

Fusing Low-Latency Data Feeds with Death Data to Accurately Nowcast COVID-19 Related Deaths

Conor Rosato* Robert E. Moore* Matthew Carter* John Heap†
Jose Storopoli‡ Simon Maskell*

Abstract

The emergence of the novel coronavirus (COVID-19) has generated a need to quickly and accurately assemble up-to-date information related to its spread. While it is possible to use deaths to provide a reliable information feed, the latency of data derived from deaths is significant. Confirmed cases derived from positive test results potentially provide a lower latency data feed. However, the sampling of those tested varies with time and the reason for testing is often not recorded. Hospital admissions typically occur around 1-2 weeks after infection and can be considered out of date in relation to the time of initial infection. The extent to which these issues are problematic is likely to vary over time and between countries.

We use a machine learning algorithm for natural language processing, trained in multiple languages, to identify symptomatic individuals derived from social media and, in particular Twitter, in real-time. We then use an extended SEIRD epidemiological model to fuse combinations of low-latency feeds, including the symptomatic counts from Twitter, with death data to estimate parameters of the model and nowcast the number of people in each compartment. The model is implemented in the probabilistic programming language Stan and uses a bespoke numerical integrator. We present results showing that using specific low-latency data feeds along with death data provides more consistent and accurate forecasts of COVID-19 related deaths than using death data alone.

Key Words: Bayesian Inference, Epidemiological Modelling, COVID-19, Stan, MAE, NEES

1. Introduction

The novel coronavirus (COVID-19), has at the time of writing, resulted in over 4.55 million deaths and 219 million confirmed cases worldwide as of 6th October 2021. By January 2020, new cases had been seen throughout Asia, and by the time the World Health Organisation (WHO) declared a global pandemic in March 2020, COVID-19 had spread to over 100 countries. Therefore, it was imperative to establish reliable data feeds relating to the pandemic so that researchers and analysts could model the ongoing spread of the disease and inform decision-making by government and public health officials. These data sets and models must be open-source to facilitate collaboration between researchers and allow for published results to be replicated and scrutinised. A popular interactive dashboard that collates total daily counts of confirmed cases and deaths for countries, and in some cases, regions within countries exists here [1]. These variables are traditionally used to calculate metrics such as the reproduction number (R_t), which is vital in understanding both the number of people on average an infected person infects and the infection growth rate or daily rate of new infections. The quality of the metrics calculated is heavily dependent on the model and ingested data.

In the United Kingdom (UK), there has been a joint effort to produce estimates of the (R_t) number, with some notable examples seen here [2]. Laboratory-confirmed COVID-19

*Department of Electrical Engineering and Electronics, University of Liverpool, United Kingdom

†Computational Biology Facility, University of Liverpool, United Kingdom

‡Department of Computer Science, Universidade Nove de Julho - UNINOVE, Sao Paulo - Brazil

diagnoses are used in [3], UK's National Health Service (NHS) Pathways data is used in [4] and hospital admissions data is used in [5]. The statistical model developed by Moore, Rosato and Maskell [6] contributed to these estimates by incorporating death, hospital admission and NHS 111 call data.

Terms such as “Infodemiology”, and “Infoveillance” described in [7] refer to the ability to process and analyse data that is created and stored digitally in real-time pertinent to disease outbreaks. The availability of these datasets, particularly at the beginning of an outbreak when very little is known, could provide a noisy but accurate representation of disease dynamics. A popular method includes extracting data from social media and, in particular, Twitter. Before the pandemic, tweets relating to influenza-like-illness symptoms were seen to substantially improve the models predicting capacity in [8] and boost nowcasting accuracy by 13% in [9]. In relation to the COVID-19 pandemic, there have been many research papers published that use social media to gain valuable information relating to the pandemic from what people tweet in real-time. Public sentiment relating to prevention strategies was analysed in [10] while [11] showed that emotion changed from fear to anger during the first stages of the pandemic. Misinformation and conspiracy theories have been shown to have propagated rapidly through the Twittersphere during the pandemic [12]. Studies have used machine learning algorithms to automatically detect tweets containing self-reported symptoms mentioned by users [13] with [14] finding that the symptoms reported by Twitter users were similar to those used in a clinical setting. To the best of our knowledge, researchers have yet to use these symptomatic tweets to calibrate epidemiological models.

