UNIVERSITY OF CAMBRIDGE

Joshua Dimasaka, Christian Geiß, Emily So

Spatial Disaggregation of Rwandan Building Exposure and Vulnerability via Weakly Supervised Conditional Census-Constrained Clustering (C4) using Earth Observation Data

Deep Conditional Census Constrained Clustering (DeepC4) for Large-scale Multi-task Spatial Disaggregation of Urban Morphology

AI4ER CURBE UNIVERSITÄT BONN DLR HIDA

Introduction

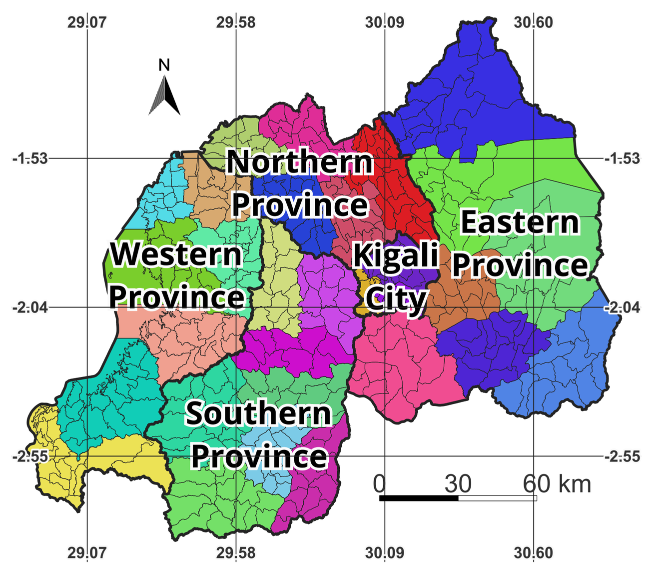

Figure 1. Geographic extent of Rwanda. The boundaries of sectors and provinces are res- pectively displayed with thinner and thicker lines. Every distinct color represents a district.

- **Understanding disaster risk** is an urgent priority of the UN Sendai Framework for Disaster Risk Reduction 2015-2030, particularly for under-resourced areas wherein data collection tools have been inequitably unaffordable.
- Recent large-scale mapping efforts of **Global Earthquake Model (GEM)** and the **METEOR** project benefited from the use of Earth Observation (EO) data. Addressing their temporal inconsistency and compounded spatiotemporal uncertainties contributes to **the reliability of EO-based data-driven efforts.**
- Due to the site-specific and heterogeneous nature of the exposure and physical vulnerability (urban morphology), we employed **a deep clustering approach** to model the EO high-dimensionality and the non-linear capability of the predictive function, **while considering coarse-grained census info as cluster-level constraints for performance guarantee.**

