Theoretical methodological frame and EO-derived information in the GHSL

Martino Pesaresi, European Commission's Joint Research Centre, Italy

The 13th ISDE International Lectures : "Advances on the Global Human Settlement Layer by joint assessment of Earth Observation and Population Survey data"

11th November 2024 (Central European Time)

European Commission

contents

- step 0 : fundamentals
- step 1 : scope and use context
- step 2 : definitions
- step 3 : data conceptual organization
- step 4 : general processing strategy
- notes on novelty, core methodological toolbox
- conclusions

step 0 : GHSL fundamentals

- The ‘Human Settlement’ (HS) abstraction adopted by GHSL is composed by two essential components: buildings and their inhabitants
- HS @ spatial quantitative modeling and data perspective

regional science

- Isard, Walter. 1949. “The General Theory of Location and Space-Economy.” The Quarterly Journal of Economics 63 (4): 476–506.

geography

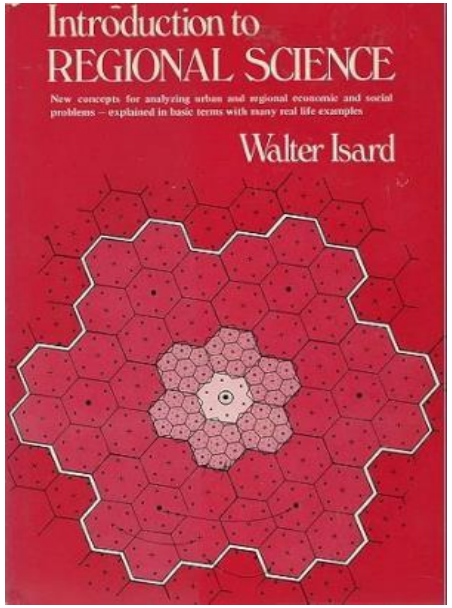
- Stone, Kirk H. 1965. “The Development of a Focus for the Geography of Settlement.” Economic Geography 41 (4): 346. https://doi.org/10.2307/141945.

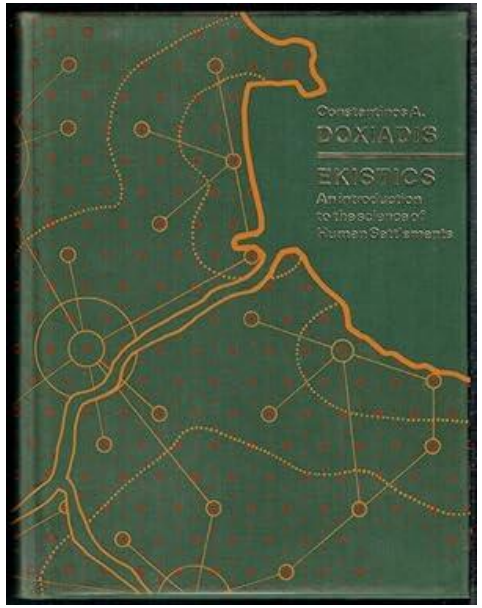
urban analysis
urban planning

- Doxiadis, Constantinos A. 1970. “Ekistics, the Science of Human Settlements: Ekistics Starts with the Premise That Human Settlements Are Susceptible of Systematic Investigation.” Science 170 (3956): 393–404.

urban science

- Bettencourt, Luís M. A. 2021. Introduction to Urban Science: Evidence and Theory of Cities as Complex Systems. Cambridge, MA and London, England: The MIT Press.





- W. Isard, (1949) cost distance, people (labor) raw materials, market
- C. Doxiadis (1970) human interactions nature, anthropos, society, shells, and networks.