The contribution of this paper is twofold: firstly, we outline how to identify symptomatic tweets that correspond to COVID-19 related symptoms in multiple languages. The geolocation information associated with each tweet, when available, is extracted, and counts per country or region are aggregated to produce estimates for the previous 24 hours. Secondly, we present a comprehensive study of how these symptomatic tweets differ from other open-source datasets when used to calibrate the extended SEIR model described in section 3 for different geographic locations. When incorporating the surveillance data, outlined in section 2, the Mean Absolute Error (MAE) and Normalised Estimation Error Squared (NEES) values are calculated when making 7-day death forecasts.

An outline of the paper is as follows. In Section 2 we describe the open-source data feeds included in the comparative study and provide the methodology for extracting the symptomatic tweets in real-time. A description of the model is presented in Section 3, with an outline of the computational experiments and results in Section 4. Concluding remarks and directions for future work are described in Section 5.

2. Data

The surveillance data used for each geographical location are summarised in Table 1. Death and positive case data for the US States and the rest of the world were downloaded from the dashboard operated by the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) [1]. It should be noted that testing methods and criteria for classifying deaths as COVID-19-related differ between geographic locations.

For NHS region-specific data, the number of deaths includes individuals with COVID-19 as the cause of death on their death certificate or those who died within 60 days of a positive test result. Patients admitted to hospital with COVID-19 symptoms and individuals that input symptoms to the ZOE COVID Symptom Study app comprise the hospital admissions and Zoe app datasets, respectively. Individuals that reported symptoms via the NHS Pathways triage and online Dashboard comprise the 111 calls and 111 online assess-

Table 1: Data Sources

Geographic Location	Data Feed	Start Date	Reference
U.S States and the rest of the world	Deaths	24th March 2020	[1]
	Tests	1st March 2020	[1]
	Twitter	13th April 2020	Section 2.1
U.K NHS Regions	Deaths	24th March 2020	[16]
	Hospital admissions	19th March 2020	[17]
	Twitter	9th April 2020	Section 2.1
	Zoe app	12th May 2020	[18]
	111 calls	18th March 2020	[19]
	111 online	18th March 2020	[19]

ments datasets, respectively. Note that the Zoe app, 111 calls and 111 online assessments may include individuals who have COVID-19 symptoms but have not tested positive and individuals who perceive they have symptoms and do not have COVID-19.

All code and datasets can be found on the CoDatMo GitHub repository [15]. The authors set this up to facilitate the sharing of code, data and ideas when modelling COVID-19.

2.1 Twitter

We created an interactive website¹ that maps symptomatic tweets to geographical locations with daily counts representing the total amount of symptomatic tweets from the previous 24 hours. Information on how to download the data as a JSON can be found on the website.

2.1.1 Pre-processing

The Twitter streaming API is filtered using keywords that align with COVID-19 symptoms from the MedDRA database [20] in English, German, Italian, Portuguese and Spanish, including terms for fever, cough and anosmia. We note that our analysis indicated that explicit COVID-19 terms (e.g. ‘coronavirus’) rarely related to individuals with symptoms. Such terms were therefore excluded. Official retweets or tweets beginning with #RT were removed to avoid duplication of tweets within the dataset.

2.1.2 Symptom Classifier Breakdown

A multi-class support vector machine (SVM) [21] was trained with a set of annotated tweets that were vectorised using a skip-gram model. The annotated tweets were labelled according to the following classes:

1. Unrelated tweet,
2. User currently has symptoms,
3. User had symptoms in the past,
4. Someone else currently has symptoms,
5. Someone else had symptoms in the past.

