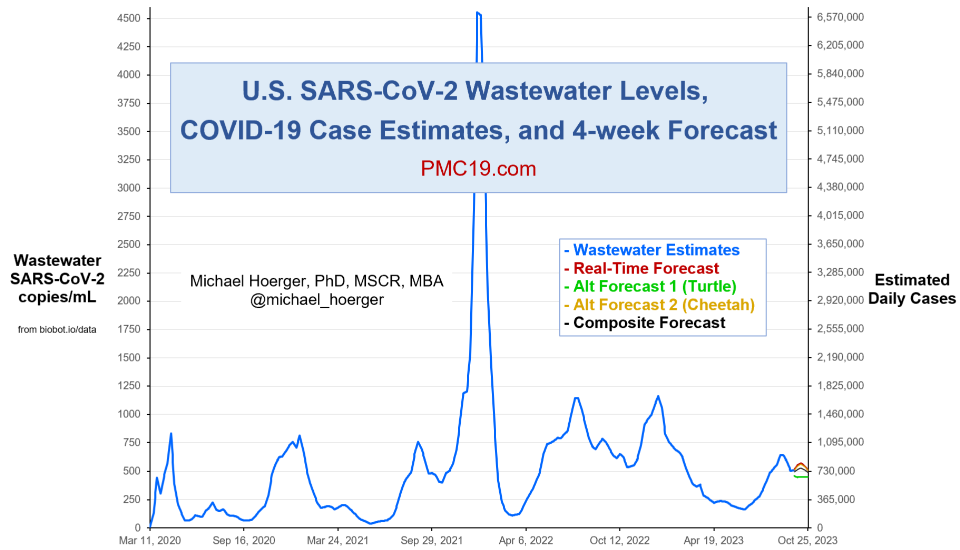
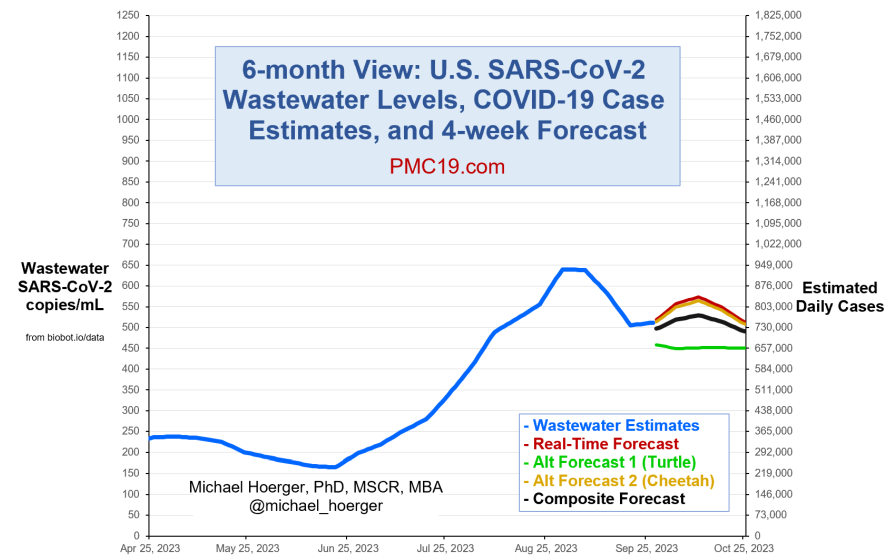
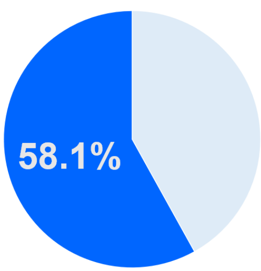
Michael Hoerger, PhD, MSCR, MBA, Pandemic Mitigation Collaborative

U.S. SARS-CoV-2 Wastewater Levels, COVID-19 Case Estimates, and 4-Week Forecast: Report for October 2, 2023, pmc19.com/data







There is more COVID-19 transmission today than during 58.1% of the pandemic.

