

PAPER • OPEN ACCESS

## Differences in power law growth over time and indicators of COVID-19 pandemic progression worldwide

To cite this article: Jack Merrin 2020 *Phys. Biol.* **17** 065005

View the [article online](#) for updates and enhancements.



**IOP | ebooks™**

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

## OPEN ACCESS

## PAPER



RECEIVED  
24 April 2020

REVISED  
12 August 2020

ACCEPTED FOR PUBLICATION  
26 August 2020

PUBLISHED  
24 September 2020

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



# Differences in power law growth over time and indicators of COVID-19 pandemic progression worldwide

Jack Merrin

IST Austria, Am Campus 1, Klosterneuburg 3400, Austria

E-mail: [jack.merrin@ist.ac.at](mailto:jack.merrin@ist.ac.at)

**Keywords:** COVID-19, power law, logistic, coronavirus, SIR model

Supplementary material for this article is available [online](#)

## Abstract

Error analysis and data visualization of positive COVID-19 cases in 27 countries have been performed up to August 8, 2020. This survey generally observes a progression from early exponential growth transitioning to an intermediate power-law growth phase, as recently suggested by Ziff and Ziff. The occurrence of logistic growth after the power-law phase with lockdowns or social distancing may be described as an effect of avoidance. A visualization of the power-law growth exponent over short time windows is qualitatively similar to the Bhatia visualization for pandemic progression. Visualizations like these can indicate the onset of second waves and may influence social policy.

## 1. Introduction

Mathematical modeling is essential to predict and control the course of pandemics [1, 2]. Many epidemiological theories derive from the SIR model or a more sophisticated version of it containing more categories [3–5]. The basic SIR model may not be sufficient to characterize COVID-19 for several reasons. The parameters of SIR, like the basic reproduction number, can change over time according to the sequence of social policies [6]. In comparison, the parameters of flu or colds remain constant except for potential seasonal variations because people continue to go about their business. The time course of parameters like the basic reproduction number may only be computable retrospectively. The SIR equations are also entirely deterministic, while in real systems, there is some stochastic element of variation. Differential equations do not take into account the integer nature of population categories over time, especially in the initial stages. The SIR model has a difficult to compute solution [7]. SIR model dynamics lumps susceptible people into one category, whereas some groups are more at risk of contracting the virus than others. Spreading the virus may also depend on differences in demographics, employment, city size, social networks, social dynamics, or individual behaviors [8]. Lastly, the SIR model does not take into account cases resulting from travel between countries.

Other approaches do not rely on mechanistic mathematical models at all. One can choose phenomenological functions like power-laws or the logistic curve that seems to fit historical data well [9]. One might infer other countries follow a similar pattern, but with slightly different parameters. If one does not know which best-fit phenomenological functions to choose, one can input historical data into a program called Eurequa, which evolves them by genetic algorithms [10]. Other approaches are visualizations that allow one to interpret the data and observe trends and make comparisons of various countries [11]. In any case, there is quite a lot of freely available worldwide data for scientists to analyze.

In this study, we analyze differences in the spread of COVID-19 hotspots using error analysis. We use the least-squares curve fitting methods to synchronize different data sets and compare power-law growth on log–log plots. The inspiration for this work is Ziff and Ziff’s hypothesis, where early growth is exponential, followed by a power-law growth [12]. The next stage is generally logistic growth and then possible second waves. The visualization and prediction of data using these methods may provide insight into when social policy works, expectations for the burden on medical infrastructure, characterizing progress towards second waves, and decision-making in the future.

## 2. Methods

### 2.1. Linear least squares curve fitting

In this paper, we use error analysis and curve fitting to analyze the progression of COVID-19 [14–17]. In weighted least squares curve fitting for linear functions one finds the fit parameters,  $A, B, \dots$ , which minimizes  $\chi^2$ , which depends on the location of data points  $x_i$ , the measured values  $y_i$ , and the uncertainties in  $y_i$  given by  $\sigma_i$ .

$$\chi^2 = \sum_{i=1}^n \frac{[y_i - A - Bx_i]^2}{\sigma_i^2}.$$

To fit a straight line model, one solves for  $A$  and  $B$  by calculating  $\frac{\partial \chi^2}{\partial A} = 0$  and  $\frac{\partial \chi^2}{\partial B} = 0$  and solving the resulting system of linear equations. We use the notation,

$$S_{ww} = \sum_{i=1}^N w_i u_i v_i \quad w_i = \frac{1}{\sigma_i^2}.$$

The uncertainties in the parameters can also be directly computed using propagation of error.

