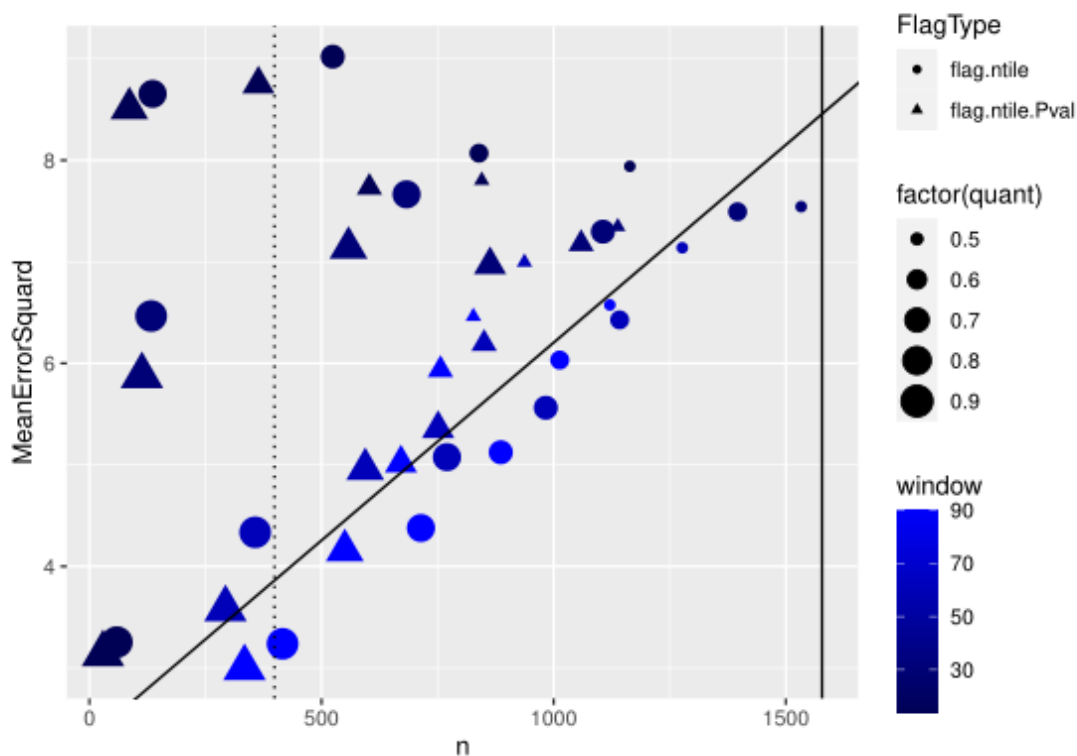


# Covid-19 Wastewater Flagging Method Optimization

Marlin Lee, Kyllan Wunder, Abe Megahed

University of Wisconsin Data Science Institute - October 2022

**Introduction:** In this analysis, we perform a comparative analysis of various wastewater flagging techniques with the goals of (1) making the wastewater flags match the case analysis flags and (2) reducing false positives in the reported wastewater flags. As the adoption of at-home tests increases and case reporting decreases, wastewater becomes a more important metric for determining genuine case numbers. Accurate flagging methods are critical for policymakers to make better-informed decisions based on predicted case changes.



The angled line is a linear regression line based on the 60 and 90-day intervals. The vertical lines represent the total number of case flags (1578) and the proportion of wastewater samples to wastewater flags if they had the same ratio as cases to case flags (389).

## **Flagging methods**

The DHS has created flagging methods based on wastewater measurements and aims to supplement existing case flagging metrics. The wastewater method uses a 5-day regression to see if the predicted change is above 100%. If so, it checks if the last 3-day average is larger than the K-day Q quantile (where K and Q are customizable parameters). Then if both are true, it is labeled as a flag.ntile. If the Linear regression has a p-value below .3, it is labeled as a flag.ntile.Pval. We explored five quantile values (.5, .6, .7, .8,..9) and four window values (14, 30, 60, 90 days) for a total of 40 ways to flag models.

## **Graph explanation**

The graph shows a distinct linear relationship between the variance and the number of flags produced (denoted on the graph by the diagonal regression line). This relationship makes it possible to balance the number of flags against the allowable variance or to determine the expected variance for a particular number of flags. We calculate the squared distance each wastewater flag is from the nearest case flag to find the optimal combination. On investigation, it becomes clear the major issue with current options is the number of flags reported. As quantile size increases, the data fits this trend more, and with a smaller quantile size, the mean squared error (MSE) is much more varied.

