

swisslandstats-geopy: Python tools for the land statistics datasets from the Swiss Federal Statistical Office

Martí Bosch

Urban and Regional Planning Community, École Polytechnique Fédérale de Lausanne, Switzerland

DOI: [10.21105/joss.01511](https://doi.org/10.21105/joss.01511)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Submitted: 20 May 2019

Published: 13 September 2019

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

Summary

The Swiss Land Statistics inventory by the Swiss Federal Statistical Office (SFSO) (Swiss Federal Statistical Office, 2017) provides land use/land cover (LULC) datasets at the Swiss national extent for a sequence of four survey periods in 1979/85, 1992/97, 2004/09 and 2013/18. The data is stored in a relational database format, where each row corresponds to one of the hectometric pixels that configure the Swiss territory, and features three groups of columns:

- Firstly, the E and N columns denote the coordinates of the pixel's centroid in the LV95 coordinate reference system (or alternatively X and Y in LV03).
- Secondly, the FJ85, FJ97, FJ09 and FJ18 columns denote the exact years when the observations for each of the four survey periods were taken. For instance, the first dataset was produced between 1979 and 1985. Accordingly, for each row/pixel, the FJ85 column will denote the exact year where its LULC category attribution was made (it can be any year within the 1979/85 period, depending on the part of Switzerland).
- Thirdly, the LULC data is provided in three different nomenclatures: the *standard nomenclature*, which feature 72 categories that combine land use and land cover information; the *land cover nomenclature* and the *land use nomenclature*. Accordingly, the LULC information for each pixel is stored in columns of the form LC85_27, where LC denotes the *land cover nomenclature*, 85 the survey period 1979/85 and 27 the number of categories considered.

The inventory for each nomenclature can be downloaded as a comma-separated value (CSV) file. For instance, the *standard nomenclature* aggregated to 17 categories can be download freely, and is of the form:

E	N	FJ85	...	FJ18	AS85_17	...	AS18_17
2485500	1109700	1980	...	2012	10	...	10
2485500	1109800	1980	...	2012	10	...	10
2485500	1109900	1980	...	2012	10	...	10
2485500	1110000	1980	...	2012	10	...	10
2485500	1110100	1980	...	2012	10	...	10
2485600	1109800	1980	...	2012	7	...	7
...

While the above structure allows storing the different survey periods into a single file which can be straightforwardly read by most data analysis packages, libraries to process geographical raster data are rarely capable of processing such a format. Therefore, the aim of the proposed

library, `swisslandstats-geopy`, is to provide an extended `pandas DataFrame` interface (McKinney & others, 2010) to the table-like LULC inventory provided by the SFSO, which includes the following features:

- Read CSV files from the SFSO into `LandDataFrame` objects, which extend the conventional `pandas DataFrame` with additional attributes that store the coordinate reference system (CRS) and pixel resolution.

```
import swisslandstats
ldf = swisslandstats.read_csv('data/AREA_NOAS04_17_181029.csv')
```

- Export the categorical LULC columns into NumPy arrays (Van Der Walt, Colbert, & Varoquaux, 2011) and GeoTIFF files

```
ldf.to_ndarray('AS18_4')

array([[3, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       ...,
       [3, 3, 2, ..., 0, 0, 0],
       [3, 3, 3, ..., 0, 0, 0],
       [3, 2, 2, ..., 0, 0, 0]], dtype=uint8)
```

- Clip `LandDataFrame` instances by Shapely vector geometries (Gillies & others, 2007) or directly by region names. The latter uses `OSMnx` (Boeing, 2017) to retrieve the geometries from the [OpenStreetMap](#)

```
vaud_ldf = ldf.clip_by_nominatim(Vaud, Switzerland')
```

- Plot the categorical LULC columns as raster images (see Figure 1)

```
ax = vaud_ldf.plot('AS09R_4')
ax.tick_params(axis='x', rotation=45)
```

