



OPEN

Decision support for the quickest detection of critical COVID-19 phases

Paolo Braca^{1✉}, Domenico Gaglione¹, Stefano Marano², Leonardo M. Millefiori¹, Peter Willett³ & Krishna Pattipati³

During the course of an epidemic, one of the most challenging tasks for authorities is to decide what kind of restrictive measures to introduce and when these should be enforced. In order to take informed decisions in a fully rational manner, the onset of a critical regime, characterized by an exponential growth of the contagion, must be identified *as quickly as possible*. Providing rigorous quantitative tools to detect such an onset represents an important contribution from the scientific community to proactively support the political decision makers. In this paper, leveraging the quickest detection theory, we propose a mathematical model of the COVID-19 pandemic evolution and develop decision tools to rapidly detect the passage from a controlled regime to a critical one. A new sequential test—referred to as MAST (mean-agnostic sequential test)—is presented, and demonstrated on publicly available COVID-19 infection data from different countries. Then, the performance of MAST is investigated for the second pandemic wave, showing an effective trade-off between average decision delay Δ and risk R , where R is inversely proportional to the time required to declare the need to take unnecessary restrictive measures. To quantify risk, in this paper we adopt as its proxy the average occurrence rate of false alarms, in that a false alarm risks unnecessary social and economic disruption. Ideally, the decision mechanism should react as quick as possible for a given level of risk. We find that all the countries share the same behaviour in terms of quickest detection, specifically the risk scales exponentially with the delay, $R \sim \exp(-\omega\Delta)$, where ω depends on the specific nation. For a reasonably small risk level, say, one possibility in ten thousand (i.e., unmotivated implementation of countermeasures every 27 years, on the average), the proposed algorithm detects the onset of the critical regime with delay between a few days to 3 weeks, much earlier than when the exponential growth becomes evident. Strictly from the quickest-detection perspective adopted in this paper, it turns out that countermeasures against the second epidemic wave have not always been taken in a timely manner. The developed tool can be used to support decisions at different geographic scales (regions, cities, local areas, etc.), levels of risk, instantiations of controlled/critical regime, and is general enough to be applied to different pandemic time-series. Additional analysis and applications of MAST are made available on a dedicated website.

With more than 57 million cases worldwide and over one and 1.36 million deaths as of November 20, 2020, the outbreak of coronavirus disease (COVID-19)¹, is undeniably one of the worst global crises since World War II. In March 2020, the exponential increase of individuals needing hospitalization in intensive care units, combined with the lack of effective cures and vaccines, pushed many governments to take extraordinary measures aimed at “flattening the curve” of infections^{2–4}. The adopted measures included the limitation of mobility and social activities, closure of schools, universities, shops, factories, and so forth, up to the extreme act of national lockdowns. Evidence that such measures achieved a reduction of the rate of new infections are gradually appearing in the scientific literature^{5–7}.

These measures contributed to keep the spread of COVID-19 under control for some time, but we are now, in November 2020, experiencing the onset of a new exponential growth of confirmed cases, the “second wave,” with severe risks for personal health and healthcare systems under severe stress. Governments and authorities

¹Research Department, NATO STO Centre for Maritime Research and Experimentation, 19126 La Spezia, Italy. ²Department of Information & Electrical Engineering and Applied Mathematics (DIEM), University of Salerno, 84084 Fisciano, SA, Italy. ³Department of Electrical and Computer Engineering, University of Connecticut, Storrs 06269, USA. ✉email: paolo.braca@cmre.nato.int

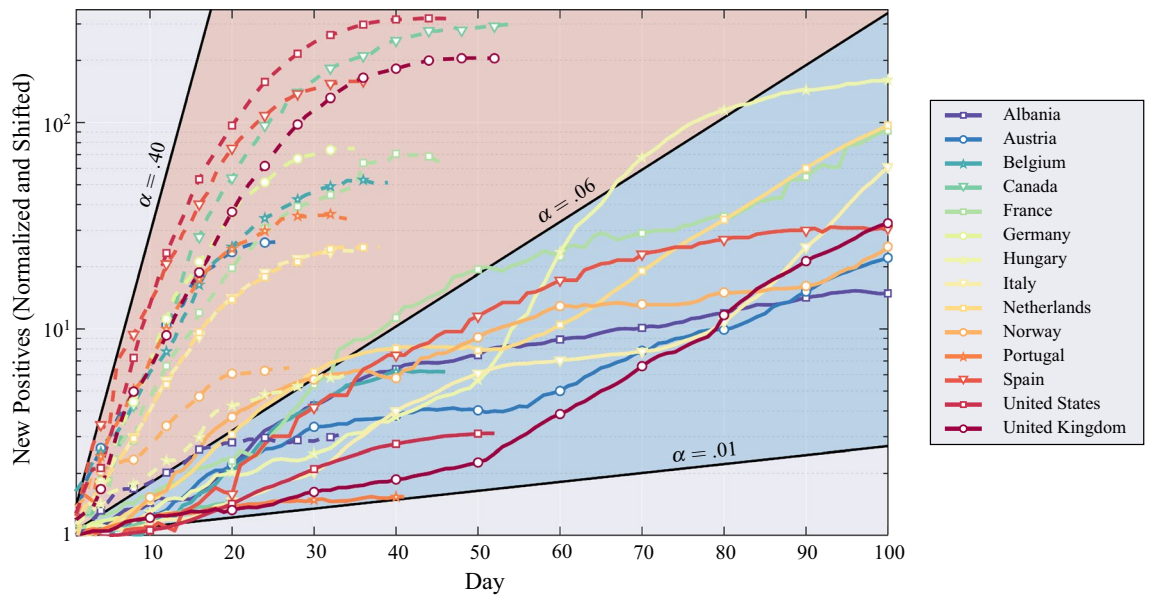


Figure 1. Number of COVID-19 daily new positives in the first and the second wave of the pandemic for several nations. Each curve is normalized by the initial value and shifted to the same initial time. The initial point of the first wave for a specific nation is when the first positives are reported, while the initial point of the second wave is when the daily positives start to increase again—growth rate larger than one. The first wave, upper region (shaded in red), was more aggressive than the second wave, mostly confined to the lower region (shaded in blue). These two regions are delimited by different values of α in Eq. (1). Curves of each nation are dashed if related to the first wave otherwise solid if related to the second wave.

are facing again the difficult task of deciding if and when new containment measures may be needed. In absence of limitations to mobility and social activities, the pandemic spreads exponentially in time, so that any delay in applying restrictions may lead to severe consequences. On the other hand, accounting for the social and economic impact of the possible countermeasures, already observed in the first half of 2020^{8–11}, restrictions should be taken only if and when it is strictly necessary. Managing the trade-off between these contrasting requirements is extremely challenging.

To address this challenge, we leverage sequential detection theory^{12–14}, and specifically *quickest detection* schemes^{15–17} to propose a rigorous methodology aimed at identifying *as quickly as possible* the onset of an exponential growth of the pandemic evolution. The proposed procedure—referred to as MAST (mean-agnostic sequential test)—is designed to minimize the average time to detect a change in regime^{15–17}: from a situation in which the pandemic is under control (\mathcal{H}_0 regime) to the onset of an exponential growth (\mathcal{H}_1 regime).

Quickest detection theory has a long history¹⁸, and has been successfully applied in several fields, including quality control, climate modeling, remote sensing, financial analysis, image analysis, security, signal and speech processing, and biomedical applications^{16,17}. In the context of public health surveillance, where the timely detection of various types of adverse health events is crucial, quickest detection techniques have found several applications¹⁹, e.g., to reveal the onset and the peak of the epidemic period²⁰ or that the peak is over²¹.

The approach pursued in this article is substantially different from most of the epidemiological models, based for instance on stochastic evolution of epidemic compartments^{5,7,22–24} and metapopulation networks^{25,26}, where the goal is to predict the mid/long-term behavior of the outbreak. For instance, in stochastic compartmental models, given an initial condition, the epidemic can have two outcomes: the number of infected individuals can increase, in which case we have a major outbreak, or decrease. The probability of a major outbreak can be computed²⁷, but it is of limited use in taking timely on-line decisions.

The importance of taking a decision as quickly as possible in an epidemic scenario can be understood by looking at the curve of daily cases of COVID-19 infection, reported in Fig. 1 for several nations²⁸. In the same figure we also report the deterministic curves of daily cases, with an exponential growth described by the following equation

$$p_{n+1} = p_n(1 + \alpha) = p_1(1 + \alpha)^n, \quad n = 1, 2, \dots, \quad (1)$$

where n is the time index (day), $1 + \alpha$ is the growth rate, and $\alpha > 0$, which corresponds to the \mathcal{H}_1 regime of a major outbreak. This exponential behavior is equivalent to a recently-proposed disease-transmission model²⁹, and the *reproduction number* defined therein is equivalent to the growth rate. All the curves in Fig. 1 are normalized to the initial value p_1 and shifted to the same initial time. Figure 1 shows that first wave was noticeably more aggressive than the second one in terms of growth rate; specifically, in the first wave α was varied by country but ranged between 0.06 and 0.40, while in the second wave it is between 0.01 and 0.06.

