Université Paris Cité

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

Basile Rousse[1,2], Sylvain Lobry[1], Géraldine Duthé[2], Laurent Wendling[1], Valérie Golaz[2,3]

1: Lipade, Université Paris Cité

2: Ined, Institut national d'études démographiques

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

Summary

Context

Context

- Impact of environmental change in demographic studies and sometimes scarce demographic data in some Sub-Saharan countries

- Demographic and Health Surveys program[1] 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

https://dhsprogram.com/

How to link remote sensing images and demography?

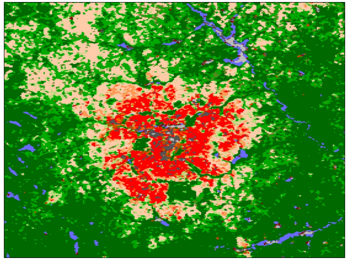
- Directly modelling demographic data using remotely sensed images.

Model

Demographic data

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

Deep Learning / NDVI ...



Demographic data

What about a complete land cover classification scheme?

About Deep Learning

Université Paris Cité

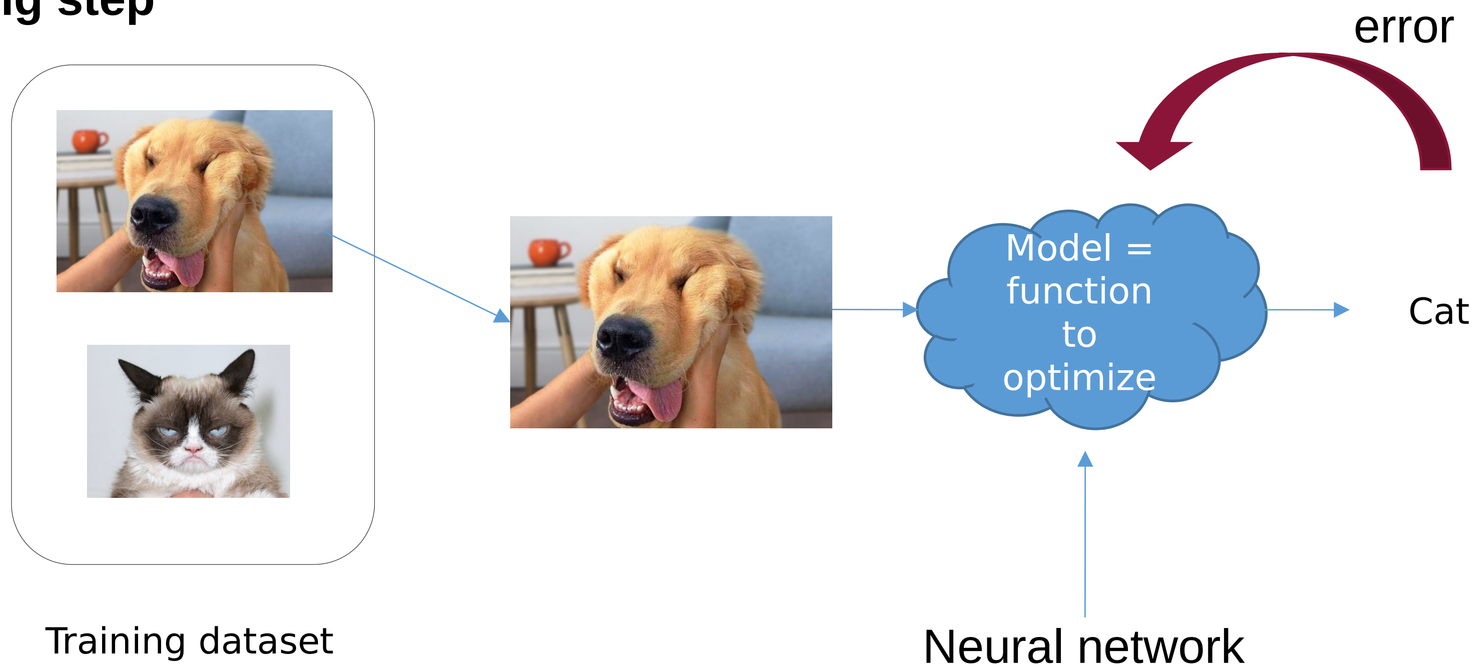
General pipeline : supervised learning

Training Step → Validation Step → Inference

Teach the model to recognize dogs and cats with **training images**

Evaluate the model on **unseen images** (different from the training images)

Training step



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

Université Paris Cité

Local Climate Zones in Sub Saharan Africa

Université Paris Cité

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)

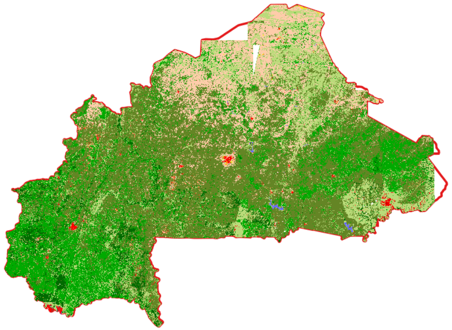
Adapting LCZ to Sub-Saharan countries : A case study of Burkina Faso

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

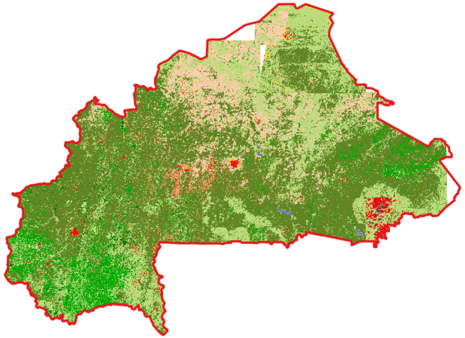
	Challenge	**Solution**
Spatial variations	Training cities are morphologically different from Ouagadougou	Train the model on morphologically **similar** cities to reduce the domain gap between training cities and Ouagadougou
Temporal variations	High variations with dry and rainy seasons not in data	Extract 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

