

lczexplore : an R package to explore Local Climate Zone classifications

## Summary

Climate change is a growing concern for city planners as Urban Heat Islands has an impact on mortality [Clarke1972], health in general [Lowe2016] and consumption of energy for building cooling [Malys2012] among other effects. A first step towards large scale study of urban climate is to define classes based on logical division of the landscape, such as Local Climate Zones (LCZ) defined by [Stewart2012].

Several workflows produce LCZ classifications from different input data and approaches. For instance, the World Urban Database and Access Portal Tool (WUDAPT) produces LCZ applying random forest models on remote sensing images, as explained in [Quan2021]. Another reference workflow called Geoclimate, [Bocher2021] uses raw geographical data (OpenStreetMap and French BDTopo) to define reference spatial units on which it defines an LCZ level.

The lczexplore package aims at comparing these different LCZ classifications produced on the same area, both in a visual way, by producing agreement maps, and analytically, by computing the percentage of area on which two classifications agree and how the levels of one classification break up into the levels of another classification. This software is available as a free and open source R package.

## Statement of need

Climate change is a growing concern for city planners with a special focus on Urban Heat Island phenomena. Stewart and Oke [Stewart2012] have proposed a classification of Local Climate Zones (LCZ) that describe rural and urban areas, with 10 built types and 7 land cover types. It has been widely used in the past decade.

Several approaches classify land units into these LCZ, with at least two main approaches :

- raster approaches process remotely sensed information, and use machine learning algorithms to take into account local experts knowledge. For instance, the WUDAPT platform produces grid models of LCZ this way.[@quan2021systematic]
- vector approaches use Geographic Information System (GIS) layers that represent the main topographic features, define spatial units, compute urban canopy parameters and use them to classify spatial units into LCZ. For instance, the GeoClimate geospatial toolbox produces LCZ classifications from OpenStreetMap or french BDTopo data [@bocher2021geoclimate].

The existence of several methods to produce LCZ classifications, or the use of a method with different input data raises the question of how these different LCZ classification agree or differ. To assess this agreement several indicator can be used :

- a general agreement, measured by the percentage of the study area on which both classifications set the same levels of LCZ,
- a geographical representation of this agreement, to allow a fast assessment of the zones where the classifications differ,
- a confusion matrix, to explore how the levels of one classification break up in levels of the other classification, showned in a graphical way to help visualize the main difference between the classifications.

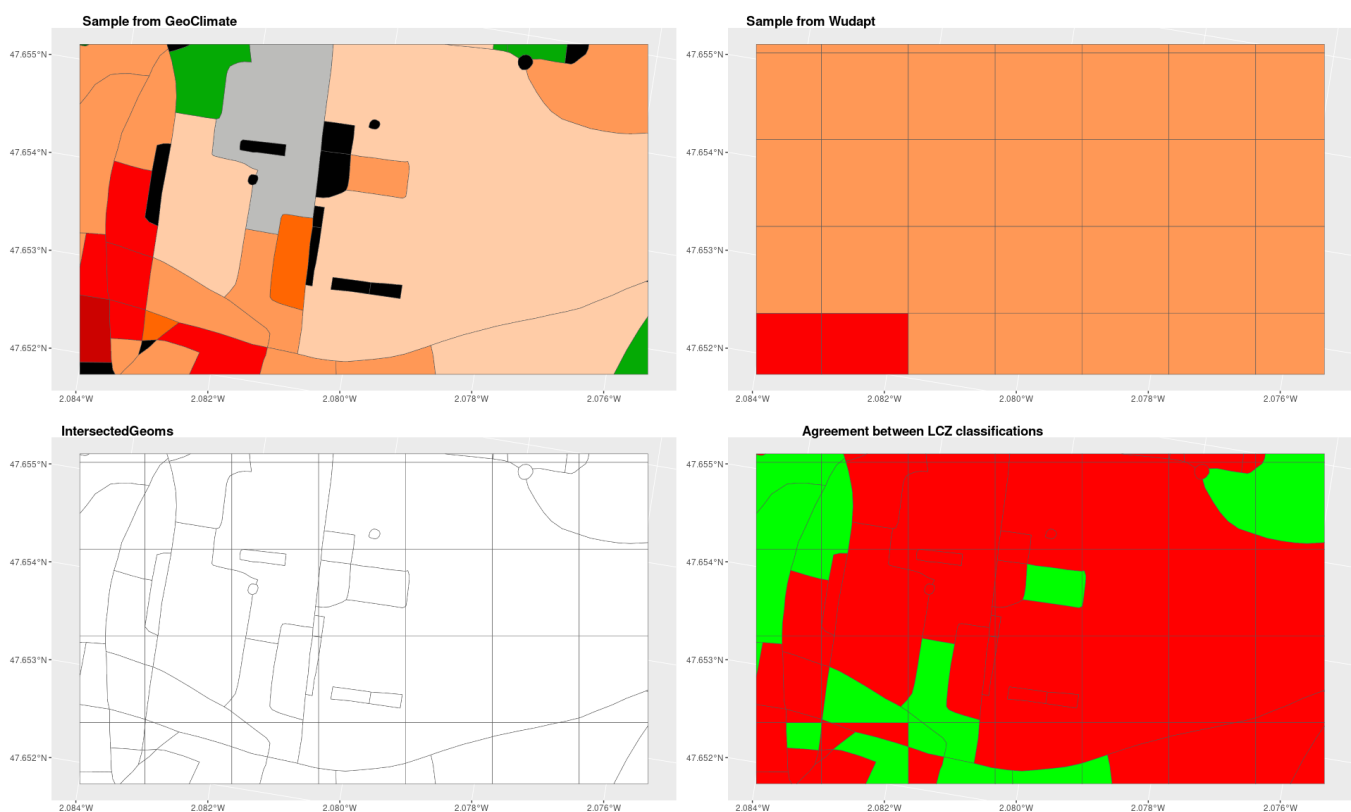
## **State of the field and feature comparison**

