Idronaut Ocean Seven 304 CTD Cookbook

CENTRE FOR EARTH OBSERVATION SCIENCE

Document Control

0.1 Version History

Version	Author(s)	Type	Date Modified	Comments
1.0	Rodgers, K.	Working Copy	2023-02-27	

0.2 Document Location

A digital copy of this document can be found on **GitLab** here:

https://cwincloud.cc.umanitoba.ca/MBGL/data/-/tree/master/Idronaut

A digital copy of this document can be found on **CanWIN DataHub** here:

https://canwin-datahub.ad.umanitoba.ca/data/project/mbgl-idronaut-data

0.3 License

Rodgers, K.(2023). Cookbook, Version 1.0. Centre for Earth Observation Science. University of Manitoba.

Contents

1 Idronaut CTD Data Workflow

1.1 Critical Summary

Conductivity, temperature, and depth are collected with water samples and deployment of instruments. Cleaning Idronaut (CTD) data will remove the stabilizing period and upcast portions. Specific conductance at 25°C is measured by the instrument; however, the script will also convert conductivity to specific conductance at 25°C. for comparison. All data cleaning workflows are done in the R markdown **Idronaut_data_workflows.Rmd**.

1.1.1 Data Downloading and Storage

Data from the Idronaut CTD is downloaded using the Iterm Application. Please refer to Idronaut Protocols for detailed data downloading instructions (protocol links below).

After downloading, raw Idronaut CTD files are stored on GitLab in the raw folder, see https://cwincloud.cc.umanitoba.ca/MBGL/data/-/tree/master/Idronaut.

1.1.2 Idronaut Use Protocols

MBGL group instrument use protocols for the Idronaut CTD are located:

1. **CanWIN DataHub**: https://canwin-datahub.ad.umanitoba.ca/data/project/mbgl-idronaut-data
2. **Gitlab**: https://cwincloud.cc.umanitoba.ca/MBGL/data/-/tree/master/Idronaut

1.1.3 Data Management Tools

Idronaut R-Project: Consists one R markdown file with three sections and each produce an intermediate file, two cleaned files, and one final curated file.

1.2 Initial Data

1.2.1 Raw Idronaut Dataset Files

Dataset Description

Individual casts are downloaded from the Idronaut CTD using Iterm application. Files are in a .txt format.

File Name(s) Format

"StationID_IdronautSN_Date_CastNubmer.txt".

File Source and Location

Raw CTD files can be found on Gitlab in the relevant repository in the folder "\raw". The raw CTD data file contains nine columns of data parameters (see Dataset Variables for parameters). Each Idronaut cast contains a stabilizing period out of the water and in the water, a downcast and an upcast.

Dataset Variables

Header	Description	Units/Format
Date	Date	DD-MM-YYYY
Time	Time	HH:mm:ss.ss
Pres	Pressure	dbar
Temp	Temperature	°C
Cond	Conductivity	μS/cm
Sal	Salinity	
Turb	Turbidity	FTU
SigmaT	Sigma T	
Cond25	Internal calculated specific conductance at 25°C	μS/cm

1.3 Intermediate Data

1.3.1 Intermediate Idronaut Dataset Files

Dataset Description

There is an intermediate output file which contains all the same data as the raw file but with an extra column that has a calculated conductivity at 25°C. The file will have two specific conductance at 25°C values. The calculated one is used to compare with the internal calculated specific conductance.

File Name(s) Format

Format of intermediate idronaut file names: "SiteID_IdronautID_YYYYMMMDD.csv"

File Source and Location

Requires raw idronaut files from "\raw" folder of the **"data\Idronaut"** repository.

Intermediate files are saved to **"processed\intermediate"** folder of the **"data\Idronaut"** repository.

