

Research



**Cite this article:** Jombart T *et al.* 2021 Real-time monitoring of COVID-19 dynamics using automated trend fitting and anomaly detection. *Phil. Trans. R. Soc. B* **376**: 20200266. <https://doi.org/10.1098/rstb.2020.0266>

Accepted: 15 February 2021

One contribution of 21 to a theme issue 'Modelling that shaped the early COVID-19 pandemic response in the UK'.

**Subject Areas:**

computational biology, health and disease and epidemiology, bioinformatics

**Keywords:**

ASMODEE, *trendbreaker*, surveillance, outbreak, algorithm, machine learning

**Authors for correspondence:**

Thibaut Jombart  
e-mail: thibautjombart@gmail.com  
Stéphane Ghozzi  
e-mail: stephane.ghozzi@helmholtz-hzi.de

<sup>†</sup>These authors contributed equally to the manuscript.

# Real-time monitoring of COVID-19 dynamics using automated trend fitting and anomaly detection

Thibaut Jombart<sup>1,2,3,†</sup>, Stéphane Ghozzi<sup>4,†</sup>, Dirk Schumacher<sup>5,10</sup>, Timothy J. Taylor<sup>1</sup>, Quentin J. Leclerc<sup>1</sup>, Mark Jit<sup>1,6</sup>, Stefan Flasche<sup>1</sup>, Felix Greaves<sup>7,8</sup>, Tom Ward<sup>7</sup>, Rosalind M. Eggo<sup>1</sup>, Emily Nightingale<sup>1</sup>, Sophie Meakin<sup>1</sup>, Oliver J. Brady<sup>1</sup>, Centre for Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Graham F. Medley<sup>1</sup>, Michael Höhle<sup>9</sup> and W. John Edmunds<sup>1</sup>

<sup>1</sup>Centre for Mathematical Modelling of Infectious Diseases, Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

<sup>2</sup>UK Public Health Rapid Support Team, London WC1E 7HT, UK

<sup>3</sup>MRC Centre for Global Infectious Disease Analysis, Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, London SW7 2DD, UK

<sup>4</sup>Department of Epidemiology, Helmholtz Centre for Infection Research, Brunswick, 38124, Braunschweig, Lower Saxony, Germany

<sup>5</sup>Department of Infectious Disease Epidemiology, Robert Koch-Institute, DE-13353 Berlin, Germany

<sup>6</sup>School of Public Health, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong SAR, People's Republic of China

<sup>7</sup>Department of Health and Social Care, Joint Biosecurity Centre, London SW1H 0EU, UK

<sup>8</sup>Department of Primary Care and Public Health, Imperial College London, London W6 8RP, UK

<sup>9</sup>Department of Mathematics, Stockholm University, 114 19 Stockholm, Sweden

<sup>10</sup>Unit for Medical Biometry and Statistics, Federal Institute for Quality Assurance and Transparency in Healthcare, Berlin, Germany

**ID** TJ, 0000-0003-2226-8692; SG, 0000-0002-3911-9573; QJL, 0000-0003-4761-001X;  
RME, 0000-0002-0362-6717; SM, 0000-0002-6385-2652

As several countries gradually release social distancing measures, rapid detection of new localized COVID-19 hotspots and subsequent intervention will be key to avoiding large-scale resurgence of transmission. We introduce ASMODEE (automatic selection of models and outlier detection for epidemics), a new tool for detecting sudden changes in COVID-19 incidence. Our approach relies on automatically selecting the best (fitting or predicting) model from a range of user-defined time series models, excluding the most recent data points, to characterize the main trend in an incidence. We then derive prediction intervals and classify data points outside this interval as outliers, which provides an objective criterion for identifying departures from previous trends. We also provide a method for selecting the optimal breakpoints, used to define how many recent data points are to be excluded from the trend fitting procedure. The analysis of simulated COVID-19 outbreaks suggests ASMODEE compares favourably with a state-of-art outbreak-detection algorithm while being simpler and more flexible. As such, our method could be of wider use for infectious disease surveillance. We illustrate ASMODEE using publicly available data of National Health Service (NHS) Pathways reporting potential COVID-19 cases in England at a fine spatial scale, showing that the method would have enabled the early detection of the flare-ups in Leicester and Blackburn with Darwen, two to three weeks before their respective lockdown. ASMODEE is implemented in the free R package *trendbreaker*.

This article is part of the theme issue 'Modelling that shaped the early COVID-19 pandemic response in the UK'.

## 1. Introduction

After a fast initial spread worldwide and large-scale epidemics in many affected countries, the trajectory of the COVID-19 pandemic is changing. In a number of severely affected countries, strong mitigation measures such as various forms of social distancing have slowed national epidemics and in many cases brought the epidemic close to control [1–3]. However, in the absence of widespread, long-lasting immunity through vaccination or natural infection [4–6], these respite are most likely temporary, and further relapses, in the form of localized outbreaks or nation-wide resurgence, remain highly likely and a very serious threat.

In the UK, a ‘lockdown’ was implemented on 23 March 2020, and gradually relaxed from the beginning of June 2020, by which point about 300 000 confirmed COVID-19 cases and 40 000 deaths had been reported [7]. Unfortunately, the risk of local flare-ups was illustrated soon after, as increased case incidence in Leicester resulted in the city being put under lockdown again on 29 June 2020 [8]. Similarly, increased restrictions were imposed in Blackburn on 9 August 2020.

In order to prevent large-scale relapses, localized COVID-19 hotspots (i.e. places with high levels of transmission) need to be detected as soon as cases occur and contained as early as possible. For such detection to be optimal, COVID-19 dynamics need to be monitored at a small spatial scale, requiring daily surveillance of multiple time series of case incidence, and prompt detection of ongoing increases. Disease surveillance algorithms have been designed for such purposes [9–12], although many of them are tailored to detecting either seasonal or point-source outbreaks and may be most effective when trained on years of weekly incidence data (e.g. Farrington algorithm [13,14]). However, careful implementation of such algorithms has proved useful as a backbone for setting up automated disease surveillance systems for endemic diseases [15].

Here, we introduce ASMODEE (automated selection of models and outlier detection for epidemics), an algorithm for detecting ongoing changes in COVID-19 incidence patterns. In order to characterize potentially very different dynamics in case incidence across a large number of locations, our approach implements a flexible time series framework using a variety of models including linear regression, generalized linear models (GLMs) or Bayesian regression. ASMODEE first identifies past temporal trends using automated model selection, and then uses outlier detection inspired by classical Shewhart control-charts to signal recent anomalous data points.

We used simulations to evaluate the potential of ASMODEE for detecting changes in incidence patterns. COVID-19 incidence dynamics were simulated using a branching-process model with realistic estimates of the time-varying reproduction number ( $R_t$ ) and serial interval, under four scenarios: steady state ( $R_t$  close to 1), relapse, lockdown and flare-up following low levels of transmission. For comparison, we also applied the modified Farrington algorithm [14], a standard method designed for the detection of point-source outbreaks and used in many public health institutions [9]. We computed a variety of scores such as the probability of detection, sensitivity and specificity for two configurations of ASMODEE.

We used our approach to design automated surveillance pipelines that monitor changes in potential COVID-19 cases reported through an online and telephone hotline used in

England, the National Health Service (NHS) Pathways system, which includes calls made to 111/999 as well as reports made through the 111-online system. We conducted the analysis at the level of Clinical Commissioning Groups (CCGs), small area divisions used for healthcare management in England’s NHS, with an average of 226 000 people each. One advantage of the NHS Pathways system is that reports occur with little delay, because no confirmatory diagnostic tests are involved. The downside is that this system suffers from the usual sensitivity and specificity issues of a syndromic surveillance system: some reported ‘potential cases’ will not be actual COVID-19 cases, and some actual cases will be unreported. We show that when applied to NHS 111/999 calls data, ASMODEE would have enabled the early detection of the flare-ups in Leicester and Blackburn with Darwen in June and July 2020, respectively. We propose that ASMODEE may be a useful, flexible complement to existing outbreak detection methods for designing disease surveillance pipelines, for COVID-19 and other diseases.

All source code necessary to run ASMODEE, analyse and visualize results, implement the NHS Pathways pipeline and reproduce the analyses is open and freely available under MIT license (see §2d).

## 2. Material and methods

Here, we first describe ASMODEE, an algorithm using automatic model selection to characterize past temporal trends, and identifying outliers in recent data points. We also provide details on different options for selecting models, and the optimal size of the outlier detection window in separate sections. We then outline different simulation scenarios and criteria for evaluating the performances of ASMODEE, with comparison to the Farrington flexible algorithm as a gold standard. Finally, we explain how ASMODEE was used for detecting step changes in potential COVID-19 cases in England using NHS Pathways data.