$$A = \frac{S_{wy}S_{wxx} - S_{wx}S_{wxy}}{S_w S_{wxx} - S_{wx}^2} \quad B = \frac{S_w S_{wxy} - S_{wx}S_{wy}}{S_w S_{wxx} - S_{wx}^2}$$

$$\sigma_A = \sqrt{\frac{S_{wxx}}{S_w S_{wxx} - S_{wx}^2}} \quad \sigma_B = \sqrt{\frac{S_w}{S_w S_{wxx} - S_{wx}^2}}.$$

There are different methods to visualize the progression of the number of positive cases. Typically one uses linear plots, semilogarithmic plots, or log–log plots. Keep in mind several things when interpreting functions on a log–log plot. Exponential growth on a log–log plot appears exponential, and power-law growth on a log–log plot appears linear. If we define two transformations then we just have to fit linear functions. Transformation 1 is

$$Y_i = \ln y_i \quad X_i = \ln x_i.$$

For a power law function  $y = \alpha x^\beta$  then

$$\ln y_i = \ln \alpha + \beta \ln x_i \rightarrow Y_i = A_1 + B_1 X_i.$$

We refer repeatedly to  $\beta$  in this paper as the power-law exponent for the data region of interest.

If we define transformation 2

$$Y_i = \ln y_i \quad X_i = x_i.$$

For an exponential function  $y = y(0)e^{rx}$  then

$$\ln y_i = \ln y(0) + rx_i \rightarrow Y_i = A_2 + B_2 X_i.$$

By propagation of error and the square root  $N$  rule, the weighting is  $\sigma_i = 1/\sqrt{y_i}$  in both transformations even though there are no explicit errorbars in reports of positive cases.

### 2.2. Logistic curves

The logistic function is the solution to the differential equation, which models many different phenomena such as population growth.

$$\frac{dN}{dt} = rN(1 - N/K).$$

We write the solution in several ways that depend on three parameters.

$$N(t) = \frac{K}{1 + e^{-r(t-\tau)}} = \frac{KN(0)}{N(0) + (K - N(0))e^{-rt}}.$$

For short times,  $N$  increases exponentially. For long times,  $N$  is exponentially approaching a plateau. The logistic curve can be fitted by the Levenberg–Marquardt method.

