High throughput wastewater SARS-CoV-2 detection enables forecasting of community infection dynamics in San Diego county

Smruthi Karthikeyan[1], Nancy Ronquillo[2], Pedro Belda-Ferre[1], Destiny Alvarado[2], Tara Javidi[2], Rob Knight[1,3,4,5]

[1] Department of Pediatrics, University of California, San Diego, La Jolla, California, USA

[2] Department. of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093, USA

[3] Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA

[4] Department of Computer Science & Engineering, University of California, San Diego, La Jolla, CA, USA

[5] Center for Microbiome Innovation, University of California, San Diego, La Jolla, CA, USA

Address correspondence to:

Rob Knight

Department of Pediatrics

University of California San Diego

9500 Gilman Drive, MC 0763

La Jolla, CA 92093, USA

robknight@ucsd.edu

+1 858-246-1184

Abstract

Large-scale wastewater surveillance has the ability to greatly augment the tracking of infection dynamics especially in communities where the prevalence rates far exceed the testing capacity. However, current methods for viral detection in wastewater are severely lacking in terms of scaling up for high throughput. In the present study, we employed an automated magnetic-bead based concentration approach for viral detection in sewage that can effectively be scaled up for processing 24 samples in a single 40-minute run. The method compared favorably to conventionally used methods for viral wastewater concentrations with a limit of detection of 8.809 viral gene copies/ml from input sample volumes as low as 10ml and can enable the processing of over 100 wastewater samples in a day. Using the high throughput pipeline, samples from the influent stream of the primary wastewater treatment plant of San Diego county (serving 2.3 million residents) were processed for a period of 13 weeks. Wastewater estimates of SARS-CoV-2 viral genome copies in raw untreated wastewater correlated strongly with clinically reported cases by the county, and enabled prediction of cases up to 3 weeks in advance. Taken together, the results show that the high-throughput surveillance could greatly ameliorate comprehensive community prevalence assessments by providing robust, rapid estimates.

Importance

Wastewater monitoring has a lot of potential for revealing COVID-19 outbreaks before they happen because the virus is found in the wastewater before people have clinical symptoms. However, application of wastewater based surveillance has been limited by long processing times specifically at the concentration step. Here we introduce a much faster method of processing the samples, and show that its robustness by demonstrating direct comparisons with existing methods and showing that we can predict cases in San Diego by a week with

excellent accuracy, and three weeks with fair accuracy, using city sewage. The automated viral concentration method will greatly alleviate the major bottleneck in wastewater processing by reducing the turnaround time during epidemics.

Main

Wastewater-based epidemiology (WBE) can facilitate detailed mapping of the extent and spread of SARS-CoV-2 in a community and has seen a rapid rise in recent months owing to its cost effectiveness as well as its ability to foreshadow trends ahead of diagnostic testing [1-4] . With over 46 million cases reported globally, 9.3 million of which are from the United States, there is an imminent need for rapid, community-level surveillance in order to identify potential outbreak clusters ahead of diagnostic data. Previous studies have reported high levels of correlation between viral concentration in sewage to clinically reported cases in a community with trends appearing 2-8 days ahead in wastewater [2, 5]. A major bottleneck in large scale wastewater surveillance is the lack of robust, high throughput viral concentration methodology. Conventional techniques for viral concentration from wastewater typically employ laborious or time-consuming processes, namely polyethylene glycol (PEG) based precipitation, direct filtration or ultra-filtration methods that severely limit throughput [6]. In the present study, we employed a magnetic bead-based viral concentration method which was incorporated on the KingFisher Flex liquid-handling robot platform robots (Thermo Fisher Scientific, USA), using a 24-plex head to process 24 samples at once in a 40 min run. RNA is then extracted on the same Kingfisher system for rapid sample processing. Reverse transcription-quantitative polymerase chain reaction (RT-qPCR) targeting N1, N2, E-gene targets was used for detection and quantification of SARS-CoV-2 RNA. Using the above pipeline, 96 raw sewage samples were processed in a period of 5 hours (concentration

to RT-qPCR detection/quantification) effectively reducing the processing time by at least 20-fold.

The efficiency of high throughput concentration method was compared to the two other commonly implemented concentration methods (electronegative membrane filtration and Polyethylene glycol, PEG-based) using nine-fold serial dilutions of heat-inactivated SARS-CoV-2 viral particles spiked into 10 mL volumes of raw sewage which was previously verified to be from a location with no SARS-CoV-2 prevalence (Suppl. Methods). The high-throughput protocol compared favorably to the conventionally used protocols demonstrating its potential implications for large scale sample processing (>100 samples/day) (Suppl. Fig. S1.A).