### **LCZ classification comparison**

A comparison of an image based approach and GIS data approach was proposed by [@muhammad2022inference], but this comparison relies on a

tile grid, and it uses several tools: the maps are produced with Q-GIS, computations are done with python scripts which don't seem to be available to the community, and the confusion matrix between different classifications were done using SAGA GIS.

The lczexplore package allows to load, show and compare LCZ classifications, even when they don't use the same spatial units to classify the area. It doesn't rely on regular tile grids, but intersects spatial units of the two classifications. On resulting units, classifications either totally agree either totally disagree, and this prevents artificial rounding effects.



{ width=100% }

All the steps of the analysis (except the production of the LCZ classification themselves) are done in the same R environment.

## Class grouping

It is sometimes useful to group some of the LCZ levels, for instance group the urban and the rural LCZ levels, or group the levels one consider as sensitive on a urban heat island point of view. The package allows the user

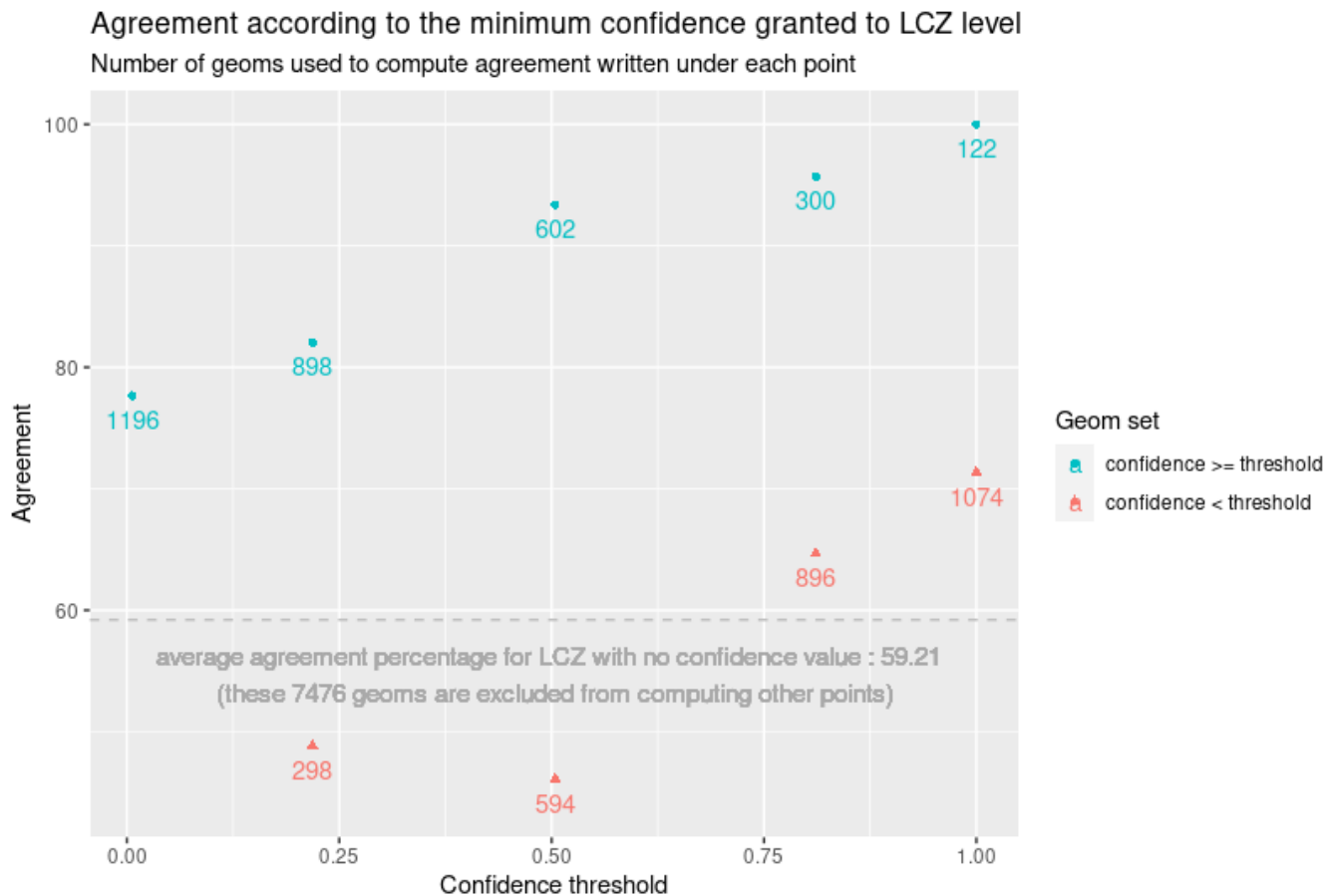
to specify the LCZ levels to group together and the colors one may use to plot them.

## **Sensitivity analysis**

In many cases, the algorithms that assign a class to a spatial unit also produces a confidence value. For instance, some algorithms compute probabilities of belonging to each class or distances to a class and assign a spatial unit to the class whose probability is the highest or the distance is the smallest, respectively. If only one class was close or likely, one can have great confidence in the assignation. If several classes were good candidates, then the confidence maybe lower.

