Universidad del País Vasco
Euskal Herriko Unibertsitatea

ZIENTZIA ETA TEKNOLOGIA FAKULTATEA
FACULTAD DE CIENCIA Y TECNOLOGÍA

Gradu Amaierako Lana / Trabajo Fin de Grado
Fisikako Gradua / Grado en Física

Analysis and Parameters Estimation of a Covid-19 Epidemic Model

Egilea/Autora:
Carmen Legarreta Gallo
Zuzendariak/Directores:
Santiago Alonso Quesada
Manuel de la Sen Parte

Leioa, 2021ko ekainaren 16a /Leioa, 16 de junio de 2021

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

During the COVID-19 pandemic quarantine and testing policies have been of vital importance, since the causative agent has been a novel virus and no vaccine was developed by the time. In this work an epidemiological model is constructed which includes quarantine periods of people with symptoms that has tested positive, and it will trigger a trace of their close contacts, who will be also tested and put in quarantine if the result is positive. Moreover, how the model's parameters affect its stability is analysed with the basic reproduction number $\mathcal{R}_0$. Since the COVID-19 outbreak in Spain (approximately the 13/03/2020) until now (25/04/2021) different restrictions have been applied, so data from the Autonomous Community of Cantabria has been gathered and used to estimate the parameters and see how the restrictions have affected their values. In the parameter estimation process it has been assumed that the constructed model follows the structure of an ARX model. Finally, considering that gathered data hold errors, the model has been validated.

Keywords: Epidemiological model, COVID-19, Quarantine, Tracing, ARX parameter estimation.

---

1 INTRODUCTION

In late December of 2019 China reported to the Word Health Organization (WHO) several cases of pneumonia apparently linked to the Huanan Wholesale Seafood Market in Wuhan City, and they point out that this disease might be caused by a new virus. Chinese scientists analysed the virus from a hospitalized person and the 11th of January China confirmed that the abnormal arose of pneumonia cases were due to a novel coronavirus, similar to SARS and MERS, so they named it SARS-CoV-2 (afterwards it will be renamed to COVID-19). It was also found out that fever and sore cough were the common symptoms, and that pneumonia was a less common consequence of the virus [1].

By late January of 2020 some cities of China lock down to slow down the increase of hospitalized people, and some airlines decided to suspend flights to (or from) China. The 30th of January the WHO reported that there were 7.818 confirmed cases all over the world, from which 7.736 cases were in China, 1.370 were severe and 170 deaths [26]. Moreover, the Centers for Disease Control and Prevention (CDC) wrote a press release in which was outlined the first case of person-to-person spread in U.S, and consequently world countries geared up [13]. All the applied prevention measures were not enough to avoid the virus spreading and on March 11 WHO determined that COVID-19 was a global pandemic [24]: there were 132.758 confirmed cases in the world and 4.955 deaths [23]. On March 17 Europe and Russia closed its borders and by the time many countries announced lock downs (e.g Spain, France,...) [1].

When Spain imposed a quarantine period the 13th of March, its epidemiological situation was critical: there were 4.209 confirmed cases from which 575 were new ones, and the 42% of the total cases were hospitalized [10]. Even if the rules were very restrictive (only on necessary situations it was allowed going outside and wearing the face mask was mandatory) the cases continue raising until the 25/03/2020 with a pick value of 720 new cases and 825 deaths [11], and it was not until the 25/05/2020, when the number of new cases was approximately 2.000 and 68 the number of deaths [11], that more permission was given to the inhabitants (e.g it was possible working out outside at certain time ranges). The 21th of June the Spanish government announced the "new normality", the mask face and social distancing (1.5m) were still mandatory though (at this point each autonomous community could impose more constrains but their capability was limited) [12]. During summer vacation the situation maintained stable, however when people started working (approximately after the 09/09/2020) the epidemiological situation worsen. To tackle this problem each Autonomous Community (AC) combined and applied differently restrictions (e.g curfew and bars close down), and considering the later results inhabitants cast doubts on whether the imposed rules were worth or not. Therefore, a mathematical model which could represent the COVID-19 behaviour over certain conditions could be very useful.

The aim of epidemiological models is to understand and predict the dynamics of a disease, and it must exhibit equilibrium between three elements [16]: accuracy, transparency and flexibility. Accuracy is the capacity to reproduce the real life behaviour and a necessary feature to establish good control actions, perhaps model's complexity is proportional to the accuracy. On the other hand, transparency considers how the model components interact in the dynamics, and it is a characteristic that is opposed to complexity and consequently to accuracy. Finally, flexibility is the ability to adapt the model to other circumstances.

If the model fits well with the real diseases dynamics it can be used to infer optimal control actions, so it makes possible implementing more efficient measures and optimizing the use of resources. In 2001 the epidemic foot-and-mouth outbreak in UK, and the result of three different models applied to this case predicted a large scale spreading and a better control of the epidemic if the infected and exposed animals were culled [16]. In many researches it has been considered a spreading of a very high death rate virus such as smallpox [16] to verify whether a massive vaccination, which could generate health problems, is advantageous or not. However, the model's results differ from each other since there are many epidemiological uncertainties.

As far as COVID-19 is a pandemic, there are many data sources that provide information about the epidemiological situation, so it is feasible to create a model parametrizable from real data and reduce epidemiological uncertainties. Thus, clear control strategies could be built up and implemented.

1.1 Objectives

In this work a deterministic mathematical model is built up, which spans the well known SEIR epidemic model by adding people in quarantine and asymptomatics. To reduce the model's dimension infected people with symptoms and asymptomatics will be gathered together, so a SEIQR epidemic model is obtained - (S) susceptible, (E) exposed, (I) infected, (Q) quarantined and (R) recovered-. In this model, it is assumed that only exposed and infected people will cause new infections since people in Q, which represent the individuals that stay in hospitals and at home during some period due to the disease, are concerned about their illness and they try not being in contact with others. On the other hand, it will be considered that some percentage of the individuals in I, more precisely the symptomatics ones, will be reported and carried to Q, and it will trigger a tracing method as it is shown in [28, 8].

After constructing the SEIQR model schematic, defining how individuals interact and defining the necessary parameters, the respective nonlinear system of differential equations will be obtained and analysed. The theoretical expression for the basic reproduction number , which is a variable dependent on the system's parameters and a threshold that determines whether the disease is out of control or not, will be obtained through the next generation matrix [29]. For stability analysis purpose the equilibrium points, *free disease equilibrium* and *endemic equilibrium*, which determine the states in which the disease is extinguished from the society and when the disease prevail (e.g current flu), respectively, are calculated. The system is linearized around the equilibrium points and the eigenvalues problem solved to evaluate when their real part become positive/negative (condition for system's local unstability/stability).

Some of the model parameters (e.g virus incubation period, probability to get infected from symptomatic/asymptomatic people) are a disease characteristic, so their values have been collected via other studies [8, 17, 18, 5, 30]. However, parameters that depend on the society behaviour (e.g average number of contacts per day) must be determined with the respective real data. Since the autonomous community of Cantabria gathers wide data about the epidemiological situation, it has been used to estimate the unknown parameters, such as the average contact rate and the tracing effectiveness. It has been assumed that the discrete differential expressions of variables such as Q and I follow an ARX model structure. So, the ARX model's parameters have been estimated with the least square method and their values have been used to infer the unknown epidemiological parameter's values. Finally, with the obtained parameters the differential equations have been solved and the results have been compared with real data.

2 SYSTEM DESCRIPTION

The system can be defined as deterministic or stochastic depending on how the population of one block transforms into the others, that is; deterministic models consider that the rates (e.g chance to be infected) maintain constant with respect to time, and therefore there is a unique solution for each initial condition. In contrast, stochastic models add random or probabilistic variables into the rates, so the model will give a set of probable solutions. Even if human behaviour and infections have stochastic components, when the real system (stochastic) is composed by a large group of people, deterministic models could be used to represent the system [16]. Then, the ratios are constant and could be determined by statistical results.

The graphical representation of the system is made by a group of blocks (or compartments), each one represents certain part of the population, and the arrows make the connection between two blocks (e.g $A \rightarrow B$) and define the way in which part of the population from A transforms into the inhabitants of the block B. To build up the model flow chart it is important to know the features of the tackling disease; which are the stages of the disease, incubation period, immunity period,...

Regarding CoVid-19, the Word Health Organization reported that there are three main ways in which the disease is transmitted; pre-symptomatic, symptomatic and asymptomatic transmissions [25]. Pre-symptomatic transmission, possible during the incubation period (time between the exposure and the symptoms onset) which is 5-6 days [25, 22], is more likely to happen 1-3 days before symptoms appear [31]. The preliminary way of transmission is via symptomatic cases, but even if the percentage of asymptomatic cases is 16%-17% [18, 5], it have been seen that they show similar viral loads [36, 18, 15]. Some researches have spotlight the importance of pre-symptomatic and asymptomatic cases since contribute to the virus high spread [35], therefore many countries were forced to include quarantine periods to reduce the spreading. The flow chart depicted in Figure 1 shows a deterministic model of CoVid-19 which includes the characteristics mentioned before.

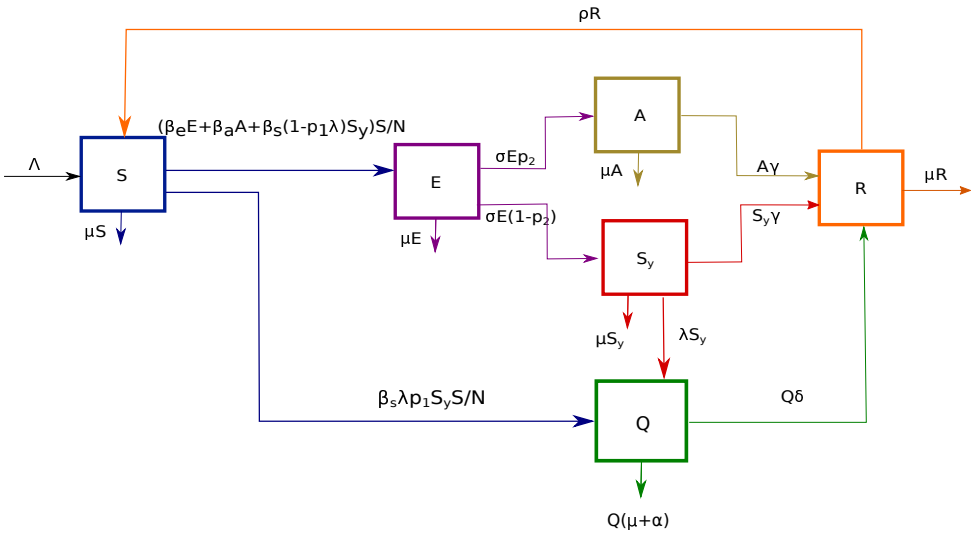


Figure 1: System flow chart.

