Track Segmentation: Stops and Movements

David C. Douglas, ddouglas@usgs.gov
July 2025, R version 4.4.1

This R script inputs animal tracking data and segments the time series for each individual into sequential periods of stops and movements. The algorithm identifies stops based on two influential user-defined thresholds: 1) a **distance threshold** (hereafter **Dist**) that defines the upper limit of spatial displacement permissible while occupying a stop, and 2) a **duration threshold** (hereafter **Dur**) that defines the minimum amount of time a stop is required to persist. Values for the **Dist** and **Dur** should be chosen in a manner that targets animal behavior(s) of interest while also accommodating the spatial accuracy and temporal frequency of the location data under analysis. Locations not classified as stops are classified as movements.

Input data:

The script is designed to input tracking data from either Movebank (www.movebank.org) or tabular comma-separated (CSV) files. Example scripts for both types of data input are provided. Six variables are required in the input data: animal_id, timestamp, latitude, longitude, lc (location class index), and species; the latter two variables can be defined globally when convenient. To input data from Movebank (direct download), your Movebank credentials (username/password) must be defined in the script, your username must have read access to the Study's data, and if applicable, you have accepted <u>in advance</u> any license terms for accessing data in the respective Movebank Study.

Input data in tabular CSV format must include (or be coerced to include) the following six variables with exactly these variable names (case sensitive):

[variable name]	[variable type]	[variable description]
animal_id	character	unique ID for each individual
timestamp	POSIXct date and time	timestamp in R format
latitude	floating point (-90 to 90)	WGS84 latitude
longitude	floating point (-180 to 180)	WGS84 longitude
lc*	character	location class (see below)
species	character	species or common name

*The location class refers to a 1-character variable that, in the case of Argos Doppler location data, is the Argos Location Class (with values '3', '2', '1', '0', 'A', 'B', or 'Z'). In the case of GPS location data, the location class should be coded as 'G'. If location class is unknown or undefined, any other character, such as 'U', could be defined for all records.

Quick start instructions:

1. Decompress the distribution software bundle into a directory (folder) of your choice. There are eight files in the bundle: three R scripts with names that begin with "_run" that you will need to modify prior to running; three R scripts with names that begin with "core" that remain unchanged; and two CSV files that contain example input data.

2. If you have tracking data in CSV format, modify either of the "_run CSV_input_example" R scripts to meet your needs. Read the internal documentation that precedes each line of code to make mandatory changes or to customize to your specifications.
3. Each of "_run" scripts are preconfigured to use example data. The only changes to the example scripts (as distributed) are to modify the 2 directory path variables to match your directories, and, if running the "Movebank example", to change the 2 variables that define your Movebank Username and Password.

Algorithm method part 1 - Identifying stops:

The user defines two thresholds: a maximum allowable distance (**Dist**, units=meters) among all pairwise distances between locations that constitute a stop, and minimum required duration (**Dur**, units=hours) that a stop must persist. The algorithm steps chronologically through the tracking time series, one animal at a time and identifies sequences of locations (i.e., periods of time) when the animal was stopped (or conversely, not stopped). Locations that constitute a stop will: 1) all be within **Dist** from one another (i.e., distances between all pairwise locations will be less than **Dist**); and 2) the elapsed time between the first and last location will be greater than **Dur** (hence, stops must be comprised of at least two locations).

In practice, the algorithm begins with the first animal location (an anchor location) and calculates the distance and elapsed time to the next location. If the distance is >**Dist**, then no stop is detected and the next location becomes the anchor and the procedure repeats. If the distance is <**Dist** and the elapsed time is <**Dur** then a "potential stop" is detected and the next location is considered. If all three pairwise distances between the three locations are <**Dist** but the elapsed time from the first to the third location is still <**Dur**, then the stop remains "potential" and the next location is considered. Locations are sequentially added to the "potential" stop until such time that: 1) all pairwise distances among the locations are <**Dist** and the elapsed time from the first to last location is >**Dur**, in which case a stop is detected; or 2) one or more pairwise distances are >**Dist** in which case no stop is detected, and the last location becomes the anchor location, and the steps start over. In the case of the former (a stop was detected), sequential locations are added one at a time until such time that either one or more pairwise distances were >**Dist**, at which point the stop is concluded with the prior location, and the last (most recently added) location becomes the anchor location, and the steps start over. After all locations for a given animal have been processed, a single location for each detected stop is calculated as a weighted average of the respective stop's locations. The weighting gives strong emphasis on GPS locations, followed by emphasis on higher quality Argos locations compared to lower quality Argos locations. Derived stop locations are assigned a timestamp equivalent to the respective stop's initial (first) tracking location. Note that "stop locations" are derived averages of the tracking location data (i.e., they are not observed data).