The `lczexplore` package allows a sensitivity analysis according on this level of confidence, in order to answer the question : is the agreement between two classification higher if one excludes the spatial units on which the confidence value associated to the LCZ is smaller than a threshold ?

This sensitivity analysis is computed on the general agreement and by LCZ levels.



## Processing steps

1. First, the LCZ classifications are imported, from a geojson file or shapefiles.
2. Each LCZ classification can then be plotted
3. The two LCZ classifications are then compared :
  1. A map of the agreement/disagreement is produced
  2. The general agreement is computed
  3. The area classified in each LCZ is computed for each classification
  4. A confusion matrix is produced : how do the levels of one LCZ classification break up into the levels of the other classification.  
The area classified in each LCZ is added to the plot in order to see if disagreement between classification concerns significant areas or not.
4. Influence of the level of confidence on the agreement between classifications is performed (sensitivity analysis)

# Coding implementation

lcz explore is an R package, all its specific functions are coded in R. It relies on reference packages :

- geographical computation requires the `sf` package for vector data and the `terra` package for raster data,
- data management mainly requires the following packages : `dplyr`, `tidyr`, `forcats`, `rlang` and `methods` packages,
- graphical production uses `ggplot2`, `grDevices`, `cowplot` and `RColorBrewer` packages,
- tests need the `tinytest` package.

