# QAQC Notes for Cheney Reservoir

**Most up-to-date files on GEISHA FTP site (as of 11 August 2020)**

Cheney Reservoir QAQC Notes\_11Aug2020.docx

RawNutrientData\_Cheney\_QAQC\_09Sep2017.csv

RawLakeHFData\_Cheney\_QAQC\_20Sep2017.csv

RawWeatherData\_Cheney\_USW00003928\_QAQC\_10Sep2017.csv

RawPhytoData\_Cheney\_QAQC\_21Jul2020.csv

SpeciesList\_Cheney\_QAQC\_20Jul2020.xslx

PhytoMetadata\_Cheney\_GEISHA\_done.docx

DailyWeatherData\_Cheney\_USW00003928\_QAQC\_06Jun2018.csv

DailyLakeHFData\_Cheney\_QAQC\_11Aug2020.csv

LakeMetadata\_Cheney\_QAQC\_09May2019.xlsx

In all files, special characters were removed and replaced i.e. Chascomús to Chascomus. See the “Log” of changes made to the files since first QAQC’d. These are located at bottom of this document.

**RawNutrientData\_Cheney\_QAQC\_09Sep2017.csv**

The following data points had suspicious values

* secchi\_m on 09-11-15 was suspiciously high at 7.3152. Data provider indicated this was incorrect and should be 0.6096m. Value was corrected.
* All values below detection were changed to 0.
* Total nitrogen was calculated by summing no2 and no3. The sum does not include nh4.
  + Explanation from Jennifer Graham
    - TN values were improperly calculated for this data set. Because we have total Kjeldahl nitrogen (which includes ammonia) and nitrate plus nitrite data for the entire period of record at the site, we typically calculate TN as total kjeldahl nitrogen plus nitrate plus nitrite (TKN+NO3+NO2). Since TKN wasn’t requested as part of the dataset, we overlooked it when describing TN. A revised data set with TN calculated as described is attached.
    - New file named: RawNutrientData\_Cheney\_061517.xlsx
      * Cells in dark blue indicate that the no3 and no2 sum was less than the analytical detection threshold and was not included in TN calculation.
* There were TP values that were less than PO4 values.
  + There are multiple values for PO4 and TP for each day, the number of values are not always equal.
  + If values are “close” TP and PO3 are set equal
  + If the difference is larger they are set to NaN
* Final file from Cheney Reservoir RawNutrientData\_Cheney\_6.26.2017.csv
  + Using Jennifer Grahams phosphorus rules the TP values on these dates were changed to NaN
    - This was done for the 12-03-03, 15-01-13, 18-01-11, 23-01-13

**RawLakeHFData\_Cheney\_QAQC\_09Sep2017.csv**

* These data were originally submitted as the RawSondeData but were actually the HF data from a single depthRaw
* Starting on the first of January 2014 the depth of the sonde measurements is ~1420 meters and continues for the entire 2014 year.
  + A copy error, corrected to 1.55m
* When the hour goes from 23:00 to 0:00 the date still stays the same.
  + The 00:00 reading was changed to the next day so it is Midnight (00:00) of the actual day.
* Some abnormally high Dissolved Oxygen Concentration percentages greater 115% (up to 160%).
  + These will be treated as true values according to the data provider
  + Could be high productivity – chla and phyco values seem high at these times
* Many negative phycocyanin values shown in “negative phyco values.csv”.
  + These are relative values and are a result of the factory calibration
  + Data provider suggests values <0.04 are set to NaN
* On 05-07-01 06:00 chlorophyll\_a\_ug\_l was -0.7. This was changed to 0.
* Conductivity data were provided, but these are not part of the data request. Have removed but could be used later if a need comes up.

**DailyLakeHFData\_Cheney\_QAQC\_11Aug2020.csv**

* Used the R package “openair”, function “timeAverage” to generate daily averages
* Some observations from raw data set had fewer than the maximum possible per day so that daily average values could be skewed if too many were missing. For any daily average values based on less than 83% of available observations in a given day, these values were switched to NaN. Data were collected at 1-hr intervals so maximum number of observations per day was 24 maximum.
* Any date where all variables had no average value was removed.

**RawPhytoData\_Cheney\_QAQC\_15Sep2017.csv**

* The date in this file has added time of day to the minute
  + These were aggregated to the day
* There are several rows starting in Feb. 7, 2007 that are colored red.
  + These days had incorrected depths. They have since been corrected with the red color removed in a new excel file.
* Abnormally high phytoplankton biomasses (great than a Kg per liter).
  + Due to conversion mistake from biovol to biomass. This has since been correct in a new RawPhytoData excel file.
* The date in this file has added time of day to the minute
* The following has cell density > 0 but the biovolume and the biomass are 0. Why?
  + 17/05/2006 Ceratium hirundinella 4407.80445 0 0
  + This was an entry mistake:
    - biovolume = 12694.464 um3/mL
    - biomass = 12.69 ug/L
* Please ask data providers about the 0.3 max depth on 12-01-10. This was an integrated sample (0-0.3 m), and secchi depth on this date was 1.2192 m. I saw in the QAQC notes there were some issues with max depth, perhaps this was missed?
  + The sample was integrated over 2.6 meters not 0.3m
* I could not find "Coelastrum americanum” in Algaebase. Data provider response was “This could be Coelastrum astroideum or Sorastrum americanum, but it is difficult to determine without additional data.” I have changed it to “Coelastrum” to be consistent with original genus.

