frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 29 September 2023
DOI 10.3389/fpubh.2023.1185720

Check for updates

OPEN ACCESS

EDITED BY
Tao-Hsin Tung,
Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University, China

REVIEWED BY
Galal Metwally,
Zagazig University, Egypt
Shalome Hanisha Anand Tatapudi,
Merck, United States

*CORRESPONDENCE
Imran Mujawar
✉ imran.mujawar@thermofisher.com

RECEIVED 30 March 2023
ACCEPTED 07 September 2023
PUBLISHED 29 September 2023

CITATION
Sharmin M, Manivannan M, Woo D, Sorel O, Auclair JR, Gandhi M and Mujawar I (2023) Cross-sectional Ct distributions from qPCR tests can provide an early warning signal for the spread of COVID-19 in communities. *Front. Public Health* 11:1185720. doi: 10.3389/fpubh.2023.1185720

COPYRIGHT
© 2023 Sharmin, Manivannan, Woo, Sorel, Auclair, Gandhi and Mujawar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Cross-sectional Ct distributions from qPCR tests can provide an early warning signal for the spread of COVID-19 in communities

Mahfuza Sharmin[1], Mani Manivannan[1], David Woo[1], Océane Sorel[1], Jared R. Auclair[2], Manoj Gandhi[1] and Imran Mujawar[1]*

[1]Thermo Fisher Scientific, South San Francisco, CA, United States, [2]Department of Chemistry and Chemical Biology, Northeastern University, Burlington, MA, United States

Background: SARS-CoV-2 PCR testing data has been widely used for COVID-19 surveillance. Existing COVID-19 forecasting models mainly rely on case counts obtained from qPCR results, even though the binary PCR results provide a limited picture of the pandemic trajectory. Most forecasting models have failed to accurately predict the COVID-19 waves before they occur. Recently a model utilizing cross-sectional population cycle threshold (Ct—the number of cycles required for the fluorescent signal to cross the background threshold) values obtained from PCR tests (Ct-based model) was developed to overcome the limitations of using only binary PCR results. In this study, we aimed to improve on COVID-19 forecasting models using features derived from the Ct-based model, to detect epidemic waves earlier than case-based trajectories.

Methods: PCR data was collected weekly at Northeastern University (NU) between August 2020 and January 2022. Campus and county epidemic trajectories were generated from case counts. A novel forecasting approach was developed by enhancing a recent deep learning model with Ct-based features and applied in Suffolk County and NU campus. For this, cross-sectional Ct values from PCR data were used to generate Ct-based epidemic trajectories, including effective reproductive rate (Rt) and incidence. The improvement in forecasting performance was compared using absolute errors and residual squared errors with respect to actual observed cases at the 7-day and 14-day forecasting horizons. The model was also tested prospectively over the period January 2022 to April 2022.

Results: Rt curves estimated from the Ct-based model indicated epidemic waves 12 to 14 days earlier than Rt curves from NU campus and Suffolk County cases, with a correlation of 0.57. Enhancing the forecasting models with Ct-based information significantly decreased absolute error (decrease of 49.4 and 221.5 for the 7 and 14-day forecasting horizons) and residual squared error (40.6 and 217.1 for the 7 and 14-day forecasting horizons) compared to the original model without Ct features.

Conclusion: Ct-based epidemic trajectories can herald an earlier signal for impending epidemic waves in the community and forecast transmission peaks. Moreover, COVID-19 forecasting models can be enhanced using these Ct features to improve their forecasting accuracy. In this study, we make the case that public health agencies should publish Ct values along with the binary positive/negative PCR results. Early and accurate forecasting of epidemic waves can inform public health policies and countermeasures which can mitigate spread.

KEYWORDS

CT model, forecasting model, epidemic trajectories, COVID-19, early detection

Introduction

The COVID-19 pandemic created an unprecedented situation where large-scale PCR testing has been performed to identify cases who were suspected to be infected with the virus. The results from such widespread testing have also helped better understand the epidemiological aspects of this disease and adapt public health interventions to reduce the spread of the virus. Current approaches for outbreak surveillance rely mainly on SARS-CoV-2 PCR testing results to estimate case counts and test positivity rates (1).While real-time or near real-time monitoring at the local and national levels can help detect changes in patterns of the pandemic and inform public health decisions, this approach, however, does not allow public health agencies sufficient lead time to prepare for upcoming surges. Public health agencies need to gain access to data predicting the evolution of the pandemic as early as possible in order to plan for surges and adapt healthcare capacities accordingly (2).

Concerted efforts from academia, industry, and government have been made to develop predictions at different forecasting horizons. Notably, these efforts have led to development of the COVID-19 Forecast Hub, which provides up-to-date data of COVID-19 cases, deaths, and hospitalizations in the US, in coordination with the Centers for Disease Control and Prevention (CDC) (3). Most of these forecasting initiatives use various compartmental models (4, 5) a modeling technique where populations are clustered into various compartments such as susceptible, infectious, and recovered, with flow patterns being simulated between them. However, almost all these compartmental models rely mainly on the binary (i.e., positive/negative) test results and provide a limited picture of the pandemic trajectory (6). Case-series trajectories, based on incidence rates alone, can exhibit biases due to variations in testing demand and supply, population sampling and reporting delay (7, 8). Most forecasting models currently available on the Forecast Hub have struggled to consistently predict COVID-19 waves before they occur (3). The ensemble model which combines the top performing models, developed by experts in the field, was unable to detect any of the variant waves in the US. This model has shown "low reliability" in predicting COVID-19 waves and is no longer used (9). These findings highlight the tremendous challenge of predicting COVID-19 surges with forecasting models using the binary test outcomes.

Recently, Hay et al. (6) have shown that using population cross-sectional viral load distributions, measured by PCR cycle thresholds (Ct—the number of cycles required for the fluorescent signal to cross the background threshold), can overcome common biases in epidemic trajectory estimates that are seen when using case series data alone. Additional studies have corroborated this observation and showed that the models with Ct information correlated to case counts (10) and can be used to better predict epidemic growth rates (11, 12). In this study, we further corroborate that Ct-based model-derived epidemic trajectories, such as incidence and reproduction number (Rt), can detect epidemic waves earlier than case-based trajectories, and that COVID-19 forecasting models can be enhanced using these Ct features to improve their forecasting accuracy of disease incidence.

