Collar Cleaning User Guide

Updated 2024-04-19

Sections

1. Processing GPS Data (Collar Cleaning)
2. Filter and Rarify Data

Collar Cleaning

1. PROCESSING GPS DATA

Processing of GPS data is brand specific.

The user location where the download happens and the time of year (DST or not) can affect local time processing for some models.

It is recommended that the user keep a log of When/Where the collar data is downloaded.

PROCESSING GPS DATA AKA *COLLAR CLEANING*

Folder Organization

- Programs assume a hierarchy
  - C/D/E
    - Wolf-Projects
      - Data
        - » locdataAll-CCv31
          - Original (cleaned collar data all in the same folder)
- cc program will generate
  - “<filename>_analysis” as input to cluster program
  - “<filename>_location” as input to merge programs
- If original files are organized within sub-folders then csvCopy.py program can be used
  - Note: change the path names to flatten all of the collar files to a single folder

cc_v35.py

- cc == CollarClean
- Note: if the files are edited with Excel versus notepad++ it is important to set the column formats of any date columns to YYYY-MM-DD
- To run program: (Python 2.7.15)
  - Copy program into folder containing raw csv files
  - Open with IDLE
  - Run Module
- Review and keep log file generated
- Change Date Range if necessary
  - # Global variables for date range selection
  - # default to no selection - program always checks for date range file
  - daterange_parsed= False - if external date range file parsed
  - daterange_flag = False - if external date range defined
  - daterange_start = "2016-01-01“ - default, can be overridden by external file
  - daterange_end = "2018-12-31“ - default, can be overridden by external file

cc_v35.py

- # Date adjustment flags – important issue
  - televiltDSTpresent = 1   # set to 1 if the file was ***downloaded*** during DST
  - televiltDSTremove = 1 # set to 1 to remove DST

```
–     if (televiltDSTpresent == 1):
–         print '   >>> '
–         print '   >>> DST Present'
–         print '   >>> '
–         if (televiltDSTremove == 1):
–           print '   >>> '
–           print '   >>> DST Removed'
–           print '   >>> '
–         else:
–           print '   >>> '
–           print '   >>> DST NOT Removed'
–           print '   >>> '
–       else:
–         print '   >>> '
–         print '   >>> DST NOT Present'
–         print '   >>> '
```

Time Terminology

- UTC – is the Universal Coordinated Time
- GMT – is a time zone in England
- UTC == GMT w.r.t. actual time
- DST
  - In regions where daylight saving time is used, it commences on the second Sunday of March, and standard time restarts on the first Sunday in November.
- Local Time – may be affected by DST
- LMT: not commonly used
  - Keep headers the same and assume LMT == LocalTime (LT)

Date/Time Fields – cc_v35

Collar Type	Date/Time Local	Local DST	Date/Time GMT	Date/Time in _Location YYYY-MM-DD	Date/Time in _Analysis (Input to Cluster Analysis) MM/DD/YYYY	Note
ATS	Y	N	N	Local	Local	
Lotek	Y	N	Y	Local	Local	
Sirtrack	N	N	Y	Calc Local from GMT	Calc Local from GMT	
Televilt	Y	Y (based on date and location of actual download)	N	Calc Local (no DST)	Calc Local (no DST)	LOCATION MATTERS! This code assumes download in same timezone as collar
Telonics	N	N	Y	Calc Local from GMT	Calc Local from GMT	

Set flags in cc program

cc program limitations (cc_v35)

- Folder names can not contain “_”
- Folder names cannot contain a collar type   e.g. Televilt
- Multiple Sirtrack Formats
  - Need to resave CSV UTC date column in YYYY-MM-DD format
- Multiple Televilt Formats
- Lotek: ignore original dateTimeGMT field
  - Instead use gmtDate, gmtTime
- Hard coded Lat and Long checks may need to be tweaked based on study area
  - Check lat outside range [50,60] ➔ error
  - Check long outside range [-108,-97] ➔ warning

The cleaned collar data _location and _analysis files may not match 1:1 given the slightly different rules in the cc program.

