On evolution model for Covid-19 - infected population. The case of New Zealand

Estate V. Khmaladze [1] **and Giorgi Kvizhinadze** [2]

[1] *Victoria University of Wellington, PO Box 600, Wellington, New Zealand e-mail:* Estate.Khmaladze@vuw.ac.nz

[2] *Capital and Coast District Health Board, Private Bag 7902, Wellington, New Zealand e-mail:* Giorgi.Kvizhinadze@ccdhb.org.nz

Abstract: The work proposes a mathematical model of the process of Covid-19 epidemic as it evolved in New Zealand. The model uses a system of differential equations which emanate form natural assumptions on some probability measure and evolution of this measure on evolving family of simplixes.

The authors did not aim at mathematical complications – the model is simple and easy to follow. The aim rather was practical – to come to justifiable estimations of important parameters like the rate of infection as function of time, thus quantifying effectiveness of the Government measures. Another parameters estimated were the probability distribution of incubation times and recovery times.

The Draft, July 10, 2020

1. General outlook

Corona virus infection arrived in New Zealand at the beginning of January 2020. At that time there was already much discussion and much anxiety around epidemics in China, and also in Italy. The public opinion was prepared for the idea that strong measures will be necessary.

It turned out that the Government understood well what was about to unfold if no measures are taken, and had a plan. According to this plan, there were four stages of social restrictions, from stage 1, least restrictive, to stage 4, the complete lockdown with strict restriction on contacts between people everywhere: in public transport, in shops and supermarkets, between neighbours and within extended families. Older people were promised essential supplies to be delivered to their homes. In principle, police was given permission to enforce the restrictions. This is not a social research, but as citizens of the country, we felt not much interference from the police.

2. Geometric model: simplexes. Differential equations

Consider a population of people, who may be vulnerable to disease, like, in our case, the population of New Zealand. Let N denote the size of this population. Suppose we start observations at some moment we call $t = 0$; say, the day zero is chosen as 1 January 2020. We describe each individual in this population by three moments in time: S is the time until the unaffected individual becomes infected; if the person was never infected, we say the S is infinite; T is the incubation period, the time which elapses from infection until the symptoms become apparent and the person is ill; Z is the time until recovery or, in rare unhappy cases death. In other words, S is calendar time after zero, when the person becomes infected, $S + T$ is the calendar time when the person becomes ill, and $S + T + Z$ is the calendar time of recovery.

These S, T and Z are for us non-negative random variables and below we say what we assume about their distribution(s). Therefore, our model starts with N triplets $(S_i, T_i, Z_i)_{i=1}^N$. Now introduce several point processes, which are of central importance in description of the state and changes of the epidemics. Three of them will be counting processes, of cumulative character, with non-decreasing trajectories, and two of them will be queues, that is differences of counting processes. Namely,

$$\begin{aligned} X_0(t) &= \sum_i \mathbb{I}_{(S_i \leqslant t)}, \\ X_{1a}(t) &= \sum_i \mathbb{I}_{(S_i+T_i \leqslant t)}, \\ X_{2a}(t) &= \sum_i \mathbb{I}_{(S_i+T_i+Z_i \leqslant t)}. \end{aligned}$$

where i labels individuals, or labels triplets in increasing order of S_i-s. Therefore S_1 is the moment when the first infection occurs, S_2 is the second such moment, and so on. Summation runs over an individuals in our initial population P. Here

$$\mathbb{I}_{(S_i \leqslant t)} = \begin{cases} 1, \text{ if } S_i \leqslant t, \\ 0, \text{ if } S_i > t, \end{cases} \qquad \text{and} \quad \mathbb{I}_{(S_i+T_i > t)} = \begin{cases} 1, \text{ if } S_i + T_i > t \\ 0, \text{ if } S_i + T_i \leqslant t \end{cases}$$

are indicator functions of events $S_i \leqslant t$ and $t - S_i \leqslant T_i$.