Materials and methods

qPCR data preparation

During the first couple years of the pandemic, the Life Science Testing Center (LTSC) conducted routine PCR screening of students, faculty and staff at Northeastern University using the TaqPath™ COVID-19 Combo Kit. In this study, PCR data were retrospectively analyzed from samples collected between August 2020 and January 2022, and prospectively up to April 2022. Deidentified.eds files from the PCR runs were analyzed to obtain the Ct values from all screened samples during the defined period. According to the TaqPath™ COVID-19 Combo Kit Instructions for Use, a sample was considered positive for SARS-CoV-2 if two or more of the three gene targets (N, S and ORF1ab) were detected ($Ct \leq 37$). Samples where only one gene target was detected ($Ct \leq 37$) were considered as inconclusive (13). Ct denotes the cycle threshold - the number of cycles required for the fluorescent signal to cross the background threshold). The data flow and processing steps are depicted in Figure 1A.

Incidence for NU campus and Suffolk county case-series

Daily campus COVID case-count was determined through counting the number of unique positive samples (adjusting for re-testing) per day. Since the data was de-identified, these campus case-counts were not adjusted for different samples taken from the same individual or for re-infection. Incidence per 100,000 population was calculated by dividing cases per day or week by the total campus population as of 2019 (14). COVID-19 cases in Suffolk county, where the university campus is located, were obtained from a publicly available database (15). Incidence per 100,000 population was calculated by dividing the county positive cases by the county's total population based on census numbers as of April 2020 (16).

Ct-based model to determine inferred incidence estimates

Incidence curves were created using the Ct values from the PCR screening tests performed on campus according to the methodology developed by Hay et al. (6). Only the N-gene Ct values from the TaqPath COVID-19 Combo Kit were selected to be included in the Ct-based model. Based on their run dates, the samples were considered as weekly occurrences. We used a Gaussian process model from the *virosolver* R-package (17) the weekly Ct-distributions to infer the incidence rates and followed the recommended parameters for the Gaussian process model. In the initial phase of the study, we retrospectively analyzed the epidemic trajectories from the Ct-based model for early detection of incidence trends of the campus and county from the samples collected between August 2020 to January 2022. In the next phase, we prospectively monitored the ability of the Ct-based model for early indicators of uptick in incidence trends in NU and Suffolk county on a weekly basis from January 2022 until April 2022.

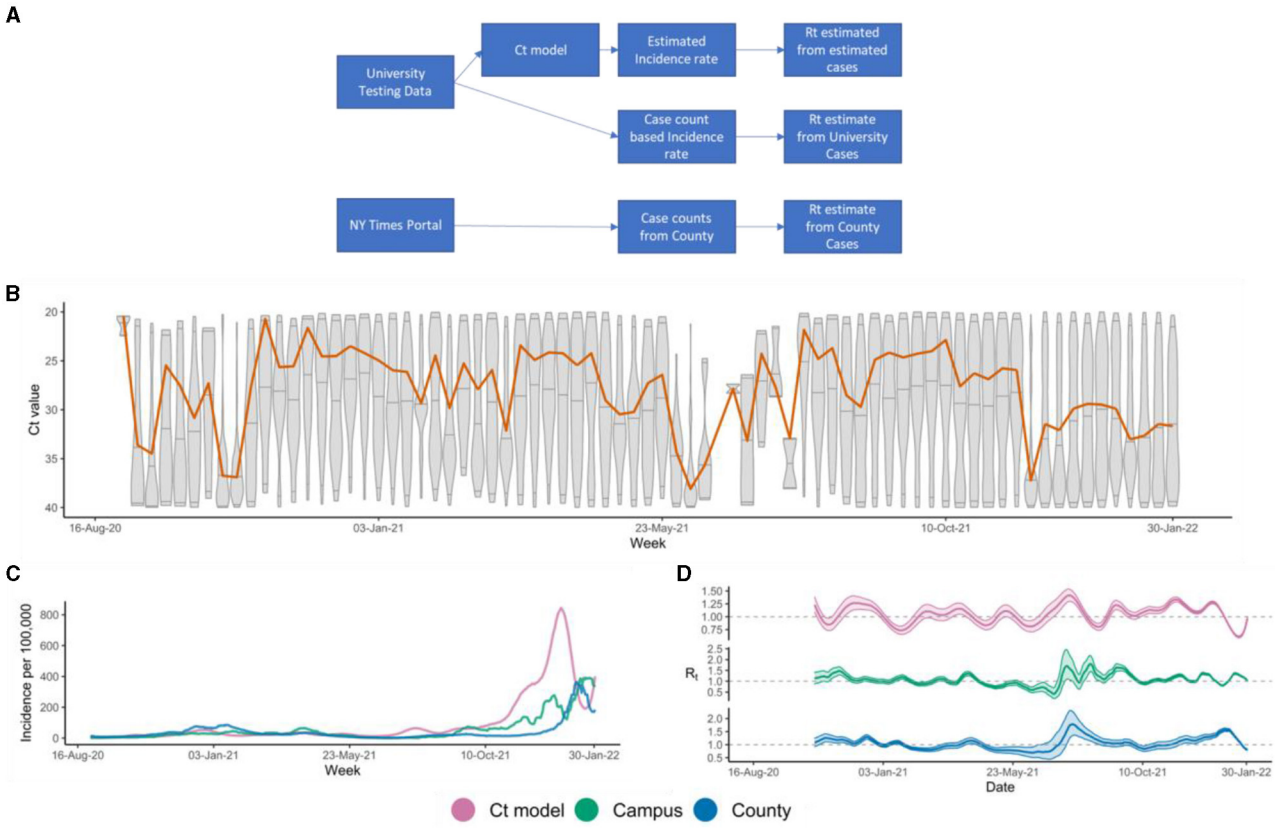

FIGURE 1
(A) Study design showing data flow and processing; **(B)** Weekly Ct value distribution represented by each boxplot. The orange line represents the rolling median of the weekly Ct distribution; **(C)** Incidence rates predicted from Ct-based model (magenta), campus case counts (green), and Suffolk County case counts (blue); **(D)** Rt estimates derived from Ct model inferred incidences (magenta), or campus (green) and Suffolk County case counts (blue).

