Seabird CTD Data Cookbook

CANADIAN WATERSHED INFORMATION NETWORK (CANWIN)

Document Control

0.1 Version History

Version	Author(s)	Type	Date Modified	Comments
1.0	Yezhova, K.	Working Copy	June 18, 2024	

0.2 Document Location

The **original copy** of this document can be found here: https://cwincloud.cc.umanitoba.ca/data_mgmt/ceos-cookbooks/-/tree/master/CEOS-cookbook-master-template?ref_type=heads

Please contact portalco@umanitoba.ca if access is required.

A **public digital copy** of this document can be found here: https://canwin-datahub.ad.umanitoba.ca/data/cookbook-url

0.3 License

Yezhova, K.(2024). Seabird CTD Data Cookbook, Version 1.0. Centre for Earth Observation Science. University of Manitoba.

Contents

1 Introduction

1.1 Critical Summary

This document discusses the processing of the autonomous CTD dataset from the 2021 James Bay Expedition.

Instrument: SBE 19plus V2 SeaCAT Profiler CTD SN7783 (Autonomous CTD)

Vessel: RV William Kennedy and its small boats

Cruise dates: August 1-17, 2021

Spatial region: James Bay

2 Seabird CTD Data Processing Workflow

2.1 Pre-Processing

2.1.1 Scripts/Software Used

Software

The SBE Processing Software was used to process the data. The manual can be found **here.** Page 20 outlines the steps for processing data.

Script Location

The CanWIN GitLab and our Data Catalogue both contain the R scripts used in data processing - (to access Gitlab, submit access requests to the data curator).

2.1.2 Environment Setup

Create the following folder structure:

1. 2021_wk_auto_ctd_sn7783
    (a) logbooks
    (b) orignials
    (c) r_scripts
    (d) seabird_psa_and_xmlcon
    (e) data
        i. 00_raw
        ii. 01_datacnv
        iii. 02_section
        iv. 03_filter
        v. 04_align
        vi. 05_ctmass

vii. 06_loopedit

viii. 07_derive

ix. 08_binavg

x. 09_split

xi. 10_final

2. Into the logbooks folder, place the ship logbook

3. Into the originals folder, place all original data from the field (data files, logbooks, etc.); zipped to prevent accidental modification

4. Ensure CTD files all follow the same naming structure, and that the names are consistent with the digital logbook

5. The .xml file of one of the first casts was compared against the .xml file of one of the last casts, no differences in configuration or calibration values were found. Only one .xmlcon file was used during the cruise to the best of Kate's knowledge, so there should not be issues.

2.2 Raw Data Processing

The raw data files are the **.hex** files. They are converted to **.cnv** files, and go through a series of steps before being converted to their final output format: **.csv**

2.2.1 Data Input

File Type: .hex

File Location: Gitlab location

2.2.2 Steps to Process Raw Data

1. Convert raw .hex files to .cnv files

   (a) SBE Data processing –> **Run** –> **Data Conversion (#1)** –> **File Setup**

      i. Open **01_DatCnv_SN7783_AUTOCTD.psa** file from the ***seabird_psa*** folder

      ii. Under Instrument configuration file, load the ***.xmlcon*** file in the ***00_raw*** folder