¹<https://pgb.liv.ac.uk/~johnheap/>

Table 2: Performance Measures

Language	Number of Data Used		Performance Measures			
	Training	Testing	F1	Accuracy	Precision	Recall
English	1105	195	0.85	0.85	0.85	0.85
German	412	260	0.89	0.89	0.90	0.89
Italian	254	260	0.97	0.96	0.97	0.96
Portuguese	3507	619	0.77	0.77	0.78	0.80
Spanish	1530	270	0.82	0.85	0.82	0.85

The sum of tweets in classes 2-5, which is the total number of tweets that mention symptoms, was calculated for each 24-hour period. Geo-tagged tweets were mapped to their location, e.g. the corresponding city, town or village, via a series of tests using shapefiles of different countries. Previous studies demonstrate that approximately 1.65% of tweets are geo-tagged [22], where the exact position of the tweeter when the tweet was posted is recorded using longitude and latitude measurements. For tweets that are not geo-tagged, we look at the author’s profile to ascertain whether they provide an appropriate location. We deemed the server offline if there were any 15 minute periods during the previous 24 hours that did not have any recorded tweets. After checking all 96 15 minute periods, the count in each geographical area was multiplied by a correction factor: reported tweet count = total tweet count * 96/(96 - downtime periods).

For each language, we labelled the corpus of tweets with native speakers, with the associated class label and randomly up- and down-sampled under- and over-represented classes such that the classifier was trained with a balanced dataset. A subset of data was used to train the classifier before testing it on the remainder. The total number of labelled tweets used for training and testing of the classifiers and the resulting performance metrics can be seen in Table 2.

3. Model

We repurpose the statistical model developed by Moore, Rosato and Maskell [6] by tweaking the observation model to be compatible with each group of surveillance data types that we use to calibrate the model in the computational experiments. We calibrate the model with a minimum of death data in all experiments, and the associated component of the observation model is unchanged. We extend the observation model to assimilate the other types of surveillance data that feature in Table 1, including Twitter and Zoe app data, by adding an extra component for each additional data type. These extra components of the observation model, the number of which can change between experiments, mirror the structure for symptom report data in the original model. More explicitly, we assume for these extra components that a generic count on day t , $x_{\text{obs}}(t)$, has a negative binomial distribution,

$$x_{\text{obs}}(t) \sim \text{NegativeBinomial}(x(t), \phi_x), \quad (1)$$

parameterised by a mean $x(t)$ and overdispersion parameter ϕ_x .

4. Computational Experiments

The time series we consider begins on 17th February 2020, with the start dates of the data feeds outlined in Table 1. We consider the end of time in our analysis for the US States and the rest of the world and NHS regions to be the 1st February 2021 and 7th January 2021,

respectively. In all cases, we consider a forecast to include seven days. For US states and the rest of the world, we include three predictions in the analysis; 9th July 2020 to 16th July 2020, 17th October 2020 to 24th October 2020 and 25th January 2021 - 1st February 2021. The analysis for UK NHS regions includes six predictions; 11th November 2020 to 18th November 2020, 21st November 2020 - 28th November 2020, 1st December - 8th December, 11th December - 18th December, 21st December - 28th December and 31st December 2020 - 7th January 2021.

Similar to the experiments in [6], the analysis in this paper was run on the University of Liverpool's High-Performance Computer (HPC). Each node has two Intel(R) Xeon(R) Gold 6138 CPU @ 2.00GHz processors, a total of 40 cores and 384 GB of memory. In the following experiments, six independent Markov Chains draw 2000 samples each, with the first 1000 discarded as burn-in. Run-time differs depending on the country and at what point in the time series the prediction is made, but it typically takes 4.5 hours per Markov Chain to complete.

We initially calibrate the model solely with death data and produce posterior predictive distributions of deaths for the following geographic locations independently:

- **US:** 50 States,
- **Rest of World:** 2 European and 16 Latin American countries,
- **UK:** 7 NHS regions.

We consider the final 7-daily deaths in this forecast to be the baseline to compare forecasts of deaths when incorporating low-latency data feeds. We use two metrics in our analysis to determine the accuracy of the resulting forecasts. Firstly, we calculate the MAE, which shows the average error over a set of predictions:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N (x^i - y^i), \quad (2)$$

where N is the number of predictions, and x_i and y_i are the predicted and true number of deaths on day i , respectively.

