12. GIS Design for Feature Extraction

Urban Functional Zone as an example

GE3238 GIS Design and Practices

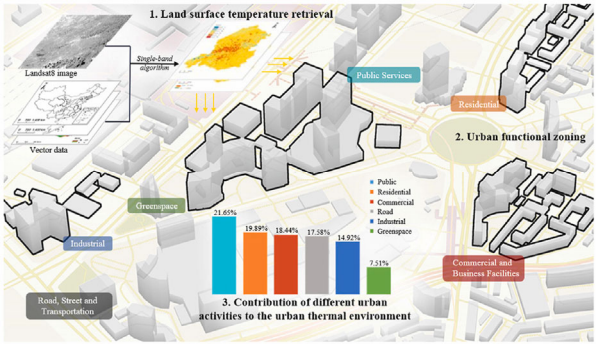
Geography@NUS

Chen-Chieh FENG

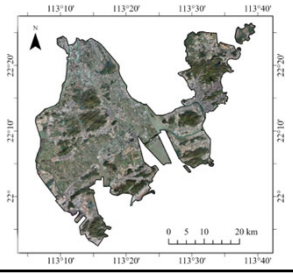
1

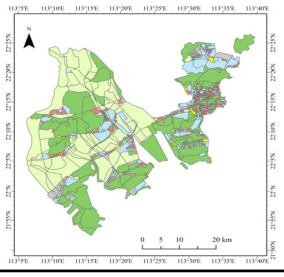
Objectives

- To look at GIS design for using geospatial technologies to retrieve urban functional zone (UFZ)
  - Technical: Classification or semantic segmentation
  - Problem set: Information retrieval and feature extraction



Source: https://ars.els-cdn.com/content/image/1-s2.0-S0360132322002426-ga1_lrg.jpg (last accessed 7 APR 2025)

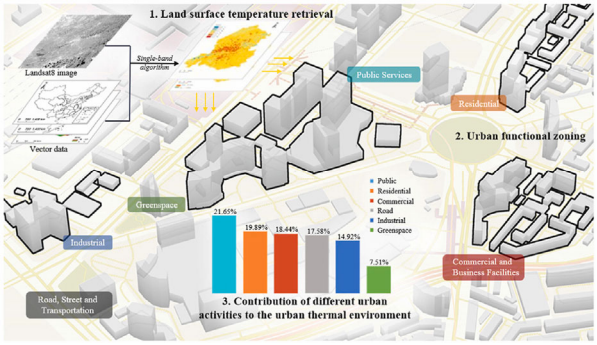




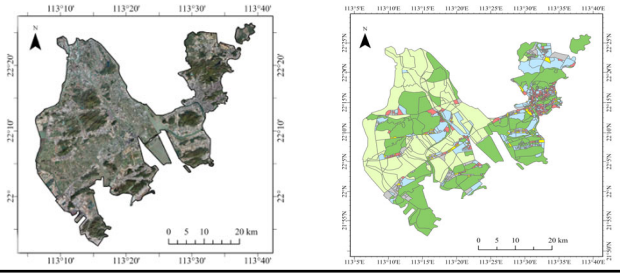
2

Objectives (cont.)

- To look at GIS design for using geospatial technologies to retrieve urban functional zone (UFZ)
  - Focus on data layers and fitness-for-use



Source: https://ars.els-cdn.com/content/image/1-s2.0-S0360132322002426-ga1_lrg.jpg (last accessed 7 APR 2025)