Algorithm method part 2 - Identifying metastops:

Metastops are one or more chronologically sequential stops in which pairwise distances between the stop locations are all <**Dist**. The script parses each animal's time series of stop locations (and their timestamps) through the same distance-detection steps described above. This may seem counterintuitive but depending on one's **Dist** and **Dur** thresholds and the weighted-averaging process, consecutive discrete stops can be <**Dist** apart. The metastop process, then, works to combine these adjacent stops in ways that satisfy the original intention of the user, while also preserving the identity of each unique stop for the sake of clarity and process. To ensure output of the metastops analysis is comprehensive, metastops are allowed to have as few as one stop (in which case the metastop and the single stop have the same attributes). When a metastop comprises more than one stop, the metastop's location is made equivalent to that of the stop with the longest duration (i.e., where the animal spent the most time). The metastop's timestamp is made equivalent to that of the first stop.

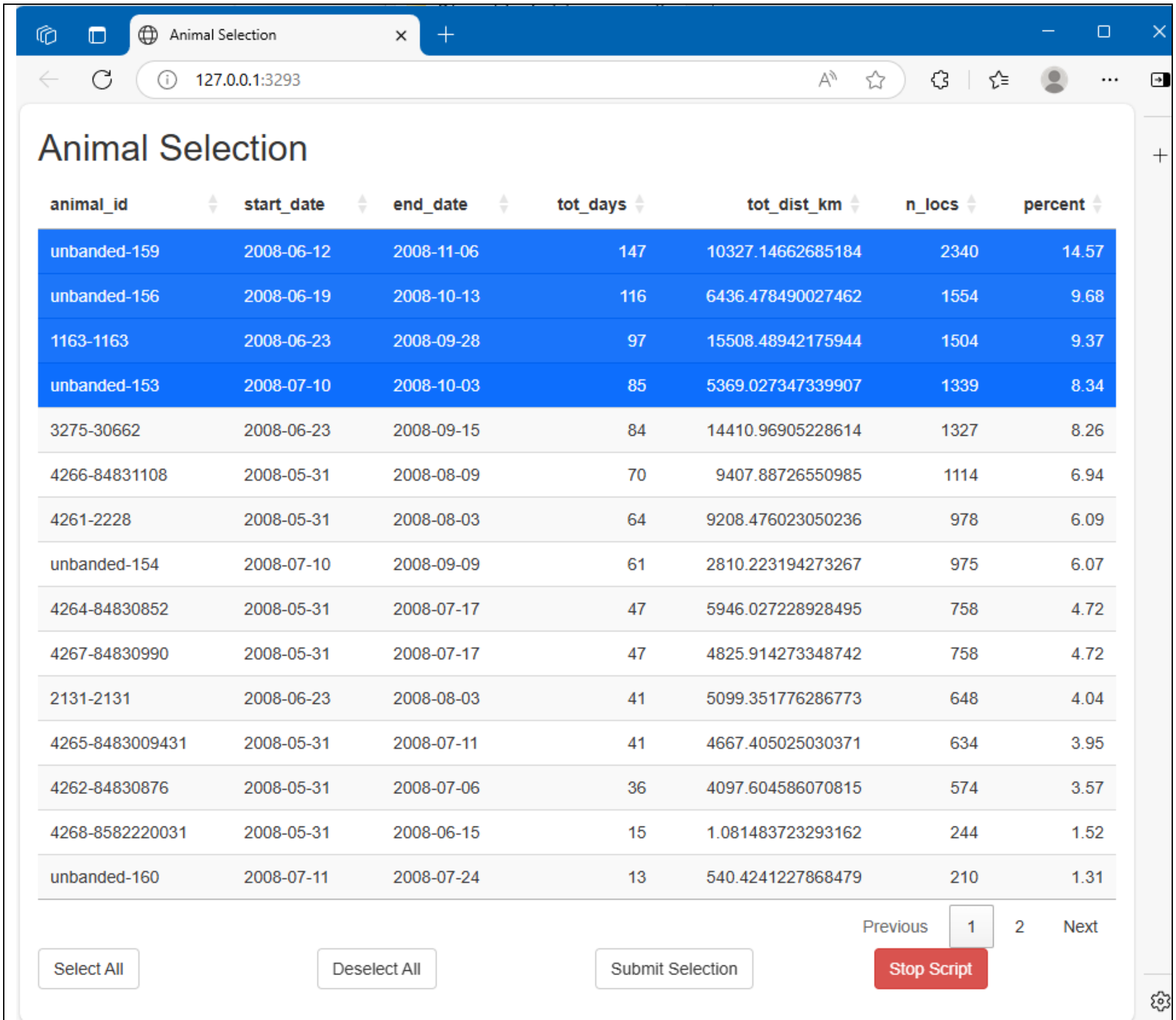
Output data files:

Three CSV output data files are written into a subdirectory that is created by the script located in the user-defined output directory path. The subdirectory name is constructed from the user-defined variable "studyName". The data file names contain studyName, **Dist**, and **Dur**. The "stopovers" output file contains one record for each stop, the "metaStops" output file contains one record for each metastop, and the "locationsAnnotated" output file contains one record for each input location and annotated with a variable denoting the animal's status at the respective place and time: Stopped, Movement, or MovementDuringStop. *The latter location type (MovementDuringStop) is only assigned to consecutive movement locations that begin and end at the same metastop.*

```
Variable descriptions in each of the three output CSV files.

STOPS:
[filename format "stopovers_v01_proxMeters_Dist_minHours_Dur_studyName.csv"]
animal_id         Unique animal ID
species           Species or common name
start_time        Time of first location (YYYY-MM-DD hh:mm:ss)
end_time          Time of last location  (YYYY-MM-DD hh:mm:ss)
stop_hours        Duration of stop (units=hours; end_time minus start_time)
latitude          Average latitude of stopped locations (LC-weighted)
longitude         Average longitude of stopped locations (LC-weighted)
locType           Location type (always "Stopover" in this file)
n_locs            Total number of locations comprising the stop
stop_id           Unique Stop ID (start_time, end_time, and animal_id concatenated)


METASTOPS:
[filename format "metaStops_v01_proxMeters_Dist_minHours_Dur_studyName.csv"]
animal_id         Unique animal ID
species           Species or common name
start_time        Time of first location of the first stop (YYYY-MM-DD hh:mm:ss)
end_time          Time of last location of the last stop   (YYYY-MM-DD hh:mm:ss)
stop_days         Duration of metastop (units=days, end_time minus start_time)
latitude          Latitude of the stop with the longest duration
longitude         Longitude of the stop with the longest duration
locType           Location type (always "Metastop" in this file)
n_stops           Total number of stops comprising the Metastop
meta_stop_id      Unique Metastop ID (start_time, end_time, and animal_id concatenated)


LOCATIONS:
[filename format "locationsAnnotated_v01_proxMeters_Dist_minHours_Dur_studyName.csv"]
animal_id         Unique animal ID
species           Species or common name
timestamp         Raw location time (YYYY-MM-DD hh:mm:ss)
latitude          Raw location latitude
longitude         Raw location longitude
lc                Raw location class (LC)
locType           Segmentation classification (Stopped, Movement, or MovementDuringStop)
stop_id           When stopped, the respective unique stop_id.
n_stops           When stopped, the total number of stops comprising the Metastop
meta_stop_id      When stopped, the unique meta_stop_id. Equals stop_id when n_stops=1.
```

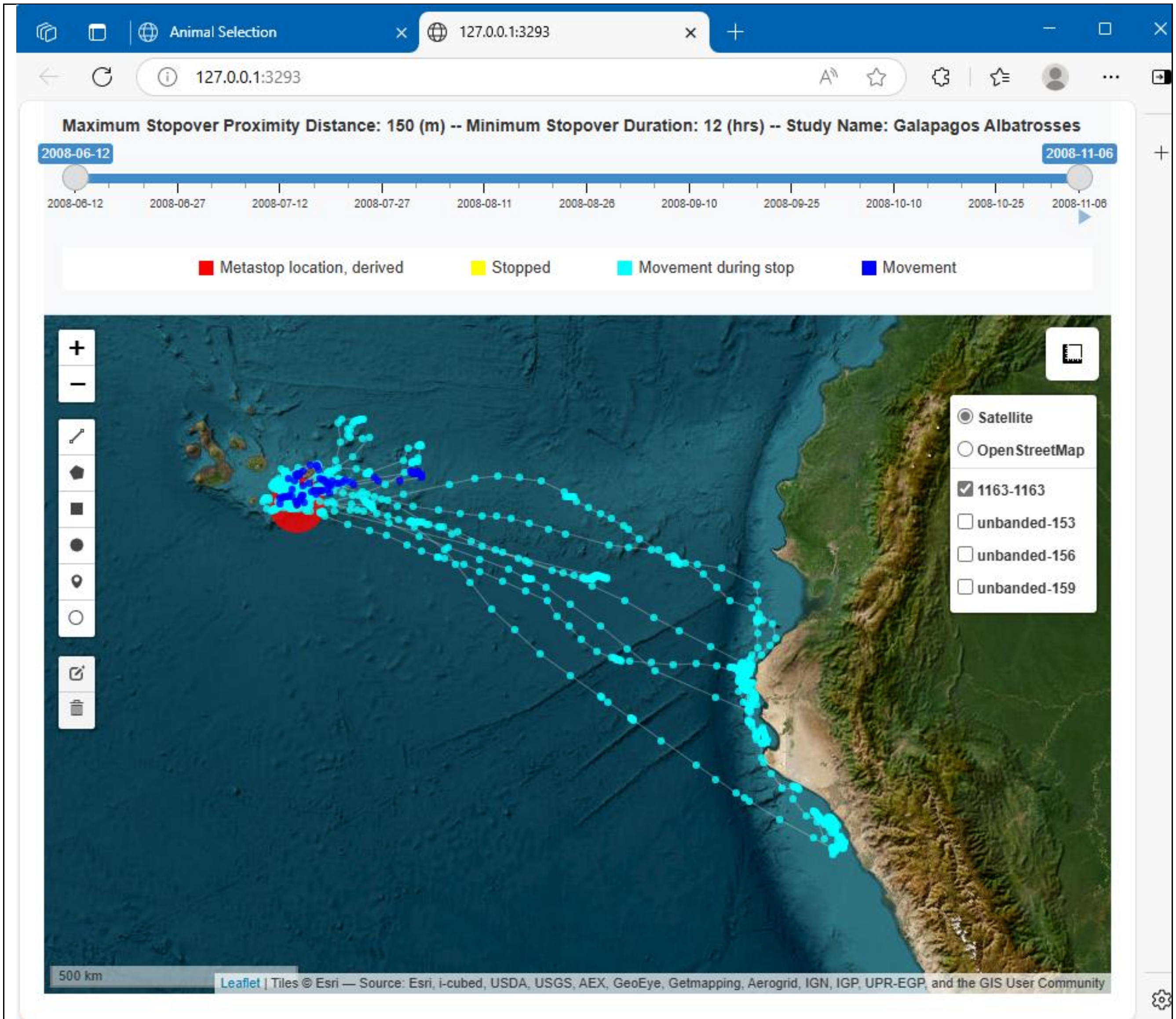
Movebank data example:

To run the "_run_detectStops_v01_Movebank_input_example01.R" script, change the "sourceDir", "resultsDir", "myUser", and "myPass" variable definitions accordingly, then run the script. Three output CSV files will be created, and the results will be presented in a browser-based interactive visualization as described below.



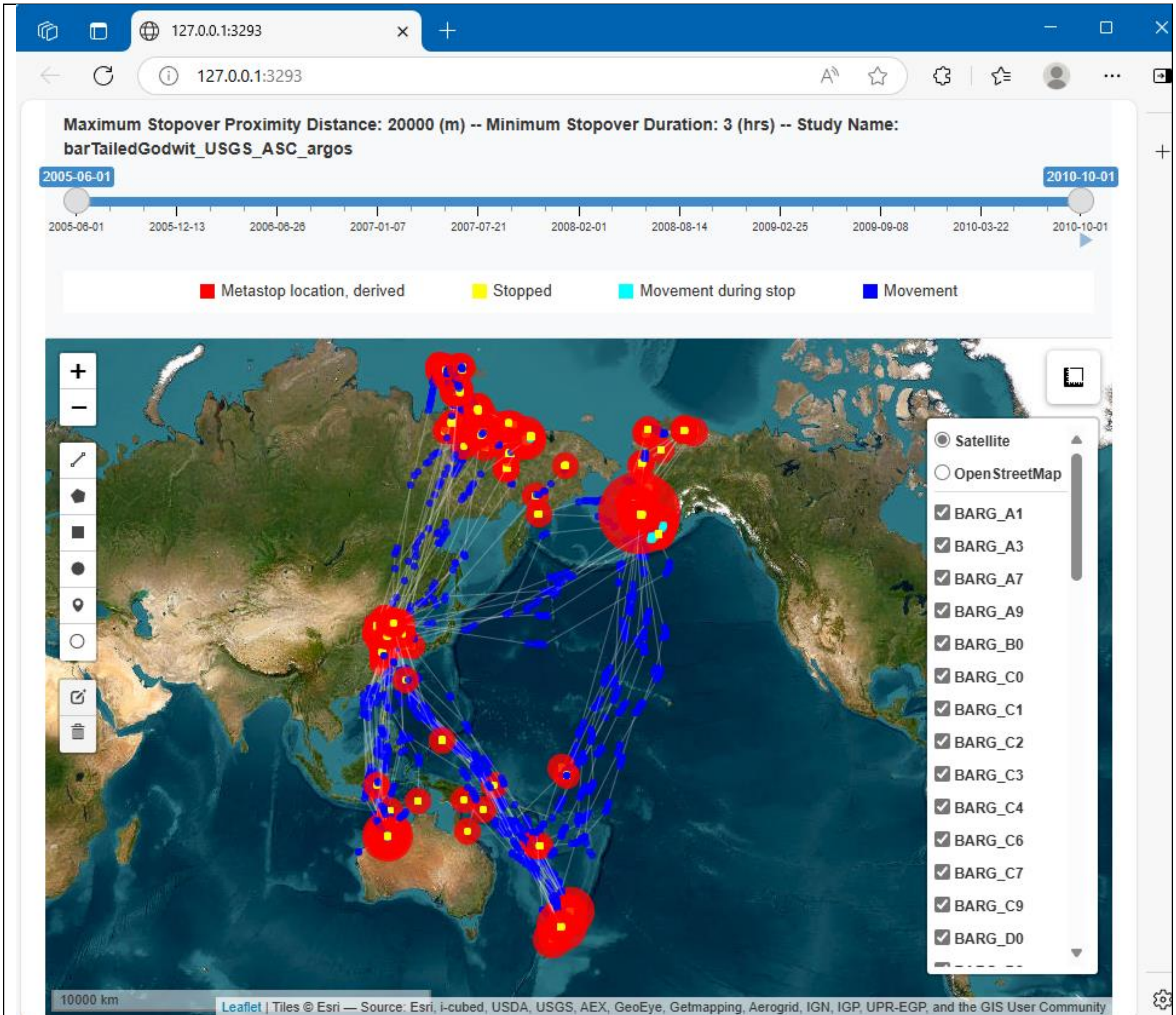
animal_id	start_date	end_date	tot_days	tot_dist_km	n_locs	percent
unbanded-159	2008-06-12	2008-11-06	147	10327.14662685184	2340	14.57
unbanded-156	2008-06-19	2008-10-13	116	6436.478490027462	1554	9.68
1163-1163	2008-06-23	2008-09-28	97	15508.48942175944	1504	9.37
unbanded-153	2008-07-10	2008-10-03	85	5369.027347339907	1339	8.34
3275-30662	2008-06-23	2008-09-15	84	14410.96905228614	1327	8.26
4266-84831108	2008-05-31	2008-08-09	70	9407.88726550985	1114	6.94
4261-2228	2008-05-31	2008-08-03	64	9208.476023050236	978	6.09
unbanded-154	2008-07-10	2008-09-09	61	2810.223194273267	975	6.07
4264-84830852	2008-05-31	2008-07-17	47	5946.027228928495	758	4.72
4267-84830990	2008-05-31	2008-07-17	47	4825.914273348742	758	4.72
2131-2131	2008-06-23	2008-08-03	41	5099.351776286773	648	4.04
4265-8483009431	2008-05-31	2008-07-11	41	4667.405025030371	634	3.95
4262-84830876	2008-05-31	2008-07-06	36	4097.604586070815	574	3.57
4268-8582220031	2008-05-31	2008-06-15	15	1.081483723293162	244	1.52
unbanded-160	2008-07-11	2008-07-24	13	540.4241227868479	210	1.31

