ClusterApp - Tutorial

The goal of the *ClusterApp* is an application to guide and streamline cluster studies based on GPS data. It is developed in `shiny` (Chang et al. 2023) within R version 4.3.1 (R Core Team, 2023) and built within the `golem` framework (Fay et al. 2023).

Start the App

The following document gives a tutorial of the basic usage of the app with the dataset *"wolf"* by the Scandinavian Wolf Research Project (SKANDULV) as well as the usage of the app when applying it to multiple individuals with the dataset "bears" (Scandinavian Brown Bear Project).

The app 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, you can download the GPS data like this and safe it in a folder of your choice. It is good practice to have a separate folder per GPS collared individual or 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.

```
write.csv(ClusterApp::wolf, “path/to/folder/data.csv”)
```

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

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

The app 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 app will now open within your Internet browser.

The app follows a structure of three steps:

1. **Upload GPS data**
   - Upload the GPS data
   - Define the necessary column names
   - Adjust date format and coordinate systems
2. **Adjust Cluster Analysis Parameters**
   - Adjust all the parameters necessary for the cluster analysis
3. **Cluster Analysis Output**
   - Look at the results, adjust cluster file and download cluster -, point- and map files

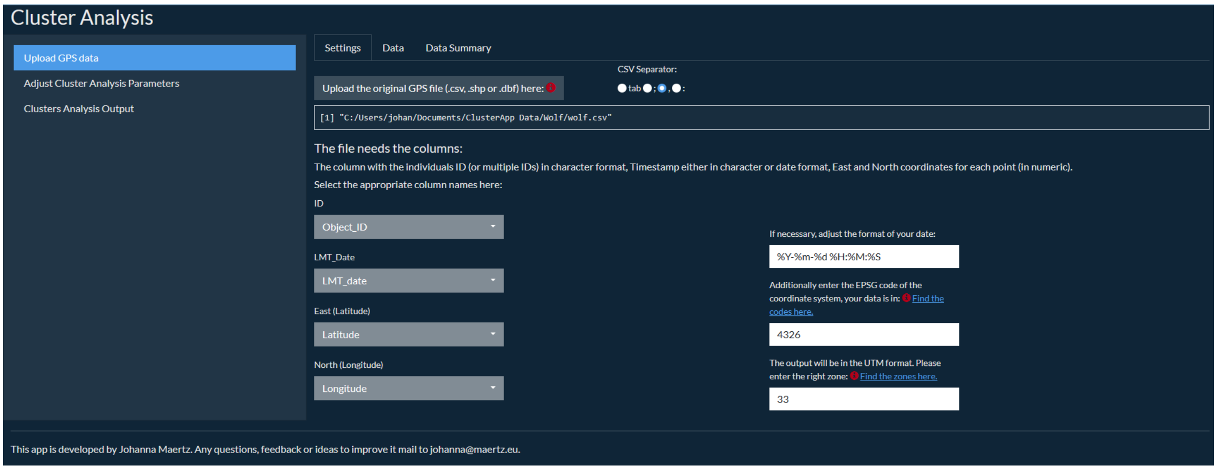
Upload GPS data

The data is loaded by selecting *"Upload the original GPS file here:"* and the path to your datafile will appear. Possible formats are *".csv", ".shp"* or *".dbf".* For the format *".csv",* you can choose the correct file separator, so that the columns are loaded correctly.

After uploading the data, the file path appears, and tabs *"Data"* and *"Data Summary"* give information about the uploaded data. These can be used to check if the data was uploaded correctly, and which column names fit as input for the ID, time stamp and coordinate variables. The timestamp has to include both date and time in one column with the default format of "**Y**ear-**m**onth-**d**ay **H**our:**M**inute:**S**econd", but this can be adjusted.

Additionally, the coordinate system your data is originally in as well as in which UTM zone the data should be downloaded, can be adjusted. Within the analysis the data is always converted to UTM, as this coordinate system has metric units, which are more intuitive within analysis involving distance measures.