Figure 1 shows 5 different compartments,

- **Susceptible** (S): this class is constructed by the group of individuals that can get the disease and are not yet infected.
- **Exposed** (E): those individuals of the group S that have been in contact with infected people and therefore are infected too, but they do not present symptoms because they are still incubating the virus. They move freely, thus they will maintain an average number of close contacts per day c_{free}.
- **Symptomatic** (S_y): this group consists on people that got the disease and present symptoms. Soon after the incubation time an individual of S_y will probably move freely (the average number

of close contacts per day will be equal to c_{free}) since the host manifests pre-symptoms, which commonly are confused with tiredness or symptoms of a common cold/flu. After, the list of symptoms will become larger and their intensity will increase too, therefore he/she will go to the doctor and put in quarantine. It is assumed that the average number of close contacts of people from S_y is lower than c_{free}.

- **Quarantined** (Q): this group is made by groups coming from S and S_y. People from S_y that presented clear symptoms will have tested positive and put in quarantine. Their positive results will trigger a tracing through their close contacts, which is assumed that are individuals from S. Traced contacts will be tested, and those with positive result will be put in quarantine. It is assumed that the average number of contacts per day of people in Q is nearly zero, because they are aware of their disease and they will interact less with others. Note that Q covers people that have tested positive and are staying at home or hospitals, therefore only people from Q might die due to the disease.

- **Asymptomatic** (A): the group of infected people that doesn't present symptoms. They move freely with an average number of close contacts equal to c_{free}.

- **Recovered** (R): when infected people recover from the diseases and have some immunity against it during a period of time.

The system's population size (N) is assumed to be large and constant; the natural deaths and births have the same rate. The newborns are introduced in the class S with a rate Λ, and from all the classes it is removed a portion with a rate μ (daily natural death rate).

It is considered that the population is homogeneously mixed, so it is possible to determine the rates β_e, β_a and β_s, which indicate the probability to infect susceptible people when they are in contact with exposed, asymptomatic and symptomatic people, respectively. The mathematical expressions are,

$$\beta_e = c_e p_e, \quad \beta_a = c_a p_a, \quad \beta_s = c_s p_s$$

where c_i is the average number of close contacts per day of a member from i, and p_i the probability to infect others when there is a close contact of a member from i, where $i \in \{E, S_y, A\}$. As it was mentioned before, it will be assumed that $c_e = c_a = c_{free}$, $c_e > c_s > 0$, and $p_a \approx p_s > p_e$. These are necessary parameters to define how susceptible people get infected; one person from S has a probability E/N, S_y/N and A/N of contacting a person from E, I and A, respectively, so the chance to get infected is $(\beta_e E + \beta_s S_y + \beta_a A)/N$. Therefore, all susceptible people that will get disease per day is,

$$(\beta_e E + \beta_s S_y + \beta_a A)\frac{S}{N}. \tag{1}$$

Exposed people after the incubation time (σ^{-1} days) will transform into asymptomatic people with a probability p_2. The regarding ones will transform into symptomatic, which at the early beginning they may not consider that they are sick. However, after λ^{-1} days (λ indicates the rate at which people from S_y go to the doctor, and it is supposed that its value will be larger as the tests stock, sanitary resources and the disease impact on society are increased) their symptoms will increase and they will go to the doctor, test positive and put in quarantine. This will start a tracing, which effectiveness is determined by p_1, of the positive tested symptomatic people close contacts ($c_s \lambda p_1 S_y S/N$), and those who test positive ($\beta_s \lambda p_1 S_y S/N$) will go to Q. Therefore, the number of new exposed per day will be equal to the susceptible people that get the disease per day, see expression (1), minus traced people that have tested positive:

$$(\beta_e E + \beta_s (1 - p_1 \lambda) S_y + \beta_a A)\,\frac{S}{N}.$$

People in Q include all people that have been tested positive, which means hospitalized ones, people in ICU, and those that are in their homes. Some people from Q will remain there during the quarantine period (δ^{-1} days), and others will die from COVID-19 with a rate α. After γ^{-1} days both asymptomatic and symptomatic individuals will recover from the disease, and they will be part of the recovered population R during a period ρ^{-1}.

6 blocks are shown in the scheme from Figure 1, so 6 differential equations are needed to represent the system. Manage with 6 equations might be difficult, therefore a new compartment I, involving A and S_y, is defined to reduce the number of equations from 6 to 5, that is;

$$I = A + S_y, \quad A = p_2 I, \quad \text{and} \quad S_y = (1 - p_2)I \tag{2}$$

and in Figure 2 the block I (infectious) substitutes the blocks A and S_y.

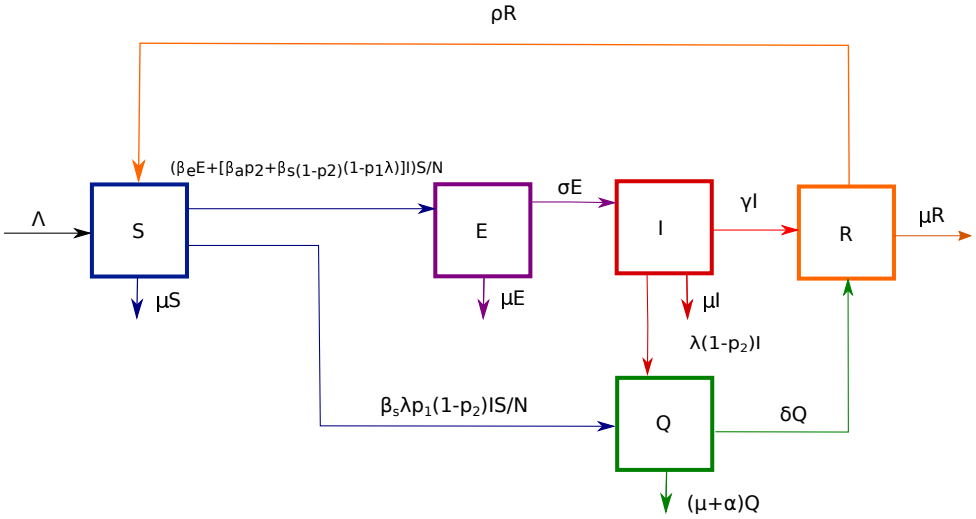


Figure 2: Reduced flow chart of an epidemiological model.

The system representing the model is given by the following set of differential equations,

$$\begin{aligned}
\dot{S} &= \Lambda + \rho R - [\beta_e E + (p_2\beta_a + (1-p_2)\beta_s)I]\frac{S}{N} - \mu S \\
\dot{E} &= [\beta_e E + (p_2\beta_a + (1-p_2)(1-p_1\lambda)\beta_s)I]\frac{S}{N} - E(\sigma + \mu) \\
\dot{Q} &= (1-p_2)\beta_s p_1 \lambda I \frac{S}{N} + \lambda(1-p_2)I - (\mu + \alpha + \delta)Q \\
\dot{I} &= \sigma E - (\lambda(1-p_2) + \mu + \gamma)I \\
\dot{R} &= \gamma I + \delta Q - (\rho + \mu)R
\end{aligned} \tag{3}$$

and regarding the total population,

$$\begin{aligned}
N &= S + Q + E + I + R \\
\dot{N} &= \Lambda - \mu N - \alpha Q
\end{aligned} \tag{4}$$

and the parameters are conditioned by,

$$\begin{gathered}
0 \leqslant \mu, \alpha, \beta_s, \beta_e, \beta_a, \sigma, \gamma, \delta, \rho, \quad \text{and} \\
0 \leqslant p_1, p_2 \leqslant 1.
\end{gathered} \tag{5}$$

To demonstrate the consistency of the equations, the sizes of the subpopulations S, Q, E, I, R and N must not be negative, it can be shown that all the solutions will maintain positive for any initial conditions $(S_0, Q_0, E_0, I_0, R_0) \in \mathbb{R}^+_5$ if the differential equations are semi-positive when $(S, Q, E, I, R) \in \mathbb{R}^+_5$ [28],

$$\begin{aligned}
\dot{S}\big|_{S=0} &= \Lambda + \rho R \geqslant 0 \\
\dot{E}\big|_{E=0} &= (p_2\beta_a + (1-p_2)(1-\lambda p_1)\beta_s)I\frac{S}{N} \geqslant 0 \\
\dot{Q}\big|_{Q=0} &= (1 + \beta_s p_1 \frac{S}{N})(1-p_2)\lambda I \geqslant 0 \\
\dot{I}\big|_{I=0} &= \sigma E \geqslant 0 \\
\dot{R}\big|_{R=0} &= I\gamma + \delta Q \geqslant 0.
\end{aligned}$$

Moreover, it is possible to prove that the entire population will be bounded: considering that nobody is dying because of the disease ($\alpha = 0$), the solution of the differential equation (4) is given by,

$$N = N_0 e^{-\mu t} + \frac{\Lambda}{\mu}$$

and the limit,

$$\lim_{t\to\infty}\left[N_0 e^{-\mu t}+\frac{\Lambda}{\mu}\right]=\frac{\Lambda}{\mu}$$

So, all the solutions of the system (3) will be bounded and the region given by:

$$\mathbf{D}=\left\{(S,E,Q,I,R)\in\mathbb{R}^5_+ : N\leqslant\frac{\Lambda}{\mu}\right\}$$

will hold all the possible solutions.