## A minimal example

The lcz explore package can be downloaded from [CHOISIR LE LIEN DE TELECH, GITHUB ORBISGIS, ZENODO, MON GITHUB]

You can install it in R with the command (à modifier selon le dépôt):

```
library(devtools)
devtools::install_github("MGousseff/lczexplore")
```

In this example we will use two LCZ classifications of Redon city, produced with the GeoClimate workflow, using the Open Street Map (OSM) data input and the french BDTopo data input.

```
dirPath<-paste0(system.file("extdata",package="lczexplore"),"/")
dirPathOSM<-paste0(dirPath,"osm/2022/Redon")
dirPathBDT<-paste0(dirPath,"bdtopo_2_2/Redon")
redonOSM<-
importLCZgen(dirPath=dirPathOSM,file="rsu_lcz.geojson",column =
"LCZ_PRIMARY")
redonBDT<-
importLCZgen(dirPath=dirPathBDT,file="rsu_lcz.geojson",column =
```

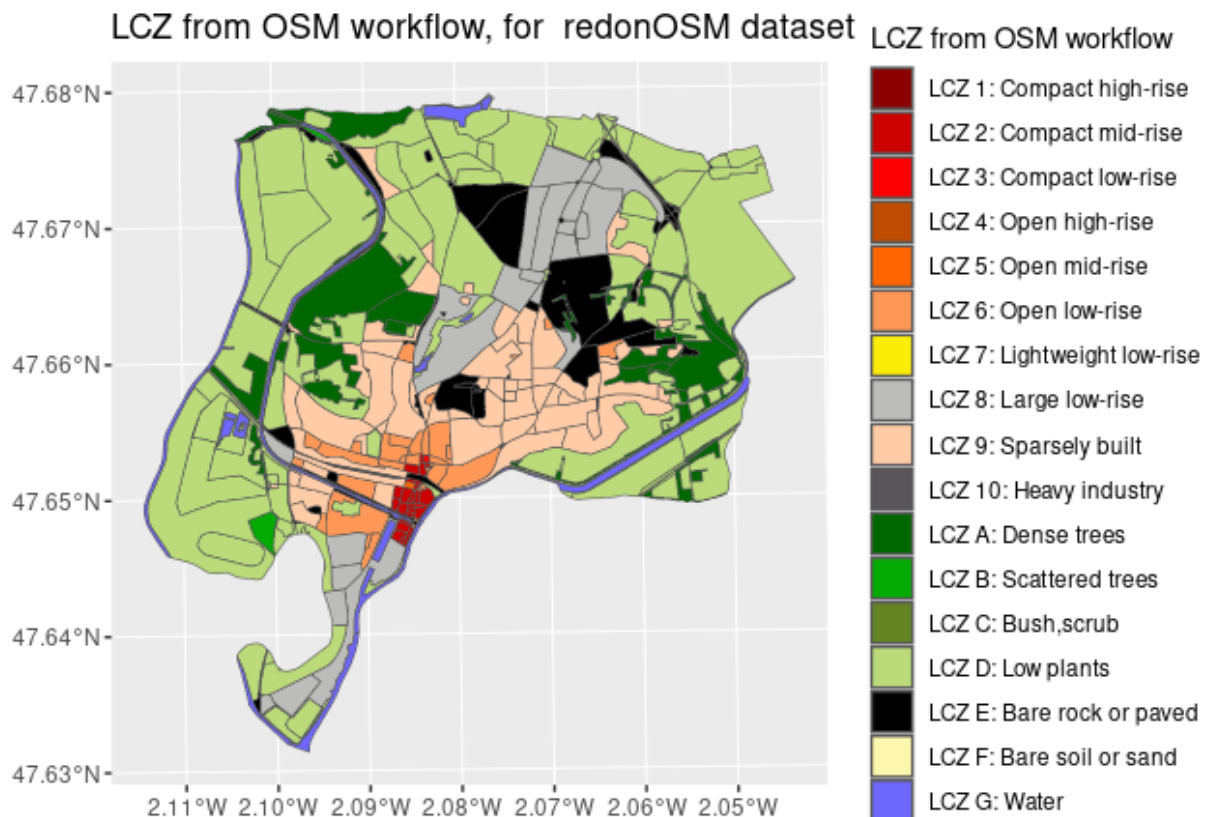
```
"LCZ_PRIMARY")
```

```
#
```

To visualize an LCZ classification, use the showLCZ function :

```
showLCZ(sf=redonOSM,wf="OSM",column="LCZ_PRIMARY",repr="brut",niveau=" ",cols=" ")
```

The result is a map of the Local Climate Zones on the area :