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



Cluster Analysis

Upload GPS data
Adjust Cluster Analysis Parameters
Clusters Analysis Output

Settings | Data | Data Summary

Upload the original GPS file (.csv, .shp or .dbf) here: ⓘ

CSV Separator:
○ tab ○ ; ◉ , ○ :

```
[1] "C:/Users/johan/Documents/ClusterApp Data/Wolf/wolf.csv"
```

The file needs the columns:
The column with the individuals ID (or multiple IDs) in character format, Timestamp either in character or date format, East and North coordinates for each point (in numeric).
Select the appropriate column names here:

ID: Object_ID
LMT_Date: LMT_date
East (Latitude): Latitude
North (Longitude): Longitude

If necessary, adjust the format of your date: %Y-%m-%d %H:%M:%S

Additionally enter the EPSG code of the coordinate system, your data is in: ⓘ Find the codes here.
4326

The output will be in the UTM format. Please enter the right zone: ⓘ Find the zones here.
33

This app is developed by Johanna Maertz. Any questions, feedback or ideas to improve it mail to johanna@maertz.eu.

Adjust cluster analysis parameters

In the first step, it is necessary to state a unique ID of the individual the input data belongs to, and which will be used throughout the entire field season. This can either be the same identifier as already chosen as an individual ID column (by **"Upload GPS data"**) or, which is especially useful when doing cluster analysis on multiple individuals, a group ID for the group of individuals that are being studied. This is advisable, as when executing consecutive cluster analyses this unique ID will make sure that earlier output files will be found, if available, in the folder of the GPS data. For the tutorial we chose the name *"wolf_demo".*

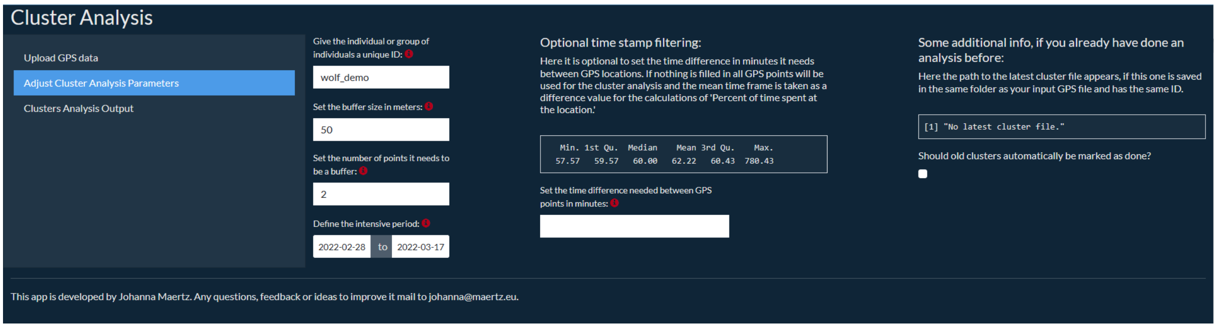
One of the important questions regarding cluster analysis and clearly related to the set research questions is setting the parameters which are necessary to identify relevant clusters. Clusters are formed by applying a buffer of *x* metres around every GPS point. If these buffers overlap (so at a maximum distance of *x* times 2), they will form clusters which will be saved as a cluster if y numbers of points are within this buffer and these points are during the intensive period that is being studied.

Therefore, it is possible to adjust the following parameters:

- Buffer size: Depending on the study species and research questions, the relevant distance criteria between movement points can be adjusted. Changing the distance, different cluster sizes are created. We keep the default set at 50 m.
- Number of GPS points: Again, depending on the study species and research questions, clusters might only be relevant when containing a minimum number of GPS points within one buffer. For this demo, it needs 2 points within a buffer to be called a cluster.
- Intensive Period: To be able to select the period during which the animal is being followed, both the starting and end date need to be set. The end date can also be in the future if the intensive period is still going on. Often the fix rate of the GPS device is adjusted to receive GPS points more frequently, therefore it is important to set the dates correctly and thus exclude points from outside this period. The intensive period of this demo data was from the 28th of February 2022 to the 17th of April 2022.

  To demonstrate a case of downloading data and then using this downloaded data in a consecutive analysis, we set the end date on the 17th of March. Now only this data will be used for the analysis and in the chapter **"Consecutive analysis"** we will further explain the column *"Some additional info, if you have done an analysis before"*.

Further optional adjustments are the filtering of the initial GPS data to only include points every x minutes. The default is to use all GPS data points that are available, which however can introduce a bias, if there are frequent changes in the fix-rate of the GPS. Therefore, the summary of the time differences between all points will be displayed, and the user can decide to specify the time between consecutive points. It is recommended to use a value close to the mean number to include maximum amount of data in the analysis. For this analysis, we keep all datapoints that are available.



