Supplementary Material for: A Rough Guide to Pre-processing High-Frequency Animal Tracking Data

Pratik R. Gupte Christine E. Beardsworth Orr Spiegel
Emmanuel Lourie Allert Bijleveld

2020-11-22

7 Contents

8	1	Pro	cessing calibration data	1
9		1.1	Prepare libraries	1
10		1.2	Access data and preliminary visualisation	2
11		1.3	Filter by bounding box	2
12		1.4	Filter trajectories	3
13		1.5	Smoothing the trajectory	5
14		1.6	Thinning the data	6
15		1.7	Residence patches	7
16		1.8	Compare patch metrics	9
17	2	Pro	cessing Egyptian fruit bat tracks	11
18		2.1	Prepare libraries	11
19		2.2	Install atlastools from Github	11
20		2.3	Read bat data	11
21		2.4	A First Visual Inspection	12
22		2.5	Prepare data for filtering	12
23		2.6	Filter by covariates	13
24		2.7	Filter by speed	13
25		2.8	Median smoothing	15
26		2.9	Making residence patches	18
27	3	Ref	erences	20

1 Processing calibration data

Here we show how the residence patch method (Barraquand and Benhamou 2008; Bijleveld et al. 2016; Oudman et al. 2018) accurately estimates the duration of known stops in a track collected as part of a calibration exercise in the Wadden Sea.

1.1 Prepare libraries

First we prepare the libraries we need. Libraries can be installed from CRAN if necessary.

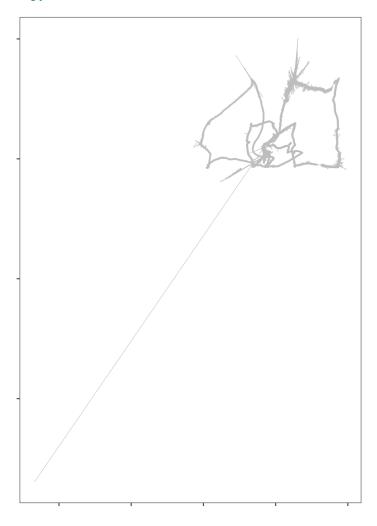
```
# load libs
library(data.table)
library(atlastools)
library(ggplot2)
library(patchwork)

# prepare a palette
pal <- RColorBrewer::brewer.pal(4, "Set1")</pre>
```

1.2 Access data and preliminary visualisation

- First we access the data from a local file using the data.table package (Dowle and Srinivasan
- ³⁶ 2020). We then visualise the raw data.

```
# read and plot example data
data <- fread("data/atlas1060_allTrials_annotated.csv")
data_raw <- copy(data)</pre>
```



1.3 Filter by bounding box

39 We first save a copy of the data, so that we can plot the raw data with the cleaned data plotted

40 over it for comparison.

37

```
# make a copy using the data.table copy function
data_unproc <- copy(data)</pre>
```

- We then filter by a bounding box in order to remove the point outlier to the far south east of
- 42 the main track. We use the atl_filter_bounds functions using the x_range argument, to
- which we pass the limit in the UTM 31N coordinate reference system. This limit is used to
- exclude all points with an X coordinate < 645,000.
- We then plot the result of filtering, with the excluded point in black, and the points that are
- retained in green.

47 1.4 Filter trajectories

48 Handle time

- ⁴⁹ Time in ATLAS tracking is counted in milliseconds and is represented by a 64-bit integer (type
- 50 long), which is not natively supported in R; it will instead be converted to a numeric, or
- 51 double.
- 52 This is not what is intended, but it works. The bit64 package can help handle 64-bit integers
- if you want to keep to intended type.
- 54 A further issue is that 64-bit integers (whether represented as bit64 or double) do not yield
- meaninful results when you try to convert them to a date-time object, such as of the class
- 56 POSIXct.
- 57 This is because as . POSIXct fails when trying to work with 64-bit integers (it cannot interpret
- this type), and returns a date many thousands of years in the future (approx. 52,000 CE) if the
- time column is converted to numeric.
- There are two possible solutions. The parsimonious one is to convert the 64-bit number to a
- 61 32-bit short integer (dividing by 1000), or to use the nanotime package.
- The conversion method loses an imperceptible amount of precision. The nanotime requires
- installing another package. The first method is shown here.
- In the spirit of not destroying data, we create a second lower-case column called time.

```
# divide by 1000, convert to integer, then convert to POSIXct
data[, time := as.integer(TIME / 1000)]
```