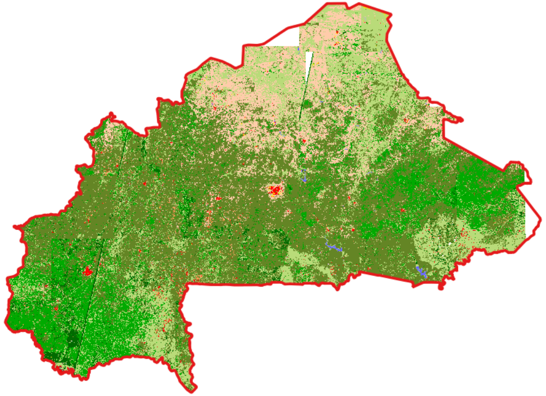


LCZ map 01/2017

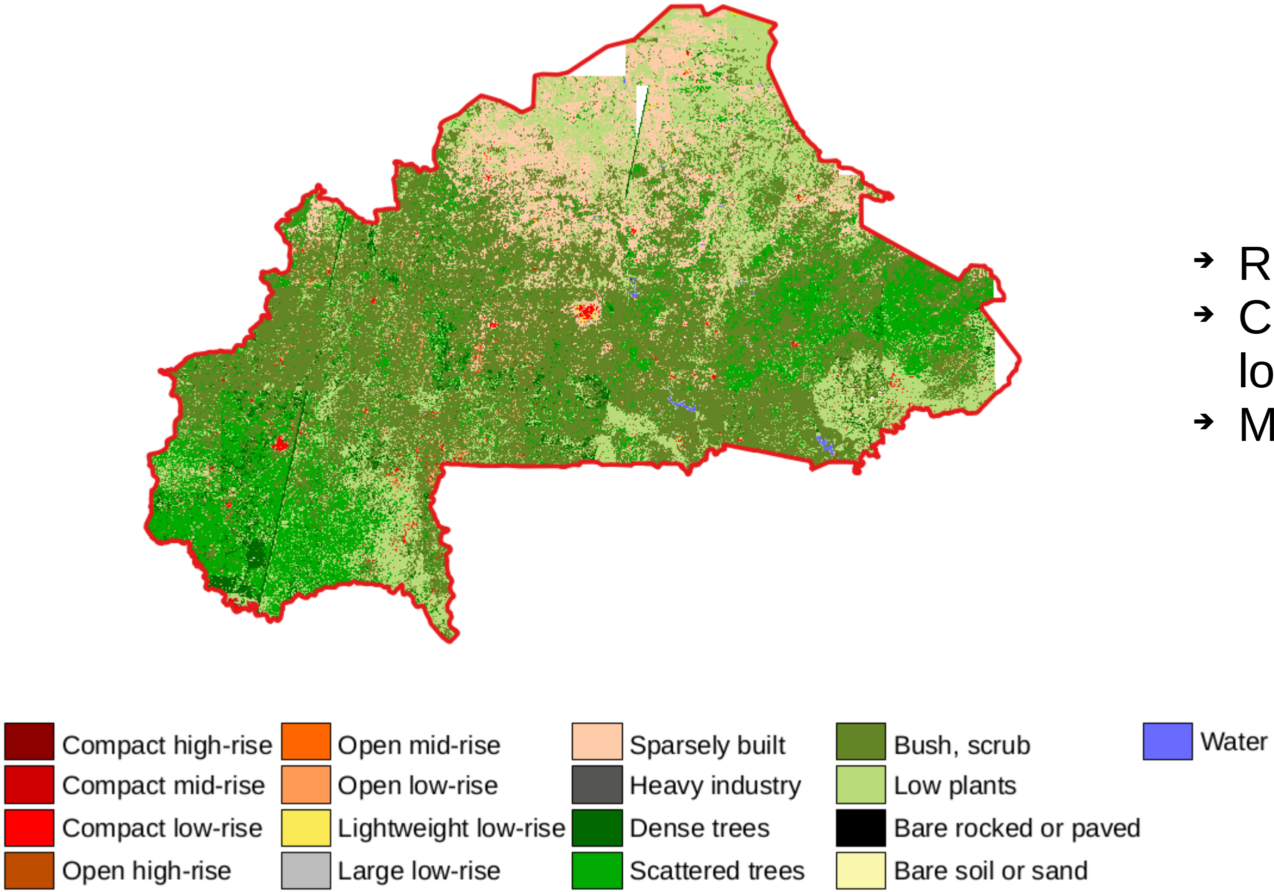


LCZ map 01/2018

Bayes’ theorem
+ Markov chain



Final results
01/2018



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

Université
Paris Cité

Local Climate Zones and demography

Université Paris Cité

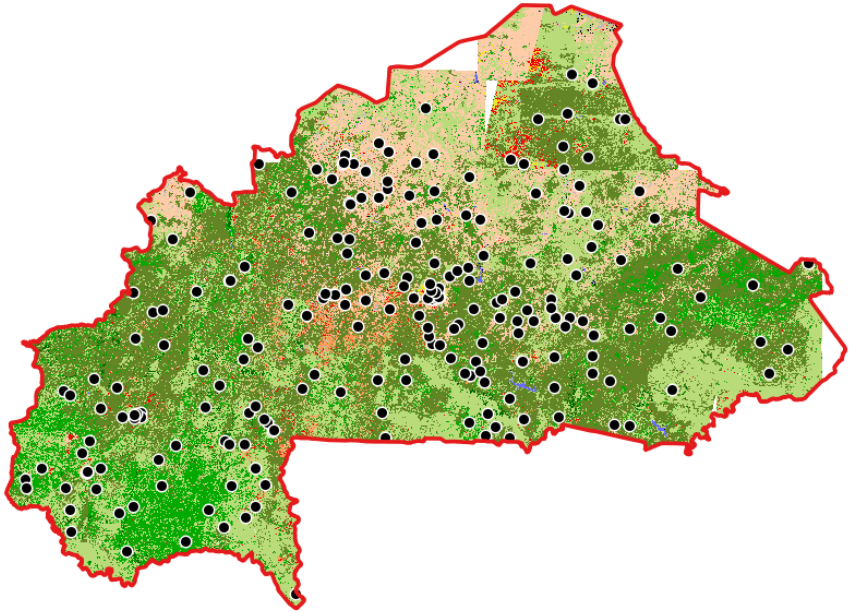
Malaria Indicator Survey 2017/2018 Burkina Faso

Objectives:

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

Informations:

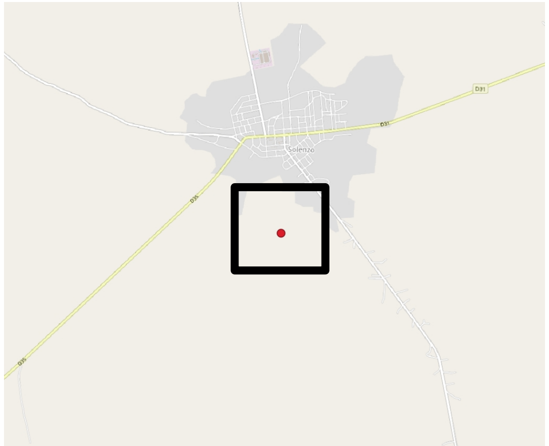
- Between November 2017 and March 2018
- 17 study areas
- 245 enumeration zones (or clusters), geolocated (<2kms for urban areas, <10kms for rural areas)
- 6322 households interviewed



⇒ 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

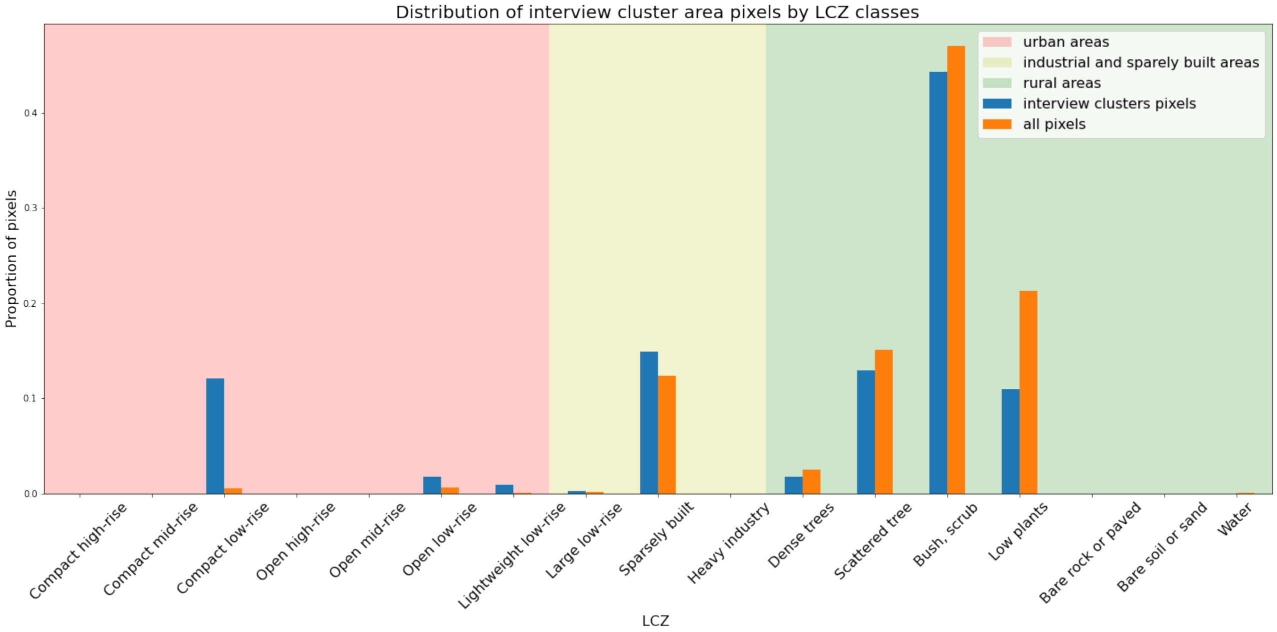
LCZ map

- 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 ?