In the latter scenario, any delay in revealing the onset of an exponential growth phase and the consequent implementation of containment measures produces a costly exponential increase of the number of new cases. In

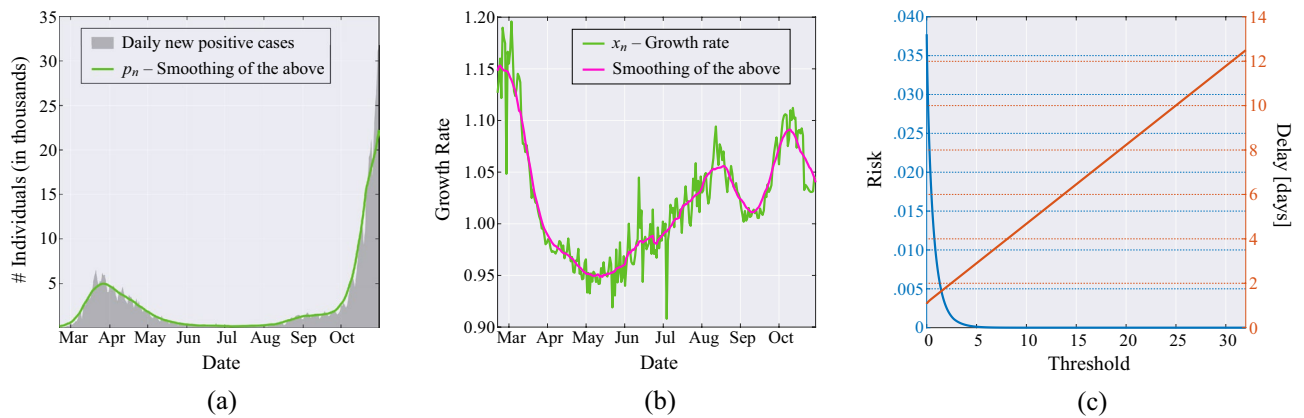


Figure 2. (a) Daily new positive individuals in Italy since February 21, 2020, and its moving average obtained with a window of 21 days (green line). (b) Growth rate of the epidemic computed from the averaged daily new positive cases (green line), and its time-varying mean obtained through a moving average that uses a window of 21 days (magenta line). (c) MAST performance, in terms of risk (left axis) versus threshold and mean delay (right axis) versus threshold, obtained from the Italian data for the COVID-19 pandemic.

other words, it is essential to reduce as much as possible the decision delay to level off the curve of infection as early as possible. At the same time, it is important to enforce restrictions only if essential, in order to avoid unnecessary social unrest and economic cost. And, inevitably, any detection procedure under the controlled regime \mathcal{H}_0 can produce false alarms, i.e., it can wrongly declare the upcoming onset of an exponential growth phase. The risk of taking a wrong decision can be quantified by the inverse of the mean time between false alarms. A false alarm could lead to the unnecessary adoption of restrictive measures. Since these have social and economic ramifications, adoption of this risk proxy quantifies how many times such an event occurs on average. Balancing between detection delay and risk represents a fundamental system trade-off. In this paper, we consider a relatively simple, but effective, mathematical model of the pandemic and develop a decision tool to quickly detect the passage from a controlled regime to a critical one. Its effectiveness in terms of delay/risk trade-off is demonstrated on publicly available COVID-19 data from several countries. It is our hope that the proposed MAST procedure can be useful in making timely and rational decisions to control the pandemic evolution.

Results

The main result of this research is the design of the MAST quickest-detection procedure, which is specifically tailored to epidemic scenarios, and its application on COVID-19 data. In this context, we show that the mean time Δ required to reveal the onset of an exponential phase is in the order of a few days (or few weeks), with a risk R that scales exponentially with the delay:

$$R \sim \exp(-\omega\Delta), \quad (2)$$

where the symbol “ \sim ” refers to asymptotic behaviour (small R , large Δ), and ω is a parameter whose value varies from country to country. This exponential relationship holds for most *optimal* quickest detection procedures under specific mathematical conditions and optimality criteria¹⁵. In this work, we show empirically that this optimality condition is verified by the MAST procedure when applied to the COVID-19 data, despite the fact that MAST is designed to work in the presence of uncertainty and non-stationarity of the data statistics.

To illustrate these concepts, let us refer to Fig. 2, where data from Italy are considered. Panel (a) shows the curve of daily positives from Johns Hopkins University COVID-19 data²⁸, along with the smoothed version thereof, shown in green. Using the latter, we obtain the sequence of growth rates x_1, x_2, \dots , shown in panel (b) (green curve) as the ratio of two consecutive values of daily positives, with its smoothed version (in magenta); see the data model in Eq. (3).

In Fig. 2a, we note that the peak of the first wave is smaller than the second one, which is likely due to the smaller number of swabs made during the first wave. In other words, in the two waves we are observing different fractions of the total infected populations. This difference does not affect our analysis because the sequence x_1, x_2, \dots , of growth rates are ratios of successive daily positives.

As discussed in “Methods” section and illustrated in the notional scheme of Fig. 5, the proposed MAST procedure takes as input the sequence x_1, x_2, \dots and recursively computes the sequence of decision statistics T_1, T_2, \dots , see Eq. (7), where T_n only depends on the observed growth rates x_1, \dots, x_n , up to day n . It is worth stressing that the growth rate process x_1, \dots, x_n , is assumed to be Gaussian distributed with *unknown* and *time-varying* mean value sequence. The MAST decision statistic is “mean-agnostic,” because it does not require knowledge of the time evolution of such a mean sequence, and it is therefore robust to its deterministic fluctuations.

The onset of the critical regime is declared by the MAST at the smallest index n such that $T_n > \chi$, where χ is a threshold level. Panel (c) of Fig. 2 shows the corresponding decision performance. The left axis reports the risk R as a function of the threshold value χ , and the right axis reports the mean delay of decision Δ , again as

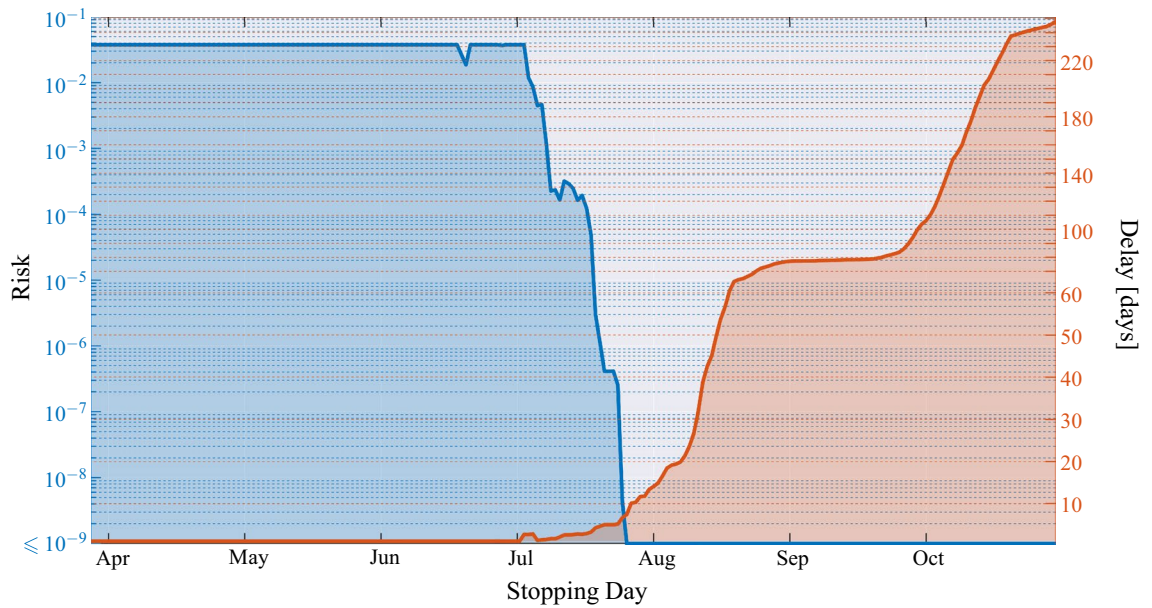


Figure 3. Application of the MAST procedure on the COVID-19 pandemic data from Italy. On the left-side vertical axis, we select a decision threshold to correspond to a desired risk, e.g., $R = 10^{-4}$. Then, the blue curve indicates the stopping day (about July 18, in the example) corresponding to the selected value of risk. Finally, the red curve referred to the right-side vertical axis shows the mean delay Δ corresponding to the selected risk R (about 3 days). For clarity, note that the right-side scale for the delay is split into two linear ranges, for a better rendering of the small- Δ range.