Give the individual or group of individuals a unique ID: wolf_demo

Set the buffer size in meters: 50

Set the number of points it needs to be a buffer: 2

Define the intensive period: 2022-02-28 to 2022-03-17

Optional time stamp filtering:

Here it is optional to set the time difference in minutes it needs between GPS locations. If nothing is filled in all GPS points will be used for the cluster analysis and the mean time frame is taken as a difference value for the calculations of 'Percent of time spent at the location.'

```
   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 points in minutes:

Some additional info, if you already have done an analysis before:

Here the path to the latest cluster file appears, if this one is saved in the same folder as your input GPS file and has the same ID.

```
[1] "No latest cluster file."
```

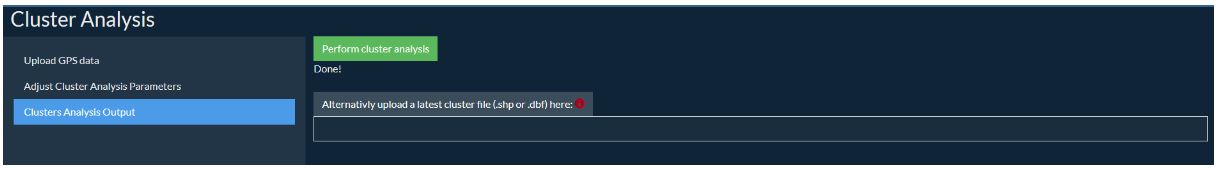
Should old clusters automatically be marked as done? □

This app is developed by Johanna Maertz. Any questions, feedback or ideas to improve it mail to johanna@maertz.eu.

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".* If the analysis ran successfully, a message would tell so and the output appears.

Alternatively, a cluster shapefile can be uploaded from one of the folders. This is an option, if the output of an earlier cluster analysis should be inspected without the need of a new analysis. When this is uploaded some following properties regarding the track and GPS points in the interactive map are not possible, as these are exclusively related to a performed cluster analysis.



Done!

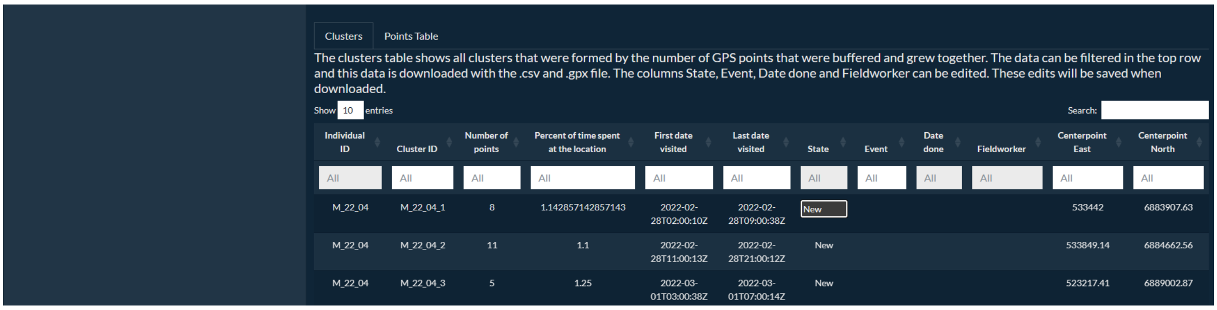
Alternativly upload a latest cluster file (.shp or .dbf) here:

During all analysis two tables are created, which will be visible in two different tabs (*"Clusters"* and *"Points Table"*).

The *"Clusters"* table appears first as the default setting and contains all the data concerning the clusters established by the analysis. The table includes columns regarding the ID of the individual (selected as "ID" column), the unique "Cluster ID" (built as ID underscore Cluster ID) and relevant information regarding the clusters (e.g., "number of points", visiting identifier). This identifier (i.e., "percent of time spent at the cluster") is calculated as the number of points the individual could have been at the cluster between the first and last time it visited (taking the mean time difference GPS points are taken) divided by the actual number of points within this cluster. This is only an approximation as the actual number of points that could have been sent can always differ from the calculated mean time difference. Close to zero indicates that the individual only visited irregularly, while an identifier close to one would suggest that the individual spent nearly the entire time frame between the first and last visit at this location. An identifier larger than one implies that more points than the mean number of points possible in this time frame were taken. This might relate to a burst of GPS positions in a short time frame, such as possible for proximity events.