Dataset Variables

Header	Description	Units/Format
Date	Date	DD-MM-YYYY
Time	Time	HH:mm:ss.ss
Pres	Pressure	dbar
Temp	Temperature	°C
Cond	Conductivity	μS/cm
Sal	Salinity	
Turb	Turbidity	FTU
SigmaT	Sigma T	
Cond25	Internal calculated specific conductance at 25°C	μS/cm
Cond_std25	External calculated specific conductance of the water body standardized to 25°C	μS/cm

1.4 Cleaned Data

1.4.1 Cleaned Idronaut Dataset Files

Dataset Description

There are two cleaned data file outputs which display the same dataset variable as the raw files however, some lines of data have been removed. The first cleaned data file will have the stabilization period removed. The second intermediate data file will have the stabilization period and upcast removed.

File Name(s) Format

Downcast and upcast data file: SiteID_IdronautID_YYYYMMMDD_cast.csv

Downcast data file: SiteID_IdronautID_YYYYMMMDD_downcast.csv

File Source and Location

Requires intermediate idronaut files from "\intermediate" folder of the **"data\Idronaut"** repository.

Cleaned files are saved to **"processed\cleaned"** folder of the **"data\Idronaut"** repository.

Dataset Variables

Header	Description	Units/Format
Date	Date	DD-MM-YYYY
Time	Time	HH:mm:ss.ss
Pres	Pressure	dbar
Temp	Temperature	°C
Cond	Conductivity	μS/cm
Sal	Salinity	
Turb	Turbidity	FTU
SigmaT	Sigma T	
Cond25	Internal calculated specific conductance at 25°C	μS/cm
Cond_std25	External calculated specific conductance of the water body standardized to 25°C	μS/cm

1.5 Final Output Data

1.5.1 Final Curated Idronaut Dataset Files

Dataset Description

This is the final dataset. Additional columns headers with associated metadata have been added. Local time has been converted to UTC time.

File Name(s) Format

SiteID_IdronautID_YYYYMMMDD_L1_2.csv

File Source and Location

Requires cleaned idronaut files from **"processed\cleaned"** folder of the **"data\Idronaut"** repository.

Final curated files are saved to **"processed\curated"** folder of the **"data\Idronaut"** repository.

Dataset Variables

See next page

Header	Description	Units/Format
Station_id	Station ID	
Latitude_dd	Station latitude coordinate	Decimal degrees
Longitude_dd	Station longitude coordinate	Decimal degrees
UTC_iso8601	Standardized collection date in UTC	YYYY-MM-DDTHH:mm:ss
Date_local	Local date	DD-MM-YYYY
Time_local	Local time	HH:mm:ss.ss
Pres_Z	Pressure	dbar
Pres_Z_result_ value_qualifier	Result value qualifier for pressure	
CTDTmp90	Temperature	°C
CTDTmp90_result_ value_qualifier	Result value qualifier for temperature	
CTDCond	Conductivity	μS/cm
CTDCond_result_ value_qualifier	Result value qualifier for conductivity	
CTDSal	Salinity	
CTDSal_result_ value_qualifier	Result value qualifier for salinity	
Turbidity	Turbidity	FTU
Turbidity_result_ value_qualifier	Result value qualifier for turbidity	
SigTheta	Sigma T	
SigTheta_result_ value_qualifier	Result value qualifier for sigma T	
CTDCond_std25_raw	Internal calculated specific conductance at 25°C	μS/cm
CTDCond_std25_raw_ result_value_qualifier	Result value qualifier for internal conductivity at 25°C	
CTDCond_std25_calc	External calculated specific conductance of the water body standardized to 25°C	μS/cm
CTDCond_std25_calc_ result_value_qualifier	Result value qualifier for external conductivity at 25 °C	

2 Scripts and Analytical Processes

2.1 Critical Summary

Overview of the general R coding steps used to create the intermediate, cleaned, and final curated dataset files from the raw Idronaut CTD data. Rmarkdown and Rscript for curating Idronaut CTD data is found on GitLab in the Idronaut directory of the MBGL Data Repository.

