WasteWater Analysis

Steve Goldstein and Marlin Lee

6/1/2021

This report is meant to give an overarching view of the data analysis done and the attempts to reduce noise and find relationships in the data. We find that the data has [I am not sure how we want to summarize this]

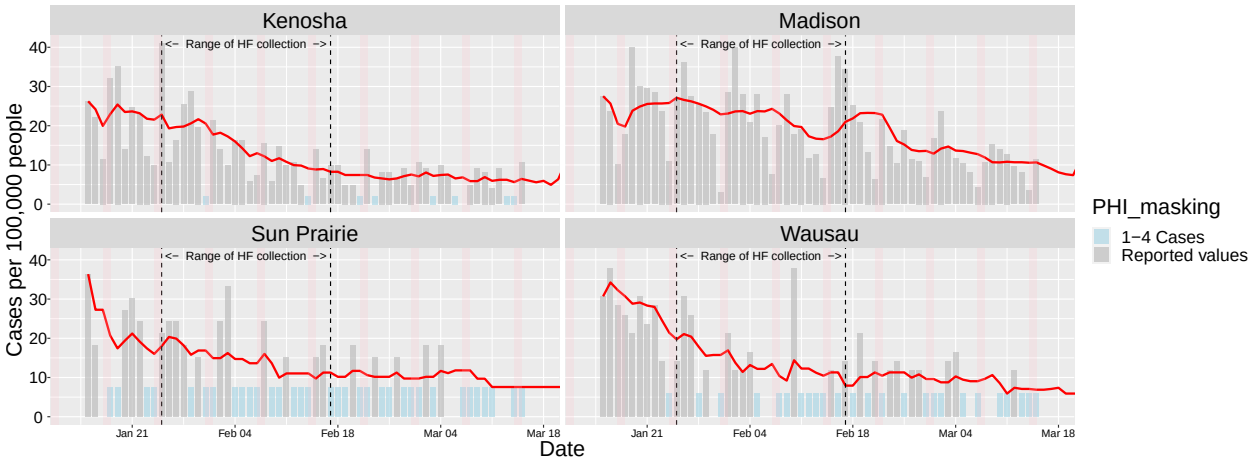
We focus on the high frequency data from January 25 2021 to February 16 2021 as it has many features which could help reduce noise in both the case data and waste water data. The data has 3 well replicates for each of the 3 filter replicates which means for each day there is 9 replicates of the sample. This make the data vital as it minimizes any error in data collection which gives the best shot for understanding an underlying trend.

We furthermore focus on the large location centers where number of cases stays mostly above 5. Due to PHI masking the majority of the case data of smaller location needs to be reduce. Additionally the lower population would be expected [Prove this?] to have higher day to day variance which makes it unsuited for showing a causal relationship.

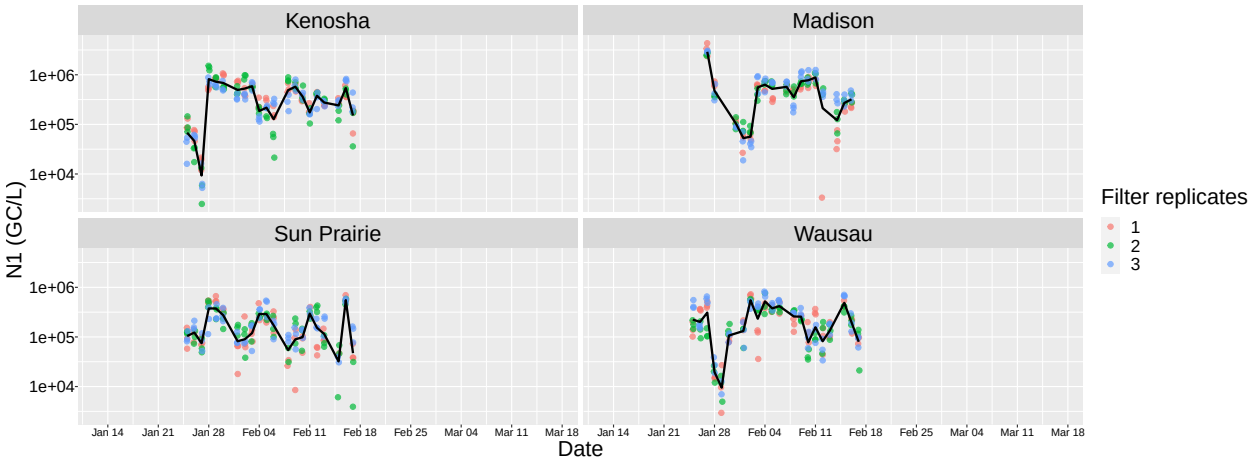
below shows that within the time frame of the high frequency measurements there is no clear 7 day average trend among the 4 Sites.