3 SYSTEM STABILITY ANALYSIS

Linear systems accomplish two properties: superposition and homogeneity. The superposition property implies that the system's output to a sum of inputs must be equal to the sum of the outputs when the respective inputs are applied individually. So, considering a system $\dot{x} = f(\mathbf{x}, \mathbf{u})$, where $\mathbf{x}$ and $\mathbf{u}$ are the state vector and the input vector respectively, if the outputs for the inputs $\mathbf{u_1}$ and $\mathbf{u_2}$ are $\mathbf{f_1}$ and $\mathbf{f_2}$ respectively, the system will fulfill the superposition property if its output is $\mathbf{f_1} + \mathbf{f_2}$ when $\mathbf{u_1} + \mathbf{u_2}$ is applied. Regarding the homogeneity property, it is satisfied if for some given inputs $\mathbf{u_1}$ and $A\mathbf{u_1}$, where A is a matrix with constant values, their outputs are $\mathbf{f_1}$ and $A\mathbf{f_1}$, respectively.

When systems are nonlinear, it is complicated to analyse their response, design profitable control inputs, etc. To deal with such a problem, it is possible to linearize the system, assuming that the signals are small, around its equilibrium states. Once the linear approximation of the system is done, the stability of the equilibrium states can be studied (e.g. with Lyapunov's theorem).

The system given by the set of equations (3) is nonlinear. With the aim of analysing its stability, first the equilibrium states have been calculated, and two have been distinguished: the *disease-free equilibrium* and the *endemic equilibrium*. The *disease-free equilibrium* considers that no infected people will remain when the disease is removed, which is the desirable situation, and simplifies calculus of the conditions that force the system go to the *disease-free equilibrium*. The *endemic equilibrium* holds all the equilibrium states that still have infected cases. Then, the system has been linearized about the equilibrium states and the Lyapunov's stability theorem applied to infer the conditions for their stability. The *basic reproduction number* $\mathcal{R}_0$ gives the condition for which the *disease-free equilibrium* is asymptotically stable (if $\mathcal{R}_0 < 1$ then it is asymptotically stable), so the methodology to calculate $\mathcal{R}_0$ has been followed and the result compared with the ones obtained via Lyapunov's theorem.

The programming language *Matlab* has been used to check whether the results match with the numerical ones. Moreover, the system has been solved with a *Matlab* function called *ode45()*, which uses the Runge-Kutta algorithm, for the cases $\mathcal{R}_0 < 1$ and $\mathcal{R}_0 > 1$. It has been seen that when $\mathcal{R}_0 < 1$ ($\mathcal{R}_0 > 1$) the disease tends to disappear (persist), as it was expected from the theoretical results.

3.1 Equilibrium States

To know how the system (3) will converge, its solution $(S^*, E^*, Q^*, I^*, R^*)$ must be obtained when it is at equilibrium ($\dot{S} = \dot{E} = \dot{Q} = \dot{I} = \dot{R} = 0$). Depending if the disease persists or not two equilibrium cases can be distinguished; *disease-free equilibrium* and *endemic equilibrium*.

Disease-free equilibrium occurs when the disease is vanished ($E^*_{dfe} = Q^*_{dfe} = I^*_{dfe} = 0$). If there is no infectious individuals R will decay to zero, and therefore S will hold all the population ($S = N = \Lambda/\mu$). This can be easily seen by substituting the mentioned conditions into the system (3),

$$\begin{aligned} 0 &= \Lambda + \rho R^*_{dfe} - \mu S^*_{dfe} \\ 0 &= -\rho R^*_{dfe} - \mu R^*_{dfe} \end{aligned} \tag{6}$$

which solution is straightforward. So, the obtained *disease-free equilibrium* point is the following one:

$$P_{dfe} = (S^*_{dfe}, E^*_{dfe}, Q^*_{dfe}, I^*_{dfe}, R^*_{dfe}) = (\Lambda/\mu, 0, 0, 0, 0).$$

Regarding *endemic equilibrium* point

$$P_{ee} = (S^*_{ee}, E^*_{ee}, Q^*_{ee} I^*_{ee}, R^*_{ee}),$$

it occurs when the disease continues affecting the population ($E^*_{ee} \neq 0$, $I^*_{ee} \neq 0$). The system to be solved is the following one,

$$\begin{aligned} 0 &= \Lambda + \rho R^*_{ee} - [\beta_e E^*_{ee} + (p_2\beta_a + (1-p_2)\beta_s) I^*_{ee}] \tfrac{S^*_{ee}}{N^*_{ee}} - \mu S^*_{ee} \\ 0 &= (\beta_e E^*_{ee} + (p_2\beta_a + (1-p_2)(1-p_1\lambda)\beta_s) I^*_{ee}) \tfrac{S^*_{ee}}{N^*_{ee}} - (\sigma + \mu) E^*_{ee} \\ 0 &= \left(1 + \beta_s p_1 \tfrac{S^*_{ee}}{N^*_{ee}}\right) \lambda (1-p_2) I^*_{ee} - (\mu + \alpha + \delta) Q^*_{ee} \\ 0 &= \sigma E^*_{ee} - (\lambda(1-p_2) + \mu + \gamma) I^*_{ee} \\ 0 &= \gamma I^*_{ee} + \delta Q^*_{ee} - (\rho + \mu) R^*_{ee} \end{aligned} \tag{7}$$

and its solution,

$$\begin{aligned} S^*_{eq} &= \frac{\Lambda}{\mu} + \frac{\sigma}{\mu}\left[\beta_2\left(\frac{\rho\delta}{\rho+\mu} - (\mu+\alpha+\delta)\right) + \frac{1}{\beta_1}\left(\frac{\rho\gamma}{\rho+\mu} + \lambda(1-p_2)\right) - \frac{\sigma+\mu}{\sigma}\right] E^*_{ee}, \\ Q^*_{ee} &= \sigma\beta_2 E^*_{ee}, \\ I^*_{ee} &= \frac{\sigma}{\beta_1} E^*_{ee}, \\ R^*_{ee} &= \frac{\sigma}{\rho+\mu}\left(\frac{\gamma}{\beta_1} + \delta\beta_2\right) E^*_{ee}, \end{aligned}$$

$\forall\, E^*_{eq} \in \mathbb{R}^+$, and

$$\begin{aligned} \beta_1 &= \lambda(1-p_2) + \mu + \gamma, \\ \beta_2 &= \frac{1}{(\mu+\alpha+\delta)}\left[\frac{\lambda(1-p_2)}{\beta_1} + \frac{(\sigma+\mu)\beta_s(1-p_2)p1\lambda}{\beta_1\beta_e+\sigma[\beta_a p_2+(\beta_s(1-p_2)(1-p_1\lambda))]}\right]. \end{aligned}$$

Note that the quarantined subpopulation is typically non-zero at the endemic equilibrium point but it could be eventually zero. In particular, note from (3) that the quarantined subpopulation at the endemic equilibrium point could be zero in the event that p2=1.

3.2 Stability of Equilibrium States

The stability of a system can be characterized by a bounded output, or a tend to an equilibrium point. However, its mathematical definition depends on the system characteristics (algebraic/dynamical, linear/nonlinear,...), and the type of stability. There are 3 common stability definitions for the case of nonlinear dynamical systems; the Lyapunov stability with respect to the equilibrium point, the structural stability, and the orbital stability. Specially in this work, it is important to calculate the stability of the system (3) around its equilibrium points P_{dfe} and P_{ee} because, by analysing the conditions for stability it will be possible to obtain indicators that could prevent from a disease growth.

Let's consider the system

$$\dot{\mathbf{x}} = f(t, \mathbf{x}) \tag{8}$$

with the equilibrium point $\mathbf{x}^*$, such that $f(t, \mathbf{x}^*) = 0$.

Given $\delta > 0$ and $\epsilon > 0$, it is said that $\mathbf{x}^*$ is stable if the next condition is satisfied [21, 7],

$$|\mathbf{x} - \mathbf{x}^*| < \delta \quad \text{at } t_0 \Rightarrow |\mathbf{x} - \mathbf{x}^*| < \epsilon \quad \forall\, t > t_0. \tag{9}$$

The system is said to be asymptotically stable around $\mathbf{x}^*$ if

$$|\mathbf{x} - \mathbf{x}^*| < \delta \quad \text{at } t_0 \Rightarrow \mathbf{x} \to \mathbf{x}^* \quad \text{as } t \to \infty. \tag{10}$$

Theorem 3.1. *If the eigenvalues of a linear system $\dot{\mathbf{x}}(t) = A\mathbf{x}(t)$ have negative real part, then the system is asymptotically stable.*

There are two Lyapunov theorems which determine the asymptotic stability of a nonlinear system [7]. The first case is needed from the jacobian of the system and its eigenvalues, and the second one uses a Lyapunov function $V(t, \mathbf{x})$, which can be interpreted as the system's energy representation. Even if in some cases it is not straightforward obtaining a representation for the system's energy, as discussed in [20, 34] approaches for Lyapunov functions are presented.

Considering that (8) is a nonlinear system, the linear approximation about the equilibrium point $\mathbf{x}^*$ gives

$$f(t, \mathbf{x}) - f(t, \mathbf{x}^*) = \frac{df(t, \mathbf{x})}{d\mathbf{x}}\Big|_{\mathbf{x}=\mathbf{x}^*}(\mathbf{x} - \mathbf{x}^*) \tag{11}$$

where $df(t, \mathbf{x})/d\mathbf{x}$ is the Jacobian matrix.

Theorem 3.2. *If the system is nonlinear and the eigenvalues of the Jacobian matrix,*

$$J = \frac{d\mathbf{f}}{d\mathbf{x}}\Big|_{x^*},$$

have negative real part, then the system is asymptotically stable in x^*.

