

Tutorial



The ‘ClusterApp’ is an application to guide and streamline field studies based on GPS data. It is developed in *Shiny* (Chang et al., 2023) within R version 4.3.1 (R Development Core Team, 2023) and built with the *golem* framework (Fay et al., 2023).

The following document gives a tutorial on the basic usage of the app with the dataset ‘wolf’ provided by the Scandinavian Wolf Research Project (SKANDULV), as well as the usage of the app when applying it on multiple individuals with the dataset ‘bears’ provided by the Scandinavian Brown Bear Research Project (SBBRP). There are five chapters describing the general usage of the app. **Chapter 1: Starting the app** describes how to download and start the app. **Chapter 2: Getting to know the app** focuses on implementing a cluster analysis using the ‘wolf’ dataset. **Chapter 3: Consecutive analysis** deals with important points when running subsequent analyses using an already downloaded shapefile or excel file from a previously executed cluster analysis. **Chapter 4: Analysis on multiple individuals** deals with performing cluster analysis on multiple individuals within the same GPS file. **Chapter 5: Handling GPS bursts** gives a short introduction on the possibilities of the app to work with GPS burst data, such as proximity-triggered GPS events.

The chosen cluster parameter settings within the tutorial are based on real cluster studies conducted on the individual(s) within the GPS data sets provided by SKANDULV and SBBRP (Ordiz et al., 2020; Wikenros et al., 2023). The wolf data follows one individual and was collected over a two-month period from 28 February – 17 April 2022. The objective of this study was to assess predation and scavenging, thus fieldworkers searched for carcasses that wolves killed or otherwise visited. Clusters were generated using 100 m buffers and minimum of 2 GPS locations. The bear data includes the GPS locations of three individuals that were followed during 1 May – 31 May 2014 (Ordiz et al., 2020). The goal of this study was to assess bear predation on moose while collecting data on other relevant bear behavior and signs of activity. Clusters were generated using 30 m buffers with a minimum of 2 GPS locations (Ordiz et al., 2020; Rauset et al., 2012).

1. STARTING THE APPLICATION

The application can be started by downloading the package from GitHub:

```
install.packages("devtools")
devtools::install_github("JohannaMz/ClusterApp")

library(ClusterApp)
```

If you want to follow the tutorial with a manual upload of the data (which will be addressed in the following chapters), you have to download the GPS data. Data for the wolf examples are loaded by the command:

```
data <- ClusterApp::wolf
```

And the data for the bear examples, including multiple individuals, are loaded by the command:

```
data <- ClusterApp::bears
```

A shortcut to accessing the demo data can be found on the first tab under ‘File upload’. This is per default set to ‘Manual upload’, however ‘Demo data wolf’ or ‘Demo data bears’ can be selected directly, which will load the data for usage within the app. The file path is set to the local working directory, so files will be downloaded into this folder.

More information regarding the two datasets can be found by searching for the datasets in the *Help* panel. To be able to fully use the data within the application and follow all steps as if done with own data, it must be saved in a folder of your choice. It is good practice to have a separate folder per GPS data file, which can include data on one or more individuals. When running multiple cluster studies, each new GPS file should be stored or replaced here. In this folder, you will also find the output of the cluster analysis.

```
write.csv(data, "path/to/folder/data.csv")
```

After loading the package, you can start the application by running the command:

```
ClusterApp::run_app()
```

The application opens in a separate window and R runs in the background. For a better view, choose ‘Open in browser’ in the upper left corner. The application will now open within your Internet browser.

The application has three main steps:

1. Upload GPS data

- Upload the GPS data
- Define the necessary column names
- Adjust date format and coordinate system format

2. Adjust Cluster Analysis Parameters

- Adjust all the parameters necessary for the cluster analysis

3. Cluster Analysis Output

- Look at the results, enter data in cluster file, and download cluster -, GPS data- and map files

2. GETTING TO KNOW THE APPLICATION

1. Upload GPS data (Figure 1)

General usage is shown with the ‘wolf’ dataset (*wolf dataset-specific settings are listed in italics in parentheses*). The data are loaded by keeping the ‘File upload’ option on ‘Manual upload’, selecting ‘Upload the original GPS file here:’, searching for your GPS file and loading it. The path to your datafile will appear. Possible formats are *.csv* or *.shp*. For the format *.csv*, the data separator can be defined as tab, semicolon, comma, or colon; making sure that the columns are loaded correctly.

After uploading the data, the file path appears, and windows ‘Data’ and ‘Data Summary’ give information about the uploaded data. These can be used to check if the data were uploaded correctly, and which column names fit as input for the ID (Animal ID), timestamp, and coordinate variables. The timestamp must include both date and time in one column. The timestamp default format is ‘**Year-month-day Hour:Minute:Second**’, but this can be adjusted based on standard R formatting.

R date formatting can for example be:

%Y-%m-%d %H:%M:%S = 1999-12-25 12:00:00

%Y-%d-%m %H:%M:%S = 1999-25-12 12:00:00

%Y-%m-%d %H:%M = 1999-12-25 12:00

%y-%m-%d %H:%M:%S = 99-12-25 12:00:00

%Y/%m/%d %H:%M:%S = 1999/12/25 12:00:00

Additionally, the coordinate system the data is provided in (*EPSG code 4326 for WGS84*) as well as in which UTM zone the data should be downloaded (*Zone 33 for Sweden*), can be adjusted. Within the analysis the data are always converted to UTM, as this coordinate system has metric units, which are more intuitive with analyses involving distance measures.

The red info sign (i) always give additional information on the necessary steps within the analysis. Blue underlined text are URLs to external websites.

Figure 1: Filling in the settings within the tab Upload GPS data for the example data ‘wolf’.