According to their definitions

$X_0(t)$ is the number of individuals, infected up to time t;
$X_{1a}(t)$ is the number of individuals, which became symptomatic up to time t;
$X_{2a}(t)$ is the number of individuals who were ill and recovered up to time t;

Now introduce two more processes,

$$\begin{aligned} X_1(t) &= \sum_i \mathbb{I}_{(S_i \leqslant t)} \mathbb{I}_{(S_i+T_i > t)} = X_0(t) - X_{1a}(t) \\ X_2(t) &= \sum_i \mathbb{I}_{(S_i+T_i \leqslant t)} \mathbb{I}_{(S_i+T_i+Z_i > t)} = X_{1a} - X_{2a}(t), \end{aligned}$$

and these are derivable from the previous three and form queues. Their heuristic meaning is also clear:

$X_1(t)$ is the number of infected but still asymptomatic individuals at time t;
$X_2(t)$ is the number of individuals who are ill (symptomatic) at the time t;

and finally

$X_3 = X_0 - X_{2a}$ is the number of individuals, recovered up to time t,

presumably with some immunity against the disease.

Let us think about t as time measured in days; intensities (or rates) below will, therefore, be daily intensities or rates. We also will use the same notations for populations themselves, not just size of this population; we may say, for example, that individual belongs to population $X_1(t)$ meaning that at time t the individual was infected but not symptomatic.

Assume now that within each triplet, random variables T_i and Z_i are independent from S_i and independent from each other. Also assume that distribution of all $T_i, i = 1, 2, \dots$, is the same and G is the distribution function of each of them. Similarly, assume that $Z_i, i = 1, 2, \dots$, are identically distributed with distribution function F. The last assumptions are some simplifications, as distributions G and F may depend on the age of the individual. However, for adequate description of the sub-populations this may not be very important, as well as particular choice of these distribution – see below on this choice.

Evolution of $B(t)$. Denote $B(t) = EX_0(t)$ the expected vale, or expected size, of $X_0(t)$. The process $X_0(t)$ is the main process, basically driving evolution of the other processes, and therefore $B(t)$ is useful to understand. We consider the following differential equation for it:

$$dB(t) = m(t)dt + r(t)[B(t) - B_{1a}(t)]dt. \tag{1}$$

where B_{1a} is the expected value of X_{1a}, see the next sub-section. The equation (1) should be considered as a textbook model; among large number of sources, let us also refer to Khmaladze [2013], Lecture 14. Note only that increment of $B(t)$, with rate $r(t)$, can be proportional only to the difference $B(t) - B_{1a}(t)$, because those who become ill were immediately isolated from the public.

In the first term, $m(t)$ is the (daily) immigration rate of infected people. Ministry of Health knows, and has published, the dates of arrival of such people. However, even in absence of this data one can model $m(t)$ as a given fraction of the total arrival to New Zealand. We consider the case of the time varying $m(t)$, because due to the Government measures, it indeed changed in time down to zero; it may change again and gradually increase to normal levels. In the second term, $B(t)$ is, as we said, the expected number of infected people thus capable for spreading the infection. The intensity of this, per infected person, is described

by $r(t)$. In fixed social conditions and habits of life this intensity depends on how aggressive the virus is and can be assumed constant. However, social conditions in New Zealand did not stay the same - we moved through levels. Therefore it also should be considered variable in time. The value of $r(t)$ can rise again from the current low level when the movements in population will become free again. But at that time $B(t)$ will be small. Solution of equation (1) is

$$B(t) = \exp\left(\int_0^t r(s)ds\right)\int_0^t \exp\left(-\int_0^y r(s)ds\right)\, m(y)dy.$$

Speaking heuristically, $B(t)$ increases for two reasons: because $\int_0^t r(s)ds$ increases for non-zero intensity of infections, and because $\int_0^t \exp(-\int_0^y r(s)ds)m(y)dy$ increases in the presence of positive inflow $m(t)$. If $m(t) = 0$ for $t > t_0$ this integral becomes constant in $t > t_0$, and $B(t)$ increases less quickly. This agrees with heuristic intuition.

In three-dimensional Euclidean space $\mathbb{R}^3$ introduce the measure ν given as direct product

$$\nu((0,t] \times (0,x] \times (0,z]) = B(t)G(x)F(z)$$

where denotes . Then introduce the following sets, which depend on t:

$$A_{1a}(t) = \{(s,x,z) : s + x \leqslant t\},\ A_{1b}(t) = \{(s,x,z) : s \leqslant t, s + x > t\},$$

$$A_{2a}(t) = \{(s,x,z) : s + x + z \leqslant t\},\ A_{2b}(t) = \{(s,x,z) : s + x < t, s + x + z > t\}.$$

and consider evolution of measure of these sets in time.

