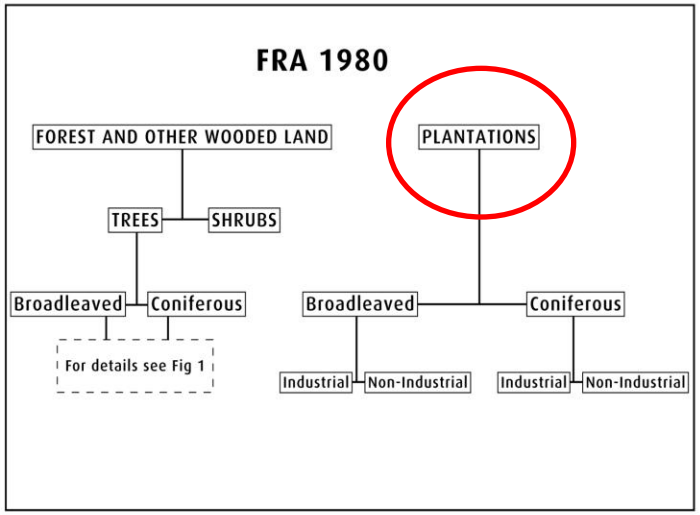
- US *Soil Taxonomy* is another international classification system
- AND
- Every country has its own national Classification encompassing the soils of the country
- National mapping agencies used to spend time preparing comparative tables of classes (as did the USDA-SCS!)
- But how many people know this?
- Other than soil surveyors? But who else would understand it?

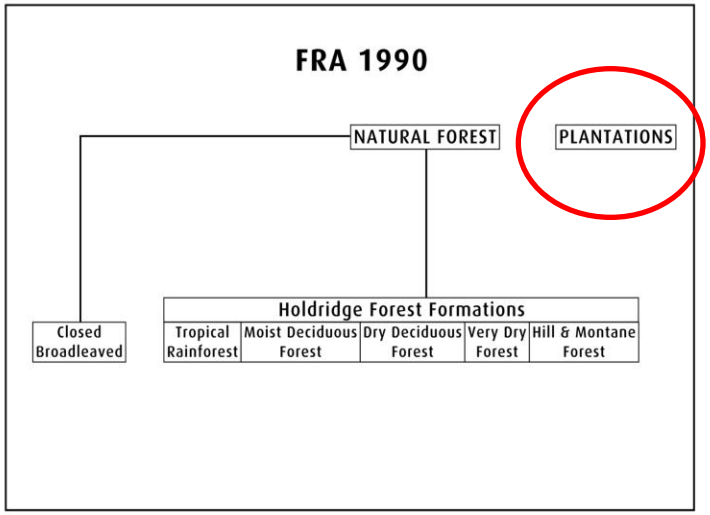
International Definitions of "Forest"



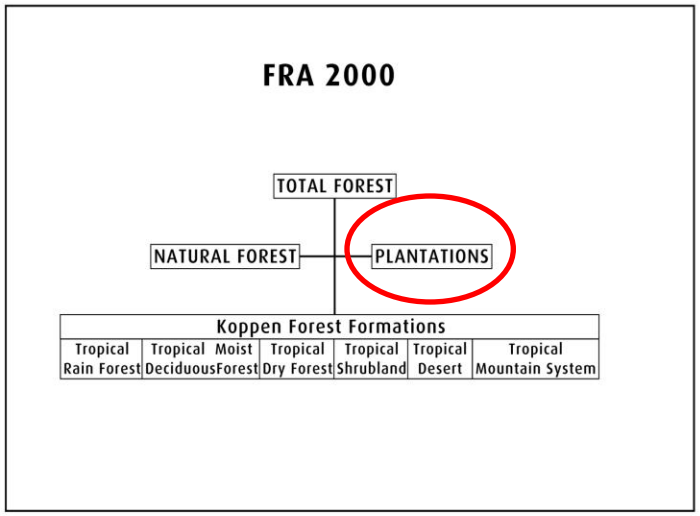
2. Examples C

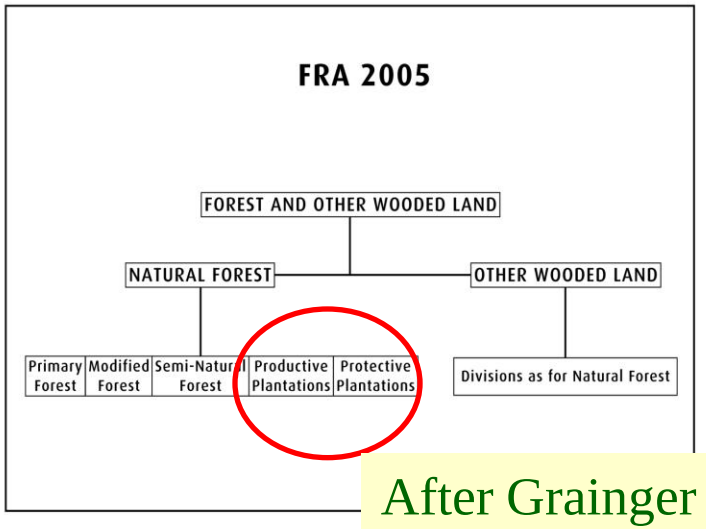
Position of "Plantation" in FAO Forest Resource Assessment





FAO - Forest Resource Assessments – categories change their meaning





After Grainger 2007

The future ?

- It is believed that computers will
- “replace the extended and often confused process by which we learn the meanings of terms and languages with precise, instantaneous translators” (Goodchild, 2006, TGIS 10, p690).
- But how? Via METADATA
- Can we discern any progress?

Metadata

- Metadata is …
  - "information that helps the user assess the usefulness of a dataset relative to their problem"
    - not usability as anything is useable but, whether it is any good is another question
- The 'big 5' of geo-spatial data quality in metadata:
  - Positional Accuracy, Attribute Accuracy, Logical Consistency, Completeness, Lineage.
- Salgé (1995) tried to introduce the concept of semantic accuracy but has largely been ignored
- SO you might expect semantics to be an issue for Metadata
- BUT

Metadata focuses on

- Allowing individuals within an organization to discover the existence of any dataset of interest and how to acquire it
- Helping organize and maintain an organization's internal investment in spatial data
- Providing information about an organization's data holdings to data catalogues, clearinghouses, and brokerages
- Providing information to process and interpret information received from an external source

Semantics in Metadata?