Secondly, we consider the uncertainties associated with the forecasts by assessing the NEES, which is a popular method in the field of signal processing and tracking [23] and recently applied to epidemiological forecasts in [6], to determine if the estimated variance of forecasts from an algorithm differs from the true variance. If the variance is larger than the true variance, then the algorithm is over-cautious, and if the estimated variance is smaller than the true variance, it is over-confident. The NEES is defined by:

$$\text{NEES} = \frac{1}{N} \sum_{i=1}^N (x^i - y^i)^T C^{i-1} (x^i - y^i), \quad (3)$$

where C^{i-1} is the estimated variance at day i , as approximated using the variance of the samples for that day. If x^i is D dimensional, then C^i should be a $D \times D$ matrix, and the NEES should be equal to D if the algorithm is consistent. Therefore, in assessing death forecasts, an ideal NEES value is $D \approx 1$.

4.1 Results

The NEES values and MAE percentage differences between the baseline, of ingesting solely deaths, and the incorporation of low-latency data feeds for US States and the rest of the world and UK's NHS regions can be seen in Tables 3 and 4, respectively. The results in these tables are averaged over the prediction periods described in section 4.

When forecasting deaths using the data available from [1], we have shown that calibrating the model with tests and tweets gives comparable increases in performance for US States, however for the rest of the world, tweets give a -17% improvement compared to just -6% for tests. For US States and the rest of the world, there is an improvement of -5% and -24%, respectively when tests and tweets are used to calibrate the model. An example of this improvement can be seen in Figure 1 for the prediction period 25th January 2021 - 1st February 2021. When comparing the mean sample, outlined in red, incorporating tests and tweets follows the deaths trend with more accuracy.

When comparing the NEES values for NHS regions in Table 4 against the baseline, it is evident that including a data feed improves the consistency of forecasts. The exception is the Zoe App data, which on average, produces estimates that are over-confident. This can be seen in Figure 2 for the prediction period 31st December 2020 - 7th January 2021. When comparing the MAE % difference, the results are similar. Ingesting hospital admissions, 111 calls, and 111 online assessments data provides improvements of -22%, -17% and -22%, respectively. However, there is a more significant difference in the MAE of 124% when including the Zoe App data than the 2% from including the Tweet data. We perceive this issue arises because, in the context of these feeds, the symptoms are self-diagnosed. Consequently, the counts may well include relatively large numbers of people who do not have COVID-19. We do not currently consider such 'false alarms' in the model described in section 3 but hope to extend it to handle these in the future.

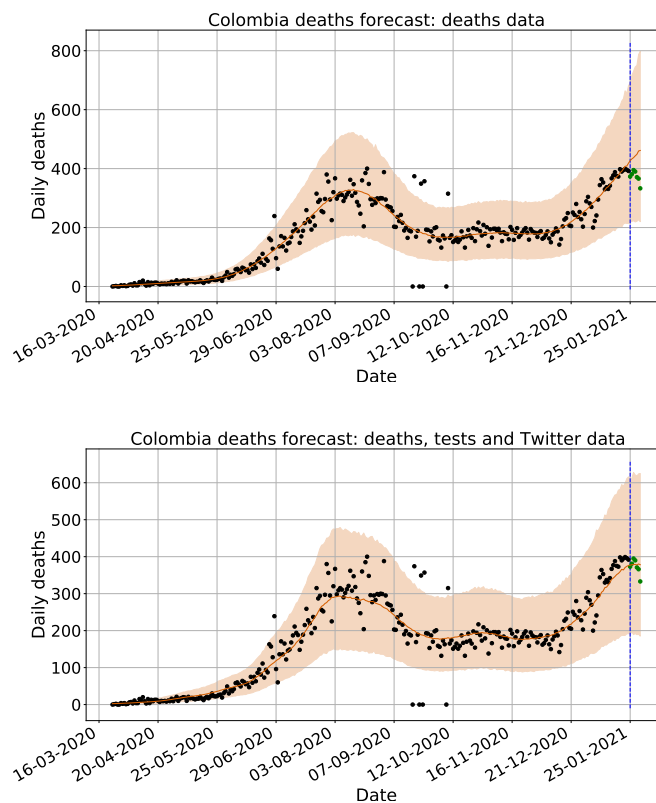


Figure 1: Deaths forecast for Colombia using solely death data (top) and deaths, tests and Twitter data (bottom). The orange ribbon is 1 standard deviation from the mean, the red line is the mean sample and the start of the predictions is the vertical dashed blue line. The black and green dots are the true deaths.

