

Matching environmental data produced from remotely sensed images with demographic data in Sub-Saharan Africa

Basile Rousse^{1,2}, Sylvain Lobry¹, Géraldine Duthé², Laurent Wendling¹, Valérie Golaz^{2,3}

1: Lipade, Université Paris Cité

2: Ined, Institut national d'études démographiques

3: Aix-Marseille Univ, IRD, LPED, Marseille, France

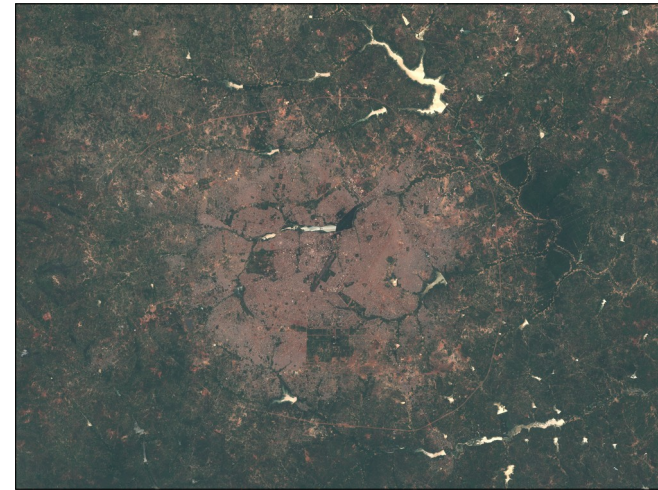
Summary

- I. Context
- II. How to link remote sensing images and demography ?
- III. About deep learning
- IV. Local Climate Zones in Sub-Saharan Africa
- V. Local Climate Zones and Malaria
- VI. Conclusion

Context

Context

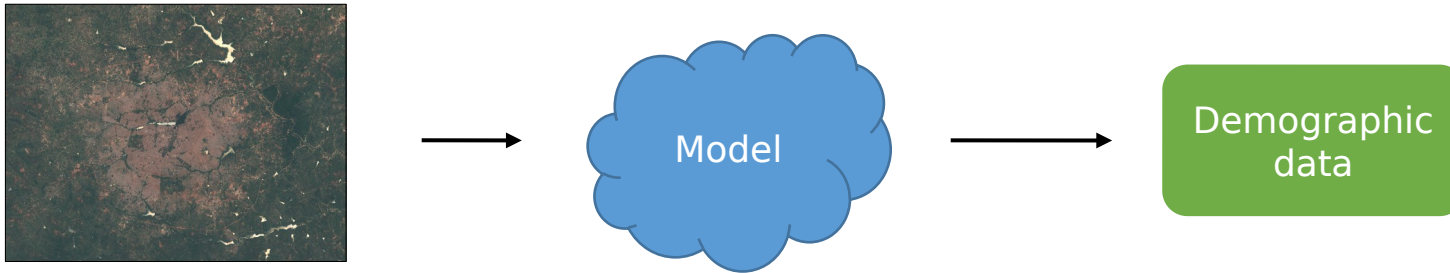
- Impact of environmental change in demographic studies and sometimes scarce demographic data in some Sub-Saharan countries
- Demographic and Health Surveys program¹ provide:
 - Demographic data: Malaria, household
 - Environmental data: rainfall, NDVI, temperature...
 - Geo-locations (2 and 10 km buffers)
- Large amount of Sentinel images (ESA) :
 - with a high refresh rate (between 2 and 5 days) that allows to monitor environmental changes
 - Resolution = 10x10m
 - Open access



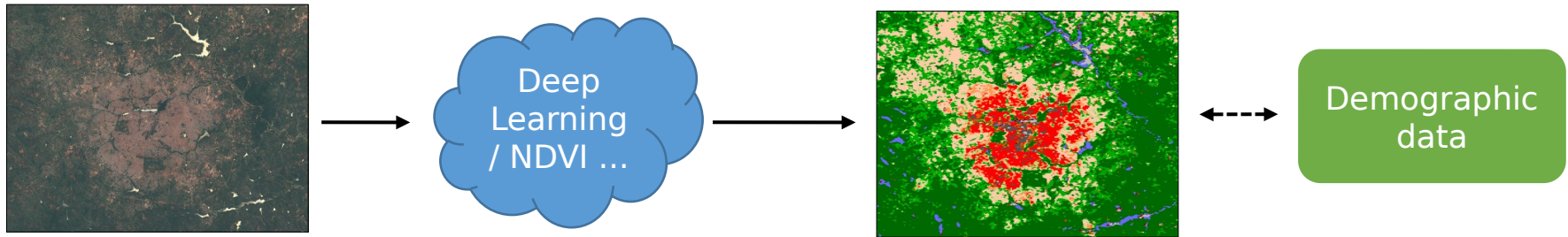
Ouagadougou, Burkina Faso, Sentinel 2 09/2021

How to link remote sensing images and demography?

→ Directly modelling demographic data using remotely sensed images.



