sftraj: A central class for tracking and movement data

Mathieu Basille

2019-10-13

Signatories

The **Project team** lists the core members of the work, who will be instrumental in progress and completion of the project. Signatories are listed by alphabetical order. **Note: R packages to which people contributed are indicated within brackets.**

Project team

- **Mathieu Basille**, Assistant Professor at the University of Florida, USA [`adehabitatHS`, `hab`, `rpostgis`, `rpostgisLT`]
- **Matt Boone**, Data Scientist at the University of Florida, USA [`refsplitr`]
- **Clément Calenge**, Statistical Analyst at the *Office national de la chasse et de la faune sauvage*, France [`adehabitatMA`, `adehabitatHR`, `adehabitatHS`, `adehabitatLT`]
- **Rocío Joo**, Postdoctoral Associate at the University of Florida, USA
- **Emiel van Loon**, Assistant Professor at the University of Amsterdam, the Netherlands [`zoon`, `RNCEP`].

The Problem

Movement defined broadly plays a central role in fields as diverse as transportation, sport, ecology, music, medicine, and data science (Gudmundsson *et al.* 2012). As a matter of fact, miniaturized tracking devices have become nearly ubiquitous, and resulted in an ever-increasing volume of *tracking data* (Joo *et al.* 2019). However, there is a critical lack of standard infrastructure to deal with movement. As a matter of fact, the Movement community in R, while very dynamic, is also heavily fragmented (Joo *et al.* 2019). We have listed 57 packages that process, visualize and analyze tracking data, half of which worked in isolation, not being linked to any other tracking package (Fig.1). While the package `adehabitatLT` has gained some traction in the community, the `ltraj` class it proposes is now fairly outdated, and for instance does not play well with recent spatial classes from `sf`.

In a nutshell, this work in isolation is for large part due to a lack of a modern infrastructure to deal with trajectories in R. We recently started the work on the `sftraj` package, a project supported by the **R Consortium**, to address this gap. `sftraj` aims to address two specific objective:

1. Provide a central class and basic functions to build, handle, summarize and plot movement data;
2. Build an ecosystem of packages that rely on `sftraj` for the data part.

The first objective is the subject of our initial stage of work, i.e. the first 6 months of the project (September 23, 2019 until March 02, 2020, already supported by the **R Consortium**). This proposal builds on the technical solution provided by `sftraj` at the end of the first stage to address the second objective, and ensure global adoption of `sftraj` by the R Movement community.

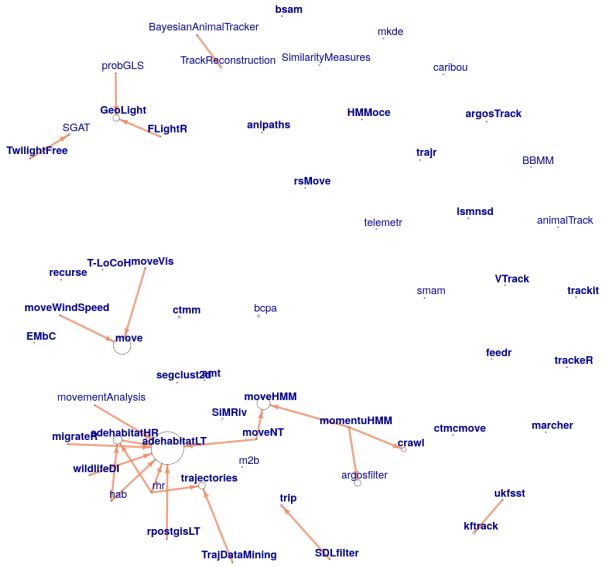

Figure 1: Network representation of the dependency between 57 tracking packages in R. Solid arrows go towards the package the others depend on. From Joo et al. 2019.

The proposal

Overview

During the first stage of the work (September 2019 until March 2020), we will work on the technical foundations of `sftraj` to deal with tracking data in R. The second stage of the work (this proposal, March–June 2020) will focus on providing solid documentation of the package and its specifications, targeting both users and package developers, which is essential for a broad adoption of the `sftraj` package.