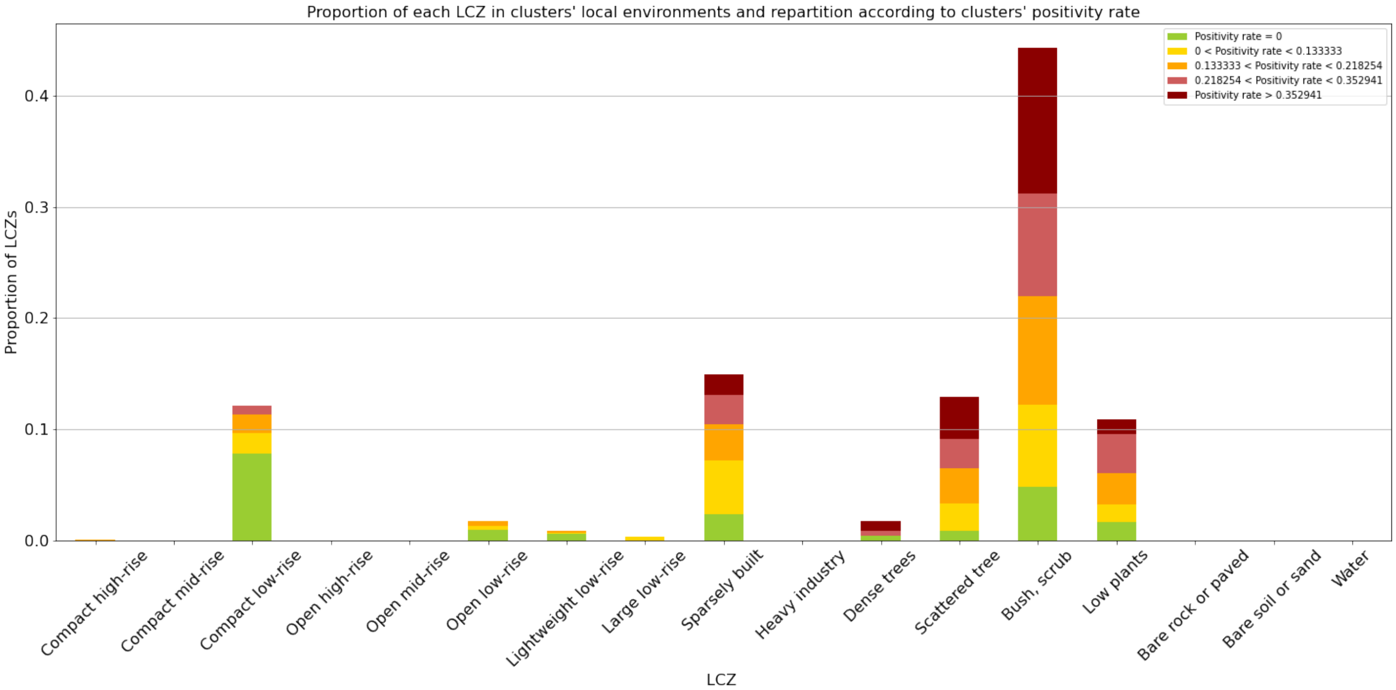
- Bush, scrub far more represented in general, as in the entire LCZ map
- Water, Bare rock and paved and bare soil in Burkina Faso but not in clusters' areas
- Very or no "high-rise" and "mid-rise": no very high buildings in Burkina Faso

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

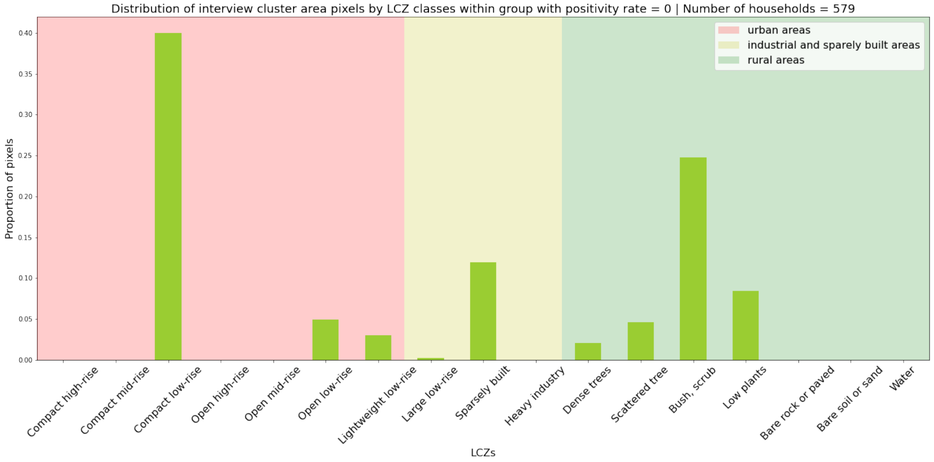
	Number of clusters	Number of households
PR = 0	44	579
0<PR<0.13	42	704
0.13<PR<0.21	48	801
0.21<PR<35	43	679
PR > 0.35	47	727

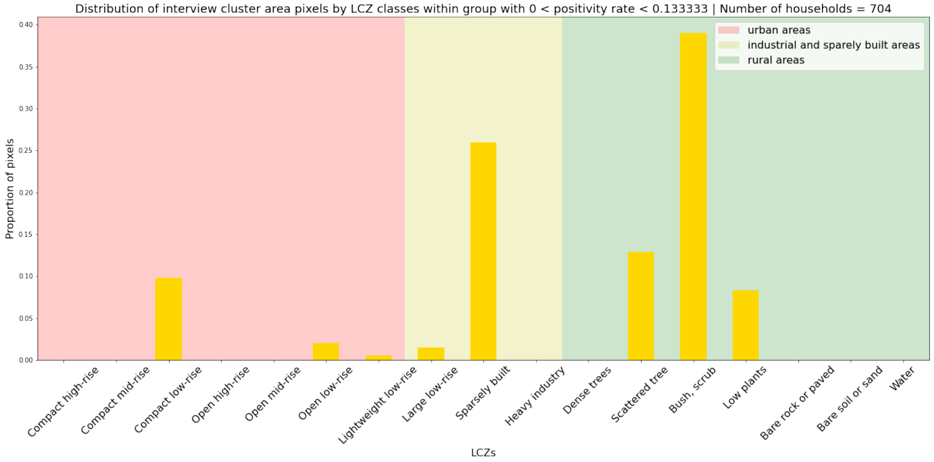
MIS 2017/2018 – LCZ classes in interview cluster areas



- 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.