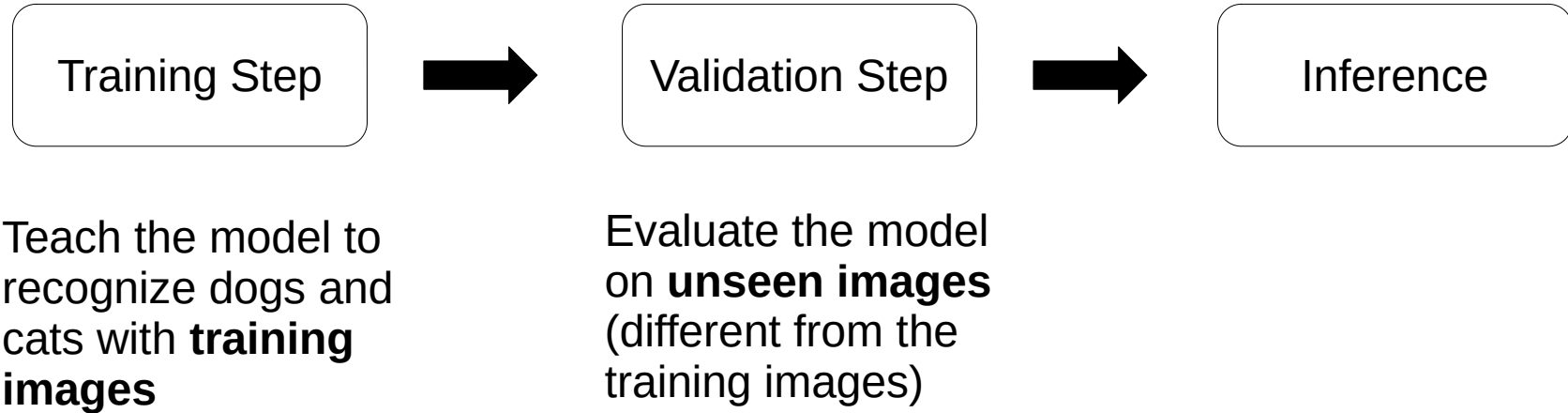
→ Predicting environmental indicators that can be linked to demographic data.



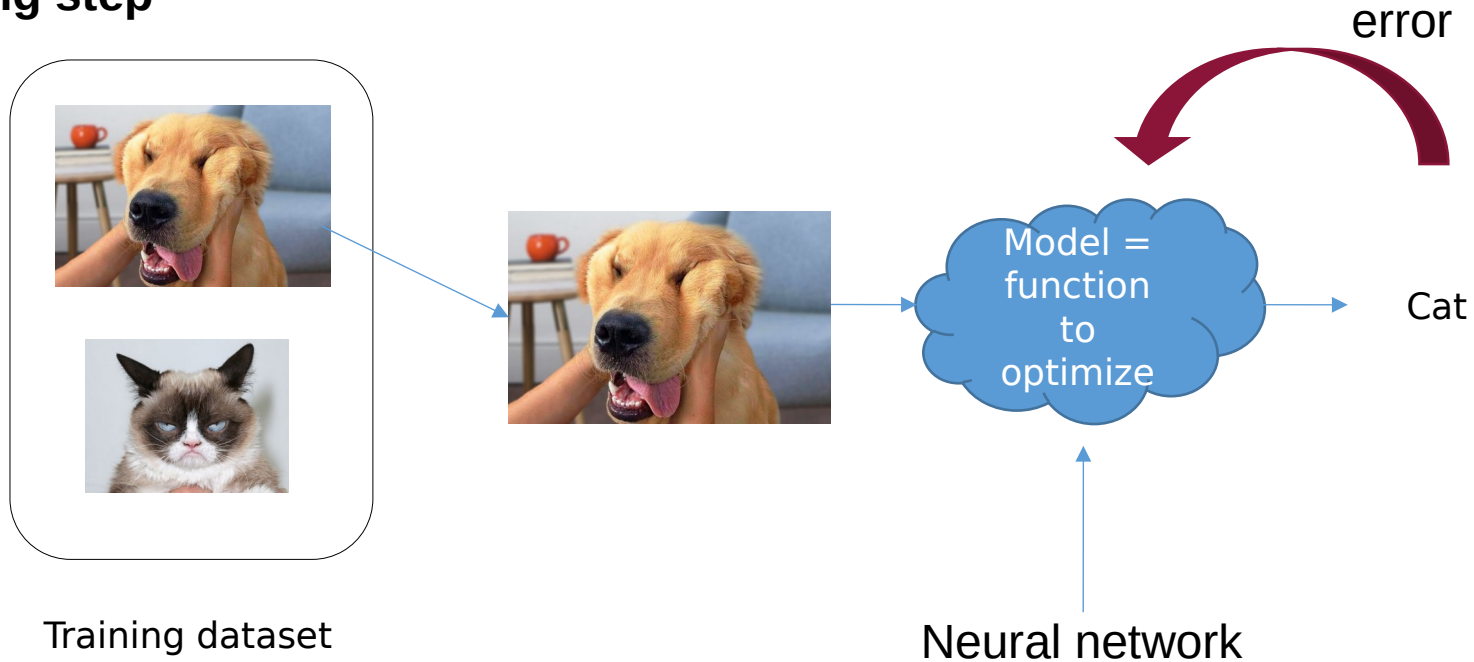
What about a complete land cover classification scheme?