- Lists of data categories are included in *some* standards

  FGDC, 1998; ANZLIC, 2001; ISO, 2003a, 2003b

- BUT they tend to be *optional* !
- They are frequently present, however
  - Short or long descriptions
  - General or technical vocabulary

Standards for semantics of GI

- Existing
  - Biological metadata standard (FGDC, 1999)
  - Vegetation classification standards (FGDC, 1997a)
  - Soil information standard (FGDC, 1997b)
  - Shoreline information standard (FGDC, 2001)
  - Cadastral mapping standard (FGDC, 2003)
  - Wetland classification standard (Cowardin *et al.*, 1979)
- Proposed:
  - UN LCCS (Di Geogorio and Jansen, 2000)
  - Geological

FGDC Metadata Standards

- For soils we have established classifications
  - In Soil Taxonomy, FAO classificationManuals of survey and laboratory methods
  - Standard emphasises how record is to be created
- Vegetation and Wetland does not have established classification
  - Standard concentrates on classification scheme

Standards are

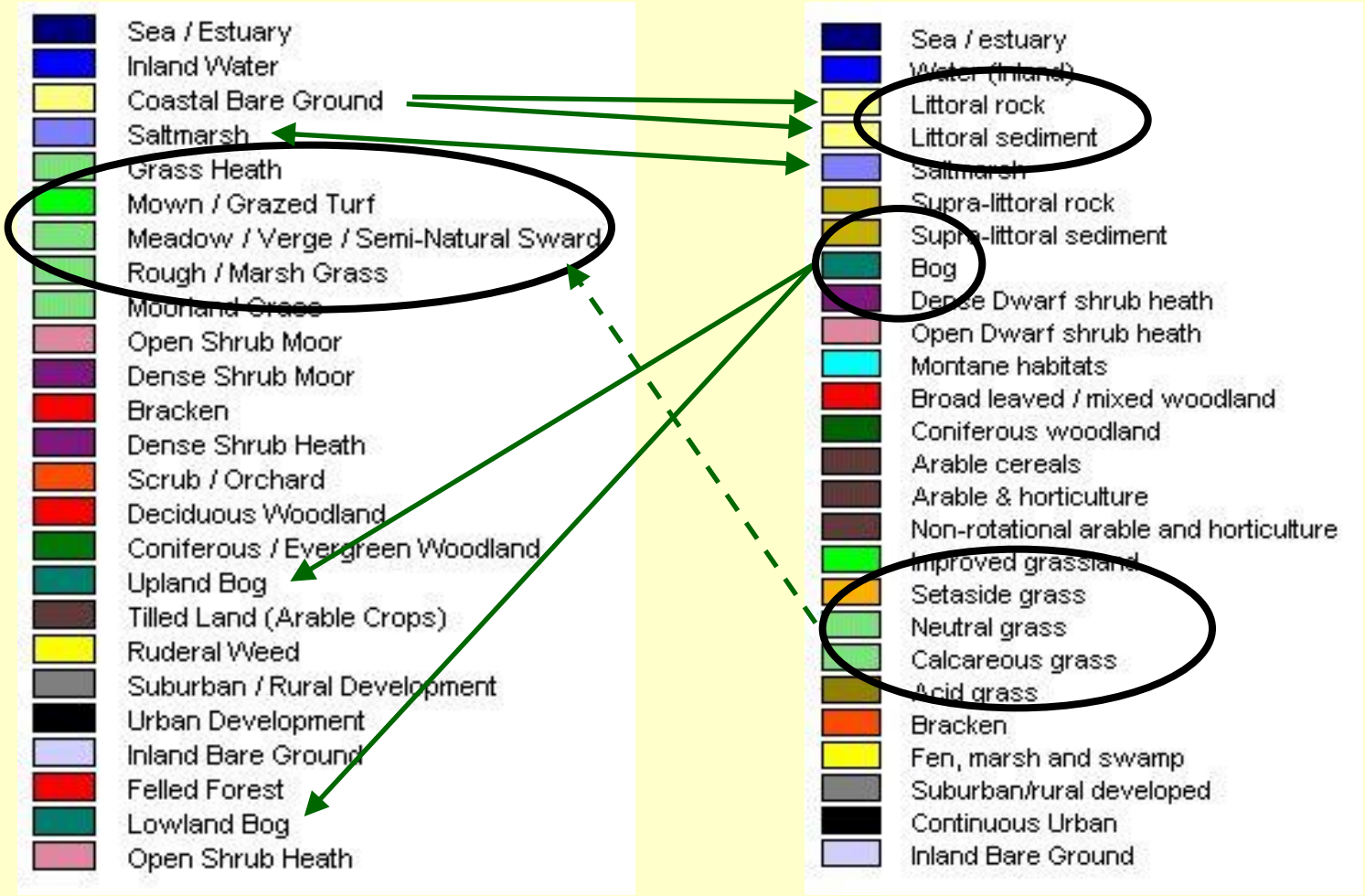
- Domain specific
- Varied in their
  - Starting point

  And so their
  - Content
- Most seem to be
  - More concerned with establishing a *de jure* ontology
  - None are concerned with reporting of *ad hoc* classifications

Discord

- Where two classifications of an area exist
- And they purport to map the same thing
- BUT
- They use completely different classification schemes
- Discord results in
  - Disagreement (mine is better)
  - Confusion (What *is* here?)
  - Violence (fight to support *my correct* interpretation)

Confused or confusing concepts!



LCMGB 1990

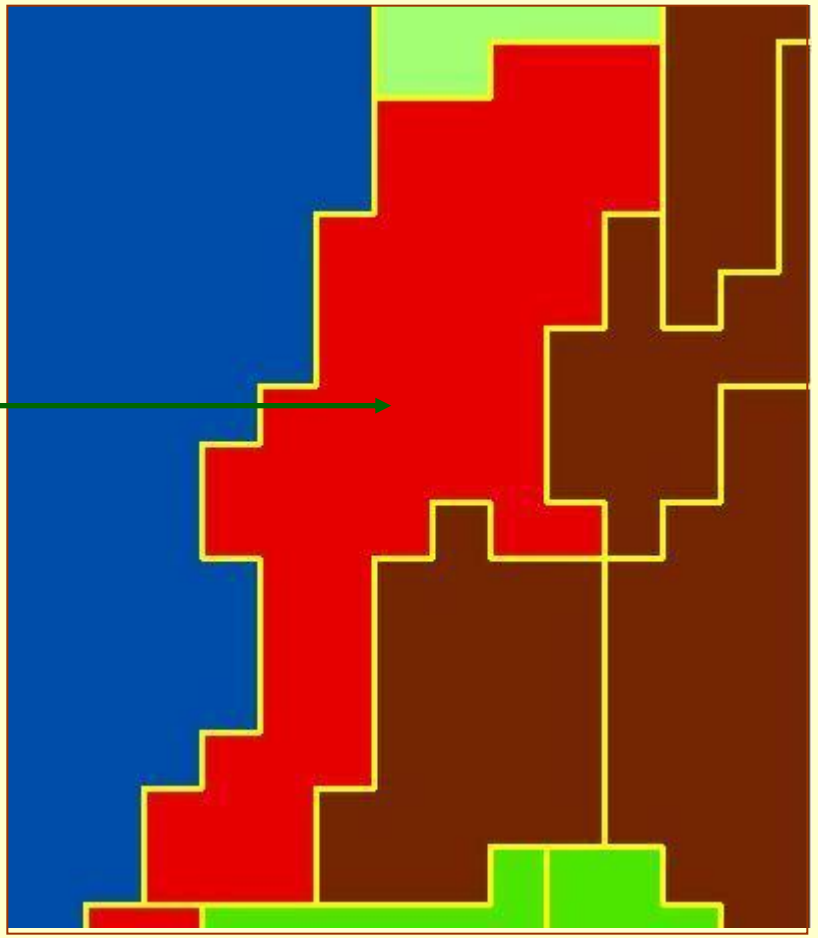
LCM 2000

We have used

- Used metadata from LCM2000
- Expert opinion on semantic, technical and change comparisons
- Dempster-Schaefer's Theory for combining the evidence
- To map incompatibility and possibility of change

Object-based Metadata for LCM 2000
Usually thrown away!!!

