



## Research article

## Some insights on the COVID-19 pandemic from Fisher information

Heriberto Cabezas<sup>a,\*</sup>, Hrvoje Štefančić<sup>b</sup><sup>a</sup> University of Miskolc, Miskolc, Hungary<sup>b</sup> Catholic University of Croatia, Zagreb, Croatia

## A B S T R A C T

We explored the application of Fisher information to the study of pandemics and illustrated the insights that can be gained using the COVID-19 pandemic, as a test case. To do so, we applied the Fisher information theory previously applied to periodic systems, to non-periodic dynamic systems. The resulting mathematical machinery was then used to compute the Fisher information measure, as the amount of information extracted from the time series for COVID-19 confirmed infections and deaths. The analysis was performed for the World as a whole and five nation-states: India, USA, Japan, Germany, and Chile. Several insights resulted from the study: (1) the information content of the time series varied widely for different time periods, over the course of the pandemic, (2) it is advisable not to fit model parameters or make policy decisions based on data from time periods with low Fisher information, (3) the most information about a wave of infections comes towards the end of the wave where the time series data has the most information about the dynamics of the pandemic, and (4) the quality of the time series data significantly affects the Fisher information value, and, therefore, what can be learned from studying the time series.

## 1. Introduction

The recent, still ongoing pandemic of COVID-19 disease has immensely influenced the modern world. In a highly interconnected world, the pandemic developed quickly, produced a healthcare crisis of unprecedented dimensions, endangered the functioning of economies, and brought measures and restrictions to which citizens of many countries in the World were completely unaccustomed.

Modeling of epidemic dynamics is often used to guide the formulation of counter-epidemic measures. These measures frequently restrict and shape the lives of entire nations and the functioning of entire economies, and, hence, the reliable forecasting of key epidemic numbers has become of central social and healthcare importance. Mathematical epidemiology [1] has suddenly received much public attention, and various models of mathematical epidemiology have been proposed and applied. These models can have very different levels of complexity, mathematical sophistication, and data acquisition methodologies [2].

Models of epidemic propagation can be primarily categorized according to the process of infection in an individual and the mechanisms of contagion between individuals [3,4]. The course of infection in individuals is usually modeled by discrete states, usually referred to as compartments [5]. As the infection evolves, each individual transits between compartments, e.g., exposed to infected. Some examples of compartments in epidemic models are susceptible (S), exposed (E), infected (I), recovered/removed (R), etc. Individuals can then be categorized into one of the subpopulations depending on the state (compartment) that they belong to. Mechanisms of contagion define how individuals in different compartments get in contact, enabling disease propagation. Homogeneous mixing models assume an equal probability of contact between all individuals belonging to different compartments [6]. Network-based models assume that social contacts that facilitate disease spreading have a graph structure presentable in the form of a complex network [7–9]. Metapopulation models combine homogeneous mixing with network structure, e.g. using network structure to describe disease spreading by air travel between cities and homogeneous mixing to describe spreading within cities [10].

\* Corresponding author.

E-mail address: [heribertocabezas@gmail.com](mailto:heribertocabezas@gmail.com) (H. Cabezas).

<https://doi.org/10.1016/j.heliyon.2024.e26707>

Received 30 November 2022; Received in revised form 14 February 2024; Accepted 19 February 2024

Available online 20 February 2024

2405-8440/Â© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Approaches to epidemic spreading relying on agent-based models (ABM), e.g. Ref. [11], simulate the motion of a large ensemble of individuals where infection can be transferred when predefined conditions are met, e.g. when individuals come close enough [12]. The dynamics of spreading need to be complemented by the processes happening at the level of an individual, such as recovery from the disease. Some frequently used compartmental models are SI, SIS, SIR, SEIR models, etc.

Epidemiological models contain parameters that encode properties of the disease such as its ease of spreading or rate of recovery of infected individuals. In most epidemiological models these parameters are taken as constant. However, models with fixed parameters prove to be inadequate for a long-term realistic description of the COVID-19 pandemic. Namely, the course of a pandemic is decisively affected by the complex and adaptive social response, counter-epidemic measures varying spatially (geographically) and temporally, and frequent mutations of the SARS-CoV-2 virus, just to name some important factors. In such circumstances, model parameters need to be frequently recalibrated and models with variable (or even stochastic) parameters would be a more adequate description. The dynamics of parameters contain cases of sudden change (such as changing mandatory counter-epidemic measures), or gradual variation (such as growth in incidence of a novel virus strain). Such variability of the environment in which the disease spreading takes place, encoded in (variable) model parameters, results in multiple waves of pandemic propagation.

In such unstable conditions, a legitimate question is what is the level of epidemic control that the society (together with all its institutions and services) actually has. A high level of control would correspond to (quickly) falling epidemic numbers and efficient functioning of counter-epidemic measures. On the other hand, a low level of control would correspond to a situation of (fast) growth of epidemic numbers and inefficient or non-existent counter-epidemic measures. The level of control is frequently identified (or at least well approximated) with the amount of available information. A model-independent measure telling us how much we currently know, i.e. which information we currently possess, about the epidemic system that we observe is, therefore, both theoretically interesting and practically needed. In this paper, we argue that Fisher information is such a measure. We further compute the Fisher information for five countries and the World, present the variability of Fisher information with the course of the pandemic, and analyze some of its properties.

Fisher information has so far been successfully applied to the analysis of ecological systems [13–17], sustainability [18–23], regime shifts in dynamic systems [24–28], the socio-political system of nation-states [24,29], and other areas [30,31]. A common property of studied systems in previous applications of Fisher information is their periodicity. In this paper we apply the concept of Fisher information to epidemic systems that do not have this property, i.e., they are not periodic. In that respect, this paper brings novelty in two directions: (i) application of Fisher information to non-periodic systems and (ii) quantification of the information content of epidemic data and the analysis of its temporal development.