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 the cluster has: all clusters that developed during this cluster analysis will be named "New", while clusters that expanded since the latest cluster analysis are stated as "Points added". In addition, the user can choose to manually adjust the state of the clusters. When clusters have been visited, the user may want to add what was found at the location, when the cluster was visited and the name of the fieldworker who visited the cluster. These adjustments can change the aesthetics of the interactive map (see below) and facilitate a streamlined data management.

The top of the data table also allows for specific filtering on all columns. This can either be useful for overview purposes within the map or downloading/looking at only specific clusters of for example specific IDs, within a specific time frame or many more.



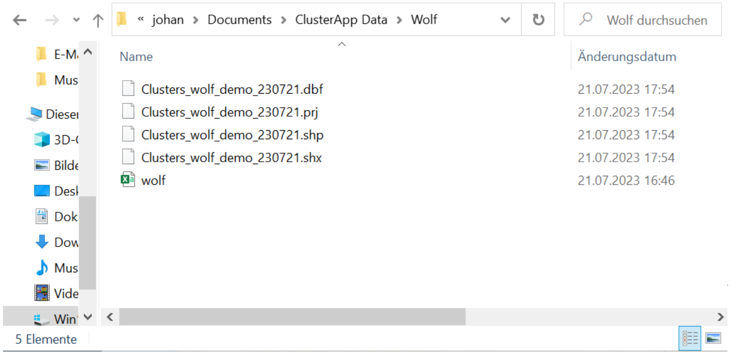
The clusters table shows all clusters that were formed by the number of GPS points that were buffered and grew together. The data can be filtered in the top row and this data is downloaded with the .csv and .gpx file. The columns State, Event, Date done and Fieldworker can be edited. These edits will be saved when downloaded.

Show 10 entries

Search:

Individual ID	Cluster ID	Number of points	Percent of time spent at the location	First date visited	Last date visited	State	Event	Date done	Fieldworker	Centerpoint East	Centerpoint North
All	All	All	All	All	All	All	All	All	All	All	All
M_22_04	M_22_04_1	8	1.142857142857143	2022-02-28T02:00:10Z	2022-02-28T09:00:38Z	New				533442	6883907.63
M_22_04	M_22_04_2	11	1.1	2022-02-28T11:00:13Z	2022-02-28T21:00:12Z	New				533849.14	6884662.56
M_22_04	M_22_04_3	5	1.25	2022-03-01T03:00:38Z	2022-03-01T07:00:14Z	New				523217.41	6889002.87

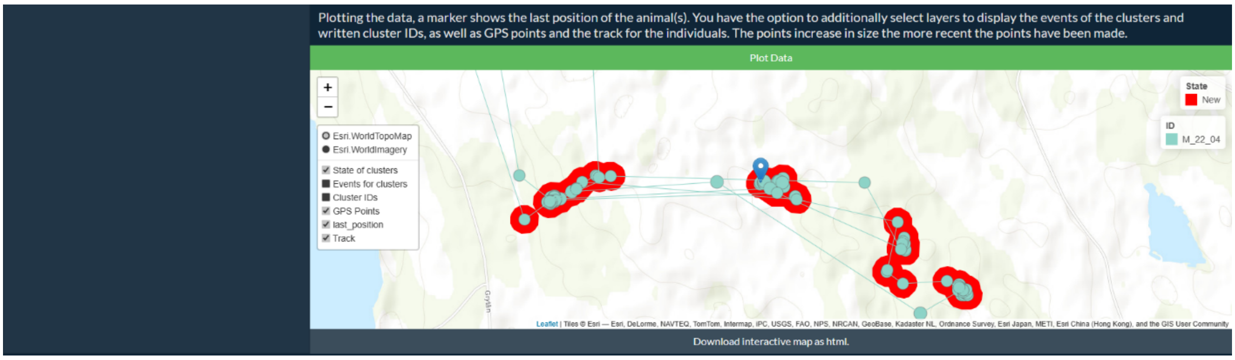
The cluster table can be downloaded in the output formats: *".shp", ".csv"* or as a GPX file *(.gpx).* Downloading the cluster file as shapefile is mandatory, if this file should be used for any consecutive analyses. This option will always download the entire cluster table in UTM and the specified zone in the folder of the original GPS data. CSV and GPX download are optional for further purposes such as easy adjustments and data management in excel or loading the file on a GPS device. The GPX is downloaded as a track in WGS 84 for direct import into any further applications. If filtering options were



Name	Änderungsdatum
Clusters_wolf_demo_230721.dbf	21.07.2023 17:54
Clusters_wolf_demo_230721.prj	21.07.2023 17:54
Clusters_wolf_demo_230721.shp	21.07.2023 17:54
Clusters_wolf_demo_230721.shx	21.07.2023 17:54
wolf	21.07.2023 16:46

applied in the table, only the filtered data frame will be downloaded as the *.csv* and/or *.gpx* file.

The *"Cluster"* table can be plotted as an interactive map by pressing *"Plot Data".* The map has the "WorldTopoMap" as a general background layer, but can be switched to "WorldImagery", 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). There are additional options to show, for example, the walking trajectory, 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.