Field	Value
FID	4284
Shape	Polygon
SEGID	SK027319r1
TOTPIXELS	59
COREPIXELS	12
BHSUBVAR	1.1.1
BHSUB	1.1
SEGID_1	SK027319r1
HIST1	35
HIST2	50
HIST3	0
HIST4	0
HIST5	0
HIST6	0
PIXA	D_c
PIX1	37
PIXB	Au_y
PIX2	19
PIXC	D_a
PIX3	15
PIXD	Dm_b
PIX4	10
PIXE	Au_d
PIX5	5



4. Solutions – Discord-based

Cross classification Look-Up Table

LCMGB vs LCM2000

+1 Expected relation

-1 Unexpected relation

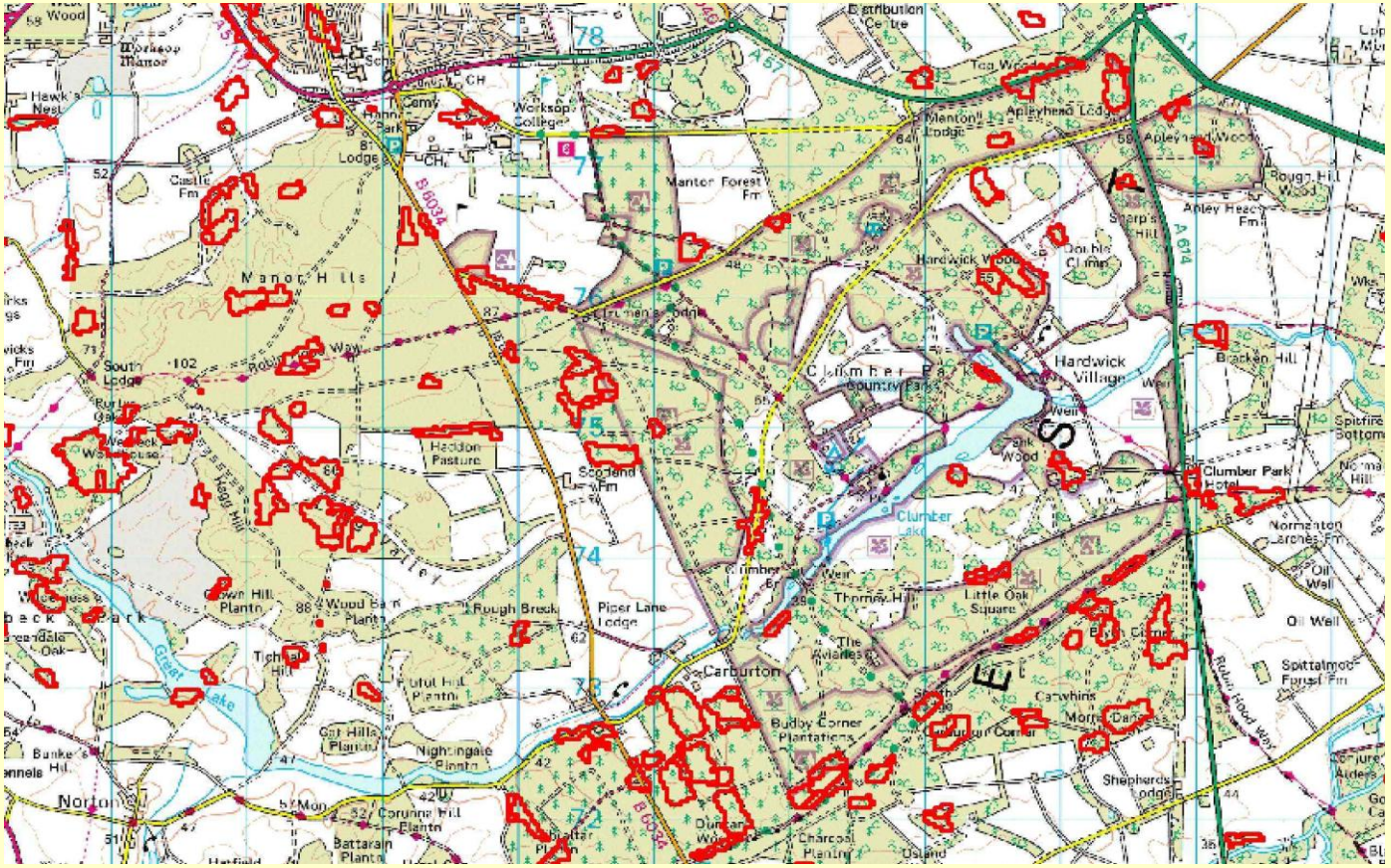
0 Uncertain relation

	Sea / Estuary	Inland Water	coastal bare?	Saltmarsh	Grass Heath	Mown / Grazed Turf	pasture / meadow / amenity grass	Rough / Marsh Grass	Moorland Grass	Open Shrub Moor	Dense Shrub Moor	Bracken	Dense Shrub Heath	Scrub / Orchard	Deciduous Woodland	Conifer	upland bog	Tilled Land	Ruderal Weed	Suburban / Rural Development	Urban Development	recently felled?	Inland Bare Ground	lowland bog	Open Shrub Heath
yew	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	-1	0	1	1	1	-1	-1	-1	0	-1	0	-1	-1	-1
	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	1	-1	-1	-1	-1	-1	0	-1	-1	-1
Arable Cereals	-1	-1	-1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	0	-1	-1	-1	0	-1	-1
Arable Horticulture	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	-1	-1	-1	-1	0	-1	-1
Non-rotational Arable and	-1	-1	-1	-1	-1	1	1	-1	-1	-1	-1	-1	-1	0	-1	-1	-1	1	0	0	-1	-1	-1	-1	-1
Improved Grasslands	-1	-1	-1	-1	0	1	1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
Setaside Grass	-1	-1	-1	0	0	0	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	1	-1	-1	0	0	-1	-1
Neutral Grass	-1	-1	-1	0	0	1	1	1	0	0	0	0	-1	-1	-1	-1	0	0	0	-1	-1	-1	-1	-1	-1
Calcareous Grass	-1	-1	-1	0	0	1	1	1	0															-1	-1
Acid Grass	-1	-1	-1	0	0	1	1	1	0															-1	-1
Bracken	-1	-1	-1	-1	-1	-1	-1	1	-1															-1	-1
Dense Dwarf Shrub heath	-1	-1	-1	-1	0	-1	-1	0	0	0	0	0	1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	0	0
Open Dwarf Shrub heath	-1	-1	-1	-1	0	-1	-1	0	0	0	0	0	0	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	0	1
Fen, Marsh Swamp	-1	0	-1	0	-1	-1	-1	1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	0
Bog (deep peat)	-1	-1	-1	0	-1	-1	-1	-1	0	0	0	-1	0	-1	-1	-1	1	-1	-1	-1	-1	-1	-1	1	0
Water	1	1	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	0	-1
Montane	-1	-1	-1	-1	-1	-1	-1	0	1	0	0	0	0	-1	-1	-1	1	-1	-1	-1	-1	-1	-1	0	0
Inland Bare Ground	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	-1	-1	-1	-1	-1	1	0	-1
Suburban/rural Development	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	0	-1	-1	-1	-1
Continuous Urban	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	1	-1	-1	-1	-1
Sea Estuary	1	1	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