The first section of this paper is the introduction in which we outline the main idea of the application of Fisher information to epidemic systems. The second section starts with the definition of Fisher information and elaborates on a specific approach to its calculation. In the third section, we display epidemic data for the World and five countries (India, United States of America, Japan, Germany, and Chile). In the fourth, section the results of Fisher information calculation are presented. The fifth section is devoted to the discussion of results and the sixth section closes the paper with a summary and conclusions.

## 2. Fisher information

### 2.1. Origins and interpretation

The information function now widely known as Fisher information was derived by the British mathematical statistician Ronald A. Fisher [32]. For a relation of Fisher information and its associated measure with other information theory quantifiers, see Martins et al. [33]. The Fisher information  $I(\theta)$  is formally a measure of the amount of information about an unknown parameter  $\theta$  obtainable from observations of a random variable  $s$ . Formally, the Fisher information is defined by,

$$I(\theta) \equiv \int \frac{1}{p(s|\theta)} \left[ \frac{\partial p(s|\theta)}{\partial \theta} \right]^2 ds \quad (1)$$

where  $p(s|\theta)$  is the conditional probability density of observing  $s$  in the presence of  $\theta$ . Note that if the random variable  $s$  contains no information about  $\theta$ , then the partial derivative of  $p(s|\theta)$  with respect to  $\theta$  is zero, and the Fisher information  $I(\theta)$  is zero as well. Hence, the choice of  $s$  in practice is dependent on what  $\theta$  represents.

The definition represented by Eq. (1), however, is difficult to use, particularly for the kinds of problems that we address in this work. The reason is that evaluating the partial derivative of  $p(s|\theta)$  with respect to an unknown parameter  $\theta$  is often not possible in practice. Fortunately, one can invoke the concept of shift-invariance to derive a simplification [34]. Hence, we first define a new variable  $\hat{s} \equiv s - \theta$  and then note that  $p(s|\theta) \equiv p_o(s-\theta|\theta) = p_o(\hat{s})$ . The first equality defines a new probability  $p_o$  law for the fluctuations of  $s - \theta$  for an observation conditional on  $\theta$ . The second equality states that for a shift-invariant system, the fluctuations  $s - \theta$  are independent of the value of  $\theta$ . This essentially states that  $p(s|\theta)$  is equal to a probability density function  $p_o(\hat{s})$  of the same form which does not depend on  $\theta$ . Lastly,  $ds = d\hat{s}$  since  $\theta$  is a constant parameter. Where again  $p_o(\hat{s})$  is a new probability density function. Using these results, we can then transform the expression in Eq. (1) into,

$$I = \int \frac{1}{p_o(\hat{s})} \left[ \frac{dp_o(\hat{s})}{d\hat{s}} \right]^2 d\hat{s} \quad (2)$$

where  $p_0(\hat{s})$  is the simple probability density for observing a value of  $\hat{s}$ . However, since the subtraction of  $\theta$  from  $s$  does not change the calculation of the Fisher information  $I$ , for the rest of this paper we tacitly use  $s$  in place of  $\hat{s}$  and  $p(s)$  in place of  $p_0(\hat{s})$ . Note that the Fisher information  $I$  is no longer dependent on the parameter  $\theta$ , and that the derivative in the integrand is the slope with respect to the variable  $s$  of the simple probability density function.

Further, Eq. (2) allows us to build an interpretation of Fisher information in terms of tangible physical concepts. Conceptually, a highly ordered system is one where the state of the system over time is not changing or changing very little, and the state of the system is, therefore, predictable with low uncertainty. As an illustration, consider that multiple measurements of a variable  $s$  for a highly ordered system will yield nearly identical values  $s$  with every observation. In such a case, a plot of  $p(s)$  versus  $s$  is a very sharp curve around the mean value  $\langle s \rangle$  of  $s$ . Then the derivative  $dp(s)/ds$  has a very high value, and the Fisher information is correspondingly high. Conversely, in the case of a highly disordered system, the state of the system changes greatly over time. In this other case, the state of the system is unpredictable or at least difficult to predict. Here, repeated measurements of the same variable  $s$  will yield values that are widely spread. A plot of  $p(s)$  versus  $s$  is then a nearly flat curve, and the derivative  $dp(s)/ds$  has a low value. The corresponding Fisher information is then low as well. Hence, the form of Fisher information represented by Eq. (2) is a measure of order, e.g., highly ordered systems have high Fisher information, and highly disordered systems have low Fisher information.

It should also be noted that Eq. (2) along with the Principle of Extreme Physical Information [34] has been used to derive many of the known laws of nature from a common principle of information transfer. This formally confirms the concept that the laws of nature are in fact present in the data from observations. In this work, however, we did not attempt to find general laws, rather we simply observed that Fisher information could offer insights into the dynamics of the COVID-19 pandemic.

### 2.2. Computing Fisher information

The next step in the development of a theory based on Fisher information is to pose the existence of a time series of  $n$  directly observable properties which adequately characterize the system. These observations  $rx_1(t), rx_2(t), \dots, rx_n(t)$  which we call “raw” data consist of measurable properties of the system. For a physical system, these could be mass, temperature, velocity, etc. In the case of a pandemic, these could be the number of infections, deaths, recoveries, etc. The  $rx_1(t), rx_2(t), \dots, rx_n(t)$  do not in general have the same dimensions, and it becomes necessary to make them nondimensional as we will see later. The method we selected was to simply divide the values of each observation  $rx_i(t)$  by the value of the average of raw observations in the time series  $\langle rx_i \rangle$  to obtain  $x_i(t)$ , e.g.  $x_i(t) = rx_i(t)/\langle rx_i \rangle$ .