CURRENT ESTIMATES FOR September 27, 2023
Wastewater Levels (copies/mL)
512
New Daily Cases
745,000
% of Population Infectious
1.56% (1 in 64 people)
New Daily Long COVID Cases
37,000 to 149,000

WEEKLY ESTIMATES FOR September 27, 2023
New Weekly Cases
5,200,000
New Weekly Long COVID Cases
261,000 to 1,043,000

2023 CUMULATIVE ESTIMATES AS OF September 27, 2023
Total 2023 Cases To Date
220,500,000
Total 2023 Long COVID Cases To Date
11,025,000 to 44,100,000

4-WEEK FORECAST FOR October 25, 2023
Wastewater Levels (copies/mL)
491 (-4% lower)
New Daily Cases
714,000
% of Population Infectious
1.49% (1 in 67 people)
New Daily Long COVID Cases
36,000 to 143,000

What's the COVID Risk in an Office or in a Classroom?

Number of People	Chances Anyone is Infectious	Number of People	Chances Anyone is Infectious
1	1.6%	25	32.5%
2	3.1%	30	37.6%
3	4.6%	35	42.3%
4	6.1%	40	46.7%
5	7.6%	50	54.4%
6	9.0%	75	69.2%
7	10.4%	100	79.2%
8	11.8%	150	90.5%
9	13.2%	200	95.7%
10	14.5%	300	99.1%
15	21.0%	400	99.8%
20	27.0%	500	>99.9%

Cite as: Hoerger, M. (2023, October 2). *U.S. SARS-CoV-2 wastewater levels, COVID-19 case estimates, and 4-week forecast: Report for October 2, 2023*. Pandemic Mitigation Collaborative. http://www.pmc19.com/data

The CDC and Biobot are in a transitionary phase on wastewater reporting. Biobot is now reporting Wednesday data on a 5-day delay instead of a 1-day delay. Expect PMC updates on late Monday or early Tuesday in the near-term. If reporting quality deteriorates, we will move to a different wastewater tracking source. That may cause a one-time delay.

General Commentary:

U.S. wastewater levels are higher than during 58% of the pandemic:

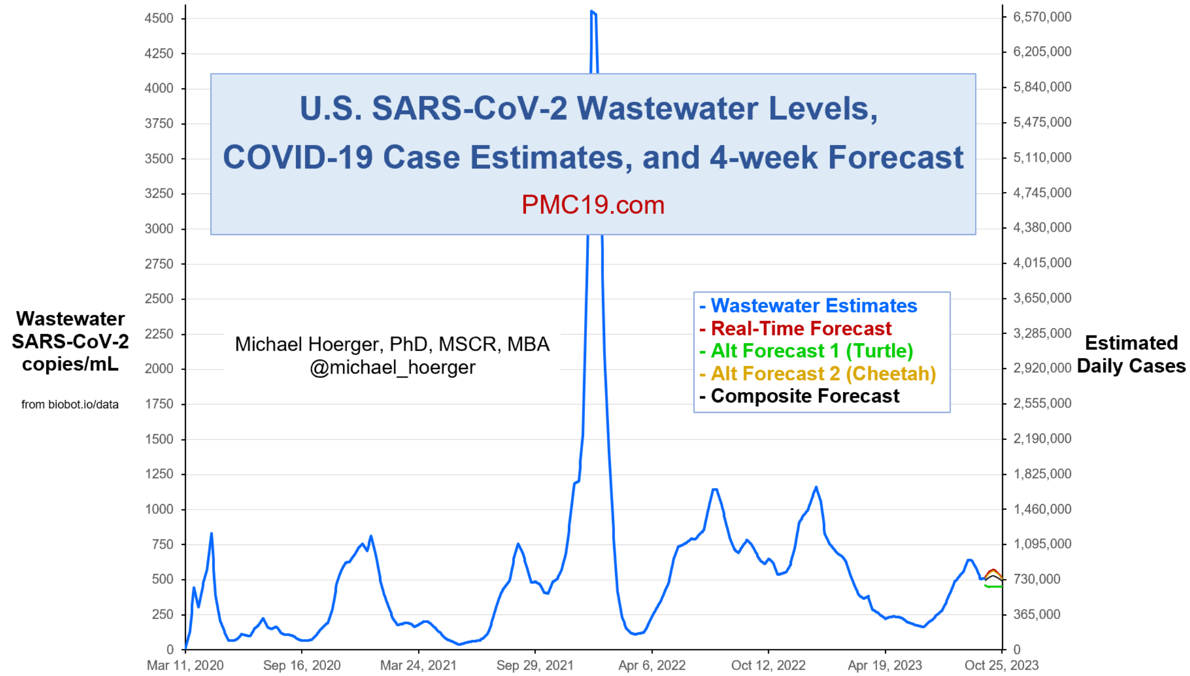
- 1.56% (1 in 64) are infectious
- >745,000 C0VID cases/day
- 37,000+ #LongCovid cases/day