3

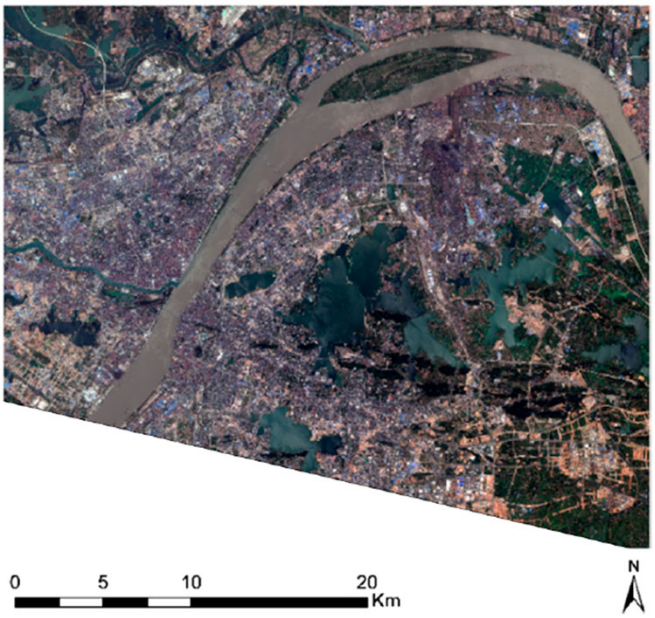
Land use/land cover (LULC) vs UFZ

Q1 -- What is/are the difference(s)?

(Why do we care about UFZ?)

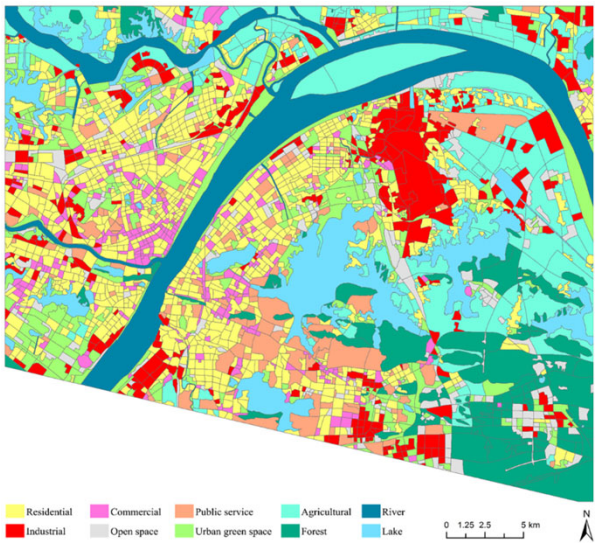
4

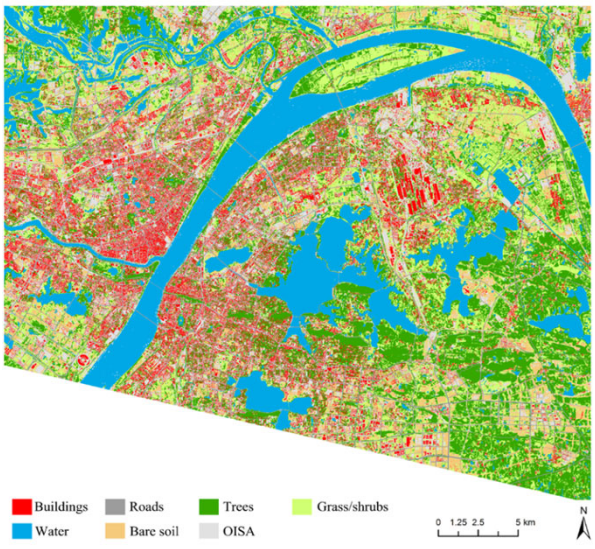
- Remote sensing imagery
  - Commonly taken to be data input for further (geo-) spatial analyses
    - Data wrangling (see slide 8 for an example)
  - Information retrieval or feature extraction
    - UFZ is the information to be retrieved



5

UFZ & Land Use/Land Cover (LULC) (cont.)