We then identify the variable  $s$  as the state of the system. In a complex system,  $s$  is a function of many independent observable and dimensionless variables  $x_i$  where  $i = 1, 2, \dots, n$ , and then  $s = s(x_1, x_2, \dots, x_n)$ . If there is a time series of observations of the system, the variables  $x_i(t)$  are functions of time  $t$  and  $s(t) = s(x_1(t), x_2(t), \dots, x_n(t))$  as well. One can then pose that  $s(t)$  represents a point at a time  $t$  in a linear space where the coordinates are the variables  $x_i(t)$  and time. In this case, the collection of observations defines a trajectory of  $s(t)$  defined by  $x_1(t), x_2(t), \dots, x_n(t)$  and  $t$ . The probability of observing a particular state of the system  $s$  is proportional to the time  $\Delta t|_s$  that the system spends in the state,

$$p(s) = \Delta t|_s / T \tag{3a}$$

where in Eq. (3a),  $p(s)$  is the probability density for observing the system in state  $s$ , and  $T$  is a characteristic time frame further discussed below. In the limit of many small states of the system, we can approximately replace the sequence of states and the sequence of time intervals with a continuous function, and then  $\Delta t|_s = dt|_s$ . To obtain Eq. (3b), we further pose that  $\Delta t|_s$  is equal to the size  $\delta s$  of state  $s$  divided by the mean speed tangential to the system path  $\langle ds/dt|_s \rangle$  while traversing the state. In the limit of small ( $\delta s$ ) states of the system, we can approximately replace the mean speed with the instantaneous speed at state  $s$  to get,

$$p(s) = \frac{1}{T} \frac{\delta s}{\left[ \frac{ds}{dt} \right]_s} \tag{3b}$$

This expression essentially says that the slower the speed while traversing a particular state, the longer the system will be in the state, and the more likely that the system will be observed in that particular state.

Finally, in a linear space, the tangential speed  $ds/dt|_s$  can be computed using the Euclidean metric to obtain Eq. (3c).

$$\left[ \frac{ds}{dt} \right]_s = \sqrt{\sum_{i=1}^n \left[ \frac{dx_i}{dt} \right]_s^2} \tag{3c}$$

where  $dx_i/dt$  is the first derivative of the dimensionless variables  $x_i$  with respect to time for the observable variables of the system. Using Eq. (3), we then develop a computable expression for the probability density function  $p(s)$  given in Eq. (4).

$$p(s) = \frac{\delta s}{T} \left[ \sqrt{\sum_{i=1}^n \left[ \frac{dx_i}{dt} \right]_s^2} \right]^{-1} \tag{4}$$

Further, it is also possible to use the chain rule to express Eq. (2) as an integral over time rather than states of the system. The resulting expression after some manipulation is,

$$I = \int \frac{1}{p(s)} \left[ \frac{dp(s)}{dt} \right]^2 \left( \frac{ds}{dt} \right)^{-1} dt \tag{5}$$

One remaining issue with Eqs. (4) and (5) is the size of the state  $\delta s$ . The parameter  $\delta s$  represents the lower limit of observability, i.e., two hypothetical states that are spaced at less than  $\delta s$  from each other are not distinguishable. It is, therefore, a property of the measuring process used to conduct the observation, but it is not a property of the system. Given a specific measuring process, it seems reasonable to assume that  $\delta s$  is a constant. In cases where the absolute value of Fisher information is important, the actual value of  $\delta s$  would be important as it appears as a constant multiplier in Eq. (4). However, in the cases that we will address in this work, we are only interested the relative value of Fisher information where the actual value of  $\delta s$  is not critical. We, therefore, tacitly set  $\delta s = 1$ .

The results from Eq. (5) can be used to derive after further manipulation a working expression for computing Fisher information of the form,

$$I(t_j) = \frac{1}{T} \int_{t_j-T/2}^{t_j+T/2} \frac{(d^2s/dt^2)^2}{(ds/dt)^4} dt \tag{6}$$

where  $T$  is again a characteristic time frame as discussed below, and where the derivatives  $ds/dt$  and  $d^2s/dt^2$  are computed from Eqs. (7a) and (7b).

$$\frac{ds}{dt} = \sqrt{\sum_{i=1}^n \left( \frac{dx_i}{dt} \right)^2} \tag{7a}$$

$$\frac{d^2s}{dt^2} = \left( \frac{ds}{dt} \right)^{-1} \sum_{i=1}^n \frac{dx_i}{dt} \frac{d^2x_i}{dt^2} \tag{7b}$$

Eqs. (6) and (7) are the working expressions for computing the Fisher information from time series for the system variables  $x_i$ . If a mathematical model exists, the derivatives can be computed analytically. Otherwise, they must be approximated numerically.

In practical terms,  $T$  is the characteristic time associated with the phenomena for which we desire to obtain the value of the Fisher information. For a periodic system,  $T$  has to be equal to at least one system period, but letting  $T$  be equal to several periods is often preferable. In the case of a non-periodic system, assigning a value to  $T$  can be more ambiguous. However, at least an estimate of the characteristic time frames for the system can often be deduced. In the case of the COVID-19 pandemic, for example, one short-term phenomenon is the rise and resolution of infections which often occurs over a period of approximately three weeks. Here, we have set  $T \approx 21$  days. Conversely, if the interest is in the seasonal course of the pandemic, then we would set  $T \approx 90$  days, about three months.