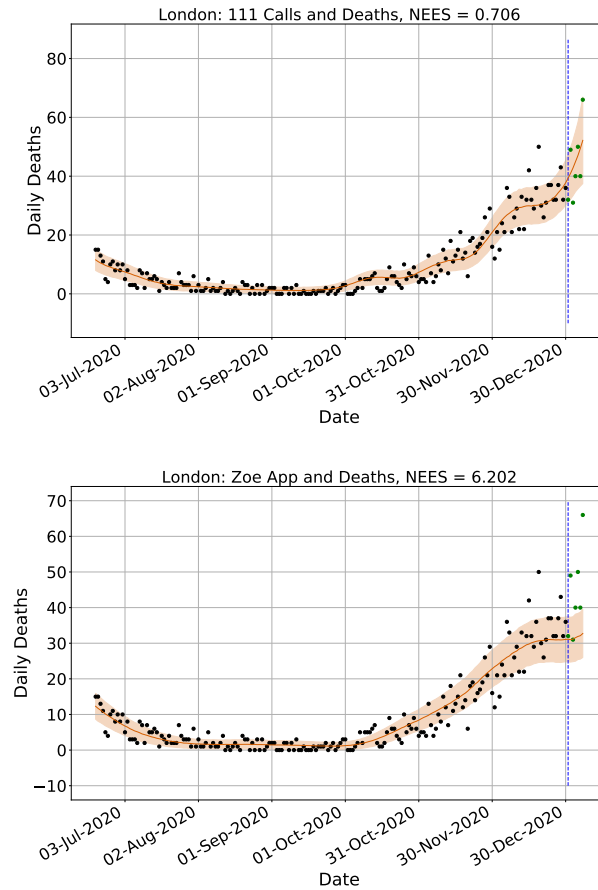


Figure 2: Deaths forecast for London using death and 111 calls data (top) and death and Zoe App data (bottom). The orange ribbon is 1 standard deviation from the mean, the red line is the mean sample and the start of the predictions is the vertical dashed blue line. The black and green dots are the true deaths.

5. Conclusions and Future Work

We have shown that calibrating the epidemiological model outlined in section 3 with certain low-latency data feeds provides more accurate and consistent nowcasts of daily deaths when compared with using death data alone.

Incorporating tweets for UK regions does not provide the same level of improvement as for other geographies. This reduced improvement could be down to many factors, including the total daily counts for NHS regions being less plentiful than for the US States or the rest of the world. We used the free Twitter streaming API for this research, which limits the number of tweets available to download to 1%. However, it is possible to pay for a Premium API that allows the user to download a higher percentage of tweets. A second way to potentially increase the hit rate of geo-located tweets is to use natural language processing techniques, such as those outlined in the review [24], to estimate the location of the tweet user. Another direction for future work is to train a more sophisticated classifier such as the Bidirectional Encoder Representations from Transformers (BERT) classifier [25].

As for the symptom report data in the original model that Moore, Rosato and Maskell introduce [6], we have assumed that all types of low-latency surveillance data are a weighted sum of current and lagged instances of the time series of new infections. We want to consider alternative structures for the observation model that link directly to the intermediate

states of the transmission model in the future.

Funding

This work was supported by a Research Studentship jointly funded by the EPSRC and the ESRC Centre for Doctoral Training on Quantification and Management of Risk and Uncertainty in Complex Systems Environments [EP/L015927/1]; an ICASE Research Studentship jointly funded by EPSRC and AWE [EP/R512011/1]; the EPSRC Centre for Doctoral Training in Distributed Algorithms [EP/S023445/1]; and the EPSRC through the Big Hypotheses grant [EP/R018537/1].