Methodology

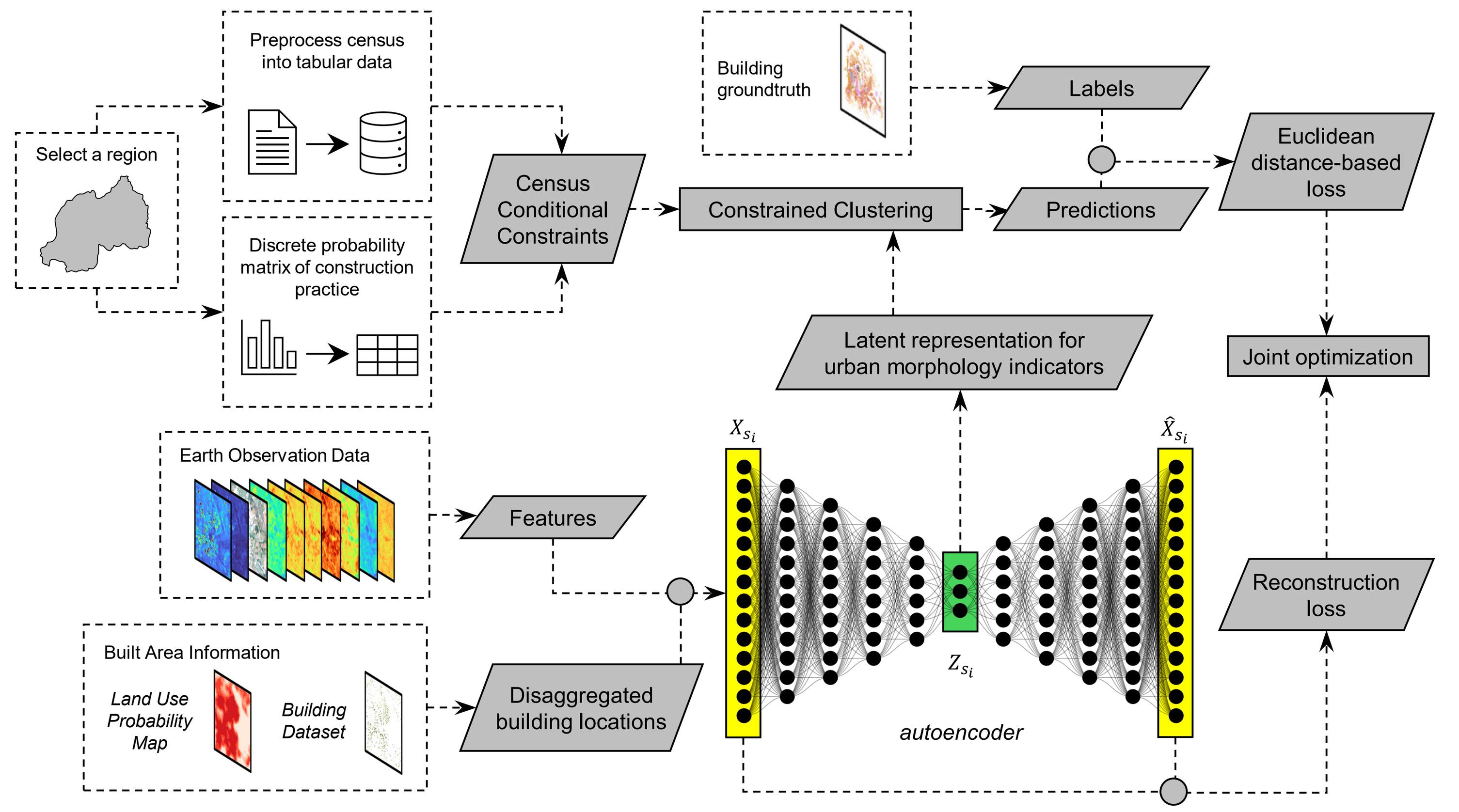

Figure 2. DeepC4 Implementation. From left to right, we began with encoding of census conditional constraints, preparation of EO features, and selection of disaggregated building locations from various imperfect sources. We then trained an AutoEncoder to obtain a set of reduced-dimension representations that were used for constrained clustering algorithms. Jointly trained using the reconstruction and prediction losses, we evaluated the performance of AutoEncoder using the available building-level groundtruth.

Acknowledgements

This work is funded by the **UKRI Centre for Doctoral Training in Application of Artificial Intelligence to the study of Environmental Risks (AI4ER) (EP/S022961/1)** and the **Helmholtz Information & Data Science Academy (HIDA)** for providing financial support enabling a short-term research stay in the German Aerospace Center.

Open Science

The data, code, and web documentation are available at: **github.com/riskaudit/DeepC4 | doi.org/10.5281/zenodo.13119552**

DeepC4 achieves the closest estimates vs official census records

- When aggregated at the national level, our findings show that the outputs of DeepC4 (that considers sector-level census records) have the lowest percent errors (or highest accuracy) in the number of dwellings and occupants. The table below also indicates that DeepC4 is followed by GEM (~50k more buildings) and then METEOR (~170k more buildings), in the increasing order of the number of buildings.