Plotting the data, a marker shows the last position of the animal(s). You have the option to additionally select layers to display the events of the clusters and written cluster IDs, as well as GPS points and the track for the individuals. The points increase in size the more recent the points have been made.

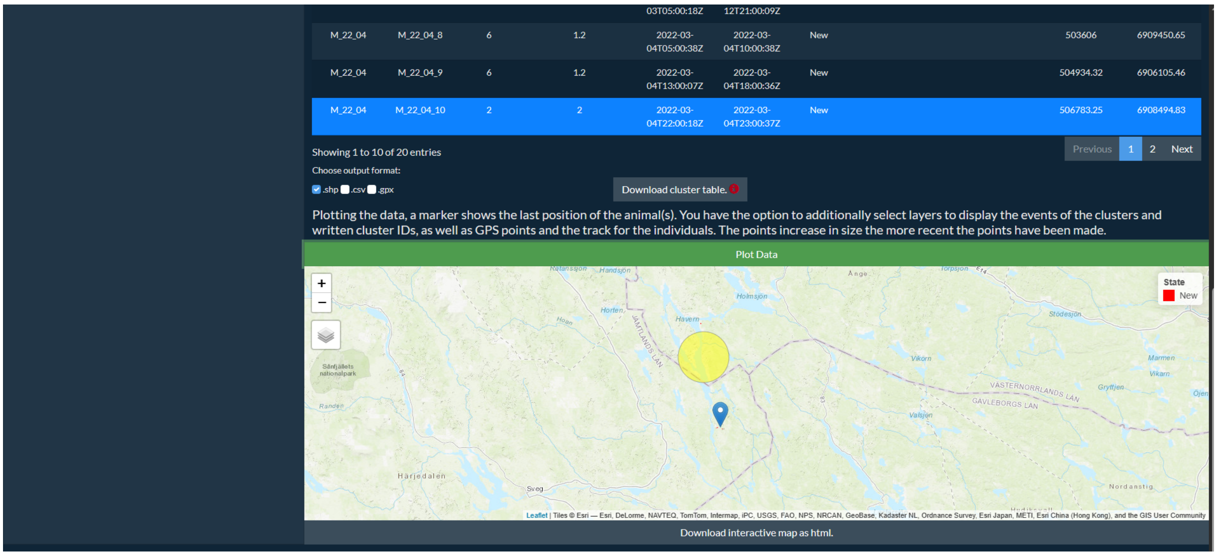
Plot Data

○ Esri.WorldTopoMap
● Esri.WorldImagery
☑ State of clusters
□ Events for clusters
□ Cluster IDs
☑ GPS Points
☑ last_position
☑ Track

Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ, TomTom, Intermap, iPC, USGS, FAO, NPS, NRCAN, GeoBase, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), and the GIS User Community

Download interactive map as html.

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 areas and understand the spatial distribution of the events that were already found surrounding new or older not visited clusters. When a row is selected within the cluster table and the map is plotted anew, this cluster is highlighted (as shown in the following screenshot). The map can be downloaded as an interactive file in "*.html*" format.



				03T05:00:18Z	12T21:00:09Z			
M_22_04	M_22_04_8	6	1.2	2022-03-04T05:00:38Z	2022-03-04T10:00:38Z	New	503606	6909450.65
M_22_04	M_22_04_9	6	1.2	2022-03-04T13:00:07Z	2022-03-04T18:00:36Z	New	504934.32	6906105.46
M_22_04	M_22_04_10	2	2	2022-03-04T22:00:18Z	2022-03-04T23:00:37Z	New	506783.25	6908494.83

Showing 1 to 10 of 20 entries

Previous 1 2 Next

Choose output format:

☑ .shp □ .csv □ .gpx

Download cluster table.

Plotting the data, a marker shows the last position of the animal(s). You have the option to additionally select layers to display the events of the clusters and written cluster IDs, as well as GPS points and the track for the individuals. The points increase in size the more recent the points have been made.

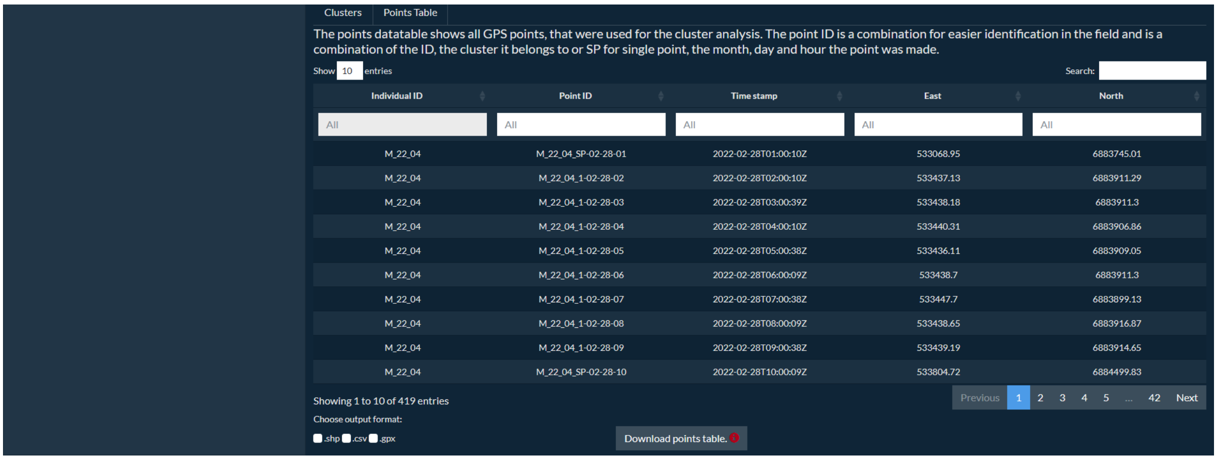
Plot Data

Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ, TomTom, Intermap, iPC, USGS, FAO, NPS, NRCAN, GeoBase, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), and the GIS User Community

Download interactive map as html.

The tab *"Points"* table gives an overview of all GPS points that were used for the cluster analysis, and which are selectable to be highlighted on the interactive map ("*GPS points*"). The point ID is a combination for easier identification in the field and consists of the individual's ID, the cluster number it belongs to or SP for single points and then month, day and hour the point was made. This way, the field personnel can get a fast overview of the distribution of the points within the visited cluster and might therefore by the experience of the species behaviour know what the event might be. The table can also be filtered to only show certain selections of the data. Again, the data can be downloaded as shapefile ("*.shp*") or comma separated values ("*.csv*") in UTM and the specified zone or as GPX file