First note that $\nu(A_{1b}(t))$ is nothing else but expected value of the process $X_1(t)$,

$$EX_1(t) = \nu(A_{1b}(t)),$$

while $\nu(A_{2b}(t))$ is expected value of the process $X_2(t)$,

$$EX_2(t) = \nu(A_{2b}(t)).$$

Indeed, if the triplet $(S_i, T_i, Z_i) \in A_{1b}(t)$ then the corresponding person belongs to the sub-population $X_1(t)$ and if $(S_i, T_i, Z_i) \in A_{1b}(t)$ then the corresponding person belongs to the sub-population $X_2(t)$. Values of $\nu(A_{1a}(t))$ and $\nu(A_{2a}(t))$ are expected values of the processes $X_{1a})(t)$ and $X_{2a}(t)$. This provides us with sufficient simplicity in description of the evolution "in expected values".

Evolution of $B_1(t)$. By definition of the measure ν, we have

$$\nu(A_{1b}(t)) = \int_0^t (1 - G(t-s))dB(s)$$

and thus for $B_1(t) = EX_1(t)$ we have

$$dB_1(t) = d\nu(A_{1b}(t)) = dB(t) - \int_0^t g(t-s)dB(s)dt. \tag{2}$$

Similarly, for $B_{1a}(t) = EX_{1a}(t)$ we fist see that

$$\nu(A_{1a}(t)) = \int_0^t G(t-s)dB(s)$$

is a convolution of G and B. It then follows that

$$dB_{1a}(t) = \int_0^t g(t-s)dB(s)dt$$

Evolution of $B_2(t)$. Again, from the definition of ν, it follows that

$$\nu(A_{2b}(t)) = \int_0^t (1-F(t-s))dB_{1a}(s)$$

and therefore, for $B_2(t) = EX_{2b}(t)$ we obtain

$$dB_2(t) = d\nu(A_{2b}(t)) = dB_{1a}(t) - \int_0^t f(t-s)dB_{1a}(s)dt. \tag{3}$$

We can represent the right hand side through B, which will reveal useful structure of the integral:

$$dB_2(t) = \int_0^t g(t-s)dB(s)dt - \int_0^t f(t-s)\int_0^s g(s-y)dB(y)ds\,dt.$$

Now introduce convolution of the densities f and g,

$$f \star g(t) = \int_0^t f(t-s)g(s)ds.$$

This is the density of the random variable which is the sum $T+Z$. Now in the second integral let us change the order of integration and integrate with respect to s first. This will lead to

$$\int_0^t f(t-s)g(s-y)ds = \int_0^{t-y} f(t-y-u)g(u)du = f \star g(t-y),$$

and altogether

$$dB_2(t) = \int_0^t g(t-s)dB(s)dt - \int_0^t f \star g(t-y)dB(y)\,dt. \tag{4}$$

Finally, for the expected value of recovered people we have

$$dB_{2a}(t) = \int_0^t f \star g(t-y)dB(y)\,dt.$$

The choice of distributions G and F. The presence of convolution here has some bearings on the practical question of how to choose densities f and g

in our model. The data on duration of incubation period, at least at present, is scant. It would be highly speculative to try and estimate g from this data. The situation with data on duration of illness may be better, but reliable estimation of density requires many data. We can, however, significantly reduce the difficulty by choosing both g and f from a family of distributions, which is closed under convolution. These are families of distributions $\{H_\alpha, \alpha \geqslant 0\}$, indexed by a scalar parameters α, and such that convolution of H_α and H_β is the distribution $H_{\alpha+\beta}$.

Now imagine that we choose as distribution of incubation periods, G, some H_α, and choose as distribution of the period of illness, F, some H_β. If the family $\{H_\alpha, \alpha \geqslant 0\}$, is sufficiently rich family, then to make such choice will be possible and will require only estimation two numbers, α and β from the data, which statistically is much more simple and stable process. And at the same time, will know the distribution $H_{\alpha+\beta}$ straightaway.