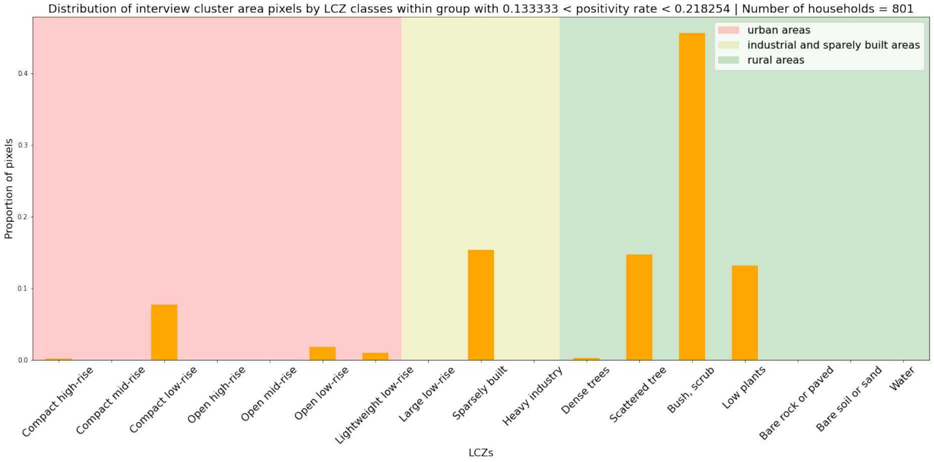
MIS 2017/2018 – LCZ distribution by group of clusters

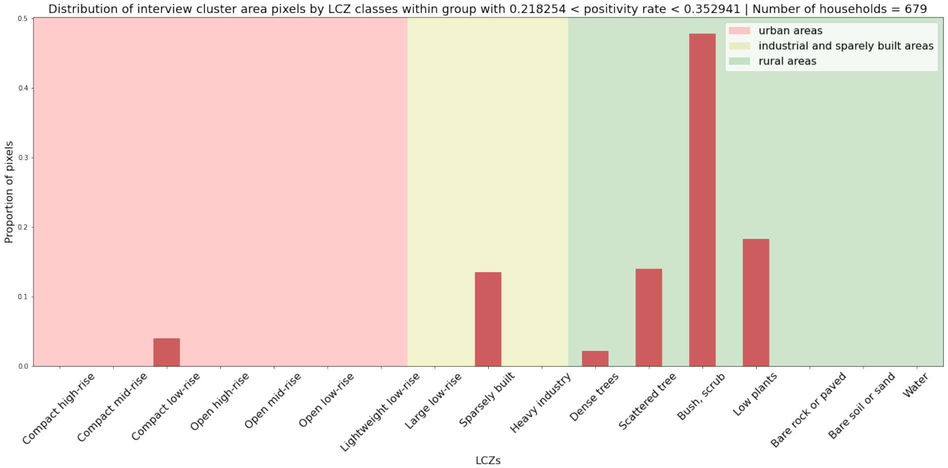




- Groups with a low PR tend to have a higher proportion of urban LCZ classes than groups with high PR
- Groups with a high PR tend to have a higher proportion of rural LCZ classes than groups with high PR

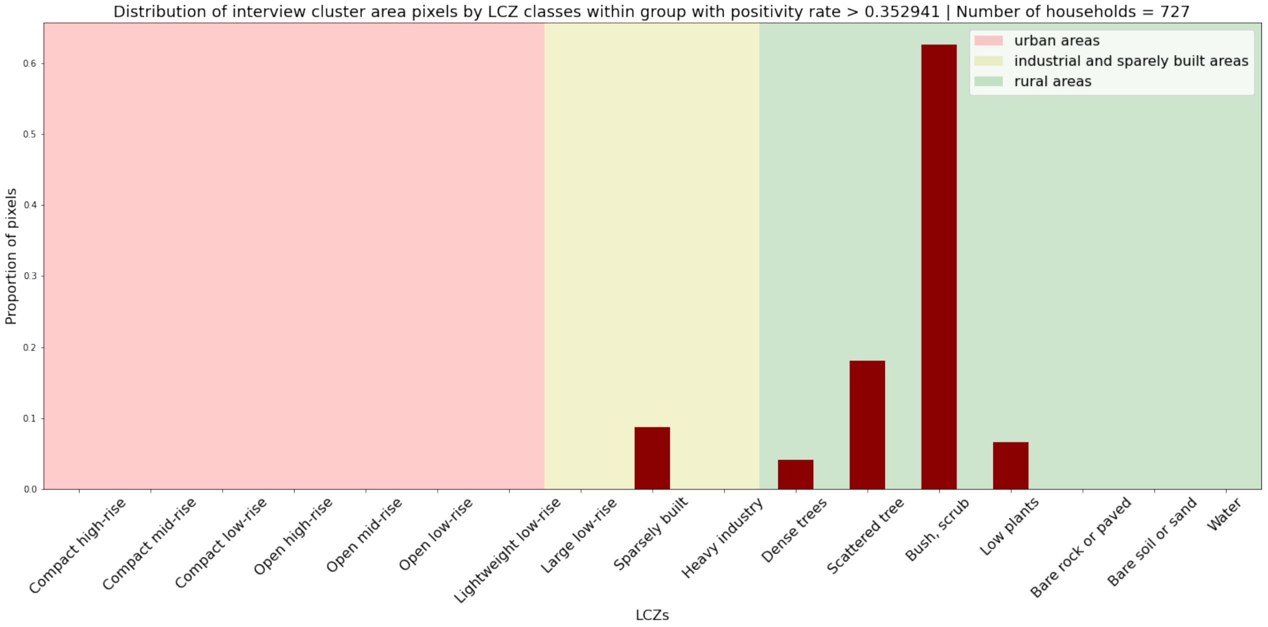
MIS 2017/2018 – LCZ distribution by group of clusters





- 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 ($\in$ R32×32) at time N .
- LCZN : the LCZ class $c_N \in [1, 17]$ an input patch at time N .
- $M \in R^{17\times17}$ a matrix where $m_{i,j \in [1..17]^2}$ is the coefficient of M at row i and column j. mi,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 :
  $P(LCZ_N = c_N \mid LCZ_{N-1} = c_{N-1}) = m_{cn-1,cn} * P(LCZ_{N-1} = c_{N-1})$

  According to the Bayes theorem :

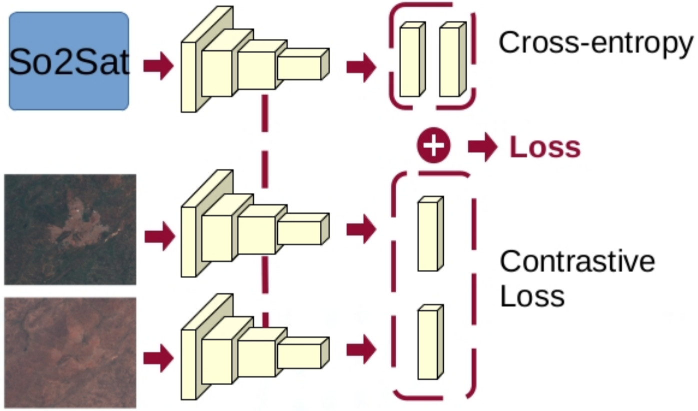
  $P(LCZ_N = c_N \mid I_N) = P(I_N \mid LCZN = c_N) * P(LCZ_N = c_N) / P(I_N)$