About Deep Learning

General pipeline : supervised learning



Training step



⇒ Supervised DL is **very sensitive** to training data

Local Climate Zones in Sub Saharan Africa

Local Climate Zones (LCZ)

- Classification scheme based on the land structure for heat island detection (Stewart et al. 2012)
- Independent from cultural aspects
 - ➔ Can be applied globally
- Interesting work:
 - ➔ So2Sat and GUL: dataset based S2 images of 42 cities very few in SSA (Zhu et al. 2019)

Built types		Land cover types	
1 	Compact highrise Dense mix of tall buildings to tens of stories. Few or no trees. Land cover mostly paved. Concrete, steel, stone, and glass construction materials.	A 	Dense trees Heavily wooded landscape of deciduous and/or evergreen trees. Land cover mostly pervious (low plants). Zone function is natural forest, tree cultivation or urban park.
2 	Compact midrise Dense mix of midrise buildings (3–9 stories). Few or no trees. Land cover mostly paved. Stone, brick, tile, and concrete construction materials.	B 	Scattered trees Lightly wooded landscape of deciduous and/or evergreen trees. Land cover mostly pervious (low plants). Zone function is natural forest, tree cultivation, or urban park.
3 	Compact lowrise Dense mix of lowrise buildings (1–3 stories). Few or no trees. Land cover mostly paved. Stone, brick, tile, and concrete construction materials.	C 	Bush, scrub Open arrangement of bushes, shrubs, and short, woody trees. Land cover mostly pervious (bare soil or sand). Zone function is natural scrubland or agriculture.
4 	Open highrise Open arrangement of tall buildings to tens of stories. Abundance of pervious land cover (low plants, trees). Concrete, steel, stone, and glass construction materials.	D 	Low plants Featureless landscape of grass or herbaceous plants/crops. Few or no trees. Zone function is natural grassland, agriculture, or urban park.
5 	Open midrise Open arrangement of midrise buildings (3–9 stories). Abundance of pervious land cover (low plants, scattered trees). Concrete, steel, stone, and glass construction materials.	E 	Bare rock or paved Featureless landscape of rock or paved cover. Few or no trees or plants. Zone function is natural desert (rock) or urban transportation.
6 	Open lowrise Open arrangement of lowrise buildings (1–3 stories). Abundance of pervious land cover (low plants, scattered trees). Wood, brick, stone, tile, and concrete construction materials.	F 	Bare soil or sand Featureless landscape of soil or sand cover. Few or no trees or plants. Zone function is natural desert or agriculture.
7 	Lightweight lowrise Dense mix of single-story buildings. Few or no trees. Land cover mostly hard-packed. Lightweight construction materials (e.g., wood, thatch, corrugated metal).	G 	Water Large, open water bodies such as seas and lakes, or small bodies such as rivers, reservoirs, and lagoons.
8 	Large lowrise Open arrangement of large lowrise buildings (1–3 stories). Few or no trees. Land cover mostly paved. Steel, concrete, metal, and stone construction materials.	VARIABLE LAND COVER PROPERTIES Variable or ephemeral land cover properties that change significantly with synoptic weather patterns, agricultural practices, and/or seasonal cycles.	
9 	Sparsely built Sparse arrangement of small or medium-sized buildings in a natural setting. Abundance of pervious land cover (low plants, scattered trees).	b. bare trees	Leafless deciduous trees (e.g., winter). Increased sky view factor. Reduced albedo.
10 	Heavy industry Lowrise and midrise industrial structures (towers, tanks, stacks). Few or no trees. Land cover mostly paved or hard-packed. Metal, steel, and concrete construction materials.	s. snow cover	Snow cover >10 cm in depth. Low admittance. High albedo.
		d. dry ground	Parched soil. Low admittance. Large Bowen ratio. Increased albedo.
		w. wet ground	Waterlogged soil. High admittance. Small Bowen ratio. Reduced albedo.

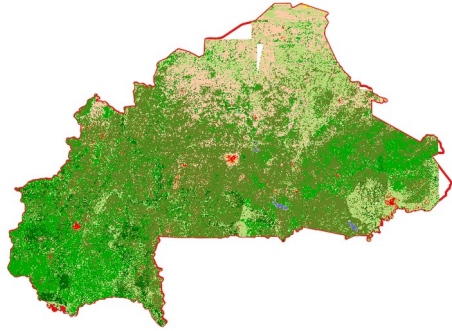
Adapting LCZ to Sub-Saharan countries : A case study of Burkina Faso

DL is very sensitive to training data ⇒ **similarity** between target and training areas

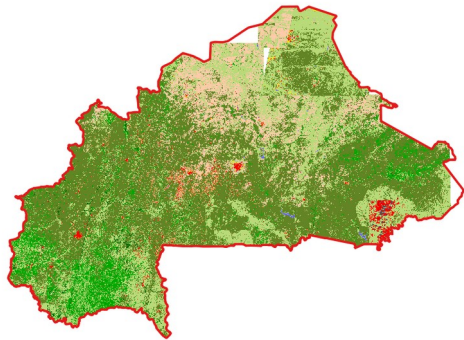
	Challenge	Solution
Spatial variations	Training cities are morphologically different from Ouagadougou	Train the model on morphologically similar cities to reduce the domain gap between training cities and Ouagadougou
Temporal variations	High variations with dry and rainy seasons not in data	Extract information from unlabelled images taken from both seasons (SimCLR, Chen et al. 2020)