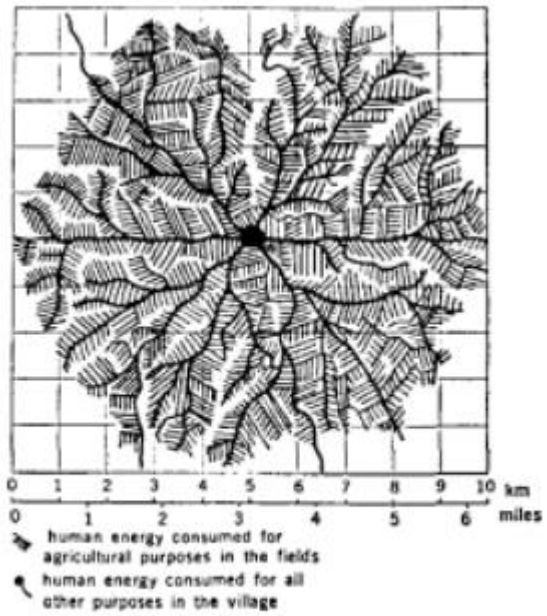
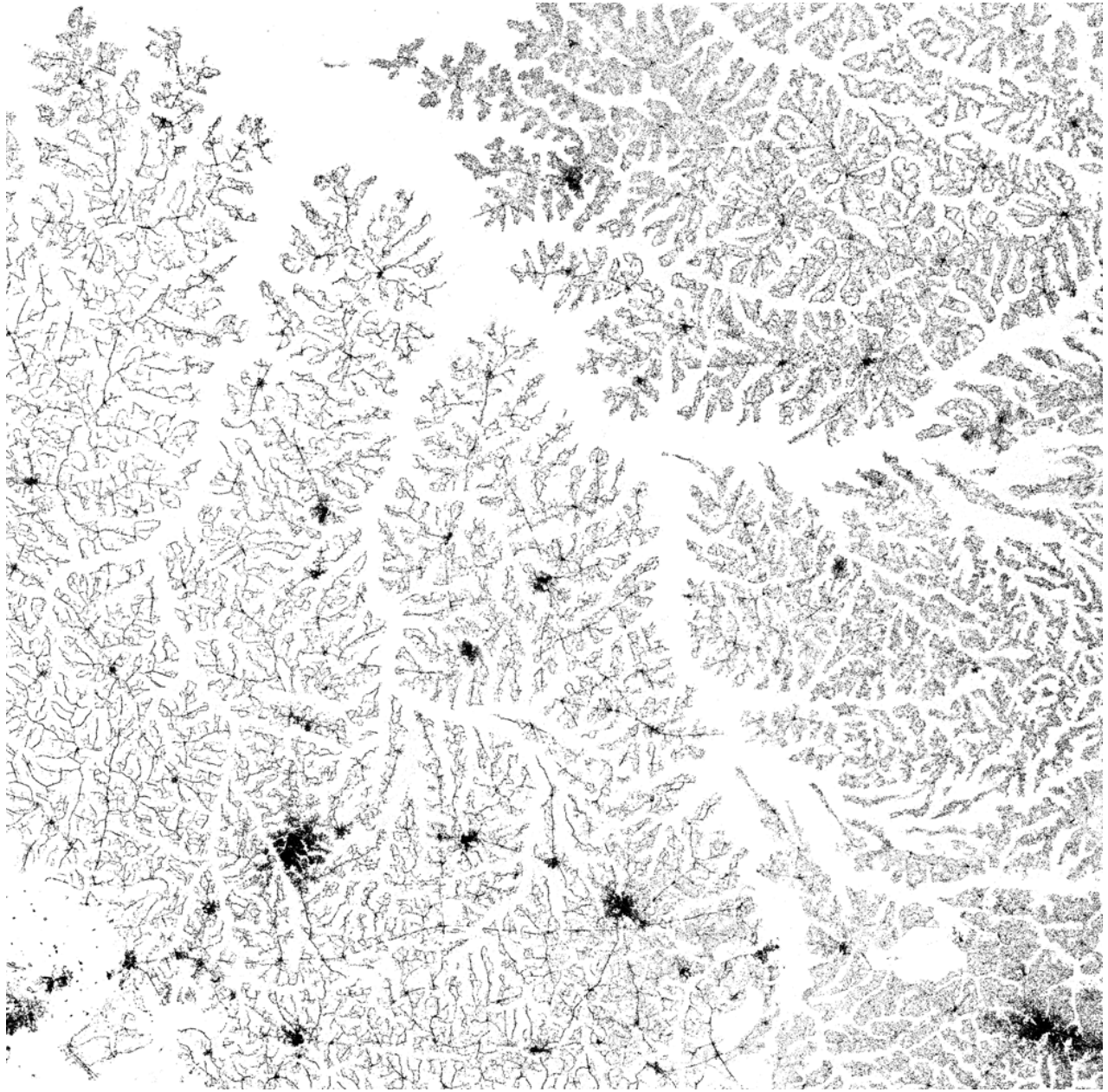


Fig. 4. Energy model of a village. Daily per capita energy consumption, 8000 calories.