Table 1. DeepC4-GEM-METEOR comparison on the total buildings, dwellings, and occupants.

Counts	Census	DeepC4	GEM	METEOR
Building	-	3,200,582	3,258,527	3,368,644
Dwelling	3,312,743	**3,350,277 (1.13%)**	3,379,883 (2.03%)	-
Occupant	13,100,600	**13,246,394 (1.11%)**	13,531,804 (3.29%)	-

- The comparative difference between DeepC4 and METEOR indicates that outputs of DeepC4 identify more grid pixels with exposure characteristics (with no METEOR estimates) and have higher numerical estimates, particularly for the City of Kigali.

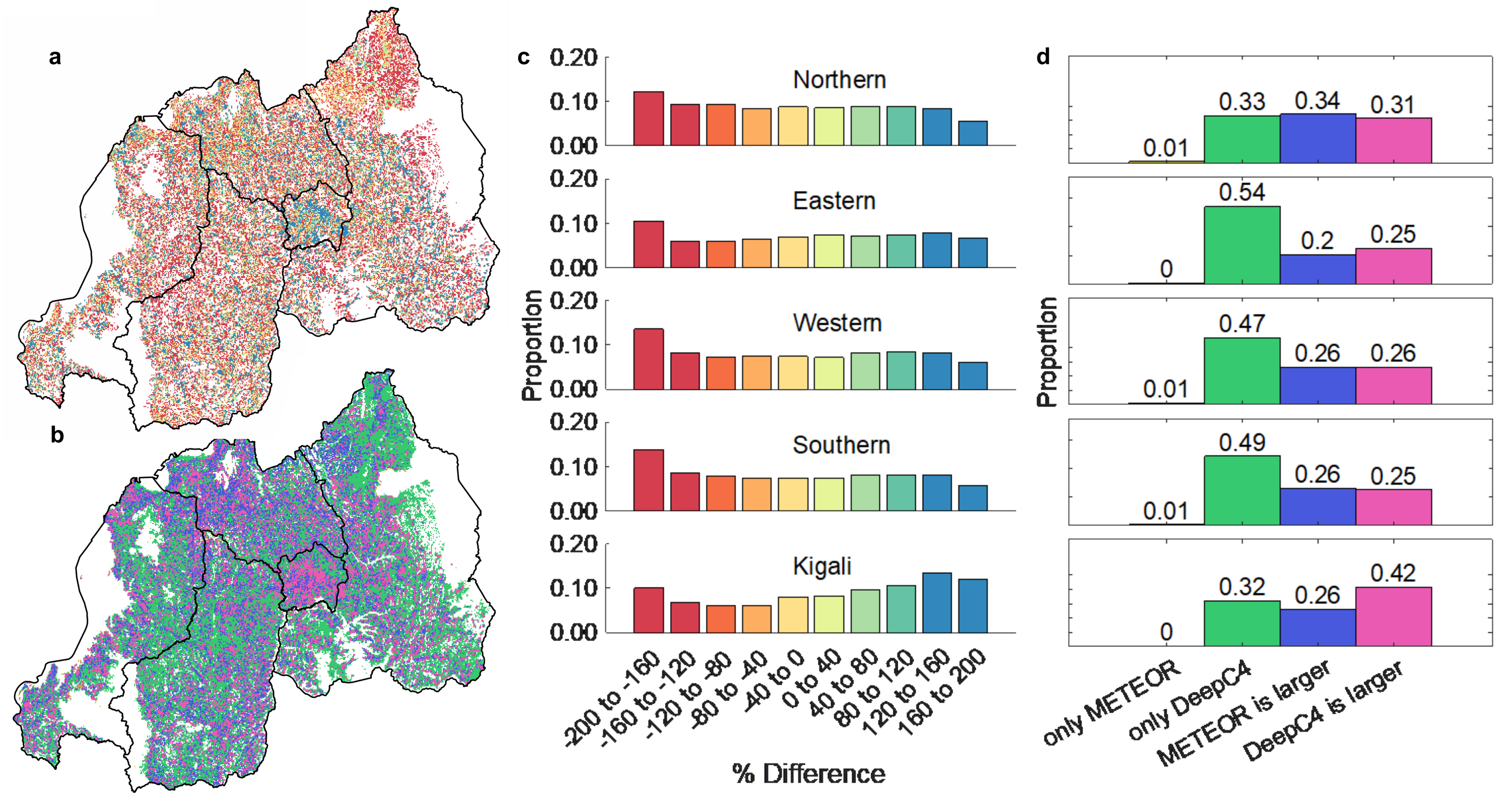

Figure 3. DeepC4-METEOR comparison of spatial disaggregation (at 500-meter pixel). Calculated percent difference **(a)** visualized on a map and **(c)** expressed as a histogram chart showing proportion of pixels per province. Categorical difference as **(b)** a map and **(d)** a histogram chart showing the proportion of pixels for each descriptive category.

- The table below consistently shows that the number of dwellings considered by the DeepC4 has **higher accuracy** than those of GEM with respect to the official census records, which can be attributed to the DeepC4's use of the sector-level census statistics.

Dwelling	Province	Census	DeepC4	GEM	% Difference
Total	North	506,064	**509,744 (0.7%)**	483,807 (-4.4%)	5.22%
	East	886,132	**896,945 (1.2%)**	950,320 (7.2%)	5.78%