2. Adjust cluster analysis parameters (Figure 2)

In the first step within the second tab, it is necessary to state a unique label of the individual(s) the input data belongs to which will be used throughout the entire field season. This should be a label useful for separating it from other cluster studies that might be done at the same time. When executing consecutive cluster analyses this unique label will make sure that earlier output files with the same label are found by the application, if available, in the folder of the GPS data (further explanation under **Consecutive analysis**). For the tutorial we choose the label 'wolf_demo'.

One of the important tasks regarding cluster analysis, which is directly related to the focal research questions, is setting the parameters necessary to generate biologically relevant clusters. Clusters are formed by applying a buffer of x metres around every GPS location. If these buffers overlap (so at a maximum distance of $x \times 2$), they will form clusters which will be saved as a cluster if y number of GPS locations are within this buffer using all locations within the study period.

Therefore, it is possible to adjust the following parameters:

- **Buffer size:** Depending on the study species and research questions, the relevant buffer distance criteria between GPS locations can be adjusted. Changing the buffer distance results in different cluster sizes. *For the wolf data, a distance of 100 m was used, which will mean locations within 200 meters of each other form a cluster.*
- **Number of GPS locations:** Depending on the study species and research questions, the minimum number of GPS locations used to generate a cluster can be set. *For the wolf data, the number of locations was set to 2.*
- **Study period:** To be able to select the period during which the animal is being followed, both the study start and end date need to be set. The study start date must be entered the same each time within a study. The end date can be in the future if the study is ongoing. Often the fix rate of the GPS device is adjusted to receive GPS locations more frequently within a study period, therefore it is important to set the dates correctly and thus exclude GPS locations from outside this time frame. The study period of this demo data was from 28 February - 17 April, 2022.

To demonstrate a case of downloading data and then using this downloaded data in a consecutive analysis, the end date was set to 17 March 2022. That means only data between 28 February and 17 March, 2022 will be used for this first analysis. In Chapter

3: Consecutive analysis we explain the column ‘Some additional info, if you have done an analysis before’ and will use the clusters, that were developed in this step.

- Number of days before and after the study period: Some studies define a number of days before and after the study period to be still taken into account for developing clusters. Points within this time frame are still taken into account for clusters, if at least one point within was also during the study period. *For this study, no time frame was added, so the default of 0 days can be left untouched.*

Further optional adjustments include the filtering of the initial GPS data to only include locations every x minutes. The default is to use all GPS locations that are available. *The summary of the time difference shows that for the wolf data, the minutes between GPS positions stay constant at approximately one hour (60 minutes). Therefore, no adjustments are necessary here, which will result in all GPS locations being used for the analysis.* More detailed explanation can be found in **Chapter 5: Handling GPS bursts**.

Only influencing overview purposes, the user can, further, decide whether columns with information on time spent at the location and the number of points within/outside of the cluster should be added to the final resulting cluster table in third tab.

Additionally, there is also the option to tick ‘Should only clusters with consecutive GPS locations be used?’. This option will generate clusters only for sequential positions in time that are within the defined buffer zone. Developed buffers can therefore overlap but will be separated if they are at different points in time.

The screenshot displays the 'Adjust Cluster Analysis Parameters' interface, organized into three columns:

- Column 1: Input Parameters**
 - Give the individual or group of individuals a label:
 - Set the buffer size in meters:
 - Set the number of GPS locations it needs to be a buffer:
 - Define the study period: to
 - Enter the number of days that will be added before and after the study period:
- Column 2: Summary and Time Filtering**
 - Summary for optional time stamp filtering:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
57.57	59.57	60.00	62.22	60.43	780.43
 - Set the time difference needed between GPS locations in minutes:
 - Should the columns related to the time spent and the number of points within/outside of the cluster be displayed in the cluster table? ☐
 - Should only clusters with consecutive GPS locations be developed? ☐
- Column 3: Previous Analysis Info**
 - Some additional info, if you already have done an analysis before:
 - Optionally load a latest file manually:
 - Should old clusters automatically be marked as done? ☐

Figure 2: Setting the necessary parameters within the tab Adjust Cluster Analysis Parameters, which is separated into three columns. The first column deals with the important parameters, the second column gives a summary of the time difference between GPS fixes and offers further (not compulsory) parameter adjustments and the third column related to previous cluster analysis files.

3. Cluster analysis output

After setting all the data file options and the cluster analysis parameters, the cluster analysis can be performed under the tab ‘Clusters Analysis Output’. A message will notify the user if the analysis ran successfully or not and the output will appear (Figure 3).

Alternatively, a cluster shapefile can be uploaded from one of the folders (and the steps in Tabs 1 and 2 skipped). This is an option if the output of an earlier cluster analysis should be inspected without the need for a new analysis. When this is uploaded, some properties regarding the track and GPS locations in the interactive map are not possible, as these are exclusively related to a newly performed cluster analysis.



Figure 3: Successfully performed cluster analysis. Alternatively, a latest cluster file can be loaded here.

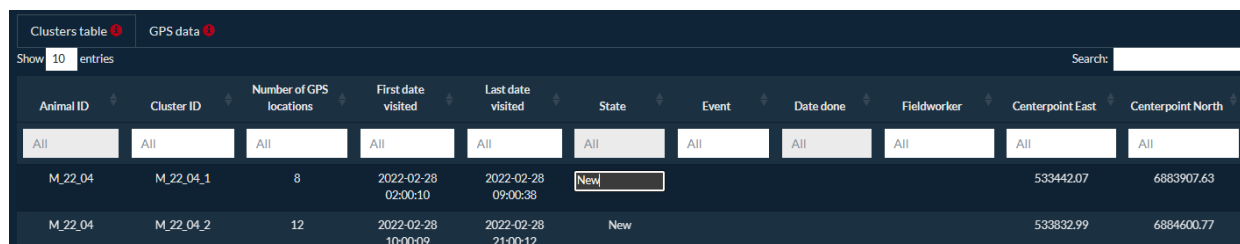
Two tables are created with the analysis which will be visible in two different windows (‘Clusters’ and ‘GPS data’).