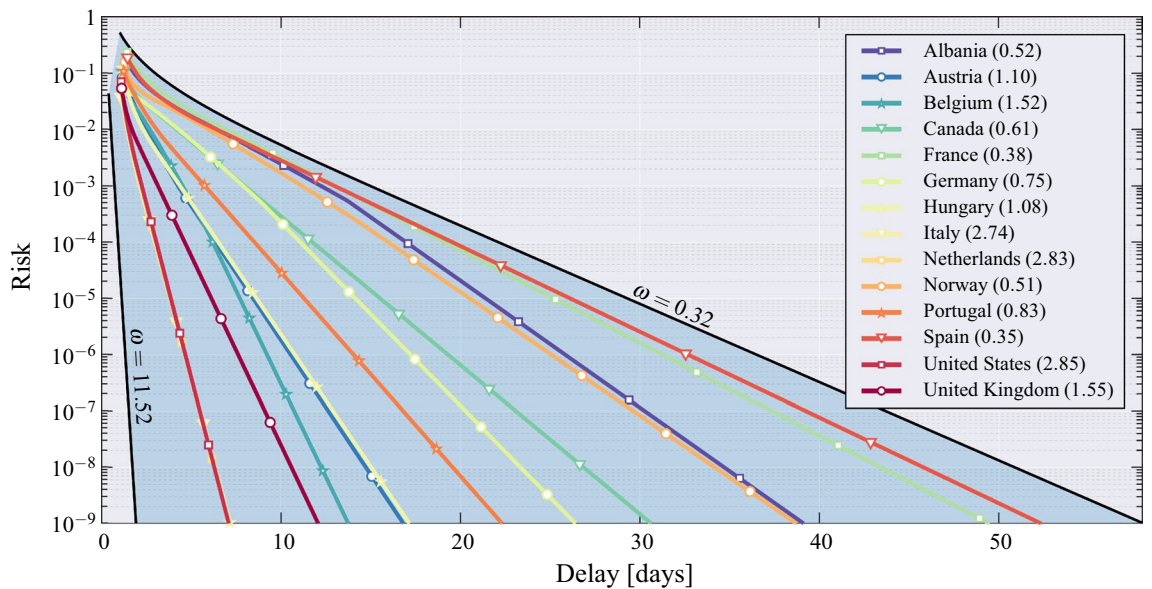


Figure 4. Operational curve—risk versus mean delay for decision—for 14 countries. For large Δ , the operational curve is described by Eq. (2), namely $R \sim \exp(-\omega\Delta)$. The outermost black curves correspond to the ideal performance of the Page’s test, assuming known and constant growth rate, with $\alpha = 0.01$ and 0.06 , respectively (extreme values of α for the second wave, see Fig. 1). Each nation is characterized by a specific value of ω (reported between brackets on the legend), and all values of ω lie in the range between $\omega = 0.32$ and $\omega = 11.52$.

a function of χ . We see that the function $R(\chi)$ is exponentially decreasing, while $\Delta(\chi)$ is linearly increasing, consistent with Eq. (2) and with known relationships for Page’s test and other quickest-detection procedures^{15,16}.

Combining the evidence shown in Fig. 2, we obtain Fig. 3 that can be interpreted as follows. The blue curve refers to the left-side vertical axis, while the red curve refers to the right-side vertical axis. Starting from a given level of risk selected on the left-side vertical axis, we find the *stopping day* of MAST, which is the day on which the

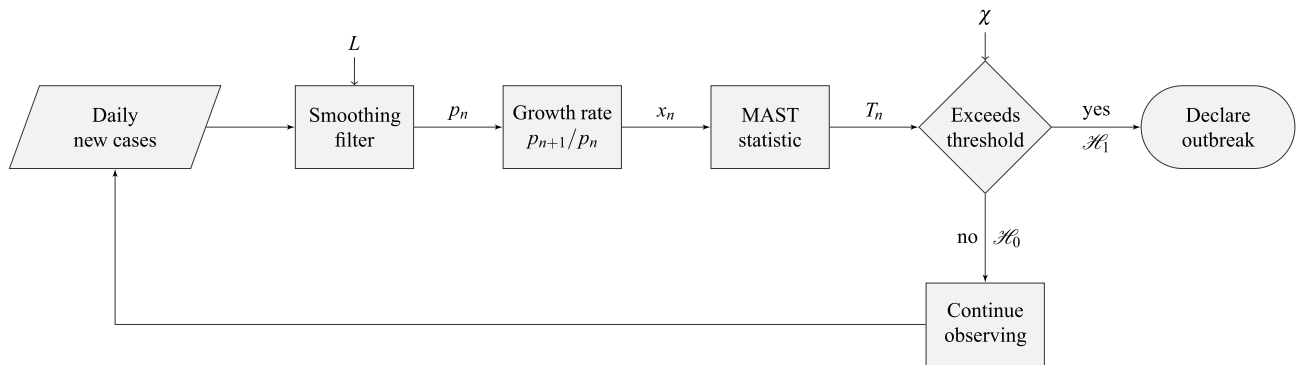


Figure 5. Flowchart of the proposed MAST procedure. Input data are the daily infected. These noisy data are filtered to mitigate imperfections in data collection, randomness, and delays. Filtered daily positive p_n are used to compute the growth rate x_n , which is used to compute the MAST statistic T_n . The statistic is then compared with the threshold χ ; if it is larger than the threshold, the outbreak is declared. Otherwise, the procedure continues to collect and process the data.

onset of the critical regime is declared (first threshold crossing). Then, referring to the red curve, one obtains the corresponding delay Δ on the right-side vertical axis. This is the mean delay incurred by the MAST procedure in declaring the onset of the critical regime. The interpretation of Δ is that the passage to the exponential growth of the pandemic takes place, on the average, Δ days before the alert produced by MAST. In Fig. 3, values of the risk smaller than 10^{-9} are collapsed to $R = 10^{-9}$, because for $R \leq 10^{-9}$ we can safely assume that risk is essentially negligible. It is worth noting that, even at negligible risk level, the stopping day is approximately July 27, and the corresponding mean delay is less than 8 days.

The analysis in Figs. 2, 3 can be repeated for other regions, yielding similar insights. We report the cases of: United States of America in Fig. 7; United Kingdom in Fig. 8; France in Fig. 9; and Germany in Fig. 10. Note that in the United States, there exists also a third wave of the pandemic, which we analyze separately in Fig. 7d, by restarting the test after declaring the onset of the second wave. (Actually, such a third wave is likely to be a delayed second wave in different geographic regions of the U.S., as a state-by-state analysis seems to imply). Additional analysis on data from more geographical regions is available online³⁰ and updated regularly.

A comparison of the MAST performance for 14 different nations is addressed in Fig. 4. Different decision performances reflect different values of the parameter ω governing the relationship between risk and delay, shown in Eq. (2). We see that, accepting a risk of $R = 10^{-4}$, the mean detection delay Δ is about 3 days for Netherlands and below 20 days for Spain. Thus, the MAST procedure detects an outbreak very expeditiously at very low risk level.

The outer black curves labeled by $\omega = 0.32$ and $\omega = 11.52$ represent the ideal performance of the Page's test (see Eq. (6)), in which the mean values of the growth rates x_1, x_2, \dots , are assumed constant, known in advance, and equal to $(1 + \alpha)$ under \mathcal{H}_1 and $(1 - \alpha)$ under \mathcal{H}_0 . For such a scenario, the parameter ω appearing in Eq. (2) is the Kullback-Leibler distance $2(\alpha/\sigma)^2$ between the distributions under the two regimes^{15,31}. In signal processing language, this quantity is often referred to as signal-to-noise ratio³². To provide performance envelopes, we use the values $\alpha = 0.01$ and $\alpha = 0.06$ corresponding to the extreme growth rates of the second wave reported in Fig. 1, and set $\sigma = 0.025$, which is the arithmetic mean of the estimated standard deviations of the 14 countries. For the two values of α , this yields $2(\alpha/\sigma)^2 = 0.32$ and 11.52 , respectively, which are the values reported in Fig. 4. By setting $\omega = 2(\alpha/\sigma)^2$ in Eq. (2) we observe that the higher is the growth rate, the better is the detection capability for a given risk. In other words, consistent with intuition, the more aggressive is the outbreak, the more quickly it can be detected. On the other hand, the larger is σ , the higher is the delay for a given risk. Also this effect is intuitive, because the standard deviation σ measures the entropy of random fluctuations in the data x_1, x_2, \dots , and reliable decisions require more time because the data is uncertain. These trade-offs are also observed for the MAST test run over COVID-19 data for different nations, and are captured by the parameter ω in Eq. (2).

Discussion

Let us focus again on Italy, and set $R = 10^{-4}$. Accepting a risk in the order of $R = 10^{-4}$ means that there is one chance in ten thousand that the countermeasures are taken too early or, in other words, unjustified actions against the pandemic are adopted every 27 years, on the average. By taking drastic countermeasures on July 18—the stopping day prescribed by our MAST procedure for $R = 10^{-4}$ —we would have left only $\Delta \approx 3$ days of uncontrolled exponential growth of the pandemic, before addressing it. In this respect, it should be noted that the adoption of severe countermeasures in Italy has been decided only at the beginning of November 2020. The delayed decision situation is analogous for other nations.

