Measuring COVID-19 spreading speed through the mean time between infections indicator

Gabriel Pena, Verónica Moreno, and Néstor Ruben Barraza
Universidad Nacional de Tres de Febrero

Abstract

Objectives: To introduce a novel way of measuring the spreading speed of an epidemic.

Methods: We propose to use the mean time between infections (MTBI) metric as obtained from a recently introduced non-homogeneous Markov stochastic model. Different types of parameter calibration are performed. We estimate the MTBI using data from different time windows and from the whole stage history and compare the results. In order to detect waves and stages in the input data, a preprocessing filtering technique is applied.

Results: The results of applying this indicator to the COVID-19 reported data of infections from Argentina, Germany and the United States are shown. We find that the MTBI behaves similarly with respect to the different data inputs, whereas the model parameters completely change their behaviour. Evolution over time of the parameters and the MTBI indicator is also shown.

Conclusions: We show evidence to support the claim that the MTBI is a rather good indicator in order to measure the spreading speed of an epidemic, having similar values whatever the input data size.

Keywords: COVID-19, Non-homogeneous Markov, Contagion, Filtering, Mean time between infections.

Introduction

The COVID-19 outbreak, declared pandemic in 2020, attracted the attention of scientists from different domains (biologists, physicists, engineers and mathematicians, among others). The urgent need to control, predict and monitor the disease progress made it essential to count with mathematical models and algorithms (see for example [8]), both to manage data on infections and deaths and to perform calculations and predictions. Not only were the well known SIR (Susceptible - Infectious - Removed) compartmental model and its derivations vastly applied [2, 4, 13–15, 18, 24, 27] but also many new models [1] and AI algorithms [9, 11, 26, 32] were proposed. One is the stochastic model presented in [3] and [20], which we call the BPM model. It is based on a contagion Markov model described by a non-homogeneous birth process (NHBP). The authors were able to obtain the functional form of the cumulative infection cases and deaths curves, which have a sub-exponential shape, which is a behaviour previously pointed out for epidemics [6, 12, 20, 31]. The model has two parameters: one that represents the power of the outbreak and another that models the immunization rate. The expected time between events can be obtained as a standard calculation in counting stochastic processes [23]. Thus, useful indicator to measure the outbreak speed is obtained: the mean time between infections (MTBI).
This indicator expresses the spreading speed, since less time between successive infections implies a more rapid disease spread. Therefore, it is expected that during the initial stage of a disease wave the MTBI indicator decreases as the peak gets closer. As we will show, taking the model parameters individually does not suffice to gain insight on the progress of the epidemic, but the MTBI actually does, which makes it a quite robust epidemiological indicator. Thus, the MTBI can be used to evaluate the impact of the actions taken by health institutions.

To calibrate the model parameters, we fit the data of the total reported cumulative infection cases to the mean value function of the process. Since this function does not have concavity changes, the data curve needs to be split into sections (which we will call stages) with a fixed concavity, and then the model can be separately applied to each of them. To achieve this, we look for the local maximums and minimums of the daily data curve, which cannot be directly observed due to sharp variations in the reported data. These high frequency jumps are commonly associated to noise in signal processing. Thus, we apply a filtering routine to smooth the daily data curve and then a standard maximum and minimum detection algorithm. Once the stages are properly separated, we fit the model for some specific stages in Argentina, Germany and the United States by the same method described in [3] to obtain the MTBI indicator which depends on the model parameters.

An interesting result of the filtering algorithm is the comparison between the filtered curves of daily cases and daily deaths, which shows that both have rather similar shapes, with a short time delay in between. This behaviour was observed in the three considered countries.