65 Add speed and turning angle

```
# add turning angle
data[, angle := atl_turning_angle(data = data)]
```

66 Get 95th percentile of speed and angle

```
# use sapply
speed_angle_thresholds <-
  sapply(data[, list(speed_in, speed_out, angle)],
       quantile, probs = 0.9, na.rm = T)</pre>
```

67 Filter on speed

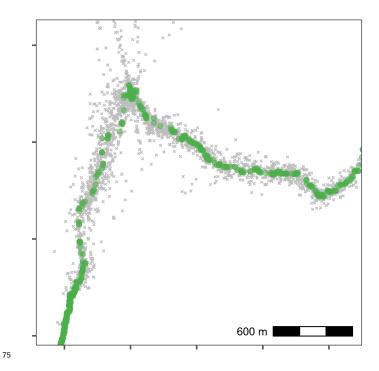
 68 Here we use a speed threshold of 15 m/s, the fastest known boat speed. We then plot the data

with the extreme speeds shown in grey, and the positions retained shown in green.



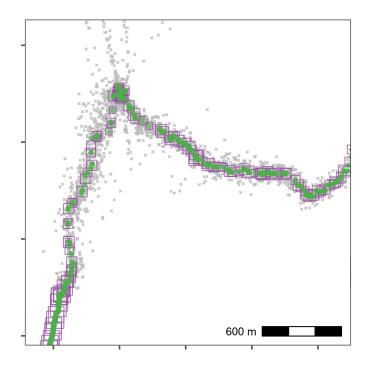
1.5 Smoothing the trajectory

We then apply a median smooth over a moving window (K = 5). This function modifies in place, and does not need to be assigned to a new variable. We create a copy of the data before applying the smooth so that we can compare the data before and after smoothin.



6 1.6 Thinning the data

Next we thin the data to demonstrate thinning by median smoothing. Following this, we plot the median smooth and thinning by aggregation.



1.7 Residence patches

Get waypoint centroids

- 82 We subset the annotated calibration data to select the waypoints and the positions around them
- which are supposed to be the locations of known stops. Since each stop was supposed to be 5
- minutes long, there are multiple points in each known stop.

```
library(stringi)
data_res <- data_unproc[stri_detect(tID, regex = "(WP)")]</pre>
```

- From this data, we get the centroid of known stops, and determine the time difference between
- the first and last point within 50 metres, and within 10 minutes of the waypoint positions' median
- 87 time.

79

- 88 Essentially, this means that the maximum duration of a stop can be 20 minutes, and stops above
- this duration are not expected.

```
# get centroid
```

```
# make a list of positions 10min before and after
```

```
wp_data <- mapply(function(1, u, mx, my) {
  tmp_data <- data_unproc[inrange(time, 1, u)]
  tmp_data[, distance := sqrt((mx - x)^2 + (my - y)^2)]</pre>
```

```
# keep within 50
tmp_data <- tmp_data[distance <= 50, ]

# get duration
return(diff(range(tmp_data$time)))

}, data_res_summary$t_min, data_res_summary$t_max,
data_res_summary$x_median, data_res_summary$y_median,
SIMPLIFY = TRUE)</pre>
```

90 Prepare data

- 91 An indicator of individual residence at or near a position can be useful when attempting to
- identify residence patches. Positions can be filtered on a metric such as residence time (Bracis,
- Bildstein, and Mueller 2018).

94 Calculate residence time

- 95 First we calculate the residence time with a radius of 50 metres. For this, we need a dataframe
- with coordinates, the timestamp, and the animal id. We save this data to file for later use.

