

Analysis of Covid-19 High-Frequency Data

Marlin Lee, Kyllan Wunder, Abe Megahed

University of Wisconsin Data Science Institute - October 2022

Introduction

This document seeks to use the high-frequency dataset (HFG) to understand the relationships between the Covid-19 measurements and the other covariates. This understanding could help to decide what normalizers to use and how to use them. The HFG data consists of 10 sites containing 9 measurements per day broken down by collection group for a little over a month.

- The main takeaways are:
- The covariates don't relate to the covid signal any more than the population.
- The detrended data seems to be largely independent.

Methods

To look for correlations, we broke the signal down into two components:

- The trend of the data, this is created using a loess smoothing.
- The detrended data.

We take the log of the data because the noise appears to be log-normal, meaning a more apparent trend. The heatmap below shows the correlation between the Covid-19 signals and the covariate components.

Results

These covariates are meant to control for the population, so it acts as the baseline. We see the data has a .25 mildly meaningful correlation, approximately the correlation the other covariants have. None of the detrended data correlates meaningfully, supporting the view that they are caused mainly by noise.

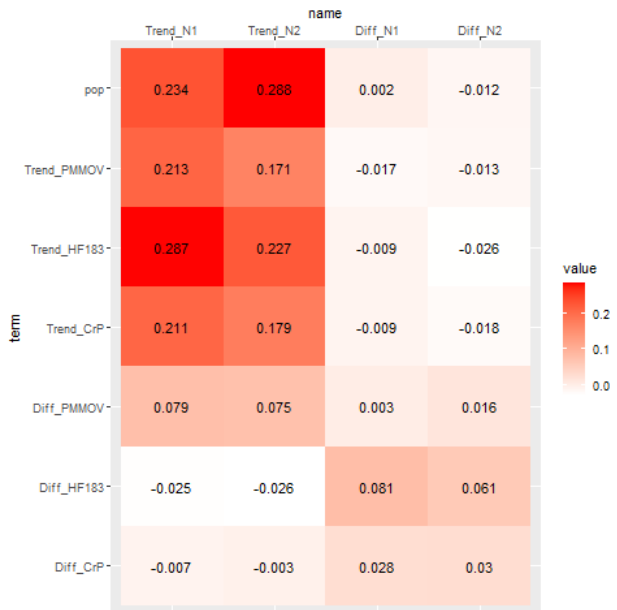


Figure 1: Correlation between Covid-19 and population covariates

Looking at the relationship between HF183 and CrP you see a distinctly strong relationship between the trend and the noise. We need more understanding of the underlying process to know if this is expected.

	log	base
trend corr	0.8801775	0.7616367
detrend corr	0.4324156	0.1648089

For a baseline, we showed the relationship between N1 and N2, which clearly shows a solid correlation due to them measuring the same thing.

	log	base
trend corr	0.9268483	0.9824389
detrend corr	0.3407199	0.3446168

We also want to look at the variance of these processes. We can capture variance at three steps of the process:

- At the lowest level (well variance), we have variance in PcR test.
- Next, we have the variance at the collection step (filter variance).
- Finally, we have the total variance of the system.

Each level contains lower levels of variance (trend variance). Below is a table showing each signal's three levels of variance.

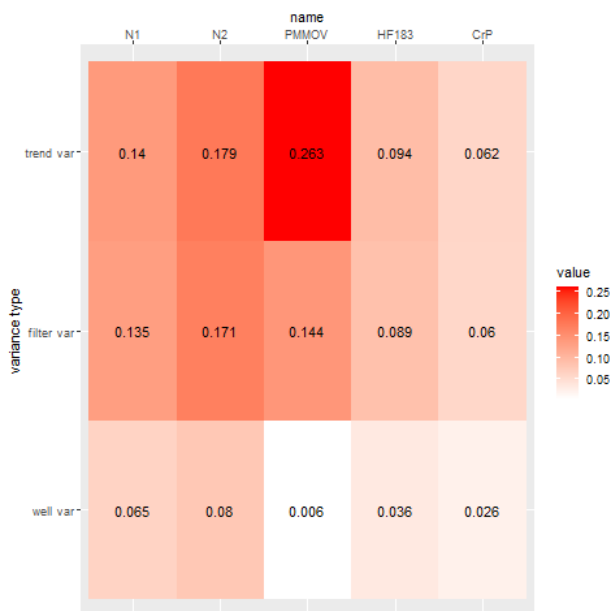


Figure 2: variance of wastewater measurements