In [3], the parameter estimation at a given time is performed using the whole history of the corresponding stage. A natural question is whether it is correct to use the full history, or just some part of it. Different approaches to this question were recently presented in the context of the SIR model [5, 7]. In all of the proposed cases there exists some arbitrariness regarding the choice of the data used to calibrate the models. In this work, we perform a comparative study between the possibilities of using different data inputs, for example using the whole stage history or short time windows of a certain number of previous days. We show statistically that the MTBI does not change significantly, whereas the model parameters do. Evolution over time of the MTBI and those parameters is also shown. The most important result obtained is that the MTBI indicator is invariant with respect to the size of the data used for calibration, which lets us conclude that the MTBI is a robust indicator with respect to the data used to estimate it.
Materials and methods

The BPM model

In this work we consider a particular model based on a non-homogeneous birth process (NHBP) which was recently proposed [3]. As it is known, NHBPs are a particular class of continuous time Markov processes [16, 29] that model the growth of a population where individuals can only be born. In a NHBP, the probability of having \( r \) individuals in a population at a given time \( t \), \( P_r(t) \) is given by Definition 1 as the solution of a recursive system of ordinary differential equations [10]:

**Definition 1.** Let \( P_r(t) \), \( r \in \mathbb{N}_0 \) denote the probability of having \( r \) individuals at time \( t \). Then, for an initial population of 0 individuals we have \( P_0(0) = 1 \) and \( P_r(0) = 0 \) \( \forall r > 0 \), whereas for \( t > 0 \) the pmf \( P_r(t) \) is defined recursively as the solution of each differential equation in the system of ODEs,

\[
P'_r(t) = -\lambda_r(t)P_r(t) + \lambda_{r-1}(t)P_{r-1}(t), \quad r > 0, \\
P'_0(t) = -\lambda_0(t)P_0(t),
\]

(1)

where \( \lambda_r(t) \) can depend on both \( t \) and \( r \). The function \( \lambda_r(t) \) is called event rate or intensity function.

The model considered here can be obtained as a particular case of the more general models presented in [16, 17, 25]. This particular process is governed by the following event rate:

\[
\lambda_r(t) = \rho \frac{1 + \frac{\gamma}{\rho} r}{1 + \rho t},
\]

(2)

where \( \gamma, \rho > 0 \) are the model parameters. \( \gamma \) measures the power of the spreading disease and \( \rho \) determines the immunization rate.

Let \( M(t) \) be the mean number of individuals of the population (mvf) for the model. In [3], the authors prove that:

\[
M(t) = \frac{\rho}{\gamma} \left[ (1 + \rho t)^{\frac{\gamma}{\rho}} - 1 \right].
\]

(3)

Quite interesting observations arise from Eq. (3). The mvf is a power of \( t \) with exponent \( \gamma/\rho \); hence, the form of the function may be quite different depending on the value of this ratio. There are three possible scenarios:

- **Case 1:** \( \frac{\gamma}{\rho} < 1 \). The curve is concave.
- **Case 2:** \( \frac{\gamma}{\rho} = 1 \). Eq. (3) reduces to that of the Polya-Lundberg process, the curve is a straight line.
- **Case 3:** \( \frac{\gamma}{\rho} > 1 \). The curve is convex.

3
A natural consequence of this is that this process can model a large variety of different scenarios, including those which were accurately described by the Polya-Lundberg process [19]. As shown in [3], the $M(t)$ function fits quite well the cumulative number of either infections or deaths from the COVID-19 pandemic, which allows estimating the parameters $\gamma$ and $\rho$ using the reported data by standard fitting techniques like Least Squares.

An important indicator useful to measure the spreading speed is obtained from the model: the expected elapsed time between the occurrence of an event and the next one, which corresponds to the Mean Time Between Infections (MTBI) or the Mean Time Between Deaths (MTBD) depending on whether the considered population consists of the infected cases or deaths, respectively. In this section we use the general notation MTBE (Mean Time Between Events). For this particular model, the formula

$$MTBE(t) = \frac{1}{\rho} \frac{1 + \rho t}{(1 + \rho t)^{\gamma} - 1}$$

(4)

gives the mean time between the event that occurred at time $t$ and the next [3].