The clusters table appears first as the default setting and contains all the data concerning the clusters generated by the analysis (Figure 4). The table includes columns regarding the animal ID (selected as ‘Animal ID’ column), the unique ‘Cluster ID’ (built as ‘Animal ID’ underscore ‘Cluster ID’) and relevant information regarding the clusters (e.g., number of locations in cluster, first and last date of visit at the cluster, centre points of the clusters). Based on the decision in the second tab, extra columns on the percent of time spent at the cluster are added. The ‘percent of time spent at the cluster’ is calculated as the number of locations the individual could potentially have been at the cluster between the first and last time it visited (taking the mean time difference GPS locations are taken) divided by the actual number of locations within this cluster plus one (further explanation in **Chapter 5: Handling GPS bursts**). Further insight into the time spent at the location are also given by the column ‘Number of GPS locations within/outside of the cluster’ which gives a summary of the number of the GPS locations inside and outside of the cluster between the first and last visit. This difference is translated into the four categories within the column ‘Classification of clusters by GPS location’: 1) ‘All locations

within cluster.’; 2) ‘More locations outside of cluster’, so during the time frame the individual spent more time outside of the cluster and returned irregularly; 3) ‘More locations inside of cluster’, translating to regular visits to the cluster but with excursions outside of it and 4) ‘Even number of locations inside and outside of cluster’.

The data in the following four columns: ‘State’, ‘Event’, ‘Date done’ and ‘Fieldworker’ can be manually adjusted by the user by double clicking on it. ‘State’ refers to the state of the cluster; all clusters that were generated during this cluster analysis will be named ‘New’, while clusters that expanded since the last cluster analysis are stated as ‘Points added’. In addition, the user can choose to manually adjust the state of a cluster. When clusters have been visited, the user can change the ‘State’ and add what was found at the location (Event), when the cluster was visited (Date done) and the name of the fieldworker who visited the cluster. These adjustments can change the aesthetics of the interactive map and streamline data management.

The top of the data table also allows for specific filtering on all columns. Plotting filtered tables anew will display only the filtered output. This can either be useful for overview purposes within the map or downloading/looking at only specific clusters, e.g., specific IDs, within a specific time frame, etc.



Animal ID	Cluster ID	Number of GPS locations	First date visited	Last date visited	State	Event	Date done	Fieldworker	Centerpoint East	Centerpoint North
M_22_04	M_22_04_1	8	2022-02-28 02:00:10	2022-02-28 09:00:38	New				533442.07	6883907.63
M_22_04	M_22_04_2	12	2022-02-28 10:00:09	2022-02-28 21:00:12	New				533832.99	6884600.77

Figure 4: Cluster table output with a selected row, which is being edited in the column State.

The cluster table can be downloaded in the output formats: *.shp*, *.xlsx* or as a *.gpx* file and will always be saved in the folder where the GPS file was initially uploaded from (Figure 5). **Downloading the cluster file as shapefile is mandatory to save data changes and if the file should be used for any consecutive analyses.** This option will always download the entire cluster table in UTM and the specified zone in the folder of the original GPS data as well as a text file with the chosen settings that were used in the cluster analysis function. Downloading *.xlsx* and *.gpx* files are optional for further purposes such as easy adjustments and data management in excel or loading the file onto a GPS device. The *.gpx* is downloaded as a track in WGS 84 for direct import into other applications or devices. If filtering options were applied

in the table, only the filtered data frame will be downloaded as the *.xlsx* and/or *.gpx* file. For this tutorial, we *choose to download the shapefile as well as the .xlsx file*.

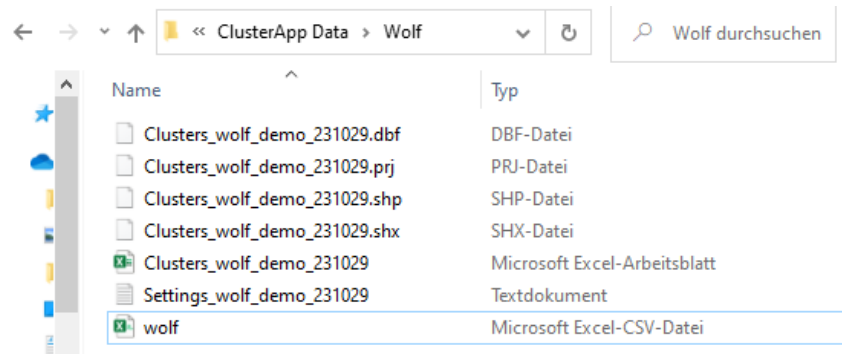


Figure 5: Folder with the initial GPS file named *wolf* and the downloaded shapefile for the clusters that were identified in this cluster analysis, as well as the same file downloaded as an excel file. The name is built according to the label given and the date that it is downloaded. The shapefile as well as the excel file can now serve as a previous cluster analysis file.

The cluster table can be plotted as an interactive map by pressing ‘Plot Data’. The map has the ‘WorldTopoMap’ (ESRI, 2021) as a general background layer, but can be switched to ‘WorldImagery’ (ESRI, 2023), and will show several layers regarding the cluster analysis. Per default the map shows cluster polygons coloured according to their state, as well as the latest GPS position of the individual(s) as a blue drop pin. There are additional options to show the walking trajectory (tracks), Events of the clusters, or the Clusters ID as text within the map. All aspects can be activated and deactivated depending on the user's choices (Figure 6).

