PITcleanr\_lite: Workflow Documentation

Last modified: 8/18/2022

Modified by: Mark Roes

Overview

The following is a step-by-step guide to compile PIT-tag mark and observation data and process them into capture histories useful for analysis of fish or animal movement, growth, survival, etc. The workflow uses a draft version of the GitHub repository [PITcleanr\_lite](https://github.com/Mount-Hood-Environmental/PITcleanr_lite) which is intended to be a user-friendly, stand-alone group of R functions and scripts that leverages functionality and is a companion to the previously developed [PITcleanr](https://github.com/KevinSee/PITcleanr) R package. PITcleanr\_lite is designed to “compress” large, sometimes unwieldy PIT-tag observation datasets for a given list of PIT-tagged fish into observation records and capture histories that are more manageable to aid in fisheries analyses. PITcleanr\_lite accommodates observation data downloaded from either the Columbia Basin PIT Tag Information System ([PTAGIS](https://www.ptagis.org/)) or from the Biomark, Inc. [BioLogic](https://data3.biomark.com/) web portal. PTAGIS is the centralized database for PIT-tagged fish in the Columbia River basin and houses observation data for many of the more permanent PIT-tag arrays located throughout the region. However, PTAGIS does not contain most observation data from either 1) observation sites outside of the Columbia River basin or 2) temporary or project-specific detection sites (e.g., litz cords). These observations can instead be found in BioLogic™ at least for Biomark, Inc. installed arrays and infrastructure.

In this document, we provide an example of the use of PITcleanr\_lite (2022) to compile a list of unique PIT-tagged fish that are marked and/or released at rotary screw traps (RST) in the Lemhi River watershed. Our focus here is on juvenile Chinook salmon *Oncorhynchus tshawystscha* and steelhead *O. mykiss* released at two RSTs located just upstream from the confluence of Hayden Creek and the upper Lemhi River: the Upper Lemhi River Rotary Screw Trap (LEMTRP) and the Hayden Creek Rotary Screw Trap (HYDTRP). Each of these traps are intended to monitor the emigration timing, abundance, and survival from subpopulations spawning upstream; juveniles are then tracked through the lower Lemhi River (downstream from Hayden Creek until the confluence with the Salmon River) including through multiple stream habitat rehabilitation projects that are “wired” with litz cords to monitor use of e.g., newly created or restored side channels. Juveniles can also be interrogated or tagged and released at the Lower Lemhi River RST (LLRTP) located below may of the stream rehabilitation projects before leaving the Lemhi River.

Required and Recommended Items

The following workflow requires:

1. Access to Biomark’s BioLogic™ web portal (<https://data3.biomark.com/>)
2. A recent version of R software installed (<https://cran.r-project.org/>)
3. A local copy of the PITcleanr\_lite GitHub repo (<https://github.com/Mount-Hood-Environmental/PITcleanr_lite>)

We additionally recommend the use of the following:

1. A PTAGIS account to create and execute queries (<https://www.ptagis.org/>)
2. [RStudio](https://www.rstudio.com/) as a useful environment for R an.

Tag or Mark Data

In our example, we use two PTAGIS queries to construct a list of unique PIT-tagged juveniles tagged and/or released at the LEMTRP and HYDTRP RSTs. Alternatively, a user could provide their own list of PIT-tagged individuals in a .txt or .csv file format. To complete the PTAGIS queries, the user will need to do the following:

1. Navigate to <https://www.ptagis.org>
2. Login (or Register for a new account, if necessary)
3. Navigate to the “Advanced Reporting Home Page”
4. Click “Create Query Builder2 Report”

This will take you to a list of standard queries that are provided by PTAGIS. We will begin with a “Tagging Detail” query.

### Tagging Detail Query

First, we will execute a “tagging detail” query in PTAGIS to identify all newly PIT-tagged fish that are released at RSTs within the Lemhi River.

1. Report type: Tagging Detail
2. Attributes – add the following attributes to the default:
   1. Length
   2. Weight
3. Filters
   1. **4** Capture Method = Screw Trap
   2. **9** Mark Site Subbasin = 17060204:Lemhi
   3. **10** Mark Year = 2020 – Present
   4. **16** Species = Chinook, Steelhead
4. Save query
   1. Give the query a descriptive name, such as “MarkDat\_Lemhi”
   2. **Important:** The query must be saved as “static”. Be sure to check the box as in the highlighted image below



1. Run Report

Recapture Query

Recapture query: All recaptures at Lemhi RSTs.