	West	671,506	**677,394 (0.9%)**	733,609 (9.2%)	7.97%
	South	760,173	**770,372 (1.3%)**	790,089 (3.9%)	2.53%
	Kigali	488,868	**495,822 (1.4%)**	422,058 (-14%)	16.07%
Rural	North	417,670	**419,807 (0.5%)**	403,373 (-3.4%)	3.99%
	East	700,049	**705,767 (0.8%)**	792,306 (13%)	11.55%
	West	522,847	**527,834 (1.0%)**	611,632 (17%)	14.71%
	South	651,454	**654,395 (0.5%)**	658,727 (1.1%)	0.66%
	Kigali	56,436	**60,973 (8.0%)**	70,176 (24%)	14.03%
Urban	North	88,394	**89,937 (1.7%)**	80,434 (-9.0%)	11.16%
	East	186,083	**191,178 (2.7%)**	158,014 (-15%)	18.99%
	West	148,659	**149,560 (0.6%)**	121,977 (-18%)	20.32%
	South	108,719	**115,978 (6.7%)**	131,362 (21%)	12.44%
	Kigali	432,432	**434,849 (0.6%)**	351,882 (-19%)	21.09%

Table 2. DeepC4-GEM comparison of spatial disaggregation (at the province level). For each province, this shows the number of urban, rural, and total dwellings, percent errors (accuracy) with respect to the official census records, and percent difference between the estimates of DeepC4 and GEM.

DeepC4 effectively jointly learns for a multi-task objective

- The learning curves and example maps below show that DeepC4 has effectively achieved a set of stable reduced-dimension representations while minimizing the prediction loss and balancing the performance trade-off across three different tasks.