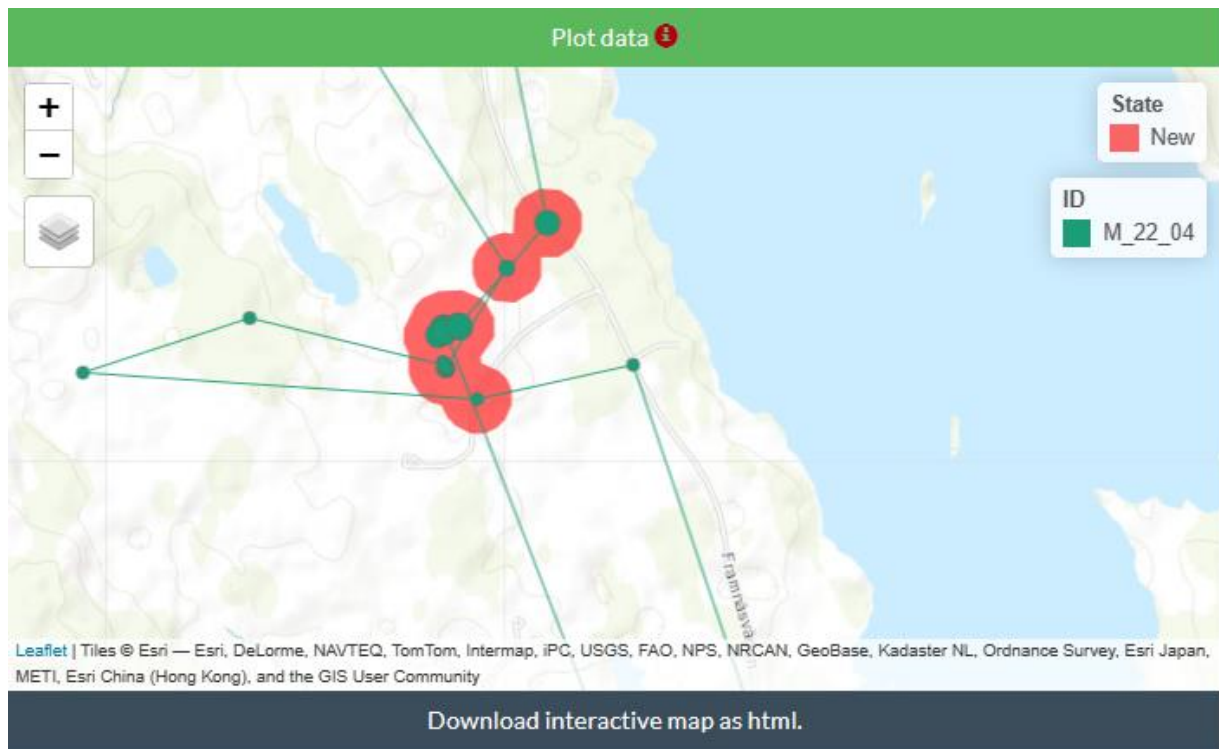


Figure 6: Interactive map output showing the clusters filled according to their state as well as the GPS locations and the track the individual walked.

Manual adjustments of the State and Event column in the cluster table can be updated by plotting the data anew. These options help to illustrate visited clusters and understand the spatial distribution of the findings or ‘Events’ around clusters. When a row is selected within the cluster table and the map is plotted anew, this cluster is highlighted (Figure 7). The map can be downloaded as an interactive file in *.html* format.

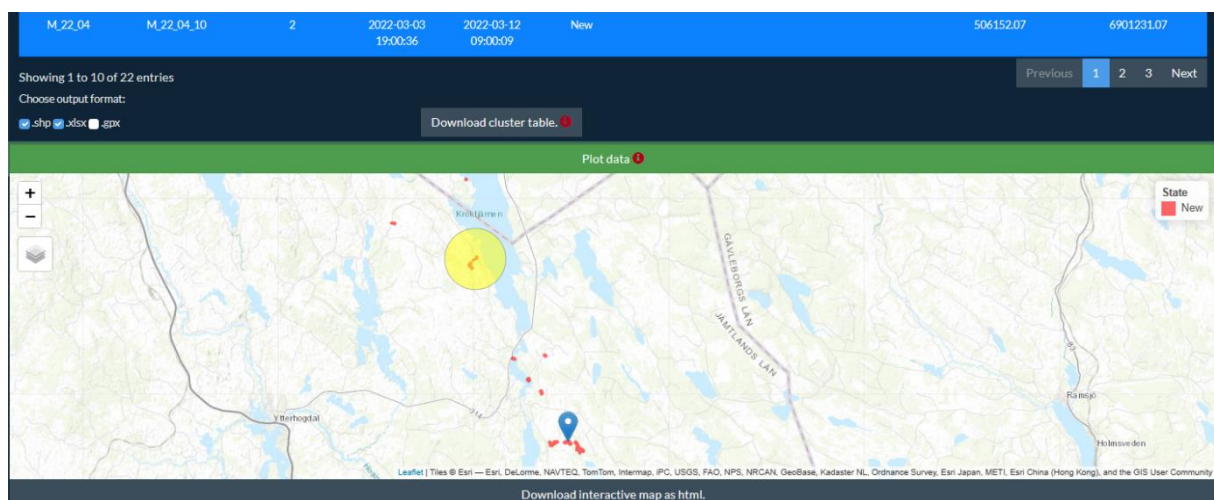
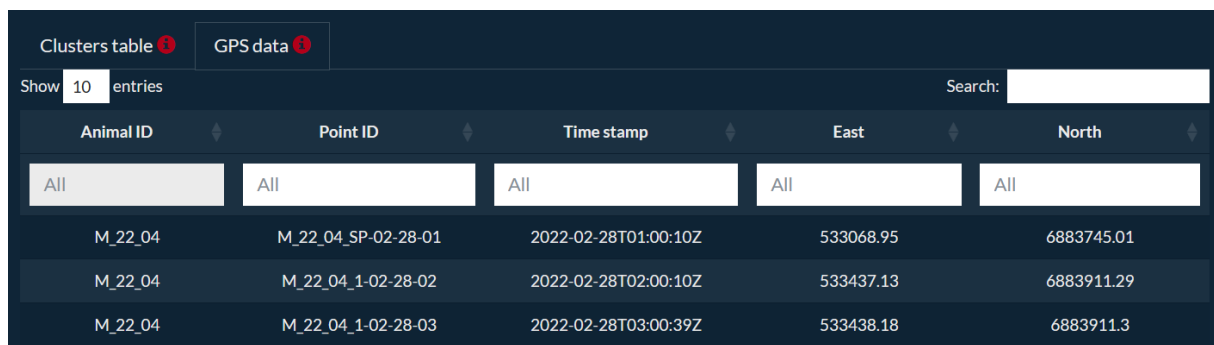