Filtering algorithm

Since the mvf curve does not have inflection points, the model needs to be applied separately in each stage having either positive or negative concavities, which correspond to the stages (early/mitigation) of each outbreak wave. This implies that a criterion for the wave and stage separation needs to be defined. We choose a rather classic approach:

- Local minimums in the daily data report represent time instants where a wave finishes and a new wave starts.
- Local maximums in the daily data report represent time instants where, within a single wave, the initial stage ends and the mitigation stage begins.

The main difficulty regarding these definitions is that the measured data (daily cases and daily deaths) is extremely noisy. This implies that finding local extrema by direct observation is not straightforward. To overcome this problem we perform a filtering procedure to suppress the noise and smooth the curve, followed by a maximum and minimum detection algorithm. In this work we considered Argentina, Germany and the United States as our study cases, but the same results shown here can be achieved for any other dataset by properly setting the filter hyper-parameters.

Some interesting observations regarding the noise can be made by looking at the data spectrum. Fig. [1] depicts the absolute value of the Discrete Fourier Transform (DFT) of the daily reported cases in the three mentioned countries. Significant peaks can be clearly distinguished at frequency $1/7$ and its second harmonic (and its third in the United States case), which is explained by the “once per week” periodic variations in the testing amount on weekends, which
are known to be lower. An analogous behaviour can be seen in the deaths dataset. These observations evidence the need of a preprocessing, namely a low-pass filtering, in order to suppress measurement noise. Consequently, we use moving-average (MA) filters, also known as sliding windows, which are the simplest low-pass FIR filters; see [28] for technical details about MA filters. We proposed to apply \( n \) MA filters of length \( L \) in series; \( L \) and \( n \) are the algorithm hyper-parameters.

![Spectrum](image1.png)

**Figure 1:** Spectrum of the daily cases data, as obtained by applying the FFT algorithm. Frequency axis has 1/day units, and is shown up to 1/2, the highest observable frequency (due to the data being sampled at 1 datum per day).

The reported cases from Argentina, Germany and the United States and their respective filtered curves are shown in Fig. 2. Hyper-parameters were chosen not without some arbitrariness, and are not the same in the three cases. Since we want to remove the 1/7 frequency component at least, \( L \) must be chosen to be \( \geq 7 \). In order to avoid harming the signal too much, we chose to fix \( L = 7 \) and increase \( n \) as much as necessary, achieving a nearly Gaussian behaviour [28] without suppressing useful signal frequencies.

Having a smooth curve, detecting maximums and minimums can be done by straightforward methods. It must be remarked that the algorithm is not completely automatic: proper hyper-parameter values must be carefully chosen by a human user, and since filtering is not perfect, in some cases human inter-
pretation is necessary to determine which minimums and maximums are “false” and which are not.

In Fig. 3 we show the filtered curves corresponding to the daily cases and daily deaths in a single plot in order to compare them. It is interesting to note that the cases and deaths curves have similar shapes, with the deaths curve delayed by a short period (approximately two weeks) as expected due to the length of the infectious period. This also shows that the propagation speed of deaths is directly related to the propagation speed of infections. However, due to the existence of measurement noise, they do not always exhibit the same amount of waves.

Results

Once the data is properly segmented into sections, the propagation speed of the disease can be assessed by analyzing the evolution over time of the MTBI indicator within each of them. In order to do this, we are going to fit the model’s mvf (Eq. 3) to the cumulative infection cases, recalibrating the parameters for each time $t$, obtaining this way two estimates $\hat{\gamma}(t)$ and $\hat{\rho}(t)$. Replacing these on
Figure 3: Daily cases (blue, scale shown on the left side axis) and deaths (red, scale shown on the right side axis) filtered curves. Filled circles indicate local maximums; × marks indicate local minimums.