Detail

In the first stage of the work, we will define a precise data model for trajectories. This is currently work-in-progress, but we already have a conceptual model to guide the work and the implementation in R (Fig.2). This model recognizes the duality of locations (i.e. *tracking data* in the form of geographic (x, y, z) and temporal (t) coordinates) and steps (i.e. the straight-line segment connecting two successive locations), and thus the need to provide collections of locations (a `Track` in our conceptual model) and steps (a `Trajectory`) in the implementation.

As we just have started the work on `sftraj`, we have opened a collaborative brainstorming with the R Movement community, as to gather as much feedback as possible to be able to account for all requirements for trajectory objects. All the development happens in the open, on the `sftraj` GitHub account (https://github.com/mablab/sftraj), where, in only 2 weeks, we have already collected feedback in the form of 6 use cases (https://github.com/mablab/sftraj/issues?q=label%3Ause-case).

As we work on a final implementation, it has already become clear to us that we need to consider features that were originally planned for later. For instance, the implementation of a measure of error (typically the error in GPS positions) has been requested, and might need to be implemented early in the development of

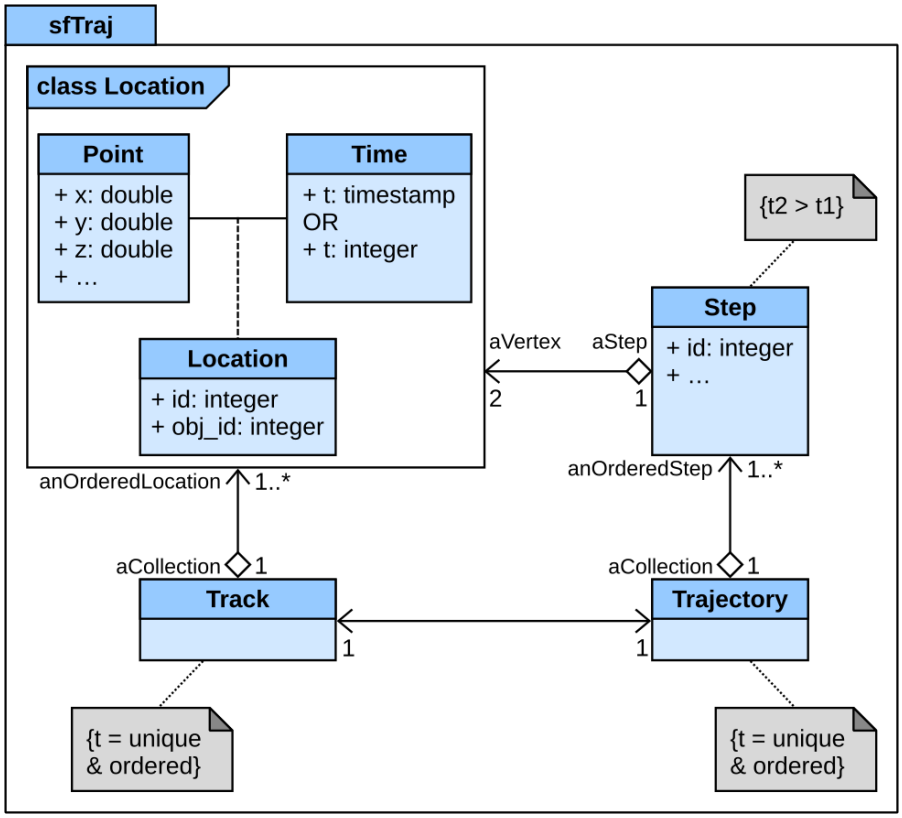

Figure 2: Proposed conceptual model for the 'sftraj' package using UML.

`sftraj`. This and other user-contributed requests, as they were considered as future work in our previous proposal, will limit the delivery of features that were only planned as optional. Among these were the addition of a vignette, and the submission of the package to CRAN (https://cran.r-project.org/).