Choosing animals for visualization: Animals are listed on separate rows (15 per page), sorted by descending number of tracking locations. Note that the table can be sorted by any column by clicking the sort tabs next to the variable name. Click on a row to select the respective animal. Blue shading shows selection. Click "Submit Selection" to continue to the visualization. Clicking "Stop Script" will abort the visualization. In the example shown, the 4 animals with the most locations have been selected for visualization (see below). When subsetAnimals <- FALSE, this selection interface is skipped entirely, and the visualization will include all animals.



Visualization (Movebank example)**:** Tracking locations for selected animals are displayed in an interactive map. Locations are colored with respect to their classified status: stopped=yellow, movement during stop=cyan, or movement=blue. Metastops are shown as red points, scaled in size by the number of stops combined into the metastop. All points are clickable for displaying specific attributes. Results for one of the four selected animals are shown above; note that the table on right side provides a visibility switch for each animal. Also note the visibility time-slider across the top (with animation capability), and a measurement tool in the upper right corner. The User's choices for **Dist** and **Dur** are documented in the title at the top. Here, one albatross made multiple foraging forays from its nest site on the Galapagos Islands. Locations collected during movements that began and ended at the same stop (i.e. the nest site in this case, and stayed for >12 h), were classified as "movement during stop" (cyan). Locations classified as "movement" (blue) occurred near the end of the track – they were not bounded (before and after) by stops at the same locale.