⇒ Model **adapted** to Burkina Faso and **robust** to seasons

Taking advantage of temporal data



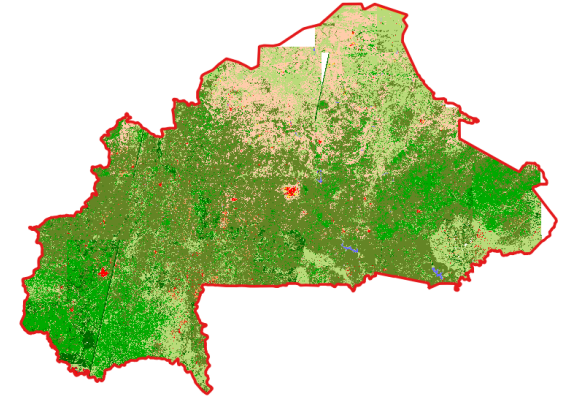
LCZ map 01/2017



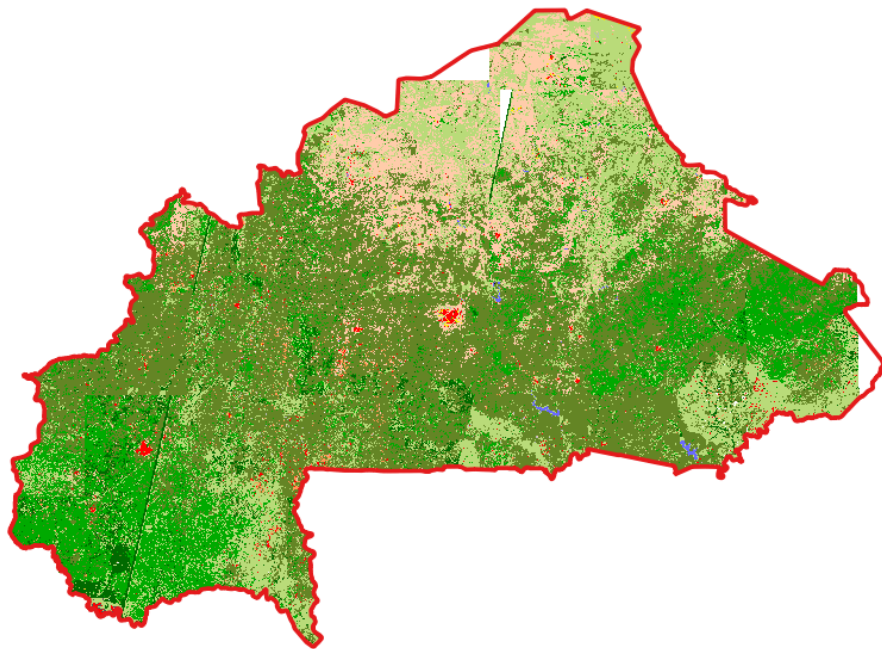
LCZ map 01/2018



Bayes' theorem
+ Markov chain



Final results
01/2018



- Resolution 320m
- Cities in red (compact low-rise)
- Mostly bush, scrub



Local Climate Zones and demography

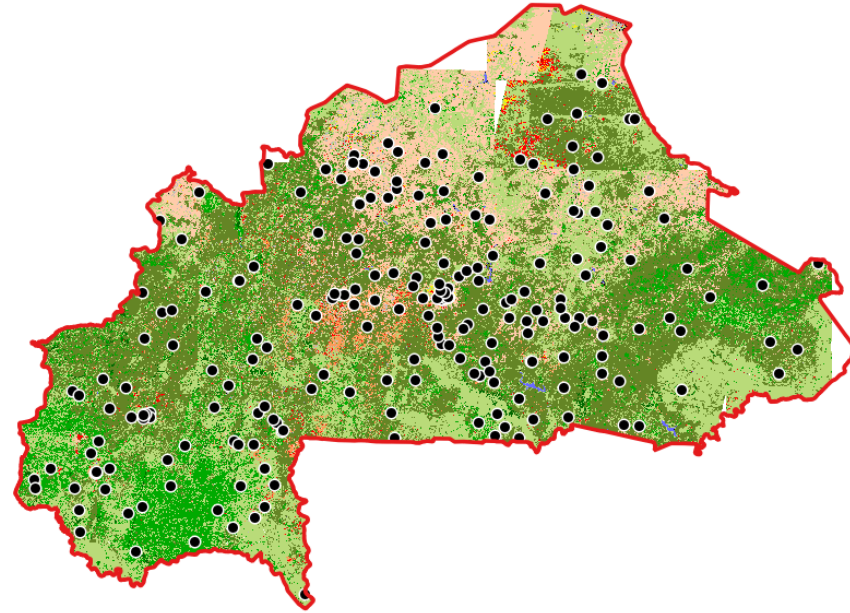
Malaria Indicator Survey 2017/2018 Burkina Faso

Objectives:

- Estimate up-to-date basic demographic and health indicators about Malaria

Informations:

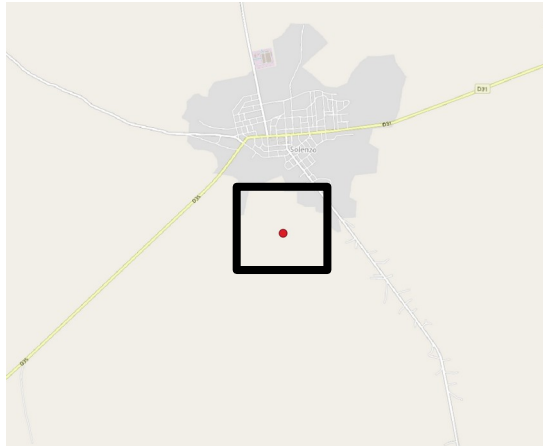
- Between November 2017 and March 2018
- 17 study areas
- 245 enumeration zones (or clusters), geolocated (<2kms for urban areas, <10kms for rural areas)
- 6322 households interviewed



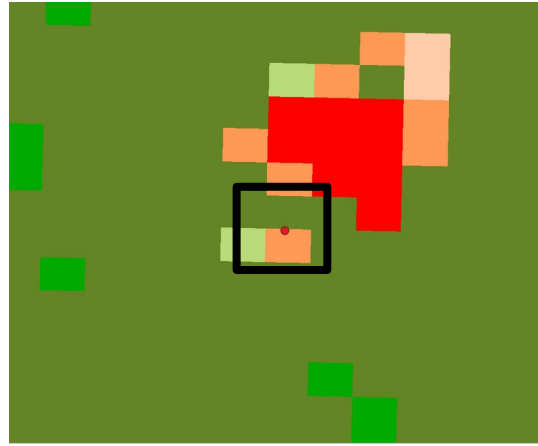
⇒ Focus on the Malaria prevalence for 6-59 months children

Linking MIS and LCZs

- The first step consists in calculating the proportion of households where at least one rapid test was positive, by cluster.
- This value is then attributed to each pixel of a 640mx640m square, centered on the centroid of the cluster.
- As we previously predicted the LCZ class of each pixel, we can link the geo-location of clusters with their malaria positivity rates and the LCZ classes in their close environments.



Solenzo, Burkina Faso, OSM 2022



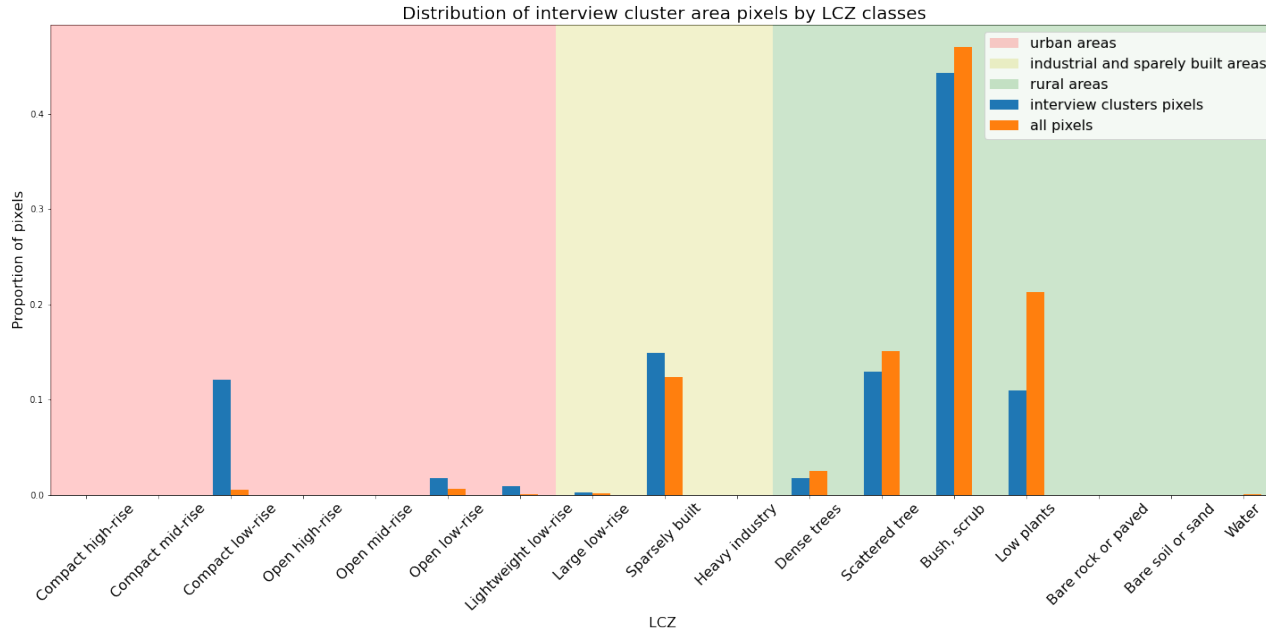
LCZ map

- Mostly bush, scrub
- City

An exemple of cluster

MIS 2017/2018 – LCZ classes in Burkina Faso and study areas

→ What LCZ classes are represented in interview clusters areas ?