This step will result in the _location files being further cleaned then filtered and rarified into 4 sub-folders

2. FILTER, CLEAN & RARIFY DATA

Manual Edits

- Manual Editing of the cleaned files may be necessary given the data collection methods adopted for the research projetc.

- Case 1: Initial period after Collaring
  - Remove all rows up to the following day GMT time (> 24 hours from local time of collaring)

- Case 2: Mortality/Collar drop-off
  - Remove all rows on that same day – local time

Program: rarifyWloc_v5.R

- Create: aaaControlFile.csv
- Run program
- Review and save R log into a file
- Check folders (R1/R2all/R2even/R2odd)

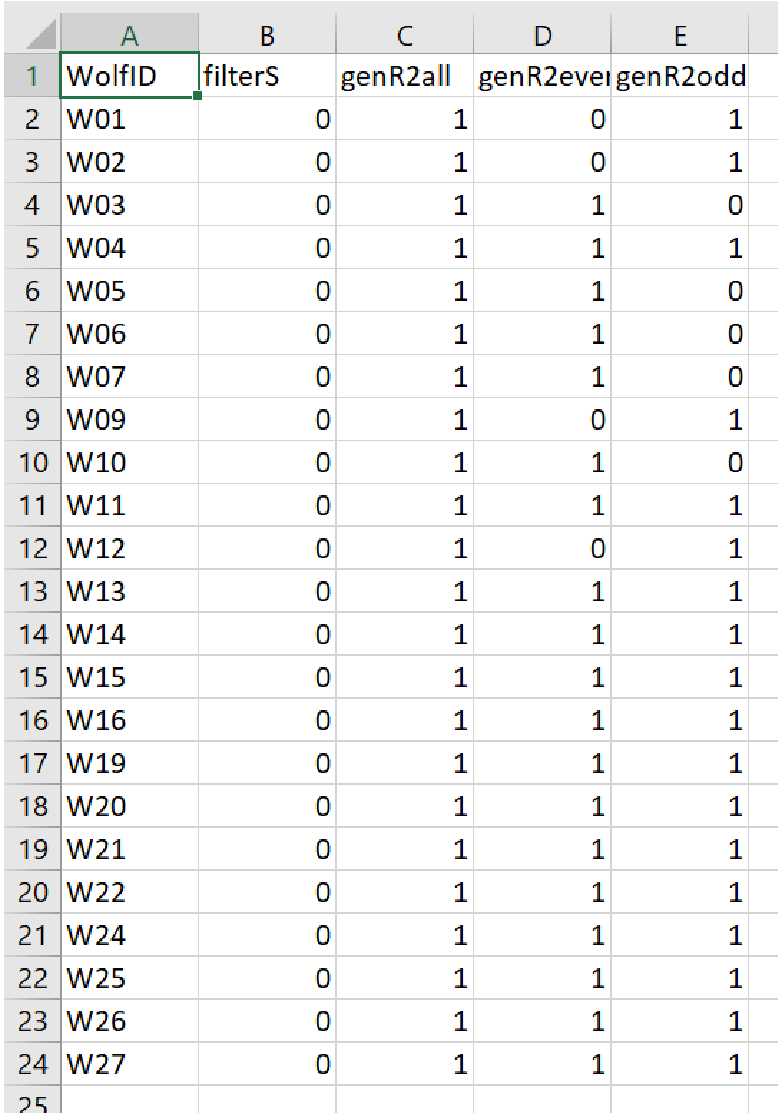
aaaControlFile

- WolfID: select wolf id
- FilterS: remove any reading <= FilterS (seconds)   default == 60s (deferred)
- gen2Hall: generate 2 hour rarified data using all points (flag (0 or 1))
- gen2Heven: generate 2 hour rarified data using even points (flag (0 or 1))
- gen2Hodd: generate 2 hour rarified data using odd points (flag (0 or 1))