      iii. Under Input directory, select all .hex files from "00_raw" folder

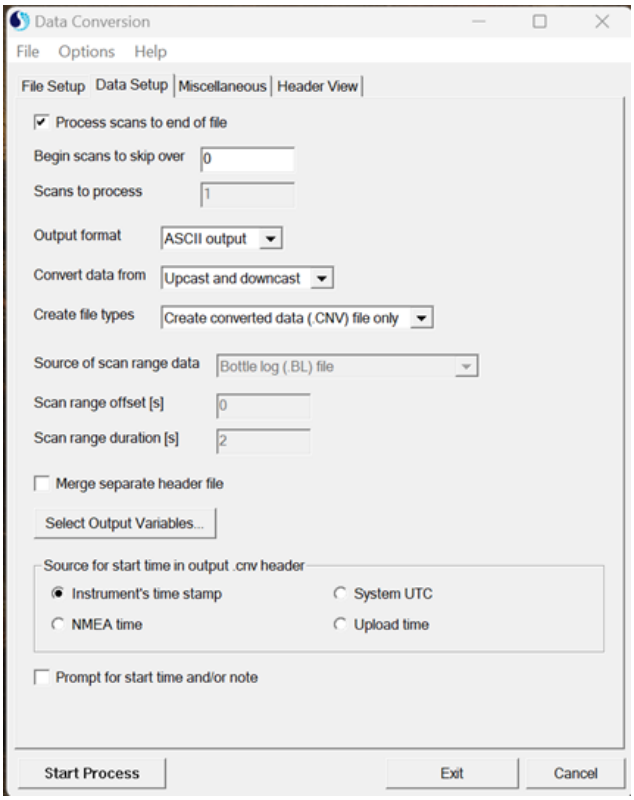
         A. Getting error: "Header line length exceeds buffer length". The issue is with "* FileName" line in the HEX files, too long. The line is 116 characters, the limit appears to be 115 characters.

- Manually change the lines from:
  - "* FileName = C:\Users\William Kennedy\Documents\William Kennedy\WK21\CTD 7783\SBE19plus. . . " to
  - "* FileName = C:\SBE19plus. . . " (with WordPad's replace function)

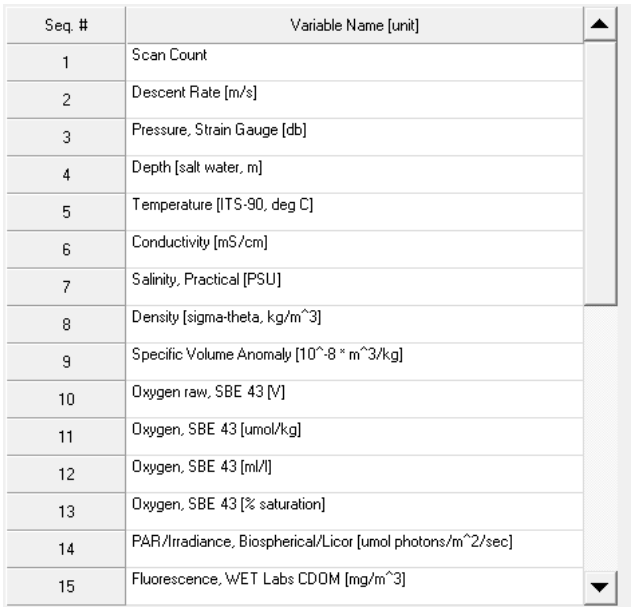
iv. Under Output directory, select "***01_datacnv***" folder

(b) ... –> Data Setup

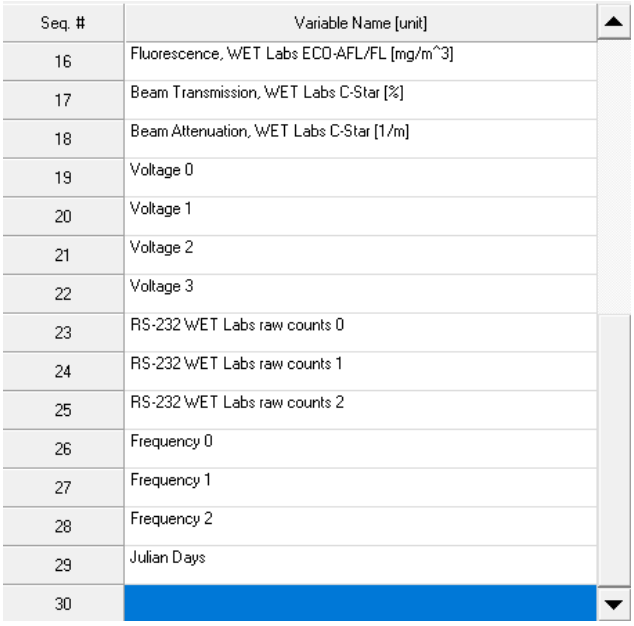
i. *Data Setup Tab*



ii. Click Select Output Variables. . . , and choose the following:



Seq. #	Variable Name [unit]
1	Scan Count
2	Descent Rate [m/s]
3	Pressure, Strain Gauge [db]
4	Depth [salt water, m]
5	Temperature [ITS-90, deg C]
6	Conductivity [mS/cm]
7	Salinity, Practical [PSU]
8	Density [sigma-theta, kg/m^3]
9	Specific Volume Anomaly [10^-8 * m^3/kg]
10	Oxygen raw, SBE 43 [V]
11	Oxygen, SBE 43 [umol/kg]
12	Oxygen, SBE 43 [ml/l]
13	Oxygen, SBE 43 [% saturation]
14	PAR/Irradiance, Biospherical/Licor [umol photons/m^2/sec]
15	Fluorescence, WET Labs CDOM [mg/m^3]



Seq. #	Variable Name [unit]
16	Fluorescence, WET Labs ECO-AFL/FL [mg/m^3]
17	Beam Transmission, WET Labs C-Star [%]
18	Beam Attenuation, WET Labs C-Star [1/m]
19	Voltage 0
20	Voltage 1
21	Voltage 2
22	Voltage 3
23	RS-232 WET Labs raw counts 0
24	RS-232 WET Labs raw counts 1
25	RS-232 WET Labs raw counts 2
26	Frequency 0
27	Frequency 1
28	Frequency 2
29	Julian Days
30	

iii. Click **Start Process**

2. Preparing an Excel sheet for taking notes

   (a) Open the ***01_datacnv*** folder

   (b) Select all files (ctrl+A)

   (c) Hold shift, right click, select copy as path

   (d) Go to excel and paste (Ctrl+V)

(e) Select **column A**, go to **Find & Select** -> **Replace** -> Type out the beginning of the paths in **find what**, and replace with blank

(f) Add title row for “cast id, start scan, end scan, notes”

(g) Save the file, titled “***section.xlsx***”, into the ***logbooks*** folder

3. Plotting casts

(a) SBE Data Processing –> **Run** –> **Sea Plot (#20)** –> **File Setup**

i. For Program setup file, choose ***02_SeaPlot_SN7783_AUTOCTD.psa***

ii. For Input directory, select all files in ***01_datacnv*** folder

iii. For Output directory, select any folder (the plots do not get automatically saved)

(b) –> Plot Setup

i. Title: datacnv

ii. For variables, choose the following:

A. y-axis: pressure

B. x-axis 1: scan count

C. hide other x-axes

(c) Click **Start Process**

(d) In the plot window select **View** –> **Show Cursor Position**

(e) Record scan # of beginning of downcast (when the CTD comes back up to surface after acclimating at ~5m depth for 3 minutes), and the end of the upcast (just before the CTD comes out of the water at the end) for each cast in the Excel sheet you created in the previous step.

(f) Some notes regarding the 2021 data:

i. For most casts, the soaking period was a few minutes just near surface, did not go down to 5 m as is the proper procedure in recent cruises.

ii. For most casts, CTD stayed for some time near the bottom before beginning the upcast.

iii. The following were not proper casts, more of surface CTD samples. The downcasts will be deleted:

A. CTD_EM_0001

B. CTD_EM_0002

C. CTD_MR_0002

D. CTD_MR_0004

E. CTD_MR_0011

iv. Use upcast only for the following cast:

A. CTD_MR_0020

4. Cutting out soaking period (must go one file at a time)

   (a) In SBE Data Processing: **Run** –> **Section (#16)** –> **File Setup**

      i. Program setup file: ***03_Section_SN7783_AUTOCTD.psa***

      ii. Input: one cast at a time from ***01_datacnv.*** (Definitely not the most time efficient method but the simplest at this point, I am running out of time to test out different MATLAB/R scripts.)

      iii. Output: ***02_section*** folder

   (b) –> Data Setup

      i. Section based on: scan count

      ii. Input minimum and maximum value for each cast and click **Start Process**, one cast at a time

5. Run the ***01_section_check.R*** script to check that the correct values were entered in the Section module and that none of the pressure/depth values ended up being negative (indicating measurements in the air).