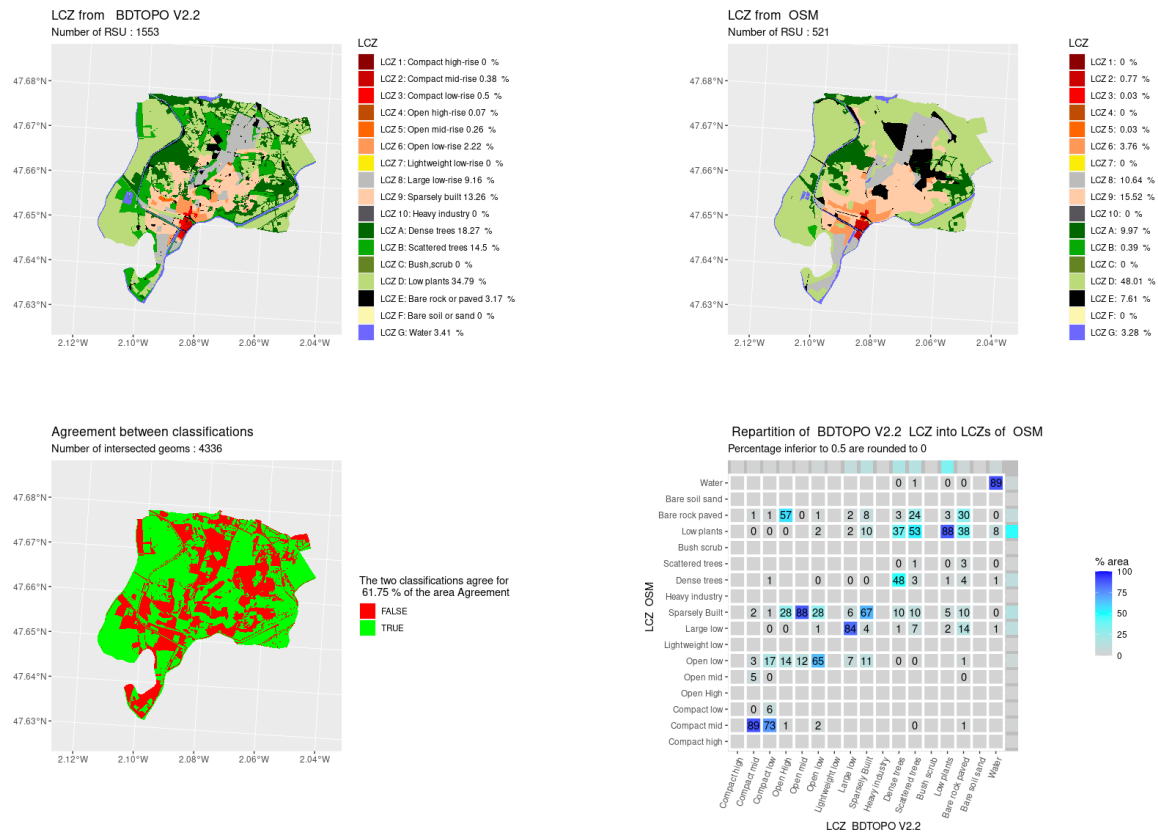
```
{ width=100% }
```

To compare the two loaded LCZ classifications, use the compareLCZ function :

```
comparaison<-compareLCZ(sf1=redonBDT,column1="LCZ_PRIMARY",  
wf1="BD TOPO v2.2",  
sf2=redonOSM,column2="LCZ_PRIMARY",wf2="Open Street  
Map", ref=1,  
repr="brut",exwrite=F,location="Redon",saveG=" ")
```

```
dirPath<-paste0(system.file("extdata",package="lczexplore"),"/")
```

The graphical results concatenate all the graphics for a quick glance :