Processing steps

- Each parcel characterised twice:
  1) Expert opinion on relations between LCM2000 and LCMGB classes
  2) Spectral relations from parcel metadata (2000)
- Difference in the characterizations indicates inconsistency in the parcel
- High degree of inconsistency might indicate change (or classification error in 1990 or 200)

Dempster-Schafer Theory – Semantic, Technical and Change combined



Results

- One expert's semantic table
- Parcels with high ΔE and ΔU → *belief*, *disbelief* in hypothesis of change
- 100% *Inconsistency* identified
  - Change / Error 1990 / Error 2000
- For 3 classes
  - 41% of inconsistent parcels had changed
  - 59% of inconsistent parcels misclassified in LCMGB or LCM2000 or both

Reporting

- Latest approach uses text mining of class descriptions
- Results published variously
  - Comber, Fisher and Wadsworth
  - PERS, IJGIS, JLUP, JoEM…etc
- Similar research using different approaches published by
  - Ahlqvist IJGIS
  - Fritz and See IJGIS

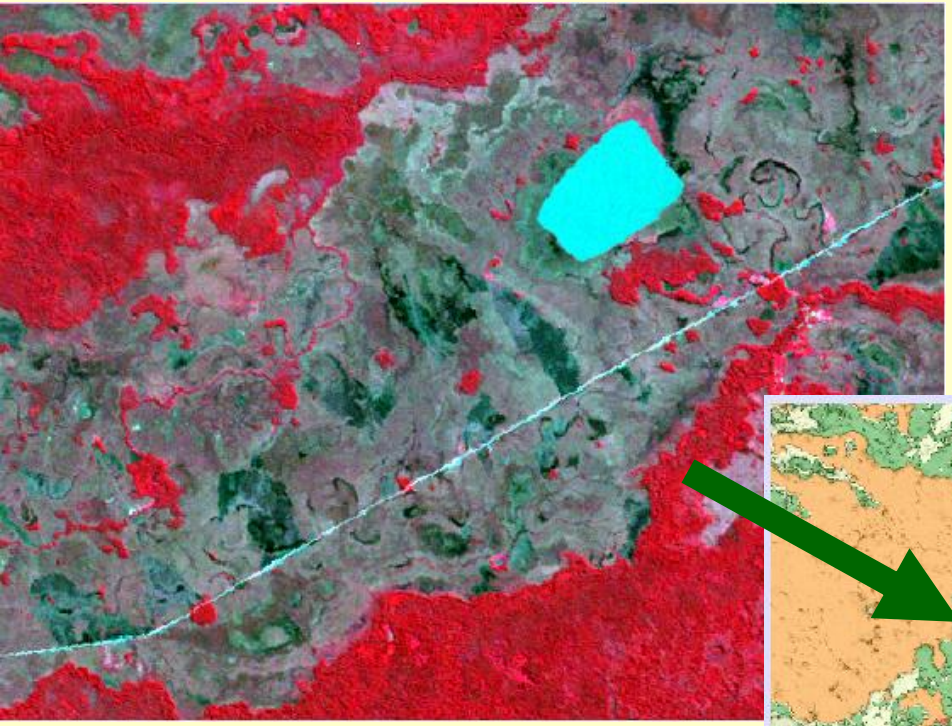
Land cover as fuzzy sets

- Even where mapped as Boolean sets

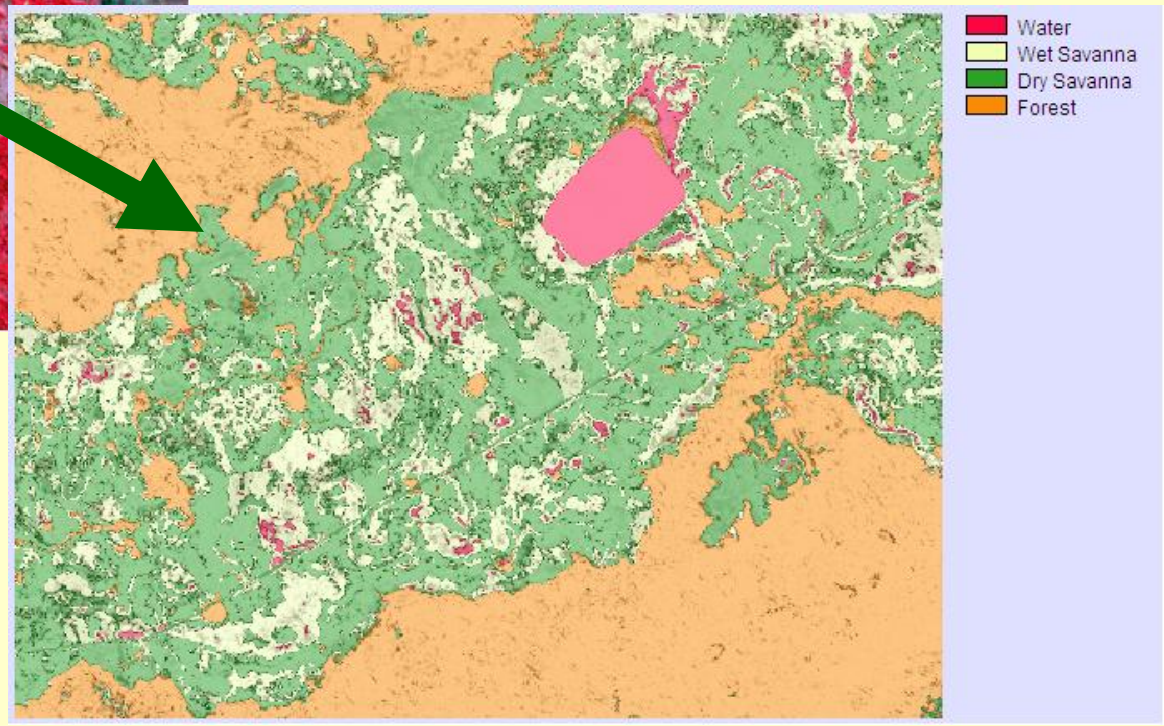
- Land cover classes are mostly poorly defined (vague)
- Descriptions involve vague terms
- Widely argued that poorly defined class boundaries are well modelled by fuzzy sets
- Applied widely to siol and land cover classes
- As type 1 Fuzzy Sets
- The concept of higher order uncertainty (vagueness of the vague) only now being explored

5. Solutions – Vagueness-based

Multi-channel satellite images



Can derive
Land Cover map, *m*=2.0

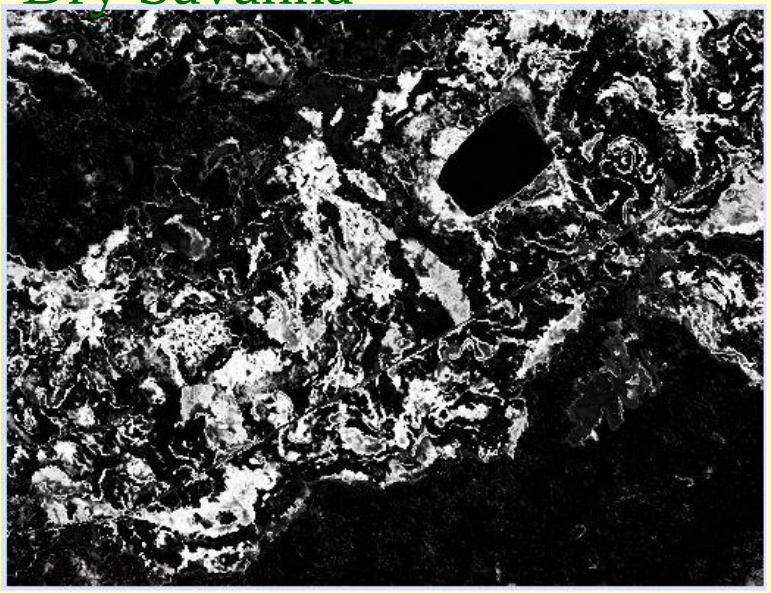


Using FCM (Bezdek et al; Lucieer)

