Export the Results tab of DASLER database as a text document with headers (I forgot to check the box for headers when exporting, so I manually added the names once uploaded in R).

Sample number (sample\_num) is formatted as follows: year, month, day, time, depth. So 200606061300012.0 would mean the sample was collected at 13:00 on the 6th of June, 2006, June, 06 at 12 meters(? units unspecified) below surface.

Location id (loc\_id) is formatted as the lake/stream code followed by the sample location, so CCK20001 is location 20001 on CAESAR CREEK LAKE , OH.

**Chlorophyll A**

There are two methods that have been used to calculate chlorophyll-A: trichromatic uncorrected and spectrophotometric acid method. The two methods don’t seem too different, with some sites giving the exact same values for both methods on a given sampling date. Generally, the spectrophotometric method is more common in earlier years, while the trichromatic method is more common in recent years (after 2000).

I’m wondering if we can combine the values a single ChlA column. Looking at the literature, it appears that the trichromatic method is a type of spectrophotometric analysis and it also uses acid. “Uncorrected” means it hasn’t been corrected for pheophytin, but there’s no indication of if the spectrophotometric method was corrected or uncorrected. Purely based on definition, I don’t see a difference. The values for each method in the data are similar. Of 5,953 samples that included both a tri and spect value, 6 samples had greater spect values (ranging from 0.02 to 39) and 126 had greater tri values (ranging from 0.13 to 56.45). Given the small percentage of disagreeing values (2%), I have combined the two chlA columns into one.

BLOOM

The EPA considers it to be a minor bloom if chlA is between 2 and 5 ug/L, a moderate bloom if chlA is between 5 and 50 ug/L, and a severe bloom if chlA is above 50. There are values as high as 500 uq/L in the dataset, but I don’t feel confident in removing them.

**Temperature**

There’s a few extreme temperature values above 40 C (104 F) including one as high as 1937. There are also -99.9 values, which I assume are meant to signify missing values. Even in the winter, you will rarely find water temps below 30 F, so I have removed temperature below -2 C and above 40 C (104 F, the standard hot tube temperature).

**Dissolved oxygen**

The amount of DO water can hold is dependent on temperature, pressure, salinity, and external processes (aeration, decomposition, and photosynthesis). Cold water can contain the highest concentration of DO at equilibrium, with roughly 14.5 mg/L of DO at 0 C. However, there are DO concentrations in the hundreds in the data. There are DO meters than can read up to 50 mg/L. While this seems like an extreme in itself, I am going to exclude values below 0 and above 50 mgL.

**Secchi depth**

There’s a ton of 0 values, which seems suspicious to me because that means as soon as they dipped in the secchi disk they couldn’t see it anymore. Some of these samples also report a high TSS so I guess it’s possible. I don’t fully buy it but I also don’t have a strong enough reason to exclude them.

**TDS**

There are some high values here (one over 7000) but when I looked into “possible” TDS concentrations, these aren’t out of the question. The EPA considers anything above 1200 mg/l to be unfit as drinking water and brackish and marine waters can well exceed 7000 so it’s a possibility especially if these samples were collected after some adverse weather event.

**Phosphorous**

Both mg/l and ug/l are given as units for total phosphorous. It additionally appears that some samples have been assigned the wrong unit and that some samples with the right unit are still not possible. Look at the data and the range of possible p levels I have decided to:

* Convert values less than from mg/l to ug/l (x \* 1000)
* Recode value between 1 and 1000 to be all ug/l
* Exclude all values above 1000 regardless of units

**Depth**

Depth is also not a solely reliable value. There are some 555, 666, 777, 888, and 999 values used by two reservoirs in the 90’s that are clearly place holders. Visualizing the data confirms this, so I removed them. Also I’m not confident in the depth units. I’m leaning towards meters but it could be feet or yards for all I know.

For the time-series plots, the SOW mentions we may need to choose the best depth for which to display data from. When you graph all of the data, the DO and temp does look especially messy, but that’s not really resolved by choosing just one depth (surface) because of the variation in temperature (and therefore DO) over the course of a year. At the DEP, we would take a temp & DO depth profile to determine where the thermocline is and then only take water from above (epilimnion) for analysis at the lab. However I don’t think filtering for samples from the epilimnion will work since some reservoirs may not stratify and since reservoirs that do stratify are sampled year round and will therefore have periods when they are not stratified in the data. The best option may be to have a default (like all data color coded for depth or just surface) and then the ability for the users to select/input a depth or depth range.

**Trophic State**

The thresholds for trophic states are a little confusing since they are given as a single value and not a range (see table 1). I took this to mean that once a value broke the threshold, it would be categorized as that trophic state. For example, a P concentration of 28 ug/l and 80 ug/L would both be classified as mesotrophic, while a P concentration of 25 ug/l would be classified as oligotrophic. Values below the threshold for oligotrophic are also classified as oligotrophic since you can’t do better. This seems reasonable given the range of values cited in the EPA’s Nutrient Criteria Technical Guidance Manual, Lakes and Reservoirs, First Edition. I could see flipping it to say that the given threshold is the upper bound of that trophic state (28 and 25 ug/l P would now by classified as eutrophic and mesotrophic, respectively) but you run into the same but flipped issue where anything above the threshold for eutrophic would still be classified as eutrophic since you can’t do worse. I plan to dig into the literature a bit more.

|  |  |  |  |
| --- | --- | --- | --- |
| Parameter | Oligotrophic | Mesotrophic | Eutrophic |
| Total Phosphorous (ug/l) | 8 | 27 | 84 |
| Total Nitrogen (mg/l) | 0.66 | 0.75 | 1.9 |
| Chlorophyll-A (ug/l) | 1.7 | 4.7 | 14 |

Regardless of which way the thresholds are used, some reservoirs have disagreeing classifications based on each parameter (e.g. a reservoir is Meso for P and N but Eutro for chlA; or it’s eutro for P, missing N data, and meso for chla). I chose to take the median value as the final classification. Reservoirs that had data for only two parameters and which disagreed by one step were given a hyphen (meso-eutro).