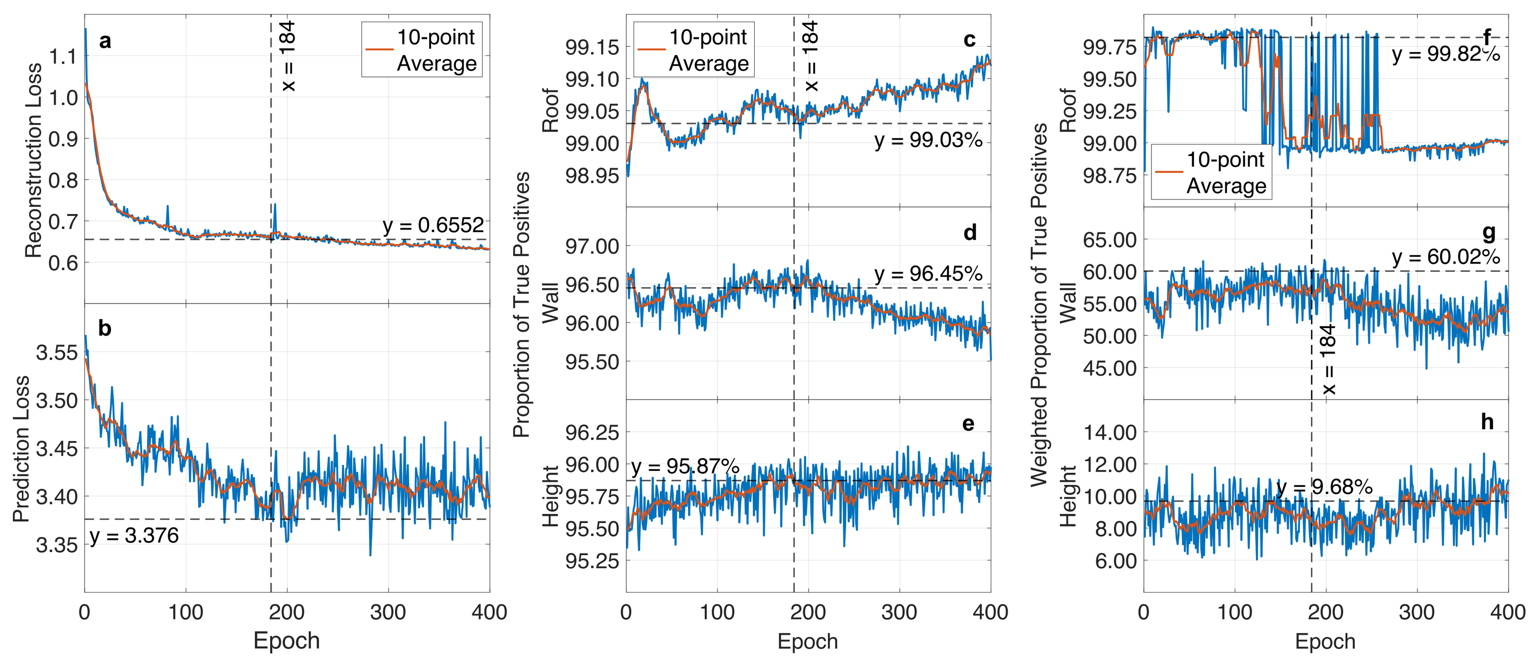

Figure 4. Learning curves. (a) reconstruction loss (how effective AutoEncoder is in reducing the EO high-dimensionality into a meaningful reduced-dimension representations); **(b)** prediction loss (how accurate the reduced-dimension representations are with respect to available building groundtruth info); **(c-e)** proportion of true positives for each prediction task (roof, wall, and height); and **(f-h)** similar plots but accounting for the imbalanced number of target classes.