[I used collected cases for this section of the report. Should I switch to reported cases to get parity with other data source]

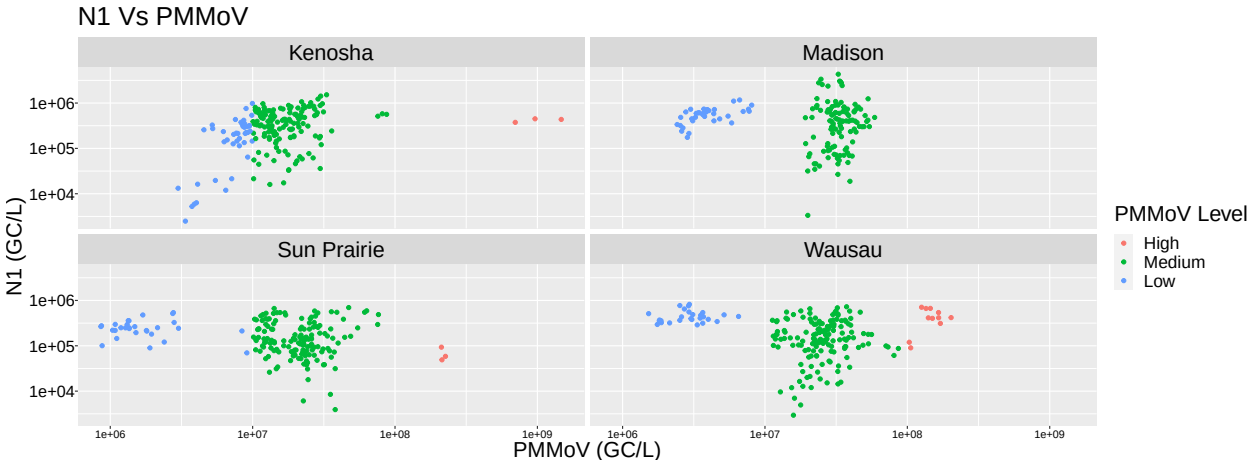
[Might want to limit the graph to just the days we have N1 data on so all date based graphics have the same ending points. Or at least a smaller buffer]

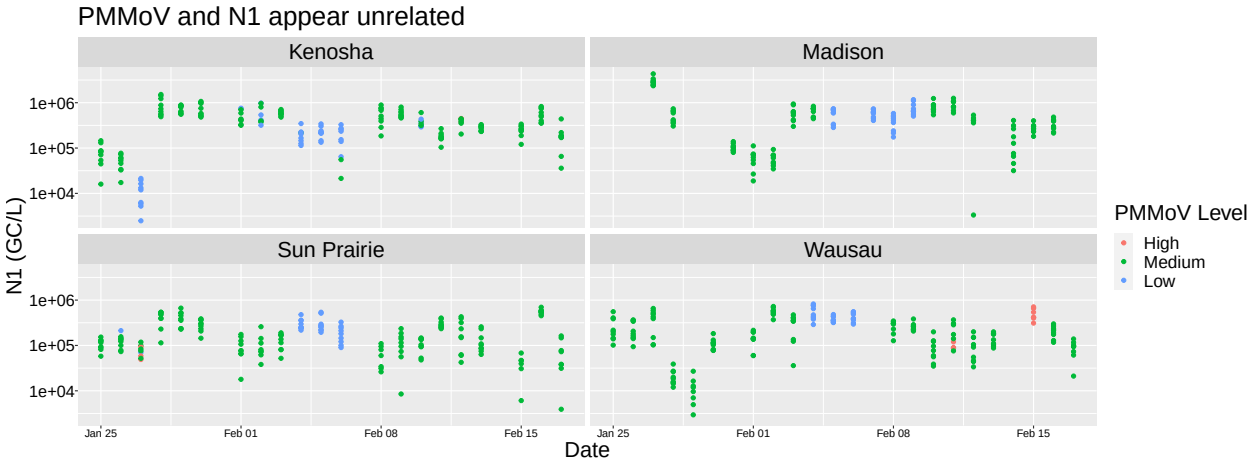


The N1 data has an unnatural dip on January 28 that should be ignore as testing error. otherwise we see chaotic movements that do not mirror the Covid-19 case data. Furthermore the the replicates have low variation which implies the problem is not in the PCR process. There are three possible explanations for these dips. The first explanations is that the waste water is naturally highly chaotic in capturing N1 loads. The second explanation is that the variation is just the true N1 relationship in the community. Finally the sampling process of waste water might fail to capture the true concentration. This might be from collection site issues or from flow issues.

