**Annie QAQC Notes**

**Most up-to-date files on GEISHA FTP site (as of 17 Jan 2019)**

Annie QAQC Notes\_26Feb2020.docx

RawWeatherData\_Annie\_AWS\_QAQC\_01Dec2018.csv

RawSondeData\_Annie\_QAQC\_04Nov2018.csv

SpeciesList\_Annie\_QAQC\_18Jun2018.xlsx

RawPhytoData\_Annie\_QAQC\_20Jun2018.csv

RawNutrientData\_Annie\_QAQC\_27May2020.csv

DailyLakeHFData\_Annie\_QAQC\_17Jan2019.csv

LakeMetadata\_Annie\_QAQC\_24Oct2018.xlsx

Bathymetry\_Annie\_QAQC\_04Nov2018.xlsx

DailyWeatherData\_Annie\_AWS\_QAQC\_26Feb2020.csv

See the “Log” of changes made to the files since first QAQC’d. These (if any) are located at bottom of this document.

**RawWeatherData\_Annie\_AWS\_QAQC\_01Dec2018.csv**

* Archbold Weather Station in filename reduced to AWS
* 210 instances of wind direction degrees of 360 were changed to 0
* 38 instances of -999 average wind speed were changed to NaN
* 850 instances of there being a wind direction with 0 average wind speed were changed to NaN
* 846 instances of there being a max wind speed >0 when the average wind speed was 0 were set to NaN
* 89 instances of air temperature being -999 °C were changed to NaN
* 25 instances of relative humidity being <0 were changed to NaN
* Raw weather data were at the 15-minute scale 🡪 hourly averages (except for rain, which was summed) were calculated on each day from the raw data to generate an hourly raw data file. Used the R package “openair”, function “timeAverage” to generate hourly averages, including the vector average for wind direction. Rain data were summed for each day. See Technical Note by Stuart K. Grange “Average wind speeds and directions”.
* Maximum wind: maximum wind on 3/21/2008 at 9:00 set to NaN (value of 166 m/s) and only one instance from before and after readings; 7/16/2008 12:00 was 162 m/s (before and after ~5 m/s), set to NaN; 5/11/2010 6:00 speed of 162 set to NaN; 7/2/2010 17:00 value of 159 set to NaN; 10/19/2010 19:00 value of 159 set to NaN; 5/31/2011 5:00 value of 160 set to NaN; 10/8/2011 9:00 value of 164 set to NaN; 11/5/2011 7:00 value of 160 set to NaN; 1/14/2012 value of 163 set to NaN; lots of values over 150 m/s, set all max wind of >150 to NaN
* Air pressure: 2/21/2008 5:00 had a value much lower than previous and after values (820 vs. 1015), which was set to NaN; 10/31/2008 through 4/14/2009 8:00 the values go from normal 1000-1015 to 600-700, these were all set to NaN; low values from 1/30/2012 4:00 through 1/30/2012 8:00 which were set to NaN; 4/1/2014 7:00 low value set to NaN; low values on 9/29/2014 8:00 and 9/30/2014 2:00 changed to NaN; 4/29/2015 4:00 low value set to NaN; 12/3/2015 12:00 low value set to NaN
* Relative humidity: values all the same from 9/10/2011 12:00 through 6/26/2014 23:00, these were set to NaN
* Many instances where maximum wind speed was less than the average wind speed for a given time, these instances of maximum wind speed were set to NaN based on the recommendation from the data providers.

**DailyWeatherData\_Annie\_AWS\_QAQC\_26Feb2020.csv – JB update**

* Rain data showed big drop in magnitude after break in data from on 21 Oct 2010 to 4 April 2011
* Corrected data sent from Hilary Swain, Monika LaPlante (Asst. data manager at Archbold)
* Changed rain data for 1 Jan 2009 to 31 Aug 2015 to data compiled by J. Doubek
* Rain gauge problems in latter half of 2015 so pulled for repairs in 2016 – data suspect after Sep. 2015
* Filled in Sep-Dec 2015 Rain data from Archbold Bio Station, NWS-80236, Main grounds

**DailyWeatherData\_Annie\_AWS\_QAQC\_01Dec2018.csv**

* Used the R package “openair”, function “timeAverage” to generate daily averages, including the vector average for wind direction. Rain data were summed for each day. See Technical Note by Stuart K. Grange “Average wind speeds and directions”.
* 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.