## **Matching wastewater flags to case flags**

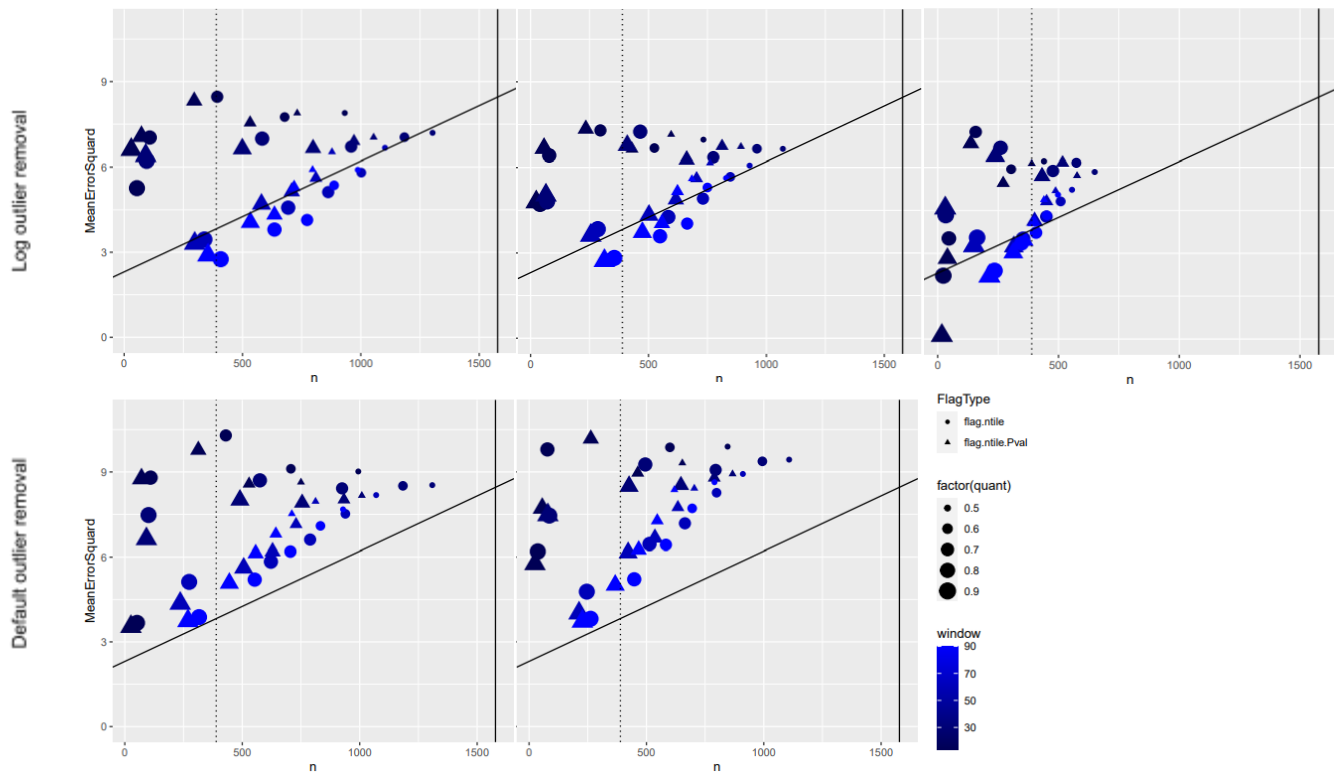
The case flag method has 1578 flags, so the chosen wastewater method should have an equal amount. None of the methods create that many flags; most produce multiple times less due to the lower frequency of wastewater sampling. The dashed line shows the number of wastewater flags we would expect if the cases to case flags ratio were the same as the wastewater samples to wastewater flags.

## **Linear model explanation**

We created two linear models to explain the relationship between the number of points and the error. First, we used the whole data set and found minimal relation (.12  $R^2$ ). Second, we excluded the 14 and 30-day windows and got a very successful linear model (.80  $R^2$ ) shown in the previous plot.

## **Removing outliers**

We tried our package outlier method (N1 / N2 nearest neighbor comparisons) to see if it improved the model error. By doing this, we aim to reduce the MSE while minimizing the decrease in the number of flags. Using our traditional method resulted in the flagging method doing meaningfully worse with fewer flags and a higher MSE. However, a modified method that instead looked for spikes in the log of the data succeeded in reducing the MSE, but at the expected cost of the number of flags. It is unclear if the increase in model complexity and the decrease in the number of flags is worth the reduction in MSE, but it is worth considering.



As we increase the threshold the number of flags decrease, especially for the more permissive methods. The default outlier removal fails to decrease the MSE as seen by the points moving left but not down. However the log outlier method succeeds at pulling the points towards the center

## Conclusion

We found no one best model because there is an inherent tradeoff between MSE and the number of flags created. However, a great candidate is the 90-day window 90% percentile flags which are around the correct number of adjusted flags and have the lowest MSE score.