COVID-19 : The «SIR-Gaussian» model

To cite : Christian Manivel* , Thierry Bourbié, Max Roche, Christian Chochillon, Corinne Fernandez, Patrick Lagadec, Patrick Hontebeyrie, Elisabeth Manivel

Christian Manivel* : Main author

Version V1 Ref : CML-SARS-CoV-2-04 French version 1 : Archive ouverte : HAL-Id : 02569180 Current version (en) : HAL-Id

Preprint submitted May 11, 2020 Preprint submitted May, 2020

Abstract

The observation of « daily new deaths in hospital » real data for COVID-19 in France has led us to find that lockdown impacted deaths starting on April 4, 2020 on one hand and led us to make the hypothesis that those data were following a Gaussian function on the other hand, both for « pre-lockdown » data and « post-lockdown » data (which have been influenced by lockdown).

We validated this hypothesis using a gaussian « fitting » for the 45 days « pre-lockdown » period from February 19, 2020 to April 3, 2020 which appeared to be of excellent quality (95 % + adequacy). We obtained an excellent adequacy for a second gaussian fitting for the « post-lockdown » period from April 4, 2020 to May 7, 2020.

We conducted the same gaussian type of fitting on similar data for Spain and Italy and obtained excellent adequacy too.

We concluded that this « gaussian behavior » is typical of the SARS-CoV-2 virus.

On the basis of this model for deaths, we deducted a model for infections (meaning the infections which would have been followed by a future death).

We then found that this gaussian model for deaths is fully compatible with a specific compartmental SIR type model that we called « SIR-Gaussian » (or SIR-G).

This SIR-G model implies that the contagiousness of this epidemic has a specific « behavior » in the way it is linearly and decreasingly dependent on time.

We have then been able to describe a modeled chronology of that epidemic.

Finally we opened some fields for future studies about the reasons why this virus has such a « behavior » and also to find what other kinds of viruses have this same « behavior ».

« *The obsession for figures without paying any attention to relevance is the best indicator of the absence of true mathematical knowledge* » Carl Friedrich Gauss (1777-1855)

COVID-19 : The « SIR-Gaussian » model Christian Manivel <u>christian.manivel.1972@polytechnique.org</u>

Table of contents

1 Real data for deaths in hospital follow a gaussian pattern

1.1 Real data

- 1.1.1 Raw data
- 1.1.2 Data collecting process
- 1.1.3 Lockdown started to impact deaths on April 4, 2020

1.2 « Pre-lockdown » data for deaths per day in hospital in France follow a gaussian pattern

- 1.2.1 General pattern of the model
- 1.2.2 Fitting of data for the period until April 3, 2020
- 1.3 « Post-lockdown » data for deaths per day in hospital also follow a gaussian pattern
- 1.4 Foreign data for deaths per day also follow a gaussian pattern
 - 1.4.1 Italy
 - 1.4.2 Spain
- 1.5 Deaths due to SARS-CoV-2 follow a gaussian model which is typical of this virus
- 1.6 Model parameters for deaths per day in hospital in France
- 2 Therefore infections (deducted from deaths in hospital) follow a gaussian pattern
- 3 This gaussian model is a specific compartmental model that we call « SIR-G » model
 - 3.1 Reminder about compartmental models
 - 3.2 Reminder about Kermack and McKendrick SIR model
 - 3.3 « Free evolution » SIR-G model
 - 3.3.1 Model compartments
 - 3.3.2 Rules
 - 3.3.3 Differential equations
 - 3.4 Contagiousness and reproduction number for SIR-G model
- 4 SIR-G model applied to COVID-19 in France
 - 4.1 « Pre-lockdown » period until March 16, 2020
 - 4.2 « Post-lockdown » period starting March 17, 2020
- 5 Modeled chronology of the epidemic
- 6 Impact of lockdown on the epidemic according to SIR-G model in France
- 7 Fields opened for further studies
 - 7.1 Contagiousness decreases « naturally » linearly over time
 - 7.2 Contagiousness is explicitly dependent on fatality rate and recovery rate
 - 7.3 What viruses have the same « behavior » as SARS-CoV-2 ?