Theorem 3.3. *If $V(t,\mathbf{x})$ is positive definite ($V(t,\mathbf{x}) > 0\ \forall\ x \neq 0,\ t > t_0$) and its time derivative is negative definite ($\dot{V}(t,\mathbf{x}) < 0\ \forall\ t > t_0$), then x^* is asymptotically stable.*

The stability of the system (3) will be analysed by using Theorems 3.2 and 3.3, so firstly the Jacobian will be calculated for a generic equilibrium point $P = (S^*, E^*, Q^*, I^*, R^*)$,

$$J = \begin{pmatrix} -(\beta_e E^* + (p_2\beta_a + (1-p_2)\beta_s)I^*)/N^* - \mu & -\beta_e S^*/N^* & 0 & -(p_2\beta_a + (1-p_2)\beta_s)S^*/N^* & \rho \\ (\beta_e E^* + (p_2\beta_a + (1-p_2)(1-p_1\lambda)\beta_s)I^*)/N & \beta_e S^*/N^* - (\mu+\sigma) & 0 & (\beta_a p_2 + \beta_s(1-p_2)(1-p_1\lambda))S^*/N^* & 0 \\ \beta_s(1-p_2)p_1\lambda I^*/N^* & 0 & -(\mu+\alpha+\delta) & \beta_s(1-p_2)p_1\lambda S^*/N^* + \lambda(1-p_2) & 0 \\ 0 & \sigma & 0 & -(\mu+\lambda(1-p_2)+\gamma) & 0 \\ 0 & 0 & \delta & \gamma & -(\rho+\mu) \end{pmatrix} \tag{12}$$

In the following two subsections, the generic equilibrium point will be substituted by the P_{dfe} and P_{ee}, respectively.

3.2.1 *Disease Free Equilibrium Point*

The disease free equilibrium point has all the population sizes equal to zero except S, so the resulting Jacobian matrix is the following one,

$$J = \begin{pmatrix} -\mu & -\beta_e S^*/N^* & 0 & -S^*(p_2\beta_a + (1-p_2)\beta_s) & \rho \\ 0 & \beta_e S^*/N^* - (\mu+\sigma) & 0 & (p_2\beta_a + (1-p_2)(1-p_1\lambda)\beta_s)S^*/N^* & 0 \\ 0 & 0 & -(\mu+\alpha+\delta) & \beta_s(1-p_2)p_1\lambda S^*/N^* + \lambda(1-p_2) & 0 \\ 0 & \sigma & 0 & -(\mu+\gamma+\lambda(1-p_2)) & 0 \\ 0 & 0 & \delta & \gamma & -(\rho+\mu) \end{pmatrix}$$

The following expression is obtained when solving the eigenvalues problem $\det(J - \lambda' I) = 0$,

$$(-\mu-\lambda')(-\rho-\mu-\lambda')(-\mu-\alpha-\delta-\lambda')\left[\left(\frac{\beta_e S^*}{N^*} - \mu - \sigma - \lambda'\right)(-\mu-\gamma-\lambda(1-p_2)-\lambda') - \sigma(p_2\beta_a + (1-p_2)(1-p_1\lambda)\beta_s)\frac{S^*}{N^*}\right] = 0$$

which three of its roots are straightforward, that is:

$$\lambda'_{1,2,3} = -\mu,\ -(\rho+\mu),\ -(\mu+\alpha+\delta),$$

which will always be negative because the implicated parameters are positive, see (5), therefore these eigenvalues are not responsible of any instability at the given disease free equilibrium point. To see how $\lambda'_{4,5}$ affect to the stability, instead of calculating their exact values, the Routh-Hurwitz criterion has been applied, which is a method to determine the conditions in which the roots become negative/positive. First, let's take the second order system and develop it,

$$\left(\tfrac{\beta_e S^*}{N^*} - \mu - \sigma - \lambda'\right)(-\mu-\gamma-\lambda(1-p_2)-\lambda') - \sigma(p_2\beta_a + (1-p_2)(1-p_1\lambda)\beta_s)\tfrac{S^*}{N^*} = 0 \rightarrow$$
$$\lambda'^2 + \left(2\mu + \lambda(1-p_2) + \gamma + \sigma - \tfrac{\beta_e S^*}{N^*}\right)\lambda + \left[\left(\mu + \sigma - \tfrac{\beta_e S^*}{N^*}\right)(\mu+\gamma+\lambda(1-p_2)) - \sigma(p_2\beta_a + (1-p_2)(1-p_1\lambda)\beta_s)\tfrac{S^*}{N^*}\right].$$

Its terms are the following ones:

$$\begin{aligned} a_2 &= 1, \\ a_1 &= 2\mu + \lambda(1-p_2) + \gamma + \sigma - \tfrac{\beta_e S^*}{N^*}, \\ a_0 &= \left(\mu + \sigma - \tfrac{\beta_e S^*}{N^*}\right)(\mu+\gamma+\lambda(1-p_2)) - \sigma(p_2\beta_a + (1-p_2)(1-p_1\lambda)\beta_s)\tfrac{S^*}{N^*} \end{aligned}$$

To get the conditions for stability the Routh Hurwitz criteria needs from the next set of determinant values

$$\begin{aligned} b_0 &= a_2 \\ b_1 &= a_1 \\ b_2 &= -\frac{\begin{vmatrix} b_0 & a_0 \\ b_1 & 0 \end{vmatrix}}{b_1} = a_0, \end{aligned}$$

and the criterion stays that the number of eigenvalues with positive real part is equal to the number of sign changes between b_0, b_1 and b_2. As $b_0 > 0$, the terms b_1 and b_2 must be positive to ensure that all eigenvalues have negative real parts, therefore the conditions for the disease free equilibrium stability are

$$-\frac{\beta_e S^*}{N^*} + 2\mu + \sigma + \gamma + \lambda(1-p_2) > 0 \quad \text{and}$$
$$\left(\mu + \sigma - \frac{\beta_e S^*}{N^*}\right)(\mu + \gamma + \lambda(1-p_2)) - \sigma(\beta_a p_2 + \beta_s(1-p_2)(1-p_1\lambda))\frac{S^*}{N^*} > 0.$$

From the second condition another case is inferred: $(\mu + \sigma - \beta_s S^*/N^*)$ must be positive, which is more restrictive than the first condition. Therefore, the stability criteria is obtained from the second condition, that is

$$\left[\frac{\beta_e}{(\mu+\sigma)} + \frac{\sigma(p_2\beta_a + (1-p_1\lambda)\beta_s)}{(\mu+\gamma+\lambda(1-p_2))(\sigma+\mu)}\right]\frac{S^*}{N^*} < 1 \tag{13}$$

where S^* and N^* are equal to Λ/μ.

The condition given by (13) shows that the *disease-free equilibrium* point might become unstable with an increase of the parameters β_e, β_a and β_s, which values are proportional to the average number of contacts per day. In contrast, an increase of p_1 can make the *disease-free equilibrium* point stable which agrees with its physical meaning.

3.2.2 *Endemic Equilibrium Point*

The complexity of the matrix (12) has been reduced by assuming that $\rho \to 0$ with respect to the other parameters involved on the system, so it makes easier analyzing the positivity of the eigenvalues of (12) when the system is at the endemic equilibrium point. The following equivalence is obtained when solving the eigenvalue problem $\det(J - \lambda' I) = 0$,

$$(\mu+\lambda')(\mu+\delta+\alpha+\lambda')\begin{vmatrix} -(\beta_e E^* + I^*(p_2\beta_a + (1-p_2)\beta_s))/N - \mu - \lambda' & -\beta_e S^*/N & -(p_2\beta_a + (1-p_2)\beta_s)S^*/N \\ (\beta_e E^* + I^*(p_2\beta_a + (1-p_2)(1-p1\lambda)\beta_s))/N & \beta_e S^*/N - (\mu+\sigma) - \lambda' & (p_2\beta_a + (1-p_2)(1-p_1\lambda)\beta_s)S^*/N \\ 0 & \sigma & -(\mu+\lambda(1-p_2)+\gamma) - \lambda' \end{vmatrix} = 0$$

where 2 eigenvalues are direct,

$$\lambda'_{1,2} = -\mu, -(\mu+\delta+\alpha).$$

The eigenvalues $\lambda'_{1,2}$ will always remain negative. The positivity/negativity of the eigenvalues $\lambda'_{3,4,5}$ will be calculated by applying the Routh Hurwitz criteria to the following equation.

$$a_3\lambda'^3 + a_2\lambda'^2 + a_1\lambda' + a_0 = 0$$

where

$$a_3 = -1$$
$$a_2 = \beta_e\left(\frac{S^*}{N^*} - \frac{E^*}{N^*}\right) - \beta_3\frac{I^*}{N^*} - \sigma - \mu - \beta_1$$
$$a_1 = \left[\beta_e\frac{E^*}{N^*} + \beta_3\frac{I^*}{N^*} + \mu\right]\left[\beta_e\frac{S^*}{N^*} - \sigma - \mu - \beta_1\right] + \beta_1\left[\beta_e\frac{S^*}{N^*} - \sigma - \mu\right] - \beta_e\left[\beta_e\frac{E^*}{N^*} - \beta_4\frac{I^*}{N^*}\right]\frac{S^*}{N^*} + \sigma\beta_4\frac{S^*}{N^*}$$
$$a_0 = \left[\beta_e\frac{E^*}{N^*} + \beta_3\frac{I^*}{N^*} + \mu\right]\left[\left(\beta_e\frac{S^*}{N^*} - \sigma - \mu\right)\beta_1 + \sigma\beta_4\frac{S^*}{N^*}\right] - \left[\beta_e\frac{E^*}{N^*} + \beta_4\frac{I^*}{N^*}\right][\beta_e\beta_1 + \sigma\beta_3]\frac{S^*}{N^*}$$

where,

$$\beta_3 = p_2\beta_a + (1-p_2)\beta_s, \quad \beta_4 = p_2\beta_a + (1-p_2)(1-p_1\lambda)\beta_s,$$

and β_1 and β_2 have been previously defined in the section 3.1.