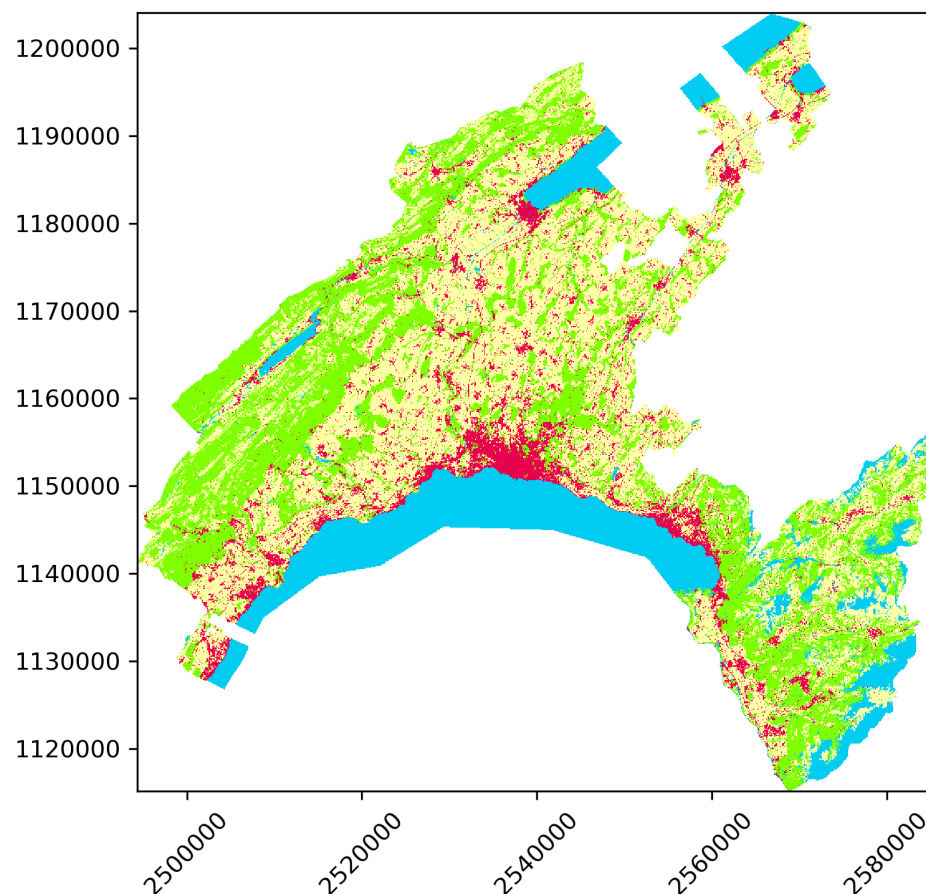


Figure 1: Plot of a categorical LULC column as a raster image.

The target audience of `swisslandstats-geopy` is researchers and developers in environmental sciences and GIS, who intend to produce repeatable and reproducible computational workflows that make use of the LULC inventory provided by the SFSO. Examples of applications of the library in the academic literature include the assessment of the carbon sequestration for the canton of Vaud (Jaligot, Chenal, & Bosch, 2019, see also [a dedicated GitHub repository](#) with the materials necessary to reproduce the results), and the evaluation of the spatio-temporal patterns of LULC change in the urban agglomerations of Zurich, Bern and Lausanne (Bosch & Chenal, 2019, see also [a dedicated GitHub repository](#) with the materials necessary to reproduce the results).

On the other hand, the `settings` module of `swisslandstats-geopy` allows changing the CRS and pixel resolution so that the library might also be used with similarly table-like raster datasets. For instance, a dataset of phenology in the Eastern US and Canadian forests for the 1984-2013 period (Melaas, Friedl, & Sulla-Menashe, 2018) has been processed seamlessly into a `LandDataFrame` instance. Nevertheless, although other table-like raster datasets such as the European land use/cover area frame statistical survey (LUCAS) (Ballin, Barcaroli, Masselli, & Scarnø, 2018) could be instantiated as `LandDataFrames`, the data could not be converted to NumPy arrays, likely because of the CRS and how their grid is sampled. Further exploration of the characteristics that a table-like raster dataset must fulfill in order to be seamlessly processed within `swisslandstats-geopy` could significantly enhance the reusability of the library.

Availability

The source code of `swisslandstats-geopy` is fully available at [a GitHub repository](#). A dedicated Python package has been created and is hosted at the [Python Package Index \(PyPI\)](#). The documentation site is hosted at [Read the Docs](#), and an example Jupyter notebook with an overview of the library's features is provided at a [dedicated GitHub repository](#), which can be executed interactively online with the SFSO free tier LULC dataset by means of the Binder web service (Bussonnier et al., 2018).

Unit tests are run within the [Travis CI](#) platform every time that new commits are pushed to the GitHub repository. Additionally, test coverage [is reported on Coveralls](#).

Acknowledgments

This research has been supported by the École Polytechnique Fédérale de Lausanne.

References

- Ballin, M., Barcaroli, G., Masselli, M., & Scarnø, M. (2018). Redesign sample for land use/cover area frame survey (lucas) 2018. Eurostat: statistical working papers. doi:[10.2785/132365](#)
- Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, 65, 126–139. doi:[10.1016/j.compenvurbsys.2017.05.004](#)
- Bosch, M., & Chenal, J. (2019). Spatiotemporal patterns of urbanization in three swiss urban agglomerations: Insights from landscape metrics, growth modes and fractal analysis. *BioRxiv*, 645549. doi:[10.1101/645549](#)
- Bussonnier, Forde, Freeman, Granger, Head, Holdgraf, Kelley, et al. (2018). Binder 2.0 - reproducible, interactive, sharable environments for science at scale. In *Proceedings of the 17th python in science conference* (pp. 113–120). doi:[10.25080/Majora-4af1f417-011](#)
- Gillies, S., & others. (2007). Shapely: Manipulation and analysis of geometric objects. [toblerity.org](#). Retrieved from <https://github.com/Toblerity/Shapely>
- Jaligot, R., Chenal, J., & Bosch, M. (2019). Assessing spatial temporal patterns of ecosystem services in switzerland. *Landscape Ecology*, 34(6), 1379–1394. doi:[10.1007/s10980-019-00850-7](#)
- McKinney, W., & others. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th python in science conference* (Vol. 445, pp. 51–56). Austin, TX.
- Melaas, E. K., Friedl, M. A., & Sulla-Menashe, D. (2018). Landsat-derived spring and autumn phenology, eastern us - canadian forests, 1984-2013. ORNL Distributed Active Archive Center. doi:[10.3334/ORNDAAC/1570](#)
- Swiss Federal Statistical Office. (2017). Statistique de la superficie selon nomenclature 2004 - standard. Available from (in French) <https://www.bfs.admin.ch/bfs/fr/home/services/geostat/geodonnees-statistique-federale/sol-utilisation-couverture/statistique-suisse-superficie/nomenclature-standard.assetdetail.4103540.html>. Accessed: 18 April 2019.
- Van Der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The numpy array: A structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2), 22. doi:[10.1109/MCSE.2011.37](#)