### 2.3. Fisher information and order (or randomness)

To further illustrate the relationship between Fisher information and order, consider the Fisher information computed from data for two independent variables  $x_1$  and  $x_2$  which follow functions of the following form as shown in Eqs. (8a) and (8b).

$$x_1 = 1000 - \left[ 20 \frac{i + 2 - \langle i \rangle}{\langle i \rangle} \right]^2 - R(0, r) \tag{8a}$$

$$x_2 = 1000 - \left[ 5 \frac{i + 2 - \langle i \rangle}{\langle i \rangle} \right]^4 - R(0, r) \tag{8b}$$

where  $i = 0, 1, 2, \dots, 23$ ,  $\langle i \rangle$  is the arithmetic mean of  $i$ ;  $R(0, r)$  is a random function that varies between 0 and  $r$ , and  $r = 1, 10, 100, 1000$ . We computed the Random Function using the standard random function available with the Microsoft Excel spreadsheet. Here  $i$  represents a generic dynamic parameter, for example, time. We then compute the value of the Fisher information for all values of  $i$  under the four separate values of  $r$ . It should be noted that we are using randomness as a proxy for disorder. The reason is that the

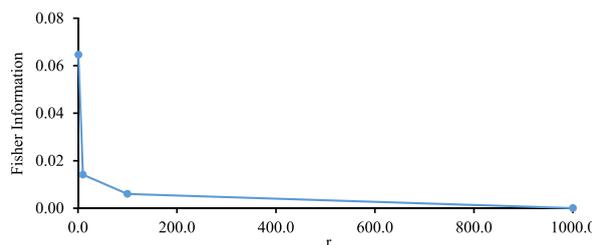


Fig. 1. shows the Fisher information computed from  $\langle i \rangle_{x_1}$  and  $x_2$  as a function of the randomness parameter  $r$  where  $T = 21$  days.

observable variables of orderly systems follow a pattern, perhaps driven by an ordering dynamic parameter, and that makes the behavior of the system generally predictable. We then show the Fisher information for one discrete case of the random function  $R(0,r)$  as shown in Fig. 1. Note that every time that  $R(0,r)$  is invoked, it generates a different sequence of random numbers between 0 and  $r$ . However, in general, the Fisher information decreases with increasing randomness as shown here.

### 3. The COVID-19 pandemic

#### 3.1. Trajectory at the global and national scale

Figs. 2–7 show the time trajectory of daily confirmed cases and daily deaths for the World, India, the United States of America, Japan, Germany, and Chile. All plots in these figures have a double y-axis where the scale on the left-hand y-axis refers to daily cases and the scale on the right-hand y-axis refers to daily deaths. The choice of countries includes a reasonably wide range of national population sizes, geographic location, and cultural backgrounds. The daily confirmed cases and daily deaths were computed by subtracting the total or cumulative cases or deaths and from those of the previous day. The actual expression for daily confirmed cases  $DC_i$  and daily deaths  $DD_i$  for day  $i$  are given by Eqs. (9a) and (9b).

$$DC_i = C_{i+1} - C_i \tag{9a}$$

$$DD_i = D_{i+1} - D_i \tag{9b}$$

where  $C_i$  and  $C_{i+1}$ , and  $D_i$  and  $D_{i+1}$  are the cumulative number of confirmed cases and deaths in days  $i$  and  $i + 1$  respectively. The primary data used in these calculations was taken from the online data repository of Johns Hopkins University in the United States of America [35]. These data were chosen because the authors deemed them to be gathered using a consistent unified methodology, reasonably accurate, and widely available.

#### 3.2. Observations

The first observation is that at both the global and the national scale, the number of confirmed infections and deaths appear to be correlated, with the deaths often slightly trailing the infections as expected, i.e., when the number of infections rises, the number of deaths usually rises later. The second observation is that the “waves” of infections and deaths are not well correlated when comparing different nations. Compare, for example, the “waves” in Fig. 3 for India to those for the United States of America in Fig. 4. The “waves” for Japan in Fig. 5 are also not well correlated to those for Germany in Fig. 6. This is to be expected because the COVID-19 pandemic did not simultaneously affect all nation-states, and the strategies of the public health measures taken also across the globe differ from nation to nation. It is also notable that on average, the number of daily cases and deaths is not highly correlated with the population of the country. For example, India has approximately four times the population of the USA, but the range of the number of daily infections and deaths is qualitatively similar. Japan has about six times the population of Chile, but the range of the number of daily cases and deaths was not very different. Also, relatively extreme events such as that for Germany around 20 November 2021 were not evident at the same magnitude elsewhere. As expected, the numbers of daily cases and deaths at the global scale are the sum of the respective numbers for all nations across the globe, including the five presented in this analysis.

At the global scale, daily death rates rose steeply with daily infections near the start of the pandemic and with succeeding waves of infections. However, the most recent data seems to point to a decoupling of daily death rates from daily infections. While the authors are not certain why this is happening, we can speculate that more widespread immunity (herd immunity) in the human population due to vaccinations and immunity resulting after recovery from infection may play a role. Additionally, it is also possible that more transmissible, yet clinically milder virus strains are becoming more prevalent in the population.

Lastly, one should note that the symptoms of COVID-19 infections can mimic those of other upper respiratory diseases, and it is

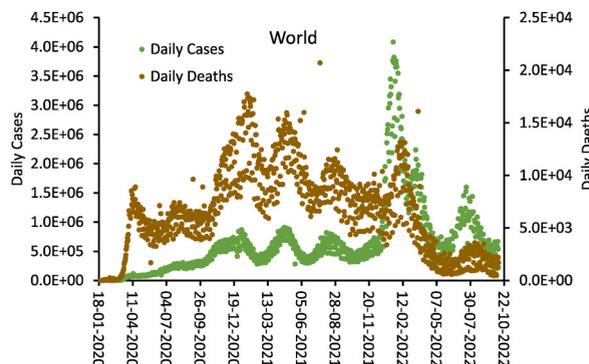


Fig. 2. Daily confirmed cases and daily deaths for COVID-19 across the World over the course of the COVID-19 pandemic up to 8 October 2022.