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

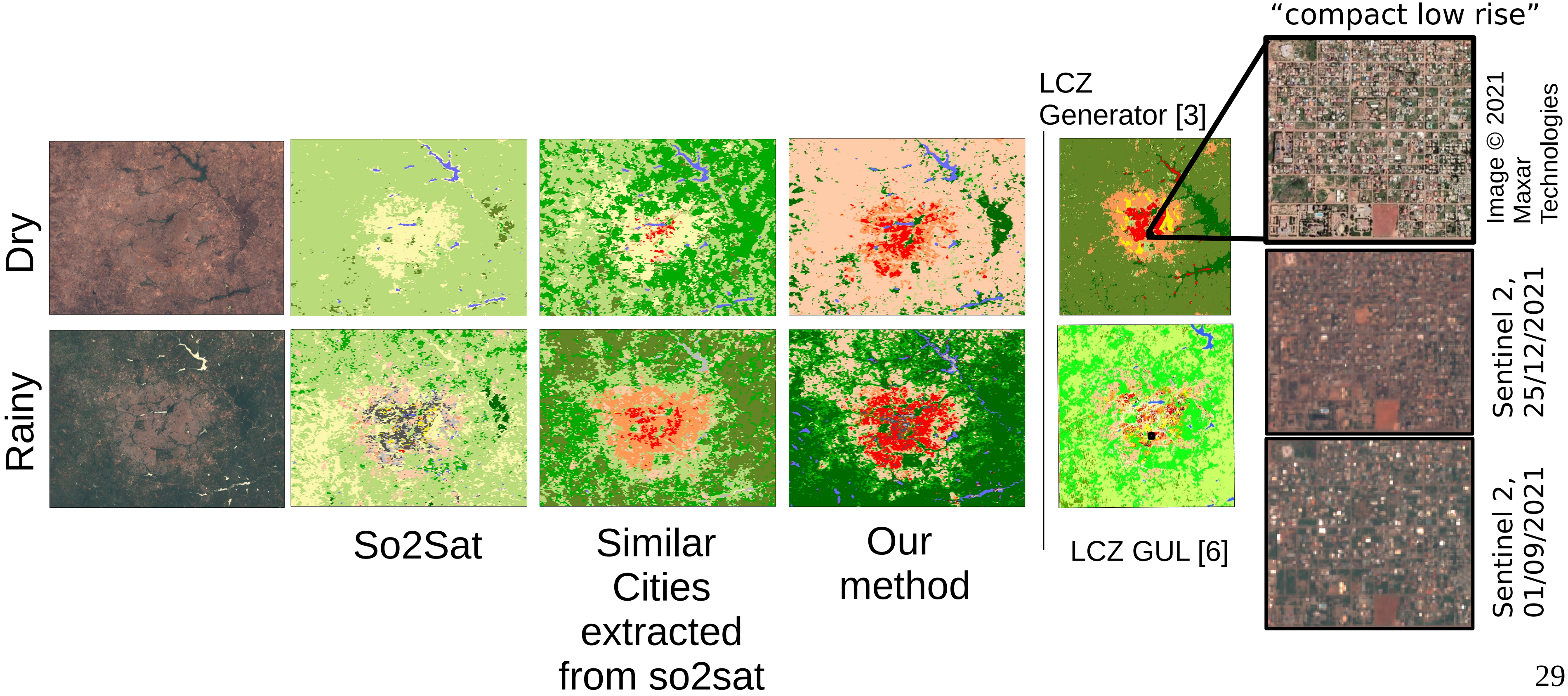
Adapting LCZ classification to Sub-Saharan countries – introducing contrastive learning

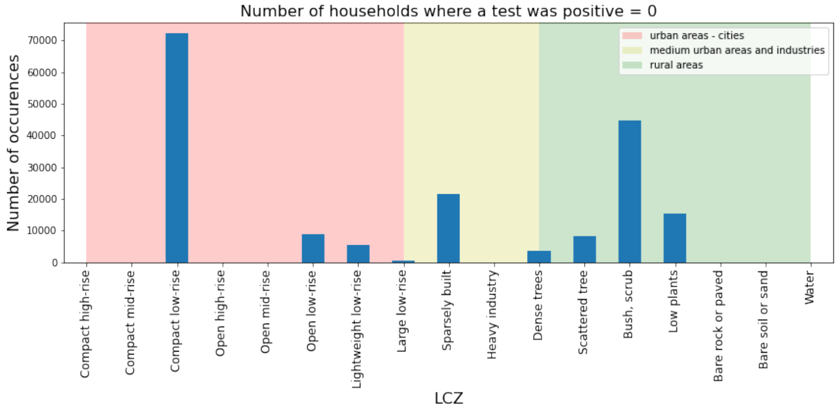
DL is very sensitive to training data ⇒ similarity between target and training areas

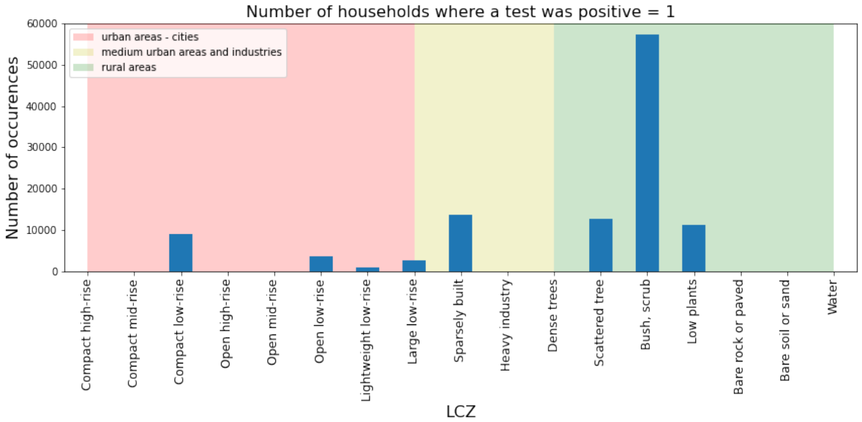
- <u>Temporal variations</u>:
  - High seasonal variations with dry and rainy seasons
  - Extract information from unlabelled images taken from both seasons (SimCLR, Chen et al. 2020),

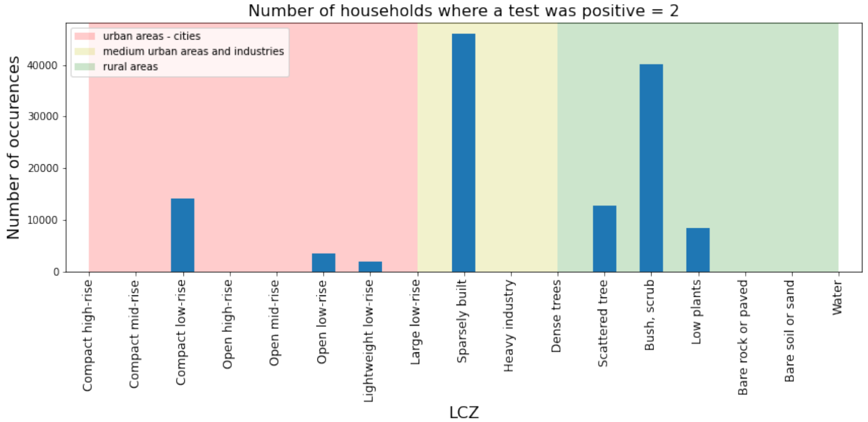


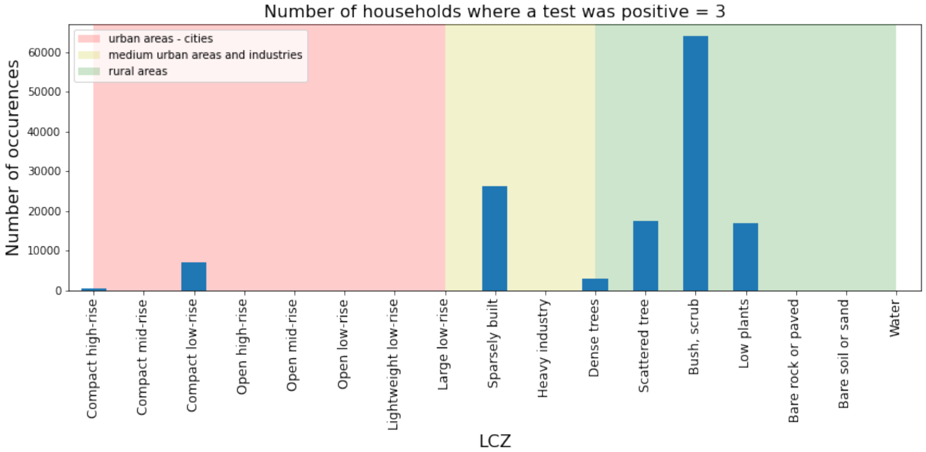
Adatpting LCZ classification to Sub-Saharan countries – different models