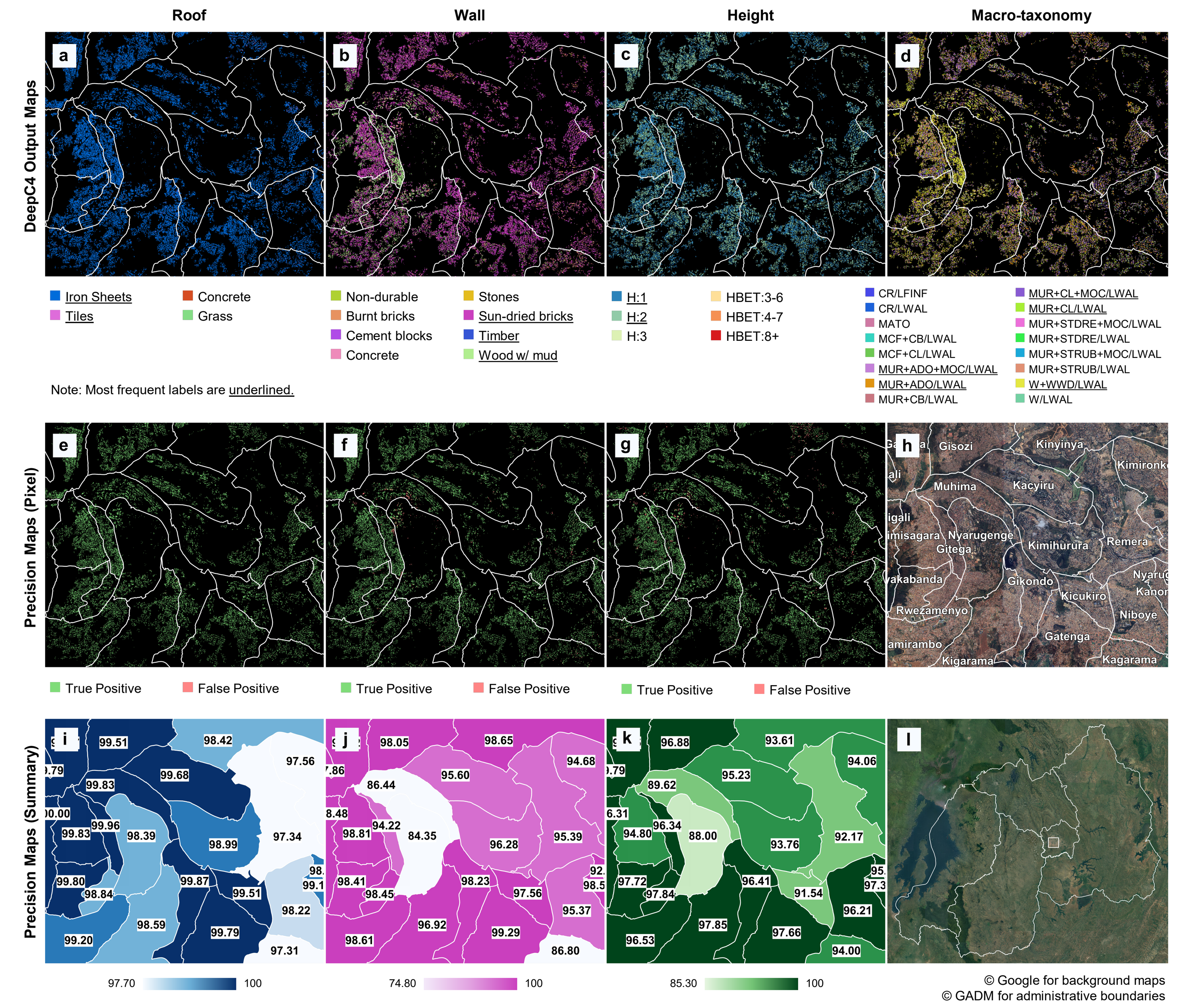

Figure 5. Usefulness of partially available building-level groundtruth. (a-d) a set of prediction maps as a result of DeepC4; **(e-g)** corresponding precision maps; **(h)** reference background map with sector labels; **(i-k)** aggregated sector-level values of precision maps; and **(l)** reference location.

DeepC4 maps Rwandan exposure and vulnerability at large scale

- Visualized at different aggregation levels below, the outputs of DeepC4 both satisfy the census records and expert belief systems in generating a more effective representation of exposure and physical vulnerability for the entire Rwanda. Various imperfect building footprint datasets have also demonstrated their flexibility and usefulness in extracting meaningful EO signals for any subsequent DeepC4 implementation.