### (a) Automated selection of models and outlier detection for epidemics

#### (i) General algorithm

ASMODEE is designed for detecting recent departures/aberrations from past temporal trends in univariate time series. The response variable typically represents case counts, but it can readily accommodate other response variables such as incidence rates or case fatality ratios. We are interested in classifying outliers in the most recent time points of the available time series. Our approach can be broken down into the following steps:

1. Partition the analysed time series into two complementary time windows: a *calibration window*, excluding the most recent  $k$  data points; and a *prediction window*, made of the last  $k$  data points. The value of  $k$  is either fixed by the user, or its ‘optimal’ value can be determined using heuristics (see §2a(iii) below). The corresponding datasets are referred to as the *calibration*, and *prediction set*, respectively. Note that the calibration window should ideally capture a single, consistent trend. Accordingly, it does not need to include all historical data, and may retain e.g. the last six weeks of data.
2. Fit a range of user-defined time series models to the calibration set, and retain the best-predicting or best-fitting model (see ‘Selecting the ‘best’ model’ below); here we describe implementations with linear regressions, various GLMs (including

- Poisson and negative binomial models) and Bayesian regression models.
3. Derive prediction intervals at a given *alpha* two-sided threshold (defaulting to 5%) for all data points included in the prediction set. This threshold is equal to the expected proportion of data points that will be classified as outliers if all points followed the same temporal trend.
  4. Identify outliers as data points lying outside the prediction interval; the number of data points showing a marked *increase* from past trends in the prediction set can be used to design an alarm system. Similarly, a *decrease* in trend can be seen in data points lying below the lower bound of the prediction interval. Points within the interval can be considered as *normal*, i.e. belonging to the recent trend.

### (ii) Selecting the ‘best’ model

The literature on model selection offers a plethora of methods for selecting models fulfilling a given optimality criteria in ‘traditional’ statistics [16–19] as well as in machine learning [20–22]. To reflect some of this diversity, we have implemented two different approaches for model selection in ASMODEE. Note that both approaches use only the calibration window, and ignore data from the prediction window.

The first approach is *K*-fold cross-validation, in which the data are randomly broken into *K* partitions of roughly equal sizes (the ‘folds’) [23]. Each fold is in turn used for a round of cross-validation: the  $K - 1$  other folds are used as a *training* set, to fit the model, while the remaining fold is used as a *testing* set, to assess the quality of the model predictions. By default, we use  $K = N$ , where  $N$  is the number of data points in the calibration window. Discrepancies between model predictions and observed data of the testing set can be measured using various metrics. Here, we chose the root mean square error (RMSE) because of its wide use in statistical modelling and ease of computation. RMSE is summed over all rounds of cross-validation to characterize a given model. The model with the lowest total RMSE is retained as the ‘best’ model. Note that this metric gives equal weights to positive and negative residuals, and thus implicitly assumes that the magnitude of deviations from the model (under- and over-estimation) is similar. Other metrics could be used with low counts, where negative residuals are typically smaller (as case counts cannot be negative). When  $K < N$ , further accuracy of measurement can be gained by repeating the cross-validation procedure several times, each time using a different set of random folds, called repeated *K*-fold cross-validation.

This approach is likely to select models with good predictive ability (in the sense of minimum RMSE), but can be computationally intensive when evaluating multiple models over many datasets. As an alternative, we also implemented model selection using Akaike’s information criterion (AIC [16]). AIC is a standard for comparing the goodness-of-fit of models while accounting for their respective complexity, and can be used for all models for which a likelihood can readily be calculated. It is defined as

$$\text{AIC} = -2 \log(L) + 2P,$$

where  $L$  is the model likelihood (so that  $-2 \log(L)$  is the model’s deviance) and  $P$  is the number of parameters of the model. AIC is calculated for each model, and the model with the lowest AIC is retained as the ‘best fitting’ model. This procedure is fast but it does not guarantee that the retained model has the best predictive abilities. In practice though, results on simulated data suggest both approaches often give similar results (results not shown), in line with recent observations on incidence forecasting [24].

### (iii) Setting the prediction window ( $k$ )

The length of the prediction window  $k$  is an important parameter in ASMODEE, as it draws the line between the time period used

for estimating past trends and the recent time period over which we identify potential anomalies. For routine surveillance, a possible approach is to fix arbitrary values for both the calibration window and the prediction window, e.g. using seven weeks of daily incidence data for calibration and looking for outliers in the last week of data. These values may need to be adjusted over time to ensure optimal detection of changes in temporal trends, and to balance the need for the calibration window to contain sufficient data points to fit the most complex time series model considered. In addition, the value of  $k$  should be chosen so that (i) it exceeds delays previously identified for detecting trend changes and (ii) it remains within the time period over which forecasts are deemed reliable (typically no more than three weeks for fast-spreading diseases such as Ebola or COVID-19).

Besides this ‘manual tuning’ approach, heuristics can be used to compare results over different values of  $k$ . Various criteria can be devised to optimize *sensitivity* (ability to detect trend changes) or *specificity* (ability to identify data points following the trend), or any trade-off of these. Here, we maximize a simple score calculated for each value of  $k$  (up to a user-specified maximum value) as the sum of two components: (i) the number of non-outliers (points within the prediction interval of the ‘best’ model) in the calibration window, and (ii) the number of outliers in the prediction window. Therefore, this criterion tends to select a value of  $k$  such that past temporal trends can be well-defined (i.e. with most pre-break points within the prediction interval), while also retaining the ability to detect recent anomalies.

## (b) Application to simulated data

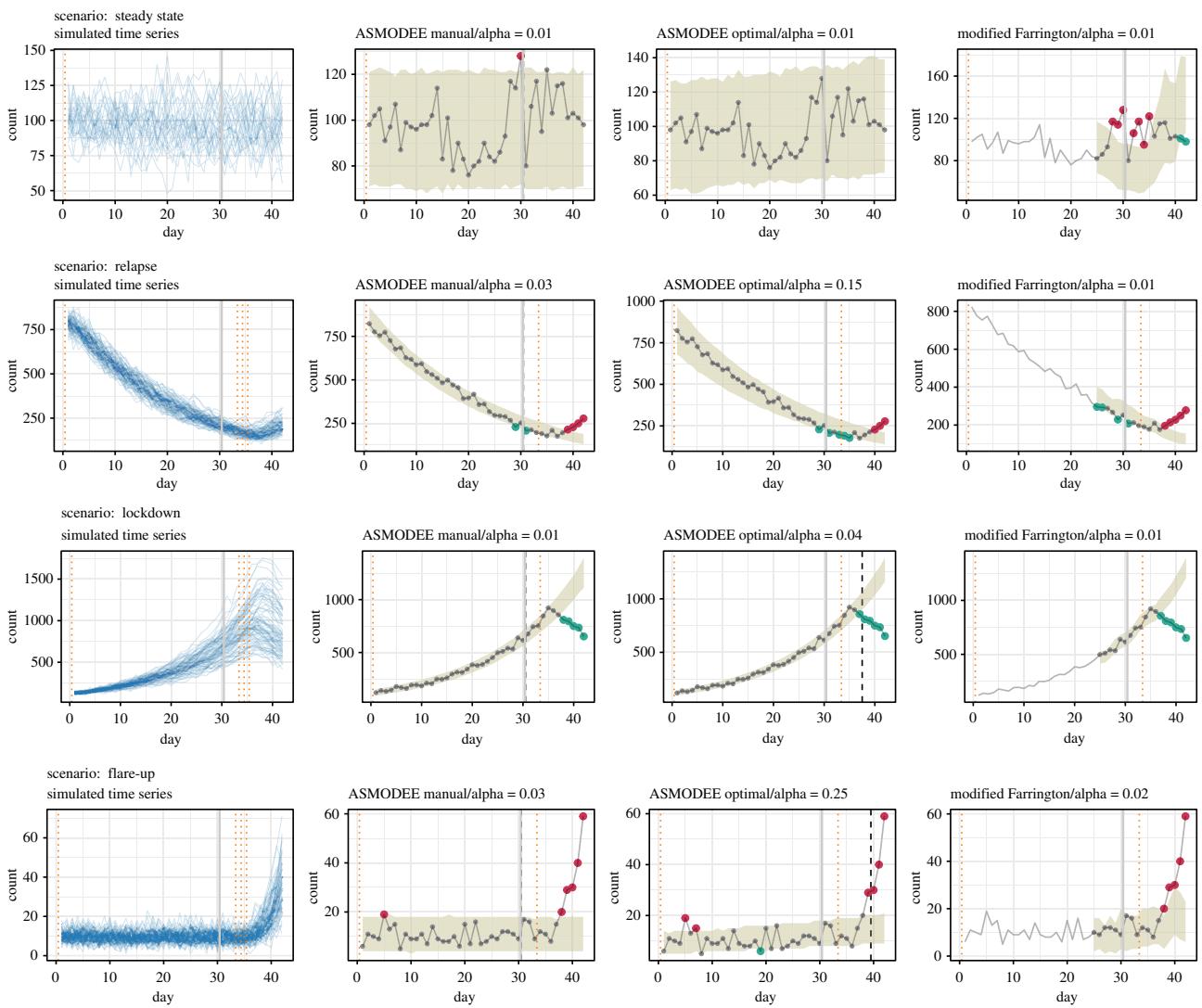
### (i) Simulated data

ASMODEE was evaluated using simulated COVID-19 incidence. All simulations used the *projections* package v. 0.5.1 [25], which implements branching-process epidemic simulations allowing for the reproduction number ( $R_t$ ) to vary by time periods. A full description of the model can be found in Jombart *et al.* [26]. Note that here and below,  $R_t$  refers to the expected number of secondary cases, and not the actual number of secondary infections drawn, so that even with  $R_t = 1$  some simulated epidemic trajectories might increase or decrease over time. Briefly, COVID-19 infections are simulated on a given day from a Poisson distribution, according to a force of infection determined by past cases, and distributions of the serial interval and  $R_t$ . Here, we used a gamma-distributed serial interval with mean 4.7 days and standard deviation 2.9 days [27].  $R_t$  was drawn from lognormal distributions with different means for different scenarios, and standard deviations 10% of the mean. We accounted for under-reporting by assuming that on average 10% of infections were reported (this value is broadly compatible with reported estimates [28]), so that the observed case count followed a binomial distribution with the number of trials, the true number of cases, and probability of success of 0.1.