GHSL (2023) data evidences in Uganda

European Commission

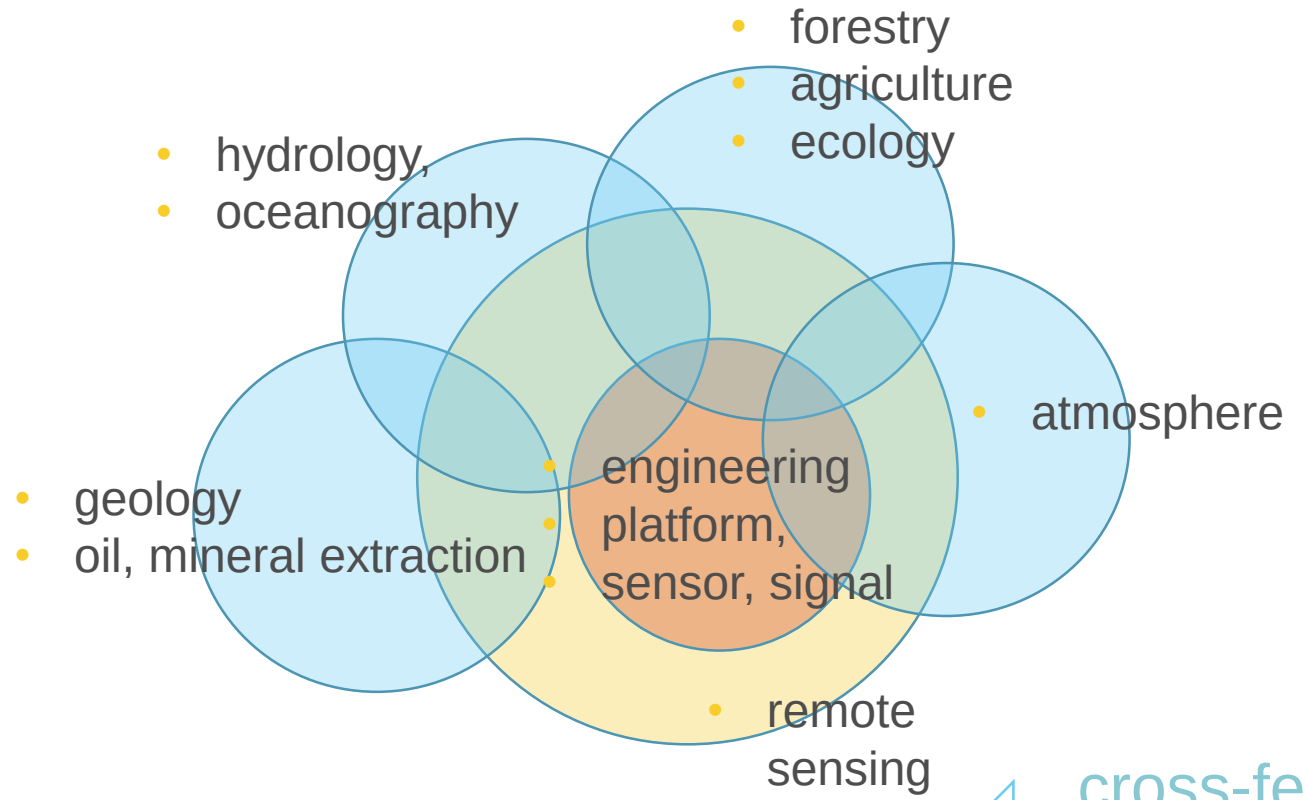
genesis of the GHSL : g-‘H’-s-l

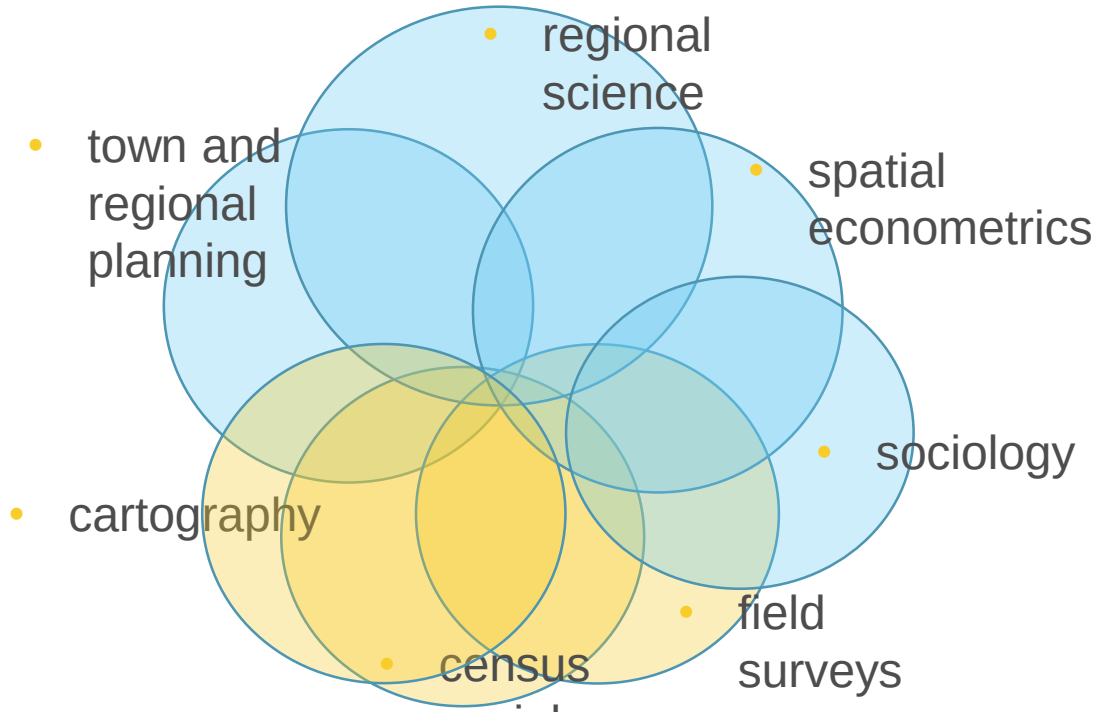
- earth sciences

→ *focus on monitoring*

- social sciences,
- econometrics,
- analysis of public policies

→ *focus on understanding the logic of the human action*





cross-fertilization

step 1 : scope and use context

- fine-scale, globally-consistent measurements that focus on human settlements, essential for international comparability and to support public decision-making
  - Pesaresi, Martino. 2018. “Principles and Applications of the Global Human Settlement Layer.” *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 2047–2050. Valencia: IEEE. https://doi.org/10.1109/IGARSS.2018.8519155.
- evidence-based policy action
  - postulation: ‘public interest’, ‘public good’
- simple and easily interpretable assumptions, thereby discouraging the use of over-parametrized models
  - postulation: models are tools for human learning facilitating human discussion, not imposing a given machine-driven solution
- full repeatability of the measures
  - National Institute for Standards and Technology. 2007. “Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results.” http://physics.nist.gov/Pubs/guidelines/contents.html.

open data and methods ecosystem

- free and open access to data, methods, and results as proposed by the FAIR Data Principles
  - Wilkinson, Mark D., Michel Dumontier, Ijsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. “The FAIR Guiding Principles for Scientific Data Management and Stewardship.” *Scientific Data* 3 (1): 160018. https://doi.org/10.1038/sdata.2016.18
- Group on Earth Observations (GEO) ‘*full and open*’ Data Sharing Principle
- Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information PE/28/2019/REV/1.

zoom on repeatability

- ‘Repeatability of measurements’ here refers to the *ability to repeat measurements made on the same subject under identical conditions*, and shall be distinguished from the ‘reproducibility of the experiment’ concept common in scientific literature.
- In case of measurements supported by a supervised classification:
  - repeatability implies that the same input data (images, training set) and same inferential system should produce always, e.g., the *exact same* built-up surface estimates. By construction, this propriety is not fully satisfied by classification methods requiring stochastic iterative processes for convergence to a given inferential solution.

step 2 : definitions (semantic transparency)

Table 1. Definitions used in the manuscript.

Term	Definition
Building	'Any roofed structure erected above ground for any human use, including structures in slums, informal settlements, and refugees/IDP camps' (Pesaresi et al. 2013) (European Commission. Joint Research Centre. 2023; Pesaresi et al. 2015a).[19]
Building height	'The distance measured from the mean ground level to the mean height of the roof' (Pesaresi et al. 2021).[20]
Built-up surface	'The gross building surface (including the thickness of the walls) bounded by the building wall perimeter' (European Commission. Joint Research Centre. 2023).
Built-up volume	'The product of the gross built-up surface by the building height'.
Residential area	'Area dedicated prevalently for residential use, including mixed with other non-conflicting uses' (European Commission. Joint Research Centre. 2023).[21]
Non-residential area	'Area dedicated exclusively to non-residential use', as typically large industrial or commercial/ retail units, confined animal feeding, indoor farming, hangars, data centers, transportation (ports/airports facilities), large gov. administration or offices units, large sport / leisure indoor facilities, exclusive religious sites and similar (European Commission. Joint Research Centre. 2023).
Place of residence	'The place where a person normally spends the daily period of rest, regardless of temporary absences for purposes of recreation, holidays, visits to friends and relatives, business, medical treatment or religious pilgrimage'[22,23,24] (Freire et al. 2016).
Urban, Suburban/Peri-urban, Rural area	'A specific combination of resident population density or built fabric density and settlement population size observed at a definite spatial generalization scale', according to the degree of urbanization (DEGURBA) methodology (Dijkstra et al. 2021).

step 3 : data conceptual organization

GHSL information is *continuous quantitative*, centered on the presence of buildings (built-up surface and building volume per spatial unit) and their inhabitants (number of residents per spatial unit).