MAIN

1 Real data for deaths in hospital follow a gaussian pattern

1.1 Real data

1.1.1 Raw data

Data used are deaths in hospital (« décès à l'hôpital »), public data from Santé Publique France. <u>https://dashboard.covid19.data.gouv.fr</u>

These deaths in hospital are not the total deaths due to SARS-CoV-2 in France, deaths in institutions for senior people or medical care centers (EPHAD et EMS) and deaths at home are not included.

However, deaths in hospital are an important part of total and therefore are to be considered as representative of the COVID-19 epidemic.

1.1.2 Data collecting process

The observation of raw data lead us to find that this process is not stable. Clearly some data are missing during week-ends and added on the next Monday and possibly Tuesday.

We have thought of correcting this issue, but it would only have increased the relevance of what we say below.

We then decided to keep those raw data as they are, being unchallengeable, as a basis for our study.

1.1.3 Lockdown started to impact deaths on April 4, 2020

A preliminary question was : since when has the lockdown impacted the deaths in hospital ?

The observation of raw data again allow us to say with great certainty that deaths per day in hospital grew continuously till April 3, 2020 but then had a regular decreasing trend. And this fact is not due to any disturbance in the data collecting process.

Therefore we concluded that the first impact on deaths in hospital started on April 4, 2020.

Considering that the first day of lockdown was March 16, and the first real massive impact on the population on March 17, we stated that the time lap between infections and deaths is of 19 days. (this in fact is consistent with what we know of the incubation / infection / death process)

Consequently, we concluded that the raw data for deaths in hospital have not been impacted by lockdown till April 3. We named those data « pre-lockdown » data.

On the contrary, raw data for deaths in hospital after April 3, are fully impacted by the lockdown. We call those data « post-lockdown » data.

NB : Had we corrected the raw data because of the week-end issue of the collecting process, it would not have modified this date of April 3.

Below the real data for deaths in hospital, « pre-lockdown » data and also « post-lockdown » data updated till May 9. (Illustration 1)

FRANCE : Deaths per day in hospital -- "Pre-lockdown" real data and "post-lockdown" data



Illustration 1

1.2 « Pre-lockdown » data for deaths per day in hospital in France follow a gaussian pattern

1.2.1 General pattern of the model

The observation of deaths per day in hospital in France till April 3, has led us to make the hypothesis that the evolution of those deaths followed a gaussian pattern.

Common gaussian function is as below :

 $m(t) = a^* \exp(bt(2r - t))$

m(t) being number of deaths per day in hospital while a, b, r are parameters such as the curve is maximum for t = r.

1.2.2 Fitting of data for the period until April 3, 2020

The hypothesis of such a gaussian type of pattern has therefore been tested on the « pre-lockdown » raw data.

The only constraint that we imposed for fitting was the respect of the April 3 data, which reported a number of deaths per day that day of 588.

This fitting was done on a reasonably long period of 45 days, meaning from February 19 (first death reported) to April 3.

The total square discrepancy between gaussian model and deaths per day real data turned out to be 180, meaning an adequation between model and real of **96,5 %**.

The adequation between model and cumulative deaths real data turned out to be 99,2 %.

Considering those very small discrepancies between model and real, we concluded with good confidence that deaths per day real data in hospital till April 3, follow a gaussian pattern.

1.3 « Post-lockdown » data for deaths per day in hospital also follow a gaussian pattern

1.3.1 General pattern of the model

Obviously b et r parameters for the « post-lockdown » model don't have the same value as those for the « pre-lockdown » pattern, but the pattern is of gaussian type for both.

1.3.2 Fitting of data for the period after April 3, 2020

The hypothesis of such a gaussian type of pattern had therefore been tested on the « pre-lockdown » raw data.

The only constraint that we imposed for fitting was the respect of the April 3 data, which reported a number of deaths per day that day of 588.

This fitting was done on a reasonably long period of 38 days, meaning starting April 3, till May 7.

The total square discrepancy between gaussian model and deaths per day real data turned out to be 496, meaning an adequation between model and real of **95,7 %**.

The adequation between model and cumulative deaths real data turned out to be 99,4 %.

Considering those very small discrepancies between model and real, we concluded with good confidence that deaths per day real data in hospital starting April 3, also follow a gaussian pattern.