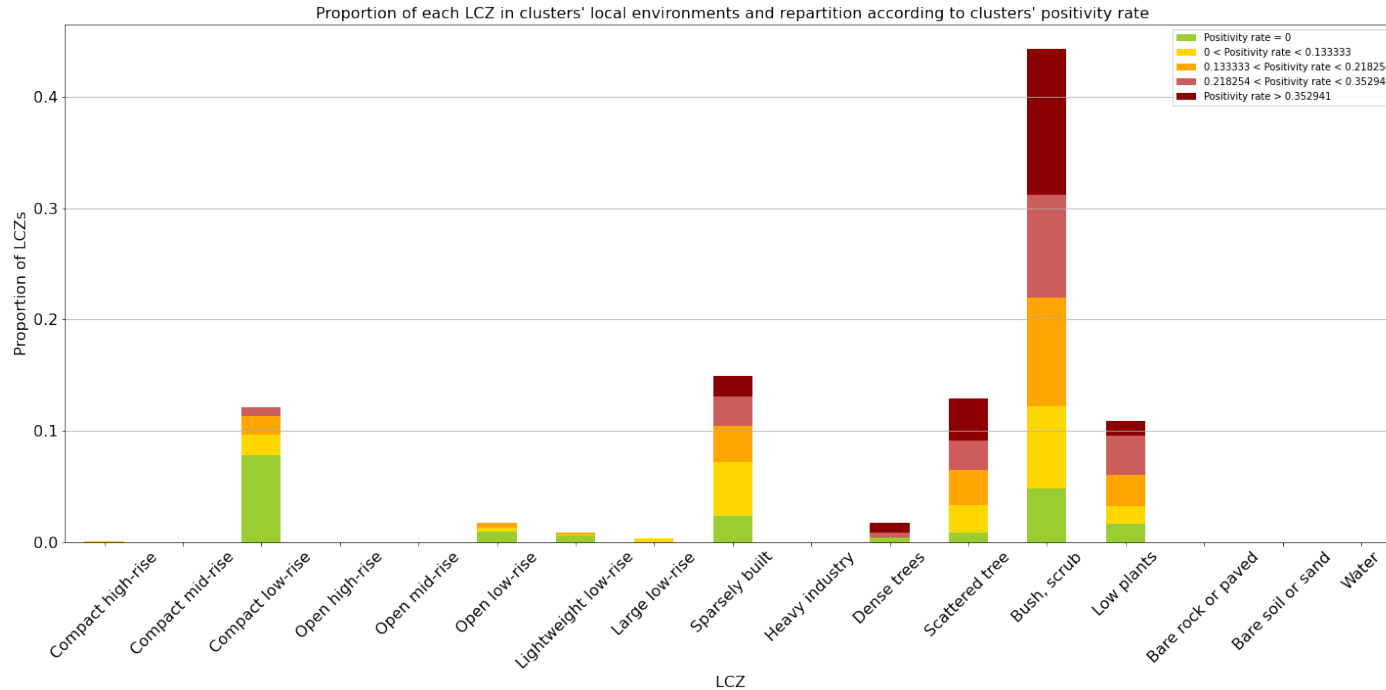
- Bush, scrub far more represented in general, as in the entire LCZ map
- Water, Bare rock and paved and bare soil in Burkina Faso but not in clusters' areas
- Very or no “high-rise” and “mid-rise”: no very high buildings in Burkina Faso

MIS 2017/2018 – Group clusters according to malaria positivity rates

- We then divide the clusters into 5 groups according to their Positivity Rate (PR)
 - $PR = 0$ | $0 < PR < 0.13$ | $0.13 < PR < 0.21$ | $0.21 < PR < 35$ | $PR > 0.35$
- We can now plot the distribution of interview cluster area pixels by LCZ class for each PR group

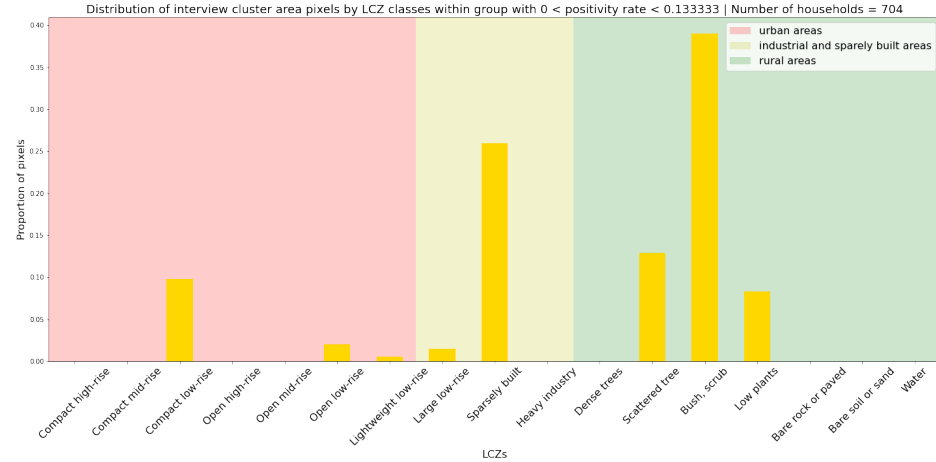
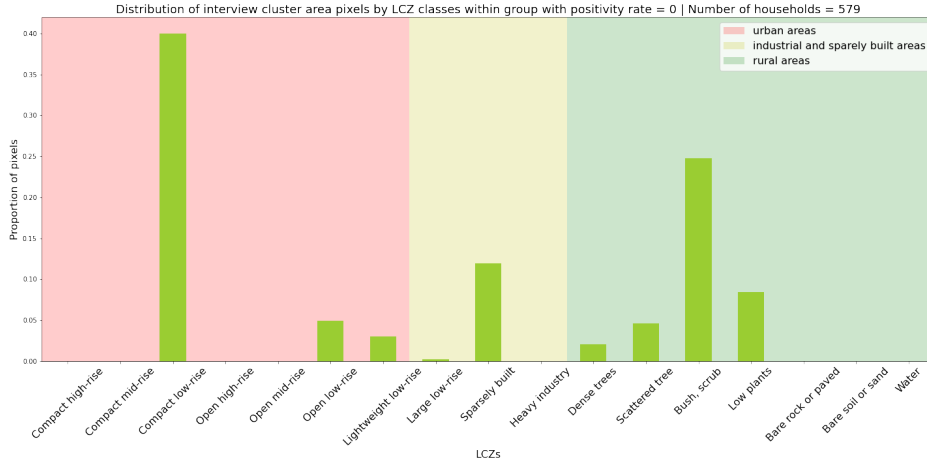
	Number of clusters	Number of households
PR = 0	44	579
0 < PR < 0.13	42	704
0.13 < PR < 0.21	48	801
0.21 < PR < 35	43	679
PR > 0.35	47	727