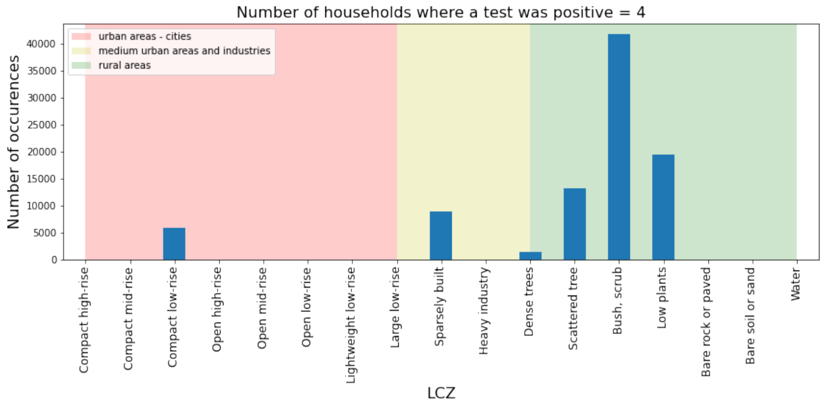


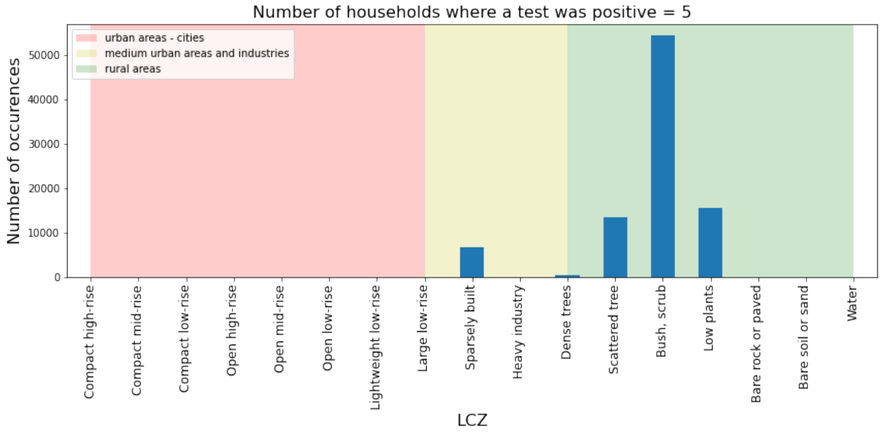

Number of households where a test was positive = 0
urban areas - cities
medium urban areas and industries
rural areas
Number of occurences
LCZ
Compact high-rise
Compact mid-rise
Compact low-rise
Open high-rise
Open mid-rise
Open low-rise
Lightweight low-rise
Large low-rise
Sparsely built
Heavy industry
Dense trees
Scattered tree
Bush, scrub
Low plants
Bare rock or paved
Bare soil or sand
Water

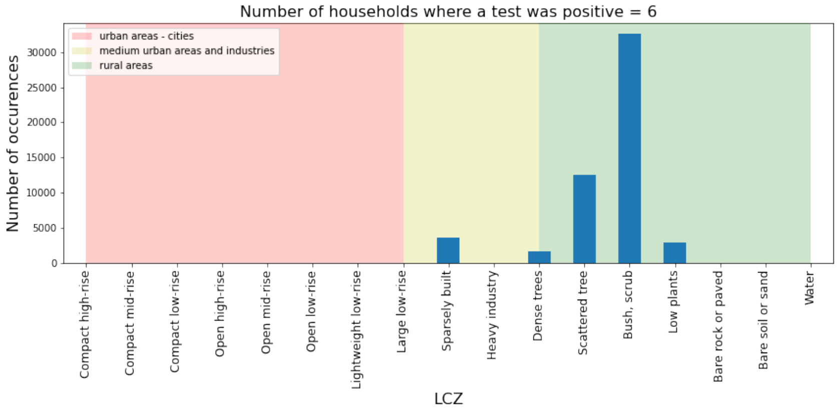

Number of households where a test was positive = 1
urban areas - cities
medium urban areas and industries
rural areas
Number of occurences
LCZ
Compact high-rise
Compact mid-rise
Compact low-rise
Open high-rise
Open mid-rise
Open low-rise
Lightweight low-rise
Large low-rise
Sparsely built
Heavy industry
Dense trees
Scattered tree
Bush, scrub
Low plants
Bare rock or paved
Bare soil or sand
Water

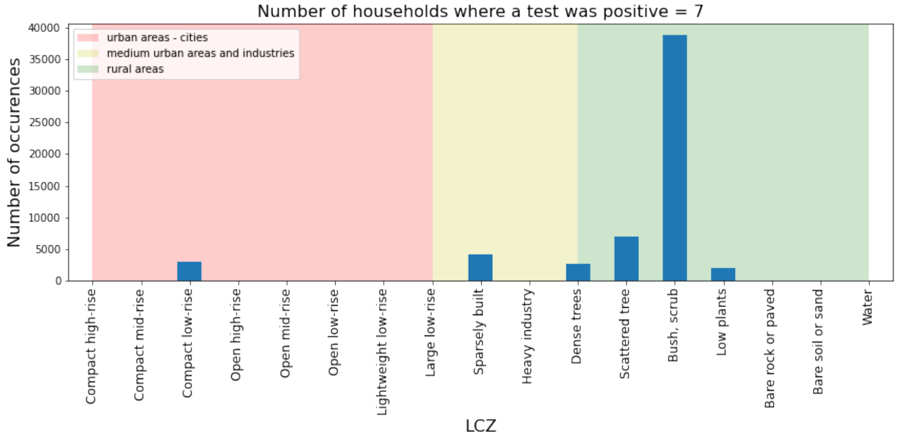

Number of households where a test was positive = 2
urban areas - cities
medium urban areas and industries
rural areas
Number of occurences
LCZ
Compact high-rise
Compact mid-rise
Compact low-rise
Open high-rise
Open mid-rise
Open low-rise
Lightweight low-rise
Large low-rise
Sparsely built
Heavy industry
Dense trees
Scattered tree
Bush, scrub
Low plants
Bare rock or paved
Bare soil or sand
Water

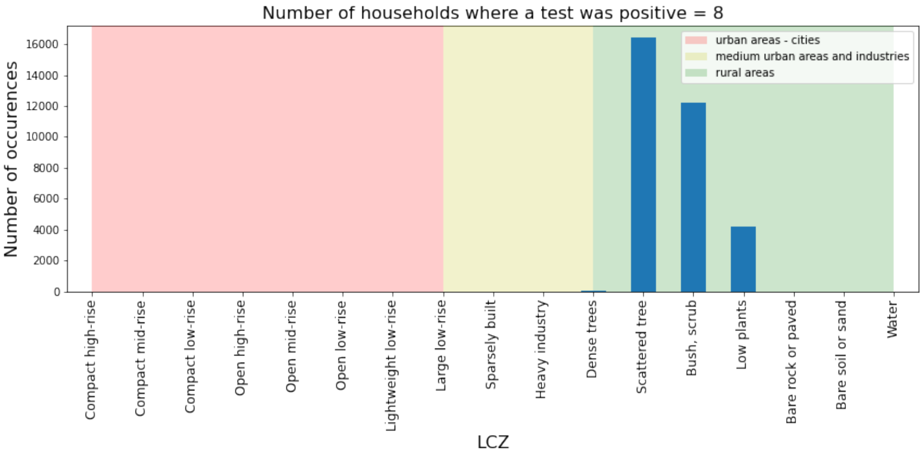

Number of households where a test was positive = 3
urban areas - cities
medium urban areas and industries
rural areas
Number of occurences
LCZ
Compact high-rise
Compact mid-rise
Compact low-rise
Open high-rise
Open mid-rise
Open low-rise
Lightweight low-rise
Large low-rise
Sparsely built
Heavy industry
Dense trees
Scattered tree
Bush, scrub
Low plants
Bare rock or paved
Bare soil or sand
Water