1. Report type: Recapture Detail
2. Attributes to add:
   1. Recap length
   2. Recap weight
3. Filters
   1. **10** Mark Site Subbasin = 17060204:Lemhi
   2. **19** Recap Capture Method = Screw Trap
   3. **27** Recap Year = 2020 – Present
   4. **28** Species = Chinook, Steelhead
4. Save query
   1. Give the query a descriptive name, such as “RecapDat\_Lemhi”
   2. **Important:** The query must be saved as “static”. Be sure to check the box as in the highlighted image below

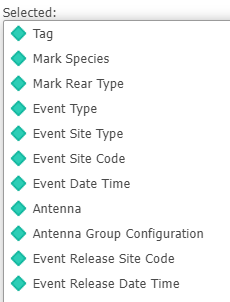


1. Run Report

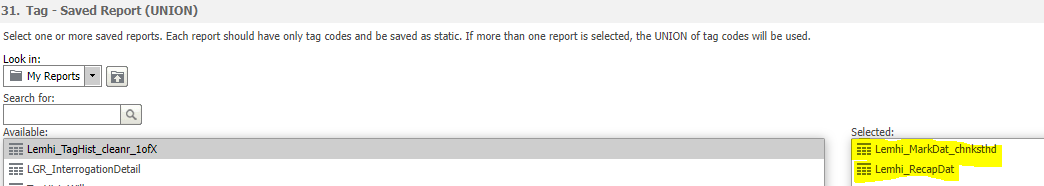
Tag History Query

This query filters the entire PTAGIS database to tags that are recorded in queries 1 or 2 and returns the associated complete tag histories to be downloaded.

1. Report type: Complete Tag History
2. Attributes: Select the attributes listed below



1. Filters
   1. **6** Event Site = HYC, LRW, EVU, S2I, S2O, BHC, S3B, S3A, EVL, LLRTP, LLR
   2. **31** Tag – Saved Report (UNION)
      1. In this window, navigate to “My Reports” and select the 1. Mark query and 2. Recapture tag history query created above. See image below for details.



1. Save report. Unlike the first two, this report can be saved as “prompted”
2. Run Report
   1. **Important: Running this query for all event sites will likely cause PTAGIS to error out due to too many records. Recommend running the report once with event site set to “HYC” only, then again for all other sites. HYC has nearly 1 million records from 2020-2022 alone.**
3. Export
   1. Whole report
   2. CSV file format
   3. Remaining as default
   4. Export
4. Name the downloaded file using the following naming convention: “TagHist\_basin\_sitenames\_year.csv” e.g., “TagHist\_Lemhi\_HYC\_20-22.csv”
5. Save the downloaded file within the “input/PTAGIS\_data” folder in PITcleanr\_lite.

BioLogic™

Observation data

PIT tag detection data downloaded from the BioLogic™ database should consist of 5 columns: site, tag, detected, reader, and antenna. BioLogic™ data downloads are saved to the “input/biologic\_data” folder and must include the string “biologic” in the filename (e.g., “0LL\_tagobs\_biologic\_07302022.csv”).

Site metadata

Site metadata information is contained in “input/site/site\_metadata.csv”. Any modifications or additions must be made within this file.

Configuration

Filtering out Tags

If any tags need to be filtered out from results (e.g., test tags), then they need to be entered into the “input/metadata/filter\_tags.csv” file. List all tags that will be filtered out in the “tag\_num” column.

Node Configuration

A node configuration file is required to convert BioLogic™ reader numbers to array names. The file is located at “input/metadata/node\_config.csv”. Modify the file as necessary to assign reader numbers to nodes.

Directionality

A directionality component can be added to the cleaned and compressed tag observation data. To do so, the directionality file (“input/metadata/node\_direction.csv”) must be configured. The configuration file includes the following columns:

* **parent:** A node directly downstream of the corresponding node in the “child” column. An alternative way to think of this is by stream order – lower stream orders (tributaries) are the “children” of mainstem reaches. This is because PITcleanr was constructed to handle both downstream migration of smolts as well as upstream migration of adults. Therefore, adults move upstream from parent to child reaches, and juveniles move downstream from child to parent reaches. The upstream or downstream direction of fish movement is defined by the “direction” argument in the **addDirectionWrap()** function of the workflow (see *Running the Scripts*).
* **child:** A node directly upstream of the node listed in the “parent” column
* **parentOrder:** Describes the hierarchy of nodes, ascending in upstream order
* **parent\_group:** Parent argument used to group together individual nodes in complex systems (e.g., Henry’s reach)
* **child\_group:** Child argument used to group together individual nodes in complex systems (e.g., Henry’s reach)