Estimating time varying reproduction number (Rt)

The three incidence trajectories: (a) Ct-based model derived estimates, (b) campus positive samples, and (c) county case-series, were smoothed by a 14-day moving average. The effective reproduction number (Rt) was derived from these three incidence trajectories, using the R-package *EpiEstim* (18) (Figure 1A). The *EpiEstim* package estimates the effective reproduction number of an epidemic, given the incidence time series and the serial interval distribution. We used a parametric serial intervention method with a mean serial interval of 6.14 and standard deviation of 3.96 following recommended parameters, and a 14-day sliding window.

Lagged correlation analysis between Ct-based model Rt curves, campus case-based Rt curves, and county case-based Rt curves

Due to the time dependent nature of the Rt curves, pairwise lagged correlation was assessed to estimate the Spearman Correlation Coefficients and the lag or lead times between the 3 Rt curves: Ct-based model, campus case-based, and county case-based. Transmission peaks were identified when the Rt estimates were >1. Transmission peaks in the Ct-based model Rt curves were mapped to subsequent transmission peaks detected in the campus and county Rt curves. The number of days between the Ct-based model peak and the corresponding presumptive subsequent peak in the campus and county curves was calculated. The median time to campus and county peak from a Ct-based model Rt peak was then estimated.

Neural relational autoregressive model (β-AR) and Ct-enhanced β-AR (Ct-β-AR)

β-AR is a case count forecasting model that was published by Facebook/Meta (19), one the best performing COVID-19 forecasting models at the time. The β-AR model requires two types of inputs: (a) time dependent data including symptom survey, mobility, doctor visits, total PCR tests per state, vaccination and masking policy, and weather, which are modeled by the recurrent neural network (RNN) part of the overall model and (b) confirmed COVID-19 cases variable is the only time independent input, which feeds into the autoregressive (AR) part of the model. We enhanced

the β-AR model using features derived directly from PCR Ct values (measures of central tendency and dispersion) and from the Ct-based epidemiological model above [Hay et al] (incidence and Rt). Each feature was measured weekly in accordance with the β-AR model measurement frequency.

Determining accuracy of Ct- β-AR to real incidence rates using RMSE and AE

We assessed performance of Ct- β-AR and β-AR from October 2020 to December 2021, by comparing the predictions at the 7- and 14-day forecast horizons to real incidence rates. To measure the performance, we calculated the root mean square error (RMSE) and mean absolute error (AE) by the following formula:

$$RMSE = \sqrt{\frac{(x_{predicted} - x_{actual})^2}{study\ period}}$$

$$AE = |x_{predicted} - x_{actual}|$$

where x_{actual} was the actual case count on a particular day and $x_{predicted}$ was the predicted case count for the same day. The difference between Ct-β-AR and β-AR models for their RMSE and AE values were statistically assessed using the paired Wilcoxon signed rank sum test.

Results

Data cohort with PCR results description

A total of 1,608,562 valid samples, collected between August 2020 and April 2022, were included in the analysis. Among those samples, 47,194 (2.93%) samples tested positive for SARS-CoV-2, and 17,677 (1.1%) samples were identified as inconclusive. Over the study period, the median N-gene detectable Ct value was around 28 with a minimum of 5 and a maximum of 40 and a skewness of −5.69. During the *wild-type* (initial Wuhan) (Aug 2020–Jan 2021) $N = 154$; *Alpha* (Jan 2021–Jun 2021) $N = 7{,}216$; *Delta* (Jul 2021–Dec 2021) $N = 11{,}790$; and *BA.1 Omicron* (Jan 2022–Mar 2022) $N = 21{,}715$ surges, samples tested positive with an approximate median Ct=28 across waves and varying skewness of −9.07, −7.38, −5.13, −3.22, respectively. The Ct distributions over time are shown in Figure 1B.

Ct-based and Case-count derived epidemic metrics were highly comparable

The predicted incidence rates, derived from either Ct-based model on campus data, Campus case count data or Suffolk County cases counts (Figure 1C) were very similar in terms of epidemic trends throughout the duration (Aug 2020 to Jan 2022). Upsurges in incidence were observed around the same periods for the 3 incidence trajectories, including when the *wild-type* variant was dominant as well as during the *Alpha*, *Delta*, and *Omicron* variant waves. The trajectories also demonstrated similar intervening troughs between the observed waves. The incidence waves were also relatively similar in size, apart from the first Omicron wave, where the Ct-based incidence wave was much larger than the other two trajectories. The three incidence rate patterns were also consistent with publicly available data for Massachusetts state (20) showing that the predicted values were in agreement with the local epidemic incidence rates. The effective reproduction number (Rt) estimates for the 3 corresponding incidence trajectories were also very similar throughout the period analyzed (Figure 1D). We observed that major Rt peaks noted at the beginning of a variant wave generally matched across the 3 trajectories, especially between the Ct-based and county Rt trajectories. The synched Spearman Correlation Coefficient between the three Rt curves (Ct-based vs. Campus case count vs. County case count) was 0.57 demonstrating that the Ct-based epidemic trajectory was highly comparable to the case-count derived epidemic trajectories (Campus and County), considering the duration analyzed (18 months) and a 3-way synched correlation.

Ct-based model predictions preceded case counts-derived estimates

Once we established that the Ct-based model was able to accurately mirror the epidemic trends, we proceeded to determine if the Ct-based model was able to portend impending surges over case-count derived models, and if so, how early could our model predict. Lagged correlation analysis of the Rt estimated curves showed that the Ct-based model had a 12-day lead time as compared to the Campus case count estimates and a 14-day lead time as compared to County case count estimates, with lagged Spearman correlation coefficients of 0.38 and 0.59 respectively. Analysis of the Rt trend peaks showed that Ct-based model peaks preceded the corresponding case counts-derived Rt peaks by a median lead time of 11 days for, both, the county (Figures 2A, B) and campus estimates (Figures 2C, D). Interestingly, prospective monitoring of the Rt trends showed that the Ct-based model derived Rt trajectory was able to detect an upcoming peak around week 4 of March 2022 which was not yet uncovered in, both, the case counts based Rt trajectories (Figure 3A). Upon further monitoring, it was discovered that the Ct-based model predicted peak was, indeed, followed by a peak in actual Campus and County case-counts after 2 weeks (week 2 of April 2022) which was due to the *Omicron* surge (BA.2) (Figure 3B).