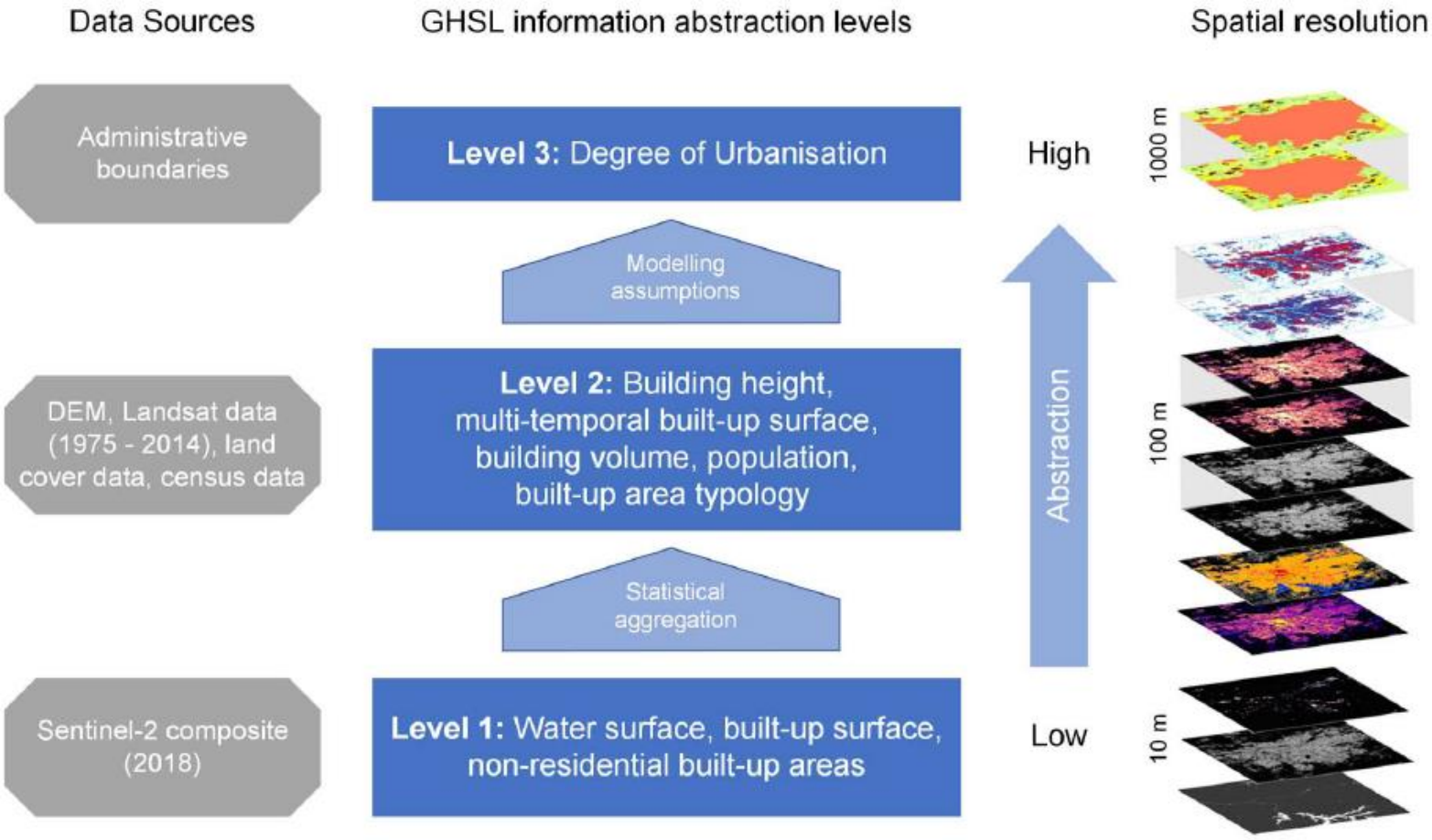


Figure 4. The GHSL hierarchical multiple-abstraction meta-model (HAMM) for spatial information production.