(“*.gpx”*) in WGS84. The shapefile will download all data, while the “*.csv”* and “*.gpx”* file will only download the selected data. This data is especially useful to load on a handheld GPS device or for any other form of orientation in the field.



The points datatable shows all GPS points, that were used for the cluster analysis. The point ID is a combination for easier identification in the field and is a combination of the ID, the cluster it belongs to or SP for single point, the month, day and hour the point was made.

Show 10 entries

Search:

Individual ID	Point ID	Time stamp	East	North
All	All	All	All	All
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
M_22_04	M_22_04_1-02-28-04	2022-02-28T04:00:10Z	533440.31	6883906.86
M_22_04	M_22_04_1-02-28-05	2022-02-28T05:00:38Z	533436.11	6883909.05
M_22_04	M_22_04_1-02-28-06	2022-02-28T06:00:09Z	533438.7	6883911.3
M_22_04	M_22_04_1-02-28-07	2022-02-28T07:00:38Z	533447.7	6883899.13
M_22_04	M_22_04_1-02-28-08	2022-02-28T08:00:09Z	533438.65	6883916.87
M_22_04	M_22_04_1-02-28-09	2022-02-28T09:00:38Z	533439.19	6883914.65
M_22_04	M_22_04_SP-02-28-10	2022-02-28T10:00:09Z	533804.72	6884499.83

Showing 1 to 10 of 419 entries

Choose output format:

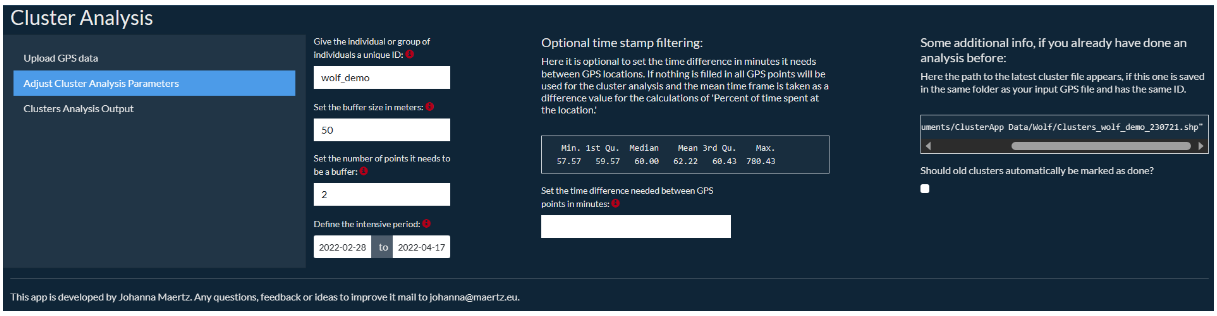
□ .shp □ .csv □ .gpx

Consecutive analysis

So, the first analysis was done. After the shapefile of the clusters is downloaded, the next analysis that is done will be a consecutive analysis. The procedure starts the same, by uploading the GPS data from the folder, setting the correct column names and the settings.

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. If the analysis is run for the first time or the program does not detect any shapefiles with the exact string it is searching for (“Clusters underscore ID underscore date point shp”) within this folder, the message will read “*No latest cluster file”* and will run the analysis as if done for the first time.

When executing a consecutive cluster analysis, it is important to check if a latest cluster file is found, as this file will make sure the earlier established cluster ID’s remain constant even with additional points that might increase, combine or create new clusters. Therefore, it is important to always download the shapefile. There can be multiple old shapefiles in the sub-folder as the program will search for the most recent file. You can optionally automatically mark all clusters from the old cluster file as “Done”, otherwise this can be done manually in the “Cluster Analysis Output” tab.



Give the individual or group of individuals a unique ID: wolf_demo

Set the buffer size in meters: 50

Set the number of points it needs to be a buffer: 2

Define the intensive period: 2022-02-28 to 2022-04-17

Optional time stamp filtering:

Here it is optional to set the time difference in minutes it needs between GPS locations. If nothing is filled in all GPS points will be used for the cluster analysis and the mean time frame is taken as a difference value for the calculations of 'Percent of time spent at the location.'