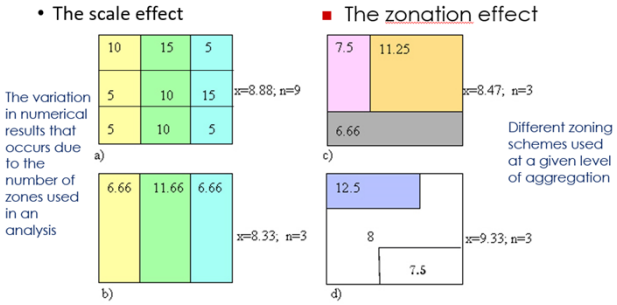




6

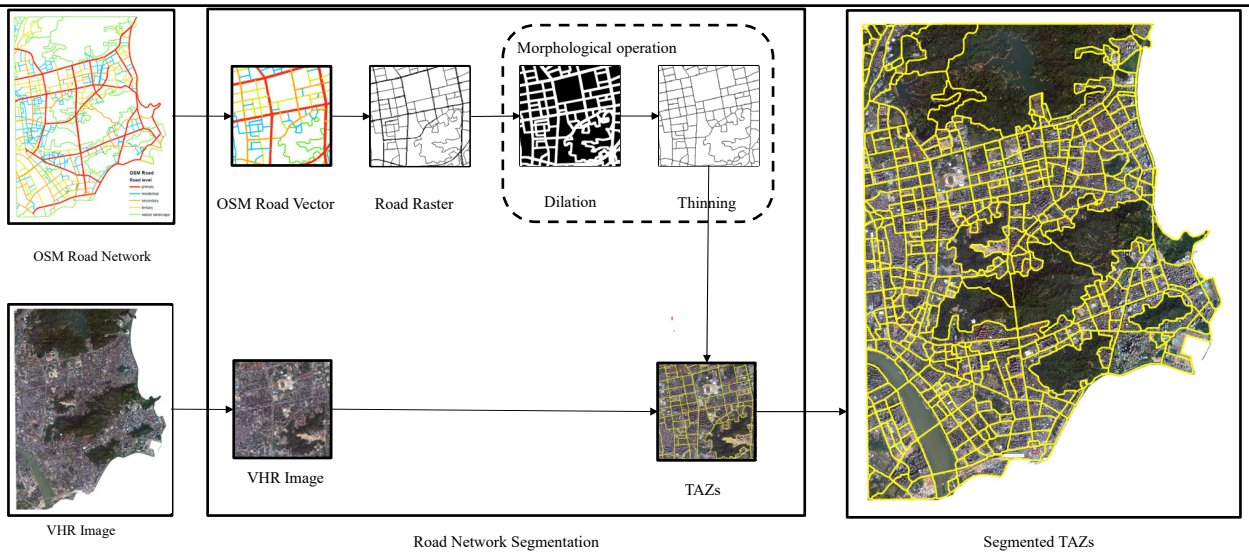
What Kind of Spatial Unit?

- Administration, e.g., planning area, subzone?
- Block demarcated by roads?
- Any other region types?



The Effects of MAUP (Modifiable Areal Unit Problem)

7



An example of spatial data wrangling for information retrieval

VHR: Very-high resolution; TAZ: Transportation analysis zone; OSM: OpenStreetMap

8

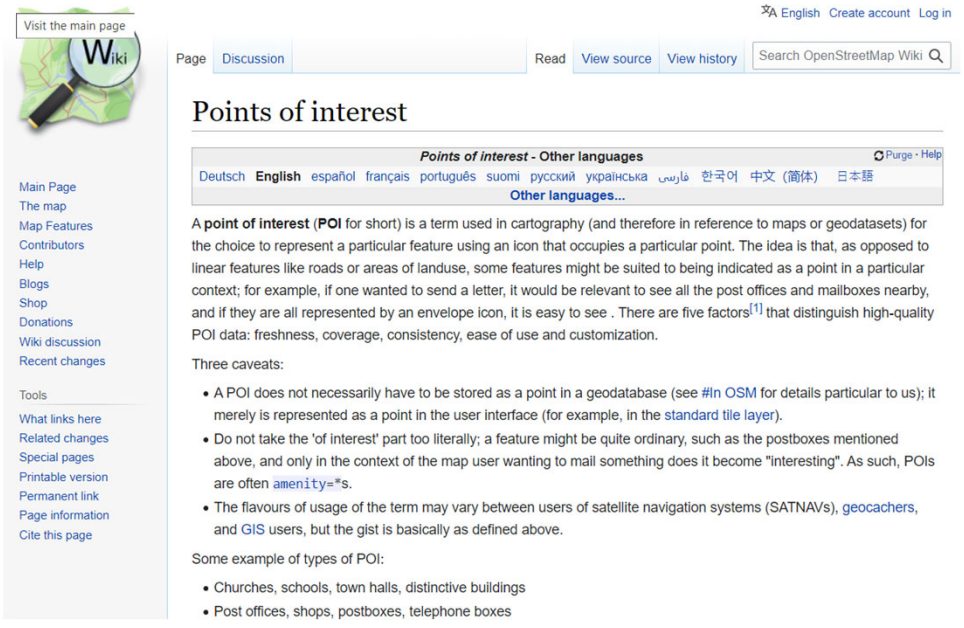
How about Capturing Human Activities?