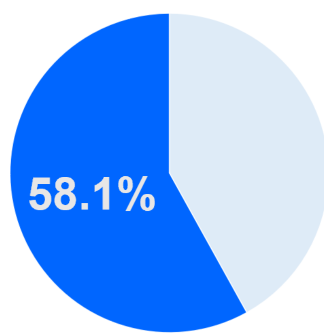
Note that Biobot is reporting Wednesday data (Sept 27) on a 5-day delay and has stopped intra-week data corrections.

What's the Current State of the Pandemic?

The 7th U.S. C0VID wave has been huge, slightly smaller than Delta, and is now at a steady but high rate of transmission, expected to sustain until about mid-November, when the 8th wave will likely begin.
Depending on estimates of repeat infections, likely half to 2/3 of Americans have been infected this year, with 220 million infections in 2023. This translates to 11 million or more Long COVID cases, when focusing on those most likely to be very serious.

We're seeing 5 million infections/week nationally, much higher than people realize, so now is a good time for advocacy.





There is more COVID-19 transmission today than during 58.1% of the pandemic.

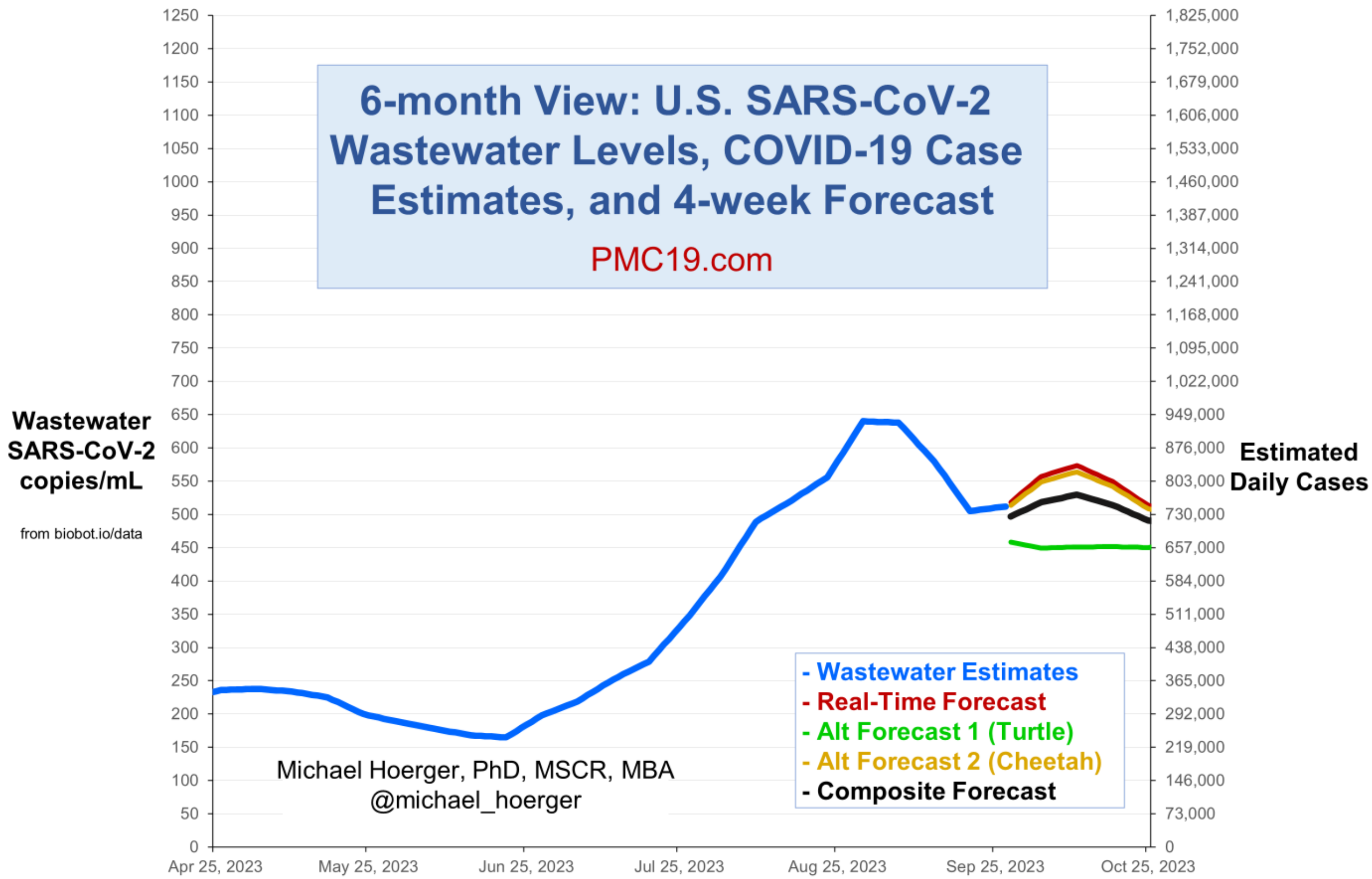
CURRENT ESTIMATES FOR September 27, 2023
Wastewater Levels (copies/mL)
512
New Daily Cases
745,000
% of Population Infectious
1.56% (1 in 64 people)
New Daily Long COVID Cases
37,000 to 149,000