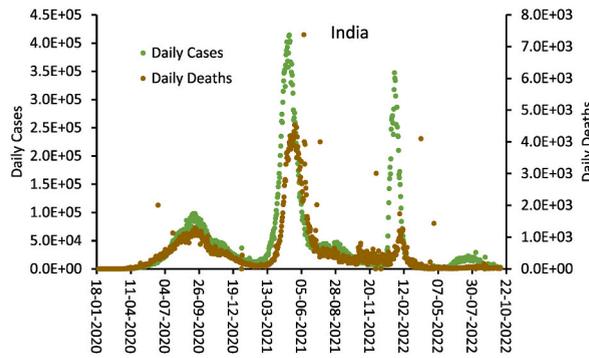


Fig. 3. Daily confirmed cases and daily deaths for COVID-19 in India over the course of the COVID-19 pandemic up to 8 October 2022.

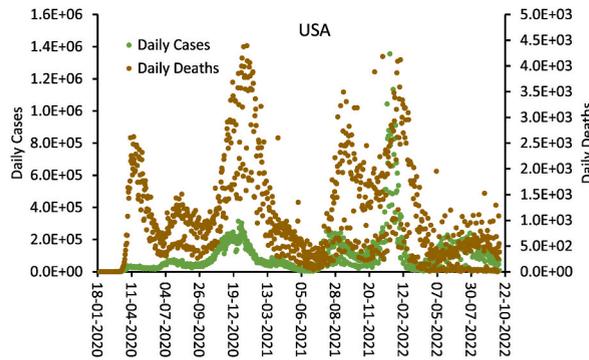


Fig. 4. Daily confirmed cases and daily deaths for COVID-19 in the United States of America over the course of the COVID-19 pandemic up to 8 October 2022.

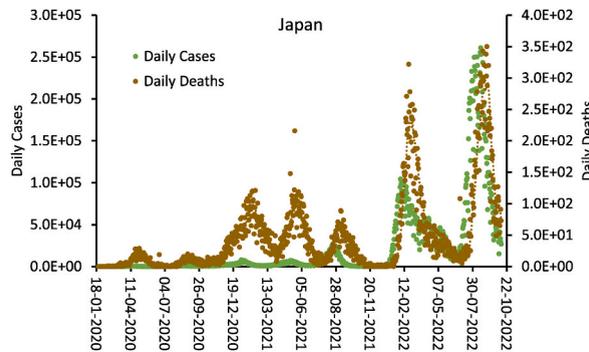


Fig. 5. Daily confirmed cases and daily deaths for COVID-19 in Japan over the course of the COVID-19 pandemic up to 8 October 2022.

possible that some COVID-19 cases may have been missed. However, the authors believe that the numbers of confirmed infections and deaths are reasonably accurate. This is supported by the wide experience of the medical profession with the COVID-19 pandemic and the relatively wide availability of tests for COVID-19 infection.

#### 4. Results from Fisher information calculations

For conducting Fisher information calculations related to the COVID-19 pandemic, we chose  $T = 21$  days or three weeks. The reason is that it normally takes about three weeks for a typical infection to go from initial infection to resolution by either recovery or death. This implies, however, that any given Fisher information value represents the information content of the data for the three-week period over which the calculation was performed. Note also, that we assigned the resulting value of the Fisher information to the middle of the three-week period, and it represents, consequently, the information in the data for the three weeks. The Fisher information at the global and national scale results are shown in Figs. 8–13.

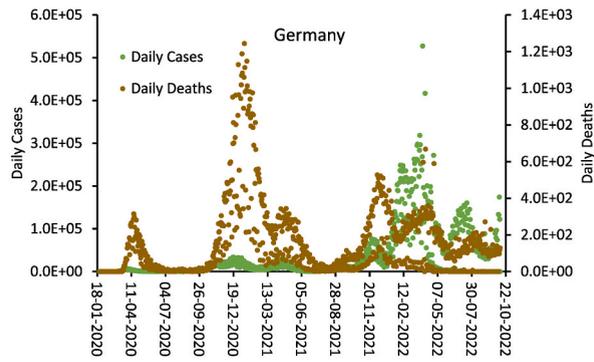


Fig. 6. Daily confirmed cases and daily deaths for COVID-19 in Germany over the course of the COVID-19 pandemic up to 8 October 2022.

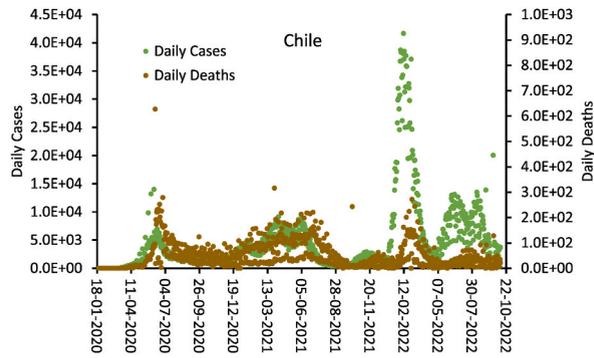


Fig. 7. Daily confirmed cases and daily deaths of COVID-19 in Chile over the course of the COVID-19 pandemic up to 8 October 2022.

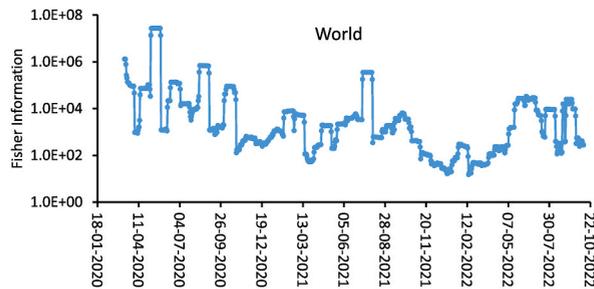


Fig. 8. The Fisher information computed from the time series for confirmed infections and deaths at the global scale over overlapping periods of three weeks ( $T = 21$  days) with the Fisher information value assigned to the middle of the three-week period. The plot is a semi-log to the base 10.