- Implied:
  - 1HourLowS == 1800
  - 1HourHiS == 2HourLowS = 5400

- Dropped: Epsilon to detect Even/Odd
  - Only 12 points were affected by rounding with Epsilon == 15 minutes

aaaControlFile – sample input



	A	B	C	D	E
1	WolfID	filterS	genR2all	genR2ever	genR2odd
2	W01	0	1	0	1
3	W02	0	1	0	1
4	W03	0	1	1	0
5	W04	0	1	1	1
6	W05	0	1	1	0
7	W06	0	1	1	0
8	W07	0	1	1	0
9	W09	0	1	0	1
10	W10	0	1	1	0
11	W11	0	1	1	1
12	W12	0	1	0	1
13	W13	0	1	1	1
14	W14	0	1	1	1
15	W15	0	1	1	1
16	W16	0	1	1	1
17	W19	0	1	1	1
18	W20	0	1	1	1
19	W21	0	1	1	1
20	W22	0	1	1	1
21	W24	0	1	1	1
22	W25	0	1	1	1
23	W26	0	1	1	1
24	W27	0	1	1	1
25					

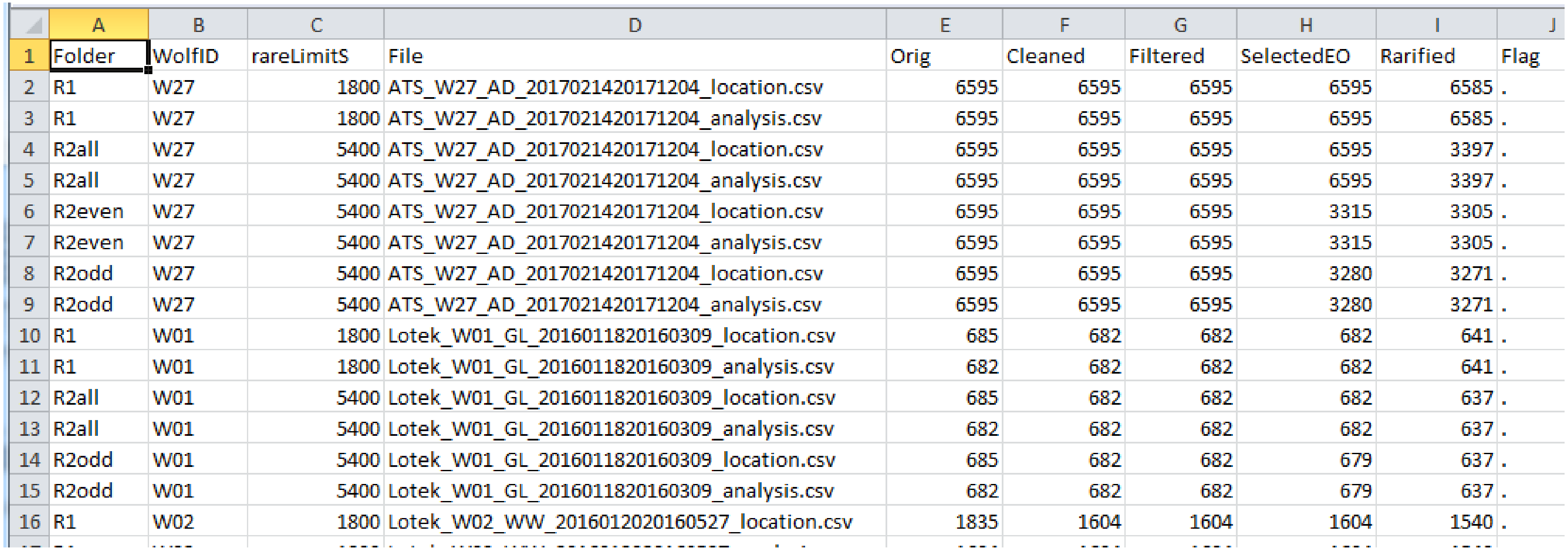
Folders

- Folders Used:
  - R1: rarified to 1 hour
  - R2all: rarified to 2 hours
  - R2even: only even points, rarified to 2 hours
  - R2odd: only odd points, rarified to 2 hours