Q2 -- What is the readily available dataset that captures human activities?

(What layers + attributes to include in your design)

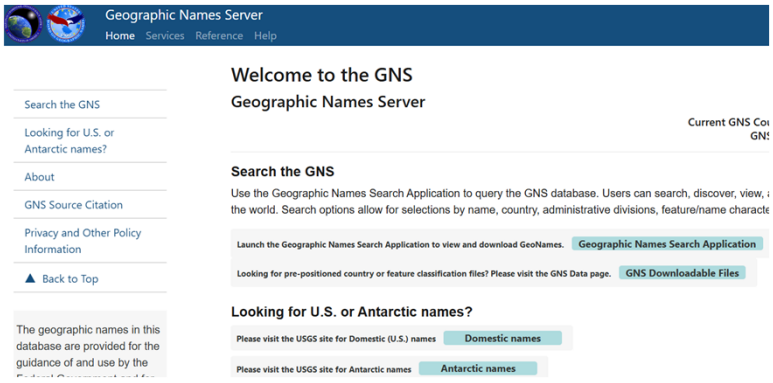
(What about their fitness-for-use?)

9



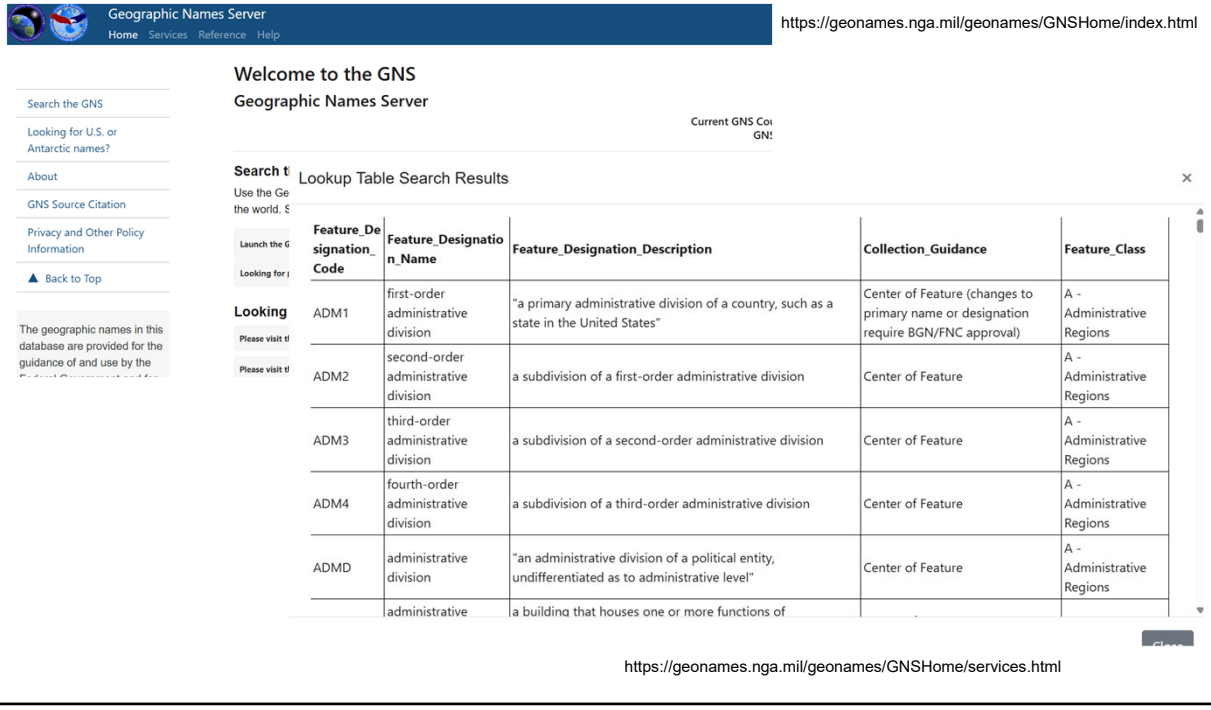
https://wiki.openstreetmap.org/wiki/Points_of_interest