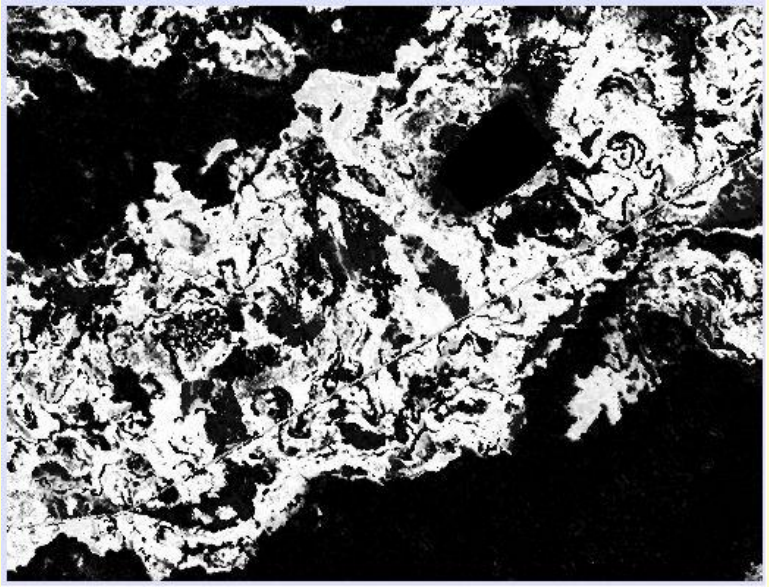
5. Solutions – Vagueness-based

Water

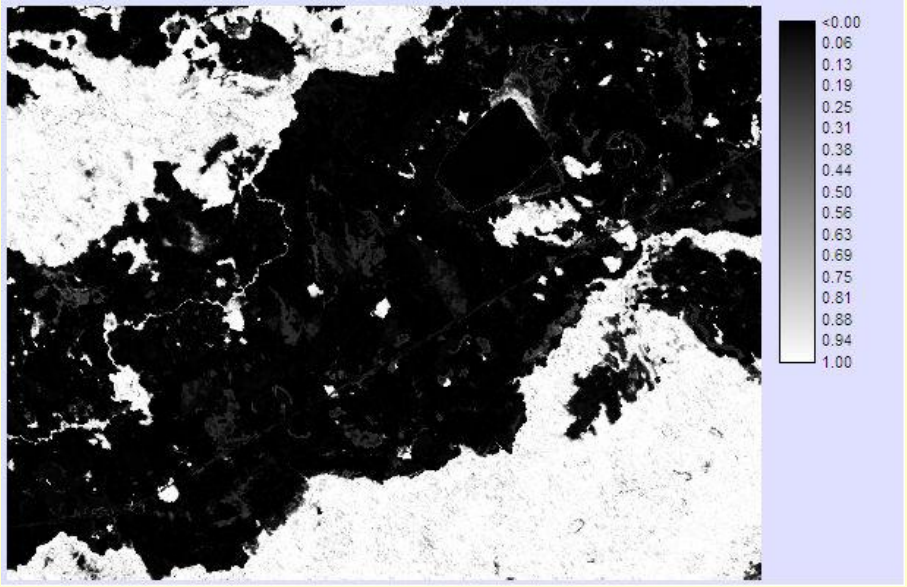
Dry Savanna



Wet Savanna

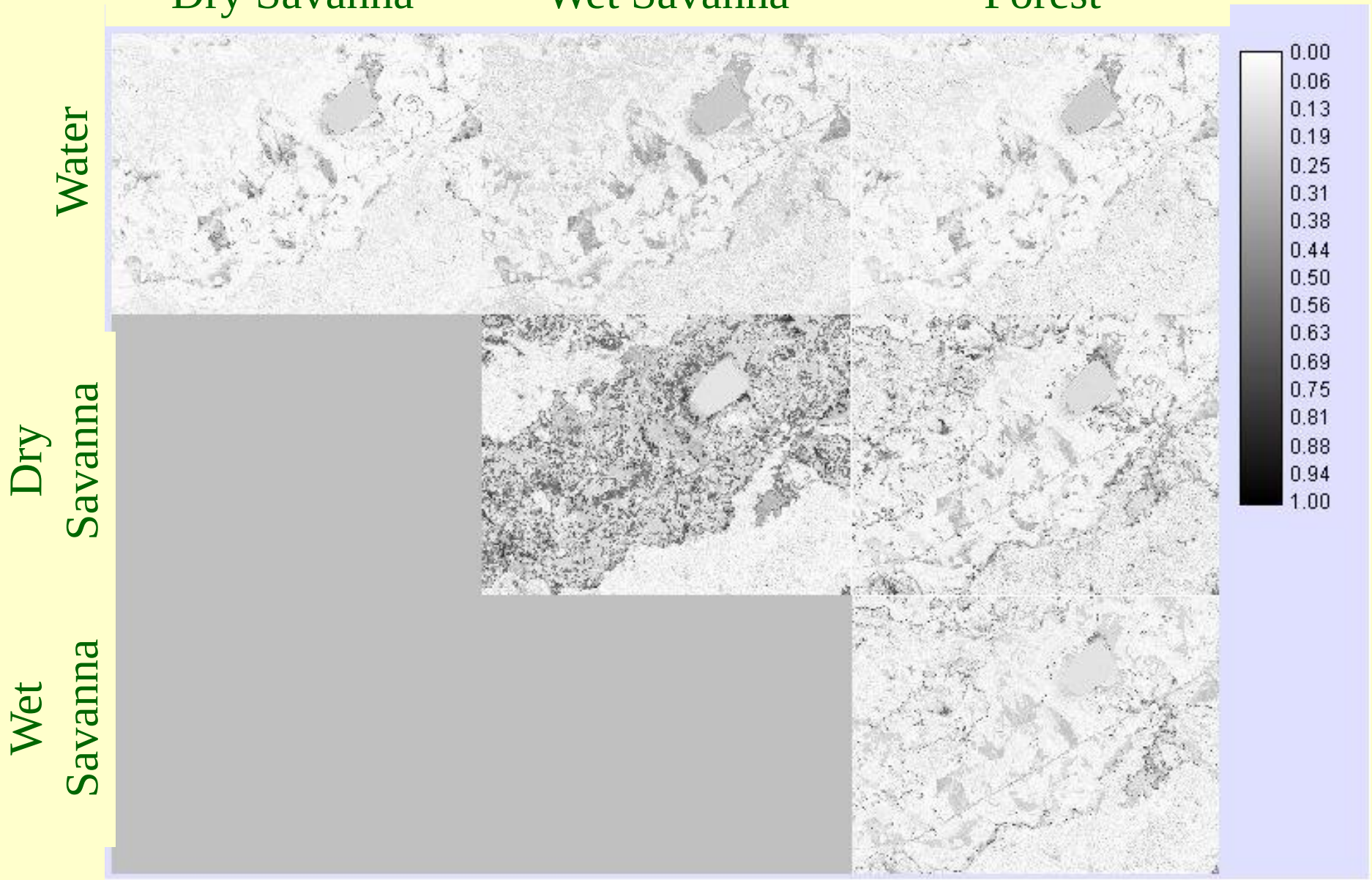


Forest



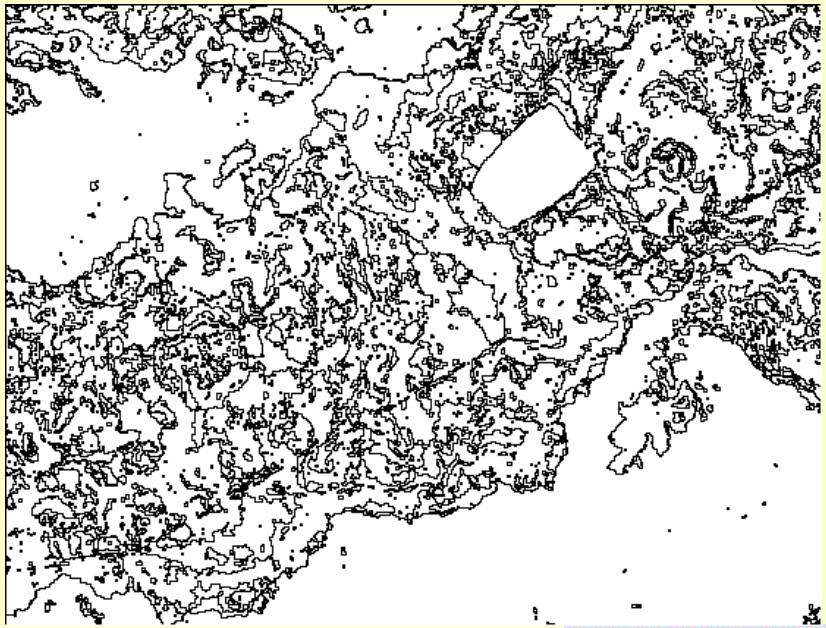
5. Solutions – Vagueness-based

Advantages of the Vague: Map the ecotones