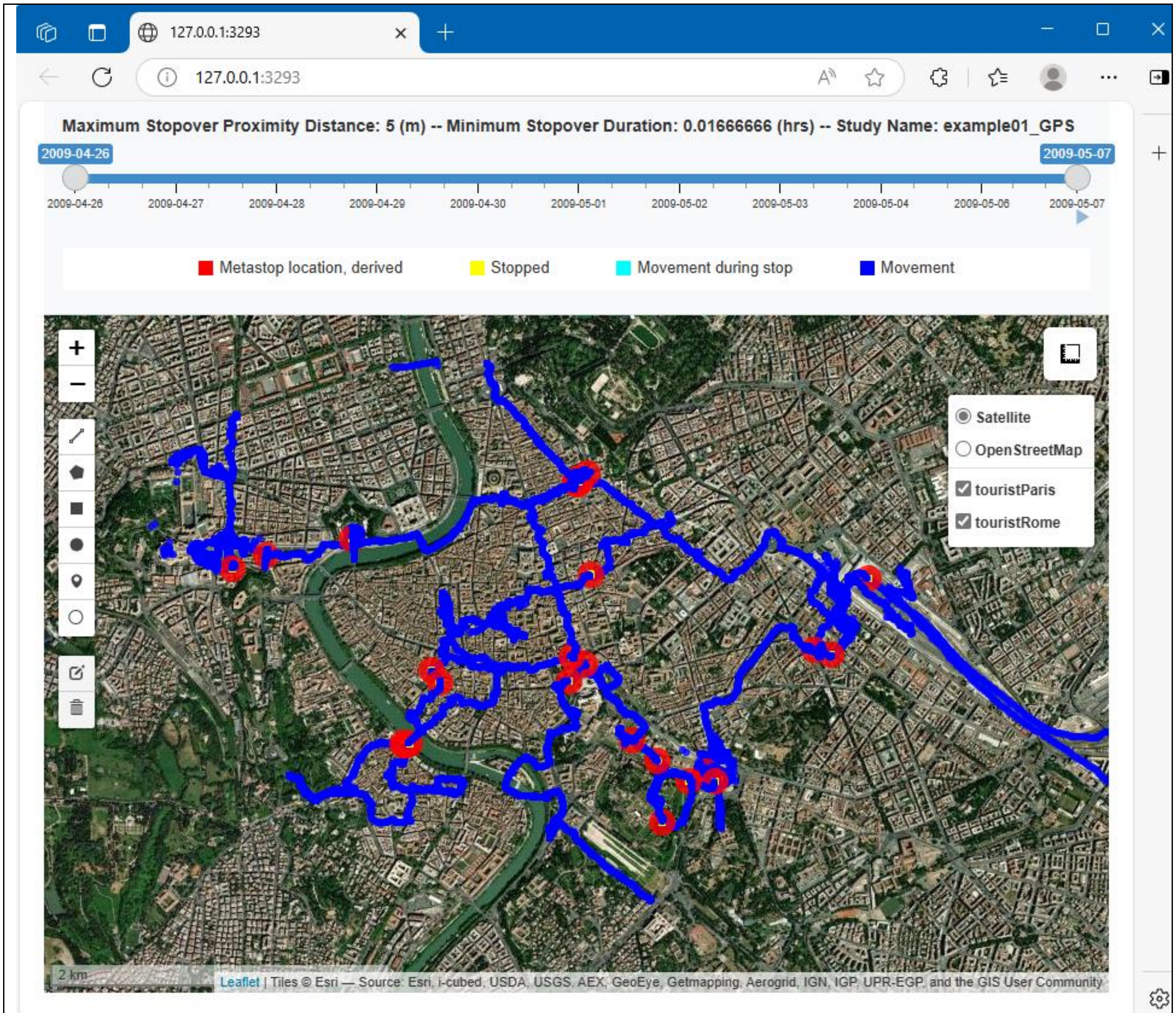
A second, ancillary Movebank example is presented at the end of the "_run_detectStops_v01_Movebank" script. Without exiting the R session, submit the ancillary code to run an analysis of bar-tailed godwit tracking data that are openly available in Movebank (i.e., no license terms need to be accepted in advance). Results for all animals are written to CSV files and displayed for visualization (because subsetAnimals <- FALSE).



Ancillary Movebank example: The analysis detected stops and movements by bar-tailed godwits (n=61). A broad spatial tolerance was defined (**Dist**=20 km) to accommodate locale-scale movements at staging areas as well as Argos location errors. A short stop duration was defined (**Dur**=3 hours) to accommodate restrictive duty cycles that were imposed to conserve battery life. Metastops (red) were detected at locales where the godwits bred, staged, and wintered. Locations along the godwits' long-distance migratory flights were classified as "movement" (blue). Rarely were locations classified as "movement during stop" (cyan), such as when a migratory flight is aborted, and the bird returns to the same locale where it started. Other isolated "movement during stop" classifications could reflect short exploratory flights or locations estimates with large error.

Tabular (CSV) data example #1:

The “_run_detectStops_v01_CSV_input_example01.R” script shows how to input tabular (CSV) tracking data and prepare the required variables. Before running the script, modify the “sourceDir” and “resultsDir” variables to conform to your directory paths. This example #1 analyzes GPS tracking data a tourist in Paris, and in Rome. GPS locations (https://www.microsoft.com/en-us/research/publication/geolife-gps-trajectory-dataset-user-guide/) were collected at very high frequency (every 1 to 10 seconds) while the tourists moved around the cities using different types of transportation. The **Dist** and **Dur** thresholds are defined in this example to detect very localized and brief stops: **Dist**=5 m and **Dur**=0.016666 h (i.e., 1 minute).



Visualization (CSV example #1)**:** Localized stops as brief as 1 minute (red) are detected in a GPS track of a tourist while moving (blue) about Rome, Italy. GPS locations were collected every 1–10 seconds. Note, it appears the data collections specifically targeted periods of movement, so only short duration stops were detectable. Nevertheless, this example illustrates that with high-frequency and high-accuracy data, very brief localized stops can be detected by the algorithm.

Tabular (CSV) data example #2:

The "_run_detectStops_v01_CSV_input_example02.R" script analyzes a simulated tracking data set that follows a hypothetical tracking scenario described below. The scenario was designed to illustrate how different choices for **Dist** and **Dur** affect the detections of stops, movements, and movements during stops. (Note the input data were not empirically collected; they were hand-generated so movement rates between areas of occupancy may sometimes be unrealistic.)

The scenario tracks a professor with GPS-accuracy locations collected every 15 minutes. The professor lives outside Seattle, Washington and commutes ~12 km on weekdays to work 9:00-5:00 at a university in the city. The professor teaches some classes in the same building as their office on Monday, Wednesday and Friday, and works in a laboratory across campus in the afternoons on Tuesday and Thursday. For lunch, the professor goes into the city's downtown district ~5 km away from campus to eat and do errands. On weekends, the professor travels to a cabin in the mountains to hike and relax as detailed later.