24-hour flow weighted composites were collected each day for 3 months between July 20th-October 21st, 2020 from the influent stream of Point Loma wastewater treatment plant. The plant processes over 175 MGD (million gallons per day) of raw sewage and is the primary treatment center for the greater San Diego area serving over 2.3 million residents. SARS-CoV-2 viral RNA was detected in all of the samples processed at an average concentration of 2,010,104.45 gene copies/L in the influent and 420,000–580,000 gene copies/L at ocean outfall (effluent). Over the course of the study, the clinically reported cases in the county increased by 29,375 (Fig. 1). Peaks in the wastewater data were frequently followed by peaks in the clinically confirmed cases at a later date. This suggests a correlation between wastewater and the number of new cases with the caveat of a time delay, where the wastewater data predicts future trends in the new number of cases. Although informative, this time-lagged correlation alone is not enough for robust predictions. This served as the main motivation to build a predictive model for forecasting the number of new cases per day in

San Diego County. Additionally, the day of the week was tracked for each data point in a third time series to capture any weekly trends for a total of three time series with 88 data points each. We used a data-driven approach to train a prediction model that utilizes wastewater data and temporal correlations (embedded in the day of the week) in order to forecast the number of new positive cases in San Diego County. The multivariate Autoregressive Moving Average (ARMA) model [7] was applied to build a prediction model for the number of new positive cases. The (predicted) number of new cases consist of lagged past values from all three series (number of new cases, wastewater data, day of the week) and each term can be thought of as the influence of that lagged time series on the number of new cases. 65 data points were used for developing the predictive model and the remaining 24 were reserved for validating our forecast estimates. Fig. 2B shows the results of the model (prediction and forecast) compared to the observed data. The Pearson correlation coefficients between the observed data and the predicted model were r = 0.84, 0.79, 0.69, 0.47 for current, 1, 2 and 3 week advance forecast values respectively (Suppl. Table S1). Our data-driven approach obtains forecasts successfully capturing general trends on the number of new cases (shown here up to 3 weeks in advance), ultimately reinforcing that wastewater analysis can be expository of previously undetected SARS-CoV-2 infections in the population which could provide useful insight for county officials for the purpose of early public health interventions. Our study demonstrates that high-throughput wastewater-based surveillance can be successfully leveraged to enable creation of a rapid, large-scale early alert system for counties/districts and could be particularly useful in community surveillance in the more vulnerable populations.

ACKNOWLEDGEMENTS

The authors would like to thank the plant supervisor and the lab team at the Point Loma Wastewater Treatment plant in San Diego for providing us with samples and are grateful for their support in this effort. We thank Dr. Jack Gilbert and the Microbiome Sample Processing Core at UC San Diego for access to qPCR equipment. This work was supported by The University of California San Diego Return to Learn program (UCSD-RTL).

FIGURE LEGENDS

Figure 1: Tracking infection dynamics in SD county. A. Map showing the San Diego sewer mains (depicted in purple) that feed into the influent stream at the primary WWTP at Point Loma. Overlaid are the cumulative cases recorded from the different zip codes in the county during the course of the study. The circles are proportional to the diagnostic cases reported from each zone and the color gradient shows the cases per 100,000 residents. **B**. Daily new cases reported by the county of San Diego. **C**. SARS-CoV-2 viral gene copies detected per L of raw sewage. All viral concentration estimates were derived from the processing of 2 biological replicates.

Figure 2: A. Daily caseload and wastewater viral concentration data shown for a period of 13 weeks, where a spline smoothing is applied to each time series to demonstrate general trends. **B.** Predictive model showing the predicted data (yellow) compared to the observed caseload (blue) and the 4-week forecast (red). Data collected from 07/07/2020 to 09/28/2020 were used as the training dataset to predict the caseload for the following weeks (up to 10/25/2020). Data (wastewater + county testing data) gathered from 09/29/2020-10/21/2020

were used for model validation. MATLAB Systems Identification toolbox was used to estimate the model order and parameters and calculate the forecasted values.

REFERENCES

1. Larsen, D.A. and K.R. Wigginton, *Tracking COVID-19 with wastewater.* Nature Biotechnology, 2020. **38**(10): p. 1151-1153.

2. Peccia, J., et al., *Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics.* Nature Biotechnology, 2020. **38**(10): p. 1164-1167.