We considered four scenarios that translated into different initial incidences and time-varying values of  $R_t$ :

- *Steady state*: 1000 cases per day (thus an average of 100 reported cases per day) with a mean  $R_t$  of 1.
- *Relapse*: 10 000 initial cases per day first declining with mean  $R_t$  of 0.8, followed by an increase with a mean  $R_t$  of 1.3.
- *Lockdown*: 1000 initial cases per day first increasing with a mean  $R_t$  of 1.3, followed by a decline with mean  $R_t$  of 0.8.
- *Flare-up*: 100 initial cases per day and a steady state with a mean  $R_t$  of 1, followed by a sudden increase with mean  $R_t$  of 2.5.

All simulations produced daily time series for 42 days. An *observation period*, over which the algorithms were evaluated, was set to be the last 12 days. This period, a bit longer than a



**Figure 1.** Simulations under four scenarios and illustration of the detection algorithm. Left column: the blue lines show time series generated under the four scenarios *steady state* (30 simulations), *relapse*, *lockdown* and *flare-up* (90 simulations each). The vertical solid grey lines mark the beginning of the observation period, the dotted orange lines the time at which a trend change occurs. Second to fourth columns: application on a single time series from each scenario of the configurations *ASMODEE manual*, *ASMODEE optimal* and the *modified Farrington* algorithms. The grey lines show the simulated time series, the ribbons show the prediction interval for the value of  $\alpha$  indicated, and the dashed black lines mark the end of the calibration period for *ASMODEE*. The red dots show data points categorized as *increase*, the green dots those categorized as *decrease*, the grey ones are considered *normal*. The values of  $\alpha$  were set so as to maximize the 6-day probability of detection. *modified Farrington* does not produce results for the first training period.

typical use case of 7 days, was chosen to offer enough variability and data points for a meaningful evaluation. Changes in  $R_t$  in the last three scenarios took place on days 34, 35 or 36. Simulations were replicated 30 times for each scenario and day of period change, resulting in 30 simulated time series for the *steady state* scenario and 90 for the others (figure 1).

## (ii) Outbreak detection

Here, we propose a general framework to evaluate COVID-19 outbreak detection algorithms. We used it to compare two configurations for *ASMODEE* together with the *modified Farrington* algorithm as a benchmark.

*ASMODEE* was applied to the simulated time series with one of three models (Poisson GLM with constant mean, and Poisson and negative binomial GLMs with log-linear trend in time) selected through minimization of the AIC. Model selection through cross-validation produced similar results (data not shown). We looked at either a fixed  $k$  of 12 days (*ASMODEE manual*) or an optimal  $k$  of at most 12 days (*ASMODEE optimal*). The rationale for setting  $k$  to (at most) the duration of the observation period is that a user would define a period in which they

are interested, implicitly assuming everything that happened before was not of direct interest and thus could be used to train the algorithms.

The *modified Farrington* algorithm [14] trains a statistical model on univariate count time series to derive bounds based on a parameter  $\alpha$ . It takes trend and seasonality into account and downweights anomalous counts in the training range. To avoid getting caught in recent trends and thus missing outbreaks, the last time steps are ignored during training. Conversely, low counts can be ignored to avoid uninteresting alarms. Lastly, it performs a one-step-ahead detection, i.e. it is retrained at every time step. It is typically used for weekly detection of point-source outbreaks with comparison to a 5-years baseline, but its parameters can be adapted to reflect other situations. Here, we use its implementation as *farringtonFlexible* in the R package *surveillance* [9] with three weeks training of a GLM with negative binomial family, weekly periodicity and a 3-days window half-size around the first day of each period, no limitation on low counts, and the last week ignored for training. Moreover, we further adapted the algorithm so that it always takes trend into account, thus removing the requirement that the estimated mean not be higher than all preceding counts.

The algorithms classified each day of a given time series as either *increase*, *decrease* or *normal* if it lay above the upper bound of the prediction interval, below its lower bound, or between the two, respectively. The application of the algorithms and the resulting classifications are illustrated for each scenario in figure 1.

### (iii) Evaluation

The three approaches (*ASMODEE manual*, *ASMODEE optimal* and *modified Farrington*) were evaluated in terms of their ability to correctly classify the last 12 days of the simulated epidemic curves. For a given simulation, each of the last 12 days could be classified as

- True positive (TP): increase (relapse and flare-up scenarios) or decrease (lockdown scenario) correctly detected by the method.
- True negative (TN): absence of change (steady-state scenario) correctly identified as such by the method.
- False positive (FP): days detected as changes (increase or decrease) in the steady state scenario.
- False negative (FN): days not correctly detected as increase (elapse and flare-up scenarios) or decrease (lockdown scenario).

This classification enables the calculation of standard performance scores (figure 2):

1. *sensitivity* TP/(TP + FN);
2. *specificity* TN/(TN + FP);
3. *balanced accuracy*,  $ba = (\text{sensitivity} + \text{specificity})/2$ ;
4. *precision* TP/(TP + FP); and *F1 score*,
5.  $F1 = (2 \times \text{precision} \times \text{sensitivity})/(\text{precision} + \text{sensitivity})$ .

The values these scores take depend on the two-sided threshold *alpha*, which was set to an optimal value as described below.

By contrast to these daily classifications, we also defined the *period detection*, which indicated whether a true change was detected over a given time period. This led to the definition of a *probability of detection within a time interval D* as the proportion of time series with the last period correctly detected within *D* days after the trend change. It thus represents a measure of *timeliness* of trend change detection. Figure 3 shows the probability of detection computed for each scenario, algorithm and selected values of the threshold parameter *alpha*.

*alpha* can be seen as an adjustment parameter: the smaller its value, the higher the proportion of days/periods classified as *normal*. For the purpose of comparing and evaluating algorithms, we varied *alpha* from 0.01 to 0.5. Its optimal value was defined as the one maximizing the probability of detection within 6 days, the longest interval available for all simulation runs (figure 3). In practice, *alpha* can be iteratively adjusted by the user to achieve the desired level of alert for a given surveillance system.

## (c) Automated pipelines for NHS pathways data

### (i) Implementation

As a real-world use case, we used the NHS Pathways data [29] reporting potential COVID-19 cases in England, broken down by Clinical Commissioning Groups (CCG). These data include all reports classified as ‘potential COVID-19 cases’ notified via calls to 111, 999 and 111-online systems [30]. These data are not confirmed cases, and are subject to unknown reporting biases. They likely include a substantial fraction of ‘false positives’ (cases classified as potential COVID-19 that are in fact due to other illness), as well as under-reporting (true COVID-19 cases not reported). Last, because these data are using self-reporting, it is likely that individual perceptions as well as ease of access to the reporting platforms impact the observed numbers.

We have discussed how these data can be interpreted and the associated caveats elsewhere and refer to these sources for more context [31]. Recent observations suggest that 111/999 calls may be more reflective of COVID-19 dynamics than 111-online [31]. Therefore, the subsequent analyses were based on these data only.

We have developed an automated pipeline to download these data, apply ASMODEE and present results. It has been implemented in a publicly accessible webpage (see §2d below) and was used in the illustration discussed in the next section.

### (ii) Illustration

The pipeline for NHS Pathways data was applied to all CCGs each day from 1 June to 10 August 2020. We highlighted the results for Leicester City and Blackburn as they experienced a large COVID-19 outbreak in the middle of June and July, respectively, with the first leading to the first local lockdown in the UK being imposed on 29 June.

