University of Manitoba

IPY-CFL Ice Beacon Cookbook

CENTRE FOR EARTH OBSERVATION SCIENCE

Document Control

0.1 Version History

Version	Author(s)	Type	Date Modified	Comments
1.3	Campbell, Y.	Working Copy	2022/04/25	None
1.0	Kerr, L., Iyakoregha, V.	Previous edition	2020/06/29	None

0.2 Document Location

A hard copy of the document can be found in the Lab 489 document cupboard.

A digital copy of this document can be found on GitLab.

0.3 License

Campbell, Y., Iyakoregha, V.(2022). IPY-CFL Ice Beacon Cookbook, Version 1.0. Centre for Earth Observation Science. University of Manitoba.

Contents

Chapter 1

Ice Beacon Data

1.1 Critical Summary

This is a data management workflow 1 process.

The type of ice beacons used were Oceanetics model 703 iridium ice tracking beacons. They were deployed during the IPY-CFL project. 22 ice beacons were launched in Franklin Bay from November 2007 to May 2008. The ice beacons record their GPS location at regular intervals set by the researcher until the ice floe breaks up and they sink. Measured parameters include latitude, longitude, speed and direction. Zonal and meridional ice velocities were computed from the speed and direction of the velocities recorded for each ice beacon, and daily averages calculated.

1.1.1 Data Management Tools

1. **Ice Beacon Python Script:** Written by Victory Iyakoregha. Inputs both raw .csv and .xlsx ice beacon files and outputs both as .csv files. The python script strips unnecessary columns, and calculates the speed and distance travelled of ice beacons using python functions that read the GPS and timestamp data. The processed ice beacon data-files are organized as a single processed ice beacon dataset.

1.2 Initial Data

Initial File Names

300034012022250.csv, 300034012818300.csv, 300034012813320.csv, 300034012917920.csv, 300034012520600.csv, 300034012023330.csv, 300034012813330.csv, 300034012520610.csv, 300034012024220.csv, 300034012814310.csv, 300034012819300.csv, 300034012817320.csv, 300034012912920.csv, 300034012819310.csv, 300034012811310.csv, 300034012525590.csv, 300034012918920.csv, 300034012025330.csv, 300034012815320.csv, 300034012612770.csv, 300034012818320.csv, 300034012618770.csv, 300034012816330.csv, 300034012615000.csv, 300034012529590.csv, 300034012526590.csv, 300034012915930.csv, 300034012817300.csv,

Initial File Source and Location

Raw CSV files are received from David Babb. These files are then stored on GitLab.

1.3 Intermediate Data

The python scripts were created originally to process BaySys Ice beacon data which can be found here, and are the same scripts previously linked. The process of calculating the speed is the same, however the variables removed or changed were different between the two datasets, and so this must be taken into consideration when reading the script.

The codebook for the scripts with the detailed analytical process performed on the **BaySys** dataset can be found here.

1.3.1 User Instructions

1. Download Python scripts **process_Ice_beacon_files.py** and **Ice_beacon_files.ini** from CanWIN's Datahub.
2. Download Python scripts _Ice_beacon_files.py here.
3. In Ice_beacon_files.ini provide the path to the ice beacon data directory in the Input Directory
4. In Ice_beacon_files.ini provide the path to the destination folder for converted Ice Beacon files in Processed Directory
5. Click "Run" at the top of the script window. The script may take several minutes to complete

1.3.2 Script Overview

Script Type: python

Libraries Used: Python os, configparser, pandas (pd), datetime, math (m), lonlat and distance (geopy.distance)

These steps were performed on the IPY-CFL data, and is somewhat different than the steps described in the codebook and the scripts for the BaySys data.

1. **Clean_csv** accepts a single parameter, taking the dataframe for the CSV file, processing it, and returning it to open_files location. Clean_csv takes the original timestamp format and changes it to match the format of year-month-day and hours-minutes-seconds.
2. "name", "lat", "long", and "timestamp" are renamed as: "Beacon ID", "Latitude", "Longitude", and "Timestamp", respectively.

3. Removed original file column headers: Elevation, Heading, Speed, HDOP, VDOP, VerticalVelocity, Pressure, TempExternal, TempInternal, BeaconAlarmState, BatteryVoltage, ModemVoltage, WindSpeed.

