Review of S17 COnductivity Data Gap from 2020

Curtis C. Bohlen

3/18/2021

# Load Libraries

(Code omitted)

#> -- Attaching packages --------------------------------------- tidyverse 1.3.0 --  
#> v ggplot2 3.3.3 v purrr 0.3.4  
#> v tibble 3.0.5 v dplyr 1.0.3  
#> v tidyr 1.1.2 v stringr 1.4.0  
#> v readr 1.4.0 v forcats 0.5.0  
#> -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
#> x dplyr::filter() masks stats::filter()  
#> x dplyr::lag() masks stats::lag()  
#> Registered S3 method overwritten by 'GGally':  
#> method from   
#> +.gg ggplot2

# Introduction

During annual data review, LCWMD staff identified a substantial gap in conductivity data from Site S17. Here we examine related data to evaluate the decision by our monitoring contractor, GZA, to omit those data from the record.

The core issue is this, from LCWMD Director, Peter Carney: > “There is a big hole from 9/11/2020 1:30PM to the end of the day on 9/30/2020.”

GZA is the monitoring contractor for LCWMD. Their data manager responded as follows:

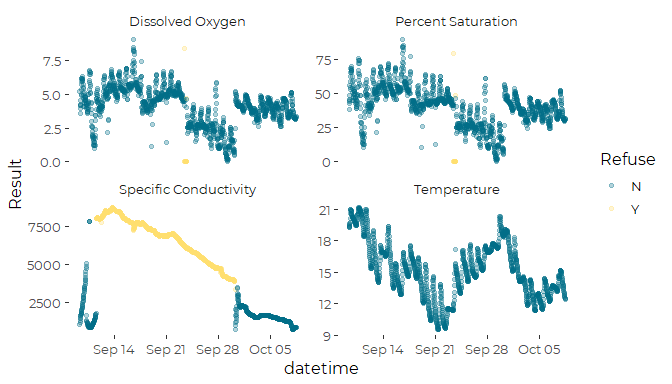
“Our upload files show that we successfully uploaded all of this data (excel sheet attached), however we did reject most of the data in the timeframe you specified. We rejected all specific conductivity (SpC) data from 9/11/2020 @ 2:00 PM through 9/30/2020 @ 5:00 AM. This data was rejected due to extremely high values recorded by the logger that were not justifiable without a reason, in conjunction with the sharp jumps to and from those high values rather than a gradual rise or decrease. In the notes column we indicated it was rejected due to “Unreliable data; suspected interference/malfunction.”

GZA shared selected logger data from S17, from period beginning on September 9th, and ending on October 8th. We review that data here quickly.

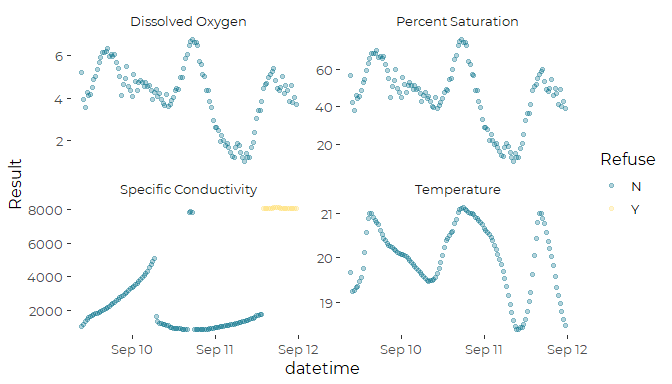
# Load Data

(Code omitted)

# Plot Data



## Zooom to the First Few Days

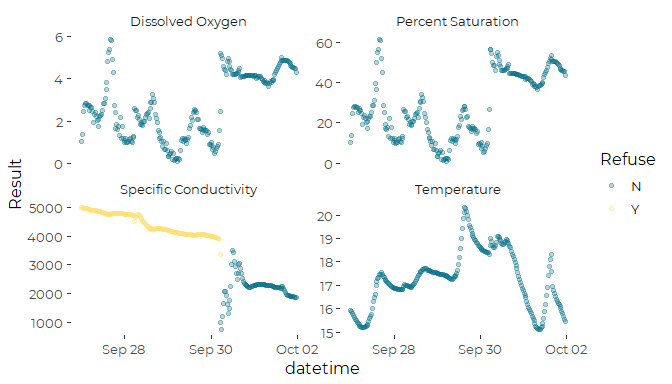


We see a diurnal pattern in DO and temperature, (both dropping an night). That matches expectation for an urban stream, with substantial respiration at night under warm water conditions. That suggests those sensors are working correctly.

On the other hand, we see an improbable climb in conductivity, a sudden drop, a leap, another drop, and then consistent elevated values. That looks highly unlikely to be real data.