The terms for the stability are

$$
\begin{aligned}
b_0 &= a_3 \\
b_1 &= a_2 \\
b_2 &= -\frac{\begin{vmatrix} b_0 & a_1 \\ a_2 & a_0 \end{vmatrix}}{a_2} = \frac{a_1 a_2 - b_0 a_0}{a_2} \\
b_3 &= -\frac{\begin{vmatrix} a_2 & a_0 \\ b_2 & 0 \end{vmatrix}}{b_2} = a_0
\end{aligned}
$$

the term $b_0 = -1$ is negative, so all remaining terms must be also negative to ensure the system's stability. In the Appendix A it is proved that a_2 and a_0 are negative for any possible parameter value, and hence the condition for stability will be given by b_2. Since a_2 is negative, b_2 will be negative if $a_1 a_2 + a_0$ is positive or $-a_1 a_2 / a_0 > 1$, that is:

$$
\frac{\left\{-\frac{\beta_e \beta_1 (\sigma+\mu)}{\beta_1 \beta_e + \sigma \beta_4}\left(\frac{\beta_5 \sigma}{\beta_1}\frac{E^*}{N^*} + \mu\right) + \left[(\beta_e \beta_1 + \sigma\beta_3)\frac{1}{\beta_1}\frac{E^*}{N^*} + \mu\right](\sigma + \mu + \beta_1)\right\}\left\{(\beta_e\beta_1 + \sigma\beta_3)\frac{1}{\beta_1}\frac{E^*}{N^*} + \frac{\beta_1^2 + \sigma\beta_4(\sigma+\mu+\beta_1)}{\beta_e\beta_1 + \sigma\beta_4}\right\}}{(\sigma+\mu)(\beta_e\beta_1 + \sigma\beta_3)\frac{E^*}{N^*}} > 1
$$

where the parameters β_3, β_4 and β_5 are defined as,

$$
\beta_3 = p_2\beta_a + (1-p_2)\beta_s, \quad \beta_4 = p_2\beta_a + (1-p_2)(1-p_1\lambda)\beta_s \quad \text{and} \quad \beta_5 = (1-p_2)p_1\lambda\beta_s.
$$

3.3 Basic Reproduction Number

The basic reproduction number $\mathcal{R}_0$ is a threshold which value determines whether the disease free equilibrium (DFE) is locally asymptotically stable or not. More precisely, when $\mathcal{R}_0 > 1$ is it said that the DFE is unstable and the disease could invade the population, on the other hand if $\mathcal{R}_0 < 1$ the DFE is locally asymptotically stable . It is explained in [29] how the basic reproduction number is defined, and it is also shown its mathematical proof and examples of how to apply it in different cases. To obtain the mathematical definition of $\mathcal{R}_0$, the theoretical development from [29] has been studied and applied; it is proved that $\mathcal{R}_0$ is equal to the spectral radius of $V^{-1}F$, see appendix C.

To obtain V and F firstly the vectors $\mathcal{F}$ and $\mathcal{V}$ which indicate the rate of appearance of new infections and the transfer (in and out) of individuals, respectively, in the compartments E, Q, and I must be defined[1],

$$
\mathcal{F} = \begin{pmatrix} (\beta_e E + (p_2\beta_a + (1-p_2)(1-p_1\lambda)\beta_s)I)S/N \\ (1-p_2)p_1\lambda\beta_s IS/N \\ 0 \end{pmatrix}, \quad \mathcal{V} = \begin{pmatrix} (\mu+\sigma)E \\ (\mu+\delta+\alpha)Q \\ (\lambda(1-p_2)+\mu+\gamma)I - \sigma E \end{pmatrix}.
$$

Note that $\mathcal{F}_3 = 0$ because in the compartment I there is no new infections, that is; E and Q present the new infections, and the infected individuals of I are a consequence (they were already infected) of people from E and Q. The matrix F and V are the derivative with respect to $Z = [E, Q, I]$ of $\mathcal{F}$ and $\mathcal{V}$, respectively, around P_{dfe},

$$
F = \begin{pmatrix} \beta_e S/N & 0 & (\beta_a p_2 + (1-p_2)(1-p_1)\beta_s)S/N \\ 0 & 0 & (1-p_2)\beta_s p_1 S/N \\ 0 & 0 & 0 \end{pmatrix}\Bigg|_{P_{dfe}},
$$

$$
V = \begin{pmatrix} \mu+\sigma & 0 & 0 \\ 0 & (\mu+\delta+\alpha) & 0 \\ -\sigma & 0 & (\lambda(1-p_2)+\mu+\gamma) \end{pmatrix}\Bigg|_{P_{dfe}}
$$

1 As far as F and V only hold the infected compartments, the expression of $\mathcal{F}$ and $\mathcal{V}$ will be reduced to those compartments too.

In the appendix D the steps to calculate the spectral radius of $V^{-1}F$ are presented, and the result is,

$$\mathcal{R}_0 = \rho(V^{-1}F) = \left[\frac{\beta_e}{(\mu+\sigma)} + \frac{\sigma(\beta_a p_2 + (1-p_2)(1-p_1\lambda)\beta_s)}{(\lambda(1-p_2)+\mu+\gamma)(\sigma+\mu)}\right]\frac{S}{N}\Bigg|_{S=N=\Lambda/\mu} \tag{14}$$

and it has been proved that it matches with the numerical results, the following repository[2] path CoVid-19-Analysis/code/checks/spectral_radius.m holds the Matlab script used, and it is equal to the result (13), so it demonstrates the consistency of the methodology used to calculate the system's stability condition via the matrices $\mathcal{F}$ and $\mathcal{V}$.

3.3.1 *Simulation*

The numerical solution of the system (3) needs the values of its parameters and the initial condition of its variables. Therefore, for a preliminary simulation, data from different sources has been gathered, see Table 1, and Table 2 presents the values used for the variables initial conditions. Note that the values of the parameters Λ/N and μ appearing in Table 1 correspond to the birth rate and natural death rate, respectively, of Spain. Moreover, the initial susceptible population, see Table 2, is approximately equal to the total population of Spain.

To complete the missing parameter's values it has been assumed that:

- The contact rate is the same for members of E and I, including symptomatic and asymptomatic people. Thus, $c_a = c_s = c_e = c_r$.
- The transmission probability of exposed people p_e is 10^{-3} times lower than p_s.
- The transmission probability of asymptomatic people p_a is equal to p_s.
- $\lambda^{-1} = 2\,days$.
- The tracing effectiveness p_1 is 0.5
- The average time in quarantine δ^{-1} is equal to $\sigma^{-1} + \gamma^{-1}$

Parameter	Definition	Value	Cite
Λ/N	Birth rate	$2.088 * 10^{-5}\,day^{-1}$	[14]
c_r	Contact rate	$10.58\,day^{-1}$	[17]
p_s	Transmission probability of infected people	0.23	[30]
p_2	Probability to be asymptomatic	0.16	[18, 5]
μ	Natural death rate	$2.282 * 10^{-5}\,day^{-1}$	[14]
σ^{-1}	Incubation period	$7\,days$	[17]
γ	Removal rate	$0.1\,day^{-1}$	[8]
α	Disease-induced death rate	$0.01\,day^{-1}$	[17]

Table 1: Values of the model parameters.

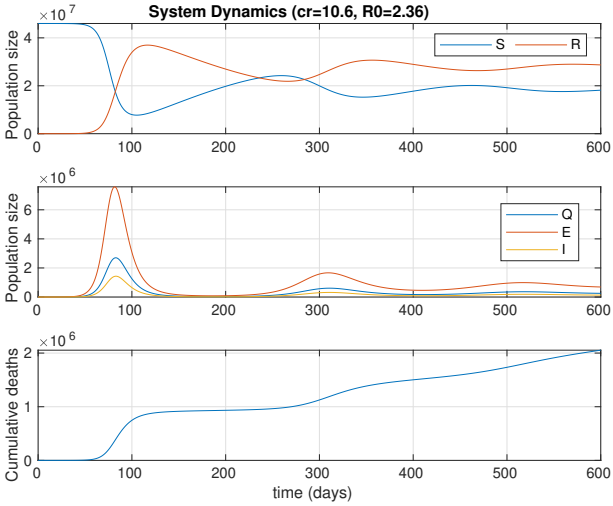
Once all the parameters values have been introduced in the code, and the system (3) solved with the function *ode45()* from *Matlab* the results have been depicted in Figure 3(a); it is shown the dynamics of the variables S, R, Q, E, I and the cumulative number of deaths with respect to time. The variables show characteristics of an underdamped dynamical system; its response is a sum of a transient response (damped sinusoidal) and the steady state. In this case the disease persists, which is an expected result since $R_0 = 2.36$. It is important to notice that, approximately 300 days after the outbreak, the total number of deaths reach approximately 1 million, and at the end of the simulation (600 days after the outbreak) it doubled.

The disease spreading can be stopped by reducing the value of c_r, therefore in the next simulation c_r was reduced to 4 ($\mathcal{R}_0 = 0.89$), and as it can be seen in Figure 3(b) the disease tends to disappear;

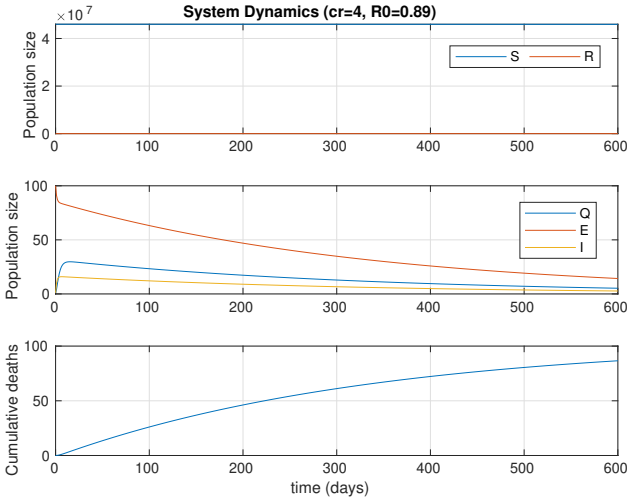
2 All the scripts created on this work have been uploaded to a GitHub repository, which link is https://github.com/CarmenLegarre/CoVid-19-Analysis

Variable	Definition	Value
S	Susceptible population	$46 * 10^6$
E	Exposed population	100
Q	Quarantine population	1
I	Infected population	1
R	recovered population	0

Table 2: Model variables initial conditions



(a) Endemic steady state



(b) Free-disease steady state

Figure 3: S, R, E, I, Q population sizes and the cumulative deaths with respect to time when $c_r = 10.6$ and $\mathcal{R}_0 = 2.360$(a), and $c_r = 4$ and $\mathcal{R}_0 = 0.89$ (b). The initial conditions in both cases are those that are shown in Table 2.

the variables Q, E and I show and overdamped response, and their values continuously decrease with respect to time. As long as, their values and the total number of deaths are small compared to S long over the simulation, the variable S seems to be unchanging, see the top graph of Figure 3(b).