FILTER RULES

- Orig == original lines read from file
- Cleaned == Orig – (Lat/Long out of range)
- Filtered == not implemented
- SelectedEO == num of points (even or odd)
- Rarified ==
  - – num of points with step size > rareLimits
- Flag == “.” iff location file == analysis file

aaaRarityLog.csv – sample output



	A	B	C	D	E	F	G	H	I	J
1	Folder	WolfID	rareLimitS	File	Orig	Cleaned	Filtered	SelectedEO	Rarified	Flag
2	R1	W27	1800	ATS_W27_AD_2017021420171204_location.csv	6595	6595	6595	6595	6585	.
3	R1	W27	1800	ATS_W27_AD_2017021420171204_analysis.csv	6595	6595	6595	6595	6585	.
4	R2all	W27	5400	ATS_W27_AD_2017021420171204_location.csv	6595	6595	6595	6595	3397	.
5	R2all	W27	5400	ATS_W27_AD_2017021420171204_analysis.csv	6595	6595	6595	6595	3397	.
6	R2even	W27	5400	ATS_W27_AD_2017021420171204_location.csv	6595	6595	6595	3315	3305	.
7	R2even	W27	5400	ATS_W27_AD_2017021420171204_analysis.csv	6595	6595	6595	3315	3305	.
8	R2odd	W27	5400	ATS_W27_AD_2017021420171204_location.csv	6595	6595	6595	3280	3271	.
9	R2odd	W27	5400	ATS_W27_AD_2017021420171204_analysis.csv	6595	6595	6595	3280	3271	.
10	R1	W01	1800	Lotek_W01_GL_2016011820160309_location.csv	685	682	682	682	641	.
11	R1	W01	1800	Lotek_W01_GL_2016011820160309_analysis.csv	682	682	682	682	641	.
12	R2all	W01	5400	Lotek_W01_GL_2016011820160309_location.csv	685	682	682	682	637	.
13	R2all	W01	5400	Lotek_W01_GL_2016011820160309_analysis.csv	682	682	682	682	637	.
14	R2odd	W01	5400	Lotek_W01_GL_2016011820160309_location.csv	685	682	682	679	637	.
15	R2odd	W01	5400	Lotek_W01_GL_2016011820160309_analysis.csv	682	682	682	679	637	.
16	R1	W02	1800	Lotek_W02_WW_2016012020160527_location.csv	1835	1604	1604	1604	1540	.

rarifyWloc: File Processing I

- Read control file
- For each input file (_location)
  - Determine Wolf ID
  - Apply existing checks
    - # Remove rows where Lat == 0 or Long == 0
    - alldata = alldata[alldata$Latitude!=0,]
    - alldata = alldata[alldata$Longitude!=0,]
    - # remove NA rows
    - alldata = alldata[!is.na(alldata$Latitude),]
    - alldata = alldata[!is.na(alldata$Longitude),]
    - # remove out of range rows
    - alldata = alldata[alldata$Longitude <= -97.0,]
    - alldata = alldata[alldata$Longitude >= -108.0,]
  - Check _location and _analysis files are the same length
  - If filterS <> 0
    - Apply FilterS

Deferred – not needed

  - Create an OrigFixNum – so that the points can be traced back to the original points
  - This is important because Cluster Analysis needs FixNum's in strict numeric order

rarifyWloc: File Processing II

- For each input file (_location, _analysis)
  - Rarify 1 hour
    - Use 1HLowS as limit
    - Write R1 files
  - If gen2Hall
    - Rarify to 2H using 2HourLowS as lower bound
    - Write R2all   files
  - If gen2Heven
    - Select even hours
    - Rarify to 2H using 2HourLowS as lower bound
    - Write R2even files
  - If gen2Hodd_f
    - Select odd hours
    - Rarify to 2H using 2HourLowS as lower bound
    - Write R2odd   files

checkWloc_v1: Check the Rarified Files

- Program checks: Id, Lat, Long

- Run program

- Review and save R log in a file

- As an extra check we want to ensure that the corresponding location and analysis files match exactly