We chose as parameters for ASMODEE a fixed *k* of 7 days and used AIC as the criterion for model selection. Here also, model selection through cross-validation produced identical results (data not shown). *alpha* was set to 0.05, low enough to avoid false alarms but high enough to rapidly detect changes. The five candidate models from which one was selected each day were: Poisson with constant mean; Poisson and negative binomial with the log-linear trend in time; the same with supplementary cofactor the day of the week (distinguishing weekends, Mondays and other days, in order to account for system closure over weekends and backlog effects); and the latter with a supplementary interaction of day and day of the week. Figure 4 shows the output of ASMODEE as applied each day, using a calibration window of 35 days. Moreover, we considered one metric to compare CCGs and highlight the peculiarity of the situations in Leicester and Blackburn with Darwen: the number of *increase* signals in a 7-day observation period (figure 5).

## (d) Availability

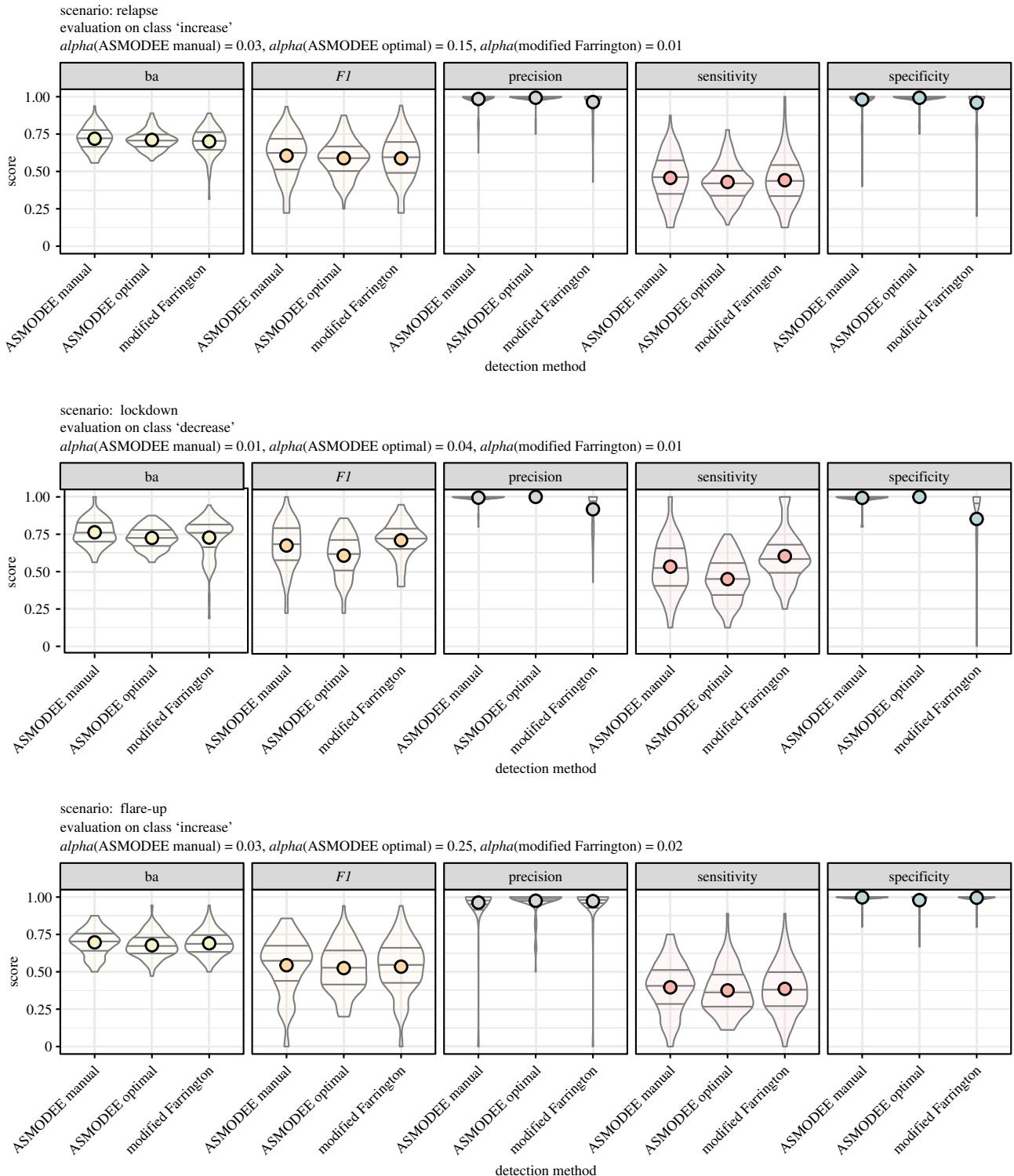
### (i) ASMODEE

ASMODEE is implemented in the new package *trendbreaker* for the R software [32], released under MIT license as part of the toolkit developed by the R Epidemics Consortium (RECON, <http://repi-demicsconsortium.org/>) for outbreak analytics [33]. While already functional and fully documented, this package is still under active development, and will be an integral part of the next generation of tools for handling, visualizing and modelling epidemic curves developed within RECON. It is available and documented at: <https://github.com/reconhub/trendbreaker>, and is scheduled for a release on CRAN by June 2021.

The package provides a unified interface for various modelling tools including linear models (function *lm*), Gaussian, Binomial, quasi-Binomial, Poisson, quasi-Poisson and Gamma GLMs (function *glm*), negative binomial GLM (function *MASS::glm.nb*) and Bayesian generalized nonlinear models using Stan [34,35] (function *brms::brm*). Various routines are implemented for automated model selection (function *select\_model*), outlier detection (function *detect\_outliers*) and the selection of *k* (function *detect\_changepoint*). The main function *asmodee* wraps these different tools and provides a full implementation of the method, including the two approaches for model selection (repeated *K*-fold cross-validation and AIC) and the automated selection of *k* using the scoring described in the previous section, plus plotting functions.

### (ii) Simulations

Scripts implementing the simulation of COVID-19 outbreaks and their analysis, as well as the adapted *farringtonFlexible* function are available at: <https://gitlab.com/stephaneghozzi/asmodee-trendbreaker-evaluation>.



**Figure 2.** Evaluation of the algorithm on simulated data. Distribution of scores over all simulation runs. Each row corresponds to a different scenario, each column to one score, computed for each algorithm. 'ba' stands for balanced accuracy, the average of sensitivity and specificity. The horizontal lines correspond to the 25th, 50th and 75th percentiles, the dots correspond to the average. The plot title indicates which class was considered positive for each scenario and which value of  $\alpha$  was considered optimal. (The trivial results for the steady state scenario, with a sensitivity very close to 1 and other scores undefined, are not shown.).

### (iii) NHS pathways analysis pipeline

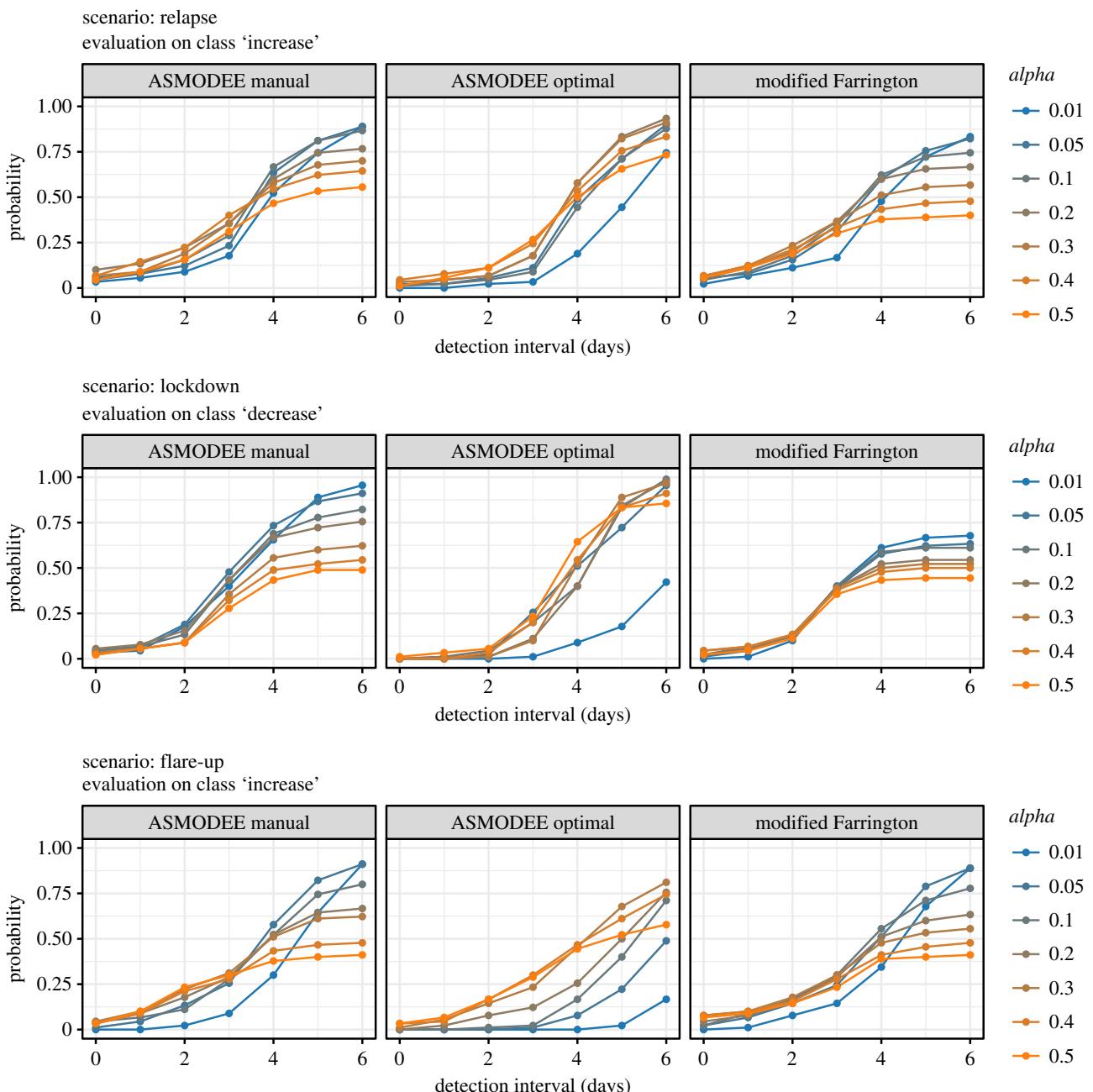
The data pipeline applying ASMODEE to NHS Pathways data is available at: [https://github.com/thibautjombart/nhs\\_pathways\\_monitoring](https://github.com/thibautjombart/nhs_pathways_monitoring). This pipeline is designed as a blogdown website [36] that automatically updates data and analyses daily using github actions. Daily data cleaning itself is done by the data pipeline hosted at: [https://github.com/qleclerc/nhs\\_pathways\\_report](https://github.com/qleclerc/nhs_pathways_report). The resulting website renders through Netlify at: <https://covid19-nhs-pathways-asmodee.netlify.app/>. Therefore, our pipeline is entirely free and can easily be replicated and adapted