### 2.3. Data sources and synchronization

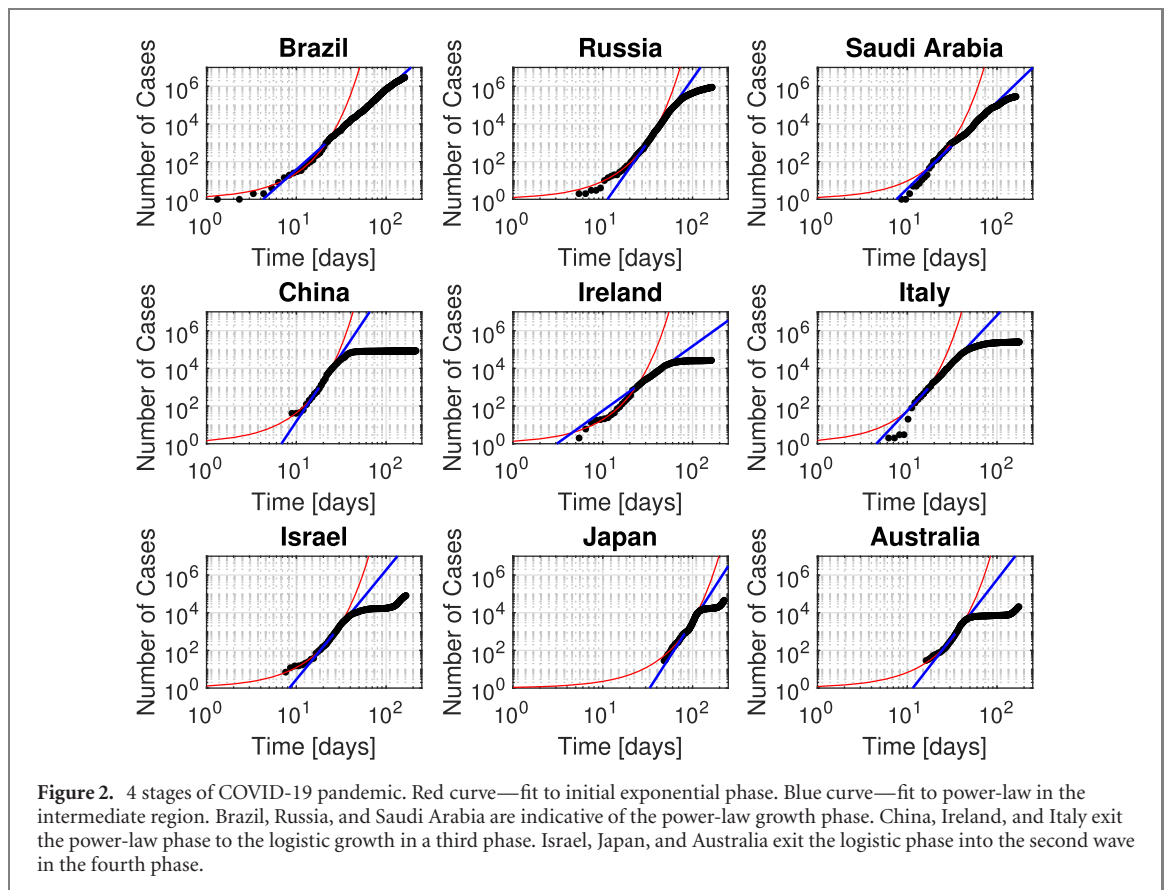
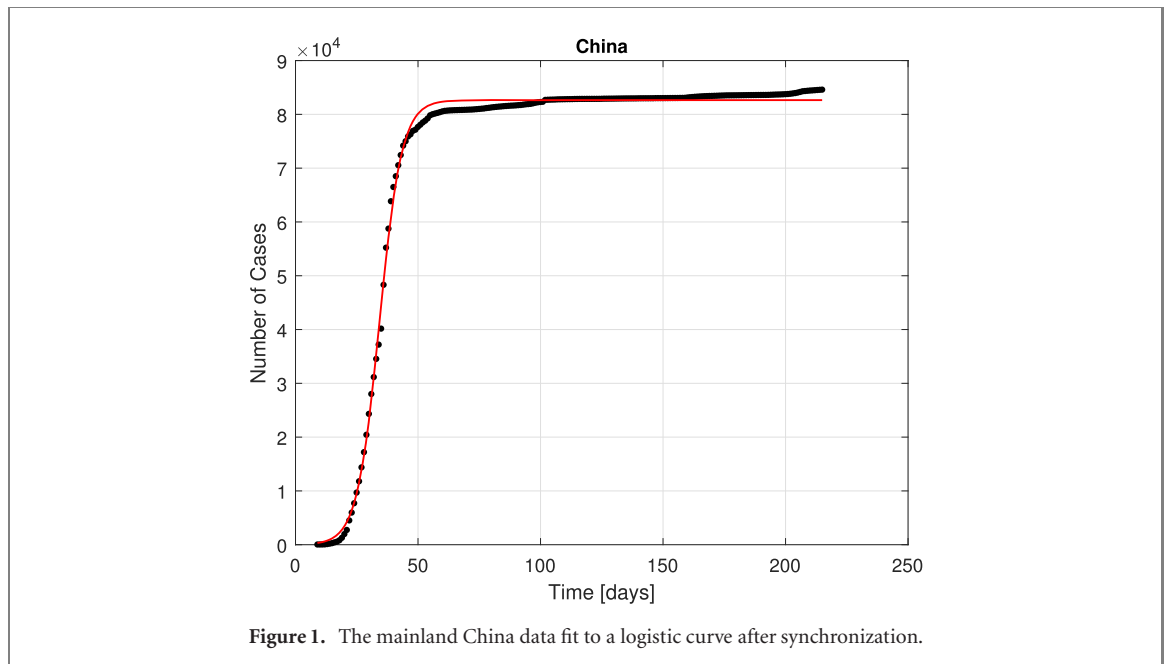
Data was collected from Johns Hopkins [18], the European Centre for Disease Prevention and Control, and Wikipedia. The data files differed on the same days, probably due to how the start of a day is defined or lack of proper reporting. A small number of errors were found in all three data sources, such as the number of cases decreasing on a successive day. Some Wikipedia pages miss the death totals or recoveries, but we consider only positive cases in this paper.