Ct-β-AR model significantly improved forecasting estimates

Over the time duration analyzed (October 2020 to December 2021), the Ct-β-AR model provided statistically significant lower AE at both forecasting horizons; 7 days (p-value < 0.001) (Figures 4A, B) and 14 days (p-value < 0.001) (Figures 4C, D). In addition, the mean AE values comparing Ct- β-AR vs. β-AR models were 273.4 vs. 322.8 and 679.6 vs. 901.1 for the 7 and 14-day forecasting horizons, respectively. Similarly, the RMSE was also significantly lower (p-value < 0.001) comparing Ct- β-AR

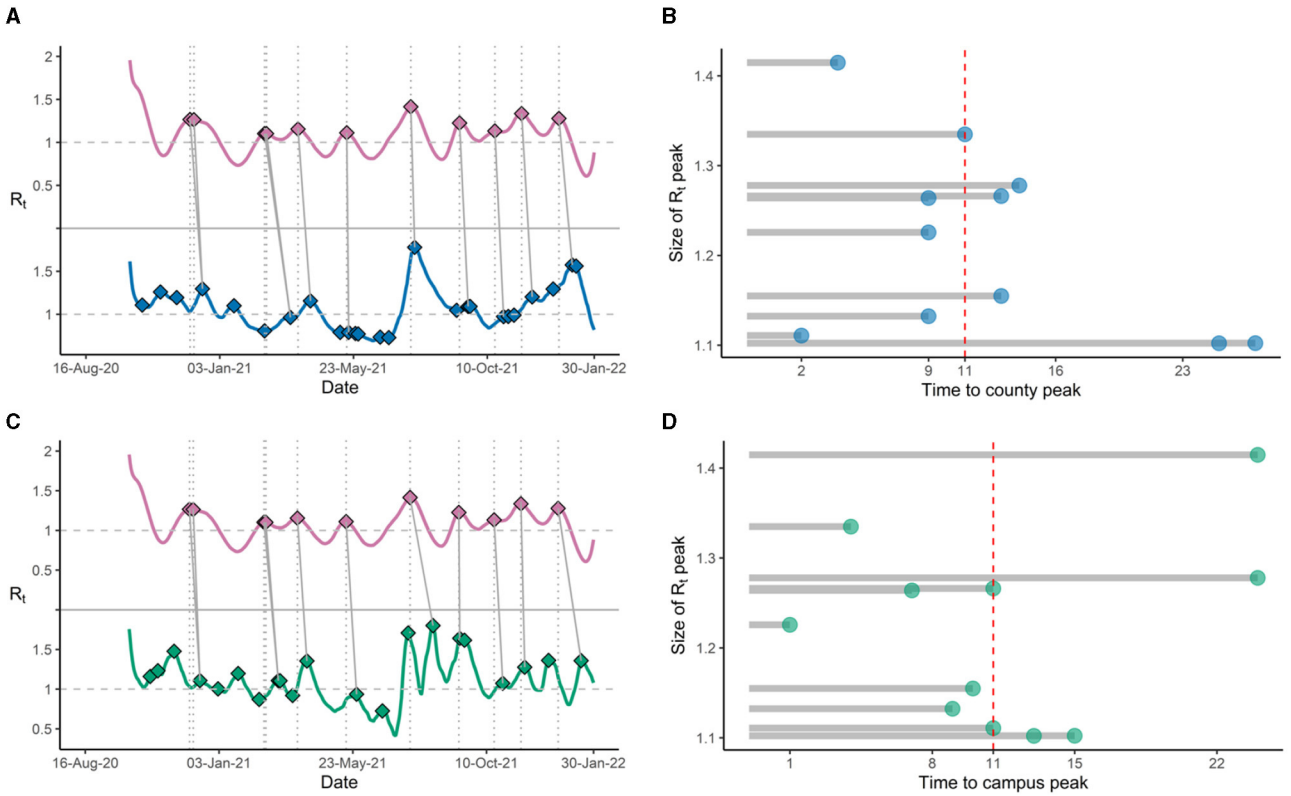

FIGURE 2
(A) R_t trends showing peaks for Ct-based model derived estimates (magenta), with subsequent peaks in R_t trends derived from county cases (blue); **(B)** Lead times (in days) between Ct model-derived R_t peaks and county R_t peaks, with median lead time (red dashed line). **(C)** R_t trends showing peaks for Ct model-derived estimates (magenta), with subsequent peaks in R_t trends derived from positive tests on campus (green); **(D)** Lead times (in days) between Ct-based model derived R_t peaks and campus R_t peaks, with median lead time (red dashed line).