It is clear that managing an unprecedented pandemic event is a huge and a multifaceted problem, which can only be addressed by taking into account many different perspectives. It is also clear that the decision when to take pandemic countermeasures depends on a large number of societal factors. The contribution of this article is limited to the analysis of the pandemic strictly from a quickest-detection viewpoint and, from this perspective, we obtain useful insights and quantitative analyses. One evidence, as just pointed out, is that critical regimes of many nations began dramatically earlier than when countermeasures were taken. In this sense, we believe that the proposed decision support tool would be a key component of a command and control (C2) system that anticipates,

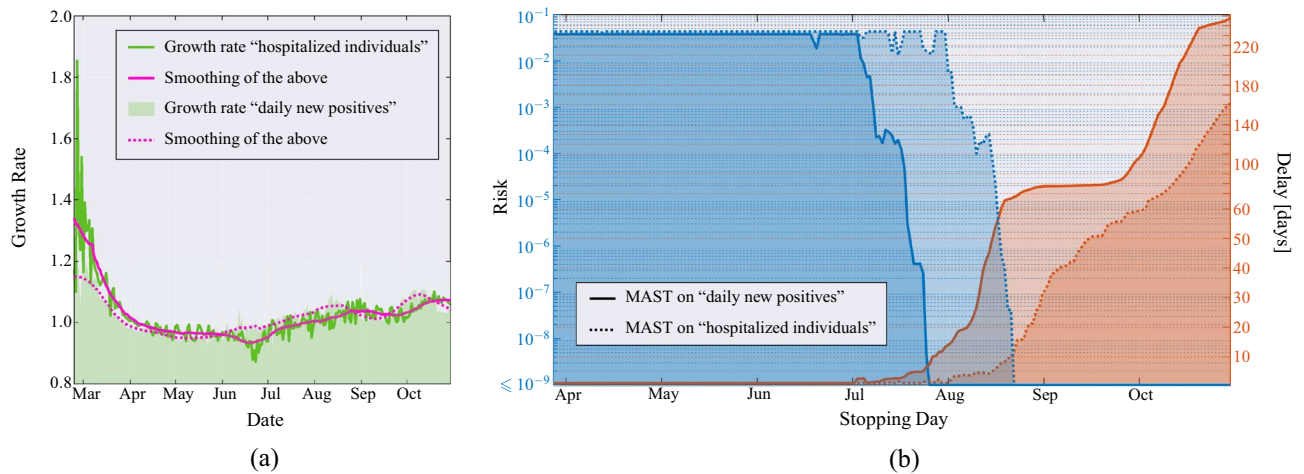


Figure 6. (a) Growth rate of the hospitalized individuals (green solid line)—and its time-varying mean obtained through a moving average that uses a window of 21 days (magenta solid line)—compared to the growth rate of the daily new positives individuals (green area)—and its time-varying mean (magenta dotted line)—in Italy since February 21, 2020. (b) Application of the MAST procedure on the growth rate sequence of the daily new positive individuals (solid lines, already shown in Fig. 4) and on the growth rate sequence of the hospitalized individuals (dotted lines). On the left-side vertical axis we select a desired risk, e.g., $R = 10^{-4}$. Then, the blue curves indicate the stopping day (about July 18 if the growth rate sequence of the daily new positive individuals is used, and August 10 if the growth rate sequence of the hospitalized individuals is used) corresponding to the selected value of risk. Finally, the red curves referred to the right-side vertical axis show the mean delay Δ corresponding to the selected risk R (about 3 days if the growth rate sequence of the daily new positive individuals is used, and below 5 days if the growth rate sequence of the hospitalized individuals is used). For clarity, note that the right-side scale for the delay is split into two linear ranges, for a better rendering of the small- Δ range.

and possibly reacts as soon as possible, to threats. Such a C2 system would be useful for both national health security and armed forces in the context of Chemical, Biological, Radiological and Nuclear (CBRN) defense^{33,34}.

Aside from the above retrospective analysis, a major contribution of the MAST quickest detection tool developed in this paper consists of providing proactive decision support for detecting future waves of the COVID-19 outbreak, and the onset of future pandemics. In these cases, precise nation-dependent performance predictions, such as those given in Fig. 4, cannot be available in advance because the curves have been obtained by exploiting estimates of the mean value of the growth rate (see e.g., the curve in magenta in Fig. 2), derived by forensic inspection of the data. However, good approximate performance bounds can be obtained by assuming constant growth rates ($1 \pm \alpha$) in the critical and controlled regime, respectively, for “reasonable” values of α , for instance, the faster and the slower rates reported in Fig. 2 and used to compute the outermost black curves in Fig. 4. The corresponding value of ω would be equal to the Kullback-Leibler information measure, which governs the performance of a clairvoyant Page’s test that knows exactly the constant growth rate.

Figure 5 shows the flowchart of the MAST procedure. The filtering operation in Fig. 5 is important to mitigate gross errors and lack or delayed reporting of the input data (for instance, thousands of positive individuals from previous days are all reported on the current day, number of recovered individuals unavailable in the US data for a long time, etc.) To compute p_n , the filtering operation used in this paper requires one to observe data samples beyond the current day n , which causes delays for on-line implementations. In these cases, alternative causal filtering strategies³⁵, such as the Savitzky-Golay filter³⁶, would be more appropriate. Another possibility to handle outliers could be the application of Huber’s robust statistic³⁷. We leave such possible enhancements to future work.

The quickest-detection tool developed can also be applied to different time-series, other than the sequence of growth rate of daily new positives $\{x_n\}$, for instance to the hospitalized individuals addressed, for the Italian case, in Fig. 6. The sequence of hospitalized individuals does not require the smoothing operation shown in the flowchart of Fig. 5, because the collected data are less affected by gross errors. These data are also normally distributed with good accuracy, as confirmed by goodness-of-fit analysis (not reported). In Fig. 6 we see that the decision taken by using the hospitalized sequence is delayed as compared to that obtained by the sequence of daily new positives. This behaviour is intuitive because the hospitalized individuals are a subset of the positive ones, and a possible hospitalization follows the onset of clinical symptoms.

Public opinion soundings suggest increased concern about new SARS-CoV-2 variants³⁸, especially related to their higher fatality rate³⁹, spreading velocity^{40,41}, and the possibility that approved vaccines might be less effective against them^{42–44}. The proposed quickest detection tool is useful also in the presence of variants of the coronavirus, which would affect the growth rate; indeed, variants were already present in the United Kingdom during the second wave, reported here in Fig. 8. It is worth noting here that the MAST statistic does not require knowledge of the time-varying mean value sequence of the growth rate.