4. **Clean_excel** accepts a single parameter, taking the dataframe for the CSV file, processing it, and returning it to open_files location. Clean_excel takes the original timestamp format and changes it to match the format of year-month-day and hours-minutes-seconds.

5. Clean_excel adds the column titles: "Beacon ID", "Latitude", "Longitude", "Timestamp".

6. **Calc_distance** accepts a single parameter, taking the dataframe for the CSV file, processing it, and returning it to open_files location.

7. This function calculates the distance between every two points, in this case, the two points being each two rows of data using the difference between the latitude and longitude of each point.

8. **Calc_speed** accepts a single parameter, taking the dataframe for the CSV file, processing it, and returning it to open_files location.

9. This function calculates the speed between every two points, in this case, the two points being each two rows of data using the difference between the distance calculated in calc_distance() and timestamp.

10. **Main()** runs the script.

Intermediate File Names

300034012022250.csv, 300034012818300.csv, 300034012813320.csv, 300034012917920.csv, 300034012520600.csv, 300034012023330.csv, 300034012813330.csv, 300034012520610.csv, 300034012024220.csv, 300034012814310.csv, 300034012819300.csv, 300034012817320.csv, 300034012912920.csv, 300034012819310.csv, 300034012811310.csv, 300034012525590.csv, 300034012918920.csv, 300034012025330.csv, 300034012815320.csv, 300034012612770.csv, 300034012818320.csv, 300034012618770.csv, 300034012816330.csv, 300034012615000.csv, 300034012529590.csv, 300034012526590.csv, 300034012915930.csv, 300034012817300.csv,

1.4 Final Data

Vocabulary Standardization and Metadata Addition

Column headers are standardized in accordance with the NERC Vocabulary, and units added after an underscore(_). Once the variables were standardized, a template was created for the final output files, and a python script was used to convert all the intermediate files to the format of this template. Additional metadata were added to the data files: project_name, platform_name, altitude_result_value_type, direction_result_value_type, speed_result_value_type

Final File Names

300034012022250_processed.csv, 300034012818300_processed.csv, 300034012813320_processed.csv, 300034012917920_processed.csv, 300034012520600_processed.csv, 300034012023330_processed.csv, 300034012813330_processed.csv, 300034012520610_processed.csv, 300034012024220_processed.csv, 300034012814310_processed.csv, 300034012819300_processed.csv, 300034012817320_processed.csv, 300034012912920_processed.csv, 300034012819310_processed.csv, 300034012811310_processed.csv, 300034012525590_processed.csv, 300034012918920_processed.csv, 300034012025330_processed.csv, 300034012815320_processed.csv, 300034012612770_processed.csv, 300034012818320_processed.csv, 300034012618770_processed.csv, 300034012816330_processed.csv, 300034012615000_processed.csv, 300034012529590_processed.csv, 300034012526590_processed.csv, 300034012915930_processed.csv, 300034012817300_processed.csv,

Final File Location

Ice beacon files can be located on CanWIN's Datahub here.

Final Dataset Variables

Table 1.1: Variables in the final data file

Header	Description	Data Type	Range or Expected Values	Units
project_name	Name of the project	string	N/A	None
platform_name	Name of the platform from which data was collected	string	N/A	None
Date_(yyyymmdd)	Date at the start of collection	datetime	N/A	None
Time_(hh24miss)	at the start of collection	datetime	N/A	None
beacon_ID	Ice beacon identification number	Index	N/A	None
latitude_DD	North-south directing lines of parallels	Rational	(0-90)	Degrees
longitude_DD	East-west directing lines of meridians	Rational	(0-180)	Degrees

altitude_m	Altitude. GPS altitude from these beacons is unreliable, ad beacon altitude should be assumed to be within 2m of mean sea level	Rational	(0-180)	Degrees
altitude_result_value_type	Identifies real or calculated data	string	N/A	None
speed_m_s	Speed calculated between two points using the difference between the distance calculated in calc_distance() and timestamp	Rational	Range	Metres/Second
speed_result_value_type	Identifies real or calculated data	string	N/A	None
direction_m	Direction	Rational	Range	Metres
direction_result_value_type	Identifies real or calculated data	string	N/A	None
fix	GPS fix acquired (1) or not (0)	Rational	0-1	N/A
sat	Number of satellites used to derive location	Rational	N/A	N/A

Chapter 2

Reference Tables

2.1 Data Levels

Level 0 – Raw data: unprocessed data and data products that have not undergone quality control. Depending on the data type and data transmission system, raw data may be available within seconds or minutes after real-time. Examples include real-time precipitation, streamflow, and water quality measurements

Level 0.1 – First pass QC: A first quality control pass has been performed to remove out of range and obviously erroneous values. These values are deleted from the record. E.g: Online Environment Canada stream-flow data, laboratory data

Level 1 – Quality Controlled Data: Data that have passed quality assurance procedures such as Level 0.1 and have been further quality controlled by data provider before being submitted to CanWIN (e.g. Idronaut data with only downwelling (upwelling data removed) data included.