3. Wu, F., et al., *SARS-CoV-2 titers in wastewater foreshadow dynamics and clinical presentation of new COVID-19 cases.* medRxiv : the preprint server for health sciences, 2020: p. 2020.06.15.20117747.

4. Ahmed, W., et al., *First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community.* Science of The Total Environment, 2020. **728**: p. 138764.

5. Medema, G., et al., *Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The Netherlands.* Environmental Science & Technology Letters, 2020. **7**(7): p. 511-516.

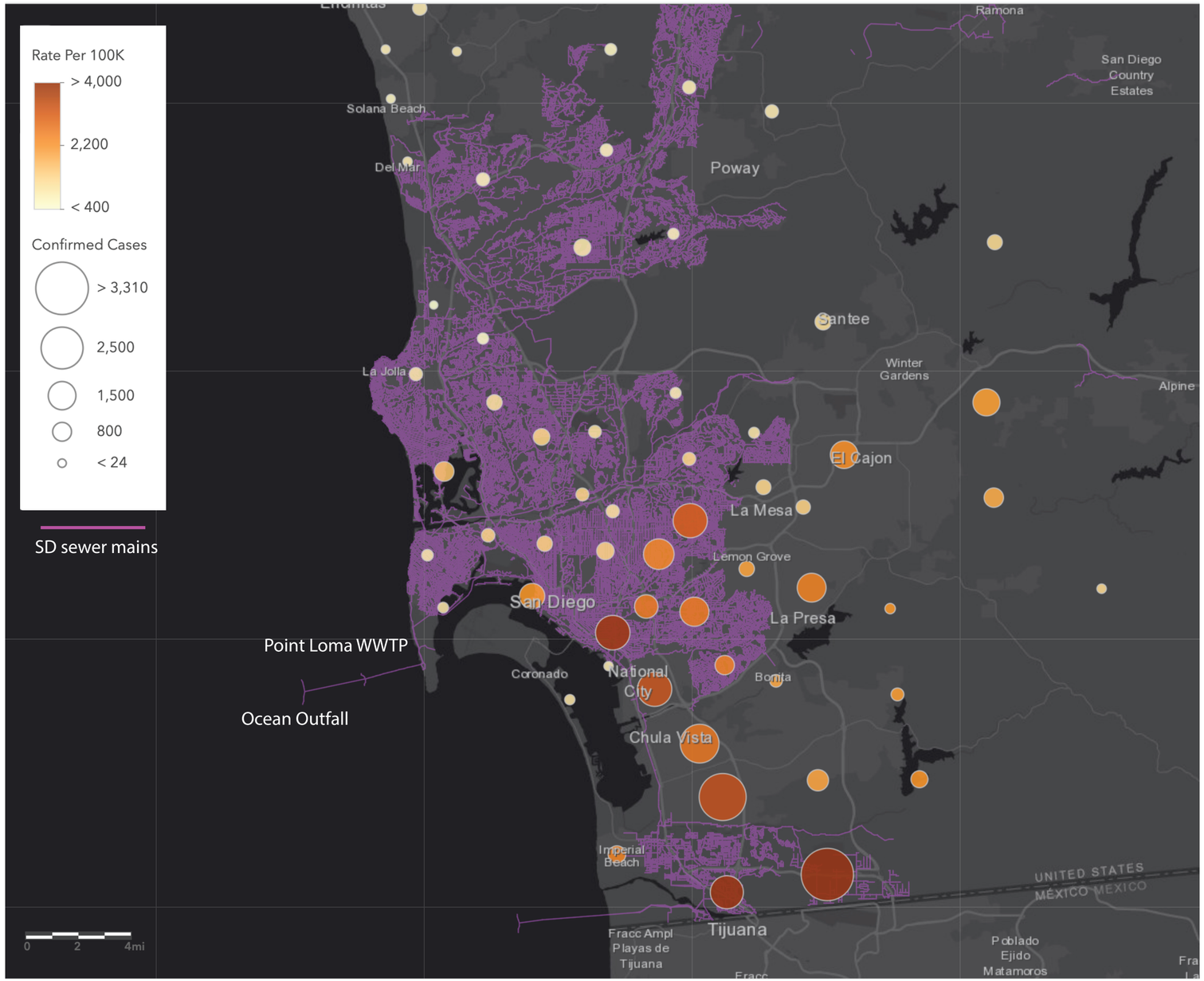
6. Ahmed, W., et al., *Comparison of virus concentration methods for the RT-qPCR-based recovery of murine hepatitis virus, a surrogate for SARS-CoV-2 from untreated wastewater.* Science of The Total Environment, 2020. **739**: p. 139960.