**RawSondeData\_Annie\_QAQC\_04Nov2018.csv**

* Water temperature: 6/11/1984 extreme temperature value at 4 m, ~37 °C when the temperatures above and below are ~27 °C, this was set to NaN; 4m on 7/16/1987 was about 6 C less than the temperature the depth above and below (23.85°C vs. ~29°C), this was set to NaN; bottom depth (19 m) on 9/13/1989 was way higher than the previous depths, set to NaN; 7/17/1996 is missing the epilimnion depths – set to NaN; also set these to NaN because of higher or lower than expected values: 06-08-84 at 8 m depth, 12-11-86 at 7 m depth, 13-05-85 at 10 m depth, 15-10-03 at 3 m depth, 15-10-96 at 18 m depth, 16-07-87 at 2 and 3 m depths, 27-06-84 at 9 m depth
* Dissolved oxygen mg\_l: 5/17/1995 do not think this cast is accurate and the prior and after dates have NaN for DO, set these to NaN; 7/17/1996 the epilimnion is missing, so set this date to NaN; 8/16/2010 some of the epilimnion and all the hypolimnion values are missing, set this date to NaN; 11/15/2013 is a crazy profile, starts at 12 mg/L then goes to 7 mg/L then up to 18 mg/L, this date was set to NaN; 8/15/2014 is missing most of the profile, this date was set to NaN; 9/15/2014 was missing most of the profile too so was set to NaN; 10/14/2014 was missing most of the profile and was set to NaN; 11/16/2015 was missing most of profile and was set to NaN; also set 03-07-01 at 10 m depth and 09-04-84 at 9 m depth to NaN
* Dissolved oxygen percent: 8/27/1984 only 3 depths with values and so set to NaN; 8/2/1985 only 3 depths and so set to NaN; 1/17/1990 some strange low values at depths 6 and 7 m at 25 and 50% lower than others and set to NaN; 7/17/1996 has most of the profile missing, set to NaN; 8/16/2010 only some values present so set these to NaN; 8/15/2014 through 10/14/2014 only a few values present so these were set to NaN; 11/16/2015 only a few values present and were set to NaN; **also set 03-07-01 at 10 m depth and 09-04-84 at 9 m depth to NaN**
* Conductivity: a value of 9999 us/cm on 12/15/2008 at 0.1m depth was set to NaN; a value of 4.01 on 5/13/2013 at 10 m depth was set to NaN.
* pH: a handful of extreme values of 1.52 and less were set to NaN (n = 7).

**RawPhytoData\_Annie\_QAQC\_18Jun2018.csv**

* Three species were removed for original file because they did not have any non-zero values: Anabaena flosaqque, Closterium cf. leibleinii, and Micrasterias spp.
* Ankistrodesum falculus is mis-spelled and should be Ankistrodesmus falcatus
* Fragillaria Nanana is mis-spelled and should be Fragilaria nanana, but the current accepted name in Algaebase is Fragilaria tenera var. nanana
* Gonatozygon cf montaenium is mis-spelled (even in the absence of the “cf”) and should be Gonatozygon monotaenium
* Groenbladia neglecta enlongatum was not found in Algaebase. Changed to Groenbladia neglecta (which there was already a recording of in the records).
* Mallomonas enlongata is mis-spelled and should beMallomonas elongate
* Mallomonas fastigiata is mis-spelled and should be Mallomonas fastigata, but the current accepted name in Algaebase is Mallomonas caudate
* Micrasterias radiata v. gracillicima is mis-spelled and should be Micrasterias radiata v. gracillima
* Pleurotaenium subcornulatum is mis-spelled and should be Pleurotaenium subcoronulatum
* Quadrigula chodati is mis-spelled and should be Quadrigula chodatii
* Quadrigula closteriodes is mis-spelled and should be Quadrigula closterioides
* Staurastrum cingulum floridense should be Staurastrum cingulum var. floridense
* Pleurotaenium acerosum could not be found in Algaebase, WORMS, or GBIF. Contacted Mike Guiry from Algaebase and he suggested Closterium acerosum
* Staurastrum rhomboideum could not be found in Algaebase, WORMS, or GBIF. Contacted Mike Guiry from Algaebase and he says the species does not exist. Reduced to Staurastrum