WEEKLY ESTIMATES FOR September 27, 2023
New Weekly Cases
5,200,000
New Weekly Long COVID Cases
261,000 to 1,043,000

2023 CUMULATIVE ESTIMATES AS OF September 27, 2023
Total 2023 Cases To Date
220,500,000
Total 2023 Long COVID Cases To Date
11,025,000 to 44,100,000

Forecast for the Next Month

According to the composite model, we should see about 650,000-800,000 infections per day for the next month. This is a very high trough between waves. In mid-November (not pictured), we expect the start of an uptick toward a larger winter wave. Each day in the remainder of 2023, at least 1.4% of Americans are anticipated to be actively infectious, though much higher as December proceeds. Let's walk through each model.



CURRENT ESTIMATES FOR September 27, 2023

Wastewater Levels (copies/mL)
512
New Daily Cases
745,000
% of Population Infectious
1.56% (1 in 64 people)
New Daily Long COVID Cases
37,000 to 149,000

4-WEEK FORECAST FOR October 25, 2023

Wastewater Levels (copies/mL)
491 (-4% lower)
New Daily Cases
714,000
% of Population Infectious
1.49% (1 in 67 people)
New Daily Long COVID Cases
36,000 to 143,000

Real-time model (red line). It assumes that real-time data reports are accurate. However, real-time data often get corrected. Biobot and the CDC are currently in a transitionary phase of modifying when and which sites report, so I take the real-time data with a grain of salt. If it is accurate, however, the model suggests a rebound in cases, peaking around Oct 11, before bottoming out around Nov 1, before cases begin rising again for the winter wave.

Alt model #1, turtle (green line). The turtle model ignores the most recent week's worth of data from Biobot, treating it as unreliable. Thus, it assumes that rather than a small increase this week, levels are actually continuing to decline and that corrections to real-time reports will later reflect that. It's essentially saying that the bump you're seeing in the real-time model is just noise. Cases will stay almost completely flat the next 6 weeks, with an official low point around Oct 25, before cases begin rising again for the winter wave.

Alt model #2, cheetah (orange line). The cheetah model aims to correct for biases in real-time data reports. If last week's real-time report overestimated levels by 10% upon correction, it assumes this week's real-time report suffers the same bias. Last week's real-time report was quite accurate, so the cheetah model just looks close to the real-time model, same mini-peak, same valley, same rise in November for the winter wave.

Composite model (black line). This is the average of the three models. It's what's used for deriving all of the statistics reported. It basically suggests that cases will be mostly level at a high rate the next 6 weeks with minor fluctuations up or down. The composite model's take-home points are 1) continued high cases the next 6 weeks, 2) minimal fluctuation on a day-to-day basis during that time, 3) a low for the remainder of the year around Oct 25, and 4) a winter wave starting to pick up in mid-November.

What’s the Risk in an Office or in a Classroom?