The second stage of work will entirely be focused on documenting and detailing the internals of `sftraj` for both users and developers. The addition of a vignette and the submission of the package to CRAN will actually form the first layer of the work, which will culminate in the preparation and submission of `sftraj` to rOpenSci (https://ropensci.org/), and the preparation and submission of a manuscript detailing `sftraj` from a technical perspective to the R Journal (https://journal.r-project.org/).

Project plan

Start-up phase

The `sftraj` package is meant to be a cornerstone for the development of a more cohesive Movement community in R. The first stage of the work, which just started, will deliver solid technical foundations for trajectories in R. Our project has already been received very positively on Twitter, which we use to channel our communication. Two blog posts (https://mablab.org/post/sftraj-intro/ and https://mablab.org/post/sftraj-plan/) finally explain what we are doing, and why we need feedback from the community. Interestingly, in a very short period of time, we already received the perspective on `sftraj` from developers of 4 tracking packages (`amt`, `crawl`, `spatsoc`, and `wildlifeDI`), which comes in addition to our early discussions with 5 developers from three other tracking packages (`VTrack`, `moveHMM` and `momentuHMM`). Altogether, feedback received only

confirms the need for a central trajectory package in R.

Technical delivery

Progress will focus on the delivery of four products of increasing complexity. They will be addressed one by one, as they all build on top of each other:

- **Vignette of the package** [month 1]: Writing a vignette will provide a reproducible tutorial which will guide through the use of `sftraj`. A well-written vignette is now essential for users to get started and see the benefits of using a package.
- **Submission to CRAN** [month 1]: During the first stage of the work, `sftraj` will become full-featured and installable from GitHub. In this second stage, we will ensure that the package is ready for CRAN and finally submit it there. Publication on CRAN will allow for easy installation on any computer, as well as enabling the dependency system of R packages, which are both mandatory for broad adoption.
- **Submission to rOpenSci** [month 2]: After the package has been submitted to CRAN, we will go one step further and actually prepare it for rOpenSci. rOpenSci emphasizes (and basically enforces) a package's quality, fit, documentation, and clarity. `sftraj` would fit in the special category "geospatial data". Becoming the first official rOpenSci tracking package (none of the 57 existing tracking packages are currently on rOpenSci) would give the package an extra exposure, and guarantee its overall usefulness and usability.
- **Submission of a companion manuscript to the R Journal** [month 3–4]: Finally, as the quality of `sftraj` increases with the previous steps, we will work on a manuscript aimed at developers to provide the rationale behind development decisions for the package, and detail its implementation. In practice, this manuscript will provide package developers the necessary information for other packages to be able to rely on classes from `sftraj`.

Other aspects

While the work on `sftraj` will happen completely in the open, we will keep communicating on our progress, directly on Twitter as well as with more detailed blog posts on https://mablab.org/. We will watch for feedback, notably through the GitHub repository, which will remain constantly open.

In parallel, we will use scientific conferences (which we already planned to attend, using other funding) to introduce `sftraj` and present our work, such as useR! 2020 (which will be held in St. Louis, Missouri in July 2020), the International Statistical Ecology Conference (ISEC, which will be held in Sydney Australia, in June 2020), or Moving2Gather (which will be held in Rennes, France in March 2020).

Requirements

People

This is the same core project team than for the first stage of the work, which will lead the work for all deliverables. This stage will likely be less collaborative than the first stage, as it focuses on writing documentation and packaging the code, although all contributors will be welcome.

In particular, we request $10,000 in salary for 1.79 months (over the course of the project) of Matt Boone, Data Scientist at the University of Florida. Matt has a solid experience in R which complements that of other team members, and will be the main developer and contributor to the codebase. Having Matt fully dedicated guarantees successful completion of the project.

Processes

We will follow the same principles of openness than before, relying on a community-based code of conduct that aims to be inclusive, and a work that will happen entirely publicly, using the GitHub repository.

Tools & Tech

No technical constraint is foreseen. All team members are already equipped with enough computer power to work on the project. The development platform (GitHub) is already set up and public, and will remain open to the entire Movement community.

Funding

We request a total of $10,000 to support 1.79 months of Matt Boone, Data Scientist (Biological Scientist II) in the MabLab at the University of Florida.