Figure 7: Output of the interactive map with the last row in the cluster table selected. This results in a light-yellow circle highlighting the cluster in the map.

The window ‘GPS data’ gives an overview of all GPS locations that were used for the cluster analysis which can be displayed on the interactive map (‘GPS locations’; Figure 8). The point ID is a combination for easier identification in the field and consists of the animal’s ID, the cluster number it belongs to, or SP for single points, and then the month, day, and hour the GPS location was made. This way, field personnel can get a quick overview of the distribution of the GPS locations within the visited cluster. Again, the data can be downloaded as *.shp* or *.xlsx* in UTM and the specified zone or as *.gpx* file in WGS84. The shapefile will download all data, while the *.xlsx* and *.gpx* file will only download selected data. This data is especially useful to load on a handheld GPS device or for any other form of orientation in the field.



Animal ID	Point ID	Time stamp	East	North
M_22_04	M_22_04_SP-02-28-01	2022-02-28T01:00:10Z	533068.95	6883745.01
M_22_04	M_22_04_1-02-28-02	2022-02-28T02:00:10Z	533437.13	6883911.29
M_22_04	M_22_04_1-02-28-03	2022-02-28T03:00:39Z	533438.18	6883911.3

Figure 8: Output of the GPS data table including all GPS locations that were used for the analysis.

Important points in this chapter:

- One folder with a unique name should include all GPS data. All downloaded files will be saved to this folder.
- The chosen label must be consistent throughout the study.
- Downloading the cluster file as shapefile is mandatory to save data changes and in order to use the file for any consecutive analyses.

3. CONSECUTIVE ANALYSIS

The first analysis has now been performed and the data downloaded.

GPS cluster studies often require field personnel to visit GPS positions of focal individuals in almost real-time, i.e., personnel must find a balance between not disturbing the animal, but

visiting the site before evidence of the individual's activity disappears. This means that field personnel routinely download GPS data and run cluster analyses at regular, commonly short (e.g., every day or few days) intervals. It is standard practice that cluster analyses are run several times within one study period.

After the shapefile of the clusters is downloaded the first time, the next analysis that is done will be a 'consecutive analysis'. Several steps are necessary to ensure the data are consistent throughout the study period: 1) New GPS data should always be saved within the same folder, replacing the old GPS data (data 'wolf.csv' in Figure 5), 2) the **uploaded GPS data must have the same starting date throughout the study**, i.e., users must upload all GPS data associated with the study each time, 3) **the study period must be defined from the same starting date each time**, 4) **the label has to be consistent for the study** (i.e. 'wolf_demo' in our case) and 5) **the cluster tables must be downloaded after every time users run a successful cluster analysis**.

The procedure starts the same by uploading the newest GPS data from the same folder and then setting the correct column names and the settings. *The study period should now range from the 28th of February 2022 to 17th April 2022.* The base folder should now additionally include a previous cluster shapefile or excel-file (data 'Clusters_wolf_demo_231029.shp' or 'Clusters_wolf_demo_231029.xlsx' in Figure 5).

Differences will appear for the third setting column within 'Adjust cluster analysis parameters', which relates to earlier cluster analysis files within the folder of the raw GPS data. With this folder path known, and the unique label entered, the application will automatically search for a shapefile within the same folder with an exact string match ('Clusters underscore label underscore date.shp', e.g., 'Clusters_wolf_demo_231029.shp'). There can be multiple old shapefiles in the sub-folder as the program will search for the most recent file. So, in order to load old cluster files correctly, **it is important to have an individual or group label which is unique during the entire field study and have all relevant files within the same folder** (see Figure 5). If this match is not found, the message will read 'No latest cluster file'. With the statement 'No latest cluster file' the analysis will be run as if done for the first time.

With the button 'Optionally load a latest file manually' in Tab 2 in the same column, the latest cluster file can also be loaded manually in .shp or .xlsx format from any folder of your choice. This can either be useful if an earlier shapefile should be taken as the previous study or if the field staff prefers to enter data in excel, rather than directly within the application. However, it is necessary to take care in this form of manual upload: the file must contain all latest clusters

(so not only a selection as could be possible with the download of selected clusters in .xlsx format) and also all necessary column names. Only changes to the four adjustable columns: State, Event, Date done, and Fieldworker will be accepted in the consecutive cluster table. **So, no column names should be changed, no additional columns can be added, and no rows should be deleted.**

When performing a consecutive cluster analysis, it is important to check if the most recent cluster file is loaded (Figure 9), as this file will make sure the earlier established cluster IDs remain constant even with additional GPS locations that might increase, combine, or create new clusters. That is why **it is important to always download the cluster table as a shape file!** You can optionally automatically mark the state of all clusters from the most recent file as 'Done'.