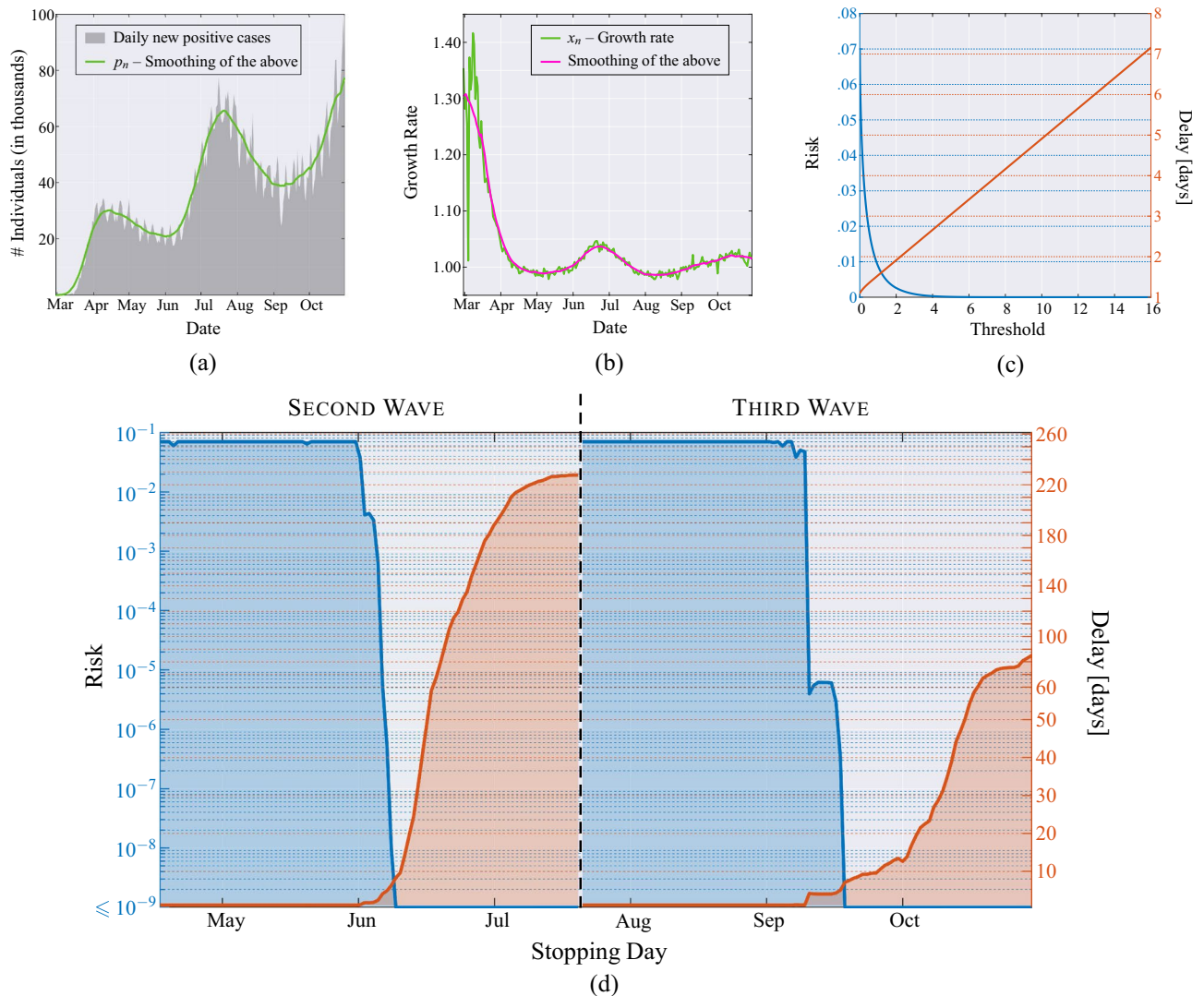


Figure 7. United States of America. **(a)** Daily new positive cases in the US since February 29, 2020, and its moving average obtained with a window of 21 days (green line). **(b)** Growth rate of the epidemic computed from the averaged daily new positive cases (green line), and its time-varying mean obtained through a moving average that uses a window of 21 days (magenta line). **(c)** MAST performance, in terms of risk (left axis) versus threshold and mean delay (right axis) versus threshold, obtained from the US data of the COVID-19 pandemic. **(d)** Application of the MAST procedure on the COVID-19 pandemic data from US. Both the second wave (left side) and the third wave (right side) are analysed. For each wave, we select a desired risk on the left-side vertical axis, e.g., $R = 10^{-4}$. Then, the blue curve indicates the stopping day (about June 6 for the first wave and September 10 for the third wave) corresponding to the selected value of risk. Finally, the red curve referred to the right-side vertical axis shows the mean delay Δ corresponding to the selected risk R (approximately 4 days for both the second and the third waves). For clarity, note that the right-side scale for the delay is split into two linear ranges, for a better rendering of the small- Δ range.

The possibility of processing different sequences of data opens the way to the design of more sophisticated decision rules based on joint processing of multiple time-series. In addition, the MAST procedure can easily accommodate different definitions of critical regime. For instance, if a pandemic growth at rate $(1 + \alpha^*)$ is considered acceptable, for some $0 < \alpha^* \ll 1$, the critical regime can be characterized by rates exceeding $(1 + \alpha^*)$, rather than 1. The necessary modifications to the MAST statistic are straightforward. We also envision that considering both the passage from a controlled to a critical regime and *vice versa* can be addressed by minor modifications to the MAST procedure^{45,46}.

Special attention is given in the literature to the evaluation of underreporting and undertesting of COVID-19 cases^{25,47}. The mortality rate is used as the main indicator to evaluate the extent of underreporting and underdetection of COVID-19 cases⁴⁷. However, in our context, the evaluation of underreporting cases would not be beneficial in terms of quickest detection, as it is provided by an estimation procedure that uses the same data processed by the quickest detection. As already mentioned, all the available data (hospitalized individuals, daily deaths, daily number of tests etc.) could be used jointly and thus improve the detection capability and reliability of the approach.

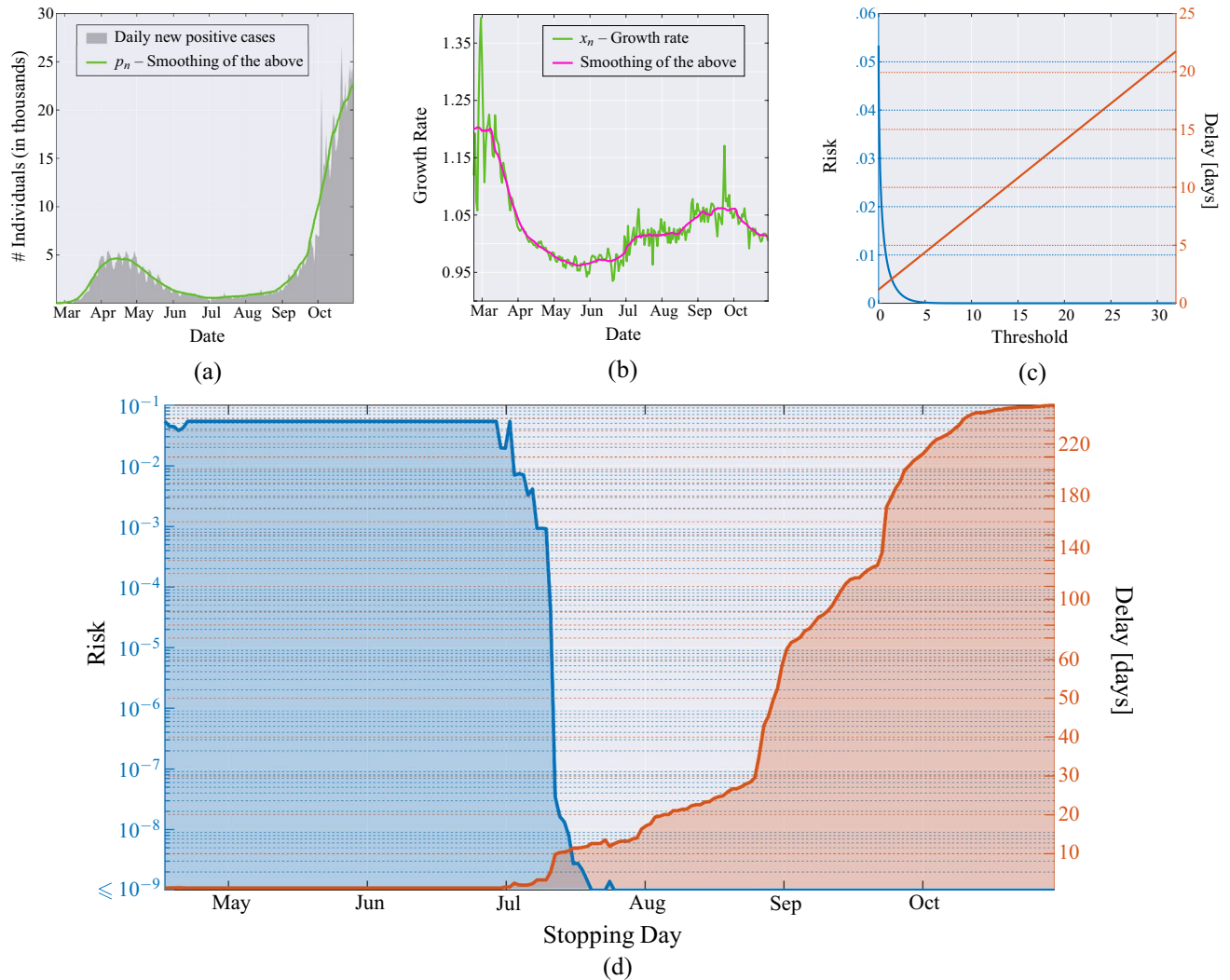


Figure 8. United Kingdom. **(a)** Daily new positive individuals in the UK since February 23, 2020, and its moving average obtained with a window of 21 days (green line). **(b)** Growth rate of the epidemic computed from the averaged daily new positive cases (green line), and its time-varying mean obtained through a moving average that uses a window of 21 days (magenta line). **(c)** MAST performance, in terms of risk (left axis) versus threshold and mean delay (right axis) versus threshold, obtained from the UK data of the COVID-19 pandemic. **(d)** Application of the MAST procedure on the COVID-19 pandemic data from UK. On the left-side vertical axis, we select a desired risk, e.g., $R = 10^{-4}$. Then, the blue curve indicates the stopping day (about July 11, in the example) corresponding to the selected value of risk. Finally, the red curve referred to the right-side vertical axis shows the mean delay Δ corresponding to the selected risk R (below 6 days). For clarity, note that the right-side scale for the delay is split into two linear ranges, for a better rendering of the small- Δ range.