97 Run residence patch method

```
We subset data with a residence time > 5 minutes in order to construct residence patches. From
```

- this subset, we construct residence patches using the parameters: buffer_radius = 5 metres,
- lim_spat_indep = 50 metres, lim_time_indep = 5 minutes, and min_fixes = 3.

```
Get spatial and summary objects
```

```
We get spatial and summary ouput of the residence patch method using the atl_patch_summary function using the options which_data = "spatial" and which_data = "summary. We use a buffer radius here of 20 metres for the spatial buffer, despite using a buffer radius of 5 metres earlier, simply because it is easier to visualise in the output figure.

# for the known and unkniwn patches
patch_sf_data <- atl_patch_summary(patch_res_known,
```

106 Prepare to plot data

We read in the island's shapefile to plot it as a background for the residence patch figure.

```
# read griend
griend <- sf::st_read("data/griend_polygon/griend_polygon.shp")</pre>
```

1.8 Compare patch metrics

We then merge the annoated, known stop data with the calculated patch duration. We filter this data to exclude one exceedingly long outlier of about an hour (WP080), which how

```
# get known patch summary
data_res <- data_unproc[stringi::stri_detect(tID, regex = "(WP)"), ]</pre>
# get waypoint summary
patch_summary_real <- data_res[, list(nfixes_real = .N,</pre>
                                        x_{median} = round(median(x), digits = -2),
                                        y_median = round(median(y), digits = -2)),
                                 by = "tID"]
# add real duration
patch_summary_real[, duration_real := wp_data]
# round median coordinate for inferred patches
patch_summary_inferred <-
  patch_summary_data[,
                      c("x_median", "y_median",
                        "nfixes", "duration", "patch")
                      [, := (x \text{ median} = round(x \text{ median}, digits = -2),
                                y_median = round(y_median, digits = -2))]
# join with respatch summary
patch_summary_compare <-</pre>
```

```
merge(patch_summary_real,
         patch_summary_inferred,
         on = c("x_median", "y_median"),
         all.x = TRUE, all.y = TRUE)
# drop nas
patch_summary_compare <- na.omit(patch_summary_compare)</pre>
# drop patch around WP080
patch_summary_compare <- patch_summary_compare[tID != "WP080", ]</pre>
7 patches are identified where there are no waypoints, while 2 waypoints are not identified as
patches. These waypoints consisted of 6 and 15 (WP098 and WP092) positions respectively,
and were lost when the data were aggregated to 30 second intervals.
Linear model durations
We run a simple linear model.
# get linear model
model_duration <- lm(duration_real ~ duration,</pre>
                        data = patch_summary_compare)
# get R2
summary(model_duration)
# write to file
writeLines(
  text = capture.output(
    summary(model_duration)
  ),
  con = "data/model_output_residence_patch.txt"
)
      19
               R^2 = 0.908
      17
   real duration (min)
      15
      13
      11
       9
                       11
                             13
                                   15
                                         17
```

inferred duration (min)

116

10

2 Processing Egyptian fruit bat tracks

We show the pre-processing pipeline at work on the tracks of three Egyptian fruit bats (*Rousettus aegyptiacus*), and construct residence patches.

20 2.1 Prepare libraries

Install the required R libraries that are required from CRAN if not already installed.

```
# load libs
library(data.table)
library(RSQLite)
library(ggplot2)
library(patchwork)

# prepare a palette
pal <- RColorBrewer::brewer.pal(4, "Set1")</pre>
```

122 2.2 Install atlastools from Github.

atlastools is available from Github and is archived on Zenodo (Gupte 2020). It can be installed using remotes or devtools. Here we use the remotes function install_github.

```
install.packages("remotes")
# installation using remotes
remotes::install_github("pratikunterwegs/atlastools")
```

125 2.3 Read bat data

Read the bat data from an SQLite database local file and convert to a plain text csv file. This data can be found in the "data" folder.

```
# write data for QGIS
fwrite(data, file = "data/bat_data.csv")
```

2.4 A First Visual Inspection

Plot the bat data as a sanity check, and inspect it visually for errors (Figure 1). The plot code is hidden in the rendered copy (PDF) of this supplementary material, but is available in the Rmarkdown file "06_bat_data.Rmd". The saved plot is shown below as Figure 1.