The screenshot shows the ClusterApp configuration window. It has a dark blue background with white text and input fields.

- Left Column (Input Parameters):**
 - "Give the individual or group of individuals a label:" with a text box containing "wolf_demo".
 - "Set the buffer size in meters:" with a text box containing "100".
 - "Set the number of GPS locations it needs to be a buffer:" with a text box containing "2".
 - "Define the study period:" with a date range from "2022-02-28" to "2022-04-17".
 - "Enter the number of days that will be added before and after the study period:" with a text box containing "0".
- Middle Column (Summary for optional time stamp filtering):**
 - A table showing summary statistics:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
57.57	59.57	60.00	62.22	60.43	780.43
 - "Set the time difference needed between GPS locations in minutes:" with a text box.
 - Two checkboxes:
 - "Should the columns related to the time spent and the number of points within/outside of the cluster be displayed in the cluster table?" (unchecked).
 - "Should only clusters with consecutive GPS locations be developed?" (unchecked).
- Right Column (Additional Info):**
 - "Some additional info, if you already have done an analysis before:" with a text box containing a file path: "han/Documents/ClusterApp Data/Wolf/Clusters_wolf_demo_231829.shp".
 - A button: "Optionally load a latest file manually".
 - A checkbox: "Should old clusters automatically be marked as done?" (unchecked).

Figure 9: A previous shapefile is loaded automatically into the third column if this is found in the folder of the initial GPS file. The name follows a set structure which must include the original label that was chosen. The application will always choose that shapefile with the most recent download date.

Afterwards, the procedure continues as explained for the first cluster analysis.

Important points in this chapter:

- Uploaded GPS data must have the same starting date throughout the study, i.e., users must upload all GPS data associated with the study each time.
- The study period must be defined from the same starting date each time for the same study.
- The cluster tables must be downloaded after every time users run a successful cluster analysis.
- The chosen label must be consistent throughout the study.
- When manually uploading a latest cluster file in .xlsx format, no columns names should be changed, no additional columns should be added and no rows deleted.

4. ANALYSIS ON MULTIPLE INDIVIDUALS

For an analysis on multiple individuals, the ‘bears’ data frame must be saved locally. Again, it is good practice to now have a separate folder for each group of individuals, where each new GPS file is stored or replaced. In this folder, you will also find the output of the cluster analysis.

All steps are identical to an analysis for only one individual. The ‘Animal ID’ column chosen in the step ‘Upload GPS data’ should now include multiple animal IDs. The label set in ‘Adjust Cluster Analysis Parameters’ is now a name for the group of individuals. We call this group of three bears ‘multiplebears_demo’. Further, we chose for a buffer size of 30 metres with 2 GPS locations needed to build a cluster within the study period 1-31 May 2014 (Figure 10).

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
57.55	59.55	60.00	60.76	60.43	180.43

Figure 10: Setting the parameters for the group of bears.

The ‘Cluster Analysis Output’ again shows the same output. Differences can be seen in the interactive map, which will now show the most recent location of the three individuals that are being followed (blue pin drops). When selecting to display the track and GPS locations of the individuals, a new legend appears which shows the three individuals in different colours. The larger the points are the more recently they have been made, ending with the location of the last sent position (Figure 11).

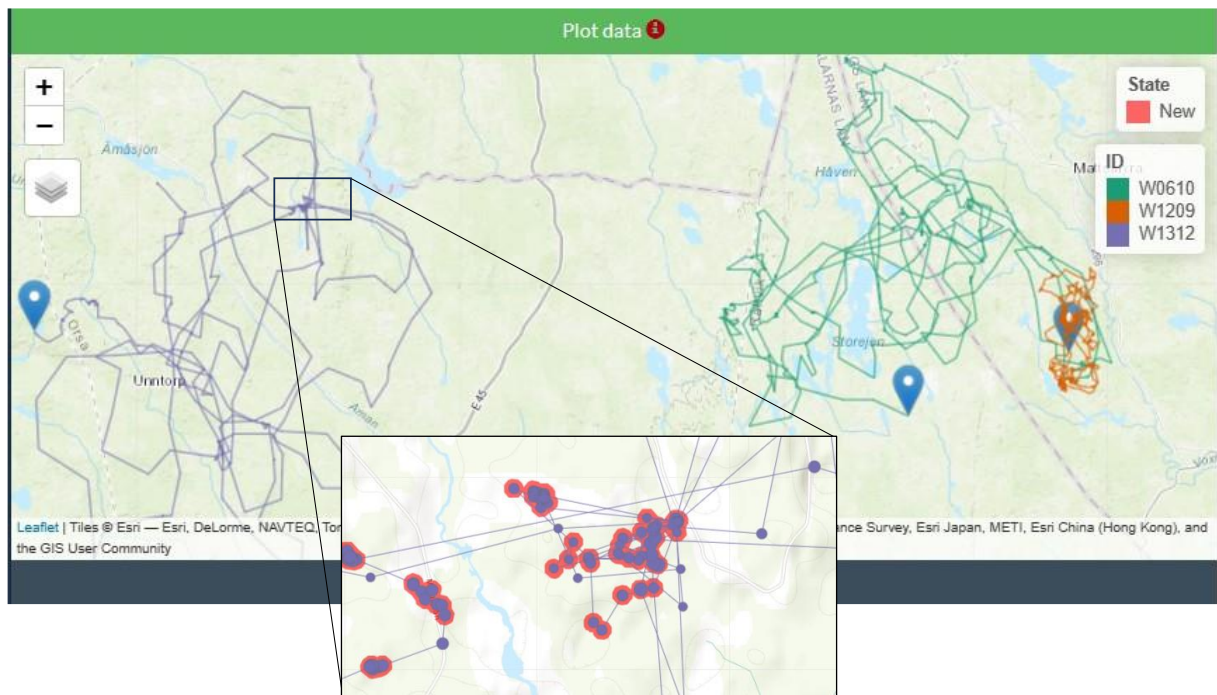


Figure 11: Cluster polygons are again all shown filled by their state. The track and GPS locations of the three bears are visualized in the interactive map with different colours. Blue pin drops show the last position of each bear.