10



https://geonames.nga.mil/geonames/GNSHome/index.html

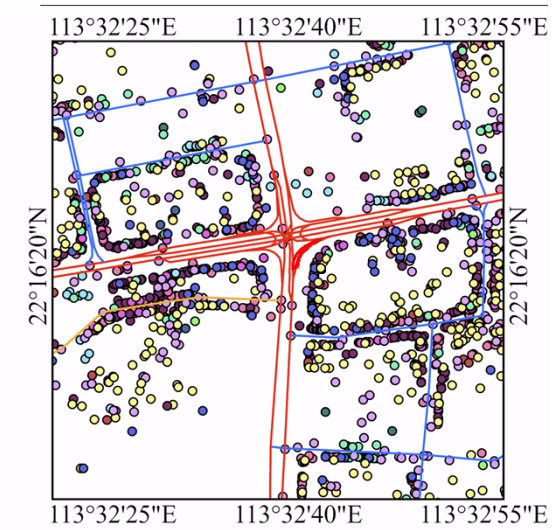
11



https://geonames.nga.mil/geonames/GNSHome/index.html

https://geonames.nga.mil/geonames/GNSHome/services.html

12



Any concerns with using POIs?

POIs (14 classes)

- Company
- Factory
- Food and beverage
- Public organization
- Health facility
- Hospital
- Living service
- Public service
- Recreational service
- Residential building
- School
- Science and education
- Shop
- Transportation

Fig 2 of Wang et al (2023) -- https://www.mdpi.com/2072-4292/15/3/730

13

Describing Human Activities (again)?

Q2 -- What is the readily available dataset that describe human activities?

(Recall the consideration on *what layers to include*?)

(What about their *fitness-for-use*?)

14

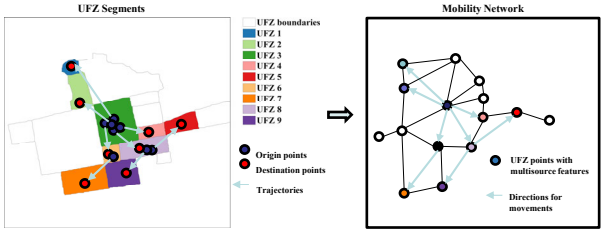
Mobility Data

- Common types
  - Taxi data
  - Tap-in and tap-out data
  - GPS tracks
  - Mobile phone data
  - Bike sharing data
  - …
- Fitness-for-use?