- Salary: $7,369
- Fringe rate (35.70 %): $2,631
- **Total award = $10,000**

Summary

Salary to support a Data Scientist is requested to have one person committed to the project, who will dedicate set chunks of time to the work. This seems required to ensure project completion in the proposed timeline. Almost 8 weeks of funded work over the course of the project is a reasonable amount of time, which matches proposed deliverables.

Success

Definition of done

This project will be successful if `sftraj` becomes the standard for tracking data in R, both from a user and a developer perspective. Outsourcing code for low level trajectory classes from other packages to `sftraj` will be the most important outcome. In this stage of the work, success is directly related to the deliverables, namely the vignette, submission of the package to CRAN, submission of the package to rOpenSci, and submission of a companion manuscript to the R Journal by the end of the 4-month period.

Measuring success

Beyond delivering the four products in the planned time frame, success will be measured essentially from the adoption of the package:

- From a user perspective: number of downloads (e.g. using RStudio download statistics);
- From a developer perspective: adoption in tracking packages that will depend on `sftraj`.

Future work

Three axes for further development will be targeted after this stage of work:

- Provide support to developers of R tracking packages to help them develop conversion tools from their own custom classes to classes from `sftraj`.
- Dynamic visualization of trajectories, allowing keyboard- and mouse-controlled exploration of trajectories, step by step (based on the solution provided in `rpostgisLT`).
- Developing tools to clean and interpolate trajectories, based on specific filters and assumptions (e.g. maximum speed allowed, or adding missing locations by interpolation, etc.).

Key risks

The main risk is actually linked to the completion of a functional package at the end of the first stage of the work. Despite strong delay to actually start the work due to administrative difficulties, we do not foresee further delay in delivering the foundations of the `sftraj` package: our team is experienced both in developing R packages and working on tracking data, and the R Movement community is responding very positively to our invitations.

Specifically for this stage of the work, core project team members have all the expertise required for the deliverables:

- **Vignette:** Clément Calenge has produced very elaborate vignettes for each package of the `adehabitat` series. The vignette for `adehabitatLT` has notably been recognized very positively in our recent survey (Joo *et al.*, 2019), with 88.6% of respondents expressing that the documentation was either good (allowing the user to do everything they wanted and needed to do with the package) or excellent (allowing users to do even more than what they initially planned because of the excellent quality of the information).
- **Submission to CRAN:** Mathieu Basille (`rpostgis`, `rpostgisLT`), Clément Calenge (`adehabitatMA`, `adehabitatHR`, `adehabitatHS`, `adehabitatLT`) and Emiel van Loon (`zoon`, `RNCEP`) all have experience preparing and submitting R packages to CRAN.
- **Submission to rOpenSci:** Matt Boone has recently been through the process for the package `refsplitr` (https://github.com/embruna/refsplitr), for which he was the lead coder. `refsplitr` is now in the latest stages of review with rOpenSci, and will be added to their list of packages very soon. As a matter of fact, Matt even wrote a detailed blog post about the whole review process at rOpenSci (https://mablab.org/post/ropensci/).
- **Companion manuscript for the R Journal:** Mathieu Basille has worked on a similar approach for the package `rpostgis`, which is detailed in Bucklin & Basille (2018), a manuscript that was prepared for the R Journal right after the package reached a stable state and was published on CRAN.

References

- Bucklin, D., & Basille, M. (2018). rpostgis: linking R with a PostGIS spatial database. The R Journal, 10(1), 251–268. https://doi.org/c7fc
- Gudmundsson, J., Laube, P., & Wolle, T. (2011). Computational Movement Analysis. *In* Kresse W, & Danko D. M. (Eds.), Springer Handbook of Geographic Information (pp. 423–438), Springer-Verlag Berlin Heidelberg. https://dx.doi.org/10.1007/978-3-540-72680-7_22
- Joo, R., Boone, M. E., Clay, T. A., Patrick, S. C., Clusella-Trullas, S., & Basille, M. (2019). Navigating through the R packages for movement. arXiv, 1901.05935, https://arxiv.org/abs/1901.05935