

Covid-19 Wastewater Data Smoothing

Marlin Lee, University of Wisconsin Data Science Institute - July, 2022

Wastewater data is inherently noisy due to a variety of factors including ambient temperature, flow rate, and precipitation. This makes determination of critical trending patterns difficult. Smoothing techniques can be used to minimize short term fluctuations so that significant longer term trends can more easily be identified.

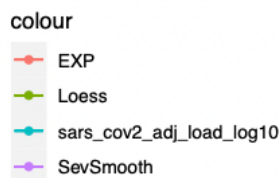
In this analysis, we look at using smoothing techniques to improve the ability to discern significant wastewater data trends and compare three different techniques - rolling average, exponential smoothing, and Loess smoothing.

The Data

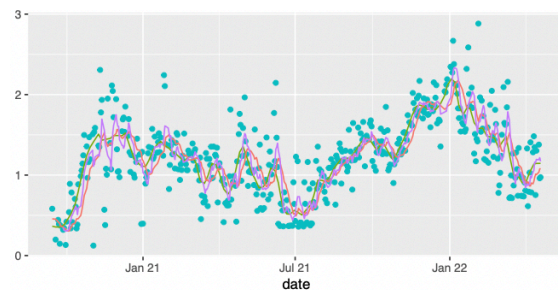
The graphs to the right display the data using various smoothing techniques as well as the data downsampled to 2 days per week, using the same smoothing techniques.

Note that the Loess (green) and exponential (orange) curves are noticeably smoother than the seven day rolling average (purple).

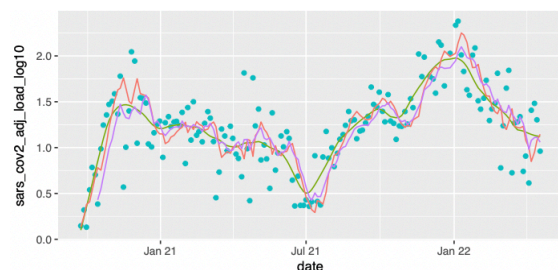
The seven day rolling average displays a periodic spikiness which is an artifact, perhaps due to a regular weekly periodic variation.



Original



Downsampled to 2 Days / Week:



Smoothing Techniques

Rolling Average

The first technique is to compute a rolling / moving average. Using this technique, we average values in proximity of the point of interest. We used a 7 day rolling average, which is computed at each point by finding the mean of the previous 7 days. https://en.wikipedia.org/wiki/Moving_average

Pros:

- simple, efficient

Cons:

- requires the choice of an arbitrarily sized window.

Exponential Smoothing

The next technique that we looked at was exponential smoothing. The exponential smoothing technique is similar to the moving average where we average the values of neighboring data except that we include all of the values in the series instead of just the values within an arbitrary window and we apply an exponentially decreasing weight factor to values that are more distant from the point of interest.

https://en.wikipedia.org/wiki/Exponential_smoothing,

https://www.researchgate.net/publication/310638788_Forecasting_with_Robust_Exponential_Smoothing_with_Damped_Trend_and_Seasonal_Components

Pros:

- Does not require the choice of an arbitrarily sized window

Cons:

- Forecasts lag since we're averaging previous results

Loess Smoothing

Loess smoothing or locally weighted scatterplot smoothing is a technique that applies a linear regression over a moving window of data points. https://en.wikipedia.org/wiki/Local_regression

Pros:

- No lag effect since we are sampling before and after each point.

Cons:

- Computationally expensive.

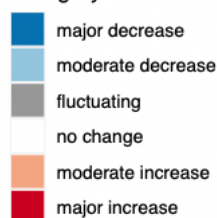
Smoothing Results Using 7 Samples / Week (Daily Sampling)

To the right, we see the results of various smoothing techniques applied to a dataset with 7 samples per week.

Generally, we see a pattern of increasing smoothness when progressing through the series of smoothing methods described below.