6. Run the ***02_pump_check.R*** script to ensure that the pump started working before the downcast began (the pump only starts working once the minimum conductivity frequency is met and the pump delay elapses).

   (a) Minimum conductivity frequency was never met during these casts, therefore, the pump never turned on. These casts will be processed further, but any measurements impacted by the pump (conductivity, temperature, DO) should NOT be used. Kate could not find entries for these CTD casts in the logbook to update the logbook.

      i. WK21_CTD_MR_0017

      ii. WK21_CTD_MR_0018

      iii. WK21_CTD_MR_0019

   (b) For the following cast, pump kicked in after beginning of the downcast. The cast was re-sectioned so that the data before the pump started working was removed, and the upcast will be used in the final data file as the downcast data now starts at approximately

7 m.

i. WK21_CTD_0033 (downcast begins at scan #229, pump turned on at scan #273)

7. Run the ***03_sal_check.R*** to check the minimum conductivity measurements to ensure none of the samples were freshwater, as the processing steps are slightly different from freshwater. SBE said in personal communication that the rough threshold for freshwater for data processing purposes is 0.6 S/m, i.e., 6 mS/cm.

   (a) The following casts captured samples with conductivity of less than or equal to 6 mS/cm; if these data are used, they will need to be flagged with a note "Cast processed following seawater guidelines, however, this cast captured freshwater samples (samples with conductivity less than or equal to 6 mS/cm).":

      i. WK21_CTD_MR_0012

      ii. WK21_CTD_MR_0014

      iii. WK21_CTD_MR_0017

      iv. WK21_CTD_MR_0018

      v. WK21_CTD_MR_0019

8. Filtering

   (a) SBE Data processing –> **Run** –> **Filter (#2)** –> **File Setup**

      i. Program setup file: ***04_Filter_SN7783_AUTOCTD.psa***

      ii. Input directory: ***02_section*** folder (all casts)

      iii. Output directory: ***03_filter*** folder

   (b) –> Data Setup

      i. Low pass filter A, time constant (s): 1.0

      ii. Low pass filter B, time constant (s): 0.5

      iii. Specify Filters. . .

         A. Clear all

         B. Pressure, Strain Gauge (db): Low pass filter A

         C. Temperature (ITS-90, deg C): Low pass filter B

         D. Conductivity (mS/cm): Low pass filter B

   (c) Click **Start Process**

9. Align CTD (advance parameters in time relative to pressure)
   (a) SBE Data processing –> **Run** –> **Align CTD (#3)** –> **File Setup**
      i. Program setup file: ***05_Align_SN7783_AUTOCTD.psa***
      ii. Input directory: ***03_filter*** folder (all casts)
      iii. Output directory: ***04_align*** folder
   (b) –> Data Setup –> Enter Advanced Values
      i. Clear all
      ii. Temperature (ITS-90, deg C): +0.5 seconds
         - This is the recommended value for SBE19plusV2 in the data processing manual
      iii. Conductivity (mS/cm): +0.5 seconds
         - Note that the manual gives contradicting statements. First statement is: "For an SBE 19plus or 19plus V2 with a standard 2000-rpm pump, do not advance conductivity." Second statement is: "If temperature is advanced relative to pressure and you do not want to change the relative timing of temperature and conductivity, you must add the same advance to conductivity."
         - Pascal applies a +0.5 second advance to both temperature and conductivity, Pascal's method will be followed.
      iv. Oxygen raw, SBE43 (V): no advance
         - The data processing manual suggests +3 to 7 seconds for an SBE19Plus
         - Janine noted that this only works if the Oxygen raw, SBE43 (V) variable is being aligned. Janine tried several delays and the 0s delay seemed best for 2021 data.
         - Pascal noted that at the beginning, he tried to estimate the right correction. You need to remove the gap between the downcast and upcast because of the long sensor response time. When you are going to apply a correction, you shift all oxygen values X seconds below their original place. Therefore, if you are moving at a speed of 1m/s, a +5 second shift would shift all values 5 m below where they were recorded. For oceanic waters with little variation, this could be okay. But for Arctic waters with chlorophyll maxima and oxygen peaks, this could create a big shift between these two events (SCM and O_2 peak). In Pascal's opinion, it is scientifically incorrect to create a gap between SCM and O_2 peak. Pascal either does not apply an oxygen correction, or he applies a 0.5s correction (same as for temperature and conductivity).
   (c) Click **Start Process**