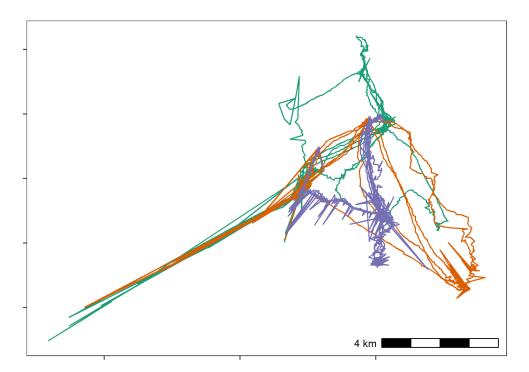


Figure 1: Movement data from three Egyptian fruit bats tracked using the ATLAS system (*Rousettus aegyptiacus*; (Toledo et al. 2020; Shohami and Nathan 2020)). The bats were tracked in the Hula Valley, Israel (33.1°N, 35.6°E), and we use three nights of tracking (5th, 6th, and 7th May, 2018), for our demonstration, with an average of 13,370 positions (SD = 2,173; range = 11,195 – 15,542; interval = 8 seconds) per individual. After first plotting the individual tracks, we notice severe distortions, making preprocessing necesary

2.5 Prepare data for filtering

Here we apply a series of simple filters. It is always safer to deal with one individual at a time, so we split the data.table into a list of data.tables to avoid mixups among individuals.

136 Prepare data per individual

```
# split bat data by tag
# first make a copy using the data.table function copy
# this prevents the original data from being modified by atlastools
# functions which DO MODIFY BY REFERENCE!
data_split <- copy(data)</pre>
```

```
# now split
data_split <- split(data_split, by = "TAG")</pre>
```

2.6 Filter by covariates

- No natural bounds suggest themselves, so instead we proceed to filter by covariates, since point outliers are obviously visible.
- We use filter out positions with SD > 20 and positions calculated using only 3 base stations, using the function atl_filter_covariates.
- First we calculate the variable SD.

```
# get SD.
# since the data are data.tables, no assignment is necessary
invisible(
  lapply(data_split, function(dt) {
    dt[, SD := sqrt(VARX + VARY + (2 * COVXY))]
  })
)
```

Then we pass the filters to atl_filter_covariates. We apply the filter to each individual's data using an lapply.

Sanity check: Plot filtered data

We plot the data to check whether the filtering has improved the data (Figure 2). The plot code is once again hidden in this rendering, but is available in the source code file.

2.7 Filter by speed

- Some point outliers remain (Figure 2), and could be removed using a speed filter.
- First we calculate speeds, using atl_get_speed. We must assign the speed output to a new
- column in the data table, which has a special syntax which modifies in place, and is shown below.
- This syntax is a feature of the data.table package, not strictly of atlastools (Dowle and
- Srinivasan 2020).

```
# get speeds as with SD, no reassignment required for columns
invisible(
  lapply(data_split, function(dt) {
```