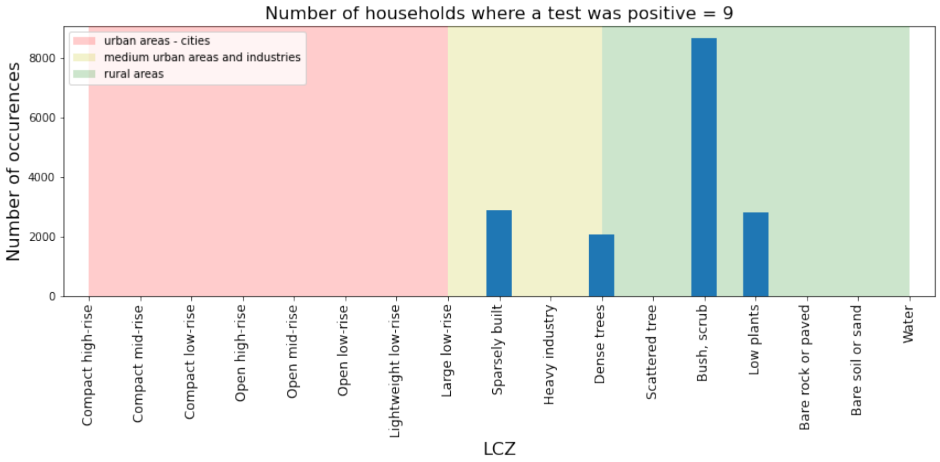

Number of households where a test was positive = 4
urban areas - cities
medium urban areas and industries
rural areas
Number of occurences
0
5000
10000
15000
20000
25000
30000
35000
40000
Compact high-rise
Compact mid-rise
Compact low-rise
Open high-rise
Open mid-rise
Open low-rise
Lightweight low-rise
Large low-rise
Sparsely built
Heavy industry
Dense trees
Scattered tree
Bush, scrub
Low plants
Bare rock or paved
Bare soil or sand
Water
LCZ

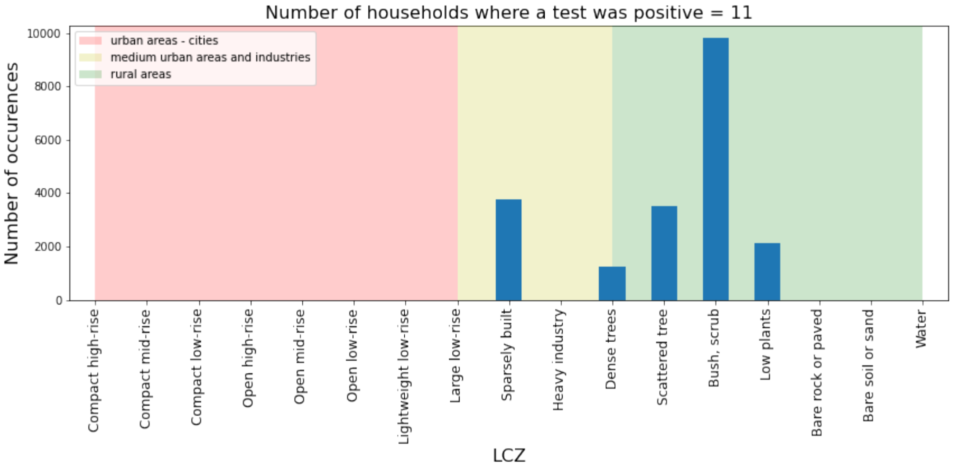

Number of households where a test was positive = 5
urban areas - cities
medium urban areas and industries
rural areas
Number of occurences
0
10000
20000
30000
40000
50000
Compact high-rise
Compact mid-rise
Compact low-rise
Open high-rise
Open mid-rise
Open low-rise
Lightweight low-rise
Large low-rise
Sparsely built
Heavy industry
Dense trees
Scattered tree
Bush, scrub
Low plants
Bare rock or paved
Bare soil or sand
Water
LCZ

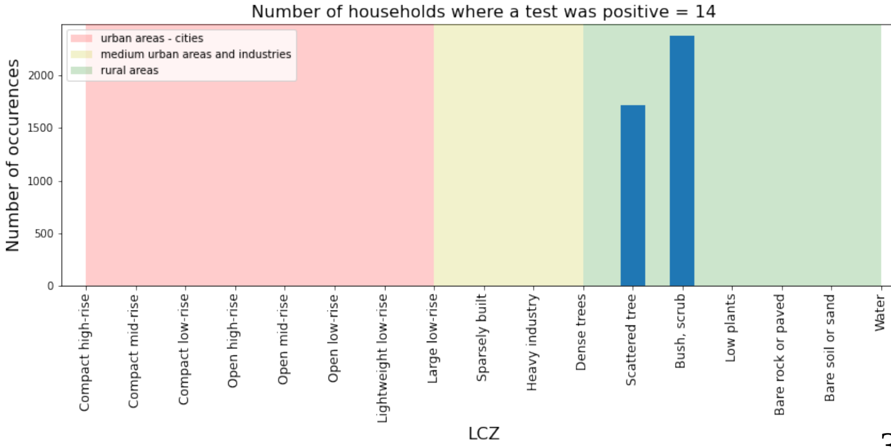

Number of households where a test was positive = 6
urban areas - cities
medium urban areas and industries
rural areas
Number of occurences
0
5000
10000
15000
20000
25000
30000
Compact high-rise
Compact mid-rise
Compact low-rise
Open high-rise
Open mid-rise
Open low-rise
Lightweight low-rise
Large low-rise
Sparsely built
Heavy industry
Dense trees
Scattered tree
Bush, scrub
Low plants
Bare rock or paved
Bare soil or sand
Water
LCZ

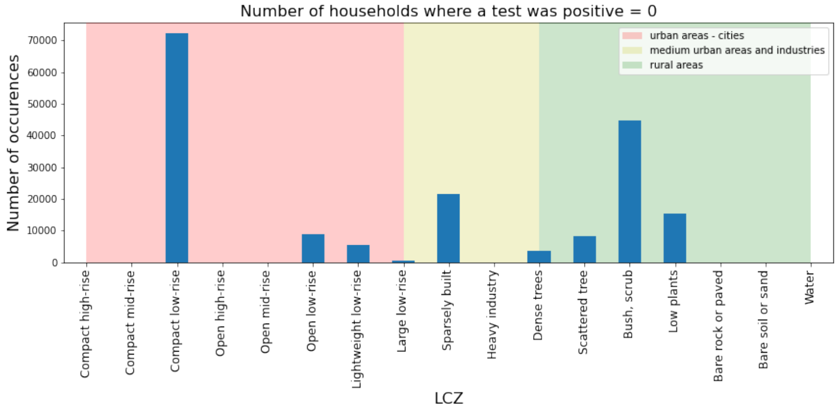

Number of households where a test was positive = 7
urban areas - cities
medium urban areas and industries
rural areas
Number of occurences
0
5000
10000
15000
20000
25000
30000
35000
40000
Compact high-rise
Compact mid-rise
Compact low-rise
Open high-rise
Open mid-rise
Open low-rise
Lightweight low-rise
Large low-rise
Sparsely built
Heavy industry
Dense trees
Scattered tree
Bush, scrub
Low plants
Bare rock or paved
Bare soil or sand
Water
LCZ