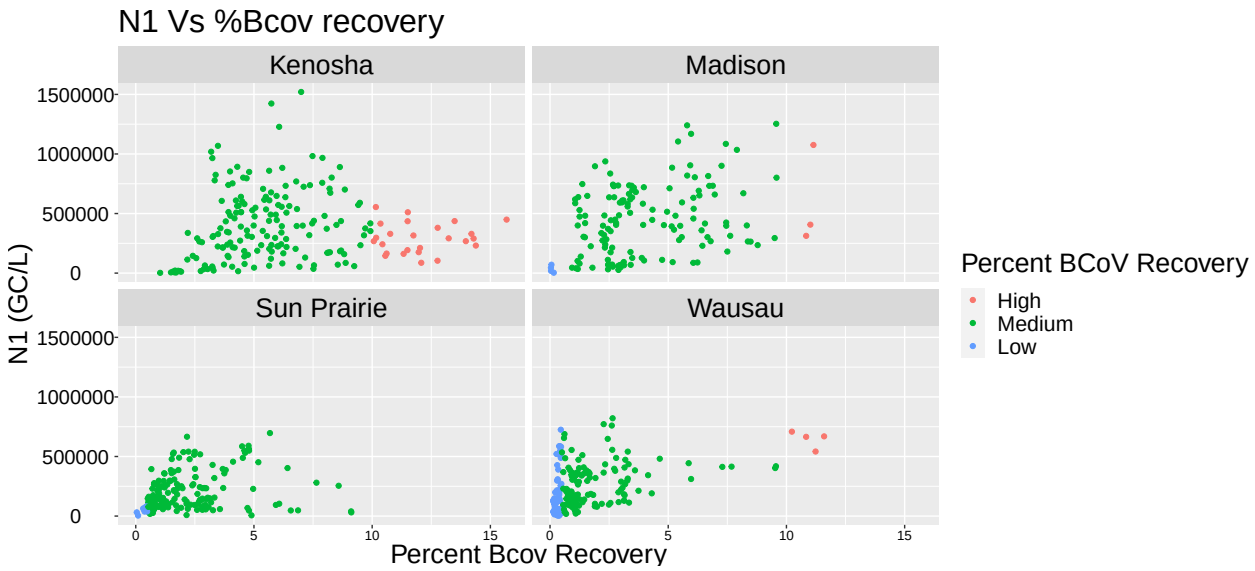


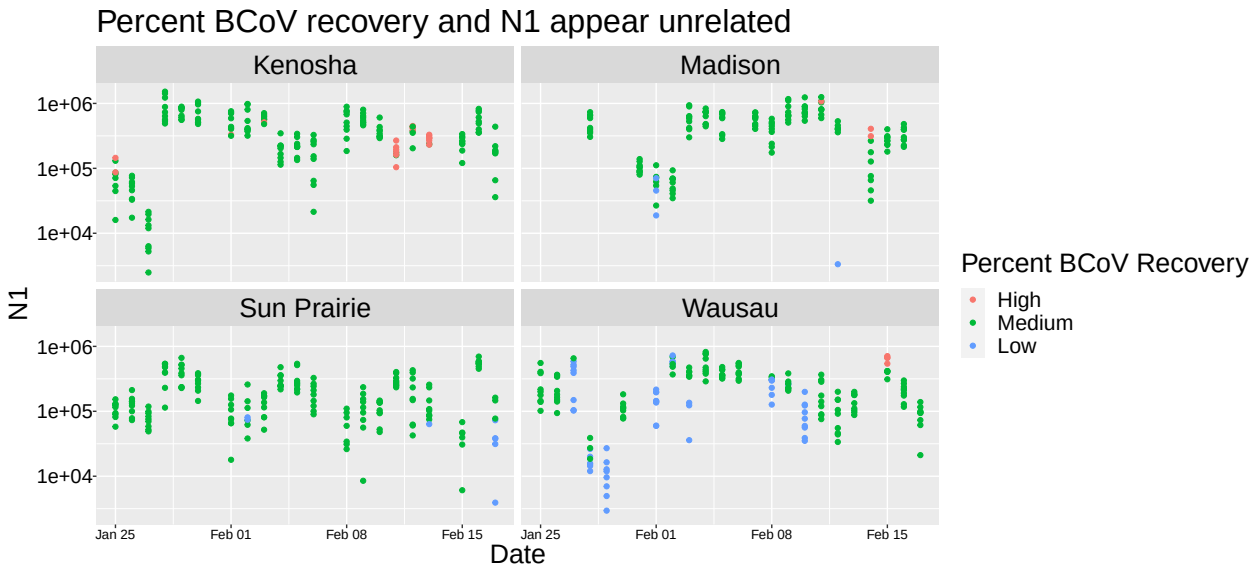
To confirm there is no way to explain the weird movements of the N1 concentration we looked to see if they could be explained by abnormal Percent BeCoV recovery or PMMoV [Use full names]. Below shows unusual PMMoV values do not explain the unusual N1 values.