Below the illustration of those fittings of « pre-lockdown » data till April 3, and « post-lockdown » data starting April 3. (Illustration 2)



Illustration 2

1.4 Foreign data for deaths per day also follow a gaussian pattern

We did the same type of fitting for « pre-lockdown » deaths data and « post-lockdown » deaths data for Italy and Spain and we confirmed the excellent adequation between a gaussian model and those real data.

1.4.1 Italy

The observation of raw data enabled us to conclude that the first impact on deaths started on March 21, 2020.

The adequation between model and deaths per day real data turned out to be **94,5 %** (on a period of 27 days).

The adequation between model and cumulative deaths real data turned out to be **97,8 %** (on a period of 49 days).

In appendix the real data for deaths in hospital, « pre-lockdown » data and also « post-lockdown » data updated till May 9. (Illustration A1)

1.4.2 Spain

The observation of raw data enabled us to conclude that the first impact on deaths started on March 25, 2020.

The adequation between model and deaths per day real data turned out to be **93,4 %** (on a period of 23 days).

The adequation between model and cumulative deaths real data turned out to be **97,1 %** (on a period of 45 days).

In appendix the real data for deaths in hospital, « pre-lockdown » data and also « post-lockdown » data updated till May 9. (Illustration A2)

1.5 Deaths due to SARS-CoV-2 follow a gaussian model which is typical of this virus

The adequation between model and deaths per day real data in hospital for « pre-lockdown » data in France turned out to be excellent (96% +) on a significant period of 45 days.

The adequation between model and deaths per day real data in hospital for « post-lockdown » data in France turned out to be excellent (96% +) on a significant period of 38 days.

The adequation between model and deaths per day real data for « pre-lockdown » and « post-lockdown » data abroad turned out to be excellent (97% +) on a significant period of 40+ days.

We can state that the adequation between a gaussian model and deaths per day real data are excellent on a significant period of 40+ days both for « pre-lockdown » and « post-lockdown » and also in France as well as abroad.

We then conclude that this excellent adequation of the COVID-19 epidemic with this gaussian model is a characteristic of this SARS-CoV-2 virus.

1.6 Model parameters for deaths per day in hospital in France

The model, as we have seen is built in two parts, first a « pre-lockdown » model till April 3, and a « post-lockdown » model starting April 3.

As a reminder : the model shape is the following : $m(t) = a^*exp(bt(2r - t))$ where m(t) is the number of deaths per day for the day t. Origin is taken on April 3, (t = 0)while a = 588

« Pre-lockdown » model figures are : r = 7 ; b = 0,004638

« Post-lockdown » model figures are : r = -10 ; b = 0,00086735

2 Therefore infections (deducted from deaths in hospital) follow a gaussian pattern

Thus, the number of new infections per day will as a consequence generate, after an average period of time, a proportional number of deaths, according to a constant fatality rate.

We stated an average delay between infections and deaths of 19 days, according to the period of time between the date when the lockdown started and the date when the lockdown impacted deaths. (see paragraph 1.1.3)

This delay is in accordance with the infections The observation period used by Institut Pasteur when they describe the « pre-lockdown » simulation published in the following study HAL Id : 02548181 In fact, they use a 45 days period from February 1, to March 17, which is exactly the 45 days « pre-lockdown » period that we used in our model for deaths, meaning from February 19, to April 3 with a delay between infections and deaths of 19 days.

The fatality rate we used is 0,53 %.

In fact we used the fatality rate calculated by Institut Pasteur HAL Id : 02548181 on the basis of the Diamond princess case taken as a representative sample.

This rate (assumed to be constant) applied to new infections per day allow us to determine the number of deaths per day with a delay corresponding to the delay between infections and consecutive deaths.

Reversely we can obtain a model for new infections per day starting from the deaths per day, applying :

- reverse delay of 19 days and,

- a multiplier of 1 / 0,53 %.

We could say that these infections obtained are the new infections per day that were followed later by a death in hospital. They are not the total number of infections for this epidemic but an important and significant part.

Therefore, real infections could have appeared before February 1, but they have not been followed by a death in hospital.

Considering the importance of that sub-population of infected individuals for this epidemic, we consider that we can adopt for the infections also a gaussian type model.