Eq. 4 yields:

$$\overline{MTBI}(t) = \frac{1}{\hat{\rho}(t)} \frac{1 + t \cdot \hat{\rho}(t)}{(1 + t \cdot \hat{\rho}(t))^{\frac{\gamma(t)}{\hat{\gamma}(t)}} - 1},$$

which gives the evolution over time of the MTBI indicator estimate.

The fittings were performed by the nonlinear Least-Squares method, using the Levenberg-Marquardt algorithm implemented in the Python language; the source code provided by the authors of [3], where they implemented the BPM model, is available at [22]. The experiments were done using the real (unfiltered) data, since the cumulative curves do not change significantly after filtering. Datasets were obtained from [21].

A natural question arises when fitting the model: is the whole history of a stage significant, at any given time instant, or should a shorter time period be considered instead. This question has already been addressed in the context of the SIR model [5, 7]. We are going to perform a comparative study of the results obtained by calibrating the model parameters (for every time index \( t \) within a stage) using the whole history of the stage up to \( t \) and those obtained by considering different time windows prior to the respective \( t \). As it is customary
in signal processing, we call this procedure “windowing”. For the analysis, we considered the following study cases:

- Argentina: initial stage of the first wave, from March 03, 2020 to October 17, 2020.
- Germany: initial stage of the fourth wave, from July 02, 2021 to November 08, 2021.
- United States: initial stage of the third wave, from September 07, 2020 to December 30, 2020.

![Figure 4](image)

**Figure 4:** Evolution over time of the averaged MTBI estimate and its variance for the chosen stage in each country. In all the study cases and with all the considered input datasets the model achieves an $R^2$ coefficient above 0.97. For the three cases considered, only the corresponding stage for Germany has a vaccination process, rising from 35 to 50 % over that period. The consequently sudden slow down in the speed propagation can be seen in the picture. Due to the population size and social mobility (between 20 and 30 % under normal) in the considered stage of the United States, the propagation speed is much higher than in the other two study cases. In the case of Argentina, the strict lockdown imposed during the first wave of the epidemic is reflected on high MTBI values at the beginning of the first wave (i.e., a slow propagation).

It can also be seen in the three plots of Fig. 4 that the MTBI curves do not change significantly in neither shape or scale when the data is subject to windowing, i.e., the MTBI indicator is invariant. This shows that the MTBI efficiently captures the behaviour of the model, and moreover it is robust to the length of the dataset used for calibration. Therefore, we propose to use the average of several estimations of the MTBI obtained through time windows of different sizes as our final MTBI estimator.
Figure 5: Evolution over time of $\rho$ (upper row) and $\gamma/\rho$ (lower row) parameters for each of the three countries in its chosen stage. Green curves show the parameters calculated using the whole stage history. Red curves show the indicators calculated using a time window of 30 days. Since the $\rho$ numerical values are rather different on each case, the green curve scale is shown in the left side axis, whereas the red curve scale is shown in the right side axis for the three $\rho$ plots.

As it can be seen in Fig. 5, the obtained $\rho$ and $\gamma/\rho$ parameters are indeed very different according to the estimation method used in each case. The resulting $\rho$ curve changes in both shape and scale (i.e., the numerical values have different orders of magnitude, which is why we plot each curve on its own $y$ axis), as it can be seen in Figs. 5a - 5c. The $\gamma/\rho$ curve changes in shape, but the numerical values do not differ much (the pairs of curves in Figs. 5d - 5f are plotted in the same scale); this implies that even if $\gamma$ and $\rho$ change individually between the two methods, they do it proportionally. This behaviour shows that computing the model parameters individually, either $\rho$ or $\gamma/\rho$, do not provide enough information about the epidemic, whereas the MTBI does.