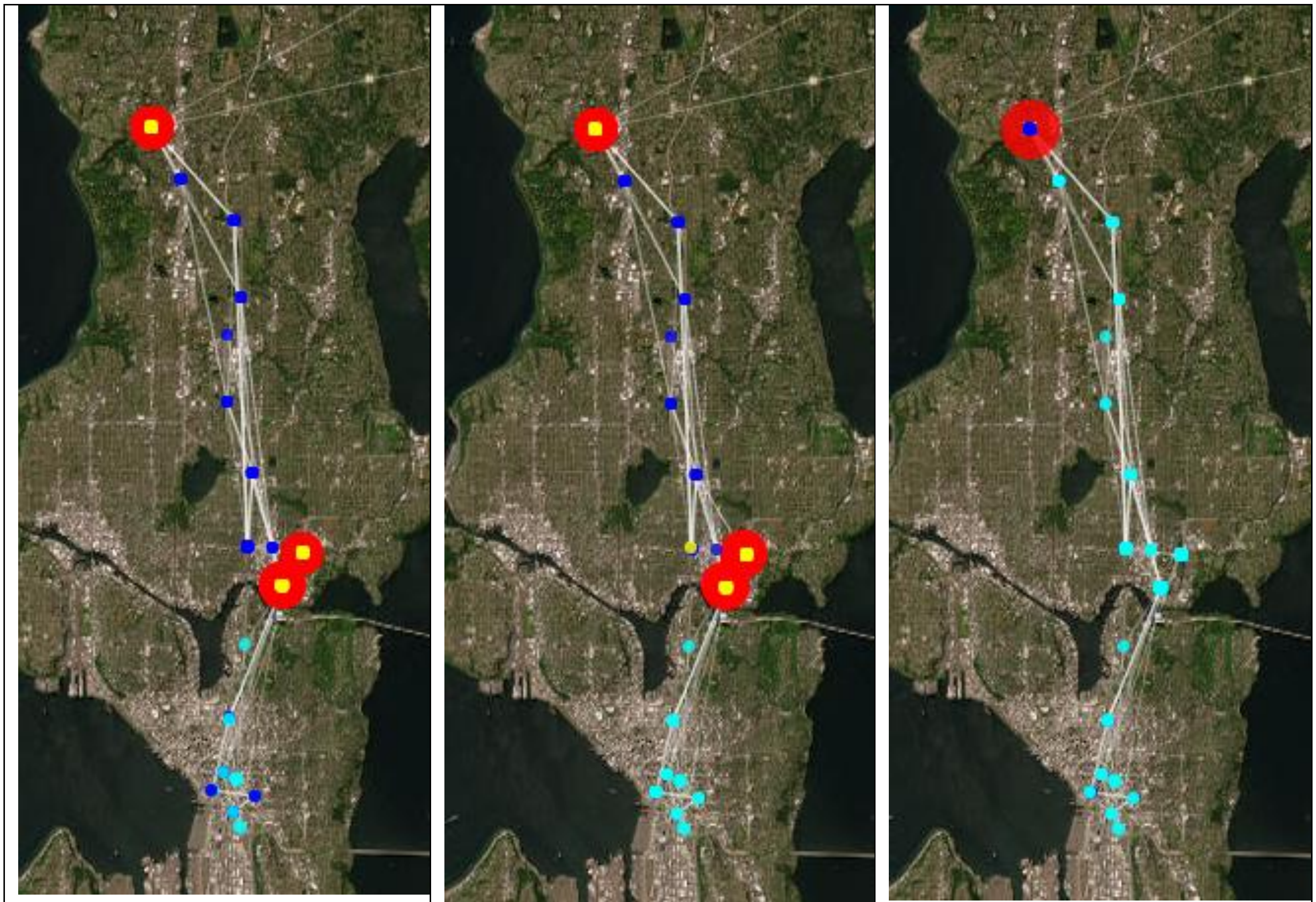
One week of input tracking data are simulated, starting at 00:00 (midnight on Monday morning). For the first example, the two key thresholds are prescribed: **Dist**=100 meters and **Dur**=3 hours.

The algorithm begins with the professor at home at midnight, after which sequential locations are added one at a time. The **Dist** test passes repeatedly (all pairwise distances are <100 m) but the **Dur** test is not achieved until after 03:00 when a stop is detected. Subsequent locations are added to that stop until the professor leaves home to go to work Monday morning. Locations collected en route to the university do not pass the **Dist** or **Dur** criteria. The first location at the professor's office anchors the beginning of the next stop and that stop ends with the last location prior to the professor leaving campus and going downtown for lunch and errands. After lunch, the next detected stop is again at the professor's office building, and it lasts until the last location before the professor commutes back home (with likely movement locations en route). The next stop is at the professor's home and it lasts until the location just prior to being located en route to work on Tuesday morning. Tuesday is like Monday, except after lunch the next stop is at the laboratory across campus (1 km away from the office building). Wednesday and Friday are like Monday, and Thursday is like Tuesday. Early Saturday morning, the Friday night stop at home ends with the last at-home location before the professor is located en route to the cabin. After eating breakfast at the cabin Saturday morning, the professor goes hiking and mid-way circles back to the cabin to get binoculars, then goes out hiking again. No stops (>3 h) are detected until Saturday evening after the professor returns from hiking and stays at the cabin for the night (the first detected stop at the cabin). The professor hikes again Sunday morning (thus ending the overnight stop) before spending all Sunday afternoon reading at the cabin (a second detected stop at the cabin). En route locations are recorded while the professor returns home Sunday evening, at which time the next stop (at home) begins upon arrival home.