MIS 2017/2018 – LCZ classes in interview cluster areas



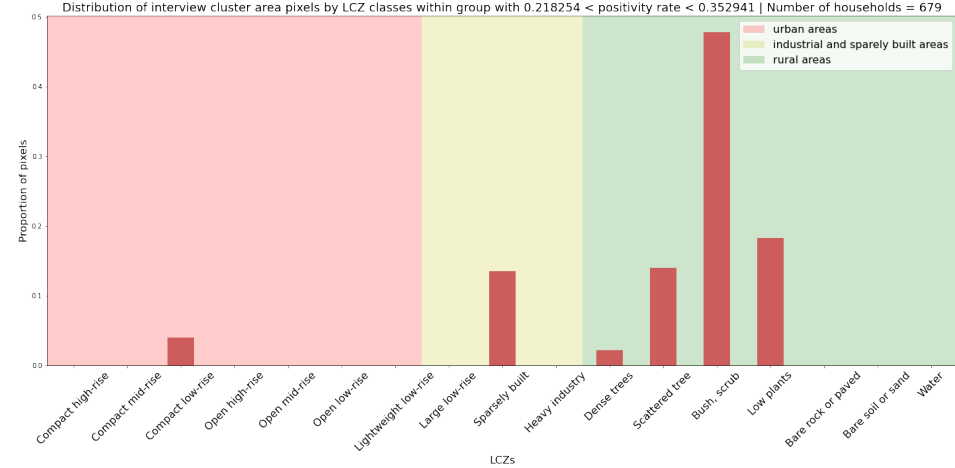
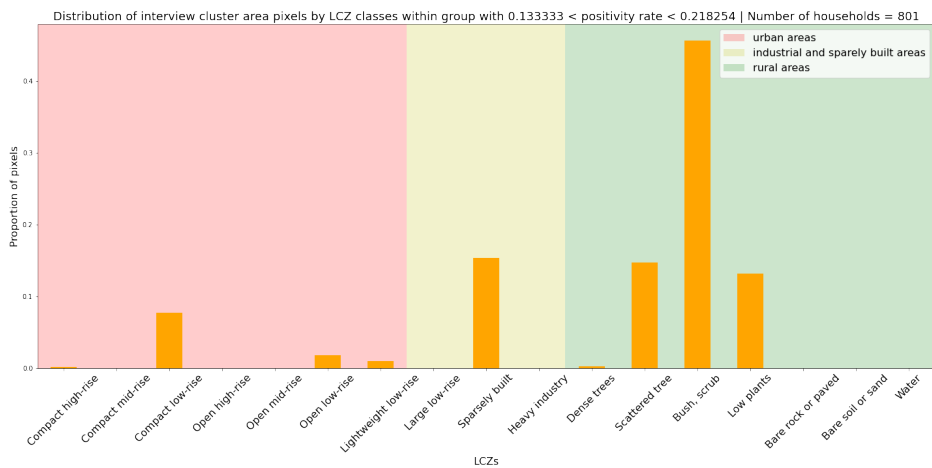
- The major part of urban pixels (compact and open low-rise) are areas with zero malaria PR.
- The major part of rural pixels (Scattered trees, Dense trees, Low plants, Bush Scrub) are areas with the highest PR.