**RawWeatherData\_Cheney\_USW00003928\_QAQC\_10Sep2017.csv**

* Changed “Cheney Lake” to “Cheney” in the lake column
* Replaced an air\_temp\_c value of -5572.78 on June 21 2004 to NaN
* No wind\_height\_m is available
  + The weather station is 402 meters above sea level
  + Cheney Reservoir is 433 meters above sea level.
* The weather station is ~33 – 38 kilometers away.
  + 33 km is the shortest distance between the edge of the reservoir and the weather station
  + 38 km is the distance between the weather station and the middle of the Reservoir
* In the daily cumulative rainfall data, T’s represent trace amounts of precipitation.
  + These were set to 0
* All wind\_dir values of 360 were changed to 0

**DailyWeatherData\_Cheney\_USW00003928\_QAQC\_06Jun2018.csv**

* Same as raw weather data (because they were already at daily time step). Just added this file with same name format to make easier for derived physics data set when that is built.

**SpeciesList\_Cheney\_QAQC\_15Sep2017.xslx**

* Contains original phyto\_name and the new phyto\_name for analyses based on Algaebase.

**Log of Changes**

20 September 2017

**RawPhytoData\_Cheney\_QAQC\_15Sep2017.csv updated to RawPhytoData\_Cheney\_QAQC\_20Sep2017.csv**

* Dates for 30-08-11 were NaN. These were correct. Thanks Clark!

**RawLakeHFData\_Cheney\_QAQC\_09Sep201.csv updated to RawLakeHFData\_Cheney\_QAQC\_20Sep201.csv**

* On 03-01-02 from 02:00 to 09:00, temperature and chlorophyll a values were 0 and everything else was NaN. All other data were NaN for the month. These 0 values were changed to NaN.

6 June 2018

**Added DailyWeatherData\_Cheney\_ USW00003928\_QAQC\_06Jun2018.csv**

* Provide data at daily time step for the derived physics data set.

17 June 2018

**RawPhytoData\_Cheney\_QAQC\_20Sep2017.csv updated to RawPhytoData\_Cheney\_QAQC\_17Jun2018.csv**

* Some original taxa were changed to new names by way of Algaebase during initial QAQC. These changes to names were made, but in cases of new duplicate names the respective values were not combined (summed) on each date for the new name (e.g., Asterionella and Asterionella sp were both in original names, Asterionella sp was changed to Asterionella). When both originals occurred on the same date, I had failed to sum the densities/biovolumes/biomass on each date. This update corrects that such that each taxa has only one (summed) observation per date.

09 May 2019

**LakeMetadata\_Cheney\_GEISHA\_done.xlsx updated to** **LakeMetadata\_Cheney\_QAQC\_09May2019.xlsx**

* Maximum, mean, and station depth were entered in units of feet not meters, these were changed

12 July 2020

**RawPhytoData\_Cheney\_QAQC\_17Jun2018.csv updated to RawPhytoData\_Cheney\_QAQC\_12Jul2020.csv**

* Original taxa name Dictyosphaerium smithii was not found anywhere, and data provider suggested the species could have been Dictyosphaerium pulchellum. However, This name is currently regarded as a synonym of [Mucidosphaerium pulchellum (H.C.Wood) C.Bock, Proschold & Krienitz](https://www.algaebase.org/search/species/detail/?species_id=139643) and so was changed in this file.

**SpeciesList\_Cheney\_QAQC\_17Jun2018.csv updated to SpeciesList\_Cheney\_QAQC\_12Jul2020.csv**

* Original taxa name Dictyosphaerium smithii was not found anywhere, and data provider suggested the species could have been Dictyosphaerium pulchellum. However, This name is currently regarded as a synonym of [Mucidosphaerium pulchellum (H.C.Wood) C.Bock, Proschold & Krienitz](https://www.algaebase.org/search/species/detail/?species_id=139643) and so was changed in this file.

20 July 2020

**SpeciesList\_Cheney\_QAQC\_12Jul2020.xlsx updated to SpeciesList\_Cheney\_QAQC\_20Jul2020.xlsx**

* Rhodomonas minuta should remain Rhodomonas minuta according to Algaebase (I had for some reason changed it to Plagioselmis nannoplanctica)
* Anabaena circinalis changed to Dolichospermum circinale according to Algaebase (not Dolichospermum sigmoideum)

**RawPhytoData\_Cheney\_QAQC\_12Jul2020.csv updated to RawPhytoData\_Cheney\_QAQC\_20Jul2020.csv**

* Rhodomonas minuta should remain Rhodomonas minuta according to Algaebase (I had for some reason changed it to Plagioselmis nannoplanctica
* Anabaena circinalis changed to Dolichospermum circinale according to Algaebase (not Dolichospermum sigmoideum)

21 July 2020

**RawPhytoData\_Cheney\_QAQC\_20Jul2020.csv updated to RawPhytoData\_Cheney\_QAQC\_21Jul2020.csv**

* Way back when, I (Jason)mistakenly changed the original Rhodomonas minuta to Plagioselmis nannoplanctica because Rhodomonas minuta (v. nannoplanctica) was changed to Plagioselmis nannoplanctica per Algaebase. However, Rhodomonas minuta should have remained Rhodomonas minuta per Algaebase. For Cheney, we had no dates when both species were present, so I went back to the original data and found the dates where Rhodomonas minuta var. nannoplanctica was present and changed these to accepted Algaebase name of Plagioselmis nannoplanctica (5 dates).