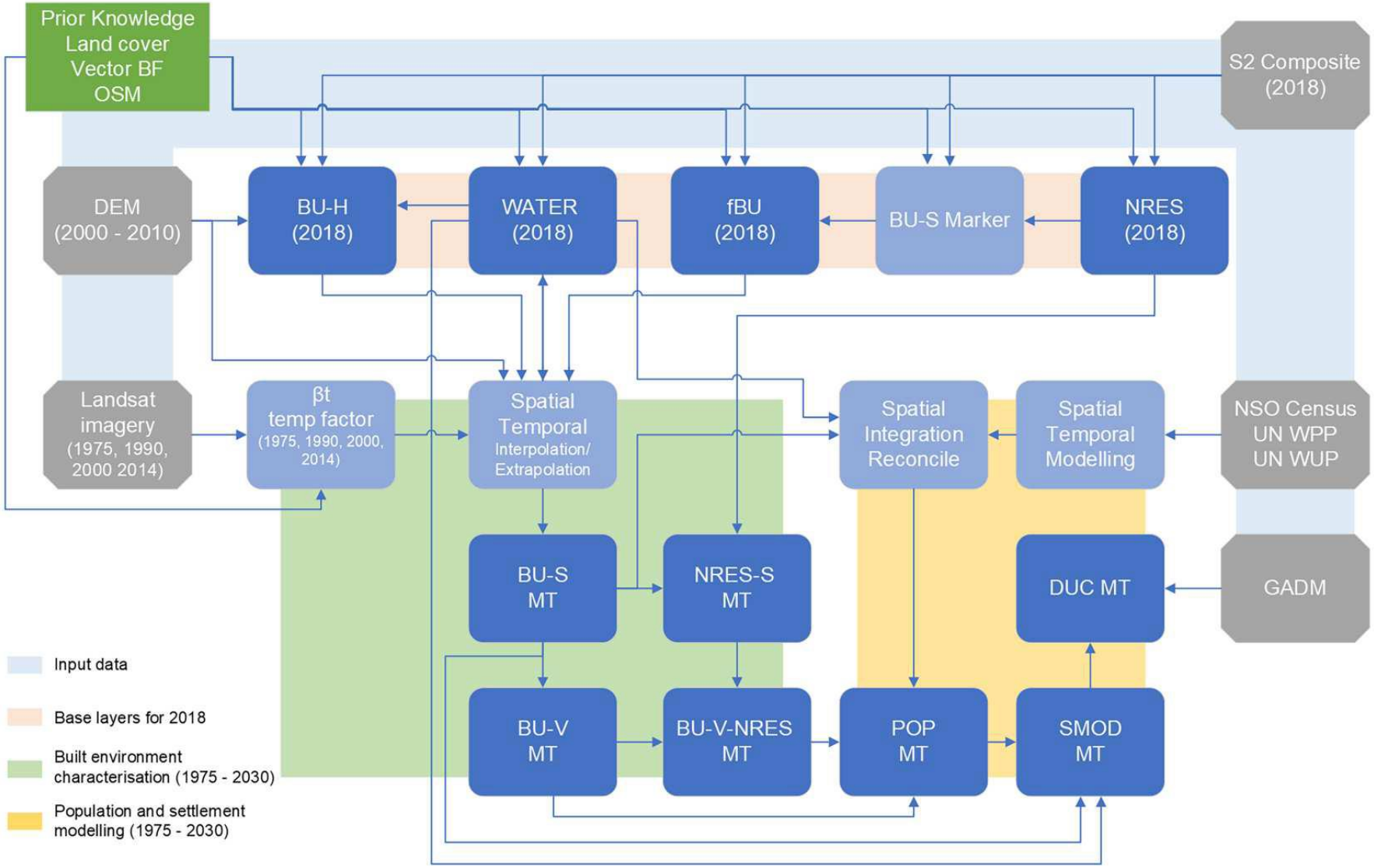
GHSL as ecosystem of information

Dataset name	Variable	Spatial resolution			Year	Source data
		10 m	100 m	1000 m		
GHS-BUILT-S	Total built-up surface [sqm]	•			2018	Sentinel-2
GHS-BUILT-S	Total built-up surface [sqm]		•	•	1975-2030	Sentinel-2, Landsat, other
GHS-BUILT-H	Average building height [m]		•		2018	Sentinel-2, ALOS-3D, SRTM, other
GHS-BUILT-V	Total building volume [m3]		•	•	1975-2030	GHS-BUILT-S, GHS-BUILT-H
GHS-BUILT-S-NRES	Non-residential built-up surface [sqm]	•	•	•	1975-2030, 2018	Sentinel-2, OSM, GHS-BUILT-S
GHS-BUILT-V-NRES	Non-residential building volume [m3]		•	•	1975-2030	GHS-BUILT-S-NRES, GHS-BUILT-H
GHS-POP	Resident population		•	•	1975-2030	CIESIN, GHS-BUILT-V-RES
GHS-LAND	Land surface [sqm]	•	•	•	2018	Sentinel-2
GHS-SMOD	Degree of Urbanisation settlement model			•	1975-2030	GHS-BUILT-S, GHS-POP
GHS-MSZ	Morphological Settlement Zone	•			2018	GHS-BUILT-S, GHS-BUILT-H, OSM, …
GHS-BUTYPE	Built-up typology		•		2018	GHS-BUILT-S, GHS-BUILT-H, OSM, …

Continuous estimates
Categorical data
Multitemporal 1975-2030

All gridded data available in World Mollweide Equal Area projection.
Selected layers in WGS84 (3-30’’), Lambert Azimuthal Equal Area Europe & Arctic (in prep.).

step 4 : general processing strategy R2023



European Commission

novelty, primacy list of the GHSL

- first global multi-temporal assessment of BU surface at decametric-res
- first MT multi-platform, integrating Sentinel-2 and historical Landsat data
- first global assessment of building height from 10m-res Sentinel-2 data
- first global assessment of non-residential built-up surfaces
- first global, 10m-resolution, sub-pixel built-up surface assessment
- first global, 10m-resolution, sub-pixel water surface assessment
- first global assessment of the building typology
- first global, internationally accepted, Urban-Rural classification

core toolbox to get the work done

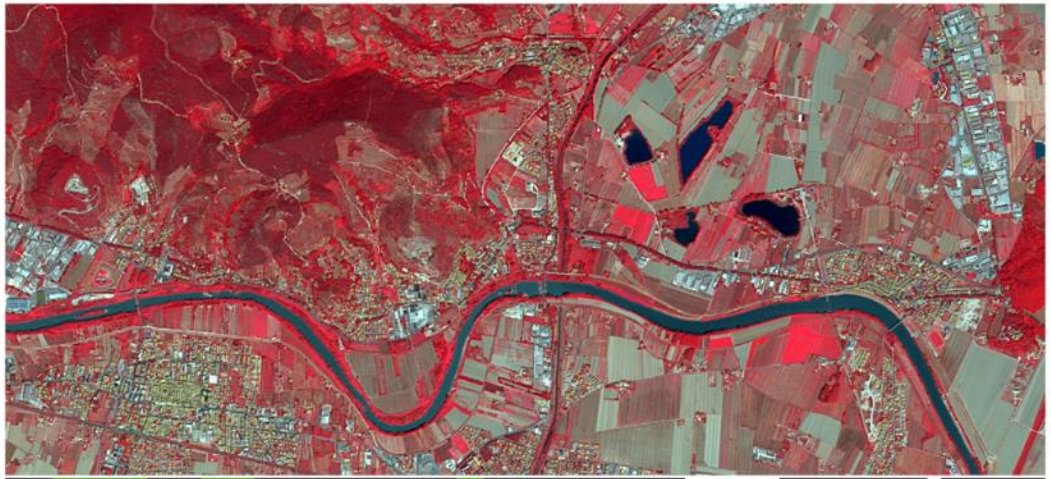
- multi-scale morphological and textural filtering,
- symbolic machine learning (SML),
- univariate, multiple-objective linear regression,
- univariate linear dasymetric mapping,
- ensemble modelling