Wikipedia data aggregation appeared to have the best-curated early time data, so it was the only data used for the analysis. Different countries have staggered outbreaks in time, and fitting power-laws depends on the origin of time. For synchronization between countries, an exponential fit to the first 1000 cases was performed. The initial time was an extrapolation of this curve to one case.

### 2.4. Moving window procedure and visualizations

To visualize COVID-19 progress, we track the slope of the number of cases on a log–log plot, which is an exponent  $\beta$ . Using a 7 d or 14 d window did not result in much difference. In this visualization, when  $\beta$  nears zero, the disease is mitigated. When  $\beta$  is large, this indicates a country is going into more danger. If  $\beta$  drops significantly below one and then increases again, this can define a second wave.

A visualization method developed by Bhatia is simple to implement and surprisingly effective [11]. In this method, one plots points of the number of cases in the last week on the logarithmic  $y$ -axis versus the total number of cases on the logarithmic  $x$ -axis. The visualization is a parametric plot eliminating time. Most of the time, the data perfectly line up and progresses upwards and to the right as things get worse for a country. If the plot moves downwards, this indicates mitigation. The noise in actual measurements only modulates the spacing between points but does not impact the general shape of the visualization much.



### 3. Results

#### 3.1. SIA model

Epidemiological mathematical models are only as good as their assumptions. Some more sophisticated models beyond SIR might include different compartments like frontline workers  $F$  in hospitals or supermarkets who have an increased rate of infectivity. Quarantined individuals  $Q$  could be isolated

from infecting others after being detected by tests. Exposed individuals  $E$  could have a time delay before switching to infected. Super spreaders  $Z$  could infect at a higher rate because they do not know they are sick and do not go into quarantine. Adding more sophisticated compartments may seem more realistic, but comes at the cost of more complicated differential equations and more model parameters to be determined.

**Table 1.** Data for the overall power law exponent  $\beta$  for the indicated case range, the synchronization time shift  $t_0$ , and the early time exponential growth constant  $r_1$ .

Country	$\beta$ case range	$\beta$	$t_0$ (d)	$r_1$ per day
Afghanistan	$10^2-10^4$	$5.52 \pm 0.02$	-29.4	$0.106 \pm 0.001$
Australia	$10^2-5 \times 10^3$	$6.13 \pm 0.04$	-16.2	$0.191 \pm 0.003$
Austria	$10^2-5 \times 10^3$	$5.11 \pm 0.04$	-6.6	$0.314 \pm 0.006$
Belgium	$10^2-10^4$	$5.88 \pm 0.03$	-11.4	$0.249 \pm 0.006$
Bolivia	$10^2-5 \times 10^4$	$5.55 \pm 0.01$	-32.3	$0.087 \pm 0.001$
Brazil	$10^3-10^6$	$4.25 \pm 0.01$	-1.3	$0.318 \pm 0.006$
Canada	$10^2-10^5$	$6.94 \pm 0.04$	-7.4	$0.215 \pm 0.003$
China	$10^2-10^4$	$7.12 \pm 0.05$	-8.9	$0.376 \pm 0.009$
Czechia	$10^2-10^3$	$5.48 \pm 0.16$	-6.4	$0.271 \pm 0.005$
Egypt	$5 \times 10^2-5 \times 10^4$	$4.98 \pm 0.01$	-35.1	$0.114 \pm 0.002$
France	$10^2-5 \times 10^4$	$5.03 \pm 0.01$	-8.7	$0.345 \pm 0.008$
Germany	$10^2-5 \times 10^3$	$5.95 \pm 0.05$	-8.6	$0.315 \pm 0.007$
Guatemala	$10^2-10^4$	$6.60 \pm 0.02$	-36.4	$0.075 \pm 0.001$
Ireland	$10^3-10^4$	$3.43 \pm 0.02$	-5.5	$0.295 \pm 0.005$
Israel	$10^2-5 \times 10^3$	$5.81 \pm 0.04$	-7.6	$0.252 \pm 0.005$
Italy	$10^2-5 \times 10^4$	$4.97 \pm 0.01$	-6.2	$0.399 \pm 0.011$
Japan	$10^2-10^4$	$7.40 \pm 0.02$	-48.2	$0.082 \pm 0.001$
Netherlands	$10^2-10^4$	$5.31 \pm 0.03$	-9.7	$0.272 \pm 0.006$
Oman	$10^2-10^5$	$4.78 \pm 0.01$	-29.9	$0.108 \pm 0.002$
Peru	$10^2-10^5$	$5.18 \pm 0.01$	-18.9	$0.162 \pm 0.003$
Portugal	$10^2-5 \times 10^3$	$5.32 \pm 0.05$	-4.1	$0.319 \pm 0.006$
Russia	$10^2-5 \times 10^4$	$6.75 \pm 0.01$	-5.4	$0.319 \pm 0.006$
Saudi Arabia	$10^2-5 \times 10^4$	$4.60 \pm 0.01$	-8.6	$0.224 \pm 0.004$
Spain	$10^2-10^4$	$6.54 \pm 0.04$	-3.3	$0.385 \pm 0.009$
Switzerland	$10^2-10^4$	$5.44 \pm 0.03$	-7.2	$0.294 \pm 0.006$
Turkey	$10^2-5 \times 10^4$	$3.78 \pm 0.01$	-1.8	$0.591 \pm 0.015$
United Kingdom	$10^2-5 \times 10^4$	$6.52 \pm 0.01$	-8.1	$0.247 \pm 0.005$
Mean value		$5.57 \pm 0.98$	$-14 \pm 13$	$0.25 \pm 0.12$
Median value		5.48	-8.6	0.252