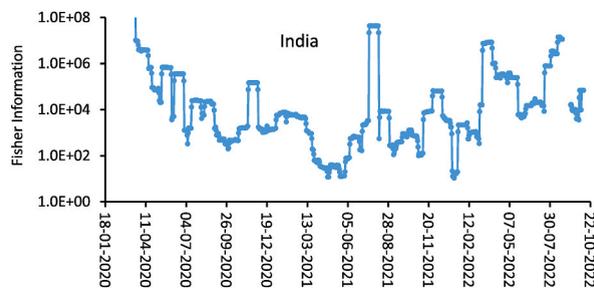
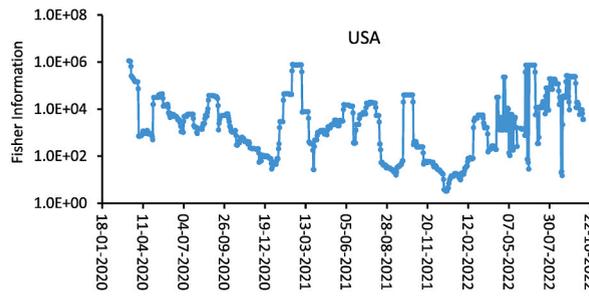
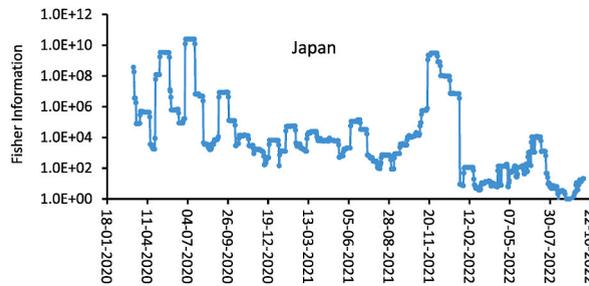


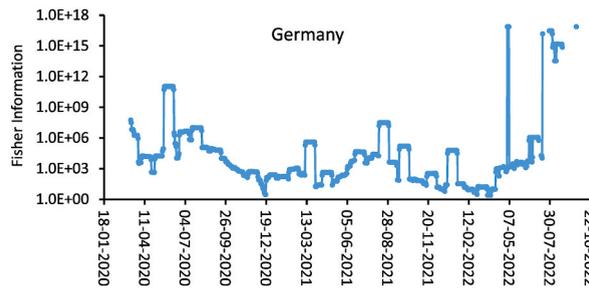
Fig. 9. The Fisher information computed from the time series for confirmed infections and deaths at the national scale for India over overlapping periods of three weeks ( $T = 21$  days) with the Fisher information value assigned to the middle of the three-week period. The plot is a semi-log to the base 10.



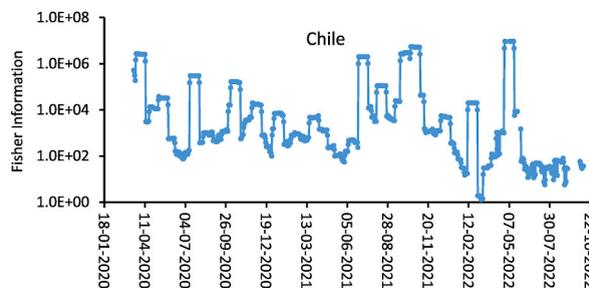
**Fig. 10.** The Fisher information computed from the time series for confirmed infections and deaths at the national scale of the USA over overlapping periods of three weeks ( $T = 21$  days) with the Fisher information value assigned to the middle of the three-week period. The plot is a semi-log to the base 10.



**Fig. 11.** The Fisher information computed from the time series for confirmed infections and deaths at the national scale of Japan over overlapping periods of three weeks ( $T = 21$  days) with the Fisher information value assigned to the middle of the three-week period. The plot is a semi-log to the base 10.



**Fig. 12.** The Fisher information computed from the time series for confirmed infections and deaths at the national scale of Germany over overlapping periods of three weeks ( $T = 21$  days) with the Fisher information value assigned to the middle of the three-week period. The plot is semi-log to the base 10.



**Fig. 13.** The Fisher information computed from the time series for confirmed infections and deaths at the national scale of Chile over overlapping periods of three weeks ( $T = 21$  days) with the Fisher information value assigned to the middle of the three-week period. The plot is a semi-log to the base 10.

## 5. Results and discussion

### 5.1. Global pandemics

The World in the early 21st Century is highly connected, perhaps a hyperconnected network. People and physical goods can be moved across the globe by air travel at speeds far greater than any seen in previous ages in history. Economic information and general news can be transmitted from any place to any other place on Earth at near the speed of light through the World Wide Web. It is not surprising then that a global event such as the SARS CoV-2 (COVID-19) impacted in some way almost everyone and almost everywhere very quickly. These are fast-moving events. One of the proposed contributions of this work is to help clarify the time frame necessary for assessing the dynamics of a pandemic such as the COVID-19 pandemic, and to help identify when optimal information can be extracted from the primary observational data of infections and deaths.