15

Mobility Data

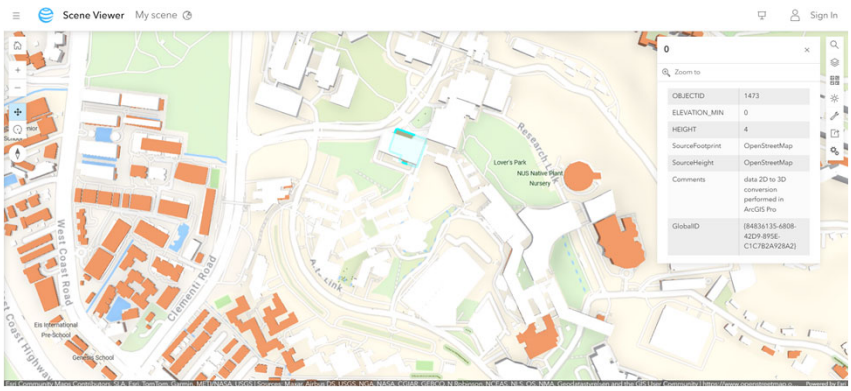
- Why do we consider mobility data for UFZ classification?
  - "Trajectories … are one of the indicators for human activities associated with socioeconomic characteristics of UFZs"
  - For example, consider where most people go in the morning and evening



16

Height Information

- Building height



https://www.arcgis.com/home/item.html?id=b5a6788bf3a141d489f36487162252af (last accessed 7 APR 2025)

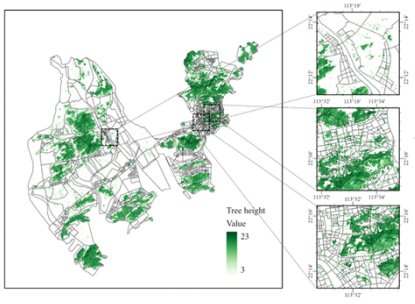
17

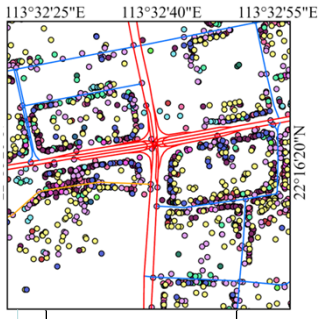
Height Information

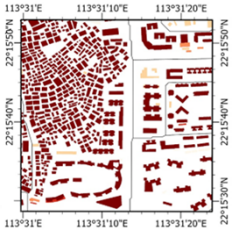
- Building height
- Canopy height

18

Input Datasets





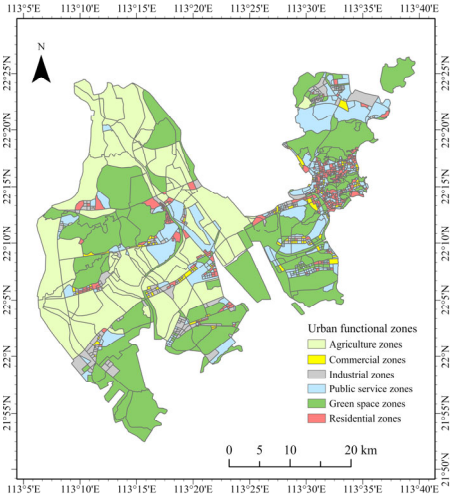


OSM data
HSR imagery
POI data
Forest canopy height data
Building data
Taxi GPS data

17,995 trajectories

19

UFZ Classification Result

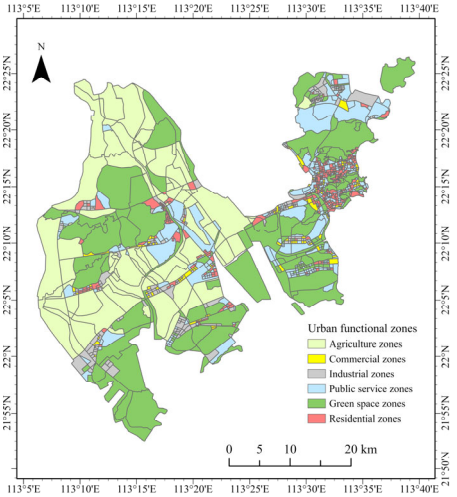


	A	C	I	P	G	R	Prod Acc (%)
A	25	0	0	1	1	1	89.29
C	0	29	1	9	0	6	64.44
I	0	0	53	7	2	2	82.81
P	0	4	5	57	7	17	63.33
G	1	0	3	10	45	0	76.27
R	0	2	1	11	1	79	84.04
User Acc (%)	96.15	82.86	84.13	60.00	80.36	75.24	OA = 75.79

A: Agriculture zone; C: Commercial zone; I: Industrial zone; P: Public service zone; G: Green space zone; R:Residential zone; and OA: Overall accuracy.