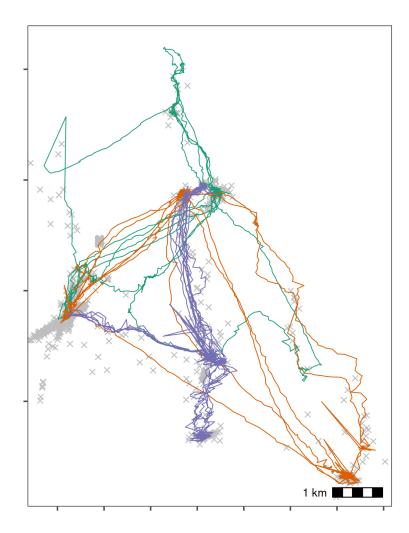


Figure 2: Bat data filtered for large location errors, removing observations with standard deviation > 20. Grey crosses show data that were removed. Since the number of base stations used in the location process is a good indicator of error (Weiser et al. 2016), we also removed observations calculated using fewer than four base stations. Both steps used the function atl_filter_covariates. This filtering reduced the data to an average of 10,447 positions per individual (78% of the raw data on average). However, some point outliers remain.

```
# first process time to seconds
         # assign to a new column
        dt[, time := floor(TIME / 1000)]
        dt[, `:=`(speed_in = atl_get_speed(dt,
                                                   x = "X", y = "Y",
                                                   time = "time",
                                                   type = "in"),
                     speed_out = atl_get_speed(dt,
                                                   x = "X", y = "Y",
                                                   time = "time",
                                                   type = "out"))]
      })
    )
   Now filter for speeds > 20 m/s (around 70 km/h), passing the predicate (a statement return
   TRUE or FALSE) to atl_filter_covariates. First, we remove positions which have NA
   for their speed_in (the first position) and their speed_out (last position).
    # filter speeds
    # reassignment is required here
    data_split <- lapply(data_split, function(dt) {</pre>
      dt <- na.omit(dt, cols = c("speed_in", "speed_out"))</pre>
      dt <- atl filter covariates(data = dt,
                                        filters = c("speed_in <= 20",</pre>
                                                       "speed out <= 20"))
    })
    Sanity check: Plot speed filtered data
   The speed filtered data is now inspected for errors (Figure 3). The plot code is once again
158
   hidden.
159
    2.8 Median smoothing
160
    The quality of this data (Figure 3) is relatively high, and a median smooth is not strictly necessary.
161
    We demonstrate the application of a 5 point median smooth to the data nonetheless.
162
    Since the median smoothing function atl_median_smooth modifies in place, we first make a
    copy of the data, using data.table's copy function. No reassignment is required, in this case.
    The lapply function allows arguments to atl_median_smooth to be passed within lapply
165
    itself.
166
    In this case, the same moving window K is applied to all individuals, but modifying this code
167
    to use the multivariate version Map allows different K to be used for different individuals. This
   is a programming matter, and is not covered here further.
    # since the function modifies in place, we shall make a copy
    data_smooth <- copy(data_split)</pre>
    # split the data again
    data_smooth <- split(data_smooth, by = "TAG")</pre>
```

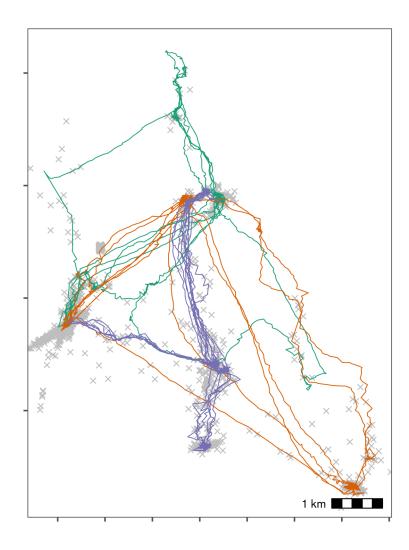


Figure 3: Bat data with unrealistic speeds removed. Grey crosses show data that were removed. We calculated the incoming and outgoing speed of each position using atl_get_speed, and filtered out positions with speeds > 20 m/s using atl_filter_covariates, leaving 10,337 positions per individual on average (98% from the previous step).

170 Sanity check: Plot smoothed data

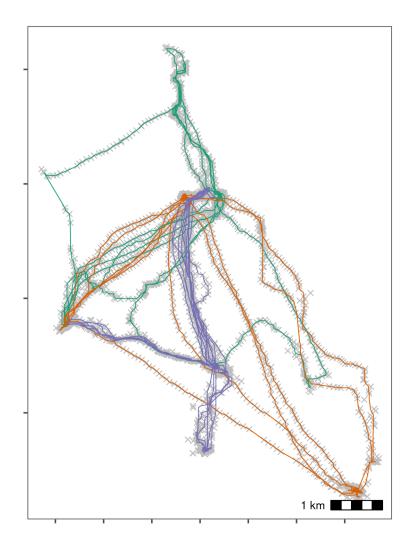


Figure 4: But data after applying a median smooth with a moving window K = 5. Grey crosses show data prior to smoothing. The smoothing step did not discard any data.

2.9 Making residence patches

2 Calculating residence time

First, the data is put through the recurse package to get residence time (Bracis, Bildstein, and Mueller 2018).

```
# load recurse
library(recurse)

# split the data
data_smooth <- split(data_smooth, data_smooth$TAG)</pre>
```