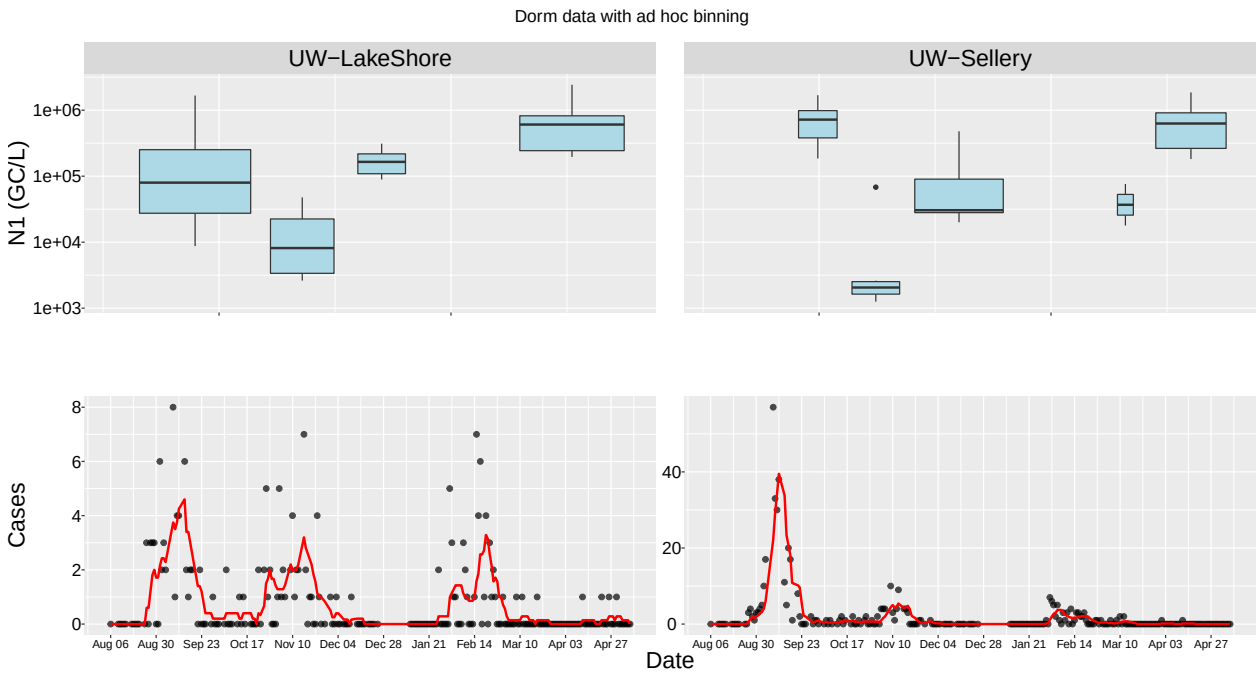


BCov has no clear outliers and the extreme values of BCov have no relation to the N1 data. This means that its use as a control is limited.





One solution to the noisy data is to instead look at boxes of days. Doing this shows a clear relationship in the US Dorms Data below. However when it comes to larger locations this method seems less effective.


Dorm data with ad hoc binning
UW−LakeShore
UW−Sellery
N1 (GC/L)
1e+06
1e+05
1e+04
1e+03
Cases
8
6
4
2
0
40
20
0
Aug 06
Aug 30
Sep 23
Oct 17
Nov 10
Dec 04
Dec 28
Jan 21
Feb 14
Mar 10
Apr 03
Apr 27
Date