20

UFZ Classification Result

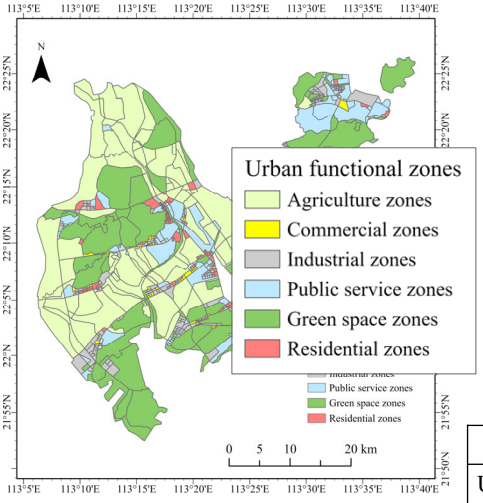


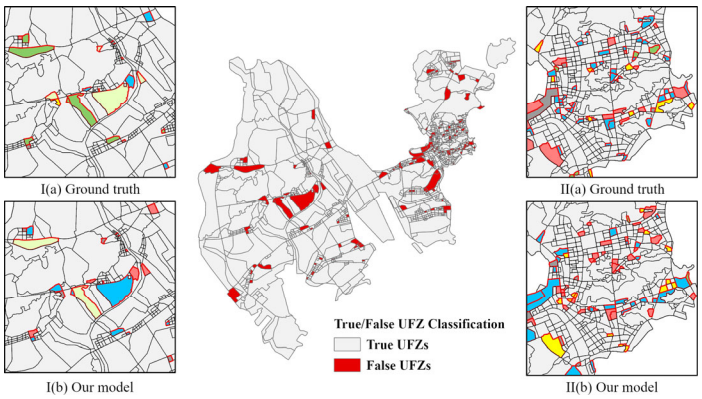
	A	C	I	P	G	R	Prod Acc (%)
A	25	0	0	1	1	1	89.29
C	0	29	1	9	0	6	64.44
I	0	0	53	7	2	2	82.81
P	0	4	5	57	7	17	63.33
G	1	0	3	10	45	0	76.27
R	0	2	1	11	1	79	84.04
User Acc (%)	96.15	82.86	84.13	60.00	80.36	75.24	OA = 75.79

A: Agriculture zone; C: Commercial zone; I: Industrial zone; P: Public service zone; G: Green space zone; R:Residential zone; and OA: Overall accuracy.

21

UFZ Classification Result





	A	C	I	P	G	R	
User Acc (%)	96.15	82.86	84.13	60.00	80.36	75.24	OA = 75.79

22

What Does the Evaluation Tell You?

- The accuracy will never reach 99.9%!
- Do not look only at overall accuracy; Look also at individual categories
  - It is expected the spatial distribution of misclassification differs by category
- Is there additional data layers to be considered? What is the rationale?
  - Alternatively, do any of the data layers introduce noises?

23

Summary

- UFZ classification as an example of information retrieval through classification (semantic segmentation)
  - Define the scope: Extraction of urban functional zone
  - Key concept: Human activities
  - Spatial unit: Transportation analysis zone (Why not pixel-based?)
  - Data wrangling: Dilution and thinning
  - Classification algorithm: out-of-scope for this lecture
  - Evaluation: Error matrix as a start; accuracy by category
  - Interpretation: Focus on implications

24