- We calculated residence time, but since bats may revisit the same features, we want to prevent confusion between frequent revisits and prolonged residence.
- For this, we stop summing residence times within Z metres of a location if the animal exited the area for one hour or more. The value of Z (radius, in recurse parameter terms) was chosen as 50m.
- This step is relatively complicated and is only required for individuals which frequently return to the same location, or pass over the same areas repeatedly, and for which revisits (cumulative time spent) may be confused for residence time in a single visit.
- While a simpler implementation using total residence time divided by the number of revisits is also possible, this does assume that each revisit had the same residence time.

```
# get residence times
```

```
data_residence <- lapply(data_smooth, function(dt) {</pre>
  # do basic recurse
  dt_recurse <- getRecursions(</pre>
    x = dt[, c("X", "Y", "time", "TAG")],
    radius = 50,
    timeunits = "mins"
  )
  # get revisit stats
  dt_recurse <- setDT(</pre>
    dt_recurse[["revisitStats"]]
  # count long absences from the area
  dt_recurse[, timeSinceLastVisit :=
          ifelse(is.na(timeSinceLastVisit), -Inf, timeSinceLastVisit)]
  dt_recurse[, longAbsenceCounter := cumsum(timeSinceLastVisit > 60),
             by = .(coordIdx)
  # get data before the first long absence of 60 mins
  dt_recurse <- dt_recurse[longAbsenceCounter < 1, ]</pre>
  dt_recurse <- dt_recurse[, list(</pre>
    resTime = sum(timeInside),
    fpt = first(timeInside),
    revisits = max(visitIdx)
```

```
),
      by = .(coordIdx, x, y)
      # prepare and merge existing data with recursion data
      dt[, coordIdx := seq(nrow(dt))]
      dt <- merge(dt,
                   dt_recurse[, c("coordIdx", "resTime")],
                   by = c("coordIdx"))
      setorderv(dt, "time")
    })
  We bind the data together and assign a human readable timestamp column.
    # bind the list
    data_residence <- rbindlist(data_residence)</pre>
    # get time as human readable
    data_residence[, ts := as.POSIXct(time, origin = "1970-01-01")]
   Constructing residence patches
186
   Some preparation is required. First, the function requires columns x, y, time, and id, which
   we assign using the data.table syntax. Then we subset the data to only work with positions
  where the individual had a residence time of more than 5 minutes.
    # add an id column
    data_residence[, `:=`(id = TAG,
                            x = X, y = Y)
    # filter for residence time > 5 minutes
    data_residence <- data_residence[resTime > 5, ]
    # split the data
    data_residence <- split(data_residence, data_residence$TAG)
190 We apply the residence patch method, using the default argument values (lim_spat_indep
191 = 100 (metres), lim_time_indep = 30 (minutes), and min_fixes = 3). We change the
buffer_radius to 25 metres (twice the buffer radius is used, so points must be separated by
  50m to be independent bouts).
    # segment into residence patches
    data_patches <- lapply(data_residence, atl_res_patch,
                             buffer radius = 25)
```

94 Getting residence patch data

We extract the residence patch data as spatial sf-MULTIPOLYGON objects. These are returned as a list and must be converted into a single sf object. These objects and the raw movement data are shown in Figure 5.

```
# get data spatials
    data_spatials <- lapply(data_patches, atl_patch_summary,
                                which_data = "spatial",
                                buffer_radius = 25)
    # bind list
    data_spatials <- rbindlist(data_spatials)</pre>
    # convert to sf
    library(sf)
    data_spatials <- st_sf(data_spatials, sf_column_name = "polygons")
    # assign a crs
    st_crs(data_spatials) <- st_crs(2039)</pre>
  Write patch spatial representations
    st_write(data_spatials,
              dsn = "data/data_bat_residence_patches.gpkg")
   Write cleaned bat data.
    data_clean <- fwrite(rbindlist(data_smooth),</pre>
                            file = "data/data bat smooth.csv")
  Write patch summary.
    # get summary
    patch_summary <- lapply(data_patches, atl_patch_summary)</pre>
    # bind summary
    patch_summary <- rbindlist(patch_summary)</pre>
    # write
    fwrite(patch_summary,
            "data/data_bat_patch_summary.csv")
    3 References
   Barraquand, Frédéric, and Simon Benhamou. 2008. "Animal Movements in Heterogeneous
   Landscapes: Identifying Profitable Places and Homogeneous Movement Bouts." Ecology 89
203
   (12): 3336–48. https://doi.org/10.1890/08-0162.1.
204
   Bijleveld, Allert Imre, Robert B MacCurdy, Ying-Chi Chan, Emma Penning, Richard M.
    Gabrielson, John Cluderay, Erik L. Spaulding, et al. 2016. "Understanding Spatial Distribu-
206
    tions: Negative Density-Dependence in Prey Causes Predators to Trade-Off Prey Quantity
207
   with Quality." Proceedings of the Royal Society B: Biological Sciences 283 (1828): 20151557.
   https://doi.org/10.1098/rspb.2015.1557.
   Bracis, Chloe, Keith L. Bildstein, and Thomas Mueller. 2018. "Revisitation Analysis Uncovers
   Spatio-Temporal Patterns in Animal Movement Data." Ecography 41 (11): 1801-11. https://doi.org/10.1011/j.j.
   //doi.org/10.1111/ecog.03618.
```