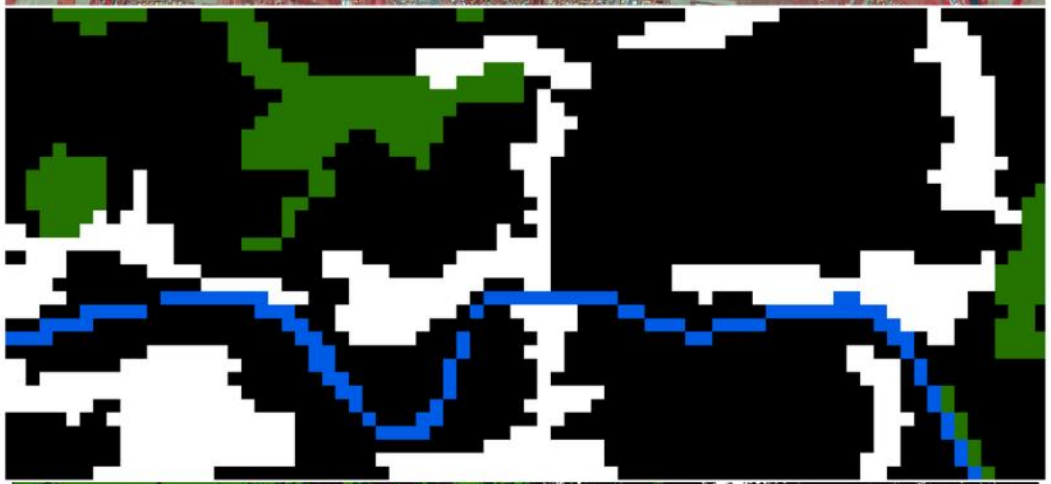
Random iterative search of inferential solution in over-parametrized discriminant function

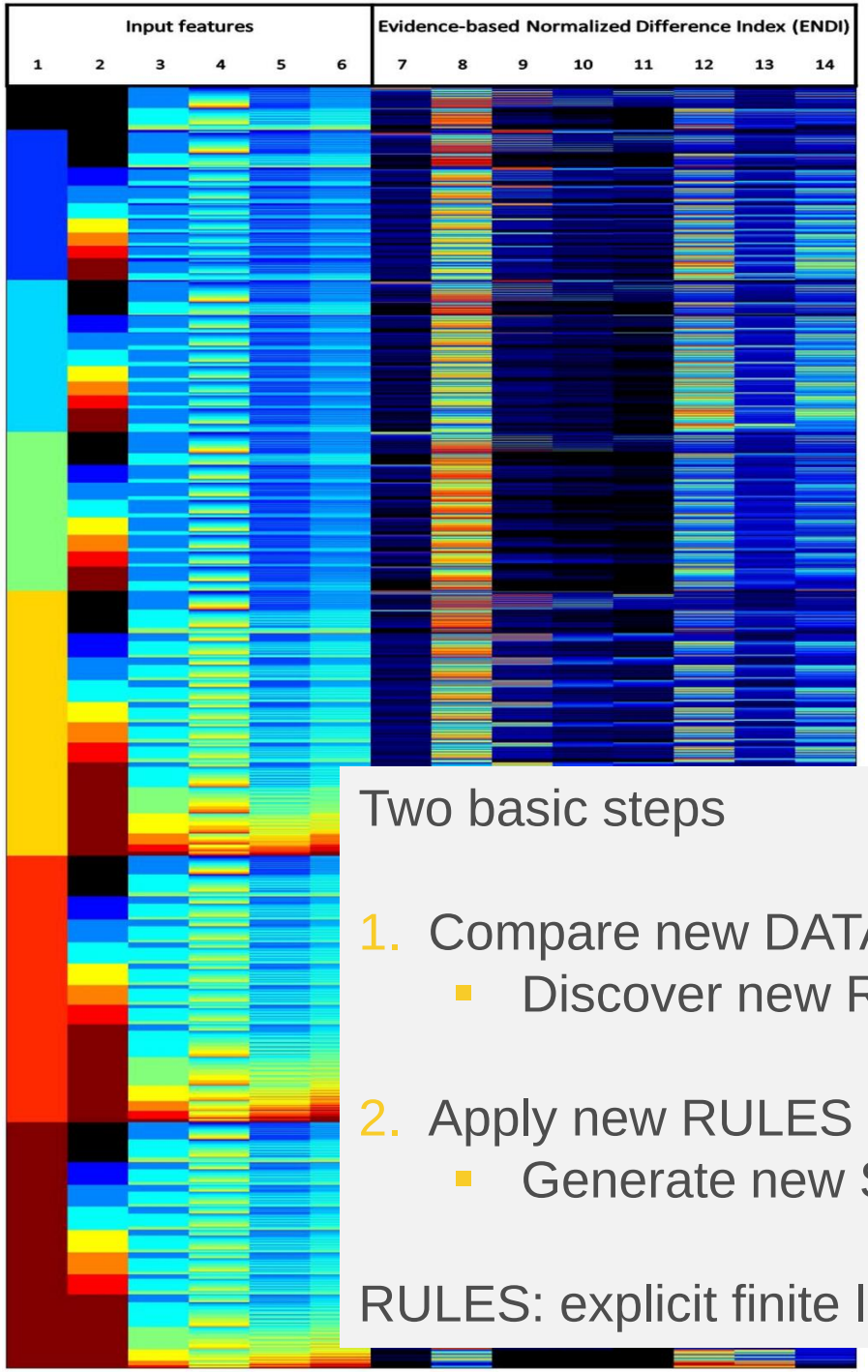
- *Support Vector Machine*
- *Random Forest*
- *Artificial Neural Network*

recall:

- evidence-based policy action
- simple and easily interpretable assumptions
- human discussability of the model logic
- full repeatability of the measures