**DailyLakeHFData\_Annie\_QAQC\_03Nov2018.csv**

* HF data for Lake Annie (temperature and oxygen) were at every 15 minutes. Therefore, daily averages were taken for each depth using R so not so much time was spent QAQC’ing profiles at every 15 minutes for 8 years
* Water temperature (each instance noted was set to NaN): 18 m on 3/14/2008 was -66 °C, which was set to NaN; this same problem was on 4/17/2008 at 18 m depth (value of -66 °C), which was set to NaN and since this is a somewhat common occurrence, all extreme negative values were set to NaN; all values 5 °C and below were set to NaN – these values cannot be all trusted as accurate because many of these instances occurred in summer months in a warm monomictic lake (1981 instances); similarly, there were many very high temperature values (up to 500 °C) that could not be trusted as accurate, values above 35 °C were set to NaN (break in data; 9148 instances); 8/16/2010 through 8/26/2010 at 17m the temperatures begin to read high (all 28 C), these were set to NaN; 8/27/2010 through 9/14/2010 temperature profile goes wonky and then the profiles are all NaN – this time period was all set to NaN too; 7/5/2011 through 9/16/2011 at 5m temperatures start to read much higher and so this depth set to NaN during this time period; 9/17/2011 through 7/18/2012 data are all NaN; after data come back online on 7/19/2012 2 m depth readings are low, 2 m depth set to NaN from 7/19/2012 through 7/25/2012; 2m from 8/18/2012 through 12/25/2012 set to NaN; 8m depth readings from 11/8/2012 through 12/25/2012 then they all go to NaN, this range set to NaN too (all values were 22.09 °C); 9/27/2013 through 12/17/2014 6m starts to drift these were set to NaN
* Dissolved oxygen: extreme low values (<3 mg/L; 96 instances) and extreme high values (> 10.1 mg/L; 114 instances) were set to NaN; there was a string of about 50 days in a row with DO values of 10.66 mg/L which is why that value was included with the break of NaN; 3/24/2009 extremely low value of 6.1 mg/L was set to NaN; 2/16/2010 one high value of 10.1 mg/L set to NaN; 7/15/2010 through 8/29/2010 all the same value set to NaN; 8/22/2011 one low value at 4.3 mg/L set to NaN; 5/16/2012 through 7/12/2012 all same value at 8.17 mg/L set to NaN; 3/15/2013 through 3/23/2013 set to NaN same value; 3/26/2013 through 6/9/2013 set to NaN same value at 8.55 mg/L; 6/26/2014 low value of 4.5 mg/L set to NaN; 6/11/2015 through 8/13/2015 same value of 8.26 mg/L set to NaN; 12/17/2015 through 12/31/2015 same value set to NaN

**RawNutrientData\_Annie\_QAQC\_24Oct2018.csv**

* Only Secchi disk depth is available – all depths set to NaN.

**RawNutrientData\_Annie\_QAQC\_27May2020.csv**

* Added depth\_m column to match other datasets and set to NaN.
* Changed year to four numbers date format = dd-mm-yyyy
* Renamed file to latest version

**SpeciesList\_Annie\_QAQC\_18June2018.xlsx**

* See name changes in the RawPhytoData\_Annie\_QAQC\_18Jun2018.csv above.

**Hypsometry\_Annie\_QAQC\_04Nov2018.xlsx**

* Added this file for Schmidt stability and lake number estimates.

**Log of Changes**

Date

**DailyLakeHFData\_Annie\_QAQC\_03Nov2018.csv updated to DailyLakeHFData\_Annie\_QAQC\_17Jan2019.csv**

* There were two dates 6/10/2013 and 6/11/2013 that had very high hypolimnion temperatures compared to days before and after (temperatures about 10 C higher than previous and post-dates). Temperatures for these two days were set to NA as recommended by the data providers.

**DailyWeatherData\_Annie\_AWS\_QAQC\_01Dec2018.csv updated to DailyWeatherData\_Annie\_AWS\_QAQC\_26Feb2020.csv**

* Rain data showed big drop in magnitude after break in data from on 21 Oct 2010 to 4 April 2011
* Corrected data sent from Hilary Swain, Monika LaPlante (Asst. data manager at Archbold)
* Changed rain data for 1 Jan 2009 to 31 Aug 2015 to data compiled by J. Doubek
* Rain gauge problems in latter half of 2015 so pulled for repairs in 2016 – data suspect after Sep. 2015
* Filled in Sep-Dec 2015 Rain data from Archbold Bio Station, NWS-80236, Main grounds