So, if there is a malfunction, it applies principally to the Specific Conductivity data, and not to the other parameters monitored by the Sonde. That limits the kinds of malfunctions that are possible. Often, when one sensor begins to malfunction, other sensors also show problems.

## Zooom to the (Near) End of the Record

I zoom into the period around September 30th, where there’s a big break in the data. 

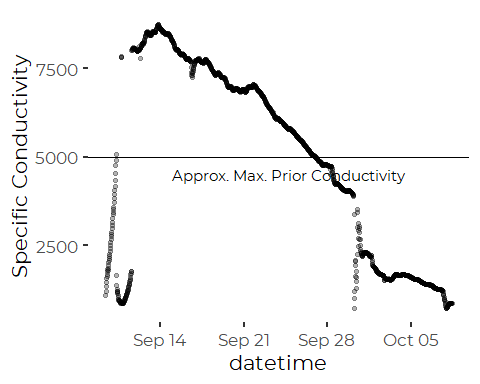
There is an obvious break early on September 30th. The break affects dissolved oxygen and conductivity, and possibly also temperature.

The jump in DO values suggests the prior dissolved oxygen data may be unreliable. GZA flagged several weeks of DO data, but did not reject it. I might chose to reject all the DO data from this period, in addition to rejecting the conductivity data.

I wonder if the sudden change on September 30th corresponds to a monthly maintenance visit or something. The data looks similar to what I see in raw data when sondes are out of the water.

Certainly it is hard to come up with any mechanisms that would cause a drop and spike in conductivity and a jump in dissolved oxygen other than some sort of equipment malfunction.

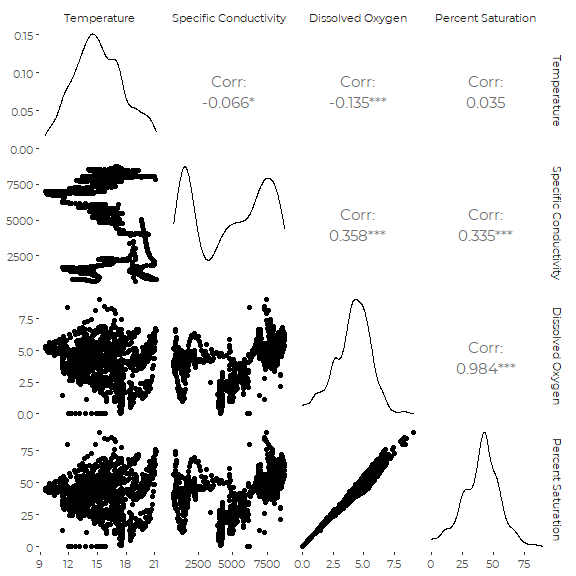
## Conductivity Levels Higher than Historic Values

The highest conductivity levels we see are generally below about 3000 microsiemens. The highest daily median was around 5000 microsiemens. 

That certainly suggests we have a problem. Our highest conductivity observations have historically been observed in spring melt events, not fall. Short of a large illicit discharge of chlorides into the stream, I can not imagine a mechanism that would generate such high values for a short period of time with sharp onset and ending.

# Look for Correlations

We can look a little closer, and examine whether there are correlations between the questionable conductivity data and the other parameters.  
Correlations may indicate a larger breakdown of electronics on the Sonde.



The only strong correlation is the expected link between DO and Percent Saturation. However, the correlations between the oxygen parameters and conductivity are troubling. It suggests the dissolved oxygen data in this period may also be suspect, but it could also be a chance temporal correlation.

# Conclusions

1. The conductivity data is almost certainly corrupted. I recommend dropping most of these conductivity from the QA/QC’d data set.
2. I would drop everything from the initial spike in conductivity values, beginning on September 9, through the (smaller) sudden drop on October 7th. That might toss out a few “good” values from September 10 and 11. Otherwise, it may be somewhat arbitrary deciding which to retain and which to drop, especially on September 9. That is a longer period of data than GZA chose to omit.
3. I am concerned about the quality of the dissolved oxygen data during this period as well. I highlighted the correlation between dissolved oxygen and the (spurious) conductivity data, above. I also note that many of the dissolved oxygen values in the period with strange conductivity data are flagged as having a high relative percent difference (Up through October 3rd). The values are already flagged with a data quality flag. We might want to be careful about trusting those dissolved oxygen values.
4. It would be beneficial to relate these observations to what was observed at other locations, to see if any other sites show wonky values at this period. I doubt they were, but it’s another good way to evaluate these data
5. It’s not at all clear when we should start accepting data after this period of malfunction. I would be interested in knowing when the equipment was maintained. It is possible that some of the problems reflect low batteries, partial fouling of the equipment, or partial burial by stream sediment. Any of those problems would get fixed the next time there was a site visit.