Discussion / Conclusions

In this work we have analyzed the evolution over time of the MTBI indicator given by a recently proposed stochastic non-homogeneous Markov model when applied to the COVID-19 pandemic, which is useful to measure the propagation speed of the epidemic. Three countries were considered as study cases: Argentina, Germany and the United States. For each of these countries, a particular stage was chosen and the model was fitted using different sets of input data. On the one hand, the full history of the stage until the day $t$ was con-
sidered to calibrate the model parameters at that day, and on the other hand, the calibration at \( t \) was made using only the data from several time windows of different lengths. We show the evolution over time of the averaged MTBI indicator, as well as the evolution of the model parameters. The most important conclusion was that the MTBI indicator (which depends on those parameters) does not change significantly using different sizes of input data, which shows this is a robust indicator independently of the dataset size.

References

[1] Ban Ghanim Al-Ani. Statistical modeling of the novel covid-19 epidemic in iraq. Epidemiologic Methods, 10(s1):20200025, 2021.

[2] Ishtiaq Ali and Sami Ullah Khan. Analysis of stochastic delayed sirs model with exponential birth and saturated incidence rate. Chaos, Solitons & Fractals, 138:110008, 2020.

[3] Néstor Rubén Barraza, Gabriel Pena, and Verónica Moreno. A non-homogeneous Markov early epidemic growth dynamics model. Application to the SARS-CoV-2 pandemic. Chaos, Solitons & Fractals, 139:110297, 2020.

[4] Zhongwei Cao, Yuee Shi, Xiangdan Wen, Huishuang Su, and Xue Li. Dynamic behaviors of a two-group stochastic sirs epidemic model with standard incidence rates. Physica A: Statistical Mechanics and its Applications, 554:124628, 2020.

[5] Guillermo Capobianco, Romina Cobiaga, Walter Reartes, and Franco Turpaud. The SIR model in the COVID-19 pandemic. In VIII Congreso de Matemática Aplicada, Computacional e Industrial, pages 715–718, 2021.

[6] Gerardo Chowell, Lisa Sattenspiel, Shweta Bansal, and Cécile Viboud. Mathematical models to characterize early epidemic growth: A review. Physics of Life Reviews, 18:66 – 97, 2016.

[7] Ermanno Cordelli, Matteo Tortora, Rosa Sicilia, and Paolo Soda. Time-window SIQR analysis of COVID-19 outbreak and containment measures in Italy. In 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), pages 277–282. IEEE, 2020.

[8] Anne Cori, Neil M Ferguson, Christophe Fraser, and Simon Cauchemez. A new framework and software to estimate time-varying reproduction numbers during epidemics. American Journal of Epidemiology, 178(9):1505–1512, 2013.

[9] Ramon Gomes da Silva, Matheus Henrique Dal Molin Ribeiro, Viviana Cocco Mariani, and Leandro dos Santos Coelho. Forecasting brazilian and american covid-19 cases based on artificial intelligence coupled with climatic exogenous variables. Chaos, Solitons & Fractals, 139:110027, 2020.
[10] W. Feller. *An Introduction to Probability Theory and Its Applications, Vol. 1.* Wiley, 3rd edition, January 1968.

[11] Athanassios S Fokas, Nikolaos Dikaos, and George A Kastis. Covid-19: predictive mathematical formulae for the number of deaths during lockdown and possible scenarios for the post-lockdown period. *Proceedings of the Royal Society A*, 477(2249):20200745, 2021.

[12] Tapiwa Ganyani, Christel Faes, and Niel Hens. Inference of the generalized-growth model via maximum likelihood estimation: A reflection on the impact of overdispersion. *Journal of Theoretical Biology*, 484, 2020.

[13] James P Gleeson, Thomas Brendan Murphy, Joseph D O’Brien, Nial Friel, Norma Bargary, and David JP O’Sullivan. Calibrating covid-19 susceptible-exposed-infected-removed models with time-varying effective contact rates. *Philosophical Transactions of the Royal Society A*, 380(2214):20210120, 2022.