SML concept

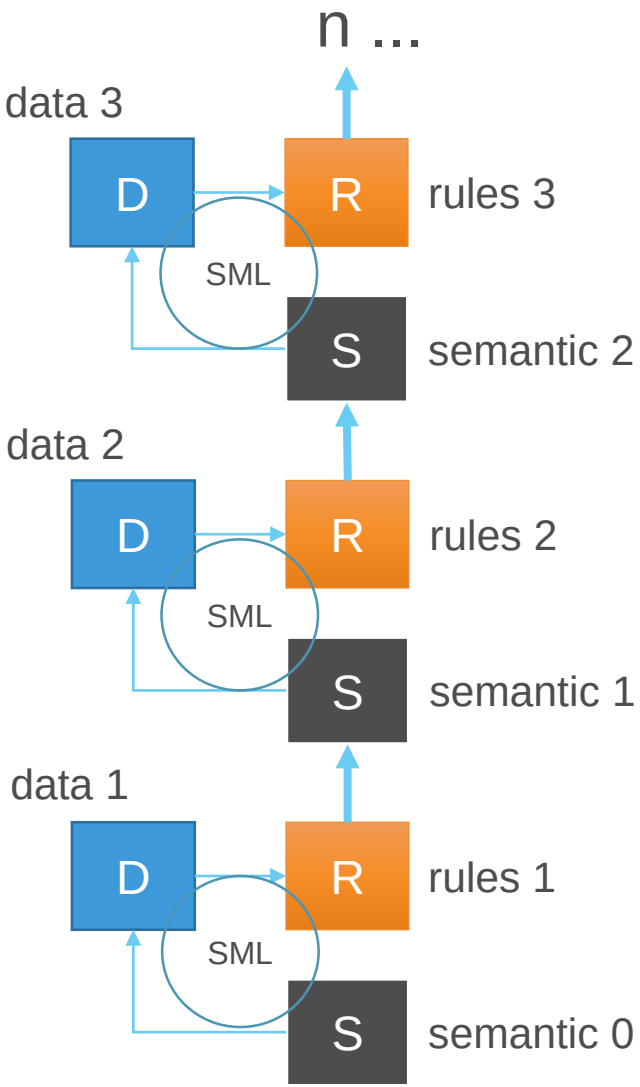
Pesaresi M, Syrris V, Julea A. A New Method for Earth Observation Data Analytics Based on Symbolic Machine Learning. *Remote Sensing*. 2016; 8(5):399. https://doi.org/10.3390/rs8050399

Two basic steps

1. Compare new DATA with existing SEMANTIC
   - Discover new RULES
2. Apply new RULES to new DATA
   - Generate new SEMANTIC