The office and classroom risks remain quite bad. In a group of 10 people (daycare, team meeting, etc.), there’s a 15% chance someone will have infectious COVID. In a group of 20-25 people (e.g., K-12 classroom, department meeting, busy hospital waiting room, etc.), there’s 30% chance someone would have infectious COVID. In a university classroom of 40-50 people, it should be assumed someone has infectious COVID. This is quite troubling for instructors or students who mix time with multiple groups of classmates each week.

Not all classrooms and meetings are the same. ***The CDC has recently approved an updated booster, available to anyone in the U.S. older than 6 months. It is becoming widely available for adults, and expect more availability for children in the next two weeks.***

Virtual meetings reduce risk close to zero. Outdoor meetings are often safer than indoors. Testing reduces risk, as do policies that encourage people to stay home when symptomatic. High-quality, well-fitting masks greatly reduce risk. Air quality monitoring and improved air cleaning reduce risk. Recent boosters reduce risk. It remains troubling that elected leaders and public health officials choose to model poor mitigation when ongoing risk is so high.

What's the COVID Risk in an Office or in a Classroom?

Number of People	Chances Anyone is Infectious	Number of People	Chances Anyone is Infectious
1	1.6%	25	32.5%
2	3.1%	30	37.6%
3	4.6%	35	42.3%
4	6.1%	40	46.7%
5	7.6%	50	54.4%
6	9.0%	75	69.2%
7	10.4%	100	79.2%
8	11.8%	150	90.5%
9	13.2%	200	95.7%
10	14.5%	300	99.1%
15	21.0%	400	99.8%
20	27.0%	500	>99.9%

General Technical Notes, Not Specific to the Current Week's Report

Status of Biobot wastewater reporting. The estimates and forecast described here use wastewater data reported by Biobot. Biobot is now updating their data less frequently, and reportedly the CDC is in the process of reauthorizing their wastewater contract to a different entity. As long as national wastewater data are being reported, the PMC reports will continue. It is administratively burdensome but analytically easy to deal with this changes, but there may be delays in the coming weeks.

Case estimates. Case estimates were used by evaluating various potential multipliers to go from wastewater levels to cases. To identify true cases, not merely just reported cases, I used the IHME's case estimates for January 1, 2021 through April 1, 2023 (https://covid19.healthdata.org/united-states-of-america?view=cumulative-deaths&tab=trend). I compared wastewater with their case estimates on the 1st of each month. The correlation was r=.94. The maximum possible correlation is 1.00, so that is freakishly high, higher than just about any of the 10,000 or so correlations I've ever run. I was hoping for a correlation of r=.70 or higher, which still would have been great. Basically, wastewater is a supreme indicator of case rates. Next, I examined multipliers. Are cases 10x the arbitrary wastewater metric? 10,000x? Something else? Take cases and divide by wastewater at each data point, then find a summary metric (mean, median, trimmed mean, etc.). The metric I found most defensible was to use a +/-10% trimmed mean (average that excludes extreme data points, where case estimates are more error-prone), where each unit of wastewater translated into 1455 cases. I would find multipliers of 1000 to 1700 (31% lower to 17% higher) also reasonable. Arguably, case rates are magnitudes (10-100 times) higher than many people expect, so these details have minimal practical significance for everyday decision making. There are also more sophisticated strategies, such as regression models, but I found those results to be counter-intuitive (e.g., positive intercept, where I would have expected zero or negative). One can set the intercept to zero, use various heteroscedasticity-related techniques, and correct for the lack of imperfect reliability, but most of that is over the heads of people using this model and would accomplish little more than the trimmed multiplier method. The multiplier method has also led to techniques (only posted on Twitter thus far) for making regional estimates using very simple multipliers. Elegant is good.