Acknowledgements

The authors would like to thank Serban Ovidiu and Chris Hankin from Imperial College London, Ronni Bowman, Riskaware and John Harris for their support and helpful discussions of this work. We would also like to thank the team at the Universidade Nove de Julho - UNINOVE in Sao Paulo, Brazil with the help they provided in labelling the Portuguese tweets. We would also like to thank Breck Baldwin for helping to progress CoDatMo.

References

- [1] Dong E, Du H, and Gardner L. An interactive web-based dashboard to track covid-19 in real time. *The Lancet Infectious Diseases*, 20(5):533 – 534, 2020.
- [2] Reproduction number (r) and growth rate: methodology. <https://www.gov.uk/government/publications/reproduction-number-r-and-growth-rate-methodology/reproduction-number-r-and-growth-rate-methodology>. Accessed: 1 October 2021.
- [3] Birrell P, Blake J, and De Angelis D van Leeuwen E, Gent N. Real-time nowcasting and forecasting of covid-19 dynamics in england: the first wave. *Phil. Trans. R. Soc. B*, 376, 2021.
- [4] Leclerc QJ, Nightingale ES, Abbott S, CMMID COVID-19 Working Group, and Jombart T. Analysis of temporal trends in potential covid-19 cases reported through nhs pathways england. *Sci Rep*, 7106, 2021.
- [5] Keeling MJ., Dyson L, Guyver-Fletcher G, Holmes A, Semple MG, ISARIC4C Investigators, Tildesley MJ., and Hill EM. Fitting to the uk covid-19 outbreak, short-term forecasts and estimating the reproductive number. *medRxiv*, 2020.
- [6] Moore RE, Rosato C, and Maskell S. Assessing the uncertainty of epidemiological forecasts with normalised estimation error squared. *Submitted to Philosophical Transactions of the Royal Society A: Mathematical, Physical, and Engineering Sciences*.
- [7] Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. *J Med Internet Res*, 40(5):154 – 158, 2009.
- [8] Achrekar H, Gandhe A, Lazarus R, Ssu-Hsin Y, and Liu B. Predicting Flu Trends using Twitter data. pages 702–707, April 2011.

- [9] Ovidiu S, Thapen N, Maginnis B, Hankin C, and Foot V. Real-time processing of social media with SENTINEL: A syndromic surveillance system incorporating deep learning for health classification. *Information Processing & Management*, 56(3):1166–1184, May 2019.
- [10] Medford RJ, Saleh SN, Sumarsono A, Perl TM, and Lehmann CU. An infodemic: Leveraging high-volume twitter data to understand early public sentiment for the coronavirus disease 2019 outbreak. *Open Forum Infectious Diseases*, 7(7), 2020.
- [11] Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, and Yang Y. Global sentiments surrounding the covid-19 pandemic on twitter: Analysis of twitter trends. *JMIR Public Health Surveill*, 6, 2020.
- [12] Sharma K, Seo S, Meng C, Rambhatla S, and Liu Y. Covid-19 on social media: Analyzing misinformation in twitter conversations. arxiv: Preprint. <https://arxiv.org/abs/2003.12309>, 2020.
- [13] Al-Garadi MA, Yang YC, Lakamana S, and Sarker A. A text classification approach for the automatic detection of twitter posts containing self-reported covid-19 symptoms. <https://openreview.net/forum?id=xyGSIttHYO>, 2020.
- [14] Sarker A, Lakamana S, Hogg-Bremer W, Xie A, Al-Garadi MA, and Yang YC. Self-reported covid-19 symptoms on twitter: an analysis and a research resource. *Journal of the American Medical Informatics Association*, 27(8):1310–1315, 2020.
- [15] Codatmo. 2021 welcome to the codatmo site. <https://codatmo.github.io>. Accessed: 1 October 2021.
- [16] Uk government. 2021 coronavirus (covid-19) in the uk. <https://coronavirus.data.gov.uk/details/deaths>. Accessed: 1 October 2021.
- [17] Uk government. 2021 coronavirus (covid-19) in the uk. <https://coronavirus.data.gov.uk/details/healthcare>. Accessed: 1 October 2021.
- [18] Zoe app: covid-public-data. <https://console.cloud.google.com/storage/browser/covid-public-data;tab=objects?prefix=&forceOnObjectsSortingFiltering=false>. Accessed: 1 October 2021.
- [19] Nhs digital. 2021 potential coronavirus (covid-19) symptoms reported through nhs pathways and 111 online. <https://digital.nhs.uk/data-and-information/publications/statistical/mi-potential-covid-19-symptoms-reported-through-nhs-pathways-and-111-online/latest>. Accessed: 1 October 2021.
- [20] Covid-19 terms and meddra. <https://www.meddra.org/COVID-19-terms-and-MedDRA>. Accessed: 1 October 2021.
- [21] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, and Duchesnay E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [22] Leetaru K, Wang S, Cao G, Padmanabhan A, and Shook E. Mapping the global twitter heartbeat: The geography of twitter. *First Monday*, 18, 2013.