```
   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 points in minutes:

Some additional info, if you already have done an analysis before:

Here the path to the latest cluster file appears, if this one is saved in the same folder as your input GPS file and has the same ID.

```
uments/ClusterApp Data/Wolf/Clusters_wolf_demo_230721.shp"
```

Should old clusters automatically be marked as done?

□

This app is developed by Johanna Maertz. Any questions, feedback or ideas to improve it mail to johanna@maertz.eu.

Afterwards, the procedure goes on as explained for the first cluster analysis. Congratulations, you executed your first and second cluster analysis for a wolf in Sweden!

Analysis on multiple individuals

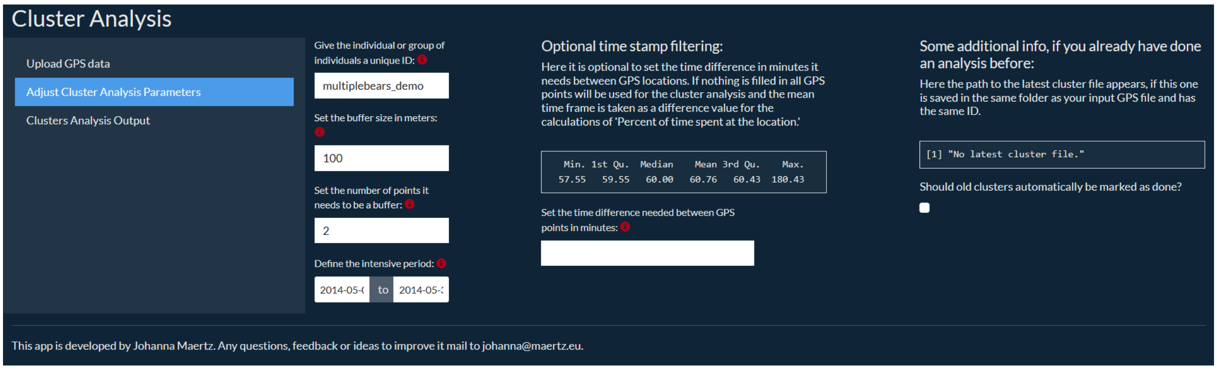
If you want to try out an analysis on multiple individuals,, you can download the GPS data and safe it in a folder of your choice. Again, it is good practice to now have a separate folder per group of individuals, where each new GPS file is stored or replaced. In this folder, you will also find the output of the cluster analysis.

```
write.csv(ClusterApp::bears, “path/to/folder/data.csv”)
```