The SIA model introduced here is probably the most simple way of getting at the logistic equation, similar to what was observed in China. This is only intended to work during lockdowns or with effective social policies.  $S$  stands for susceptible,  $I$  stands for been infected in the past, and  $A$  stands for avoiders. Avoiders are good with hygiene, staying away from sick people, and wearing masks not to get infected. Using avoiders is similar to neglecting recoveries in the SIR model with a different population size.

$$S + I + A = P \quad S + I = P - A = K.$$

Then

$$\frac{dI}{dt} = \mu IS = \mu I(K - I).$$

One can use  $r = \mu K$  which gives

$$\frac{dI}{dt} = rI \left(1 - \frac{I}{K}\right).$$

This derivation gives a logistic curve with reduced carrying capacity than the entire population, but does not suggest a way how to estimate  $K$  *a priori*.

In reality, there are probably different categories of people, depending on a complicated way of social interaction networks and behavior that cannot easily be quantified. This model provides a logistic curve, but many SIR model situations are also a similar form. Often in SIR, the effective value of  $K$  is of the same order of magnitude as  $P$ , whereas in SIA, one can easily achieve  $K \ll P$ , which

is different. Once restrictions are eased, then the curve would restart by not considering the previous cases anymore with new parameters in a possible second wave. Also, if there is significant travel between countries, one would expect this model to breakdown.

### 3.2. China and logistic growth of disease spread

The SIA model may describe the logistic growth of positive cases in China. The reported cases follow a smooth curve once the correction for chest x-ray diagnoses is considered [13]. After the synchronization procedure, one finds in figure 1 the following parameter values.