- With no smoothing, we can't distinguish any trend information since the classification merely jumps between the two categories of “no change” and “major increase”.
- The rolling average starts to identify a few trend regions, showing a significant region of moderate increase to the far left, the middle right and a smaller region to the far right.
- The exponential smoothing does a good job of identifying broader regions of both moderate increase and moderate decrease.
- The Loess smoothing produces similar results to the exponential smoothing, which is a good indicator that the trends that are identified are valid. It identifies fewer small regions of “no change” than the exponential smoothing.

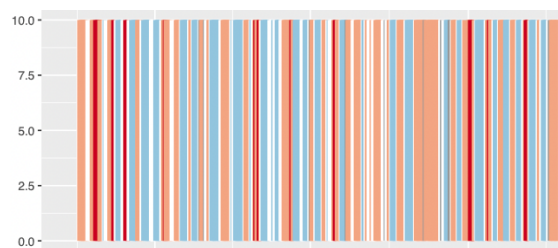
Category



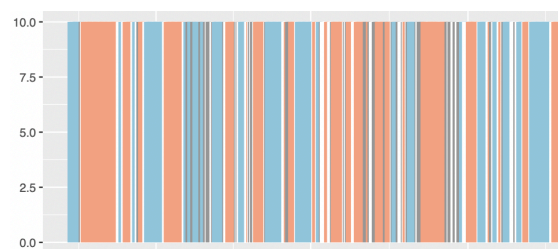
No Smoothing



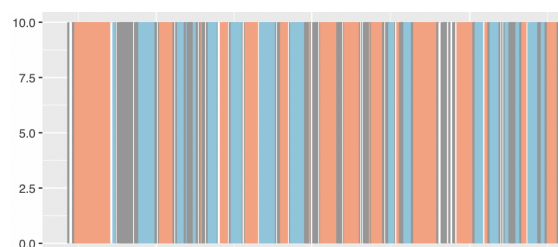
21 Day Rolling Average



Exponential



Loess



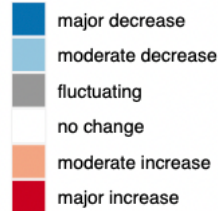
Smoothing Results Using 2 Samples / Week (Downsampling)

To the right, we see results from application of the smoothing techniques to downsampled data (2 days per week).

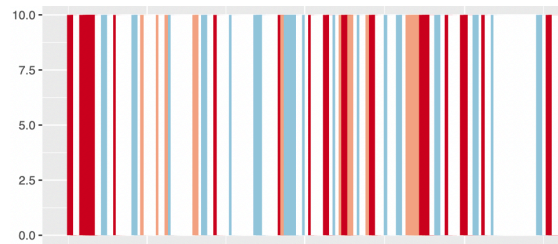
Since there are fewer samples, the data shows lower frequency variation than the full dataset. However, the general trending patterns are similar.

This implies that smoothed and downsampled or lower frequency data is a valid indicator of trending information.

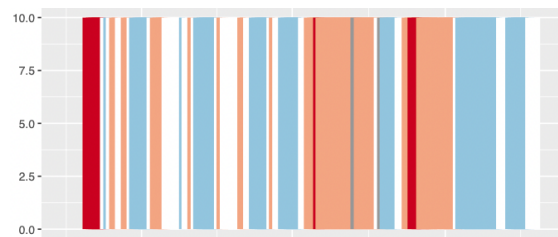
Category



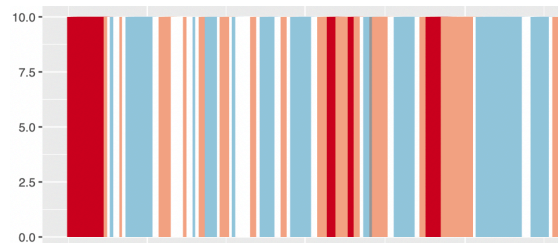
No Smoothing



21 Day Rolling Average



Exponential



Loess