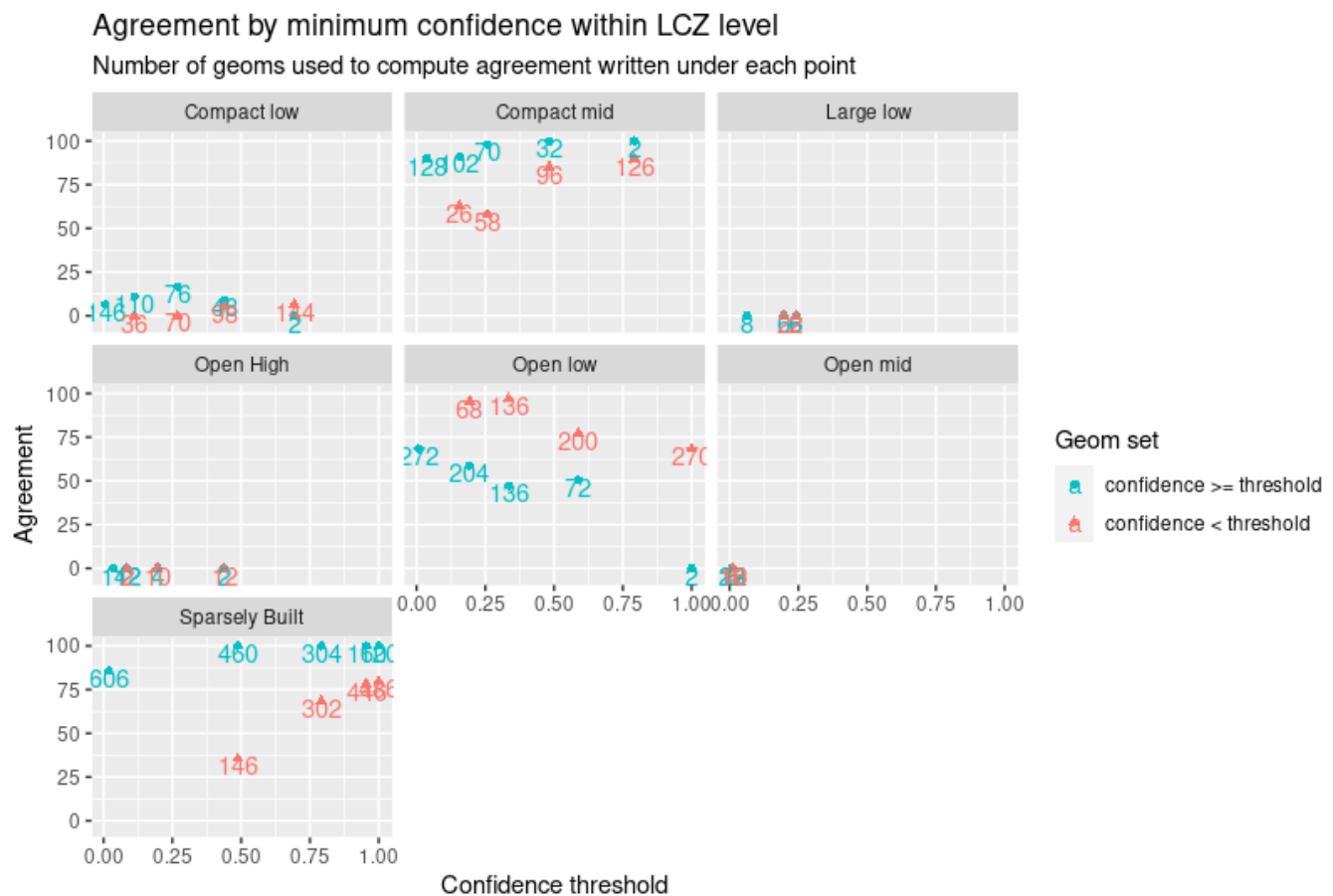
{ width=100% }

The following code produces an example of confidence sensitivity analysis :

```
mainPath<-system.file("extdata", package = "lczexplore")
testSourceFact<-
read.csv(paste0(mainPath,"/bdtopo_2_2_osm.csv"),
sep=";",header=TRUE,stringsAsFactors = TRUE)
confidSensib(inputDf=testSourceFact, filePath="",
nPoints=5, wf1="bdtopo_2_2", wf2="osm",
geomID1="ID_RSU", column1="LCZ_PRIMARY",
confid1="LCZ_UNIQUENESS_VALUE",
geomID2="ID_RSU.1",column2="LCZ_PRIMARY.1",
confid2="LCZ_UNIQUENESS_VALUE.1",
sep=";", repr="brut", plot=TRUE, saveG=mainPath)
```



It results in the following graphics, for the agreement per LCZ levels present in the datasets.



## Research projects involving GeoClimate

The lczexplore package was developed within the PAENDORA 2 project (2022-2023) funded by ADEME

## References