- [23] Zhaozhong Chen, Christoffer Heckman, Simon Julier, and Nisar Ahmed. Weak in the nees?: Auto-tuning kalman filters with bayesian optimization. In *2018 21st International Conference on Information Fusion (FUSION)*, pages 1072–1079, 2018.
- [24] Huang C-Y, Tong H, He J, and Maciejewski R. Location prediction for tweets. *Frontiers in Big Data*, 2:5, 2019.
- [25] Devlin J, Ming-Wei C, Kenton L, and Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805, 2018.

Table 3: US States and the rest of the world: nowcasting mean absolute error per geographic location when using only death data and when using deaths and different low-latency data feeds.

Geographic Location	Deaths		Tests		Twitter		Tests and Twitter	
	NEES	MAE % Diff	NEES	MAE % Diff	NEES	MAE % Diff	NEES	MAE % Diff
U.S States								
Alaska	0.329	-36	0.334	-29	0.301	-92	0.302	
Alabama	0.684	-29	1.874	-29	1.723	-2	1.000	
Arkansas	0.275	3	0.317	-1	0.288	-1	0.313	
Arizona	0.337	20	0.334	18	0.344	-20	0.244	
California	0.611	6	0.709	9	0.802	5	1.206	
Colorado	1.886	-25	0.401	-41	0.457	10	1.278	
Connecticut	13.406	-8	1.922	-2	0.875	2	11.459	
Delaware	3.020	-3	0.918	16	1.046	12	0.727	
Florida	0.406	-24	0.179	13	0.353	-20	0.454	
Georgia	0.550	9	0.325	41	0.891	-48	0.255	
Hawaii	11.459	-12	28.114	-4	24.695	17	10.149	
Iowa	19.176	5	7.720	4	1.476	-3	1.600	
Idaho	0.914	0	0.809	2	1.791	7	0.986	
Illinois	0.573	9	0.350	13	0.319	-116	1.091	
Indiana	0.561	-17	0.652	-40	0.781	0	0.481	
Kansas	1.021	1	1.037	-2	1.835	1	0.488	
Kentucky	0.355	-4	0.374	10	0.548	-15	0.214	
Louisiana	0.298	-7	0.305	-2	0.341	9	0.234	
Massachusetts	0.351	3	0.342	-3	0.365	14	0.409	
Maryland	0.485	-3	0.619	10	0.581	31	0.313	
Maine	0.488	1	0.567	-28	0.796	-9	0.952	
Michigan	0.592	-6	0.445	-7	0.453	4	0.850	
Minnesota	0.683	9	1.019	11	1.200	51	0.747	
Missouri	0.810	-7	1.165	-27	1.609	20	0.475	
Mississippi	0.683	12	0.721	2	0.997	-15	0.320	
Montana	5.034	4	2.244	-1	1.538	-5	5.189	
North Carolina	0.908	-1	0.453	9	0.877	-19	0.570	
North Dakota	0.513	-32	0.521	-18	0.544	-8	0.661	
Nebraska	0.259	5	0.253	7	0.570	5	0.286	
New Hampshire	0.252	-74	0.240	-148	0.430	-36	0.288	
New Jersey	0.901	-7	0.788	-6	0.926	10	3.177	
New Mexico	0.832	-28	0.738	-12	0.969	0	0.489	
Nevada	2.129	-24	0.353	-12	0.425	-13	1.904	
New York	0.496	31	0.146	3	0.135	-17	0.418	
Ohio	0.263	63	0.675	54	0.468	3	0.337	
Oklahoma	0.301	-5	0.369	0	0.621	8	0.256	