### 5.2. Interpretation of Fisher information results

An interpretation of Fisher information as an information measure of the level of control can be nicely demonstrated in several prominent examples. One would expect that at the end of a wave, before a new wave has started, we have the largest degree of control. By that point, we have accumulated the greatest amount of information on the biological properties of the disease and ways and measures how it can be controlled and restrained along with the response of the human population. The biggest epidemic wave in India happened in the spring of 2021, see Fig. 3., and by July 2021 it was largely over. As one can see in Fig. 9 in July 2021 Fisher information for India peaked at a high level. Another example is the end of the first epidemic wave in Germany in May 2020, see Fig. 6., which is accompanied by a peak in Fisher information in Fig. 12.

The results for the World and the United States of America display numerous peaks in Fisher information, but they are of less prominent height. This fact may be related to the composite nature of results for epidemic numbers for these relatively large geographical areas. Namely, the results for the World are a combination of national results where countries may adopt different strategies in the combat against the pandemic or new virus strains may reach them at different times. A similar situation is present in the United States of America where states have adopted different responses to the pandemic. Furthermore, these two geographical areas are spatially big and contain areas of very different demographic properties such as population density. In both these cases, epidemic waves are also less prominent.

It is worth noting that Fisher information tends to reach a “peak” towards the end of a pandemic wave of infections and deaths as already mentioned. This implies that the most information about the dynamics of the pandemic only becomes available towards the end of the wave. However, it should also be noted that the information content of the data quickly tends to collapse to a low value shortly after the end of a wave. The implications here are that most of the information needed to analyze the pandemic dynamics or to fit a model parameter is contained in the data between the start and the end of a wave. One would also caution about making major policy changes based on pandemic data with low Fisher information.

Lastly, there are a few discontinuities in the plots of Fisher information, e.g., a point all by itself or a gap between two Fisher information trajectories. These are due to numerical problems arising from data quality issues. A few examples of these issues are the daily numbers for new cases or new deaths being zero on several consecutive days, and then appearing to be unexpectedly high on the day after. Given the challenges imposed by the COVID-19 pandemic, such errors should not be unexpected. These data issues cause various problems such as giving undefined values of Fisher information due to a division by zero and other problems. However, despite these difficulties, the authors feel that they could still show that it is possible to conduct an insightful analysis using Fisher information. In fact, it is important to demonstrate that the analysis can be conducted with actual and realistic field data.

## 6. Summary and conclusions

In this paper, we extend to non-periodic systems the concept of Fisher information that has been previously applied to (quasi) periodic systems. As a contemporary example of major interest, we calculate Fisher information for the ongoing pandemic of COVID 19. Using the data for daily numbers of newly infected and deceased individuals from the Johns Hopkins database [35], we calculate Fisher information as a function of time for the World and five geographically, culturally, demographically, and economically different countries: The United States of America, India, Germany, Japan, and Chile. It is found that Fisher information exhibits intensive dynamics where it acquires values differing by many orders of magnitude. In particular, the Fisher information peaks after the end of an epidemic wave for all studied geographic regimes where epidemic waves are clearly present. The behavior of Fisher information also shows different patterns for different nation-states, e.g., India and Germany, where epidemic control measures were applied relatively homogeneously. There are also unique patterns for areas such as the USA or the World where measures were applied more heterogeneously. Finally, it is argued that Fisher information is a measure of the level of control that a society has in the battle with the pandemic because information is a prerequisite to the formulation and application of epidemic control measures.

In conclusion, the extension of the Fisher information concept used in this work from periodic to non-periodic systems seems reasonable in that it can produce useful results. However, more research is required to establish the general validity of our approach to non-periodic systems. The Fisher information values, calculated from the COVID-19 empirical time series data for various countries and regions, show that the Fisher information can vary by many orders of magnitude and that periods of (very) high Fisher information exist. Finally, Fisher information is found to be a good measure of the level of control over the pandemic process.

## Data availability statement

The data that support the findings of this study are available on request from the corresponding author.

## CRediT authorship contribution statement

**Heriberto Cabezas:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **Hrvoje Štefančić:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Hrvoje Stefančić reports financial support was provided by Catholic University of Croatia. Heriberto Cabezas reports a relationship with University of Miskolc Research Institute of Applied Earth Sciences that includes: employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

The research of HS presented in this paper is supported by the European Regional Development Fund under the grant KK.01.1.1.01.0009 (DATACROSS).

The research contribution by HC was carried out in the GINOP-2.3.2-15-2016-00010 framework “Development of enhanced engineering methods with the aim at utilization of subterranean energy resources” project at the Research Institute of Applied Earth Sciences of the University of Miskolc, the Széchenyi 2020 Plan, partially funded by the European Union, cofinanced by the European Structural and Investment Funds.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e26707>.