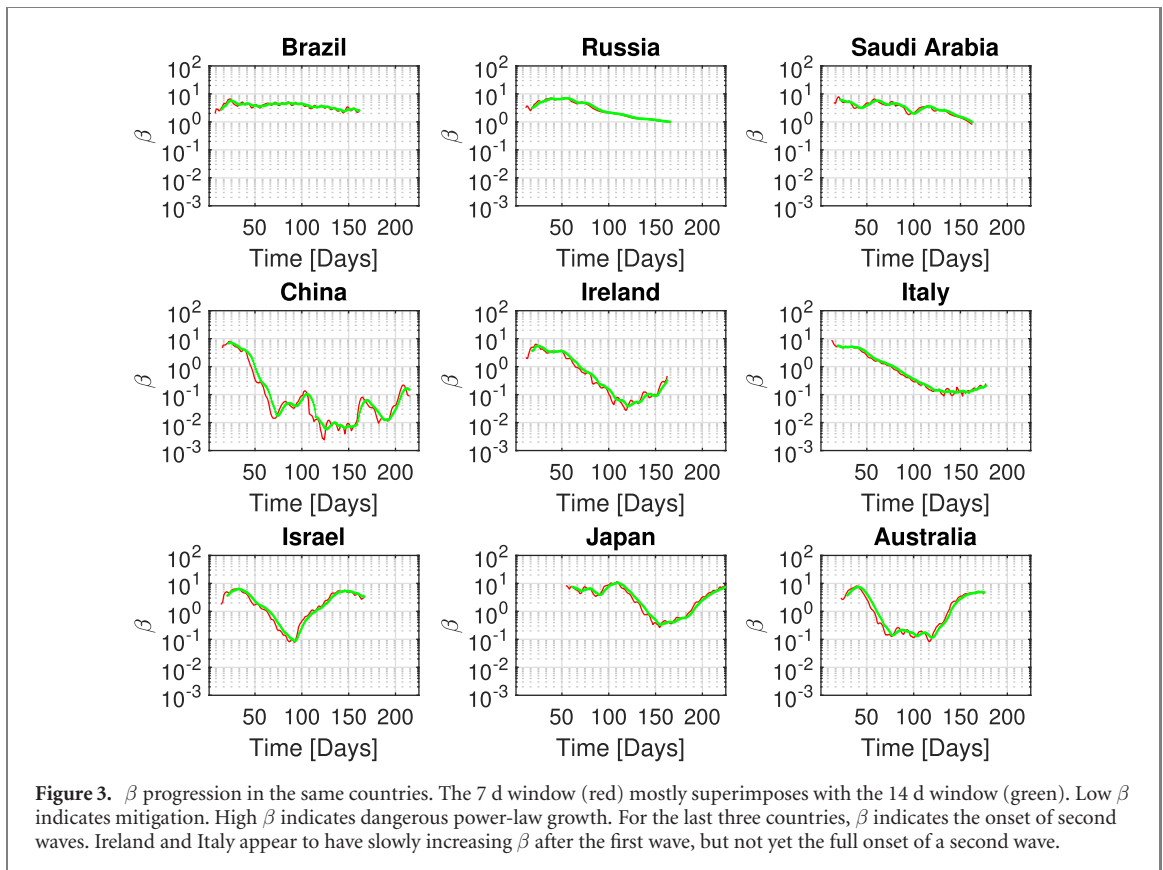
$$K = 82\,646 \pm 92 \text{ people} \quad r = 0.219 \pm 0.003 \text{ d}^{-1}$$

$$\tau = 34.26 \pm 0.08 \text{ d.}$$

The fit seems reasonable on order 200 d. The genetic algorithm Eureqa also found a logistic curve for the simplest best fit function when running overnight. Most of the increase happens during a 15–20 d period, which may be why neglecting recoveries in SIR is reasonable. China is very strict on travel, but new cases trickle in at later times.

### 3.3. Representative analysis of nine countries

The number of positive cases was investigated with four analyses: early time exponential synchronization,



overall power-law growth, the Bhatia visualization, and the  $\beta$  visualization. Nine representative countries are shown here in the results and all 27 in the supplementary materials (<https://stacks.iop.org/PB/17/065005/mmedia>). Brazil, Russia, and Saudi Arabia are examples of countries that are primarily still in the power-law phase or just exiting it. China, Ireland, and Italy have progressed to the logistic phase. Israel, Japan, and Australia are examples of a second wave after the logistic phase without a saturation.

Figure 2 illustrates the early time exponential phase and power-law growth. In no case does COVID-19 transmit like a pure runaway exponential function to the entire population. The power-law phase extends the longest for Brazil over many powers of 10 because they were slow to implement effective policies. The initial synchronization is not always possible in a satisfactory way due to a lack of data or irregularities. The mean initial exponential growth rate is about  $0.25 \pm 0.11$  per day for 27 countries. An overall power-law exponent was computed for the region of cases that appears linear on the synchronized log–log plot in table 1. The mean overall power-law exponent for all 27 countries is  $5.57 \pm 0.98$  suggesting there is a similarity in the cases' trajectories after synchronization.