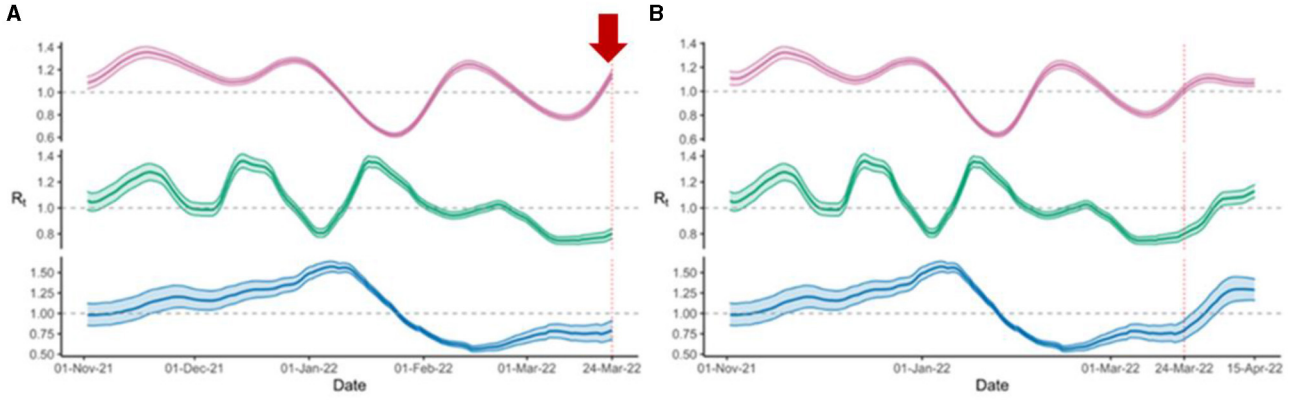

FIGURE 3
Prospective monitoring of the Rt trajectories showed **(A)** R_t trends with an early uptick (red arrow) in Ct model-derived estimates (magenta) around 24 March 2022, while the other two trajectories, including R_t trends derived from county cases (blue), and positive tests on campus (green) are flat; **(B)** Later upticks noted in R_t trends derived from county cases (blue), and positive tests on campus around 15 April 2022.

vs. β-AR with RMSE values of 366.3 vs. 406.6 and 828.8 vs. 1045.9 for the 7 and 14-day forecasting horizons, respectively. Regardless of which analytical method was used, RMSE or AE, enhancing the model with Ct features reduced the error to the real incidence value and provided more accurate forecasting than the β-AR model.

Discussion

In this study, we demonstrate that Ct-derived information from routine PCR testing can detect epidemic waves earlier than case-based trajectories, and by doing so, provides an earlier indication of upcoming COVID-19 waves in the community. We took the β-AR

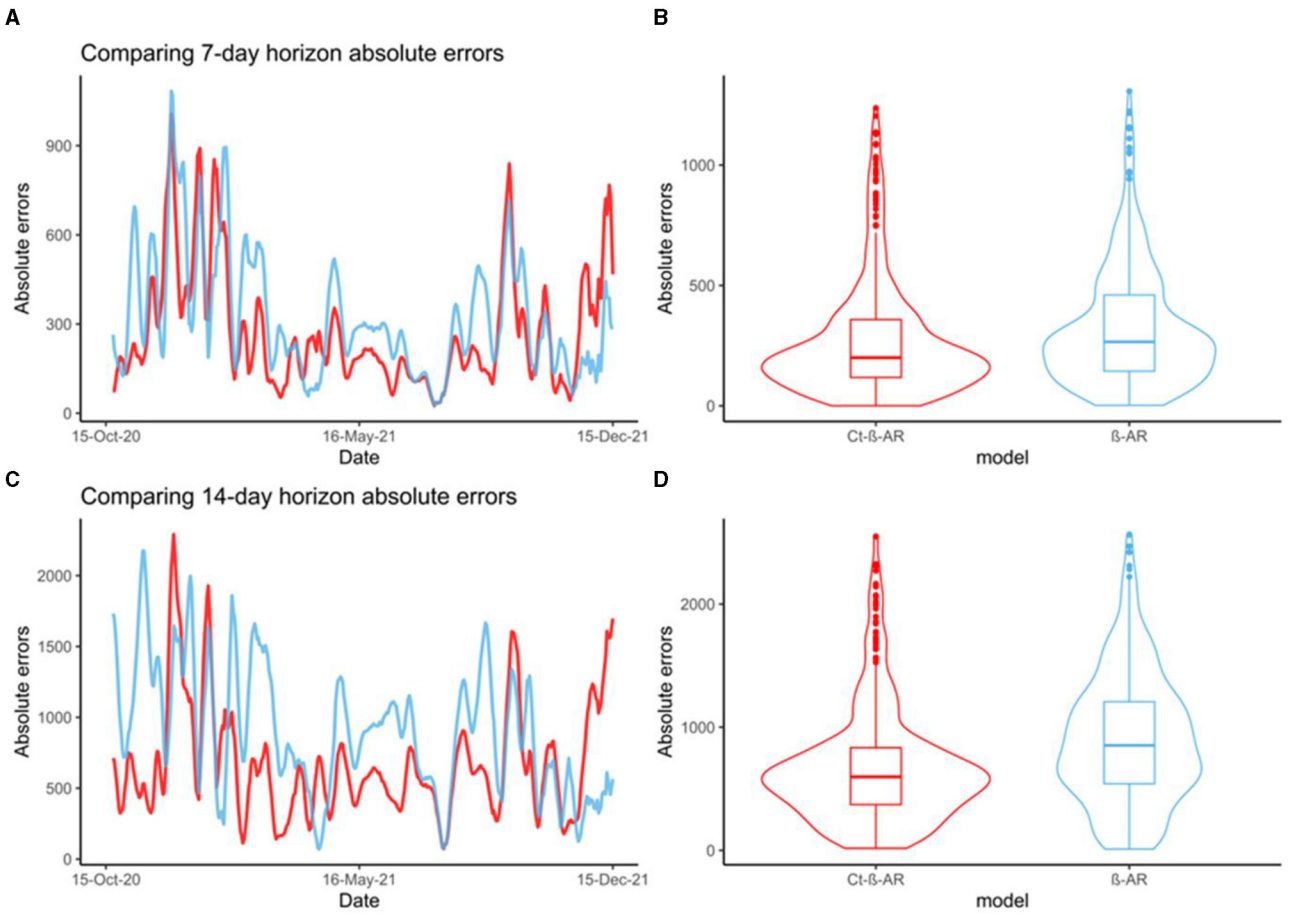

FIGURE 4
(A) Comparing trends of 7-day horizon absolute errors for Ct-β-AR (red) and β-AR (blue) models; **(B)** Comparing distributions of 7-day horizon absolute errors for Ct-β-AR (red) and β-AR (blue) models; **(C)** Comparing trends of 14-day horizon absolute errors for Ct-β-AR (red) and β-AR (blue) models; **(D)** Comparing distributions of 14-day horizon absolute errors for Ct-β-AR (red) and β-AR (blue) models.

forecasting model which uses deep learning and autoregression to produce weekly forecasts up to a 4-week horizon and enhanced it with Ct features from PCR test results. The β-AR model enhanced by Ct-derived features provided more accurate forecasts (lower RMSE/AE) than the β-AR model without Ct features.