to other data streams. The script for illustration on CCG data is available at: <https://gitlab.com/stephaneghozzi/asmodee-trend-breaker-evaluation>.

## 3. Results

### (a) Simulated data

ASMODEE detected anomalies in the three scenarios with trend changes in the simulated time series for both the



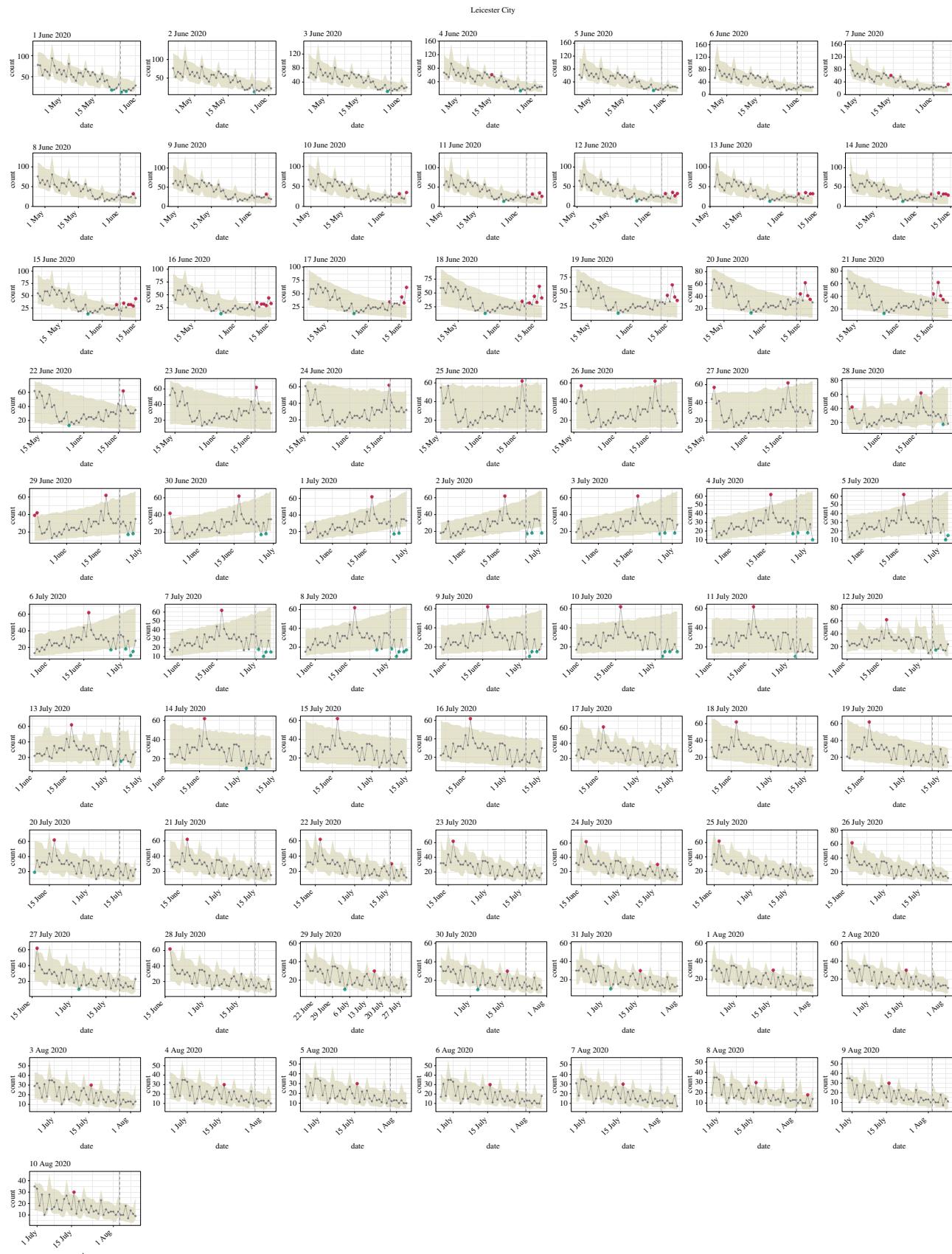
**Figure 3.** Delay to trend change detection. This figure illustrates the delay between the change in  $R_t$  and the first detected change (*increase* or *decrease*) by the different methods. The probability of detection on the y-axis is defined as the proportion of simulations where a trend change was detected, conditionally on no changes having been detected (false positive) before. Results are provided for selected values of the threshold parameter  $\alpha$ , represented in colour. (The trivial results for the *steady state* scenario are not shown.).

manual and optimal tuning (figure 1). Very few false alarms were generated in the *steady-state* scenario.

The five performance scores varied substantially within and across scenarios. All approaches showed almost perfect sensitivity in the *steady state* scenario, where typically more than 99% of ‘normal’ days are classified as such. Results were more heterogeneous in the *relapse* and *lockdown* scenarios where both ASMODEE configurations showed high precision and specificity, with varying sensitivity, translating into high overall balanced accuracy and  $F_1$  scores (figure 2). ASMODEE manual performed somewhat better in sensitivity and thus in  $F_1$  scores. Both ASMODEE approaches performed similarly in the *flare-up* scenario. Overall, the methods exhibited high specificity and precision, but with lower sensitivity than in the *lockdown* and *relapse* scenarios.

modified Farrington performed very similarly to ASMODEE manual, with slightly higher sensitivity and lower specificity in the *lockdown* scenario.

It should be noted that results are not only a reflection of the methods evaluated, but are also largely conditioned by the simulations and evaluation settings. Because we used a branching process to simulate COVID-19 outbreaks, changes in  $R_t$  are typically reflected after at least one serial interval (4–5 days on average). The resulting lag in the incidence time series, which can be seen in figure 1, imposes an upper bound on sensitivity: algorithms can only detect changes in transmissibility once these are reflected in the case counts. With the last trend period encompassing 7–9 days in our simulations, around half of the data points at best can be expected to be correctly detected as anomalies.