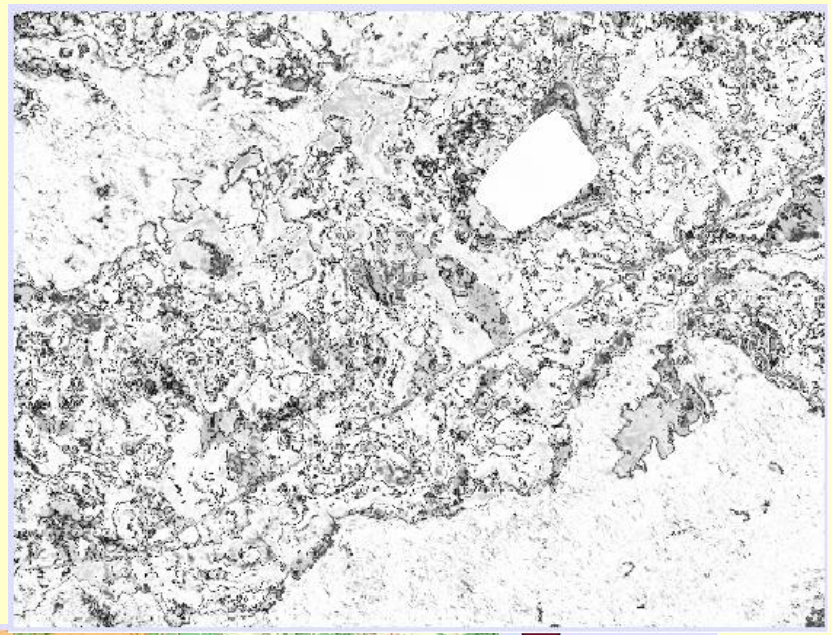


5. Solutions – Vagueness-based

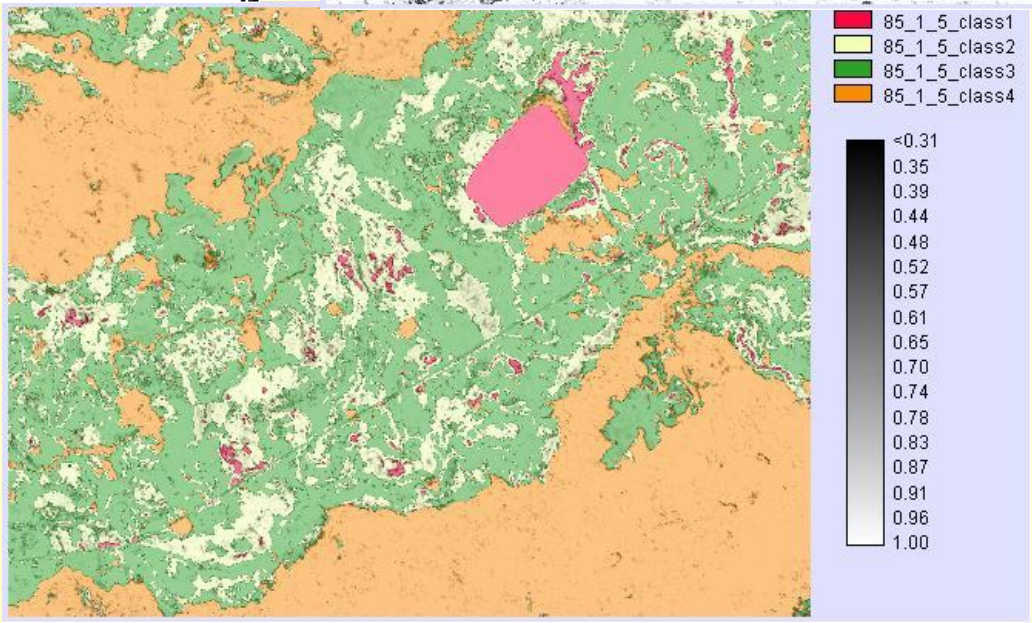
Hard boundaries



Union of Ecotones when m = 2.0



Boolean classes and boundary uncertainty



5. Solutions – Vagueness-based

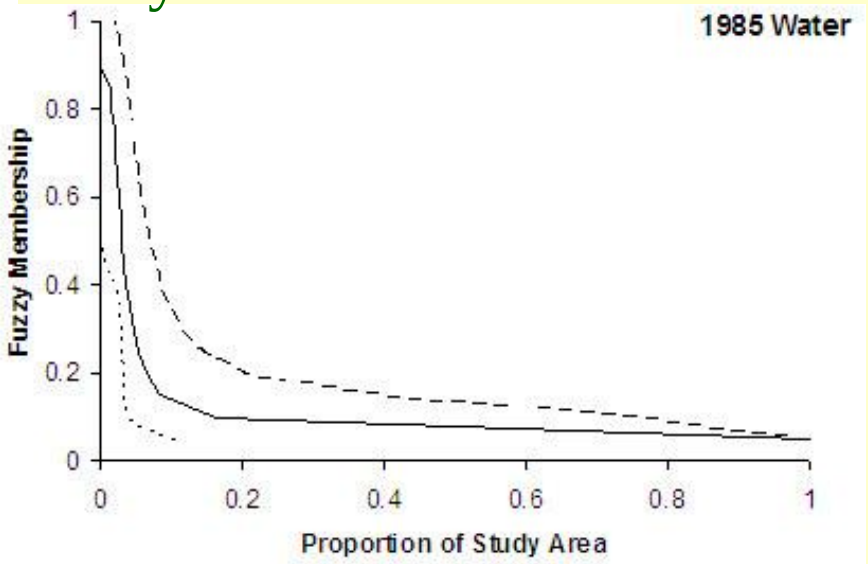
Advantages of the vague: Fuzzy Area of land covers as Type 2 FS