Important points in this chapter:

- The analysis steps stay the same for one or multiple individuals. The chosen ‘Animal ID’ column in the step ‘Upload GPS data’ should include multiple animal IDs.

5. HANDLING GPS BURSTS

GPS collars are commonly designed to transmit GPS locations at regular intervals (fix rates). Therefore, the default and the most common setting in cluster analysis is to use all GPS data that are available within the study period. Some research projects incorporate the use of proximity sensors in collars to detect interactions between individuals of the same species or between different species. These proximity or burst settings can provide insights into social behaviour, group dynamics or potential contacts between species. GPS data including proximity events will therefore include GPS locations at irregular time frames, which makes an additional identifier of these clusters necessary.

Loading the data into the application, the optional time stamp filtering already gives an indication that this data sends at irregular time steps with the mean at 26.4 minutes, however with a span of minimum 0.7 to 120 minutes. The minimum time frame between GPS locations is also visible in the density plot on the day of the 26th of May 2023, where a burst of GPS locations was detected while during the rest of the period, fixes were constantly sent every half an hour (Figure 12).

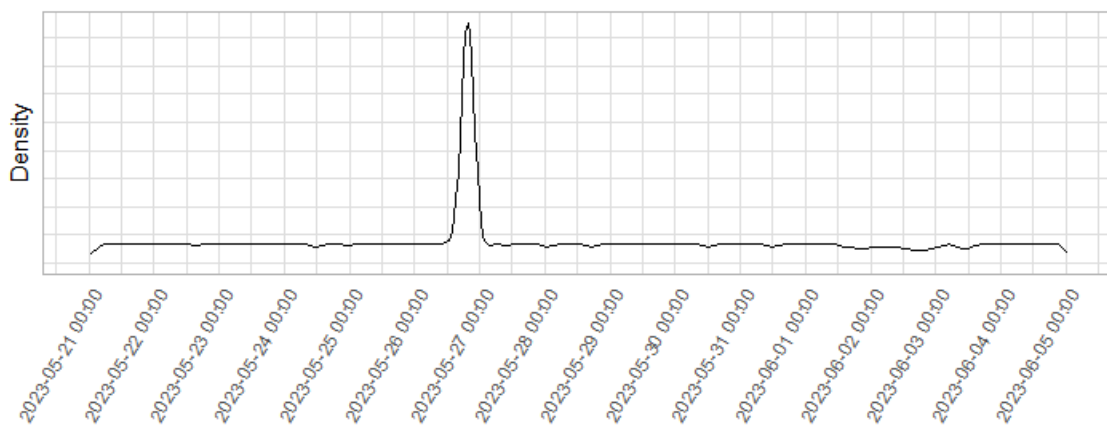


Figure 12: Density plot of send GPS positions over the time frame the individual was monitored.

Based on this information and on the research questions asked, this value can be left blank which will result in using all GPS locations for the analysis. Alternatively, for the user who has

strong deviation of the mean time difference and is only interested in generating clusters that are comparable with each other, a necessary time difference between fixes can be specified. This will result in selecting only one GPS fix per set time frame and the exclusion of all additional GPS locations from the analysis, which will be a cluster analysis comparable to the examples above.

However, if clusters of proximity data are also interesting for the research question, all datapoints can be used for the analysis. By doing so, the percent of time spent at the cluster in the clusters table output becomes more relevant; this value is calculated as the number of locations the individual could have been at the cluster between the first and last time it visited (taking the mean time difference GPS locations are taken) divided by the actual number of locations within this cluster plus one. This is only an approximation as the actual number of locations that could have been sent can always differ from the calculated mean time difference, due to for example satellite availability or environmental conditions disrupting the device's ability to transmit data on schedule. A value close to 1 implies that the individual spent most of the time between the first and last visit at this cluster, which can be related to for example bed sites used for a longer timeframe. Values close to 0 indicate that the individual only visited irregularly over a longer timeframe, such as for example baiting sites. A value larger than 1 implies that more locations than the mean number of locations possible in this timeframe were taken. This might relate to a burst of GPS positions in a short timeframe, such as proximity events. In order to see all columns related to the time spend, this box has to be ticked under tab 2: "Should the columns related to the time spent and the number of points within/outside of the cluster be displayed in the cluster table?".

Figure 13 shows the resulting cluster of the proximity event for the bear individual followed in this example. The percent of time spent at the cluster is 20.7, with 145 positions sent between 18:26 to 20:50 on the 26th of May 2023 all within the cluster and none outside of it.

While it is still important to visit clusters with a value higher than 1, it needs to be considered that these are not relatable to the clusters with the mean time difference between locations. If the research question does not cover these specific clusters, it is advisable to set a time difference in the step cluster analysis parameters and exclude these clusters that developed from GPS bursts.