Our observations are consistent with three previous studies and further corroborate and highlight the value of Ct values in epidemic forecasting (10–12). The study conducted in Madagascar showed that epidemic growth curves [Hay et al] integrating Ct-derived features (median, skewness) slightly preceded incidence data-derived epidemic growth curves (12). Likewise, the study conducted in Belgium that correlated daily/weekly Ct value distributions (mean) and case counts showed that epidemic trajectories incorporating Ct values preceded those derived by case counts by 17 days (11). Lastly, the Hong Kong study, observed 7 days of leading Rt waves from the log-regression model derived from the Ct statistics (mean, skewness) as compared to the Rt waves derived from classic case counts (10). Our study adds to this body of evidence that highlights the importance of Ct values for epidemic trajectories and how they could be leveraged to predict surges in a timely manner.

Some of the more widely used forecasting models use features such as population mobility data, international flight data, government restrictions, non-pharmaceutical interventions, vaccination, and seasonality, as well as COVID-19 symptoms survey data (19, 21–24). With the exception of COVID-19 symptom data, most of these features such as international flight data, government restrictions etc. are not grounded in the inherent biology of the infection. There are other forecasting methods that have been developed that take some of the biological characteristics of the virus into account; for example, wastewater surveillance (25) that measures concentrations of the virus in wastewater to infer community-level disease dynamics, or self-reported COVID-19 symptoms (26), or detecting pre-symptomatic cases using digital wearable devices (27, 28), or a combination of self-reporting and digital wearable devices (29). However, as compared to these models, Ct values would still be the most readily available since they are automatically generated with most routine COVID-19 RT-PCR tests. Nevertheless, features such as wastewater surveillance data can be used in conjunction with Ct data to enhance forecasting models as future work, but is currently beyond the scope of this paper.

While our Ct-enhanced model showed better forecasting than the case-count derived forecasting approach, our study is not without limitations. Firstly, our data were obtained from a single PCR assay. However, Hay et. al demonstrated that the

using different PCR methods does not influence the ability the Ct-based model to accurately estimate the epidemic trajectory. Secondly, currently used forecasting models can make predictions at different geographical levels such as at the national, state (23) [and/or the county level (30–32) while we assessed our model at a smaller geographical scale, at the community and the county level. Thirdly, we enhanced only one forecasting model, the β-AR model, to assess the usefulness of incorporating Ct values. There are other publicly available forecasting models employing a compartmental model methodology (21, 23, 24, 33–38), machine learning techniques (30, 39, 40), deep learning approaches (41–45), ensemble methods (9, 46, 47), statistical methods (48–51), hybrid approaches (52–57) that could be enhanced to validate our approach. Additionally, campus case-counts were not adjusted for different samples taken from the same individual or for re-infection, but this would only impact campus case-based epidemic trajectories. Future work is needed to assess the utility of Ct values in early detection of other pathogens and across diverse geographical locations. Lastly, we acknowledge limitations around cost and availability of PCR tests, especially during lockdowns or in low-resource settings. To that point, we advocate that when resources are spent on conducting these tests, we get the most out of them by not only just using the binary positive/negative result, but also the generated Ct values which give a better picture of the true epidemic trajectory in the community.

More importantly, what this study, along with others, has shown is the potential of utilizing PCR Ct values for surveillance of other respiratory infectious diseases such as seasonal influenza or its broader use in the field of other infectious diseases. The value of PCR Ct data has often been overlooked for public health surveillance due to lack of awareness and/or due to the lack of data availability. There is now growing evidence that Ct-derived forecasting methods are an effective method for sentinel surveillance and early identification of surges. Many groups have advocated for public reporting of Ct information for surveillance, especially in countries with limited public health surveillance resources with routine testing laboratories even acting as "epidemic sensors" for the geographical regions they serve (11, 12). Given the compelling evidence for the use of Ct values in disease monitoring and forecasting, we strongly urge public health agencies to make PCR Ct values publicly available along with the binary test results.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Northeastern 223 University Office of Human Subject Research Protection (HSRP). The studies were conducted in accordance with the local legislation and institutional requirements. The study used deidentified samples that were not requiring consent and this exemption was approved by Northeastern 223 University Office of Human Subject Research Protection (HSRP).

Author contributions

MS, MM, DW, and IM designed research. JA provided the data. MS and IM performed research and analyzed data. MS, IM, OS, and MG wrote the paper. All authors contributed to the article and approved the submitted version.

Conflict of interest

MS, MM, DW, OS, MG, and IM are employed by Thermo Fisher Scientific. The funder had the following involvement with the study: research support (salaries of employees only).

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Martín J, Tena N, Asuero AG. Current state of diagnostic, screening and surveillance testing methods for COVID-19 from an analytical chemistry point of view. *Microchem J.* (2021) 167:106305. doi: 10.1016/j.microc.2021.106305

2. Jewell NP, Lewnard JA, Jewell BL. Predictive mathematical models of the COVID-19 pandemic: underlying principles and value of projections. *Jama.* (2020) 323:1893–4. doi: 10.1001/jama.2020.6585

3. Cramer EY, Huang Y, Wang Y, Ray EL, Cornell M, Bracher J, et al. The United States COVID-19 forecast Hub dataset. *Scientific Data.* (2022) 9:462. doi: 10.1038/s41597-022-01517-w

4. Ross R. An application of the theory of probabilities to the study of a priori pathometry—Part I. *Proceedings of the Royal Society of London Series A, Containing Papers of a Mathematical and Physical Character.* (1916) 92:204–30. doi: 10.1098/rspa.1916.0007

5. Harko T, Lobo FSN, Mak MK. Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. *Appl Math Comput.* (2014) 236:184–94. doi: 10.1016/j.amc.2014.03.030

6. Hay JA, Kennedy-Shaffer L, Kanjilal S, Lennon NJ, Gabriel SB, Lipsitch M, et al. Estimating epidemiologic dynamics from cross-sectional viral load distributions. *Science.* 373:eabh0635. doi: 10.1126/science.abh0635

7. Díaz-Pachón DA, Rao JS, A. simple correction for COVID-19 sampling bias. *J Theor Biol.* (2021) 512:110556. doi: 10.1016/j.jtbi.2020.110556

8. Kucirka LM, Lauer SA, Laeyendecker O, Boon D, Lessler J. Variation in false-negative rate of reverse transcriptase polymerase chain reaction-based SARS-CoV-2 tests by time since exposure. *Ann Intern Med.* (2020) 173:262–7. doi: 10.7326/M20-1495

9. National Center for Immunization and Respiratory Diseases (NCIRD) DoVD. COVID-19 Forecasting and Mathematical Modeling. (2020). Available online at: https://www.cdc.gov/coronavirus/2019-ncov/science/forecasting/mathematical-modeling.html (accessed March 03, 2023).