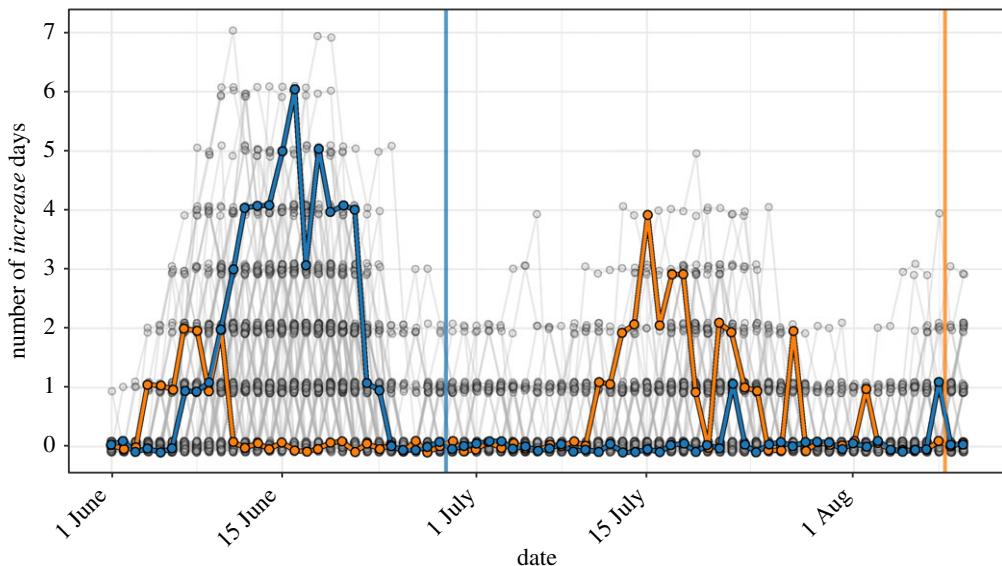


**Figure 4.** ASMODEE applied to the COVID-19 cases reported through the NHS Pathways (111/999 calls) for the CCG Leicester City during June and July and the first 10 days of August 2020. Each plot corresponds to a different reference date, with the algorithm applied to the preceding 42 days (including the reference day).  $k$  was fixed to 7 days,  $\alpha$  was set to 0.05. Considered temporal trends models included: Poisson GLM with constant mean, linear regression with linear trend in time, negative binomial GLM with log-linear trend in time, the same with a 'day of the week' effect (distinguishing weekends, Mondays and other days), and the latter with an additional interaction to allow for different slopes by day of the week. Automated model selection using AIC was used for all analyses.

The probability of detection within 6 days or less is less sensitive to the choice of the observation period, as it does not depend on its duration. As figure 3 shows,  $\alpha$  has a non-

monotonic effect on the detection performance, leading to an optimal value maximized after 6 days. In the three scenarios with a trend change, a 50% probability of detection for

how Leicester City and Blackburn with Darwen compare to other CCGs.



**Figure 5.** Leicester City and Blackburn with Darwen stand out among all CCGs from 1 June until 10 August 2020. Number of days classified as *increase* within 7 days of the reference date. Each dot represents one CCG on one reference day, the lines connect the dots corresponding to the same CCGs from one day to the next. Leicester City is highlighted in blue, Blackburn with Darwen in orange. A small jitter was applied for better readability. The blue and orange vertical lines indicate the dates at which increased restrictions were imposed in Leicester and Blackburn, respectively.

ASMODEE is typically reached only after 3–5 days, again due to the delay between changes in  $R_t$  and its actual reflection in the incidence time series. Overall, all three algorithms show good probabilities of detection, reaching almost 100% and both ASMODEE approaches outperforming *modified Farrington* in the *lockdown* scenarios.

### (b) Detection of flare-ups of COVID-19 cases in England

Looking at the particular case of Leicester in June 2020, we see that *increase* signals are consistently triggered for reference dates ranging from 10 to 22 June (figure 4). ASMODEE indicates one anomalous day from 7 to 9 June, two on 10 June and three on 11 June. Consistent with simulation results, ASMODEE exhibited good specificity, with very few *increase* signals generated before or after the relapse. Similar patterns were observed for Blackburn from 11 until 18 July 2020 (data not shown). The algorithm correctly flagged the weeks following the lockdown in Leicester on 29 June and the associated drop in incidence as *decreases* (figure 4).

To understand the context of the increase in the Leicester and Blackburn with Darwen CCGs, we analysed the time series of cases in all 136 CCGs that reported potential COVID-19 cases in June and July 2020. Leicester showed one of the longest periods of consecutive increase from 13 to 21 June (figure 5). Similarly, Blackburn stands out from 13 to 18 July. Taken together, these results show that ASMODEE could have detected the outbreaks in both regions early on.

## 4. Discussion

We have presented ASMODEE, a novel algorithm for trend change and anomaly detection. ASMODEE unifies two important features of an early warning system: gauging temporal trends in incidence time series and detecting recent anomalies, exceptionally higher case counts than expected based on recent data. The main advantage of the approach may be its flexibility

for modelling temporal trends. Rather than imposing a single model, it allows for a range of candidate models to be specified, and automatically selects the best-predicting (using repeated K-fold cross-validation) or fitting (using AIC) model. The current implementation provides support for linear regression, various different types of GLMs, as well as Bayesian regression implemented in Stan, and should as such provide ample flexibility for modelling different epidemic time series.

We tested ASMODEE using simulated COVID-19 epidemics, but this approach can be readily extended to other diseases and contexts. For instance, ASMODEE could be useful for detecting new hotspots of Ebola cases in large-scale epidemics such as the West-African Ebola outbreak [37,38] or the more recent one in North Kivu/Ituri, Democratic Republic of the Congo [26,39]. In a different context, it could complement existing systems for tracking nosocomial outbreaks of norovirus at a national level [40], to help identify wards or hospitals with a sudden increase of cases that would warrant prompt intervention. The fact that ASMODEE is model-agnostic and can include additional predictors beyond time grants the method flexibility to accommodate additional factors impacting incidence, such as changes in data quality, surveillance or testing capacity. For instance, changes in testing capacity could be accounted for by including the daily numbers of tests as a covariate in candidate models, so that any detected departure from past trends would indeed reflect changes in transmission, and not testing. Similarly, periods of disrupted surveillance, e.g. due to heightened insecurity [26], could be accounted for by including the presence or absence of disruption as an additional predictor.

We compared two variants of ASMODEE (with or without automatically setting the parameter  $k$ ) to a state-of-the-art outbreak detection algorithm in *modified Farrington*. Considering a number of metrics that reflect how a user would typically interpret detection results, we found that all three methods performed well overall, with very high specificity and varying sensitivity. Using a fixed value of  $k$  in ASMODEE (ASMODEE

*manual*) seemed to increase sensitivity in the considered scenarios. As this approach is also more computationally efficient, it may be a reasonable default for setting up surveillance pipelines. The *ASMODEE manual* configuration performed very closely to *modified Farrington*. This similarity may not be surprising: without seasonality and past, localized outbreaks, both approaches are based on similar statistical modelling and aberration detection approaches. Interestingly, when considering the delay to change detection, ASMODEE optimal showed more consistent, and better performances across different values of *alpha* than the other two methods in two scenarios out of three. Importantly, the low sensitivity exhibited in some results can be largely explained by the branching process used to simulate outbreaks: changes in  $R_t$  are only actually reflected in incidence time series after a lag of at least one serial interval. Further work may focus on evaluating ASMODEE using alternative models to simulate time series without such lag.

ASMODEE, as implemented in the package *trend-breaker*, is both much simpler than *modified Farrington* (requiring fewer parameters to be fine-tuned for specific contexts and diseases) and much more flexible (many kinds of statistical models can be easily integrated and automatically selected), as illustrated by the fact that we had to adapt *modified Farrington* to account for upward trends and produce meaningful results. However, it is not the case that ASMODEE would always be the approach of choice: *modified Farrington* has been developed for very specific use cases and practical needs, namely the longer-term surveillance of infections causing point-source outbreaks, such as food- or water-borne diseases. Nevertheless, we believe it to be a suitable benchmark for COVID-19 as, for lack of more specific standard tools, it would still have been the default method employed for COVID-19 surveillance.

Other aspects of the ASMODEE remain to be investigated. In our simulations, model selection using repeated K-fold cross-validation and AIC gave near identical results (results not shown). This may not be the case in other settings, and further work will be needed to investigate the automated model selection step. In particular, one may look for optimal fold sizes and numbers of repeats in the cross-validation procedure, and may also consider other goodness-of-fit statistics such as the Bayesian information criteria (BIC [17]). Further developments may also be considered for defining the optimal value of  $k$ , beyond the simple scoring system introduced here. We also note that our approach assumes that the calibration window contains a single trend, which can be adequately captured by at least one candidate model. In practice, this means temporal trends may be best characterized by selecting a subset of the most recent weeks of data. Further work may be devoted to optimizing this aspect as well, e.g. using absolute measures of fit to ensure the recent trends are indeed well captured.

The proposed method has a number of limitations. The main one relates to reporting delays, which are an intrinsic feature of most epidemiological data. ASMODEE takes incidence data on face value, i.e. without accounting for the potential effect of reporting delays, which typically cause incidence time series to artificially decrease over the last few days of data. While this limitation is not specific to ASMODEE, it will clearly hinder the method's capacity to

detect recent increases in case counts. A possible improvement of the method would be to characterize reporting delays, and then use augmented data/nowcasting to simulate the true underlying incidence, on which ASMODEE would be run. This approach would undoubtedly increase computational time, but would be easy to parallelize and most likely still fast enough to be used in daily surveillance of hundreds of geographical locations.