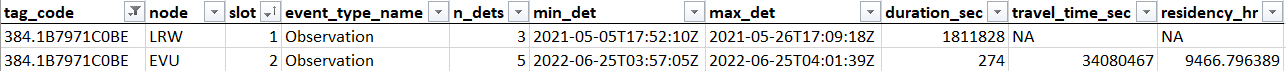
Running the Scripts

1. In the PITcleanr\_lite parent folder, open the “PITcleanr\_lite” R project file using R or Rstudio.
2. Within R, open the “Workflow.r” script located within the “scripts” folder. This can be done by navigating to “File → Open File”, ctrl+o, or using the “Files” tab in the files and help pane.
3. Highlight the entire script and click the “run” button.
   1. Packages listed under the “load packages” will automatically be downloaded from the CRAN repository if they have not been installed. These are required for the script to run.
   2. **Note:** Some function arguments can be modified for different types of outputs
      1. **AddDirectionWrap()**
         1. group\_nodes = T will use the “parent\_group” and “child\_group” columns from “node\_direction.csv” instead of individual nodes
         2. build\_diagram = T will output a directionality diagram located in “output/figures”
         3. generate\_map = T will generate a map of detection nodes using downloaded NHD flowlines
         4. downstream\_site states the furthest downstream site to map when “generate\_map = T”
         5. The “direction” argument states the direction of fish movement. direction = ‘d’ is for downstream fish movement (i.e. outmigration) and direction = ‘u’ is for upstream migration (i.e. adult returns)
4. This will output three .csv files with cleaned tag observation data, and a fourth optional .csv with the final paths for all detected fish. These are located in the “output” folder.
   1. “TagObs\_Compressed\_YY-MM-DD.csv” contains cleaned and compressed tag observation data. Each row describes the number and duration of observations for a single tag at one array.
   2. “TagObs\_Wide\_YY-MM-DD.csv” summarizes compressed tag observation data and pivots it wide. Each row is a unique tag and the associated number of detections at all observation sites.
   3. “TagObs\_Directionality\_YY-MM-DD.csv” contains tag records with associated movement direction. Observations are limited to arrays/nodes are listed in the “input/metadata/node\_direction.csv” configuration file.
5. **Optional:** There is a code chunk following the primary workflow that subsets the “TagObs\_Directionality” dataset to the final detections for each tag and writes it to “TagObs\_FinalPaths\_YY-MM-DD.csv”. This provides the complete detection path for each fish through nodes specified in the “node\_direction” file.

Data Interpretation

“TagObs\_Compressed\_YY-MM-DD.csv”

This file contains all tag observation data. Each record is the number and duration of observations for a single tag at one node. For example, the first record below reads as: The first observation (slot = 1) for tag “384.1B7971C0BE” was at node LRW. There were 3 detections from 2022-05-05 at 17:52:10Z to 2022-05-26 at 17:09:18Z (which is 1,811,828 seconds). Then, the second node that the fish was detected at was EVU (slot = 2). The amount of time that elapsed between the final detection at LRW and first detection at EVU was 9,466 hours (or 34,080,647 seconds). As default, each detection of a unique tag at a new node is given a “slot”. In the example below, the fish was first detected at LRW, so this set of 3 detections at LRW is assigned slot #1. Next, the fish went to EVU for 5 detections, which is assigned slot #2. If the fish was detected at LRW **after** the 5 detections at EVU, then the next set of LRW detections would get a new row and have slot #3.



“TagObs\_Wide\_YY-MM-DD.csv”

This file takes the nodes and number of detections from the “compressed” file and pivots it wide. Here, each row describes the number of detections that the tag had at all nodes. In the example below, tag “3DD.003D57FB85” was detected 1 time at HYC, 2 times at LRW, 7 at SRSC1, 1 at HRSC6, and so on.



“TagObs\_Directionality\_YY-MM-DD.csv”

This file uses the “node\_direction” metadata file to estimate fish movement. Like the “compressed” data, each record contains the number and duration for a set of detections at a node. “path” and “direction” columns are also added. In the case that node locations were grouped when adding directionality, a “node\_id” column is also present to detail the specific node where the fish was detected. In the “direction” column there are 1 of 5 possible values, depending on the previous node where the fish was detected:

* **Start:** The first observation of the fish within the zone designated by the “node\_direction” file.
* **Forward:** Between the previous observation and the current observation, the fish has moved in the direction designated by the “node\_direction” file.
* **Backward:** Between the previous observation and the current observation, the fish has moved in the opposite direction as designated by the “node\_direction” file.
* **Unknown:** The direction between the previous and current observations could not be determined.
* **No movement:** The previous fish detection was at the same location, but under a different “slot” number.

“TagObs\_FinalPaths\_YY-MM-DD.csv”

This file subsets data from the “directionality” file so that there is a maximum of one record per unique tag. This record is the last node location where the fish was observed, so that the “path” column shows a possible path that the fish may have followed.