## References

- [1] F. Brauer, *Mathematical epidemiology: past, present, and future*, *Infectious Disease Modelling* 2 (2017) 113–127.
- [2] J. Bedson, et al., A review and agenda for integrated disease models including social and behavioural factors, *Nat. Human Behav.* 5 (2021) 834–846.
- [3] R.M. Anderson, R.M. May, *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press, 1991. ISBN: 9780198540403.
- [4] M.J. Keeling, P. Rohani, *Modeling Infectious Diseases*, Princeton Univ. Press, 2008. ISBN 978-0-691-11617-4.
- [5] W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics, *Proc. Roy. Soc. Lond.* 115 (1927) 700–721.
- [6] S. Bansal, B.T. Grenfell, L. Ancel Meyers, When individual behavior matters: homogeneous and network models in epidemiology, *Journal of the Royal Society Interface* 4 (2007) 879–891.
- [7] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Complex networks: structure and dynamics, *Phys. Rep.* 424 (2006) 175–308.
- [8] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, A. Vespignani, Epidemic processes in complex networks, *Rev. Mod. Phys.* 87 (2015) 925–980.
- [9] A. Vespignani, Modelling dynamical processes in complex socio-technical systems, *Nat. Phys.* 8 (2012) 32–39.
- [10] V. Colizza, R. Pastor-Satorras, A. Vespignani, Reaction–diffusion processes and metapopulation models in heterogeneous networks, *Nat. Phys.* 3 (2007) 276–282.
- [11] N. Hoertel, et al., A stochastic agent-based model of the SARS-CoV-2 epidemic in France, *Nat. Med.* 26 (2020) 1417–1421.
- [12] M. Tizzoni, P. Bajardi, A. Decuyper, G. Kon Kam King, C.M. Schneider, V. Blondel, Z. Smoreda, M.C. González, V. Colizza, On the use of human mobility proxies for modeling epidemics, *PLoS Comput. Biol.* 10 (2014) e1003716.
- [13] B.D. Fath, H. Cabezas, C.W. Pawłowski, Regime changes in ecological systems: an information theory approach, *J. Theor. Biol.* 222 (4) (2003) 517–530.
- [14] B.D. Fath, H. Cabezas, Exergy and Fisher information as ecological indices, *Ecol. Model.* 174 (1–2) (2004) 25–35.
- [15] A.L. Mayer, C.W. Pawłowski, H. Cabezas, Fisher information and dynamic regime changes in ecological systems, *Ecol. Model.* 195 (1–2) (2006) 72–82.
- [16] A.L. Mayer, C. Pawłowski, B.D. Fath, H. Cabezas, Applications of Fisher information to the management of sustainable environmental systems, in: *Exploratory Data Analysis Using Fisher Information*, Springer, London, 2007, pp. 217–244.
- [17] E.S. Rawlings, J.C. Barrera-Martínez, V. Rico-Ramírez, Fisher information calculation in a complex ecological model: an optimal control-based approach, *Ecol. Model.* 416 (2020) 108845.
- [18] H. Cabezas, B.D. Fath, *Towards a Theory of Sustainable Systems*, vol. 194, Fluid phase equilibria, 2002, pp. 3–14.
- [19] V. Rico-Ramírez, P.A. Quintana-Hernández, J.A. Ortiz-Cruz, S. Hernández-Castro, Fisher information: a generalized sustainability index?, in: *Computer Aided Chemical Engineering* (Vol. 25, Pp. 1155–1160) Elsevier, 2008.
- [20] T. Eason, H. Cabezas, Evaluating the sustainability of a regional system using Fisher information in the San Luis Basin, Colorado, *J. Environ. Manag.* 94 (1) (2012) 41–49.
- [21] A. González-Mejía, L. Vance, T. Eason, H. Cabezas, Recent developments in the application of Fisher information to sustainable environmental management, *Assessing and measuring environmental impact and sustainability* (2015) 25–72.
- [22] L. Vance, T. Eason, H. Cabezas, An information theory-based approach to assessing the sustainability and stability of an island system, *Int. J. Sustain. Dev. World Ecol.* 22 (1) (2015) 64–75.
- [23] Y. Wang, J.E. Taylor, M.J. Garvin, Measuring the resilience of human–spatial systems to disasters: framework combining spatial-network analysis and Fisher information, *J. Manag. Eng.* 36 (4) (2020) 04020019.
- [24] A.T. Karunanithi, A.S. Garmestani, T. Eason, H. Cabezas, The characterization of socio-political instability, development, and sustainability with Fisher information, *Global Environ. Change* 21 (1) (2011) 77–84.

- [25] A.M. Gonzalez-Mejia, T. Eason, H. Cabezas, M.T. Suidan, Computing and interpreting Fisher Information as a metric of sustainability: regime changes in the United States air quality, *Clean Technol. Environ. Policy* 14 (5) (2012) 775–788.
- [26] T. Eason, A.S. Garmestani, H. Cabezas, Managing for resilience: early detection of regime shifts in complex systems, *Clean Technol. Environ. Policy* 16 (4) (2014) 773–783.
- [27] E. König, H. Cabezas, A.L. Mayer, Detecting dynamic system regime boundaries with Fisher information: the case of ecosystems, *Clean Technol. Environ. Policy* 21 (7) (2019) 1471–1483.
- [28] T. Eason, W.C. Chuang, S. Sundstrom, H. Cabezas, An information theory-based approach to assessing spatial patterns in complex systems, *Entropy* 21 (2) (2019) 182.
- [29] L. Vance, T. Eason, H. Cabezas, M.E. Gorman, Toward a leading indicator of catastrophic shifts in complex systems: assessing changing conditions in nation-states, *Heliyon* 3 (12) (2017) e00465.
- [30] D. Spichak, A. Kupetsky, A. Aragonese, Characterizing complexity of non-invertible chaotic maps in the Shannon–Fisher information plane with ordinal patterns, *Chaos, Solit. Fractals* 142 (2021) 110492.
- [31] L.R. Moreno-Torres, A. Gómez-Vieyra, M. Lovallo, A. Ramírez-Rojas, L. Telesca, Investigating the interaction between rough surfaces by using the Fisher–Shannon method: implications on the interaction between tectonic plates, *Phys. Stat. Mech. Appl.* 506 (2018) 560–565.
- [32] R.A. Fisher, On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, Series A, Containing Papers of a Mathematical or Physical Character* 222 (1922) 309–368, <https://doi.org/10.1098/rsta.1922.0009>.
- [33] A.M. Martins, L.H. Fernandes, A.D. Nascimento, Scientific progress in information theory quantifiers. *Chaos, Solitons & Fractals* 170 (2023) 113260.
- [34] B.R. Frieden, *Science from Fisher Information: A Unification*, Cambridge Univ. Press, 2004. ISBN 0-521-00911-1.
- [35] Johns Hopkins University. [https://github.com/CSSEGISandData/COVID-19/tree/master/csse\\_covid\\_19\\_data](https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data). Accessed on 8 October 2022..