RULES: explicit finite list

self-evolutionary GHSL production system



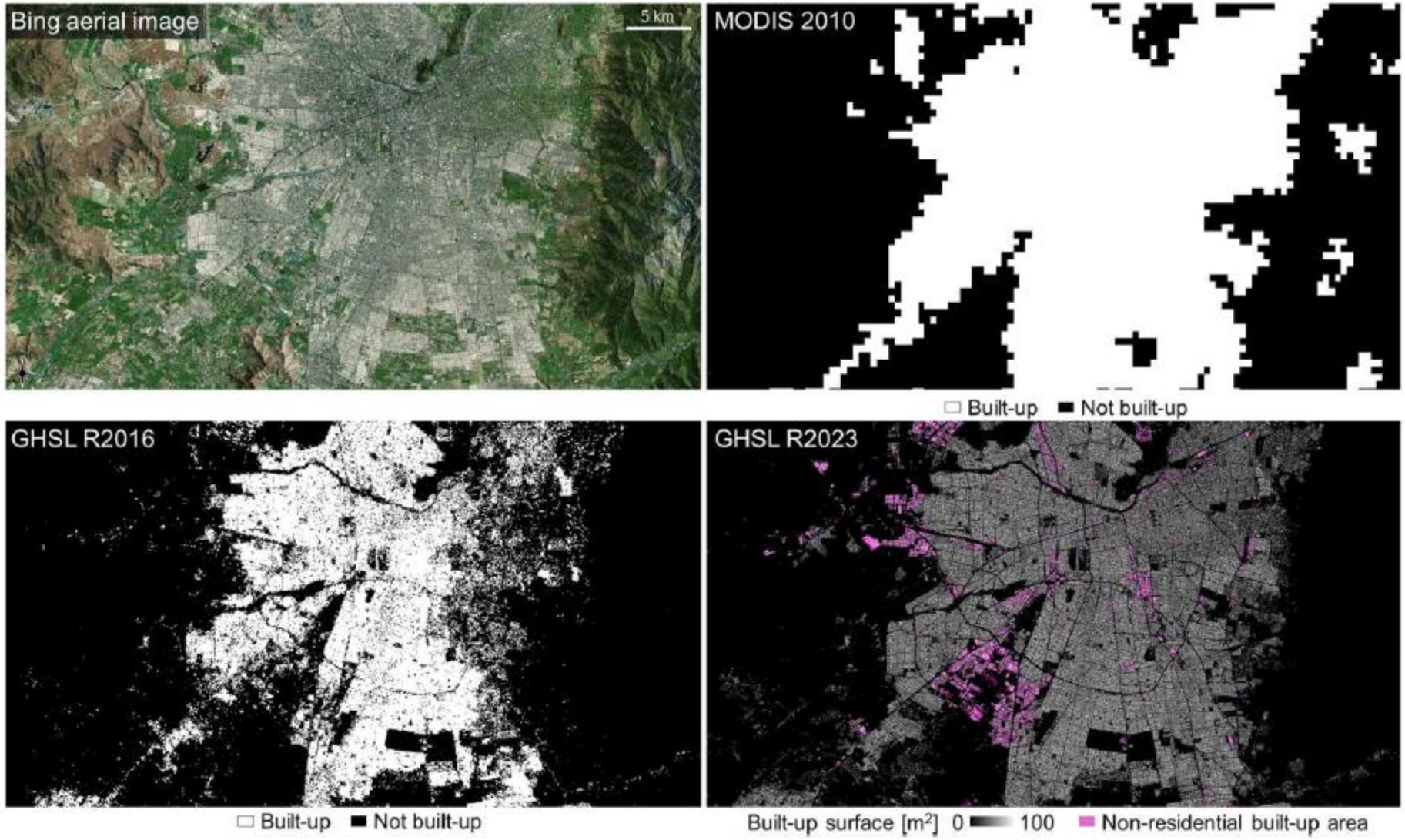
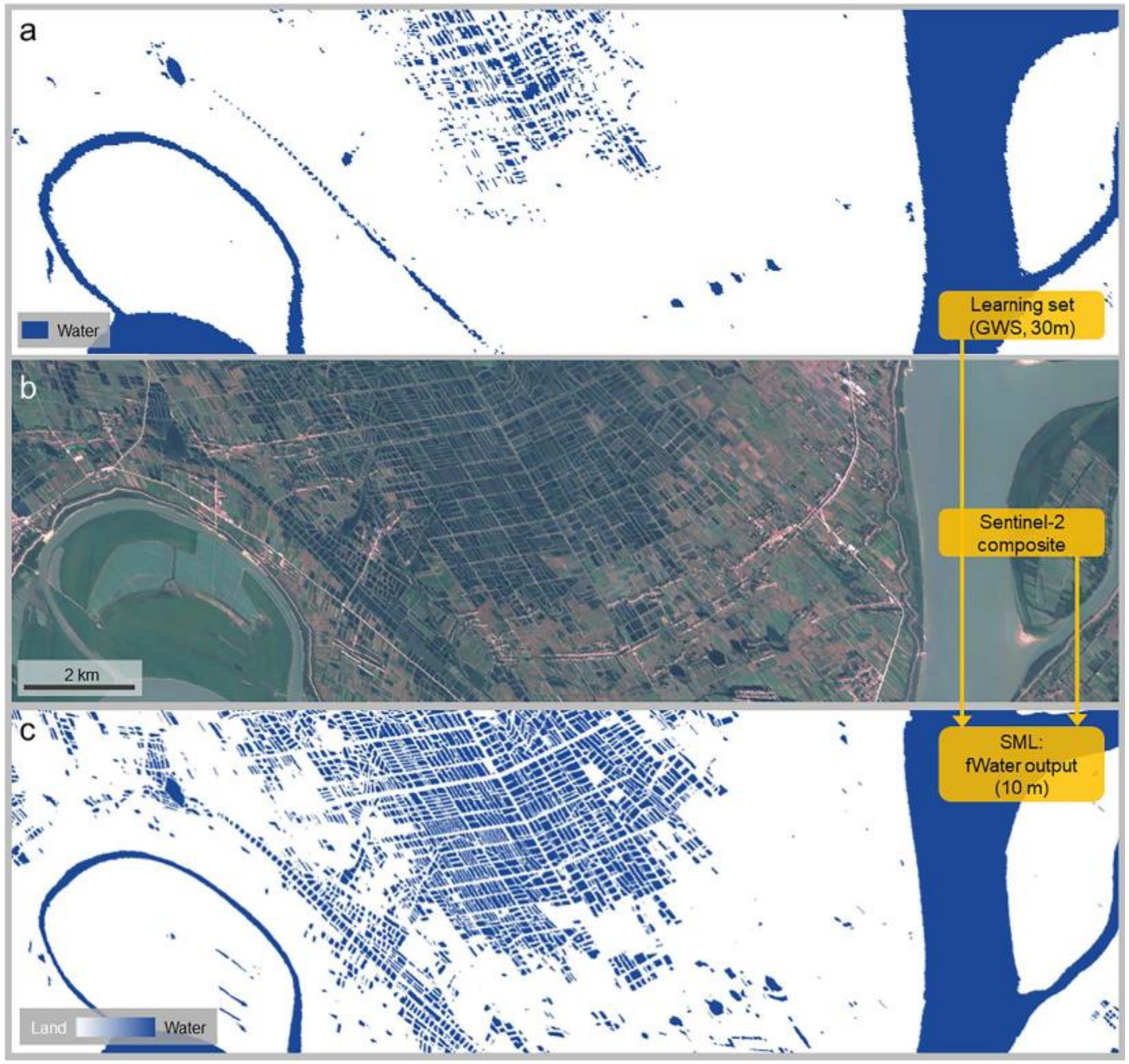


Figure 8. Self-evolutionary GHSL information production system. Santiago de Chile, satellite image (Microsoft ™ Bing © 2024, Earthstar Geographics), MODIS-derived prior knowledge available before the GHSL, the Landsat-derived GHSL BU information generated by using MODIS as SML learning set, and the Sentinel-2-derived information using the Landsat GHSL (R2016 R2019) as SML learning set.

European Commission



Illustrating the production of fWATER for the region south of Wuhan, west of the Yangtze River, China.

- (a) Prior information about water presence from the GSW at 30 m resolution as extracted from Landsat data classification, which supports the SML classification of Sentinel-2 data as train set,
- (b) Sentinel-2 composite imagery,
- (c) fWATER, the output of the SML classification of 10 m resolution Sentinel-2 imagery.

Note the presence of thin fish farm structures that are identified by the SML in the data, despite them largely contradicting the available training set.

conclusions

- GHSL theoretical and methodological background was illustrated
- Multi-faceted novelty landscape of the GHSL was delineated
- Too ambitious to discuss about all aspects in the short time available
- Details on the population spatial distribution and quality control of the GHSL R2023 in the next presentations
- More humbly, the main purpose here was to tease your curiosity and inspiration – please make your own explorative path in the paper!

Pesaresi, M., Schiavina, M., Politis, P., Freire, S., Krasnodębska, K., Uhl, J. H., … Kemper, T. (2024). Advances on the Global Human Settlement Layer by joint assessment of Earth Observation and population survey data. *International Journal of Digital Earth*, *17*(1). https://doi.org/10.1080/17538947.2024.2390454

Thank you
and keep in touch

Global Human Settlement Layer
https://human-settlement.emergency.copernicus.eu/
jrc-ghsl@ec.europa.eu

EU Science Hub
Joint-research-centre.ec.europa.eu

European Commission