An extended analysis of COVID-19 infection data from more countries than those covered in this paper is available on the web³⁰. We hope that in the near future the publicly available data can be: (i) more reliable so as to mitigate bias effects due to, e.g., false positives, contrasting multiple test outcomes for the same individual, markedly different contagion incidence in close geographical areas, etc.; and (ii) released with finer granularity so as to allow for analyses stratified by population age, comorbidity, etc. These aspects are also relevant for effective vaccination policies.

Methods

The observation model used in this paper can be formally obtained by replacing the constant growth rate $(1 + \alpha)$ appearing in Eq. (1) by the sequence of random variables x_1, x_2, \dots , yielding

$$p_{n+1} = p_1 \prod_{k=1}^n x_k, \quad n = 1, 2, \dots \tag{3}$$

where p_n is the number of new cases on day n , while $x_n = p_{n+1}/p_n$, a time-series that makes explicit the growth rate that we seek to coopt. The model in Eq. (3) is validated empirically. In particular, in the accompanying

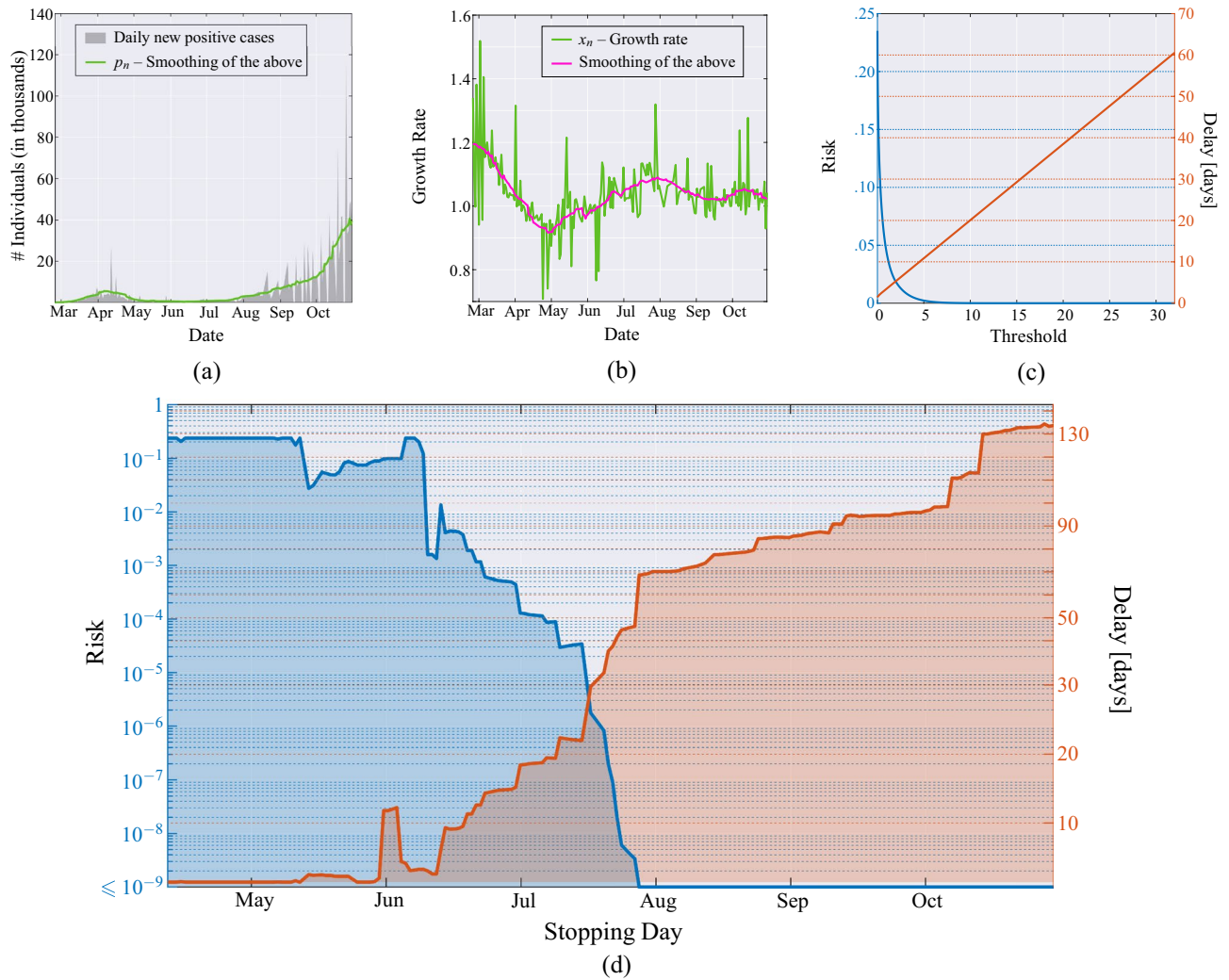


Figure 9. France. **(a)** Daily new positive individuals in France since February 25, 2020, and its moving average obtained with a window of 21 days (green line). **(b)** Growth rate of the epidemic computed from the averaged daily new positive cases (green line), and its time-varying mean obtained through a moving average that uses a window of 21 days (magenta line). **(c)** MAST performance, in terms of risk (left axis) versus threshold and mean delay (right axis) versus threshold, obtained from the French data of the COVID-19 pandemic. **(d)** Application of the MAST procedure on the COVID-19 pandemic data from France. On the left-side vertical axis, we select a desired risk, e.g., $R = 10^{-4}$. Then, the blue curve indicates the stopping day (about July 7, in the example) corresponding to the selected value of risk. Finally, the red curve referred to the right-side vertical axis shows the mean delay Δ corresponding to the selected risk R (below 20 days). For clarity, note that the right-side scale for the delay is split into two linear ranges, for a better rendering of the small- Δ range.

Supplementary Information, we elaborate on the classic susceptible-infected-recovered (SIR) compartmental epidemic model to motivate the usage of the sequence $\{x_n\}$ as observable process for quickest detection. Using the model in Eq. (3), we assume to have available the sequence of daily new positives, for a certain region of interest; the analysis presented here relies on the data provided by the Johns Hopkins University²⁸. Referring for instance to Fig. 2, such a sequence is shown in gray in the left panel. To address gross errors, missing values and delays in reporting the data, the sequence is smoothed by a moving average filter with uniform weights (the moving average filter of length L is always assumed to have equal $1/L$ weights). The smoothed sequence $\{p_n\}$ so obtained is shown in green in Fig. 2a. The growth rate process used as observable is then computed as $\{x_n = p_{n+1}/p_n\}$, and is represented by the green curve in Fig. 2b. The length of the filter is selected to $L = 21$ as to obtain a convenient mitigation of the effects of gross errors in all analyzed countries. After such a pre-processing, we implement Kolmogorov-Smirnov tests to check data Gaussianity, see details in the Supplementary Information.

The time-varying statistical mean $\{\mu_n\}$ of the sequence can be estimated by low-pass filtering of the sequence $\{x_n\}$, and for this we use again a moving average filter of length $L = 21$ days. By subtracting from each x_n the estimated mean value $\hat{\mu}_n$, that is, the curve in magenta in Fig. 2b, one obtains the sequence $x_n - \hat{\mu}_n$. Statistical analysis conducted by Kolmogorov-Smirnov goodness-of-fit test⁴⁸ reveals that $x_n - \hat{\mu}_n$, for each $n = 1, 2, \dots$, can be modeled by a zero-mean Gaussian random variable with (country-dependent) standard deviation $\sigma \ll 1$.