Figure 3 illustrates the  $\beta$  visualization for the same nine countries. Using a 7 d (red) or 14 d (green) window does not lead to much difference. The  $\beta$  always starts high then decreases. One can see the second

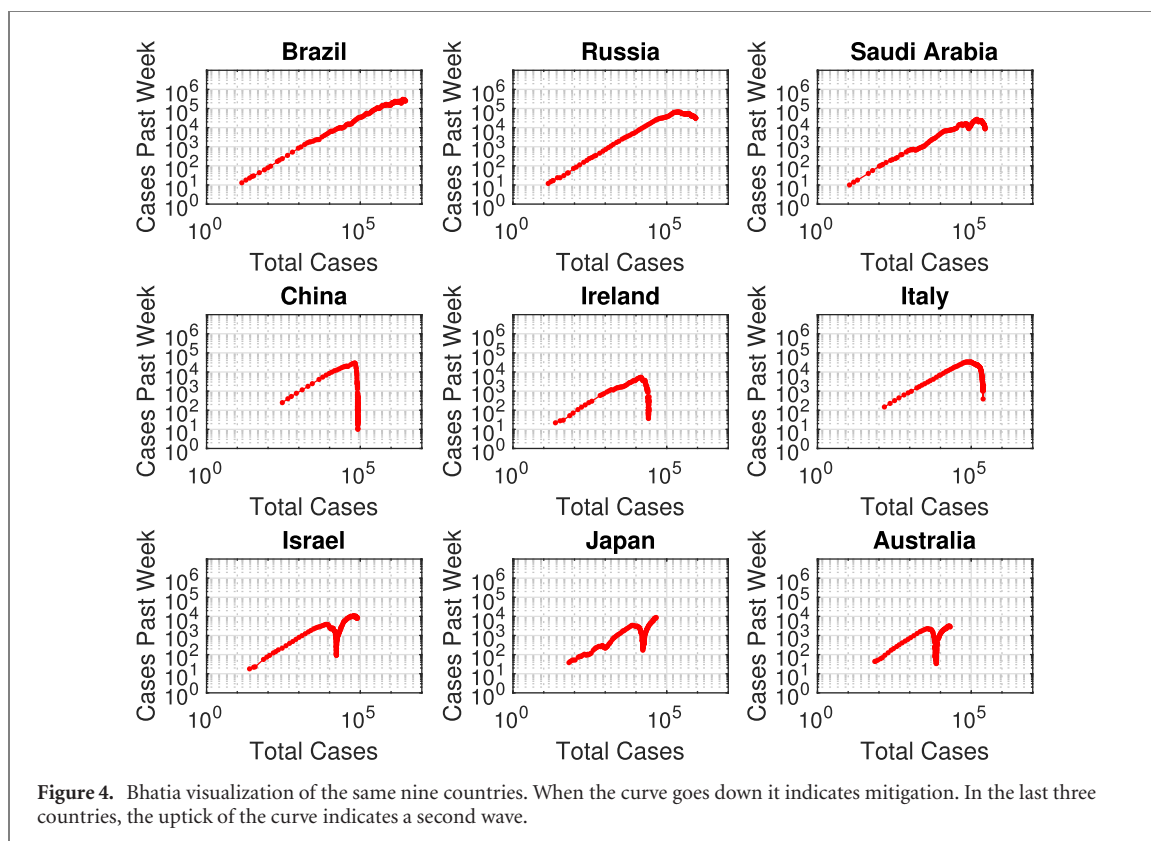
wave as a decrease in the power-law exponent that later increases significantly above one. China seems to avoid a second wave due to strict policies, but second waves remain to be seen in several more months for most other countries.

Figure 4 illustrates the Bhatia visualization. There is no time parameter on the plots, but one sees the second wave as a drop then an increase in the curve. Most countries fall along the same initial line during the power-law phase. If the reopening of countries is only a delay in the progression, the curve will rise back up and continue along the same initial line. The  $\beta$  visualization is noisier than the Bhatia visualization because it depends explicitly on time.

#### 4. Discussion

The status of the COVID-19 pandemic is better understood by how new cases change over time rather than the absolute totals. The Bhatia and  $\beta$  visualizations are independent indicators of where countries stand and progression into possible second waves. The optimistic SIA model leading to saturating logistic growth only appears to work while lockdowns and strict social measures are maintained like in China. Every time restrictions change significantly in a region, new model parameters, and a reset of the initial conditions are required.

In many affected countries, a pattern of exponential growth, power-law growth, logistic growth, and possible second waves emerges from this



survey. About half the hotspot countries above 10 000 cases follow a more irregular pattern that is difficult to model. Early data irregularities in reporting and noise make it often hard to synchronize countries, but an intermediate power-law growth exponent of about 5.6 appears similar between countries over that phase. Since countries besides China have similar policies, it seems likely that most countries who had COVID-19 under control will potentially see at least a reduced second wave because they act similarly. Travel restrictions should be an important consideration.