The results of both simulations corresponds to the expected behaviour in the cases of $\mathcal{R}$ greater and lower than 1. Besides, considering the cumulative number of deaths of both simulations, it is clear that reducing the value of c_r could prevent a huge increase of the total deaths. See the next repository folder CoVid-19-Analysis/code/solve1 to see Matlab script used to simulate and visualize the results.

4 DATA ACQUISITION AND PROCESSING

4.1 Data Acquisition

The official database from Cantabria [9] offers a unique file (*COVID19_historico.csv*) which holds information about the cumulative number of performed tests $t_c(k)$, the cumulative number of positive results $p_c(k)$, the daily number of hospitalized people $ph_d(k)$, the daily number of people at the intensive care unit (ICU)[3] $i_d(k)$, the cumulative number of deaths $d_c(k)$ and the cumulative number of recovered people $r_c(k)$ (the subindexes d and c are used to indicate if the data is given in daily or cumulative form, respectively). Moreover, the mentioned information can easily be extracted from *COVID19_historico.csv* and it doesn't present data gaps. Considering that $y_d(k)$, where $y(k) = \{t(k), p(k), d(k), r(k)\}$, is approximately equal to the time derivative of $y_c(k)$, the Euler forward approximation can be applied to calculate $y_d(k)$ from $y_c(k)$, that is:

$$y_d(k+h) = \frac{y_c(k+h) - y_c(k)}{h}, \tag{15}$$

where h is the step size and its equal to 1 day. After applying the expression (15), next data is obtained; the number of total tests per day $t_d(k)$, the positive results per day $p_d(k)$, death people per day $d_d(k)$ and recovered ones $r_d(k)$ per day, and it has been depicted in Figure 4 within $ph_d(k)$ and $i_d(k)$. If the context is clear, the subindex d is omitted.

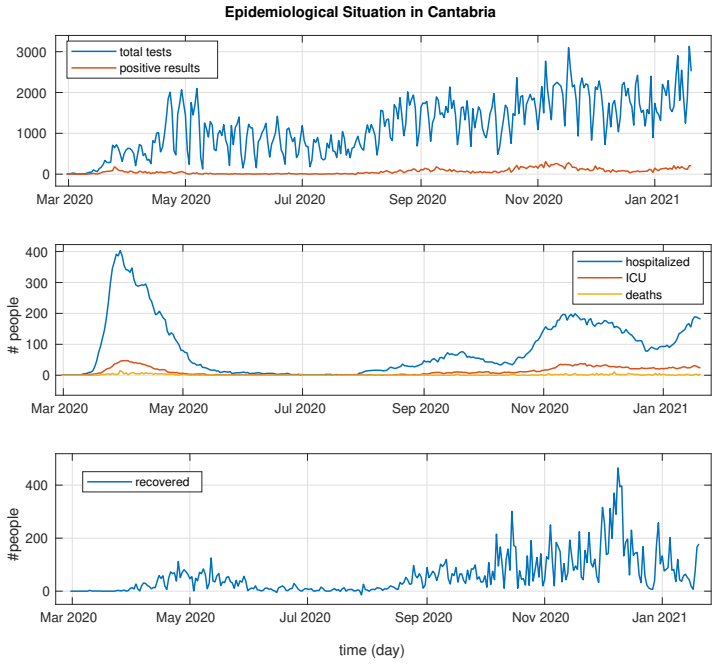


Figure 4: Epidemiological situation in Cantabria.

A variable of interest, and as it will be later proved, achievable from the given data, is the number of people in quarantine $Q(k)$. This subpopulation is important because it can be used to infer the values of many parameters (e.g the number of death and recovered people per day are equal to $\alpha Q(k)$ and $\delta Q(k)$, respectively, so the parameters α and δ can be estimated by applying the least-squares method to the subsets $\{Q(k), d(k)\}$ and $\{Q(k), r(k)\}$).

To estimate the number of people in quarantine $\hat{Q}(k)$, the discrete form of expression $\dot{Q}(t)$ described in (3) has been considered. From (3) it can be seen that $\dot{Q}(t)$ depends on the term S/N, however by the 14/02/2021 the sum of the cumulative number of recovered and death people was approximately 25.500, which is only a 5% of the total population of Cantabria. Therefore, it has been assumed that

3 In system (3) hospitalized people and the number of people at ICU do not appear, however it is interesting showing it since most of the control measurements are implemented to avoid higher occupation

$S/N \approx 1$ during the period from the 29/02/2020 to the 14/02/2021. Regarding the outputs $p(k)$, $d(k)$ and $r(k)$, they might hold some delays (i.e. when a symptomatic individual tests positive and is put in quarantine, it is after some days when the positive result is reported in official data) and noise w. Thus,

$$p(k) = \lambda(1-p_2)(1+p_1 p_s c_r)I(k-\tau_1)+w, \quad d(k) = (\alpha+\mu)Q(k-\tau_2)+w, \quad \text{and} \quad r(k) = \delta Q(k-\tau_3)+w, \tag{16}$$

where τ_1, τ_2 and τ_3 are positive (necessary condition to ensure causality). Note that in the expression (16), β_s has been substituted by $p_s c_r$ since c_r (the average number of close contacts per day) will be a parameter of interest. To reduce noise effects of the outputs, the centered moving mean filter, which takes 14 samples, has been applied to $p(k)$, $d(k)$ and $r(k)$ so the estimations $\hat{p}(k)$, $\hat{d}(k)$ and $\hat{r}(k)$, respectively, have been derived. Considering the discrete expression of $\dot{Q}(t)$,

$$\dot{Q}(k) = \lambda(1-p_2)(1+p_1 p_s c_r)I(k) - (\alpha+\delta+\mu)Q(k) \tag{17}$$

and the filtered data, the estimation $\hat{\dot{Q}}(k)$ is inferred

$$\hat{\dot{Q}}(k) = \hat{p}(k+\tau_1) - \hat{d}(k+\tau_2) - \hat{r}(k+\tau_3). \tag{18}$$

The Euler forward approximations applied to (18) gives,

$$\hat{Q}(k+1) = \hat{Q}(k) + h\left[\hat{p}(k+\tau_1) - \hat{d}(k+\tau_2) - \hat{r}(k+\tau_3)\right] \tag{19}$$

where h is the step size and is equal to 1 day.

By applying the delay τ_2 to (19), it can be seen that without further information only the relative delays are possible to calculate, that is:

$$\hat{Q}(k+1-\tau_2) = \hat{Q}(k-\tau_2) + \hat{p}(k+\tau_1-\tau_2) - \hat{d}(k) - \hat{r}(k+\tau_3-\tau_2) \tag{20}$$

since $\hat{d}(k)$ and $\hat{r}(k+\tau_3-\tau_2)$ are proportional to $Q(t-\tau_2)$, the delay τ_4 between the signals $\hat{d}(k)$ and $\hat{r}(k)$ will be equal to $\tau_2-\tau_3$. With $\hat{d}(k)$ and $\hat{r}(k-\tau_4)$, if a delay τ_5 is applied to $\hat{p}(k)$, where $\tau_5 = \tau_1-\tau_2$, the estimated $\hat{Q}(k-\tau_2)$ should be proportional to $\hat{d}(k)$ (or $\hat{r}(k-\tau_4)$). This knowledge makes possible to create an algorithm that calculates the best τ_4 and τ_5 values. Note that it is not possible to infer τ_2. At different time periods distinct delays might appear, that is: as far as the coexistence with Covid-19 increases, it is expected more rapid responses so the delay to report the number of positive tests, the number of death and recovered people is decreased.

The next steps have been used to determine the delays:

1. Divide the dataset $\{p(k), d(k), r(k)\}$ into subsests (e.g. first data set $\{p^{(1)}(k), d^{(1)}(k), r^{(1)}(k)\}$ belongs to the period from the 29/02/2020 to the 17/07/2020, the second data set $\{p^{(2)}(k), d^{(2)}(k), r^{(2)}(k)\}$ to the period from the 18/07/2020 to the 31/12/2020, etc), so the delays τ_4 and τ_5 will be calculated for each subset. As far as the delays might change with respect to time, this step could ensure a more accurate calculation of τ_4 and τ_5.

2. For each subset, seek for any delay between $\hat{r}(k)$ and $\hat{d}(k)$ with the Matlab's function *finddelay()*. The result given by the function *finddelay()* will be equal to τ_4.

3. Determine a set of possible values T_5 and calculate $\hat{p}(k-\tau_5)$ for all $\tau_5 \in T_5$, and then calculate $\hat{Q}(k-\tau_2)$

4. Apply the linear regression method to the data sets $\{\hat{Q}(k-\tau_2), \hat{r}(k-\tau_4)\}$ and $\{\hat{Q}(k-\tau_2), \hat{d}(k)\}$, compute the coefficient of determination $R_i^{2(r)}$ and $R_i^{2(d)}$, respectively. The coefficient of determination (R^2) is the proportion of variability in the set of the given data explained by the prediction [32, 27], and it is a variable used to measure the goodness of the linear regression; when its value is near 1, it means that there is a low variability with respect to the linear regression, so there is a good prediction. In contrast, when its value is near 0, there is a high variability and a bad prediction.