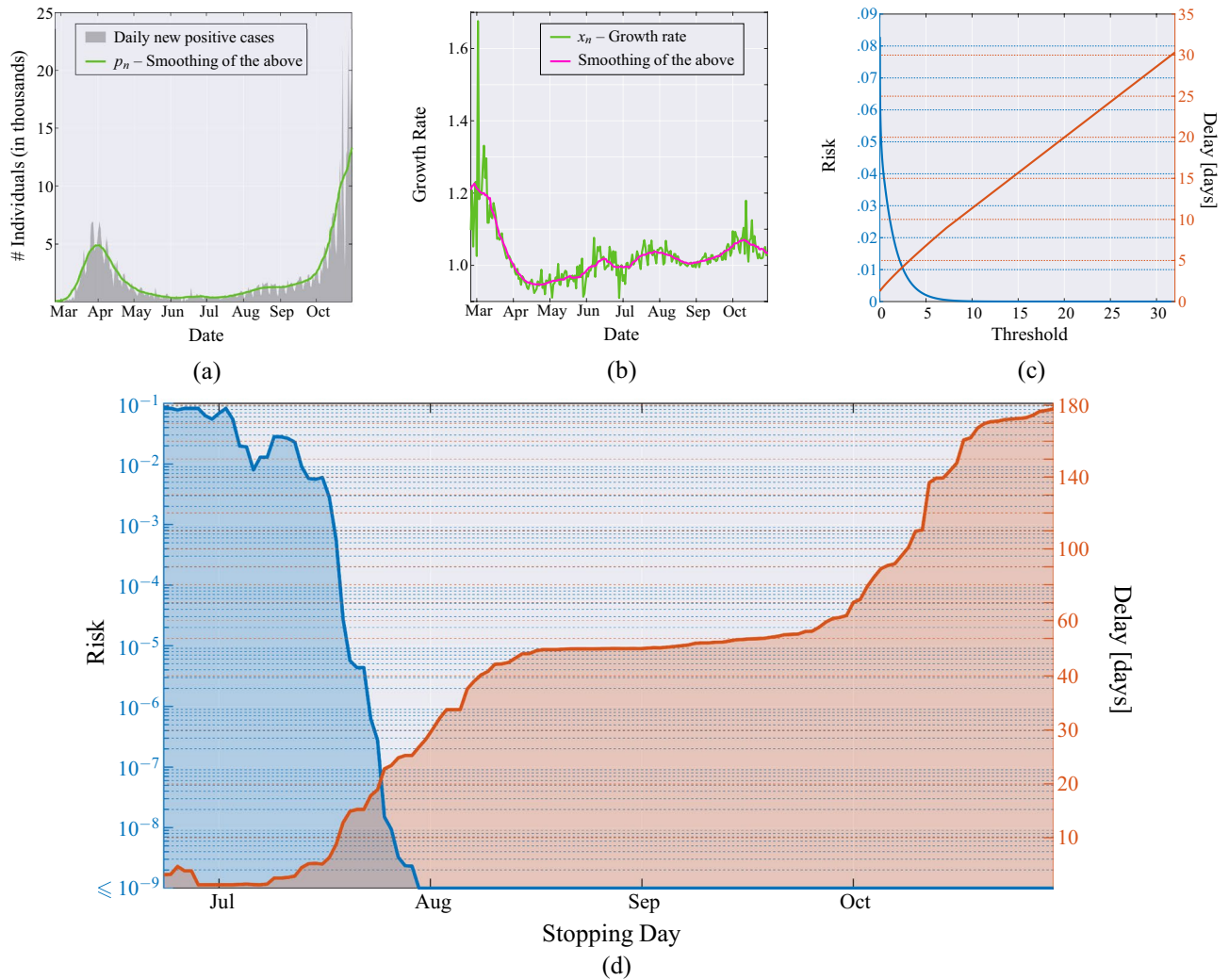


Figure 10. Germany. **(a)** Daily new positive individuals in Germany since February 25, 2020, and its moving average obtained with a window of 21 days (green line). **(b)** Growth rate of the epidemic computed from the averaged daily new positive cases (green line), and its time-varying mean obtained through a moving average that uses a window of 21 days (magenta line). **(c)** MAST performance, in terms of risk (left axis) versus threshold and mean delay (right axis) versus threshold, obtained from the German data of the COVID-19 pandemic. **(d)** Application of the MAST procedure on the COVID-19 pandemic data from Germany. On the left-side vertical axis, we select a desired risk, e.g., $R = 10^{-4}$. Then, the blue curve indicates the stopping day (about July 19, in the example) corresponding to the selected value of risk. Finally, the red curve referred to the right-side vertical axis shows the mean delay Δ corresponding to the selected risk R (below 13 days). For clarity, note that the right-side scale for the delay is split into two linear ranges, for a better rendering of the small- Δ range.

Since μ_n is close to unity, we see that $x_k < 0$ with negligible probability, hence the Gaussian approximation should not be problematic.

By observing the sequence $\{x_n\}$ as time n elapses, we want to detect the passage from the controlled regime \mathcal{H}_0 , to the critical regime \mathcal{H}_1 , when there is the exponential growth. In the controlled regime, the mean value of the random variable x_n is below one, i.e., $\mu_n = \mu_{0,n} \leq 1$, meaning that the number of new positives remains approximately stable or decreases. Conversely, in the critical regime, the mean value is greater than unity, $\mu_n = \mu_{1,n} > 1$, which implies an explosion of daily new positives in the long run. Formally:

$$\text{controlled regime } \mathcal{H}_0 \Rightarrow x_n \sim \mathcal{N}(\mu_{0,n}, \sigma), \quad \mu_{0,n} \leq 1, \tag{4}$$

$$\text{critical regime } \mathcal{H}_1 \Rightarrow x_n \sim \mathcal{N}(\mu_{1,n}, \sigma), \quad \mu_{1,n} > 1. \tag{5}$$

We assume that the x_k 's are mutually independent under either regime. As we show in the accompanying Supplementary Information, slight deviations from the condition of perfect independence do not significantly affect the results of the proposed MAST. Different nations are characterized by different values of σ , hence for

each country σ is assumed known, because in practice it can be estimated on-line from the data. Conversely, the quantities $\{\mu_{0,n}\}$ and $\{\mu_{1,n}\}$ are modeled as deterministic but *unknown* sequences.

To detect the change of regime, we rely on the Generalized Likelihood Ratio Test (GLRT) approach^{12–14}, a milestone of decision theory in scenarios where the statistical distributions of the data contain unknown parameters—in our case, the sequences of mean values $\{\mu_{0,n}\}$ and $\{\mu_{1,n}\}$, and the time at which the passage of regime occurs. If we assume the sequences of mean values in the two regimes to be constant and known, say $\mu_{0,n} = 1 - \alpha$ and $\mu_{1,n} = 1 + \alpha$ (cf. Eq. (1)), the GLRT solution to the quickest-detection problem would be the celebrated Page's test: compare to an appropriate threshold the CUSUM statistic^{15,16,18}

$$\begin{cases} Q_0 = 0, \\ Q_n = \max \left\{ 0, Q_{n-1} + \frac{2\alpha(x_n-1)}{\sigma^2} \right\}, \quad n \geq 1. \end{cases} \quad (6)$$

In the presence of unknown and time-varying sequences complying with the constraints on $\mu_{0,n}$ and $\mu_{1,n}$, shown in (4)–(5), the GLRT solution to the quickest-detection problem amounts to compare the MAST statistic T_n to a threshold χ ⁴⁹:

$$\begin{cases} T_0 = 0, \\ T_n = \max \left\{ 0, T_{n-1} + \frac{(x_n-1)^2 \text{sign}(x_n-1)}{2\sigma^2} \right\}, \quad n \geq 1, \end{cases} \quad (7)$$

and declaring the change of regime at the first occurrence of the threshold crossing. It is worth noting that the MAST statistic is formally obtained by replacing the unknown value of α appearing in the CUSUM statistic, with the estimate $\hat{\alpha}_n = |x_n - 1|$ (constant factors can be incorporated in the threshold). The reader educated in detection theory will recognize the analogy with the energy detector arising in testing the presence of an unknown time-varying deterministic signal buried in Gaussian noise³².

The threshold χ employed in the MAST procedure is selected to trade-off decision delay Δ and risk R , two quantities that are defined as follows. The mean delay Δ is the mean value of the difference between the time at which the MAST statistic crosses the threshold χ and the time of passage from the controlled to the critical regime. The risk R is defined as reciprocal of the mean time between successive false alarms (in fact, “false alarm probability” is perhaps a more familiar jargon to readers with background in detection theory), where the false alarm is defined as a threshold crossing during the controlled regime (i.e., the regime in which there is no need for intervention). In this paper, the functions $R(\chi)$ and $\Delta(\chi)$ are obtained by standard Monte Carlo simulations^{13,27,32} for relatively small values of χ , see e.g., Fig. 2c. We found that the first function is essentially exponential, and the second is essentially linear, consistent with known expressions for the Page's test^{15,16}. This allows us to extrapolate the behaviour of the two functions for values of χ that would be problematic to obtain from real data or by computer experiments, such as those used in Fig. 3 and similar figures for other nations.

Data availability

A visualization of MAST performance on updated COVID-19 infection data from an extended set of countries is available at <https://covid-mast.github.io>.