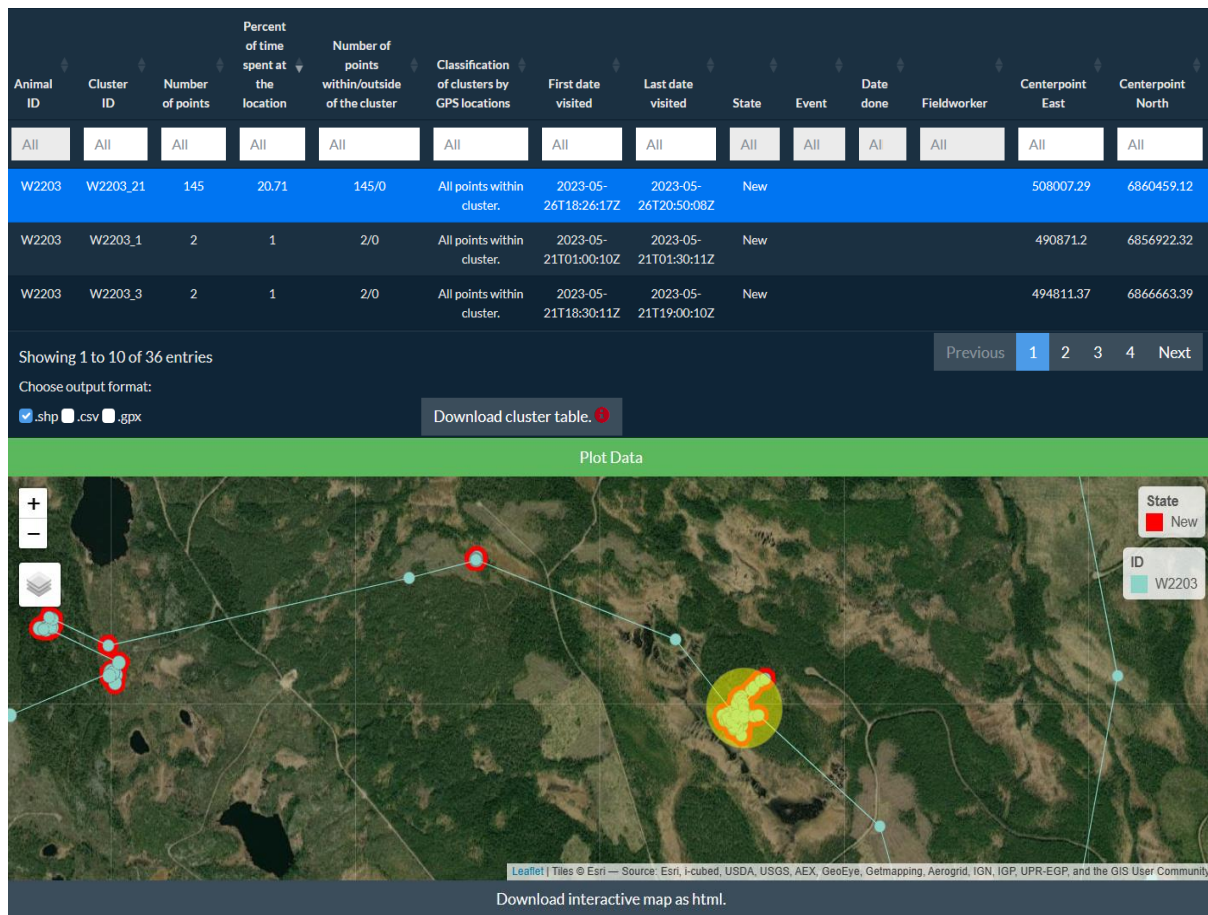


Figure 13: Output of clusters that include proximity events in the GPS data. The selected cluster was generated from a GPS burst, which can be easily recognised in the column 'percent of time spent at the location'.

Important points in this chapter:

- In order to avoid clusters from irregular GPS fix rates, check the 'the optional time stamp filtering' summary. If the minimum and maximum value are far from the mean, consider setting a fixed time difference between GPS locations. If all GPS locations are taken for the analysis be aware of the 'Percent of time spent at the location' value in the clusters table.

REFERENCES

- Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J.,
Dipert, A., & Borges, B. (2023). *shiny: Web Application Framework for R* [Computer
software]. <https://shiny.posit.co/>
- ESRI (2021). "World Topographic Map" [basemap].
<https://www.arcgis.com/home/item.html?id=30e5fe3149c34df1ba922e6f5bbf808f>.
(August 11, 2023).
- ESRI (2023). "World Imagery Map" [basemap].
https://services.arcgisonline.com/ArcGIS/rest/services/World_Imagery/MapServer.
(August 11, 2023).
- Fay, C., Guyader, V., Rochette, S., & Girard, C. (2023). *A Framework for Robust Shiny
Applications* [Computer software].
- Ordiz, A., Milleret, C., Uzal, A., Zimmermann, B., Wabakken, P., Wikenros, C., Sand, H.,
Swenson, J. E., & Kindberg, J. (2020). Individual Variation in Predatory Behavior,
Scavenging and Seasonal Prey Availability as Potential Drivers of Coexistence
between Wolves and Bears. *Diversity*, 12(9), 356. <https://doi.org/10.3390/d12090356>
- R Development Core Team. (2023). *R: a language and environment for statistical computing*
[Computer software]. R Foundation for Statistical Computing.
- Rauset, G. R., Kindberg, J., & Swenson, J. E. (2012). Modeling female brown bear kill rates
on moose calves using global positioning satellite data. *Journal of Wildlife
Management*, 76(8), 1597–1606. <https://doi.org/10.1002/jwmg.452>
- Wikenros, C., Di Bernardi, C., Zimmermann, B., Åkesson, M., Demski, M., Flagstad, Ø.,
Mattisson, J., Tallian, A., Wabakken, P., & Sand, H. (2023). Scavenging patterns of an
inbred wolf population in a landscape with a pulse of human-provided carrion.
Ecology and Evolution, 13(7). <https://doi.org/10.1002/ece3.10236>