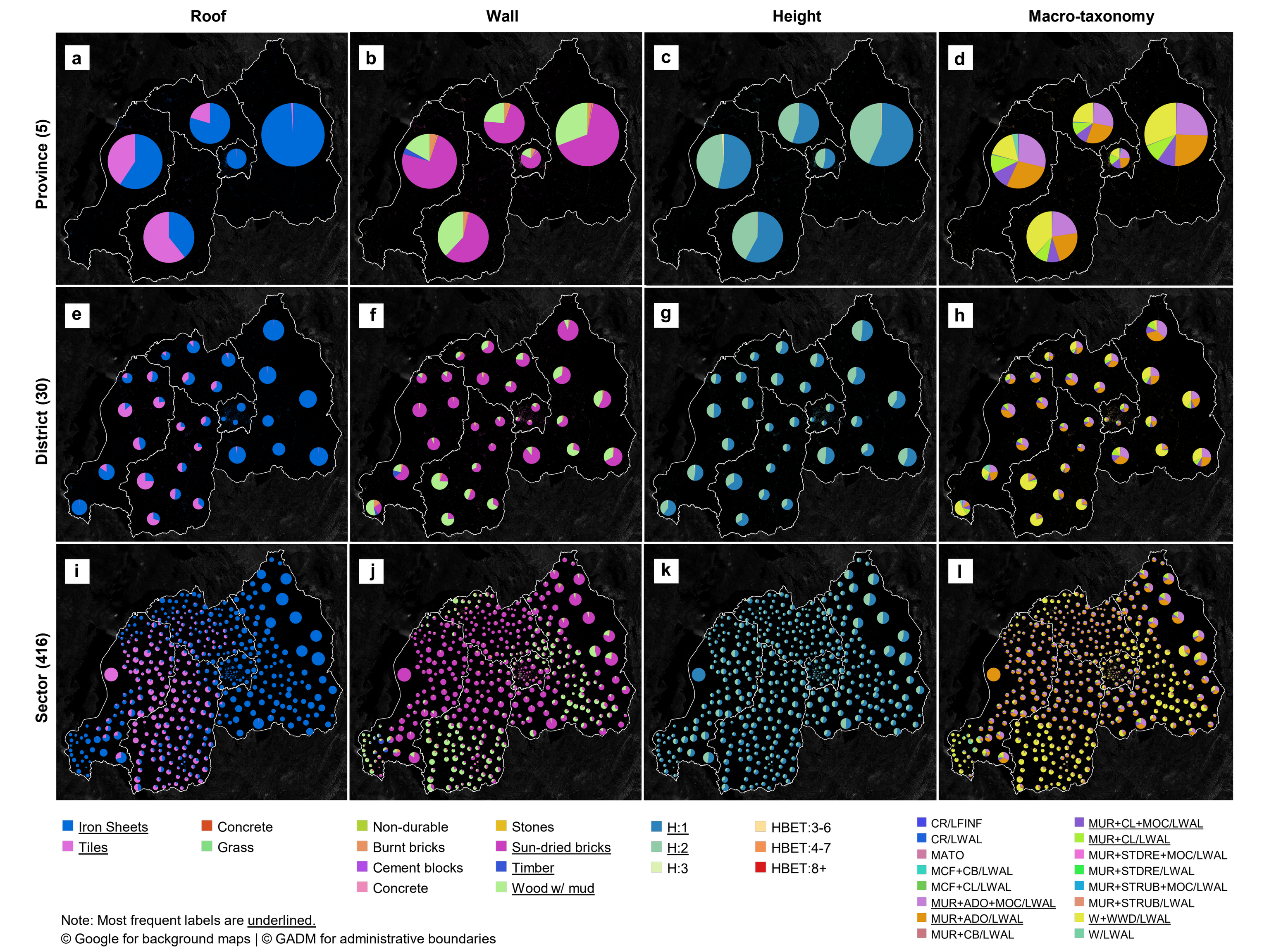

Figure 6. Summary diagrams of large-scale 10-meter maps of roof, wall, height, and macro-taxonomy classes at the: (a-d) province, **(e-h)** district, and **(i-l)** sector levels.