10. Cell Thermal Mass

    (a) As per the data processing manual, “Perform conductivity cell thermal mass correction if salinity accuracy of better than 0.01 PSU is desired in regions with steep gradients. Note: do not use Cell Thermal Mass for freshwater data.” Note that Cell Thermal Mass is being used here on several casts that captured freshwater data (see above). These freshwater data should be flagged if they are used.

    (b) SBE Data processing –> **Run** –> **Cell Thermal Mass (#4)** –> **File Setup**

        i. Program setup file: ***06_CTMass_SN7783_AUTOCTD.psa***

        ii. Input directory: ***04_align*** folder (all casts)

        iii. Output directory: ***05_ctmass*** folder

    (c) –> Data setup –> Correct primary conductivity values

        i. Thermal anomaly amplitude (alpha): 0.04

        ii. Thermal anomaly time constant (1/beta) = 8.0

    (d) Click **Start Process**

11. Loop Edit (flags scans with very low and backward velocity)

    (a) SBE Data processing –> **Run** –> **Loop Edit (#5)** –> **File Setup**

        i. Program setup file: ***07_LoopEdit_SN7783_AUTOCTD.psa***

        ii. Input directory: ***05_ctmass*** folder (all casts)

        iii. Output directory: ***06_loopedit*** folder

    (b) –> Data Setup

        i. Minimum velocity type: Fixed minimum velocity

        ii. Minimum CTD velocity (m/s): 0.05. Note that Pascal recommended using velocity < 0.1 m/s (as opposed to the SBE recommended 0.25 m/s).

        iii. Uncheck **“Remove surface soak”**

        iv. Check **“Exclude scans marked bad”**

    (c) Click **Start Process**

12. Derive (computes thermodynamic properties based on EOS-80 (practical salinity))

    (a) SBE Data processing –> **Run** –> **Derive (#6)** –> **File Setup**

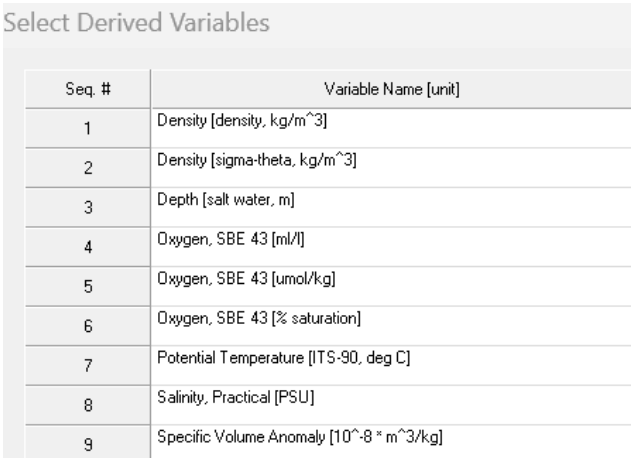
        i. Program setup file: ***08_Derive_SN7783_AUTOCTD.psa***

ii. Instrument configuration file: ***AUTO_7783_w_DO_ECO_Cstar.xmlcon***

iii. Input directory: ***06_loopedit*** folder (all casts)