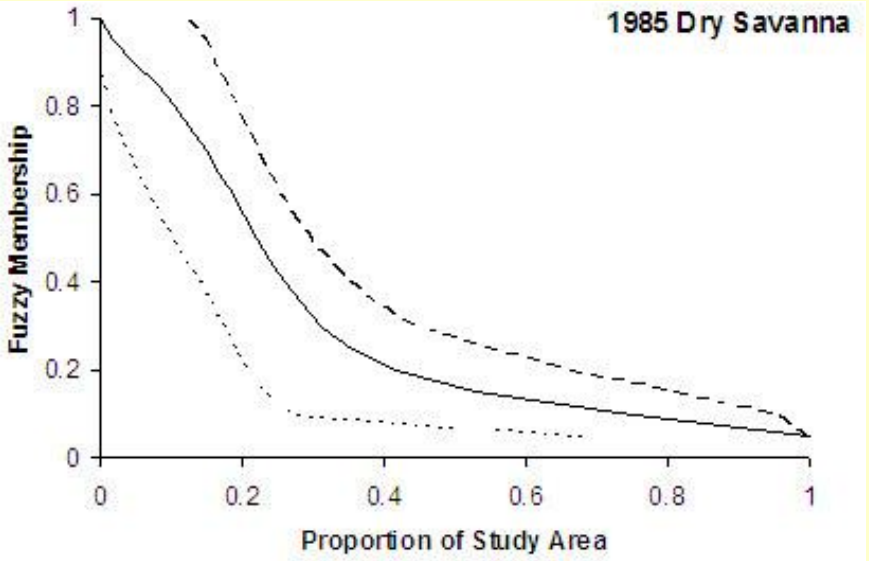
	Water	Dry savanna	Wet savanna	Forest
Max	14.95	37.40	47.59	41.11
Mean	5.69	24.93	35.82	33.57
Min	1.36	10.70	23.64	24.63