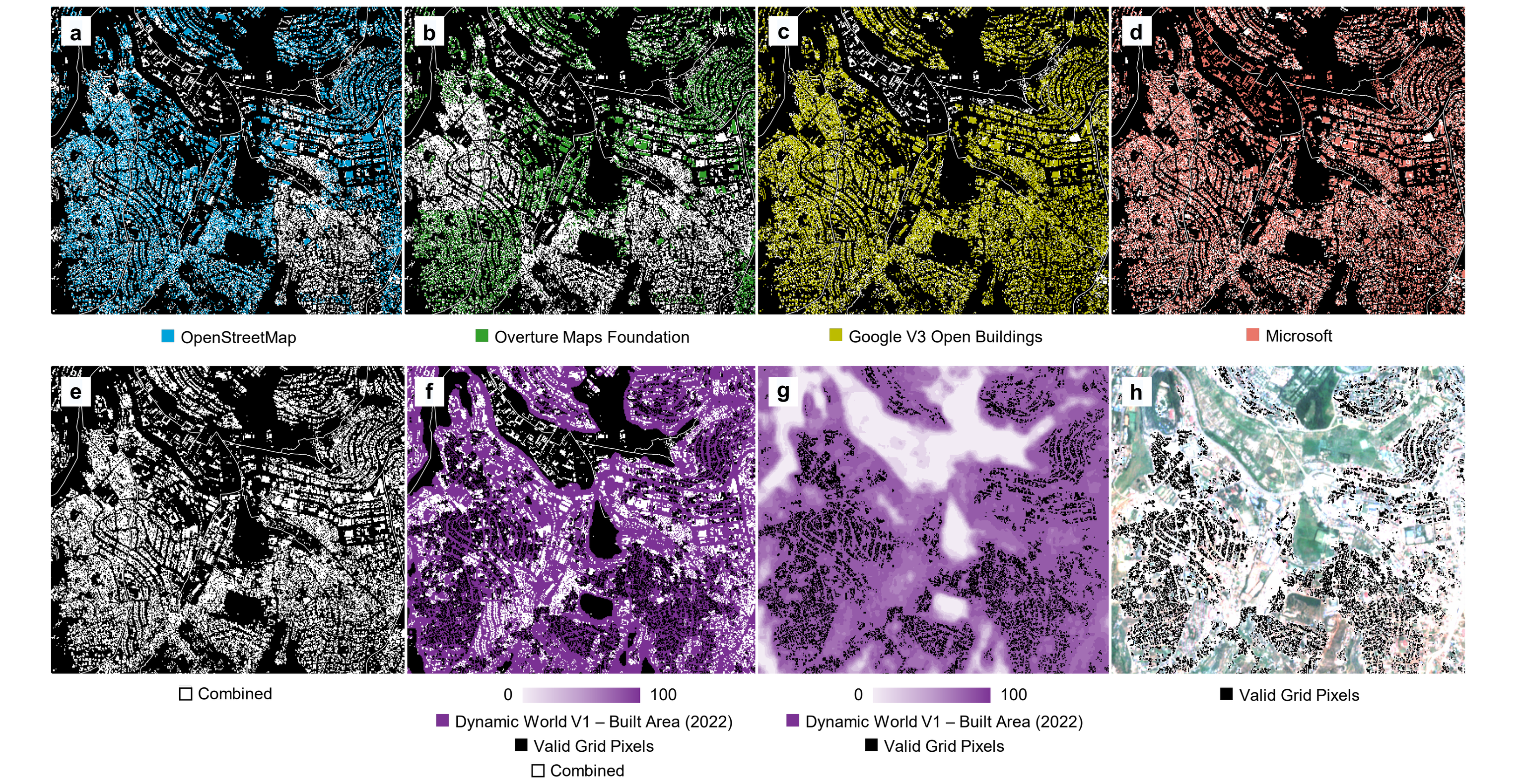

Figure 7. Comparison of built area information. (a-e) various imperfect building footprint sources and **(f-g)** the use of Dynamic World V1 near-real-time land use land cover as additional information for **(h)** possible building locations for subsequent clustering implementation in DeepC4.