10. Lin Y, Yang B, Cobey S, Lau EHY, Adam DC, Wong JY, et al. Incorporating temporal distribution of population-level viral load enables real-time estimation of COVID-19 transmission. *Nat Commun*. (2022) 13:1155. doi: 10.1038/s41467-022-28812-9

11. Yin N, Dellicour S, Daubie V, Franco N, Wautier M, Faes C, et al. Leveraging of SARS-CoV-2 PCR cycle thresholds values to forecast COVID-19 trends. *Front Med*. (2021) 8:743988. doi: 10.3389/fmed.2021.743988

12. Andriamandimby SF, Brook CE, Razanajatovo N, et al. Cross-sectional cycle threshold values reflect epidemic dynamics of COVID-19 in Madagascar. *Epidemics*. (2022) 38:100533. doi: 10.1016/j.epidem.2021.100533

13. Administration USFaD. *TaqPath COVID-19 Combo Kit—Instructions for Use—FDA*. (2020). Available online at: https://www.fda.gov/media/136112/download (accessed December 1, 2023).

14. University N. *Facts and Figures 2021*. (2021). Available online at: https://facts.northeastern.edu/ (accessed December 1, 2023).

15. Times TNY. *Coronavirus (COVID-19) Data in the United States*. (2021). Available online at: https://github.com/nytimes/covid-19-data (accessed December 1, 2023).

16. Bureau UC. *Quick Facts Suffolk County, Massachusetts*. (2023). Available online at: https://www.census.gov/quickfacts/suffolkcountymassachusetts/ (accessed December 1, 2023).

17. James A. Hay LK-S. *jameshay218/virosolver*. Zenodo: Publication release (v102) (2021).

18. Cori A, Cauchemez S, Ferguson NM, Fraser C, Dahlqwist F, Demarsh PA, et al. *Estimate Time Varying Reproduction Numbers from Epidemic Curves*. (2021). Available online at: https://cran.r-project.org/web/packages/EpiEstim/index.html (accessed March 03, 2023).

19. Maximilian Nickel LS, Mark Ibrahim, Matt Le, Timothee L. *Neural Relational Autoregression for High-Resolution COVID-19 Forecasting*. FB Data for Good. (2020). Available online at: https://ai.meta.com/research/publications/neural-relational-autoregression-for-high-resolution-covid-19-forecasting/

20. Gangavarapu K, Latif AA, Mullen JL, Alkuzweny M, Hufbauer E, Tsueng G, et al. Outbreak.info genomic reports: scalable and dynamic surveillance of SARS-CoV-2 variants and mutations. *medRxiv*. (2022) 20:512–22. doi: 10.1038/s41592-023-01769-3

21. Li ML, HTB, OSL. *Overview of DELPHI Model V3—COVID Analytics*. In: Massachusetts Institute of Technology, Brown University. (2020).

22. Fox Spencer J, Lachmann M, Tec M, et al. Real-time pandemic surveillance using hospital admissions and mobility data. *Proc Nat Acad Sci*. (2022) 119:e2111870119. doi: 10.1073/pnas.2111870119

23. Lemaitre JC, Grantz KH, Kaminsky J, et al. A scenario modeling pipeline for COVID-19 emergency planning. *Sci Rep*. (2021) 11:7534. doi: 10.1038/s41598-021-86811-0

24. Yamana T, Pei S, Kandula S, Shaman J. Projection of COVID-19 cases and deaths in the US as individual states re-open may 4, 2020. *medRxiv*. (2020) 4:2020–5. doi: 10.1101/2020.05.04.20090670

25. Zhu Y, Oishi W, Maruo C, Saito M, Chen R, Kitajima M, et al. Early warning of COVID-19 via wastewater-based epidemiology: potential and bottlenecks. *Sci Total Environ*. (2021) 767:145124. doi: 10.1016/j.scitotenv.2021.145124

26. Canas LS, Sudre CH, Pujol JC, Polidori L, Murray B, Molteni E, et al. Early detection of COVID-19 in the UK using self-reported symptoms: a large-scale, prospective, epidemiological surveillance study. *Lancet Digit Health*. (2021) 3:e587–98. doi: 10.1016/S2589-7500(21)00131-X

27. Mishra T, Wang M, Metwally AA, Bogu GK, Brooks AW, Bahmani A, et al. Pre-symptomatic detection of COVID-19 from smartwatch data. *Nat Biomed Eng*. (2020) 4:1208–20. doi: 10.1038/s41551-020-00640-6

28. Alavi A, Bogu GK, Wang M, Rangan ES, Brooks AW, Wang Q, et al. Real-time alerting system for COVID-19 and other stress events using wearable data. *Nat Med*. (2022) 28:175–84. doi: 10.1038/s41591-021-01593-2

29. Brakenhoff TB, Franks B, Goodale BM, van de Wijgert J, Montes S, Veen D, et al. A prospective, randomized, single-blinded, crossover trial to investigate the effect of a wearable device in addition to a daily symptom diary for the remote early detection of SARS-CoV-2 infections (COVID-RED): a structured summary of a study protocol for a randomized controlled trial. *Trials*. (2021) 22:694. doi: 10.1186/s13063-021-05241-5

30. Sundar S. MIT_Crit_Data-GBCF: a non-mechanistic approach to real-time forecasting of US COVID-19 mortality using machine learning and novel digital data (v.1.2.0). *Zenodo*. (2021). doi: 10.5281/zenodo.5559076

31. Bosch J, Wilson A, O'Neil K, Zimmerman PA. COVID-19 predict—Predicting pandemic trends. *medRxiv*. (2020) 5:1593. doi: 10.1101/2020.09.09.20191593

32. Toledo. GISaAGGCatUo. *COVID-19 Case Prediction*. Available online at: https://gisag.maps.arcgis.com/apps/dashboards/c1e2502c67204f5683baf07bbc387867 (accessed March 03, 2023).