One such family, we believe, is available for us. It is the family of Gamma distributions with the same scale parameter λ:

$$h_\alpha(x) = \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x}$$

3. What was found. Statistical discussions

4. Results as shown through graphs

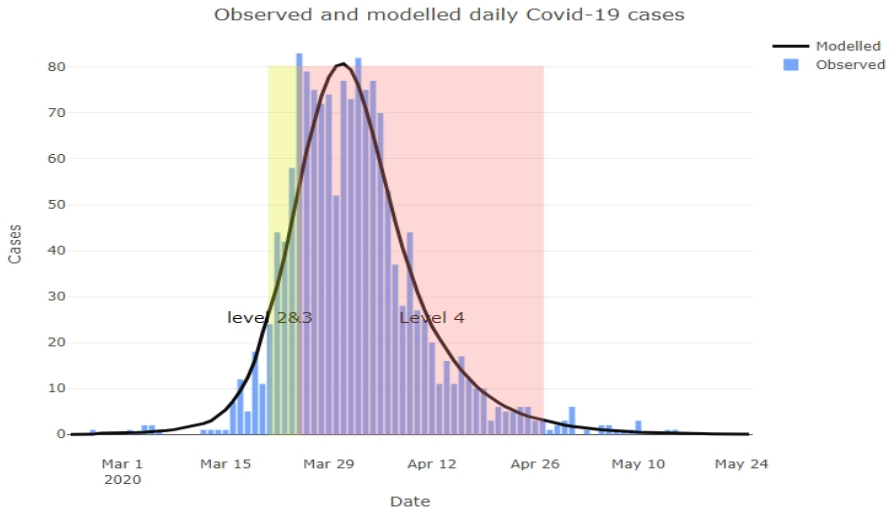

FIG 1. *The graph shows daily reported cases and its expected value, from the model. In our notations these are daily increments of X_{1a} and B_{1a}, respectively. Overlaid on the graph are periods of New Zealand levels 2 and 3, and then the strongest level 4 (the "lockdown") of social restrictions.*

References

Ottar N. Bjornstad, *Epidemics: Models and data using R*, Springer, 2018

Pierre Brémaud, *Point processes and queues. Martingale Dynamics*, Springer-Verlag, 1981

Estate Khmaladze, *Statistical Methods with Applications to Demography and Life Insurance*, CRC Press, Chapman and Hall Book, 2013

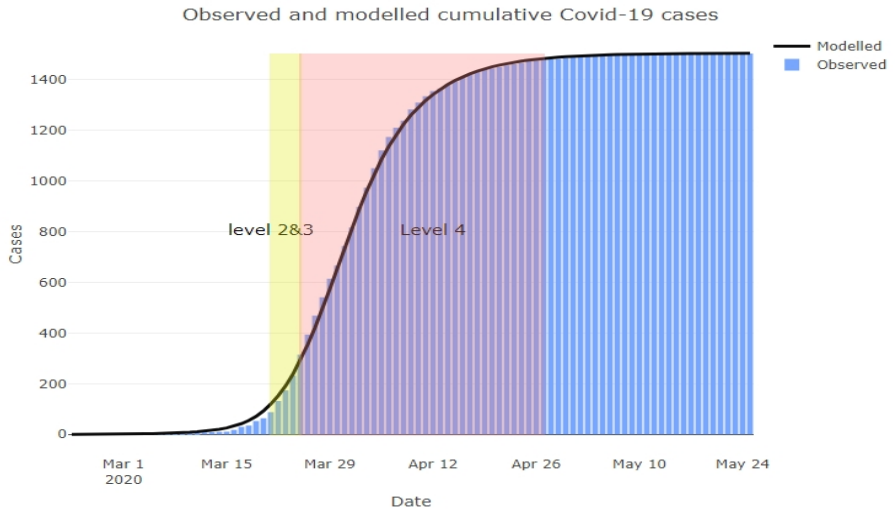

FIG 2. *The graph shows cumulative number of reported and its expected value, from the model. In our notations these are X_{1a} and B_{1a}, respectively. Again, overlaid on the graph we show periods of New Zealand levels 2 and 3, and then the strongest level 4 of social restrictions.*