Received: 10 December 2020; Accepted: 2 March 2021

Published online: 20 April 2021

References

- Merriam Webster. Definition of COVID-19. <https://www.merriam-webster.com/dictionary/COVID-19>.
- World Health Organization (WHO). Coronavirus disease (COVID-19) pandemic. <http://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/novel-coronavirus-2019-ncov> (2020).
- Anderson, R. M., Heesterbeek, H., Klinkenberg, D. & Hollingsworth, T. D. How will country-based mitigation measures influence the course of the COVID-19 epidemic?. *Lancet* **395**, 931–934. [https://doi.org/10.1016/S0140-6736\(20\)30567-5](https://doi.org/10.1016/S0140-6736(20)30567-5) (2020).
- Hellewell, J. *et al.* Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob. Health* **8**, e488–e496. [https://doi.org/10.1016/S2214-109X\(20\)30074-7](https://doi.org/10.1016/S2214-109X(20)30074-7) (2020).
- Maier, B. F. & Brockmann, D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science* **368**, 742–746. <https://doi.org/10.1126/science.abb4557> (2020).
- Dehning, J. *et al.* Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science* <https://doi.org/10.1126/science.abb9789> (2020).
- Gaglione, D. *et al.* Adaptive Bayesian learning and forecasting of epidemic evolution—Data analysis of the COVID-19 outbreak. *IEEE Access* **8**, 175244–175264. <https://doi.org/10.1109/ACCESS.2020.3019922> (2020).
- Nicola, M. *et al.* The socio-economic implications of the coronavirus pandemic (COVID-19): A review. *Int. J. Surg.* **78**, 185–193. <https://doi.org/10.1016/j.ijsu.2020.04.018> (2020).
- Sharif, A., Aloui, C. & Yarovaya, L. COVID-19 pandemic, oil prices, stock market, geopolitical risk and policy uncertainty nexus in the US economy: Fresh evidence from the wavelet-based approach. *Int. Rev. Financ. Anal.* **70**, 101496. <https://doi.org/10.1016/j.irfa.2020.101496> (2020).
- Guan, D. *et al.* Global supply-chain effects of COVID-19 control measures. *Nat. Hum. Behav.* **4**, 577–587. <https://doi.org/10.1038/s41562-020-0896-8> (2020).
- Millefiori, L. M. *et al.* COVID-19 impact on global maritime mobility. *Nat. Commun.* (2020). Under review. [arXiv:2009.06960](https://arxiv.org/abs/2009.06960).
- Poor, H. V. *An Introduction to Signal Detection and Estimation* (Springer, 1988).
- Lehmann, E. L. *Testing Statistical Hypotheses* (Springer, 2005).
- Shao, J. *Mathematical Statistics* 2nd edn. (Springer, 2003).
- Basseville, M. & Nikiforov, I. V. *Detection of Abrupt Changes: Theory and Application* Vol. 104 (Prentice Hall Englewood Cliffs, 1993).
- Poor, H. V. & Hadjiladis, O. *Quickest Detection* (Cambridge University Press, 2009).

17. Truong, C., Oudre, L. & Vayatis, N. Selective review of offline change point detection methods. *Signal Process.* **167**, 107299. <https://doi.org/10.1016/j.sigpro.2019.107299> (2020).
18. Page, E. Continuous inspection schemes. *Biometrika* **41**, 100–115 (1954).
19. Frisén, M. Optimal sequential surveillance for finance, public health, and other areas. *Seq. Anal.* **28**, 310–337. <https://doi.org/10.1080/07474940903041605> (2009).
20. Frisén, M., Andersson, E. & Schiöler, L. Robust outbreak surveillance of epidemics in Sweden. *Stat. Med.* **28**, 476–493. <https://doi.org/10.1002/sim.3483> (2008).
21. Bock, D., Andersson, E. & Frisén, M. Statistical surveillance of epidemics: Peak detection of influenza in Sweden. *Biometrical J.* **50**, 71–85. <https://doi.org/10.1002/bimj.200610362> (2008).
22. Kermack, W. O., McKendrick, A. G. & Walker, G. T. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond* **115**, 700–721. <https://doi.org/10.1098/rspa.1927.0118> (1927).
23. Skvortsov, A. & Ristic, B. Monitoring and prediction of an epidemic outbreak using syndromic observations. *Math. Biosci.* **240**, 12–19. <https://doi.org/10.1016/j.mbs.2012.05.010> (2012).
24. Hu, Z. *et al.* Evaluation and prediction of the COVID-19 variations at different input population and quarantine strategies, a case study in Guangdong province. *China. Int. J. Infect. Dis.* **95**, 231–240. <https://doi.org/10.1016/j.ijid.2020.04.010> (2020).
25. Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel Coronavirus (SARS-CoV-2). *Science* **368**, 489–493. <https://doi.org/10.1126/science.abb3221> (2020).
26. Chinazzi, M. *et al.* The effect of travel restrictions on the spread of the 2019 novel Coronavirus (COVID-19) outbreak. *Science* **368**, 395–400. <https://doi.org/10.1126/science.aba9757> (2020).
27. Allen, L. J. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infect. Dis. Model.* **2**, 128–142. <https://doi.org/10.1016/j.idm.2017.03.001> (2017).
28. Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE). COVID-19 data repository. <https://github.com/CSSEGISandData/COVID-19> (2020).
29. Tuite, A. & Fisman, D. Reporting, epidemic growth, and reproduction numbers for the 2019 novel coronavirus (2019-nCoV) epidemic. *Ann. Intern. Med.* **172**, 567–568. <https://doi.org/10.7326/M20-0358> (2020).
30. Braca, P. *et al.* MAST: COVID-19 pandemic onset test—Multi-country analysis and visualization. <https://covid-mast.github.io> (2020).
31. Cover, T. M. & Thomas, J. A. *Elements of Information Theory* 2nd edn. (Wiley-Interscience, 2006).
32. Kay, S. M. *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory* (Prentice Hall PTR, 1998).
33. Kumar, V., Goel, R., Chawla, R., Silambarasan, M. & Sharma, R. K. Chemical, biological, radiological, and nuclear decontamination: Recent trends and future perspective. *J. Pharm. Bioallied Sci.* **2**, 220. <https://doi.org/10.4103/0975-7406.68505> (2010).
34. Ramesh, A. C. & Kumar, S. Triage, monitoring, and treatment of mass casualty events involving chemical, biological, radiological, or nuclear agents. *J. Pharm. Bioallied Sci.* **2**, 239. <https://doi.org/10.4103/0975-7406.68506> (2010).
35. Oppenheim, A. V. & Willsky, A. S. *Signals & Systems* 2nd edn. (Prentice Hall International Inc., 1997).
36. Schafer, R. W. What is a Savitzky-Golay filter? [lecture notes]. *IEEE Signal Process. Mag.* **28**, 111–117. <https://doi.org/10.1109/MSP.2011.941097> (2011).
37. Huber, P. J. *Robust Statistics* Vol. 523 (Wiley, 2004).
38. Alteri, C. *et al.* Genomic epidemiology of SARS-CoV-2 reveals multiple lineages and early spread of SARS-CoV-2 infections in Lombardy, Italy. *Nat. Commun.* **12**, 1–13. <https://doi.org/10.1038/s41467-020-20688-x> (2021).
39. Becerra-Flores, M. & Cardozo, T. SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. *Int. J. Clin. Pract.* **74**, e13525. <https://doi.org/10.1111/ijcp.13525> (2020).
40. Kirby, T. New variant of SARS-CoV-2 in UK causes surge of COVID-19. *Lancet Respir. Med.* **9**, e20–e21. [https://doi.org/10.1016/S2213-2600\(21\)00005-9](https://doi.org/10.1016/S2213-2600(21)00005-9) (2021).
41. Elettrey, R., Zhuang, Y., Carley, K. M., Yağan, O. & Poor, H. V. The effects of evolutionary adaptations on spreading processes in complex networks. *Proc. Natl. Acad. Sci.* **117**, 5664–5670. <https://doi.org/10.1073/pnas.1918529117> (2020).
42. Conti, P. *et al.* The British variant of the new coronavirus-19 (SARS-CoV-2) should not create a vaccine problem. *J. Biol. Regul. Homeost. Agents* <https://doi.org/10.23812/21-3-E> (2021).
43. Kemp, S. A. *et al.* Neutralising antibodies in spike mediated SARS-CoV-2 adaptation. *medRxiv* <https://doi.org/10.1101/2020.12.05.20241927> (2020).
44. Kupferschmidt, K. New mutations raise specter of immune escape. *Science* **371**, 329–330. <https://doi.org/10.1126/science.371.6527.329> (2021).
45. Soldi, G. *et al.* Quickest detection and forecast of pandemic outbreaks: Analysis of COVID-19 waves. *IEEE Commun. Mag.* (2021). Under review. [arXiv:2101.04620](https://arxiv.org/abs/2101.04620).
46. Marano, S. & Sayed, A. H. Decision-making algorithms for learning and adaptation with application to COVID-19 data. *IEEE Trans. Signal Process.* (2020). Under review. [arXiv:2012.07844](https://arxiv.org/abs/2012.07844).
47. Lau, H. *et al.* Evaluating the massive underreporting and undertesting of COVID-19 cases in multiple global epicenters. *Pulmonology* **27**, 110–115 (2021).
48. Feller, W. On the Kolmogorov-Smirnov limit theorems for empirical distributions. *Ann. Math. Stat.* **19**, 177–189. <https://doi.org/10.1214/aoms/1177730243> (1948).
49. Braca, P. *et al.* Quickest detection of COVID-19 pandemic onset. *IEEE Signal Process. Lett.* (2021). Under review. [arXiv:2011.10502](https://arxiv.org/abs/2011.10502).

Acknowledgements

PW was supported by AFOSR under contract FA9500-18-1-0463.

Author contributions

S.M. developed the approach and supervised the work. P.B., L.M.M., and D.G. did most of the analysis. D.G. prepared the figures. L.M.M. prepared the website. P.W. and K.P. supervised the work. All authors contributed in writing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-86827-6>.

Correspondence and requests for materials should be addressed to P.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© NATO 2021, corrected publication 2021