iv. Output directory: ***07_derive*** folder

(b) –> **Data Setup** –> Select **Derived Variables**

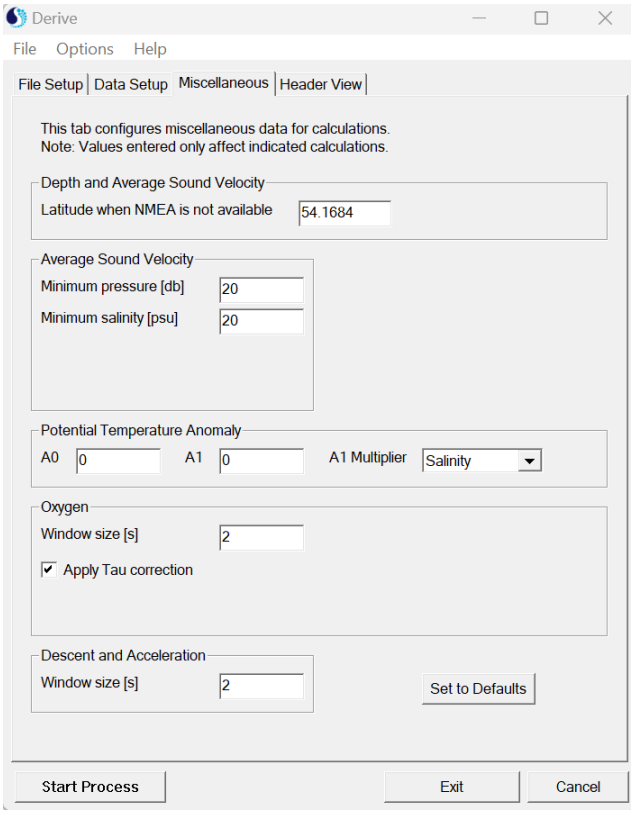

Select Derived Variables

Seq. #	Variable Name [unit]
1	Density [density, kg/m^3]
2	Density [sigma-theta, kg/m^3]
3	Depth [salt water, m]
4	Oxygen, SBE 43 [ml/l]
5	Oxygen, SBE 43 [umol/kg]
6	Oxygen, SBE 43 [% saturation]
7	Potential Temperature [ITS-90, deg C]
8	Salinity, Practical [PSU]
9	Specific Volume Anomaly [10^-8 * m^3/kg]

i.

(c) –> Miscellaneous

i. Latitude when NMEA is not available: 53.3114 (average latitude of all casts in the logbook).



(d) Click **Start Process**

(e) Optional step that was not done: Use Derive TEOS-10 (absolute salinity) module to derive variables based on TEOS-10.

13. Bin Average

(a) SBE Data processing –> **Run** –> **Bin Average (#8)** –> **File Setup**

i. Program setup file: ***09_BinAvg_SN7783_AUTOCTD.psa***

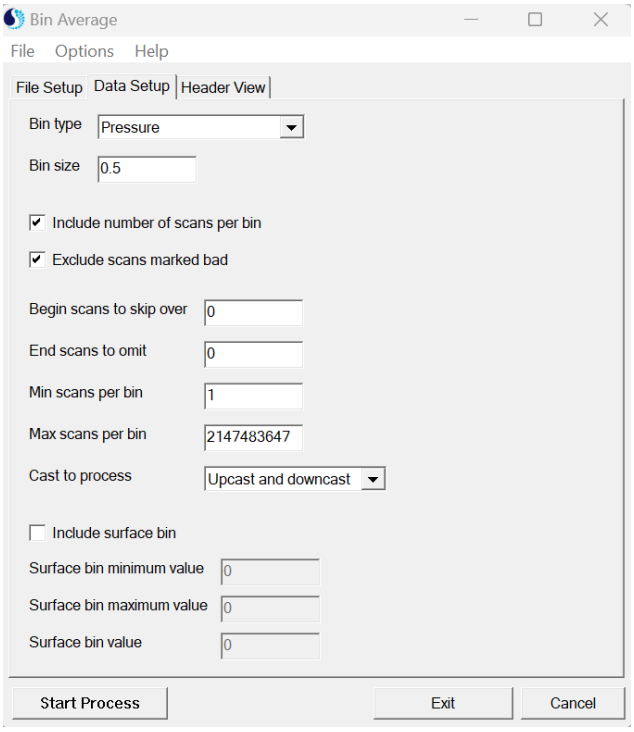
ii. Input directory: ***07_derive*** folder (all casts)