Dowle, Matt, and Arun Srinivasan. 2020. Data. Table: Extension of 'data. Frame'. Manual.

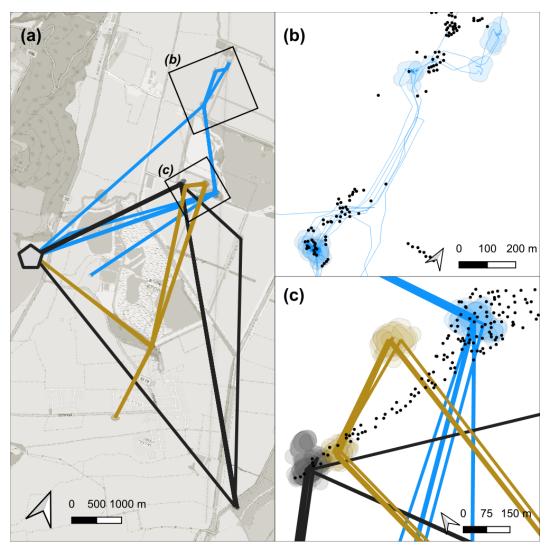


Figure 5: A visual examination of plots of the bats' residence patches and linear approximations of paths between them showed that though all three bats roosted at the same site, they used distinct areas of the study site over the three nights (a). Bats tended to be resident near fruit trees, which are their main food source, travelling repeatedly between previously visited areas (b, c). However, bats also appeared to spend some time at locations where no fruit trees were recorded, prompting questions about their use of other food sources (b, c). When bats did occur close together, their residence patches barely overlapped, and their paths to and from the broad area of co-occurrence were not similar (c). Constructing residence patches for multiple individuals over multiple activity periods suggests interesting dynamics of within- and between-individual overlap (b, c).

- Gupte, Pratik Rajan. 2020. "Atlastools: Pre-Processing Tools for High Frequency Tracking Data." Zenodo. https://doi.org/10.5281/ZENODO.4033154.
- Oudman, Thomas, Theunis Piersma, Mohamed V. Ahmedou Salem, Marieke E. Feis, Anne
- Dekinga, Sander Holthuijsen, Job ten Horn, Jan A. van Gils, and Allert I. Bijleveld. 2018.
- ²¹⁸ "Resource Landscapes Explain Contrasting Patterns of Aggregation and Site Fidelity by Red
- Knots at Two Wintering Sites." Movement Ecology 6 (1): 24-24. https://doi.org/10.1186/
- 220 s40462-018-0142-4.
- 221 Shohami, David, and Ran Nathan. 2020. "Cognitive Map-Based Navigation in Wild Bats Re-
- vealed by a New High-Throughput Tracking System." Dryad. https://doi.org/10.5061/DRYAD.
- G4F4QRFN2.
- Toledo, Sivan, David Shohami, Ingo Schiffner, Emmanuel Lourie, Yotam Orchan, Yoav Bartan,
- 225 and Ran Nathan. 2020. "Cognitive MapBased Navigation in Wild Bats Revealed by a New
- 226 High-Throughput Tracking System." Science 369 (6500): 188–93. https://doi.org/10.1126/
- science.aax6904.
- ²²⁸ Weiser, Adi Weller, Yotam Orchan, Ran Nathan, Motti Charter, Anthony J. Weiss, and Sivan
- Toledo. 2016. "Characterizing the Accuracy of a Self-Synchronized Reverse-GPS Wildlife
- 230 Localization System." In 2016 15th ACM/IEEE International Conference on Information Pro-
- cessing in Sensor Networks (IPSN), 1–12. https://doi.org/10.1109/IPSN.2016.7460662.