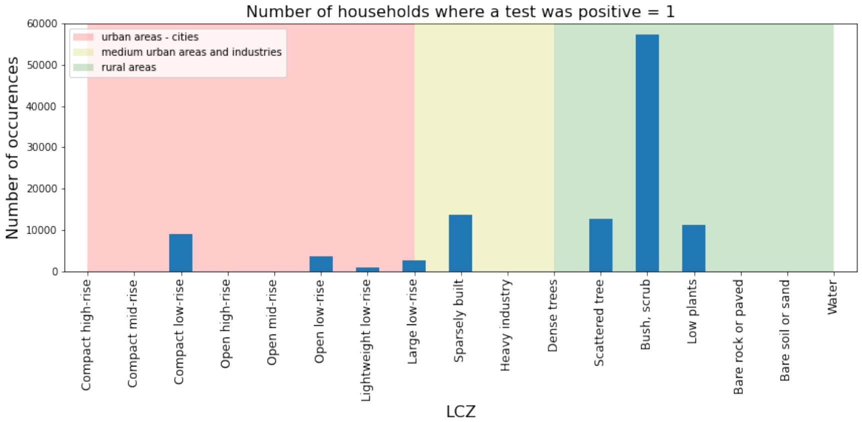

Number of households where a test was positive = 8
urban areas - cities
medium urban areas and industries
rural areas
Number of occurences
LCZ
Compact high-rise
Compact mid-rise
Compact low-rise
Open high-rise
Open mid-rise
Open low-rise
Lightweight low-rise
Large low-rise
Sparsely built
Heavy industry
Dense trees
Scattered tree
Bush, scrub
Low plants
Bare rock or paved
Bare soil or sand
Water

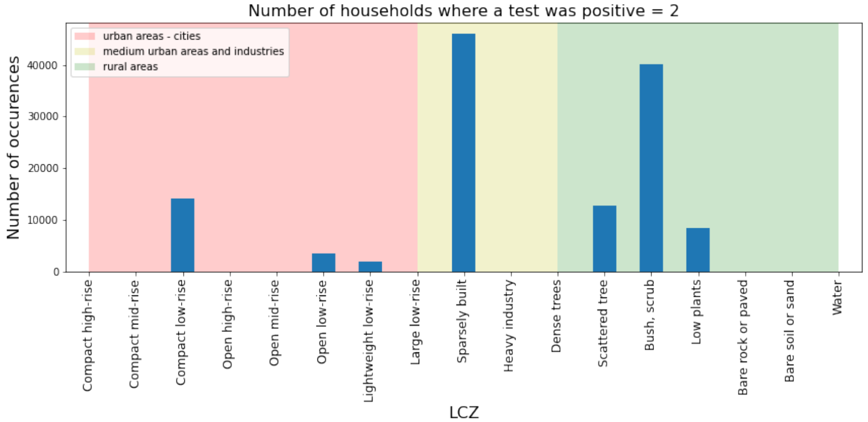

Number of households where a test was positive = 9
urban areas - cities
medium urban areas and industries
rural areas
Number of occurences
LCZ
Compact high-rise
Compact mid-rise
Compact low-rise
Open high-rise
Open mid-rise
Open low-rise
Lightweight low-rise
Large low-rise
Sparsely built
Heavy industry
Dense trees
Scattered tree
Bush, scrub
Low plants
Bare rock or paved
Bare soil or sand
Water

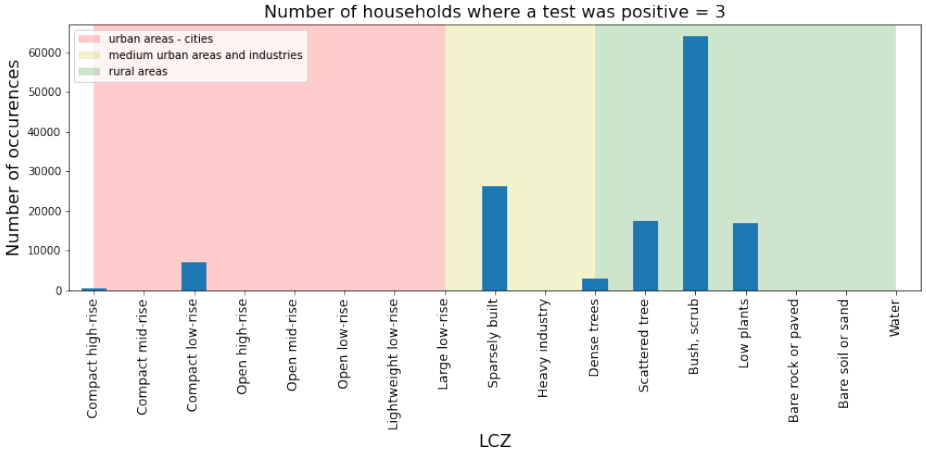

Number of households where a test was positive = 11
urban areas - cities
medium urban areas and industries
rural areas
Number of occurences
LCZ
Compact high-rise
Compact mid-rise
Compact low-rise
Open high-rise
Open mid-rise
Open low-rise
Lightweight low-rise
Large low-rise
Sparsely built
Heavy industry
Dense trees
Scattered tree
Bush, scrub
Low plants
Bare rock or paved
Bare soil or sand
Water


Number of households where a test was positive = 14
urban areas - cities
medium urban areas and industries
rural areas
Number of occurences
LCZ
Compact high-rise
Compact mid-rise
Compact low-rise
Open high-rise
Open mid-rise
Open low-rise
Lightweight low-rise
Large low-rise
Sparsely built
Heavy industry
Dense trees
Scattered tree
Bush, scrub
Low plants
Bare rock or paved
Bare soil or sand
Water


Number of households where a test was positive = 0
urban areas - cities
medium urban areas and industries
rural areas
Number of occurences
0
10000
20000
30000
40000
50000
60000
70000
Compact high-rise
Compact mid-rise
Compact low-rise
Open high-rise
Open mid-rise
Open low-rise
Lightweight low-rise
Large low-rise
Sparsely built
Heavy industry
Dense trees
Scattered tree
Bush, scrub
Low plants
Bare rock or paved
Bare soil or sand
Water
LCZ


Number of households where a test was positive = 1
urban areas - cities
medium urban areas and industries
rural areas
Number of occurences
0
10000
20000
30000
40000
50000
60000
Compact high-rise
Compact mid-rise
Compact low-rise
Open high-rise
Open mid-rise
Open low-rise
Lightweight low-rise
Large low-rise
Sparsely built
Heavy industry
Dense trees
Scattered tree
Bush, scrub
Low plants
Bare rock or paved
Bare soil or sand
Water
LCZ


Number of households where a test was positive = 2
urban areas - cities
medium urban areas and industries
rural areas
Number of occurences
0
10000
20000
30000
40000
Compact high-rise
Compact mid-rise
Compact low-rise
Open high-rise
Open mid-rise
Open low-rise
Lightweight low-rise
Large low-rise
Sparsely built
Heavy industry
Dense trees
Scattered tree
Bush, scrub
Low plants
Bare rock or paved
Bare soil or sand
Water
LCZ


Number of households where a test was positive = 3
urban areas - cities
medium urban areas and industries
rural areas
Number of occurences
0
10000
20000
30000
40000
50000
60000
Compact high-rise
Compact mid-rise
Compact low-rise
Open high-rise
Open mid-rise
Open low-rise
Lightweight low-rise
Large low-rise
Sparsely built
Heavy industry
Dense trees
Scattered tree
Bush, scrub
Low plants
Bare rock or paved
Bare soil or sand
Water
LCZ