Oregon	0.729	0	1.032	-2	1.692	-4	0.793	
Pennsylvania	0.411	-7	0.385	0	0.426	10	0.402	
Rhode Island	0.609	-9	0.546	-31	0.446	-2	1.699	
South Carolina	2.072	-3	2.157	-4	5.601	-39	0.429	
South Dakota	1.259	14	1.080	-2	1.089	2	5.050	
Tennessee	0.794	15	1.191	14	1.687	-11	0.600	
Texas	0.585	6	0.784	1	0.750	-71	0.706	
Utah	0.499	-98	0.716	-127	1.196	13	0.632	
Virginia	0.731	-10	0.396	6	0.864	9	0.676	
Vermont	0.142	59	0.300	-1	0.163	40	0.043	
Washington	0.608	-8	0.561	19	1.787	-1	0.782	
Wisconsin	0.842	6	1.028	25	3.921	8	0.850	
West Virginia	0.650	-6	0.547	2	1.042	7	0.291	
Wyoming	1.939	5	0.951	-15	1.126	25	0.395	
Average	1.696	-5	1.409	-6	1.483	-5	1.269	
Rest of the World								
Argentina	0.567	3	0.695	-17	0.904	-19	0.765	
Bolivia	0.339	-85	0.207	-117	0.182	-118	0.195	
Brazil	0.396	-4	0.405	11	0.578	4	0.493	
Chile	0.371	15	0.439	14	0.506	10	0.425	
Colombia	0.154	17	0.243	-46	0.164	-115	0.223	
Costa Rica	0.423	6	0.583	18	3.060	2	0.786	
Ecuador	0.156	-26	0.195	-99	0.234	-69	0.234	
Guatemala	0.557	-19	0.670	-31	0.815	-31	0.713	
Honduras	0.405	-8	0.381	-27	0.915	-41	0.541	
Mexico	0.766	16	0.939	11	1.100	11	1.110	
Nicaragua	0.091	-13	0.207	-24	1.340	-22	0.364	
Panama	0.550	-20	0.421	-4	0.451	-7	0.368	
Paraguay	0.535	28	0.877	-7	2.615	8	1.473	
Peru	0.507	33	0.103	26	1.630	16	0.515	
Uruguay	0.619	11	0.742	-13	0.899	-7	0.643	
Venezuela	0.610	-14	0.713	-49	0.890	-91	0.603	
Germany	0.379	5	0.613	15	2.131	14	1.570	
Italy	0.360	17	0.557	29	3.149	34	1.991	
Average	0.433	-6	0.500	-17	1.198	-24	0.723	

Table 4: NHS Regions: nowcasting mean absolute error per geographic location when using only death data and when using deaths and different low-latency data feeds.

Geographic Location	Deaths		Hospital		Twitter		Zoe App		111 Calls		111 Online	
	NEES	MAE % Diff	NEES	MAE % Diff	NEES	MAE % Diff	NEES	MAE % Diff	NEES	MAE % Diff	NEES	MAE % Diff
NHS Regions												
East of England	0.435	-13	0.419	-7	0.655	38	2.908	-15	0.820	-19	0.795	
London	0.878	-36	0.666	-7	1.163	131	3.150	-43	0.750	-47	0.754	
Midlands	0.635	-16	0.466	13	0.569	132	3.330	-19	0.418	-47	0.404	
North East and Yorkshire	0.753	5	1.188	-4	0.824	153	2.325	-16	0.860	-14	0.888	
North West	0.735	-1	0.756	17	1.408	129	3.285	-25	0.932	-25	0.934	
South East	0.652	-24	0.805	-3	1.255	126	4.390	8	1.018	6	0.957	
South West	0.545	-69	0.474	2	1.432	160	2.729	-8	1.617	-6	1.653	
Average	0.662	-22	0.682	2	1.044	124	3.160	-17	0.916	-22	0.912	