It shouldn't be expected a delay τ_6 between $\hat{Q}(k-\tau_2)$ and $\hat{r}(k-\tau_4)$ (or $\hat{d}(k)$), however $\hat{Q}(k-\tau_2)$ is calculated from filtered data and the applied delays might not be exact to the real ones. As long as a nonzero τ_6 could deteriorate severally the coefficient of determination values, before applying the linear regression the function *finddelay()* will be used to find τ_6 and align $\hat{Q}(k-\tau_2)$ with respect to $\hat{d}(k)$ and $\hat{r}(k-\tau_4)$.

5. Determine the function

$$f_i = \frac{R_i^{2(r)} + R_i^{2(d)}}{2}$$

so τ_5 is equal to the value τ_i that maximize the function f_i. Thus,

$$\tau_5 = \arg\max_{\tau_i} f_i$$

Two data sets have been selected: the first one gathers information from the 29/02/2020 to the 25/07/2020, and the second one from the 25/07/2020 until the 14/02/2021. For each data set the previously enumerated steps have been applied to determine the delays and Q. Thus,

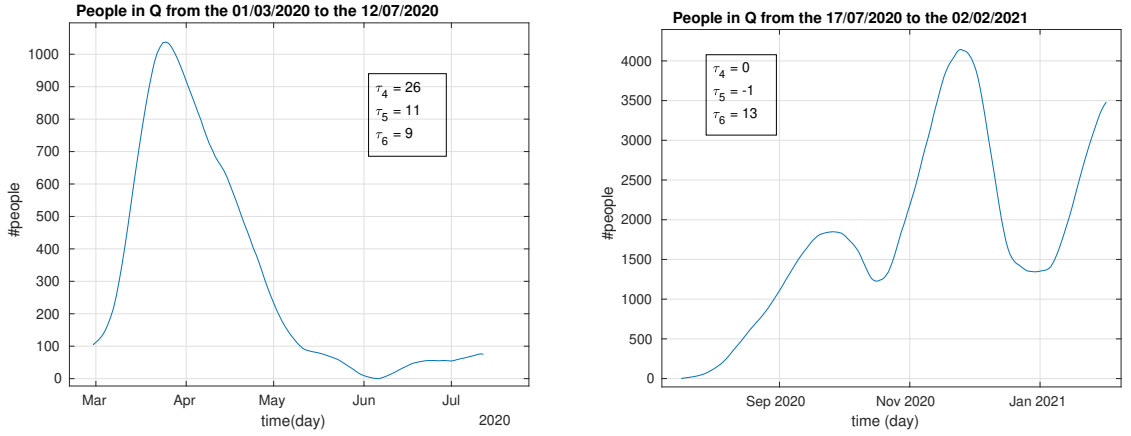
$$\tau_4, \tau_5, \tau_6 = \begin{cases} \tau_4^{(1)}, \tau_5^{(1)}, \tau_6^{(1)} \text{ for } k \in [29/02/2020, 12/07/2020] \\ \tau_4^{(2)}, \tau_5^{(2)}, \tau_6^{(2)} \text{ for } k \in [17/07/2020, 02/02/2021] \end{cases}$$

4.2 Estimated delays

After following the steps described in the previous section the next delays have been obtained,

$$\tau_4^{(1)} = 26 \text{ days}, \quad \tau_5^{(1)} = 11 \text{ days} \quad \text{and} \quad \tau_6^{(1)} = 9 \text{ days}, \tag{21}$$

and the number of people in quarantine due to a positive test result is displayed in Figure 5 (a).



(a) Period between the 01/03/2020 and 12/07/2020.

(b) Period between the 17/07/2020 and 02/02/2021.

Figure 5: Estimated people in quarantine after testing positive for CoVid-19 at different time periods.

Regarding the second data subset no delay was found between $d(k)$ and $r(k)$, and after analysing the function f_i for different delays values, the delays

$$\tau_4^{(2)} = 0 \text{ days}, \quad \tau_5^{(2)} = -1 \text{ days} \quad \text{and} \quad \tau_6^{(2)} = 13 \text{ days}, \tag{22}$$

have been obtained. The number of people in quarantine during this period is depicted in Figure 5 (b). The value of $\tau_5^{(2)}$ implies that the estimated output $\hat{d}(k)$ is one sample delayed with respect $\hat{p}(k)$, and its negative value does not generate a causality problem as far as $\tau_2^{(2)} - 1 > 0$.

See the next repository paths CoVid-19-Analysis/code/parameter_estimation/data_processing/first_wave and CoVid-19-Analysis/code/parameter_estimation/data_processing/nex_waves to see the Matlab scripts.

4.3 Discussion

The high values of $\tau_6^{(1)}$ and $\tau_6^{(2)}$ might signalize that the model described by (3) is not able to reproduce all the real epidemiological behaviour. However a hypothesis that could explain that delays will be expound, and it will be also proved that it is also possible to continue with the constructed model. It will be assumed that the delays are caused because the stage Q holds sub-stages, as it is shown

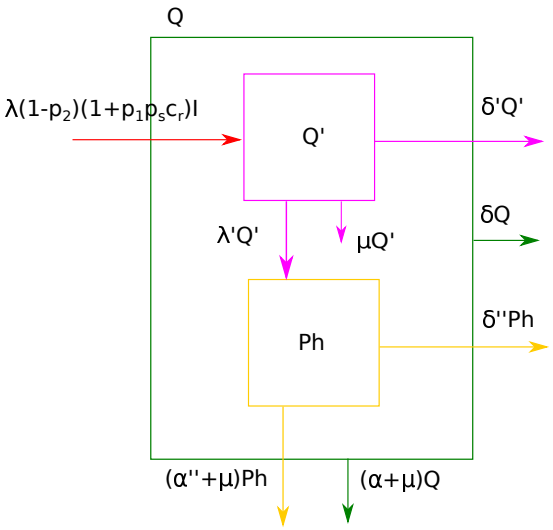


Figure 6: Flow chart of the compartment Q (green box) and its sub-stages Q' (pink box) and Ph (yellow box).

in Figure 6; symptomatic people and traced ones $(\lambda(1-p_2)(1+p_1p_sc_r)I)$ go to Q', which represents people in quarantine that stay at home, some of them $(\delta' Q')$ will recover and others $(\lambda' Q')$ will go to Ph, which embodies people that are at hospitals due to a health worsen. Then, some people $(\delta'' Ph)$ in Ph will recover and go to R after δ''^{-1} days and others $((\alpha''+\mu)Ph)$ will die. The dynamics of the subsystem Q' and Ph are described for the next differential equations,

$$\begin{aligned} \dot{Q}' &= \lambda(1-p_2)(1+p_1p_sc_r)I-(\lambda'+\delta'+\mu)Q' \\ \dot{Ph} &= \lambda' Q'-(\alpha''+\delta''+\mu)Ph. \end{aligned} \tag{23}$$

First the Laplace transform of $Q(s)$ and $Ph(s)$ of the functions $Q(t)$ and $Ph(t)$, respectively, will be calculated to show a possible existence of a delay τ_6 between $(\alpha''+\mu)Ph(t)$ and $(\alpha+\mu)Q(t)$, so $(\alpha+\mu)Q(t-\tau_6)=(\alpha''+\mu)Ph(t)$. Considering that $S(t)/N(t)\approx 1$ the following expression for $Q(s)$ is obtained from (3):

$$Q(s)=\frac{\lambda(1-p_2)(1+p_1p_sc_r)}{s+\alpha+\delta+\mu}I(s) \tag{24}$$

With the Laplace transform of the functions given in (23), $Ph(s)$ in terms of $I(s)$ is calculated,

$$Ph(s)=\frac{\lambda'\lambda(1-p_2)(1+p_1p_sc_r)}{(s+\lambda'+\delta'+\mu)(s+\alpha''+\delta''+\mu)}I(s). \tag{25}$$

Note that the model depicted in Figure 2 estimates that the number of death people due to Covid-19 is αQ. However, a more realistic model, see Figure 6, shows that the number of death people is given by $\alpha'' Ph$. The transfer function $H_1(s)=\alpha'' Ph(s)/\alpha Q(s)$ represents the relationship between the signals $\alpha'' Ph$ and αQ, and therefore, if it assumed that $\alpha'' Ph$ describes accurately the real number of death people, then the analysis of $H_1(s)$ could show how well the prediction αQ fits with reality, whether the relation is linear/nonlinear, etc. The transfer function $H_1(s)$ is given by

$$H_1(s)=\frac{\alpha''\lambda'(s+\alpha+\delta+\mu)}{\alpha(s+\lambda'+\delta'+\mu)(s+\alpha''+\delta''+\mu)} \tag{26}$$

by substituting $s = j\omega$, the amplitude response $|H_1(j\omega)|$ and phase response $\angle H_1(j\omega)$ are determined,

$$\begin{aligned} |H_1(j\omega)| &= \frac{\alpha''\lambda'}{\alpha}\sqrt{\frac{\omega^2+(\alpha+\delta+\mu)^2}{(\omega^2+(\lambda'+\delta'+\mu)^2)(\omega^2+(\alpha''+\delta''+\mu)^2))}} \\ \angle H_1(j\omega) &= \arctan\left(\frac{\omega}{\alpha+\delta+\mu}\right) - \arctan\left(\frac{\omega}{\lambda'+\delta'+\mu}\right) - \arctan\left(\frac{\omega}{\alpha''+\delta''+\mu}\right) \end{aligned} \tag{27}$$

Given a signal $Q(k)$ in terms of cosine and sine basis $\{\cos(\omega k), \sin(\omega k)\}_{\omega=0}^{\infty}$, and if all the parameter's values appearing in (27) were known, the amplitude response and the phase difference could be calculated mathematically. Supposing that the power spectrum of the signal $Q(k)$ is contained mainly in the frequency ω_0, the expression of $Q(k)$ can be written as follows:

$$Q(k) = A\cos(\omega_0 k + \phi),$$

and assuming that $\alpha''Ph$ describes accurately the number of death people $d(k)$, so $\alpha''Ph(k) \approx d(k)$, then,

$$d(k) \approx \alpha''Ph(k) = \alpha|H_1(j\omega_0)|A\cos(\omega_0 k + \phi + \tau_6)$$

where $\tau_6 = \angle H_1(j\omega_0)$.