iii. Output directory: ***08_binavg*** folder

(b) –> Data Setup

i. Bin type: Pressure

ii. Bin size = 0.5



(c) Click **Start Process**

14. Split (splitting the downcast from upcast)

(a) SBE Data processing –> **Run** –> **Split (#17)** –> **File Setup**

i. Program setup file: ***10_Split_SN7783_AUTOCTD.psa***

ii. Input directory: ***08_binavg*** folder (all casts)

iii. Output directory: ***09_split*** folder

(b) –> Data Setup

i. Output files: upcast and downcast (it will rename each file for downcast with a “d” and upcast with a “u” in front of the file name)

ii. Check “Exclude scans marked bad”

(c) Click **Start Process**

(d) The following files gave errors, could not be processed through the split module. Errors:

i. Header entry ’nvalues’ must be > 0 in WK21_CTD_EM_0001.cnv.

ii. Header entry ’nvalues’ must be > 0 in WK21_CTD_MR_0004.cnv.

iii. Header entry ’nvalues’ must be > 0 in WK21_CTD_MR_0011.cnv.

iv. Header entry 'nvalues' must be > 0 in WK21_CTD_MR_0020.cnv.

(e) Resulting downcast files were deleted for the following casts, see explanation above:

i. CTD_EM_0002

ii. CTD_MR_0002

15. Merging with logbook (using R)

(a) Run the ***04_final_file.R*** script to merge CTD data with the logbook and output Excel and ODV files.

(b) Important notes:

i. Janine noted that CTD casts 1-7 had incorrect internal timestamps. Timestamps were adjusted in the final file by adding 7 hours to CTD's internal time for these casts.

A. CTD 1 Internal time: 8/3 21:50 Logbook time: 8/4 4:52

B. CTD 2 Internal time: 8/3 22:53 Logbook time: 8/4 5:54

C. CTD 3 Internal time: 8/3 23:42 Logbook time: 8/4 6:42

D. CTD 4 Internal time: 8/4 00:31 Logbook time: 8/4 7:32

E. CTD 5 Internal time: 8/4 01:18 Logbook time: 8/4 8:19

F. CTD 6 Internal time: 8/4 02:04 Logbook time: 8/4 9:06

G. CTD 7 Internal time: 8/4 6:24 Logbook time: 8/4 13:25

ii. Janine noted that four of the Moose River casts (MR_0002, MR_0004, MR_0011, MR_0020) were experimental measurements that involved holding the CTD just below the surface of the water while the small boat moved across tidal fronts. The same may be true for some EM casts, it was not clear from the notes. Janine also noted that for some of these MR/EM casts, she did not apply loop edit, bin averaging, or split modules when she processed the data. Given that many casts from these two transects are "abnormal" in how they were collected (lacking soaking period, too short of a soaking period, surface sample only, collecting dynamic tidal front data), and given that many casts captured freshwater, some had pump issues, and some were not able to make it through all stages of processing (see 19.d above), MR/EM casts will not be part of the final file. Users can decide based on their needs from which stage of data processing to take the MR/EM data from.

2.3 Final Output Data

This section describes the final output files produced. The final output is the last dataset produced after processing/analysis is complete.

2.3.1 File Details

File Type: .csv

Data Level: 1

File Location: The processed files are located here:

- CanWIN Gitlab

- CanWIN Data Catalogue

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://canwin-datahub.ad.umanitoba.ca/content/dataset-level.

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