2.1.1 Script Overview

1. **Idronaut_data_workflow.Rmd**: This R markdown has 3 sections of scripts. Use under the **Idronaut R-Project** R Project.
   (a) **Intermediate Idronaut Dataset**: Add external calculated specific conductance at 25°C
   (b) **Clean Idronaut Datasets**: Creates two sets of cleaned data by removing stabilization period and upcast
   (c) **Final Curated Idronaut Dataset**: Standardizes variable names and applies any additional metadata for curation
2. **plot_temp_pres_script.R**: This R script plots a temperature and pressure profile. Use under the **Idronaut R-Project** R Project.

2.1.2 Idronaut_data_workflow.Rmd script

Script type: R-project, R markdown

Libraries used: tidyverse, lubridate, here

This Rmarkdown contains 3 sections.

Section 1. Intermediate Idronaut Dataset

File(s) in: .txt Idronaut CTD file

File(s) out: .csv intermediate Idronaut CTD file

Analytical Process:

1. Imports raw .txt file from **"raw"** folder.
2. Adds a new column with the external calculated specific conductance at 25°C.

3. Exports new .csv file to **"processed\intermediate"** folder.

Section 2: Clean Idronaut Datasets

This script has two parts and needs adjustments for every data file that is imported.

File in: .csv intermediate Idronaut CTD file

File out: two independent.csv cleaned Idronaut CTD files

Analytical Process:

1. Part 1 - Subset cast period (downcast and upcast)
   (a) Imports intermediate .csv file from **"processed\intermediate"** folder.
   (b) Subset rows that are part of the upcast and downcast. This will remove the stabilizing period and times idronaut is out of the water. Modify line 79.
   (c) Exports new .csv file to **"processed\cleaned"** folder.
2. Part 2 - Subset downcast period
   (a) Imports first cleaned .csv file from **"processed\cleaned"** folder.
   (b) Subset rows that are part of the downcast period. Modify line 96.
   (c) Exports new .csv file to **"processed\cleaned"** folder.
3. Part 3 - Extracting downcast only
   (a) Imports intermediate .csv file from **"processed\intermediate"** folder.
   (b) Subset rows that are only part of the downcast. Modify line 113.
   (c) Exports new .csv file to **"processed\cleaned"** folder.
4. Part 4 - Checking Downcast data
   (a) Imports cleaned downcast .csv file from **"processed\cleaned"** folder.
   (b) Plot temperature against pressure.

Section 3: Final Curated Idronaut Dataset

This script creates the final curated file in a .csv file. Currently, files are converted into .txt files when added to the CanWIN Datahub site.

File in: .csv cleaned Idronaut CTD file

Files out: .csv curated Idronaut CTD file

Analytical Process:

1. Imports cleaned .csv file from **"processed\cleaned"** folder.
2. Converted local time (America/Winnipeg) to UTC time.
3. Standardizes names and adds any additional metadata columns.
4. Exports new .csv file to **"processed\curated"** folder.

2.1.3 plot_temp_pres_script.R script

This R script creates a temperature versus pressure profile plot of the downcast data.

1. Imports downcast .csv file from **"processed\cleaned"** folder.
2. Creates plot of temperature versus pressure profile using ggplot2.
3. Exports .png file to **"graphs"** folder.

A Reference Tables

A.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.

Content retrieved from https://lwbin.cc.umanitoba.ca on July 06, 2020.

A.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

B Glossary of Options and Packages

B.1 R Packages

Visit https://cran.r-project.org/web/packages/available_packages_by_name.html to learn more about R packages

- **tidyverse** - set of packages that make it easier to share common data representations and 'API' design.
- **lubridate** - makes it easier to work with dates and times.
- **janitor** - allows for simple functions for examining and cleaning dirty data.
- **ggplot2** - creating graphics, based on "The Grammar of Graphics".
- **hms** - storing and formatting time-of-day values.
- **here** - provides simplier way to find/load files.