MovingPandas: A Python package for the analysis of movement data

Anita Graser[1, 2]

1 AIT Austrian Institute of Technology, Vienna, Austria **2** University of Salzburg, Austria

Summary & Purpose

`MovingPandas` is a package for dealing with movement data. Movement data appear in many different scientific domains, including biology, ecology, chemistry, transport and logistics, astrophysics, remote sensing, and more.

`MovingPandas` implements a Trajectory class and corresponding methods based on `GeoPand as`. A trajectory has a time-ordered series of point geometries. These points and associated attributes are stored in a GeoDataFrame. `MovingPandas` implements spatial and temporal data access and analysis functions (covered in (Graser, 2019)) as well as plotting functions.

When researchers and data analysts are working with movement data, there are a suite of data exploration activities that are often performed including:

1. Creating and visualizing trajectories from consecutive time-stamped point location records
2. Computing and visualizing movement characteristics, such as time between records, speed, and direction
3. Extracting trip start and end locations, for example to generate origin destination matrices or filter by these location
4. Clipping trajectories to an area of interest, thus limiting the geographic extent of the analysis
5. Summarizing movements using aggregations, such as flow maps

The above operations are crucial to understanding a dataset and identifying issues that may need to be addressed with further data processing when beginning an analysis. However, there isn't a tool that makes these tasks easy for users in the `Python` open source package landscape.

MovingPandas Tutorial Notebooks

In addition to detailed API documentation and example code executed by pytest, `MovingPand as` documentation includes multiple tutorial notebooks that demonstrate its functionality using case studies. These notebooks can be executed locally or online using MyBinder These tutorial notebooks document how to use `MovingPandas` to explore different types of movement datasets (including, for example, a day of vessel movements with high temporal resolution and multiple years of bird migration data with low temporal resolution).

```
( flows.hvplot(title='Generalized aggregated trajectories', geo=True, hover_cols=['weight'], line_width='weight', alpha=0.5, col
  clusters.hvplot(geo=True, color='red', size='n')
+
  GeoDataFrame(od_matrix).hvplot(title='OD flows', geo=True, tiles='OSM', hover_cols=['n'], line_width='n', alpha=0.5, frame_heig
    GeoDataFrame(matrix_nodes).hvplot(c='n', size='symbol_size', hover_cols=['cluster', 'n'], geo=True,  cmap='RdYlGn')
)
```
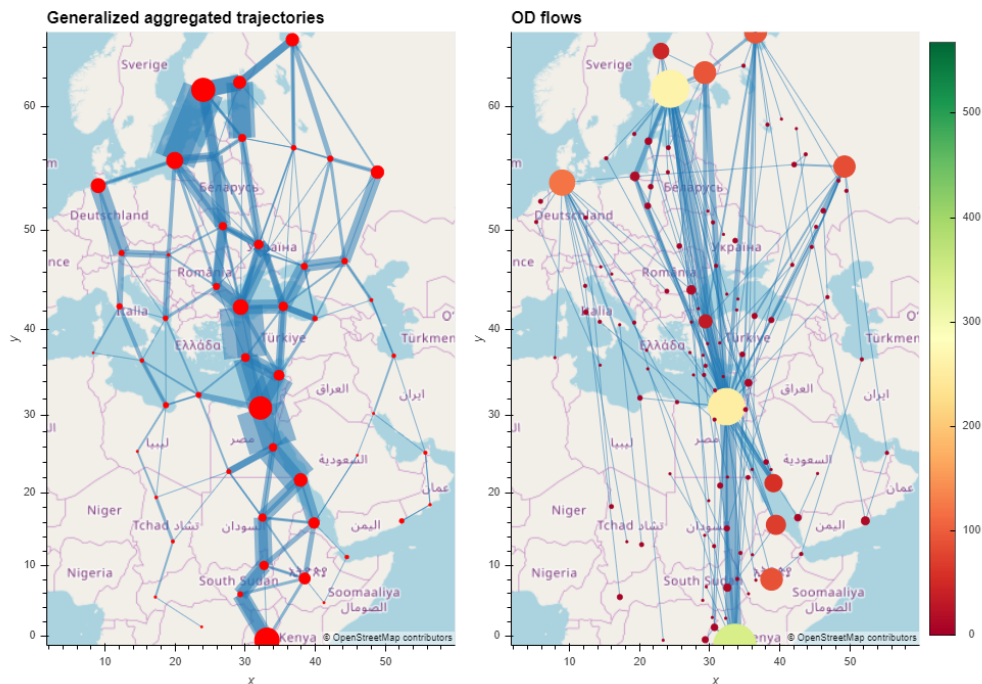

Figure 1: Example trajectory summarizations included in the interactive notebooks

MovingPandas in Context

`MovingPandas`'s goal is to make working with movment data easier for scientists and data analysist who want to use open source `Python` tools for analysis and visualization.

`MovingPandas` depends upon GeoPandas (Jordahl et al., 2019), which facilitates reading and writing movement data from common spatial data formats. It also requires `Matplotlib` for static plots and `hvplot` (Rudiger et al., 2020) for interactive plots.

To simplify dependency management and installation for non-experts, we maintain a version of `MovingPandas` on the `conda-forge` channel, which installs the system libraries upon which `MovingPandas` depends. This lowers the barrier to entry for people, particularly scientists, who are learning how to work with movement data in `Python`.

`scikit-mobility` (Pappalardo, Barlacchi, Simini, & Pellungrini, 2019) is a similar `Pyth on` package which is in an early development stage and also deals with movement data. They implement TrajectoryDataFrames and FlowDataFrames on top of Pandas instead of GeoPandas. There is little overlap in the covered use cases and implemented functionality. `MovingPandas` focuses on spatio-temporal data exploration with corresponding functions for data manipulation and analysis. `scikit-mobility` on the other hand focuses on computing human mobility metrics, generating synthetic trajectories and assessing privacy risks.

Acknowledgements

References

Graser, A. (2019). MovingPandas: Efficient structures for movement data in python. *GI_Forum Journal of Geographic Information Science 2019*, *1-2019*, 54–68. doi:10.1553/giscience2019_01_s54

Jordahl, K., Bossche, J. V. den, Wasserman, J., McBride, J., Fleischmann, M., Gerard, J., Tratner, J., et al. (2019). *Geopandas/geopandas: V0.6.2*. Zenodo. doi:10.5281/zenodo.3545747

Pappalardo, L., Barlacchi, G., Simini, F., & Pellungrini, R. (2019). Scikit-mobility: An open-source python library for human mobility analysis and simulation. *arXiv preprint arXiv:1907.07062*.

Rudiger, P., Signell, J., Bednar, J. A., Andrew, Stevens, J.-L., B, C., Samuels, J., et al. (2020). *Holoviz/hvplot: Version 0.5.2*. Zenodo. doi:10.5281/zenodo.3634720