[14] Norden E. Huang, Fangli Qiao, Qian Wang, Hong Qian, and Ka-Kit Tung. A model for the spread of infectious diseases compatible with case data. *Proceedings of the Royal Society A*, 477, 2021.

[15] William Ogilvy Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.

[16] S. A. Klugman, H. H. Panjer, and G. E. Willmot. Loss models: Further topics. John Wiley & Sons, 2013.

[17] Hidetoshi Konno. On the exact solution of a generalized Pólya process. *Advances in mathematical physics*, 2010, 2010.

[18] Yingfen Liu, Yan Zhang, and Qingyun Wang. A stochastic sir epidemic model with lévy jump and media coverage. *Advances in Difference Equations*, 2020, 12 2020.

[19] O. Lundberg. On Random Processes and Their Application to Sickness and Accident Statistics. Almqvist & Wiksells boktryckeri-a.-b., 1964.

[20] Verónica Moreno, Gabriel Pena, and Néstor Rubén Barraza. Procesos de nacimientos no homogéneos y su aplicación a la pandemia del COVID-19. In VIII Congreso de Matemática Aplicada, Computacional e Industrial, pages 711–714, 2021.

[21] Our World In Data. Data on COVID-19 (Coronavirus). [https://github.com/owid/covid-19-data/tree/master/public/data](https://github.com/owid/covid-19-data/tree/master/public/data), 2021. Accessed: 2021-07-29.
[22] G. Pena. Epydemics. https://github.com/GabrielPenaU3F/epydemics/releases/tag/version-1.0, 2021.

[23] G. Pena, V. Moreno, and N. R. Barraza. Stochastic modeling of the mean time between software failures: A review. In Prashant Johri, Adarsh Anand, Juri Vain, Jagvinder Singh, and Mohammad Quasim, editors, System Assurances: Modeling and Management, Emerging Methodologies and Applications in Modelling, Identification and Control, chapter 20. Elsevier Science, 2022.

[24] S Rojas. Comment on “estimation of covid-19 dynamics “on a back-of-envelope”: Does the simplest sir model provide quantitative parameters and predictions?”. Chaos, Solitons & Fractals: X, 5:100047, 2020.

[25] K. Sendova and L. Minkova. Introducing the non-homogeneous compound-birth process. Stochastics, pages 1–19, 09 2019.

[26] Petrónio C.L. Silva, Paulo V.C. Batista, Hélder S. Lima, Marcos A. Alves, Frederico G. Guimarães, and Rodrigo C.P. Silva. Covid-abs: An agent-based model of covid-19 epidemic to simulate health and economic effects of social distancing interventions. Chaos, Solitons & Fractals, 139:110088, 2020.

[27] Matthieu Simon. Sir epidemics with stochastic infectious periods. Stochastic Processes and their Applications, 130(7):4252 – 4274, 2020.

[28] Steven W. Smith. Moving average filters. In Steven W. Smith, editor, Digital Signal Processing, chapter 15, pages 277–284. Newnes, Boston, 2003.

[29] Daniel W. Stroock. An Introduction to Markov Processes, volume 230 of Graduate Texts in Mathematics. Springer-Verlag Berlin Heidelberg, 1st edition, 2005.

[30] S. Triambak, D.P. Mahapatra, N. Mallick, and R. Sahoo. A new logistic growth model applied to covid-19 fatality data. Epidemics, 37:100515, 2021.

[31] Cécile Viboud, Lone Simonsen, and Gerardo Chowell. A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks. Epidemics, 15:27 – 37, 2016.

[32] Di Xiong, Lu Zhang, Gregory L. Watson, Phillip Sundin, Teresa Bufford, Joseph A. Zoller, John Shamshoian, Marc A. Suchard, and Christina M. Ramirez. Pseudo-likelihood based logistic regression for estimating covid-19 infection and case fatality rates by gender, race, and age in california. Epidemics, 33:100418, 2020.