Level 1.5 – Advanced Quality Controlled Data: Data have undergone complete data provenance (i.e. standardized) in CanWIN. Metadata includes links to protocols and methods, sample collection details, incorporates CanWIN's or another standardized vocabulary, and has analytical units standardized. Note: Process still under development in CanWIN (as of May 13, 2020).

Level 2 – Derived Products: Derived products require scientific and technical interpretation and can include multiple data types. E.g.: watershed average stream runoff derived from stream-flow gauges using an interpolation procedure.

Level 3 – Interpreted Products: These products require researcher (PI) driven analysis and interpretation and/or model-based interpretation using other data and/or strong prior assumptions. E.g.: watershed average stream runoff and flow using streamflow gauges and radarsat imagery

Level 4 – Knowledge Products: These products require researcher (PI) driven scientific interpretation and multidisciplinary data integration and include model-based interpretation using other data and/or strong prior assumptions. E.g.: watershed average nutrient runoff concentrations derived from the combination of stream-flow gauges and nutrient values.

2.2 Result Value Qualifiers

ADL - Above Detection Limit

BDL - Below Detection Limit

FD - Field Duplicate

LD - Lab Duplicate

$ - Incorrect sample container

EFAI - Equipment failure, sample lost

FEF - Field equipment failed

FEQ - Field Equipment Questionable

FFB - Failed. Field blank not acceptable

FFD - Failed. Field Duplicate

FFS - Failed. Field spike not acceptable

H - Holding time exceeded

ISP - Improper sample preservation

ITNA - Incubation time not attained

ITNM - Incubation temperature not maintained

JCW - Sample container damaged, sample lost

NaN - Value is missing and reason is not known

NC - Not collected

ND - Not detected

NR - Sample taken/measured on site but information in this field not recorded

NS - Sample collected but not submitted

OC - Master Coordinate List Used

P - Analysis requested and result pending

prob_good - probably good value. Data value that is probably consistent with real phenomena but this is unconfirmed or data value forming part of a malfunction that is considered too small to affect the overall quality of the data object of which it is a part

prob_bad - probably bad value. Data value recognised as unusual during quality control that forms part of a feature that is probably inconsistent with real phenomena

Interpolated - This value has been derived by interpolation from other values in the data object

Q - Below limit of quantification (LOQ). The value was below the LOQ of the analytical method. The value in the result field is the limit of quantification (limit of detection) for the method

Chapter 3

Options and Packages

3.1 Python

3.1.1 Python Script-Specific Options

- **os** -os is a Python module, part of the standard library on Python. Therefore, it comes with Python but still needs to be imported.
- **Configparser** - Configparser is a Python class which implements a basic configuration language for Python programs. It provides a structure similar to Microsoft Windows INI files. ConfigParser allows to write Python programs which can be customized by end users easily.
- **pandas** - Pandas is another Python module that allows the user to import CSV (comma seperated values) files. Importing pandas using the pd prefix avoids overlap with other Python tools.
- **datetime** - Datetime is the Python module containing: date, time, and datetime.
- **math** - m is a mathematical calculations module of Python.
- **lonlat, distance** - Importing lonlat, distance from geopy.distance takes x-longitude, y-latitude, and optionally, z-altitude values and gives you the coordinates as a point.

3.1.2 Python Packages

Visit https://docs.python.org/3/library/ to learn more about python packages

- **DateTime** - Date/time values
- **geopy** - Uses geocoding to locate addresses, cities, towns, etc.

Example: Section 2.1 from Victory's semi-hemi codebook.

Glossary

Data Management Workflow 1 - one of three data management workflows used at CEOS. Workflow 1 is used for data that is managed directly by CEOS from instrument collection to data sharing. 1