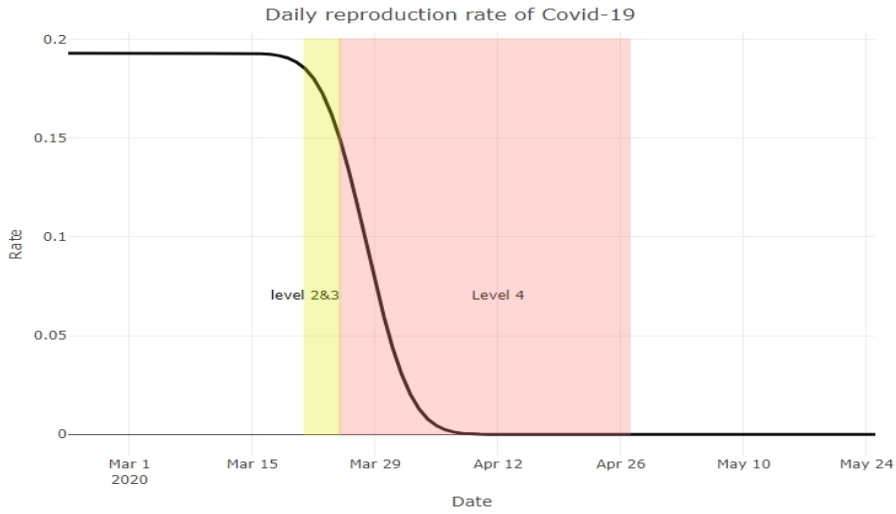

FIG 3. *The graph shows the infection rate according to the model. In our notations it is $r(t)$. Heuristically speaking, it is the expected number of people an infected person will infect in a day.*

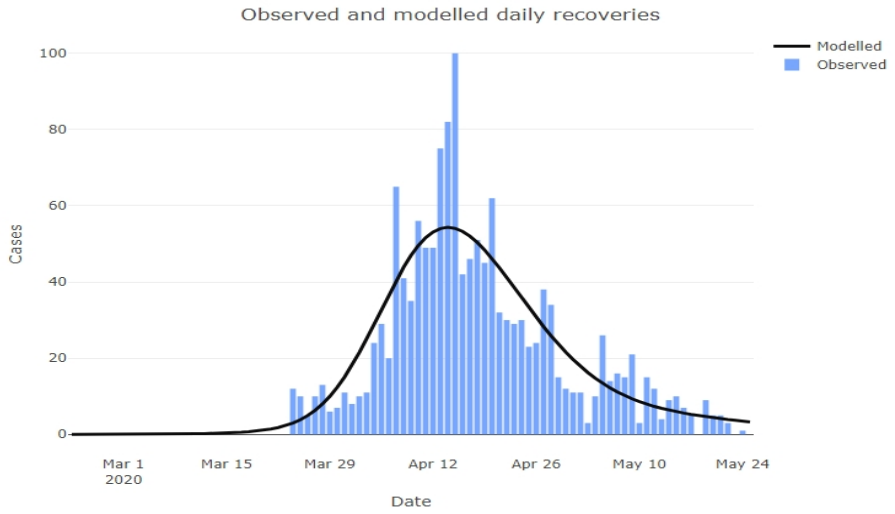

FIG 4. *The graph shows the daily recovery rate and its model. In our notations it is* $r(t)$. *?? Heuristically speaking, it is the expected number of people an infected person will infect in a day.*

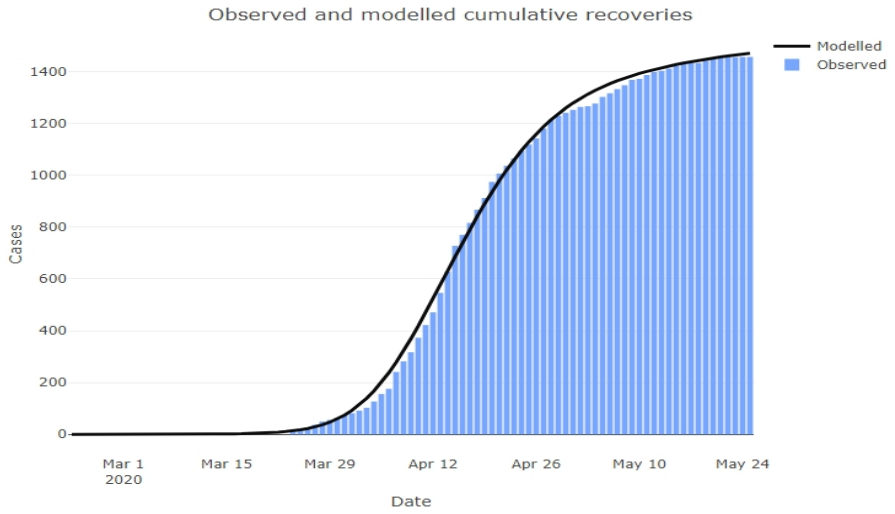

FIG 5. *The graph shows cumulative number of recovered people and its model. In our notations these are* X_{2a} *and* B_{2a}, *respectively ??* .