A second limitation of our approach is that ASMODEE does not consider the spatial spread of epidemics. While multiple locations can be analysed separately as illustrated in our analysis of NHS 111/999 data, the approach does not account for transmission across different locations. The general framework used in ASMODEE could in theory be extended to multivariate time series models incorporating spatial dependency, but the current implementation would need additional work to support such features. In practice, we expect this may only be a substantial limitation when very good data on patient locations and movement are available. An alternative worth exploring would be to run ASMODEE on spatially smoothed data [41] to reduce noise in data typically observed in small geographical areas with low case numbers and therefore facilitate anomaly detection.

We expect ASMODEE will be most useful for surveying potentially large numbers of incidence time series, e.g. to detect flare-ups of COVID-19 in small geographical units and/or specific population demographics. It may be best used in conjunction with human judgement rather than as a purely automated algorithm. For instance, ASMODEE could rank incidence times series according to their respective numbers of 'increase' days, and then the highest-ranked series could be examined by epidemiologists to decide whether further investigation is warranted. As such, ASMODEE could form the basis for a daily COVID-19 surveillance system, and be regularly refined, e.g. changing the duration of the detection window (parameter  $k$ ) or the *alpha* threshold to meet required alert levels. The automated pipeline we have developed for NHS 111/999 data shows that such surveillance systems can be built using free, open-source tools, and readily automated for daily updates.

It is our hope that ASMODEE will form a useful complement to existing point-source outbreak detection methods such as the *modified Farrington* algorithm [14] and scan statistics [42]. We believe its inherent simplicity and flexibility for modelling time series, together with its availability in the free, open-source R package *trendbreaker*, will facilitate further improvements and adoption by the community, and its broader use for disease surveillance for COVID-19 outbreaks and beyond.

## 5. Context: real-time monitoring of COVID-19 dynamics using automated trend fitting and anomaly detection

The focus of the work of Scientific Pandemic Influenza Group on Modelling (SPI-M) has evolved as the COVID-19 epidemic unfolded in the UK. Before the first wave peaked, most questions revolved around the speed at which the epidemic was growing and spreading geographically, and estimates of future bed occupancy to assess whether, or when, hospital capacity might be exceeded.

After the first wave, as cases were globally falling in the UK, interest started shifting to detecting local flare-ups of cases, and more generally any new trend in COVID-19 cases that may warrant tailored interventions. This type of question generally falls within the realm of *outbreak detection* methods that have been developed over the years for routine disease surveillance. Unfortunately, local dynamics of COVID-19 cases posed a number of challenges that existing methods struggled to accommodate, ranging from weekly reporting artefacts (e.g. cases going down on weekends and up on Mondays) to changes in testing intensity or age-specific mortality.

ASMODEE was designed to overcome these issues. When developing the method, we followed three guiding principles. First, we wanted to avoid estimating changes in the reproduction number ( $R$ ), which was already the focus of many other groups in SPI-M. The underlying idea was that showing significant changes in  $R$  may require weeks of data, and we wanted our approach to be able to detect trend changes as soon as they happened. This is why we preferred a strategy relying on characterizing past trends, and identifying recent departures from these trends using an outlier detection approach. Second, we designed our approach to be more flexible than existing standard surveillance tools. Indeed, it is able to use a variety of models, and to account for confounders such as testing or the age distribution of cases by including them as covariates in candidate models. We found that using many (hundreds of) simple candidate models, retaining the best one for each locality, was a much more efficient approach for COVID-19 surveillance than attempting to design a one-size-fits-all model.

Last, we wanted this approach to become a new standard in infectious disease surveillance. We spent a vast amount of time designing free, open-source software packages that can be used for other disease surveillance tasks beyond COVID-19. These tools include *incidence2*, *trending*, *trendeval* and *trendbreaker*, which are now part of the base of the reconvert (https://www.reconvert.org/), a new suite of tools for outbreak analysis. ASMODEE has been deployed in the UK for detecting trend changes in potential COVID-19 cases reported through calls to 111 and 999 and 111-online. While it could have no doubt been applied to other data sources, we favoured transparency and reproducibility, and used these data sources as they were all publicly available early on. Since then, ASMODEE has been used outside the UK, for COVID-19 surveillance in Germany at the Robert Koch Institute and at the World Health Organization, where it is used for monitoring changes in COVID-19 trends in countries across the world.

**Data accessibility.** Full details of data accessibility are provided within the text.

**Authors' contributions.** In alphabetic order: S.G., T.J., D.S. developed the methodology. S.G., T.J., D.S., T.J.T. contributed code. S.G., T.J. performed the analyses. S.G., T.J.T. reviewed code. S.G., T.J. wrote the first draft of the manuscript. O.J.B., W.J.E., R.M.E., S.F., F.G., S.G., M.H., T.J., M.J., Q.J.L., G.F.M., S.M., E.N., D.S., T.W., T.J.T. contributed to the manuscript. CMMID COVID-19 Working Group gave input on the method, contributed data and provided elements of discussion. The following authors were part of the Centre for Mathematical Modelling of Infectious Disease 2019-nCoV working group: Arminder K. Deol, Kathleen O'Reilly, Charlie Diamond, David Simons, Petra Klepac, Christopher I. Jarvis, Sebastian Funk, Nicholas G. Davies, Yung-Wai Desmond Chan, Damien C. Tully, Nikos I. Bosse, Simon R. Procter, Kaja Abbas, Amy Gimma, Jon

C. Emery, Billy J. Quilty, Kevin van Zandvoort, Stéphane Hué, Rosanna C. Barnard, Timothy W. Russell, Sam Abbott, Kiesha Prem, Adam J. Kucharski, Akira Endo, Fiona Yueqian Sun, James W. Rudge, Katharine Sherratt, Yang Liu, Katherine E. Atkins, Rein M. G. J. Houben, Matthew Quaife, Joel Hellewell, Gwenan M. Knight, Carl A. B. Pearson, Georgia R. Gore-Langton, Anna M. Foss, Megan Auzenbergs, Alicia Rosello, Samuel Clifford, C. Julian Villabona-Arenas, Hamish P. Gibbs, Alicia Showering, Jack Williams, Frank G. Sandman, Naomi R. Waterlow. Each contributed in processing, cleaning and interpretation of data, interpreted findings, contributed to the manuscript and approved the work for publication.

**Competing interests.** We declare we have no competing interests.

**Funding.** The named authors (O.J.B., W.J.E., R.M.E., T.J., M.J., S.M., E.N., T.J.T.) had the following sources of funding: O.J.B. was funded by a Sir Henry Wellcome Fellowship funded by the Wellcome Trust (grant no. 206471/Z/17/Z). R.M.E. receives funding from HDR UK (grant no. MR/S003975/1). S.F. is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (grant no. 208812/Z/17/Z). T.J. receives funding from the Global Challenges Research Fund (GCRF) project 'RECAP' managed through RCUK and ESRC (ES/P010873/1), the UK Public Health Rapid Support Team funded by the United Kingdom Department of Health and Social Care and from the National Institute for Health Research (NIHR)—Health Protection Research Unit for Modelling Methodology. M.J. receives funding from the Bill and Melinda Gates foundation (grant no.: INV-003174) and the NIHR (grant no.: 16/137/109 and HPRU-2012-10096). E.N. receives funding from the Bill and Melinda Gates Foundation (grant no.: OPP1183986). M.J. and W.J.E. receive funding from European Commission project EpiPose (grant no. 101003688). S.M. receives funding from the Wellcome Trust (grant no. 210758/Z/18/Z). S.F. receives funding from the Wellcome Trust (grant no. 208812/Z/17/Z). T.J. and T.J.T. receive funding from the MRC (grant no. MC\_PC\_19065). The following funding sources are acknowledged as providing funding for the working group authors. BBSRC LIDP (BB/M009513/1: D.S.). This research was partly funded by the Bill & Melinda Gates Foundation (INV-001754: M.Q.; INV-003174: K.P., M.J., Y.L.; NTD Modelling Consortium OPP1184344: C.A.B.P., G.F.M.; OPP1180644: S.R.P.; OPP1191821: K.O'R., M.A.). BMGF (OPP1157270: K.A.). DFID/Wellcome Trust (Epidemic Preparedness Coronavirus research programme 221303/Z/20/Z: K.v.Z.). DTRA (HDTRA1-18-1-0051: J.W.R.). Elrha R2HC/UK DFID/Wellcome Trust/this research was partly funded by the National Institute for Health Research (NIHR) using UK aid from the UK Government to support global health research. ERC Starting Grant (#757699: J.C.E., M.Q., R.M.G.H.). This project has received funding from the European Union's Horizon 2020 research and innovation programme—project EpiPose (101003688: K.P., M.J., P.K., R.C.B., Y.L.). HDR UK (MR/S003975/1: R.M.E.). NIHR (16/136/46: B.J.Q.; 16/137/109: B.J.Q., C.D., F.Y.S., M.J., Y.L.); Health Protection Research Unit for Immunisation NIHR200929: N.G.D., F.G.S.; Health Protection Research Unit for Modelling Methodology HPRU-2012-10096: T.J., F.G.S.; NIHR200929: M.J.; PR-OD-1017-20002: A.R.). UK DHSC/UK Aid/NIHR (ITCRZ 03010: H.P.G.). UK MRC (LID DTP MR/N013638/1: G.R.G.-L., Q.J.L., N.R.W.; MC\_PC\_19065: A.G., N.G.D., R.M.E., S.C., T.J., Y.L.; MR/P014658/1: G.M.K.). Authors of this research receive funding from UK Public Health Rapid Support Team funded by the United Kingdom Department of Health and Social Care (T.J.). Wellcome Trust (206250/Z/17/Z: A.J.K., T.W.R.; 206471/Z/17/Z: O.J.B.; 208812/Z/17/Z: S.C., S.F.; 210758/Z/18/Z: S.F., S.A., J.H.). No funding (A.K.D., A.M.F., C.J.V.-A., D.C.T., S.H., Y.-W.D.C.). The UK Public Health Rapid Support Team is funded by UK aid from the Department of Health and Social Care and is jointly run by Public Health England and the London School of Hygiene and Tropical Medicine. The University of Oxford and King's College London are academic partners.