3 This gaussian model is a specific compartmental model that we call « SIR-G » model

3.1 A reminder about compartmental models

Nowadays, epidemiology is strongly influenced by dynamic compartmental models since Ross studies in 1911 (paludism) then Kermack and McKendrick studies in 1927.

The principle consists of modeling of the behavior of a pathogen (virus, ...) by categorizing subpopulations (compartments) within the global population and defining flows of population between those compartments and the rules associated to them.

Compartmental dynamic models have evolved towards more sophisticated models (multiplicity of compartments, contact matrices, ... etc.) but most of them have kept the Kermack and McKendrick SIR model principles.

3.2 A reminder about Kermack and McKendrick SIR model

There are 3 compartments in this model :

- **S** : Susceptible individuals = Non-infected individuals that could be contaminated
- I : Infected individuals, who can contaminate others
- R : Removed individuals meaning cured and non-contaminating or deceased individuals

The flows between compartments are the following :



Contamination of Susceptibles follow a rate β proportionally to the probability of contact between Susceptibles and infected individuals and to S.I

 β is the infection transmission parameter

The Removed (non contaminating cured individuals or deceased ones) follow a rate γ proportionally to I.

 γ is the removal rate (cured and deceased)

The model could be described as follows:

∂S/∂t = -βSI	(1)
$\partial I/\partial t = \beta SI - \gamma I$	(2)
∂R/∂t = γl	(3)

N = S + I + D + R

with N the total population assumed as constant (no demographic effect), meaning no flow of birth or deaths outside the epidemic.

Struggling against epidemics consists therefore in modifying the behavior of contacts between individuals to decrease virus transmission (by diminishing β SI) and/or improving healing through the health system (by increasing γ)

The underlying hypotheses are the following :

Without any intervention to modify the evolution of the epidemic, β and γ are constants. Even in more sophisticated models, this hypothesis is used for the flows between compartments.

This hypothesis is of importance as we will see thereafter, **B** parameter for the COVID-19 epidemic is not a constant.

3.3 « Free evolution » « SIR-G » model

We call « SIR-G » model a compartmental model where deaths per day follow a gaussian pattern.

We call « free evolution » epidemic (or model) an epidemic or model for which is influenced by no intervention or measure to attenuate the impact of the epidemic (or to modify the pattern of the model) like protective measures or lockdown.

Having said that, the issue now is to determine the characteristics of such a model.

3.3.1 SIR-G model compartments

We keep the principle of a simple SIR model.

- S : Susceptible individuals = Non-infected individuals that could be contaminated
- I: Infected individuals, who can contaminate others

R : Removed individuals split into 2 sub-compartments :

- D: Deceased individuals due to infection
- NC : Cured and non-contaminating individuals

The flows between compartments are the following :



3.3.2 Rules

Infections are caused by contact between individuals infected and non-infected We assume that the population is homogeneous and that the transmission is of mass action law type.

S is the density of non-infected hosts

I is the density of infected hosts

 β is the transmission parameter, i.e. the transmission rate of the illness at each contact or the infectious contact rate.

Its unit is rate / individual / time unit

No assumption is made on β

Infected individuals are according to a γ rate. γ is the healing rate. Its unit is rate / time unit γ is assumed to be constant.

Infected individuals die due to infection according to a fatality rate α Its unit is rate / time unit α is assumed to be constant.

3.3.3 Differential equations

Mathematically speaking, this leads to the following system, describing over time the evolution of the densities of non-infected and infected individuals.

∂S/∂t = -βSI	(1)
∂I/∂t = βSI - αI – γI	(2)
$\partial \mathbf{D}/\partial t = \alpha \mathbf{I}$	(3)
$\partial \mathbf{NC} / \partial \mathbf{t} = \mathbf{\gamma} \mathbf{I}$	(4)

Total population is assumed to be constant (no demographic evolution)

Then we have to add to this system the gaussian model for deaths per day and therefore for new infections per day (see paragraph 3.2) which has this shape :

 $I(t) = \exp(t(2T_m - t) / 2T_s^2)$

Time origin is the date of first infection case : At t = 0 we have $I(0) = I_0 = 1$

maximum of infection is obtained for $t = T_m$ and its value is $I_m = \exp(T_m^2 / 2T_s^2)$ The return to a level of 1 infection per day is obtained when $t = 2T_m$