There are issues with choosing positive cases or deaths to track COVID-19. Usually, only sick people receive tests. Some people are positive but exhibit mild or no symptoms, so never seek out a test. Many regions may only have a limited throughput of testing, but it does not seem to be a bottleneck in most countries. Positive test data requires the same standards and protocols over time to be reliable. Positive testing results must also be made publicly available and not falsified to be useful. Deaths occur delayed from the original infections, and there are lower totals than positive cases. As infrastructure becomes overburdened, death rates can increase. Death rates also depend on demographics and individuals' general quality of health. Whether considering positive cases or death rates, scientists should keep their options open in tracking COVID-19.

It appears there is no simple social policy allowing most countries to go back to complete normality.

The current pressing issues are how prevalent reinfection can be, the future availability of vaccines, the development of drugs that stop death, and what to do about future waves. Concerning vaccines, questions arise about efficacy, their mass production, delivery, or even if everyone will accept to use them. Modeling suggests that the number of actual positive cases has been significantly underestimated, but is well below 50 percent [6]. Therefore, there is still more modeling to be done and science to do to save lives.

## Acknowledgments

I would especially like to thank Michael Sixt for encouraging me to think about these problems while working at home due to restrictions in place. I want to thank Nick Barton, Katka Bodova, Matthew Robinson, Simon Rella, Federico Sau, Ivan Prieto, and Pradeep Kumar for useful discussions.

## ORCID iDs

Jack Merrin  <https://orcid.org/0000-0001-5145-4609>

## References

- [1] Brauer F 2017 *Mathematical epidemiology: past, present, and future* *Infect. Dis. Model.* **2** 113–27
- [2] Brauer F and Castillo-Chavez C 2012 *Mathematical Models in Population Biology and Epidemiology* vol 2 (Berlin: Springer)

- [3] Kermack W O and McKendrick A G 1927 A contribution to the mathematical theory of epidemics *Proc. R. Soc. A* **115** 700–21
- [4] Hethcote H, Ma Z and Liao S 2002 Effects of quarantine in six endemic models for infectious diseases *Math. Biosci.* **180** 141–60
- [5] Numberophile Ben Sparks The coronavirus curve <https://youtube.com/watch?v=k6nLfCbAzgo>
- [6] Ferguson N *et al* 2020 Report 9: Impact of Non-pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand *Imperial College COVID-19 Response Team* (London: Imperial College London) (<https://doi.org/10.25561/77482>)
- [7] Harko T, Lobo F S N and Mak M K 2014 Exact analytical solutions of the susceptible-infected-recovered (SIR) epidemic model and of the SIR model with equal death and birth rates *Appl. Math. Comput.* **236** 184–94
- [8] Stier A, Berman M and Bettencourt L 2020 COVID-19 attack rate increases with city size *medRxiv* (<https://doi.org/10.1101/2020.03.22.20041004>)
- [9] Wu K, Darcet D, Wang Q and Sornette D 2020 Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world (arXiv:2003.05681)
- [10] Schmidt M and Lipson H 2009 Distilling free-form natural laws from experimental data *Science* **324** 81–5
- [11] Bhatia A Trajectory of COVID-19 confirmed cases <https://aatishb.com/covidtrends/>
- [12] Ziff A L and Ziff R M 2020 Fractal kinetics of COVID-19 pandemic *medRxiv* (<https://doi.org/10.1101/2020.02.16.20023820>)
- [13] Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, Qian T, Sun Z and Xia L 2020 Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases *Radiology* **296** 200642
- [14] Merrin J 2017 *Introduction to Error Analysis (The Science of Measurements, Uncertainties, and Data Analysis)* (Scotts Valley, CA: CreateSpace)
- [15] Taylor J 1997 *Introduction to Error Analysis (The Study of Uncertainties in Physical Measurements)* (Sausalito, CA: University Science Books)
- [16] Bevington P and Robinson D K 2002 *Data Reduction and Error Analysis for the Physical Sciences* (New York: McGraw Hill)
- [17] Hughes I and Hase T 2010 *Measurements and Their Uncertainties: A Practical Guide to Modern Error Analysis* (Oxford: Oxford University Press)
- [18] Dong E, Du H and Gardner L An interactive web-based dashboard to track COVID-19 in real time *Lancet Inf. Dis.* **20** 533–4