MIS 2017/2018 – LCZ distribution by group of clusters



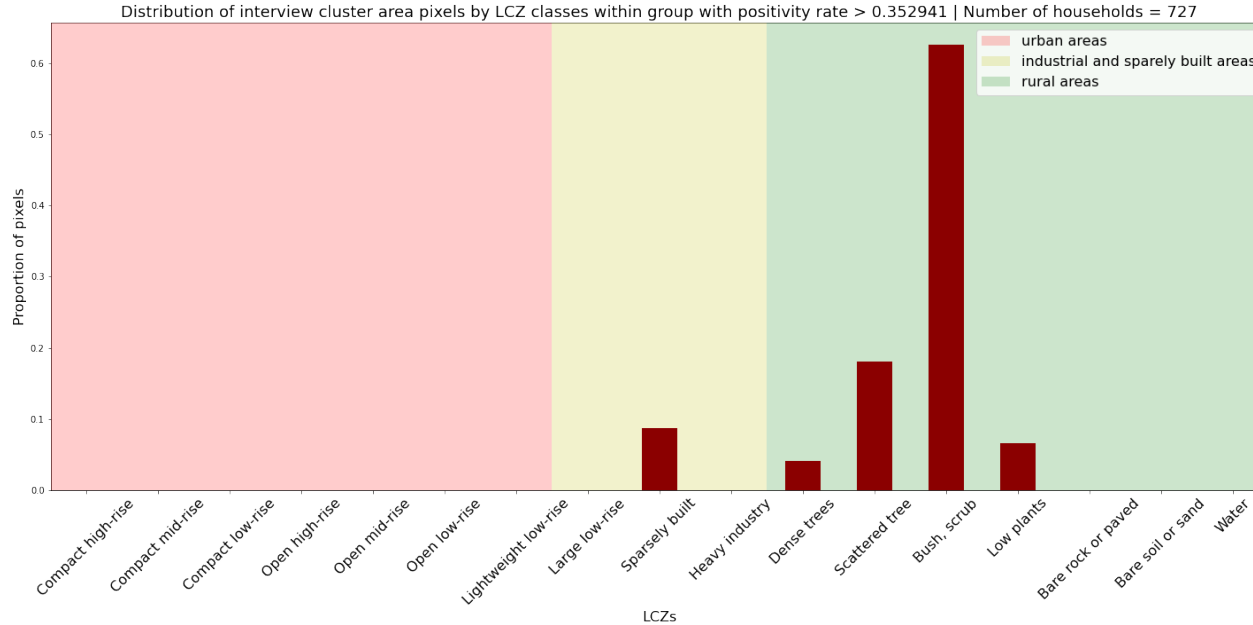
- Groups with a low PR tend to have a higher proportion of urban LCZ classes than groups with high PR
- Groups with a high PR tend to have a higher proportion of rural LCZ classes than groups with high PR

MIS 2017/2018 – LCZ distribution by group of clusters



- Groups with a low PR tend to have a higher proportion of urban LCZ classes than groups with high PR
- Groups with a high PR tend to have a higher proportion of rural LCZ classes than groups with high PR

MIS 2017/2018 – LCZ distribution by group of clusters



- Groups with a low PR tend to have a higher proportion of urban LCZ classes than groups with high PR
- Groups with a high PR tend to have a higher proportion of rural LCZ classes than groups with high PR

Conclusion

- Creation of a mapping pipeline based on Sentinel images:
 - Adapted to Burkina Faso, but can be adapted to other countries
 - Can generate maps every 5 days
- Link local environment to Malaria studies: what can we learn from LCZ ?

References

- [1] Stewart, I. D. et al. , Local climate zones for urban temperature studies. Bulletin of the American Meteorological Society, 93(12), 1879-1900, 2012
- [2] Zhu, X.X., et al., So2Sat LCZ42: A benchmark dataset for global local climate zones classification. IEEE Geosci. Remote Sens. Mag 2019
- [3] Demuzere, M., et al., LCZ Generator: a web application to create Local Climate Zone maps. Frontiers in Environmental Science 9:637455, 2021
- [4] Chen, T., et al. , A simple framework for contrastive learning of visual representations. ArXiv:2002.05709, 2020
- [5] Zhu, X.X., et al. "The urban morphology on our planet–Global perspectives from space." Remote Sensing of Environment 269 (2022): 112794.
- [6] Brousse O, et al. Can we use local climate zones for predicting malaria prevalence across sub-Saharan African cities? , Environ. Res. Lett. 15 124051, 2021

Appendices

Markov Chain and Bayes theorem

Let's define :

- I_N : the observation of the model ($\in \mathbb{R}^{32 \times 32}$) at time N .
- LCZ_N : the LCZ class $c_N \in [1, 17]$ an input patch at time N .
- $M \in \mathbb{R}^{17 \times 17}$ a matrix where $m_{i,j \in [1..17]}$ is the coefficient of M at row i and column j . $m_{i,j}$ is the probability in the first order Markov process to go from $LCZ_{N-1} = i$ to $LCZ_N = j$, $(i, j) \in [1..17]^2$
- (LCZ_N) follows a Markov process at the first order. Then, for all N :

$$P(LCZ_N = c_N | LCZ_{N-1} = c_{N-1}) = m_{c_{N-1}, c_N} * P(LCZ_{N-1} = c_{N-1})$$

According to the Bayes theorem :

$$P(LCZ_N = c_N | I_N) = P(I_N | LCZ_N = c_N) * P(LCZ_N = c_N) / P(I_N)$$

$$\Rightarrow P(LCZ_N = c_N | I_N) = \underbrace{P(I_N | LCZ_N = c_N) / P(I_N)}_{\text{Prediction scores given by the model}} * \underbrace{m_{c_{N-1}, c_N} * P(LCZ_{N-1} = c_{N-1})}_{\text{First order Markov chain term}}$$

Prediction scores
given by the model

First order Markov
chain term