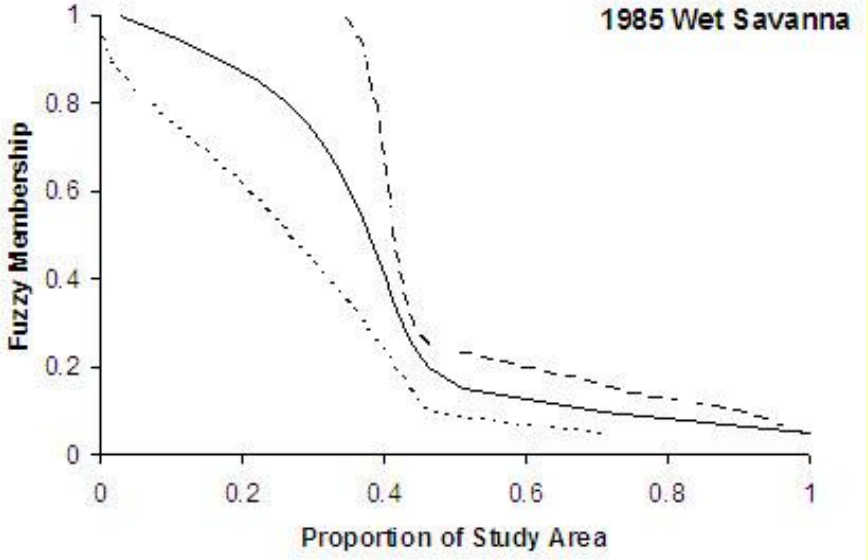
Boolean	2.91	24.50	36.67	35.92

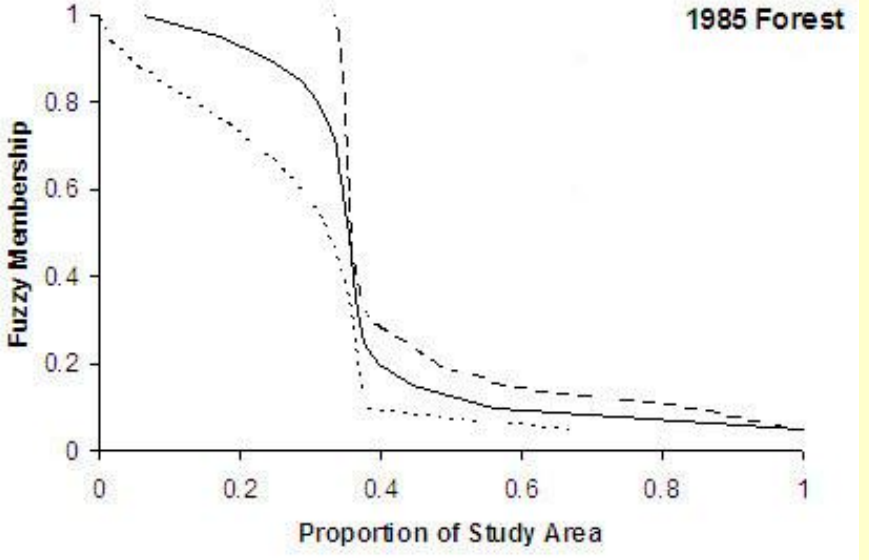
5. Solutions – Vagueness-based

Advantages of the vague: Fuzzy Area as fuzzy numbers and as type 2 fuzzy sets



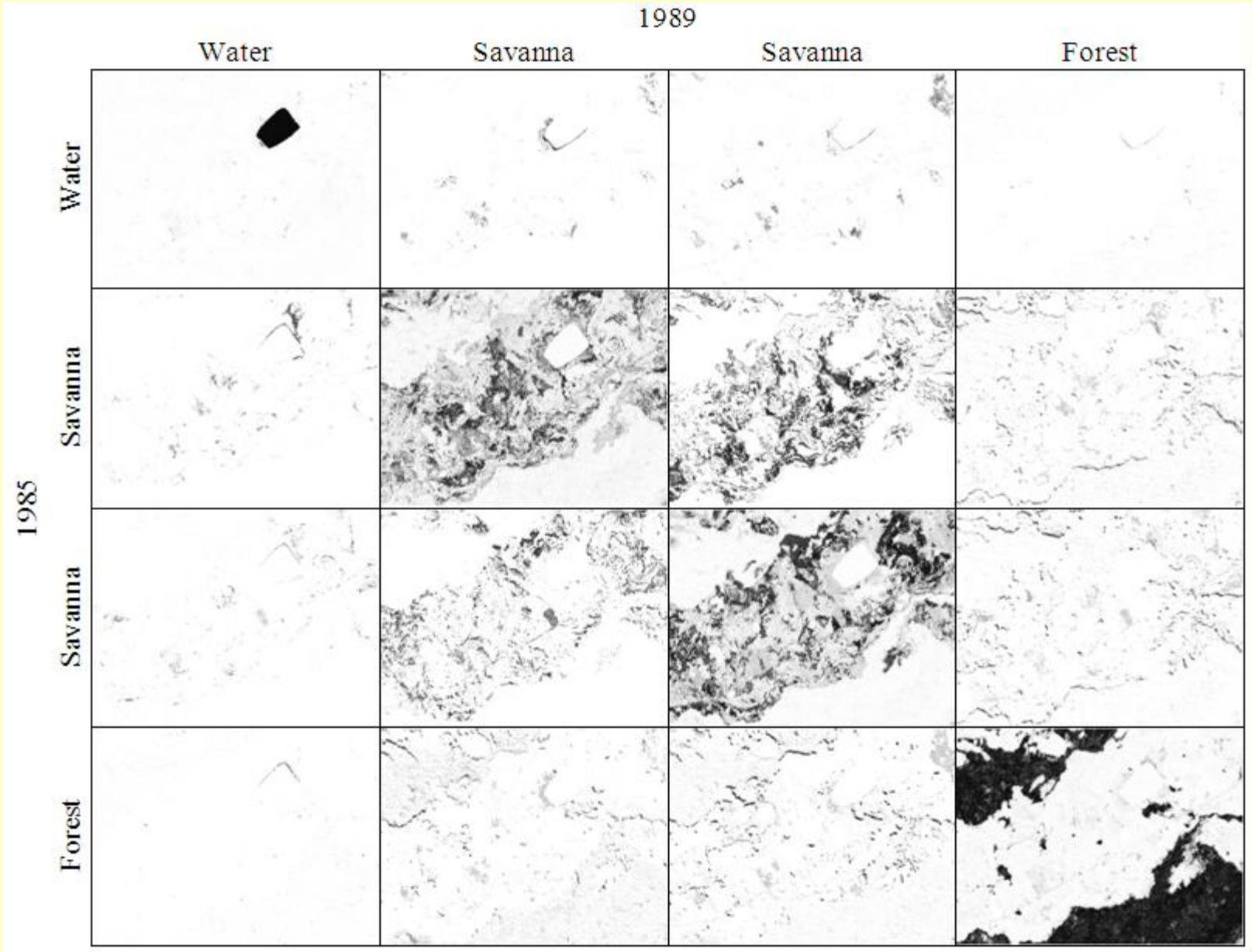






5. Solutions – Vagueness-based

Advantages of the vague: Fuzzy Change analysis – Ecotone change



5. Solutions – Vagueness-based

- Further work:

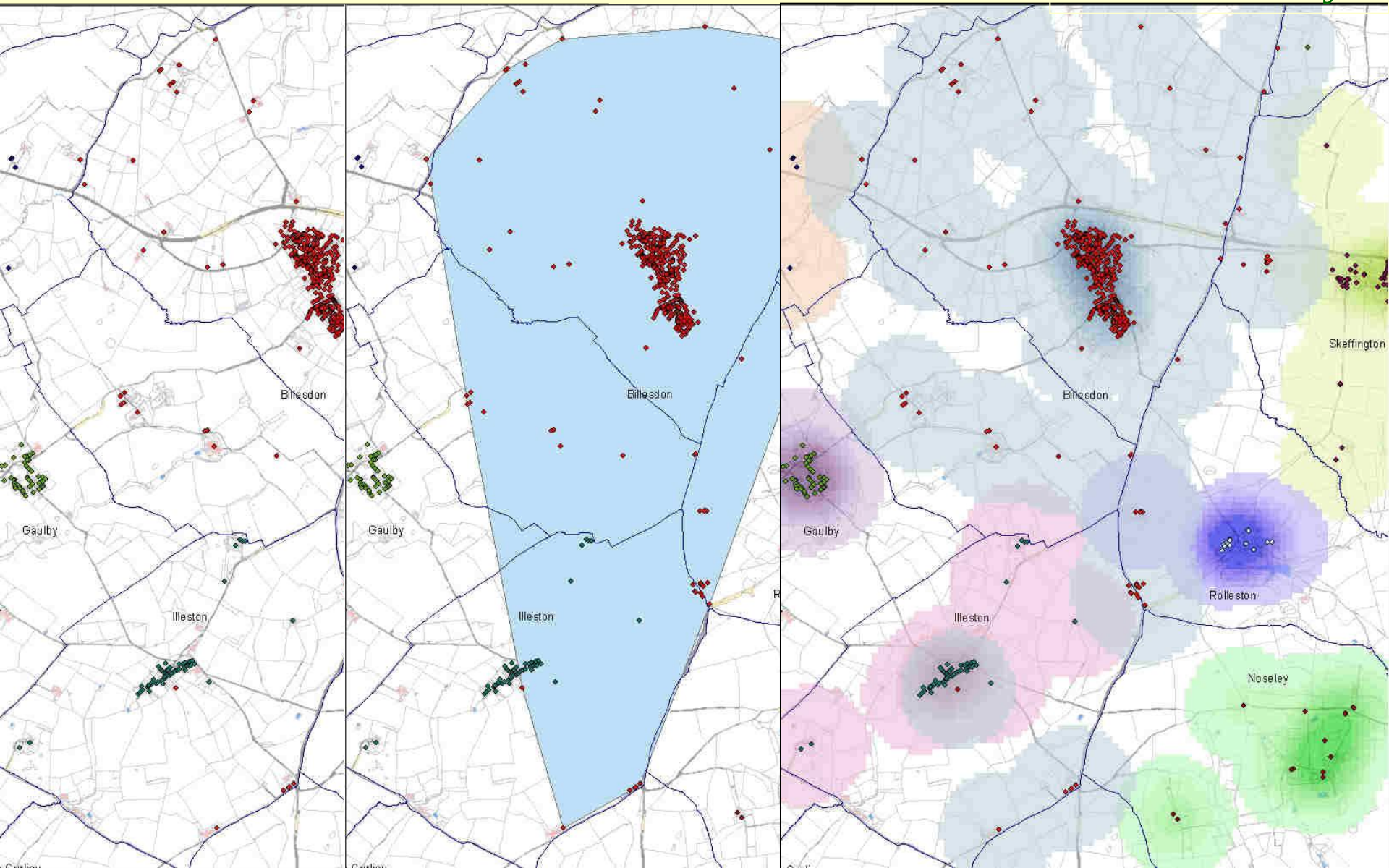
- The probablistic assumption:
  - $\sum_{l=1}^{k} \mu_{il} = 1$ for all i and j

- Possibilistic c-means may be more appropriate to use

5. Solutions – Vagueness-based

Where is Billesden ?

Original Points | Convex Hull | Kernel Density



Conclusion

- A profound uncertainty issue in GI is around the meaning of words used to describe the information

1. Less usual methods may need to be more widely adopted to address this and they need to be researched
2. That meaning needs urgently to be clarified through expanded metadata

Accuracy 2010

- International conference on Spatial Accuracy in Environmental Information
- Leicester
- July 20th-23rd
- www.accuracy2010. ???
- Just use Google!

Questions?

- Peter Fisher – pff1@le.ac.uk