Percentage infectious. After estimating the current number of new infections, it is relatively straightforward to estimate the percentage of the U.S. population actively infectious with COVID-19, but there are several caveats worth noting. One, the U.S. population is assumed to be 334,565,848. This was the CDC-estimated U.S. population on the final day of the IHME case estimation model. The number of new daily cases divided by the population tells one the percentage of the population newly infected today, often small at around 0.3% or less. Two, consider the infectious window. The percentage of the population infectious depends on the percentage of new people infected but also the duration people stay infectious. The model assumes people stay infectious for 7 days. Low estimates are that people are infectious for an average of 5 days (this defies the preponderance of the evidence, in my view), and high estimates are more like 10 days (too high in my view, based on a preference for round numbers). Other compelling estimates are more like 8-8.5 days. This duration may change over time, based on new variants, new vaccines, vaccine utilization rates, and treatments. If assuming the infectiousness duration is 10% longer, multiply by 1.10. If assuming 20% shorter, multiply by 0.80. New cases divided by the population equals new daily infections. Note also, these are merely averages and do not reflect individual variation, as some get infected and are not contagious, whereas others get infected and remain infectious likely for months (extremely rare). New daily infections multiplied by the number of days infectious indicates the percentage of the population actively infectious.

Long COVID. Long COVID case estimation. The lower and upper bounds for Long COVID case estimates assume that 5-20% of people infected with SARS-CoV-2 will develop Long COVID as a result of that infection. Some published reports and analysts have suggested lower (1%) or higher (40%) values. A useful framework for thinking about these estimates is that the low value is more indicative of people experiencing serious, enduring, known harms, whereas the upper estimates are closer to the number experiencing disruptive symptoms for at least several months, perhaps with full or partial recovery. These estimates do not indicate unknown long-term harms. For example, if infections increase the risk of cancer or cardiovascular disease substantially and with increasing risk over 10-30 years, that is not captured well by these metrics. The metrics also do not encompass the 1.2 to 1.8 million Americans who have died of COVID-19. Future models may incorporate estimates of mortality. Finally, the estimates project the number who will ultimately experience Long COVID from a new infection, but that is several months down the line. The estimates reflect future implications. For simplicity of interpretation, they are not modeling the number of new Long COVID cases today that resulted from infections three months ago.

General forecasting model specification. The forecasting models are elegant, meaning simple and effective. In regression analyses using historical pandemic wastewater data, the model explains 96% of the variance in the following week's forecast. The model is simple. It includes the year (2020, 2021, 2022, or 2023). It includes the historical average for the current half month; imagine the year sliced into 26 pieces, and it incorporates data on the historical average for that half month (e.g., second half of September). The model also incorporates four lagged variables, the wastewater levels 1, 2, 3, and 4 weeks ago. Overall, you can think of the model as having two main processes. One incorporates what we know historically. The other incorporates what has been happening the past several weeks. The historical data are useful because transmission mostly, but not always, follows a particular monthly pattern. It is not seasonal in that there are not just three bad months a year, but there is month-to-month variation, and sometimes even useful differences between the first versus second half of the month. The use of recent wastewater estimates helps in several ways. It lets the model know if something about the current point in time differs dramatically from the historical data, and it quickly adapts the model to changes, such as if a wave is starting or ending,

Real-time model (red line). This model assumes that real-time data reports of wastewater levels are accurate. However, real-time data often get corrected. Some sites may be slow reporting, and if there is a bias built in, such as places with high transmission being late to report, that would be a problem. Often, the real-time reports are quite accurate, but occasionally they have been corrected substantially a week later. The general model places a lot of weight on the most recent data, so any errors here can lead the model to assume a wave is picking up that really is not (false alarm) or that things are improving better than expected (false hope).

Alt model #1, turtle (green line). The turtle model moves slow and steady. It completely ignores the most recent week's worth of data from Biobot, treating it as unreliable. It will ignore false fluctuations inferred from inaccurate real-time reporting. However, it will be slower to respond to real changes, such as the onset in a new wave or the decline in a wave that has peaked.

Alt model #2, cheetah (orange line). The cheetah model moves fast. It aims to correct for biases in real-time data reports. If last week's real-time report overestimated levels by 10% upon correction, it assumes this week's real-time report suffers the same bias. If last week’s real-time report underestimated true levels, it assumes the same for this week. If last week’s real-time report was accurate, it will look similar to the real-time model. This model is very good if there is a bias, such as if areas with high transmission experience delays in reporting. However, it can also be overreactive. If there was some error in a real-time report that was just “random” rather than biased in a particular correction, it will tend to overcorrect the next week’s model.

Composite Model (black line). This is the arithmetic average of the three models. It's what's used for deriving all of the statistics reported. When all of the individual models are very close to the average, that suggests high confidence. When the models make vastly different predictions, that suggests more uncertainty in the data, largely based on perceptions of the accuracy of real-time wastewater reporting.