There are two instances of >1 stops getting combined into metastops in the tracking scenario above. One instance occurs on Monday, Wednesday and Friday when two consecutive stops are detected each day at the professor's office building (a morning stop and an afternoon stop, separated by the lunchtime movement downtown). Because the two stops were consecutive (no intervening stops) and the derived stop locations were <**Dist** apart, the two stops get combined into a metastop that begins and ends with the timestamps of the first (morning) and last (afternoon) office locations. The lunchtime excursions into the city that occurred on Monday, Wednesday and Friday are attributed as locType = 'MovementDuringStop'. Thus, movements that begin and end at the same stop locale (without stopping somewhere else along the way) get uniquely classified, providing a useful technique to detect movements associated with forays from a central locale, such as foraging forays say from a den or colony. The morning office stop and afternoon laboratory stop on Tuesday (and Thursday) are not pooled into a metastop because those two consecutive stops were >**Dist** apart. On those days, the professor's lunchtime locations get attributed as locType = 'Movement'. The other instance of a metastop occurs on Sunday when the Saturday overnight stop at the cabin is pooled with the Sunday afternoon stop at the cabin, and locations collected during the professor's Sunday morning hike get attributed as locType = 'MovementDuringStop'. The hike on Saturday is attributed as locType = 'Movement' because the hike occurred before any stop at the cabin had been detected.

Now consider an analysis of the same tracking data after increasing the distance threshold to: **Dist**=1500 m and keeping **Dur**=3 hours. During the weekdays, morning and afternoon stops on campus would be detected as before (above) because they are separated by movements downtown during lunch, however on Tuesdays and Thursdays, the morning office and afternoon laboratory stops would be combined into metastops (just like the other weekdays) because the distance between the professor's office and laboratory is <**Dist**. This means that the lunchtime excursions downtown all get classified as 'MovementDuringStop' every day of the week. Essentially, with **Dist**=1500 m, everywhere on the entire campus is considered spatially equivalent (i.e., the same locale). When the professor goes to the cabin, the hike Saturday morning never got >**Dist** from the cabin until after the return to get binoculars so a Saturday morning stop (>**Dur**) gets detected. The second part of the Saturday hike does get >**Dist** from the cabin, so a second Saturday night stop is detected when the professor returns to the cabin after hiking Saturday and rests there until hiking again Sunday. A third stop at the cabin is detected Sunday at the cabin after hiking and before departing for home. All those stops are combined into one metastop that begins upon arrival at the cabin and ends upon departure, with parts of both hikes on Saturday and Sunday classified as "movementDuringStop". There are actually four stops combined into the metastop because the Saturday morning locations get divided into two stops when a pairwise distance between a location en route to the cabin and a hiking location was >**Dist**, which ended the initial stop, then the subsequent location initialized the other stop. Both of these stops were attributed with averaged locations near the cabin and within **Dist** of one another. It's a subtle tracking sequence that is difficult to tease apart, but this illustrates how localized movements can generate series of consecutive localized stops which ultimately get combined into a the more intuitive and arguably more useful metastop.

Finally, consider the same tracking data but analyzed with: **Dist**=100 m and **Dur**=10 h. During weekdays, the only periods during which the professor remained at a locale for >10 hours was at home, so all the full 10+ hour weekday overnights at home get combined into one metastop and each weekday movements (beginning Tuesday) to campus, downtown, back to campus and back home in the evening get classified as bouts of 'MovementDuringStop'. Note the first stop (>10 h) is not detected until Tuesday night because the tracking data do not begin until midnight Monday morning. On the weekend, the only stop >**Dur** occurs Saturday night while sleeping, and since there is no other stop at the cabin, all locations before and after that Saturday night stop get classified as "Movement".



Visualization (CSV example 2): With **Dist**=100 m and **Dur**=3 h (left panel), the southward lunchtime movements from campus to downtown are classified as “MovementDuringStop” (cyan) on Monday, Wednesday and Friday because the professor returns to his office after lunch, but on Tuesday and Thursday the professor goes to a laboratory after lunch that is 1 km from the office, so the lunchtime locations are classified as “Movement” (blue). Keeping **Dur**=3 h, but changing **Dist**=1500 m (center panel), all weekday lunchtime movements downtown are classified as “MovementDuringStop” (cyan) because all buildings on campus are within **Dist** of one another so all stops there are treated as spatially equal. Using **Dist**=100 m and Dur=**10** h (right panel), no stops on campus are detected because they are too short; only stops at home are of sufficient duration for detection, and thus all daytime movements away from home during weekdays are classified as “MovementDuringStop” (cyan).

Questions or feedback:

David Douglas at ddouglas@usgs.gov