**Disclaimer.** The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR or the UK Department of Health and Social Care (K.v.Z.).

## References

- Lai S *et al.* 2020 Effect of non-pharmaceutical interventions for containing the COVID-19 outbreak in China. *medRxiv*. (doi:10.1101/2020.03.03.20029843)
- Courtemanche C, Garuccio J, Le A, Pinkston J, Yelowitz A. 2020 Strong social distancing measures in the United States reduced the COVID-19 growth rate: study evaluates the impact of social distancing measures on the growth rate of confirmed COVID-19 cases across the United States. *Health Aff.* **10**, 1377. (doi:10.4324/9781003141402-20)
- Davies NG *et al.* 2020 Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *Lancet Public Health* **5**, e375–e385. (doi:10.1016/S2468-2667(20)30133-X)
- Long Q-X *et al.* 2020 Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat. Med.* **26**, 1200–1204. (doi:10.1038/s41591-020-0965-6)
- Grifoni A *et al.* 2020 Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell* **181**, 1489–1501. (doi:10.1016/j.cell.2020.05.015)
- Baum A *et al.* 2020 Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science* **369**, 1014–1018. (doi:10.1126/science.abd0831)
- Home - Office for National Statistics. 2021 See <https://www.ons.gov.uk/>.
- Department of Health and Social Care. 2020 Leicester lockdown: what you can and cannot do. See <https://www.gov.uk/guidance/leicester-lockdown-what-you-can-and-cannot-do>.
- Salmon M, Schumacher D, Höhle M. 2016 Monitoring count time series in R: aberration detection in Public Health Surveillance. *J. Stat. Softw.* **70**, 1–35. (doi:10.18637/jss.v070.i10)
- Yuan M, Boston-Fisher N, Luo Y, Verma A, Buckeridge DL. 2019 A systematic review of aberration detection algorithms used in public health surveillance. *J. Biomed. Inform.* **94**, 103181. (doi:10.1016/j.jbi.2019.103181)
- Morbey R *et al.* 2018 Comparison of statistical algorithms for syndromic surveillance aberration detection. *Online J. Public Health Inform.* **10**, e4. (doi:10.5210/ojphi.v10i1.8302)
- Sonesson C, Bock D. 2003 A review and discussion of prospective statistical surveillance in public health. *J. R. Stat. Soc. Ser. A Stat. Soc.* **166**, 5–21. (doi:10.1111/1467-985X)
- Farrington CP, Andrews NJ, Beale AD, Catchpole MA. 1996 A statistical algorithm for the early detection of outbreaks of infectious disease. *J. R. Stat. Soc. Ser. A Stat. Soc.* **159**, 547–563. (doi:10.2307/2983331)
- Noufaily A *et al.* 2013 An improved algorithm for outbreak detection in multiple surveillance systems. *Stat. Med.* **32**, 1206–1222. (doi:10.1002/sim.5595)
- Salmon M *et al.* 2016 A system for automated outbreak detection of communicable diseases in Germany. *Euro Surveill.* **21**, 30180. (doi:10.2807/1560-7917.ES.2016.21.13.30180)
- Akaike H. 1974 A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**, 716–723. (doi:10.1109/TAC.1974.1100705)
- Schwarz G. 1978 Estimating the dimension of a model. *Ann. Stat.* **6**, 461–464. (doi:10.1214/aos/1176344136)
- Kullback S, Leibler RA. 1951 On information and sufficiency. *Ann. Math. Stat.* **22**, 79–86. (doi:10.1214/aoms/1177729694)
- McCullagh P, Nelder JA. 1989 *Generalized linear models*, 2nd edn. Boca Raton, FL: CRC Press.
- He X, Zhao K, Chu X. 2021 AutoML: A survey of the state-of-the-art. *Knowledge-Based Systems* **212**, 10662. (doi:10.1016/j.knosys.2020.106622)
- Kotthoff L, Thornton C, Hoos HH, Hutter F, Leyton-Brown K. 2017 Auto-WEKA 2.0: automatic model selection and hyperparameter optimization in WEKA. *J. Mach. Learn. Res.* **18**, 826–830. (doi:10.1007/978-3-030-05318-5\_4)
- Cawley GC, Talbot NLC. 2010 On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.* **11**, 2079–2107.
- Rodriguez JD, Perez A, Lozano JA. 2010 Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**, 569–575. (doi:10.1109/TPAMI.2009.187)
- Nightingale ES *et al.* 2020 A spatio-temporal approach to short-term prediction of visceral leishmaniasis diagnoses in India. *PLoS Negl. Trop. Dis.* **14**, e0008422. (doi:10.1371/journal.pntd.0008422)
- Jombart T *et al.* 2020 Projection of future incidence. R package version 0.5.1. (doi:10.5281/zenodo.3923626)
- Jombart T *et al.* 2020 The cost of insecurity: from flare-up to control of a major Ebola virus disease hotspot during the outbreak in the Democratic Republic of the Congo, 2019. *Eurosurveillance* **25**, 1900735. (doi:10.2807/1560-7917.ES.2020.25.2.1900735)
- Nishiura H, Linton NM, Akhmetzhanov AR. 2020 Serial interval of novel coronavirus (COVID-19) infections. *Int. J. Infect. Dis.* **93**, 284–286. (doi:10.1016/j.ijid.2020.02.060)
- Hauser A *et al.* 2020 Estimation of SARS-CoV-2 mortality during the early stages of an epidemic: a modeling study in Hubei, China, and six regions in Europe. *PLoS Med.* **17**, e1003189. (doi:10.1371/journal.pmed.1003189)
- NHS Pathways coronavirus triage. 2021 See <https://digital.nhs.uk/dashboards/nhs-pathways>.
- NHS 111 online. See <https://111.nhs.uk/>.
- Leclerc QJ *et al.* 2020 Analysis of temporal trends in potential COVID-19 cases reported through NHS Pathways England. *Sci. Rep.* **11**, 7106. (doi:10.1038/s41598-021-86266-3)
- R Core Team. 2020 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing
- Polonsky JA *et al.* 2019 Outbreak analytics: a developing data science for informing the response to emerging pathogens. *Phil. Trans. R. Soc. B* **374**, 20180276. (doi:10.1098/rstb.2018.0276)
- Gelman A, Lee D, Guo J. 2015 Stan: a probabilistic programming language for Bayesian inference and optimization. *J. Educ. Behav. Stat.* **40**, 530–543. (doi:10.3102/1076998615606113)
- Bürkner P-C. 2017 brms: an R package for bayesian generalized linear mixed models using Stan. *J. Stat. Softw.* **80** (1). (doi:10.18637/jss.v080.i01)
- Xie Y, Hill AP, Thomas A. 2017 *Blogdown: creating websites with R markdown*. Boca Raton, FL: CRC Press.
- WHO Ebola Response Team. 2014 Ebola virus disease in West Africa—the first 9 months of the epidemic and forward projections. *N. Engl. J. Med.* **371**, 1481–1495. (doi:10.1056/NEJMoa1411100)
- WHO Ebola Response Team. 2015 West African Ebola epidemic after one year—slowing but not yet under control. *N. Engl. J. Med.* **372**, 584–587. (doi:10.1056/NEJMrc1414992)
- Kalenga Ol, Sparrow A, Nguyen V-K, Lucey D, Ghebreyesus TA. 2019 The ongoing Ebola epidemic in the Democratic Republic of Congo, 2018–2019. *N. Engl. J. Med.* **381**, 373–383.
- Norovirus Reporting. <https://hnors.phe.gov.uk/>.
- Elliott P, Wartenberg D. 2004 Spatial epidemiology: current approaches and future challenges. *Environ. Health Perspect.* **112**, 998–1006. (doi:10.1289/ehp.6735)
- Kulldorff M, Heffernan R, Hartman J, Assunção R, Mostashari F. 2005 A space–time permutation scan statistic for disease outbreak detection. *PLoS Med.* **2**, e59. (doi:10.1371/journal.pmed.0020059)