7. Brockwell, P.J., R.A. Davis, and S.E. Fienberg, *Time series: theory and methods: theory and methods.* 1991: Springer Science & Business Media.

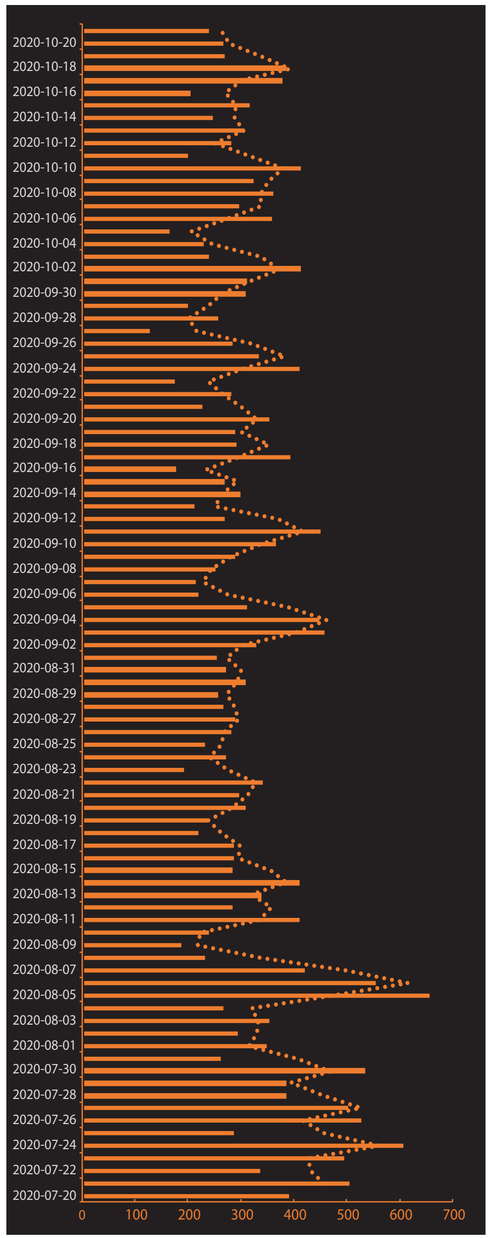
8. Barclay, R.A., et al., *Nanotrap particles improve detection of SARS-CoV-2 for pooled sample methods, extraction-free saliva methods, and extraction-free transport medium methods.* bioRxiv, 2020: p. 2020.06.25.172510.

9. Corman, V.M., et al., *Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR.* Eurosurveillance, 2020. **25**(3): p. 2000045.