Therefore

 $\partial \mathbf{I}/\partial t = \mathbf{I} (\mathbf{T}_m - t) / \mathbf{T}_s^2$ (6)

 T_s parameter is homogeneous to time. Inflection points for the curve I(t) are symmetrical to the maximum on either side of the curve T_1 = $T_m - T_s$ T_2 = $T_m + T_s$

3.4 SIR-G model contagiousness and reproduction number

Using (2) et (6) it results in :

$$\beta S = \alpha + \gamma + (T_m - t) / T_{S^2}$$

Infection duration Ti of an individual is $1 / (\alpha + \gamma)$ This parameter depends only on the illness and is a characteristic of this virus.

Finally we obtain

 β S = 1 / T_i + (T_m - t) / T_{s²}

Usually we call reproduction number the ratio $\beta S / (\alpha + \gamma)$, more easily understandable.

Reproduction number is then :

$$R = 1 + T_i(T_m - t) / T_s^2$$

And the **initial value** of this reproduction number (**corresponding to t = 0**), noted R₀, and called basis reproduction number, is then :

 $R_0 = 1 + T_i T_m / T_s^2$

We call « **theoretical end of reproduction** » the moment corresponding to $\beta = 0$. It happens when $t = T_r = T_m + T_s^2 / T_i$

ATTENTION : **T**_r is not the end of the virus's life (the virus is still present in infected individuals at that date)

Consequently:

4 SIR-G model applied to COVID-19 in France

4.1 « Pre-lockdown » period until March 16, 2020

The figures we obtained when we created our model with the « Pre-lockdown » real data in France, are the following :

- $T_i = 1 / (\alpha + \gamma) = 19$
- model time origin (Io = 1) is February 1.
- **T**m = 51 (**I** = **I**m)
- Im = 139 252 and therefore,
- **T**s = 10,48

Important reminder : these data (and particularly I_m value) correspond to infections deducted from deaths per day in hospital, i.e. the daily infections which have been followed by death in hospital. Therefore, real infections could have appeared before February 1, but they have not been followed by a death in hospital.

The chronology of the key events of the epidemic is:

February 1, : time origin $= t = 0 = 10 = 1$		R ₀ = 9,88
March 8, : first inflection of the daily infections curve	=> t = 36	R 1 = 2,81
March 16 : declaration of the beginning of lockdown	=> t = 44	R c = 2,21

4.2 « Post-lockdown » period starting March 17, 2020

A reminder: We call « Post-lockdown » data the data impacted by lockdown

4.2.1 Notation change

The equations have the same shape :

$$\beta S = 1 / T_i + (\theta_m - t) / \theta_{s^2}$$

Infection duration Ti of an individual is unchanged = $1 / (\alpha + \gamma)$ because it is a characteristic of the virus

Reproduction number is :

$$\mathbf{R} = \mathbf{1} + \mathbf{T}_{i}(\mathbf{\theta}_{m} - \mathbf{t}) / \mathbf{\theta}_{s^{2}}$$

We call « **theoretical end of reproduction** » the moment corresponding to $\beta = 0$. It happens when $t = T_r = T_m + T_s^2 / T_i$

 $t = \theta_r = \theta_m + \theta_s^2 / T_i$

ATTENTION : θ_r is not the end of the virus's life (the virus is still present in infected individuals at that date)

Consequently :

 $\mathbf{R}_{1} = \mathbf{R}(\mathbf{\theta}_{m} - \mathbf{\theta}_{s}) = 1 + \mathbf{T}_{i} / \mathbf{\theta}_{s}$ $\mathbf{R}_{m} = \mathbf{R}(\mathbf{\theta}_{m}) = 1$

 $\mathbf{R}_2 = \mathbf{R}(\boldsymbol{\theta}_m + \boldsymbol{\theta}_s) = 1 - \mathbf{T}i / \boldsymbol{\theta}_s \\ \mathbf{R}_r = \mathbf{R}(\boldsymbol{\theta}_r) = 0$

4.2.2 « Post-lockdown » figures

The figures we obtained when we created our model with the « Post-lockdown » real data in France are the following :

- $T_i = 1 / (\alpha + \gamma) = 19$
- model new time origin (Io = 1) is November 7, 2019
- $\Theta_{m} = 119$ (I = Im)
- Im = 120 996 and therefore,
- **θ**s = 24,6

Important reminder : these data (and particularly I_m value) correspond to infections deducted from deaths per day in hospital, i.e. the daily infections which have been followed by death in hospital. Therefore, real infections could have appeared before February 1, but they have not been followed by a death in hospital.