33. Pei S, Shaman J. Initial simulation of SARS-CoV2 spread and intervention effects in the continental US. *medRxiv*. (2020) 5:2020–3. doi: 10.1101/2020.03.21.20040303

34. Cramer Estee Y, Ray Evan L, Lopez Velma K, et al. Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States. *Proc Nat Acad Sci*. (2022) 119:e2113561119. doi: 10.1073/pnas.2113561119

35. Khan ZS, Van Bussel F, Hussain F. *A Predictive Model for COVID-19 Spread Applied to Eight US States*. (2020). Available online at: https://ui.adsabs.harvard.edu/abs/2020arXiv200605955K (accessed June 01, 2020).

36. Budzinski J. *COVID-19 Analyses*. (2021). Available online at: https://github.com/joebud/covid-19-analyses (accessed March 03, 2023).

37. Zou D, Wang L, Xu P, Chen J, Zhang W, Gu Q. Epidemic model guided machine learning for COVID-19 forecasts in the United States. *medRxiv*. (2020) 25:2020–5. doi: 10.1101/2020.05.24.20111989

38. Pagano B. *COVID-19 Modeling*. (2022). Available online at: https://bobpagano.com/covid-19-modeling/ (accessed March 03, 2023).

39. EOD. *Random-Walks*. (2020). Available online at: https://github.com/e3bo/random-walks (accessed March 03, 2023).

40. Quoc Tran HH, Lam Ho. COVID-19 Prediction. (2020). Available online at: https://covid19.aipert.org/ (accessed March 03, 2023).

41. Rodríguez A, Tabassum A, Cui J, et al. Deep COVID: an operational deep learning-driven framework for explainable real-time COVID-19 forecasting. *Proceed AAAI Conf Artif Intell*. (2021) 35:15393–400. doi: 10.1609/aaai.v35i17.17808

42. Chung-Yan Fong D-YF. *HKUST-DNN, State-level COVID-19 Trend Forecasting Using Deep Neural Networks. GitHub page*. (2022). Available online at: https://cfong32.github.io/HKUST-DNN/ (accessed March 03, 2023).

43. Wu D, Gao L, Xiong X, Chinazzi M, Vespignani A, Ma YA, et al. *DeepGLEAM: A Hybrid Mechanistic and Deep Learning Model for COVID-19 Forecasting*. (2021):arXiv:2102.06684. Available online at: https://ui.adsabs.harvard.edu/abs/2021arXiv210206684W (accessed February 01, 2021).

44. Zadeh K. *SLSTM (Covid-lstm-hosp)*. (2022). Available online at: https://github.com/geohai/covid-lstm-hosp (accessed March 03, 2023).

45. Zheng S, Gao Z, Cao W, Bian J, Liu T-Y. HierST: A Unified Hierarchical Spatial-temporal framework for COVID-19 trend forecasting. In: *Proceedings of the 30th ACM International Conference on Information &amp Knowledge Management Association for Computing Machinery*. (2021) 5:4383–92. doi: 10.1145/3459637.3481927

46. Lami OS. *Markov Decision Process Model for Covid Prediction*. (2020). Available online at: https://github.com/oskali/mit_cassandra (accessed March 03, 2023).

47. University CM. *Developing the Theory and Practice of Epidemiological Forecasting*. (2020). Available online at: https://github.com/cmu-delphi/ (accessed March 03, 2023).

48. Andrea Kraus DK. *COVID Global Dashboard*. Available online at: https://krausstat.shinyapps.io/covid19global/#section-about (accessed March 03, 2023).

49. Ali Jadbabaie AS, Devavrat S. *Real Time Mixture-Based Predictions*. (2020). Available online at: https://idss.mit.edu/vignette/real-time-mixture-based-predictions/ (accessed March 03, 2023).

50. Karlen D. *Characterizing the Spread of COVID-19*. (2020) arXiv:2007–56. Available online at: https://ui.adsabs.harvard.edu/abs/2020arXiv200707156K (accessed July 01, 2020).

51. Walraven R. *RobertWalraven-ESG*. Available online at: http://rwalraven.com/COVID19/Model (accessed March 03, 2023).

52. Wilson DJ. Weather, Mobility, and COVID-19: *A Panel Local Projections Estimator for Understanding and Forecasting Infectious Disease Spread*. Federal Reserve Bank of San Francisco Working Paper 2020-23. (2021). doi: 10.24148/wp2020-23

53. Chen Y, Lin YT, Miller EF. Impacts of vaccination and severe acute respiratory syndrome coronavirus 2 variants alpha and delta on coronavirus disease 2019 transmission dynamics in the 15 most populous metropolitan statistical areas in the United States. *medRxiv*. (2021) 2021–10. doi: 10.1101/2021.10.19.21265223

54. Chen YC, Lu PE, Chang CS, Liu TH, A. Time-dependent SIR model for COVID-19 with undetectable infected persons. *IEEE Transact Network Sci Engin*. (2020) 7:3279–94. doi: 10.1109/TNSE.2020.3024723

55. Pottier L. Forecast of the covid19 epidemic in France. *medRxiv*. (2021) 3:2021–4. doi: 10.1101/2021.04.13.21255418

56. Ben-Nun M, Riley P, Turtle J, Bacon DP, Riley S. Forecasting national and regional influenza-like illness for the USA. *PLoS Comput Biol*. (2019) 15:e1007013. doi: 10.1371/journal.pcbi.1007013

57. UCF TA-EMga. UCF-AEM. (2020). Available online at: https://github.com/UCF-AEM/UCF-AEM