A

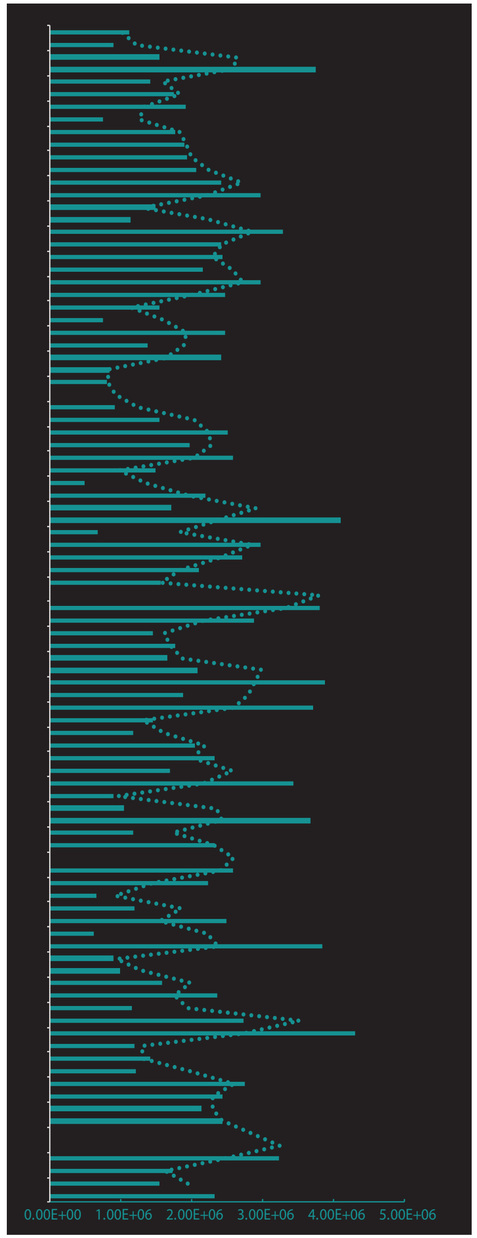


B

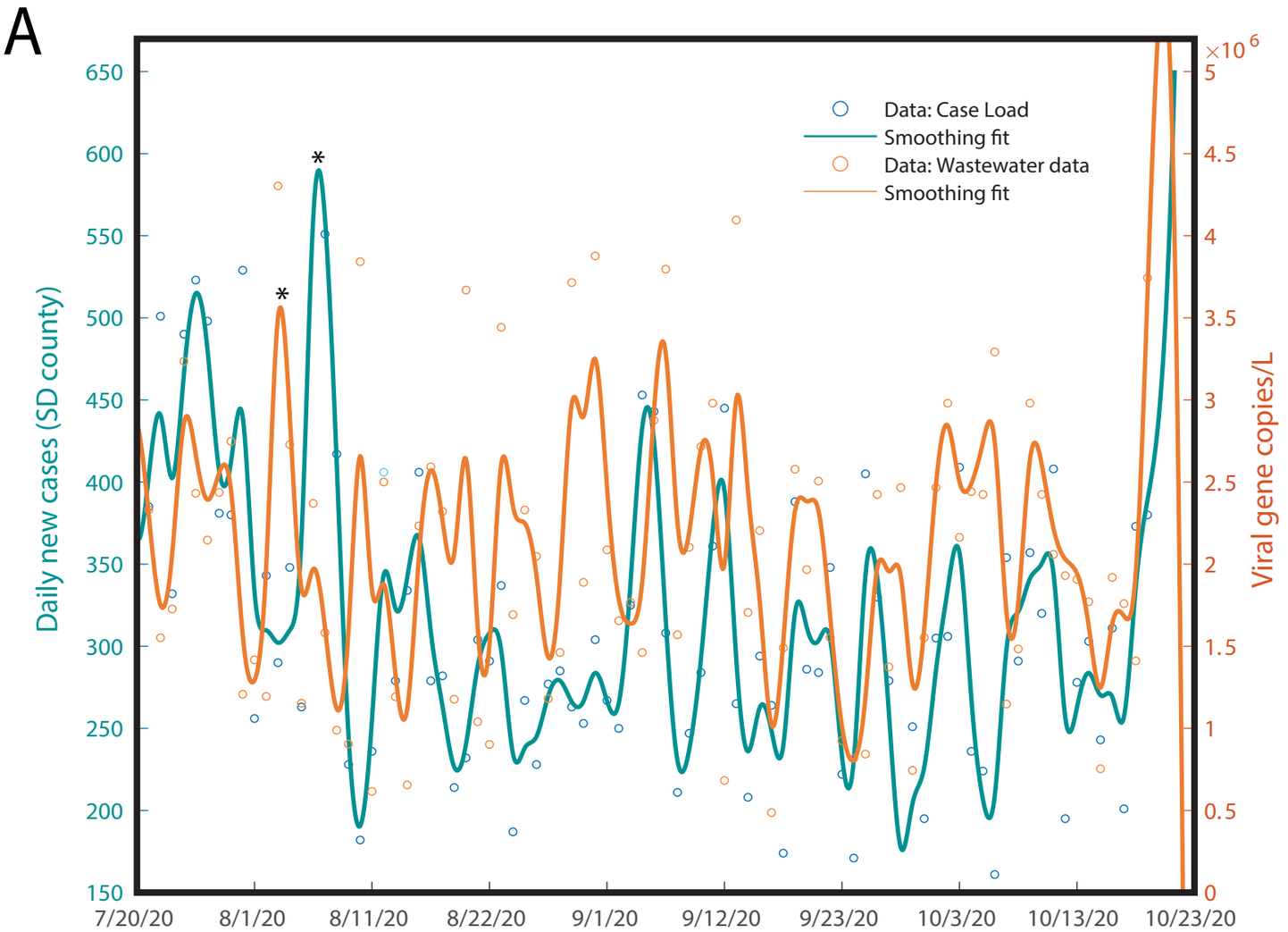


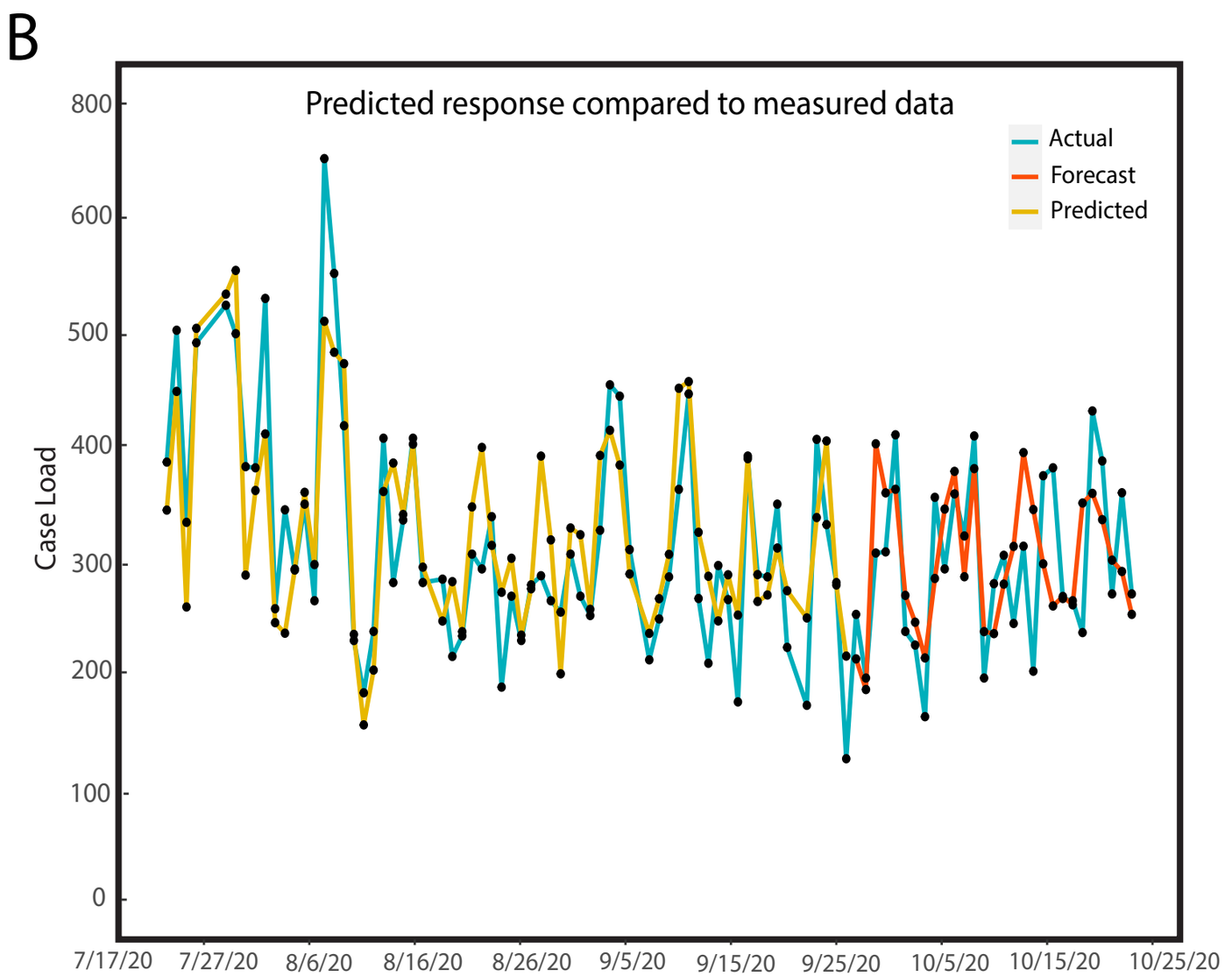
Reported cases

C



Viral gene copies/L


A
Data: Case Load
Smoothing fit
Data: Wastewater data
Smoothing fit
Daily new cases (SD county)
Viral gene copies/L
×10^6
650
600
550
500
450
400
350
300
250
200
150
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
7/20/20
8/1/20
8/11/20
8/22/20
9/1/20
9/12/20
9/23/20
10/3/20
10/13/20
10/23/20


B
Predicted response compared to measured data
Actual
Forecast
Predicted
Case Load
800
600
500
400
300
200
100
0
7/17/20
7/27/20
8/6/20
8/16/20
8/26/20
9/5/20
9/15/20
9/25/20
10/5/20
10/15/20
10/25/20