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

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

All steps are identical to an analysis for only one individual. The column chosen in the step **Upload GPS data** should now include multiple IDs. The ID name 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 choose for a buffer size of 100 metres with 2 points needed to build a cluster within the month of May 2014.



Cluster Analysis

Give the individual or group of individuals a unique ID: multiplebears_demo

Set the buffer size in meters: 100

Set the number of points it needs to be a buffer: 2

Define the intensive period: 2014-05-(to 2014-05-:

Optional time stamp filtering:

Here it is optional to set the time difference in minutes it needs between GPS locations. If nothing is filled in all GPS points will be used for the cluster analysis and the mean time frame is taken as a difference value for the calculations of 'Percent of time spent at the location.'

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

Set the time difference needed between GPS points in minutes:

Some additional info, if you already have done an analysis before:

Here the path to the latest cluster file appears, if this one is saved in the same folder as your input GPS file and has the same ID.

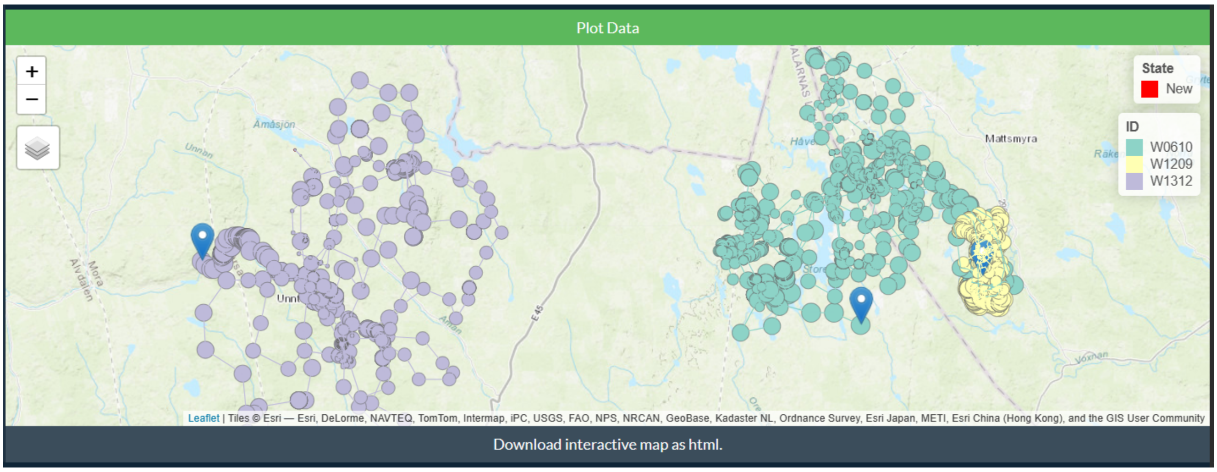
```
[1] "No latest cluster file."
```

Should old clusters automatically be marked as done?

□

This app is developed by Johanna Maertz. Any questions, feedback or ideas to improve it mail to johanna@maertz.eu.

The **Cluster Analysis Output** again shows the same output. Differences can be seen in the interactive map, which will now show the last location of the three individuals that are being followed. When selecting to display the track and GPS points of the individuals, a new legend appears which shows the three individuals in different colours. The larger the points are the more recents they have been made, ending with the location of the latest position.



Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ, TomTom, Intermap, iPC, USGS, FAO, NPS, NRCAN, GeoBase, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), and the GIS User Community

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_. R package version 1.7.4.1, <https://CRAN.R-project.org/package=shiny>.

Fay C, Guyader V, Rochette S, Girard C (2023). _golem: A Framework for Robust Shiny Applications_. R package version 0.4.1, <https://CRAN.R-project.org/package=golem>.

R Core Team (2923). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.