The chronology of the key events of the epidemic is:

March 17, : lockdown starts=> t = 131 $\mathbf{R}_c = 0,62$ March 26, : second inflection point of the daily infections curve=> t = 144 $\mathbf{R}_2 = 0,22$ April 6, : « theoretical end of reproduction »=> t = 151 $\mathbf{R}_r = 0$

We can observe the breakpoint in the growth of the epidemic in the model due to lockdown decision :

March 16, $R_c = 2,21$ March 17, $R_c = 0,62$

5 Modeled chronology of the epidemic

<u>A reminder</u> :

The data used for this modeling are the public deaths in hospital.

These deaths in hospital are not the total deaths due to SARS-CoV-2 in France but are an important part (50%+) of the total and therefore are to be considered as representative of the COVID-19 epidemic.

The adequation between model and deaths per day real data in hospital for « pre-lockdown » data in France turned out to be excellent (daily data 96,5% and cumulated data 99,2%) on a significant period of 45 days.

The adequation between model and deaths per day real data in hospital for « post-lockdown » data in France turned out to be excellent (daily data 95,7% and cumulated data 99,4%) on a significant period of 38 days.

We can state that the adequation between model and reality have been verified on long enough periods of time and that this adequation is excellent.

Important reminder : these data (and particularly I_m value) correspond to infections deducted from deaths per day in hospital, i.e. the daily infections which have been followed by death in hospital. Therefore, real infections could have appeared before February 1, but they have not been followed by a death in hospital.

<u>Recap</u>

February 1, : time origin $\Rightarrow t = 0 \Rightarrow t_0 = 1$		R 0 = 9,88
March 8, : first inflection of the daily infections curve		R 1 = 2,81
March 16 : declaration of the beginning of lockdown		R c = 2,21
March 17, : lockdown starts $=> t = 131$		R c = 0,62
March 26, : second inflection point of the daily infections curve	=> t = 144	R ₂ = 0,22
April 6, : « theoretical end of reproduction » => t = 151		R r = 0
May 11, : partial end of lockdown		
June 1, : Return to a level of 10 deaths per day in hospital		
June 17, : Return to a level of 1 daily death in hospital		

<u>NB :</u>

The modeling we established doesn't prejudge the fact that new clusters could appear in the future if necessary conditions are present, such as:

- presence of an important viral load,

- in a closed space,

- with an important density of non immune individuals,

- and having permanent contacts between them,

- during a significant period of time.

As an illustration, a conference of 2 000 people in a congress type place with lots of interactions between participants during 8 hours and among them a few seriously infected individuals ... or a « pub crawl » during 4 or 5 hours visiting 10 crowded pubs and meeting a few seriously infected individuals.

If a new phase of the epidemic had to start caused by a virus of this kind, the current model could be used to assess the duration, the intensity of that phase and to decide for proportioned measures.

6 Impact of lockdown on the epidemic according to SIR-G model in France

This is presented separately via another publication

7 Opened fields for further studies

The following suggestions for further studies are opened for a « free-evolution » epidemic, outside any measure to lessen its development.

7.1 Contagiousness decreases « naturally » linearly over time

We need to fully understand what this explicit dependency on time means.

We might contemplate the idea that through the process of successive contaminations and replications of the virus, this contagiousness diminishes.

We could think of successive genetic mutations modifying a specific part of its RNA involved in this contagiousness.

We intend to go further into this matter later.

7.2 Contagiousness is explicitly dependent on fatality rate and recovery rate

Here we are probably contemplating the subject of a virus searching for some sort of compromise between death (or cure) and transmission.

We intend to go further into this matter later.

7.3 What viruses have the same « behavior » as SARS-CoV-2 ?

We have begun to look into the process of finding which pathogens fit the same model or « behavior ».

Thank you for your time reading.

ANNEXES



Figure A1