In the same manner, it is possible to determine the phase difference between the number of recovered people given by $\delta Q(t)$ and $\delta'Q'(t) + \delta''Ph(t)$. In this case the transfer function $H_2(s) = (\delta'Q'(s) + \delta''Ph(s))/\delta Q(s)$ is defined as

$$H_2(s) = \frac{(s+\alpha+\delta+\mu)\left[\delta'(s+\alpha''+\delta''+\mu)+\delta''\lambda'\right]}{\delta(s+\lambda'+\delta'+\mu)(s+\alpha''+\delta''+\mu)} \tag{28}$$

where the amplitude and phase responses are the following ones,

$$\begin{aligned} |H_2(j\omega)| &= \frac{1}{\delta}\sqrt{\frac{[\omega^2+(\alpha+\delta+\mu)^2][(\delta'\omega)^2+(\delta'(\alpha''+\delta''+\mu)+\delta''\lambda')^2]}{(\omega^2+(\lambda'+\delta'+\mu)^2)(\omega^2+(\alpha''+\delta''+\mu)^2))}} \\ \angle H_2(j\omega) &= \arctan\left(\frac{\omega}{\alpha+\delta+\mu}\right) + \arctan\left(\frac{\omega}{(\alpha''+\delta''+\mu)+(\delta''/\delta')\lambda'}\right) - \arctan\left(\frac{\omega}{\lambda'+\delta'+\mu}\right) - \arctan\left(\frac{\omega}{\alpha''+\delta''+\mu}\right) \end{aligned} \tag{29}$$

Comparing $\angle H_1(j\omega)$ and $\angle H_2(j\omega)$ gives that the number of death people per day $\alpha''Ph(t)$ will be delayed with respect to the recovered people per day $\delta'Q'(t) + \delta''Ph(t)$. The algorithm used previously to infer τ_4, τ_5 and τ_6 doesn't consider the presence of such a delay, perhaps the computed $R^{(r)}$ and $R^{(d)}$ in the first period (from the 01/03/2020 to the 12/07/2020) are near 1. Therefore,

$$\arctan\left(\frac{\omega}{(\alpha''+\delta''+\mu)+(\delta''/\delta')\lambda'}\right) \approx 0.$$

It has been proved that the model described by the set of DE (3) doesn't fit with the real epidemiological behaviour, since delays appearing in real data haven't been taken into account. Nevertheless, a new "dummy" variable D can be introduced in the set of DE to include the delays, that is:

$$\begin{aligned} \dot{S} &= \Lambda + \rho R - [\beta_e E + I(p_2\beta_a + (1-p_2)\beta_s)]\frac{S}{N} - \mu S \\ \dot{E} &= (\beta_e E + I(p_2\beta_a + (1-p_2)(1-p_1\lambda)\beta_s))\frac{S}{N} - (\sigma+\mu)E \\ \dot{Q} &= (1-p_2)\beta_s p_1 \lambda I\frac{S}{N} + \lambda(1-p_2)I - (\mu+\alpha+\delta)Q \\ \dot{I} &= \sigma E - (\lambda(1-p_2)+\mu+\gamma)I \\ \dot{D} &= (\alpha+\delta+\mu)Q, \quad \text{for } 0 < t \leqslant \tau_6 \\ \dot{D} &= (\alpha+\delta+\mu)\left[Q - D(t-\tau_6)\right], \quad \text{for } t \geqslant \tau_6 \\ \dot{R} &= \gamma I + \delta D(t-\tau_6) - (\rho+\mu)R \end{aligned} \tag{30}$$

and in this case the number of recovered and death people per day will be given by $\delta D(t-\tau_6)$ and $(\alpha+\mu)D(t-\tau_6)$, respectively.

5 PARAMETERS ESTIMATION

5.1 Ready Made Models

Depending on the case different parametric models can be distinguished: tailor made models and ready made models [19]. In the case of tailor made models, the constructed models are based on physical principles, so their parameters have physical meaning. On the other hand, to create a ready made model it is necessary to guess which model structure fits better with the real system's responses, determine the model's order, which might be a difficult task without a previous knowledge of the system, and with the recorded set of inputs and outputs infer its parameters values. However these type of model responses are constrained to the set of inputs used for its creation.

In this study case, the model described by the set of equations (3) is a tailor made model; it describes the dynamics of population with different characteristics (e.g how infected population infects susceptible people). As far as some of its parameters, such as λ, δ, α, p_1 and c_r depend on the epidemic situation, the objective is to create a ready-made model from each situation, get its parameters and relate them with the tailor-model's ones.

Usually ready made models are based in discrete models since the data to be used for the estimation are sampled data. The output of generic linear discrete time system can be defined as follows,

$$y(k) = G(q,\theta)u(k) + H(q,\theta)w(k) \tag{31}$$

where $y(k)$, $u(k)$ and $w(k)$ are the output, input and noise, respectively, at the sampling time k, and $G(q,\theta)$ and $H(q,\theta)$ are rational functions of the shift operator q. Thus,

$$\begin{aligned} G(q,\theta) &= \frac{B(q,\theta)}{F(q,\theta)} = \frac{b_1 q^{-nk} + b_2 q^{-nk-1} + ... + b_{nb} q^{-nk-nb+1}}{1 + f_1 q^{-1} + ... + f_{nf} q^{-nf}}, \\ H(q,\theta) &= \frac{C(q,\theta)}{D(q,\theta)} = \frac{1 + c_1 q^{-1} + ... + c_{nc} q^{-nc}}{1 + d_1 q^{-1} + ... + d_{nd} q^{-nd}}. \end{aligned} \tag{32}$$

The parameters nb, nc, nd, nk and nf determine the model's order which must be chosen before starting with the parameter estimation, and the parameters vector θ, which contains b_i, f_i, c_i and d_i, are calculated so the best fit with the data is obtained.

In [19] different model structures are shown, such as ARMAX, which considers that $F(q,\theta) = D(q,\theta) = A(q,\theta) = 1 + a_1 q^{-1} + ... + a_{na} q^{-na}$, or ARX which is equal to ARMAX but considering that $C(q,\theta) = 1$. Let's consider the ARX model,

$$\begin{aligned} &A(q,\theta)y(k) = B(q,\theta)u(k) + w(k) \rightarrow \\ &y(k) + a_1 y(k-1) + ... + a_{na} y(k-na) = b_1 u(k-nk) + ... + b_{nb} u(k-nk-nb+1) + w(k) \end{aligned} \tag{33}$$

and it can be rewritten in terms of $\theta = [a_1, ..., a_{na}, b_1, ..., b_{nb}]^T$ and $\varphi = [-y(k-1), ..., -y(k-na), u(k-nk), ..., u(k-nk-nb+1)]$,

$$y(k,\theta) = \theta^T \varphi(k) + w(k). \tag{34}$$

5.2 Least-squares Method

Considering that the systems output $y(k)$ is ruled by an ARX model, it is possible to build a model that follows the same structure which will give a prediction $\hat{y}(k,\theta)$ of the real output, that is:

$$\hat{y}(k,\theta) = \hat{\theta}^T \varphi(k) \tag{35}$$

so the prediction error is defined as

$$e(k,\theta) = y(k) - \hat{y}(k). \tag{36}$$

Considering that inputs and outputs have been sampled at $k = 1, ..., N$, the expression for the least square error is

$$V_N(\theta) = \frac{1}{N}\sum_{k=1}^{N} e(k,\theta)^2. \tag{37}$$

The parameter vector $\hat{\theta}$ will be defined by those values that minimize (37),

$$\hat{\theta} = \arg\min_{\theta} V_N(\theta). \tag{38}$$

which solution is given by,

$$\hat{\theta}(k) = \left(\frac{1}{N}\sum_{k=1}^{N}\varphi(k)\varphi(k)^T\right)^{-1}\left(\frac{1}{N}\sum_{k=1}^{N}\varphi(k)y(k)\right) \tag{39}$$

5.3 Estimation of α and δ with Linear Regression

Since in the section 4.2 data was separated in two subsets, the estimation of the parameters α and δ will be carried at each data set. Thus,

$$\alpha, \delta \equiv \begin{cases} \alpha_1, \delta_1 & \text{for } k \in [29/02/2020,\ 12/07/2020] \\ \alpha_2, \delta_2 & \text{for } k \in [17/07/2020,\ 02/02/2021] \end{cases} \tag{40}$$

So, the previously estimated $\hat{Q}(k)$, $\hat{d}(k)$ and $\hat{r}(k)$, and their respective delays have been considered. In Figure 7 (a) and (b) the data sets $\{\hat{Q}(k-\tau_2), \hat{d}(k-\tau_4^{(1)}-\tau_6^{(1)})\}$ and $\{\hat{Q}(k-\tau_2), \hat{r}(k-\tau_6^{(1)})\}$ are depicted, respectively, for the period 01/03/2020-16/07/2020. In the same way, Figure 8 (a) and (b) show the data sets $\{\hat{Q}(k-\tau_2), \hat{d}(k-\tau_4^{(2)}-\tau_6^{(2)})\}$ and $\{\hat{Q}(k-\tau_2), \hat{r}(k-\tau_6^{(2)})\}$, respectively, but from the 17/07/2020 to the 02/02/2021. The linear regression method has been applied and the estimated parameter's values within their respective error and coefficient to determination are shown in Table 3. See the next repository paths CoVid-19-Analysis/code/parameter_estimation/data_processing/first_wave and CoVid-19-Analysis/code/parameter_estimation/data_processing/next_waves to see the used Matlab scripts.

Parameter	value	error	R^2
α_1	0.0063	0.0002	0.96
α_2	$7.1 * 10^{-4}$	$0.6 * 10^{-4}$	0.75
δ_1	0.057	0.002	0.97
δ_2	0.043	0.003	0.80

Table 3: Estimated parameters via linear regression.

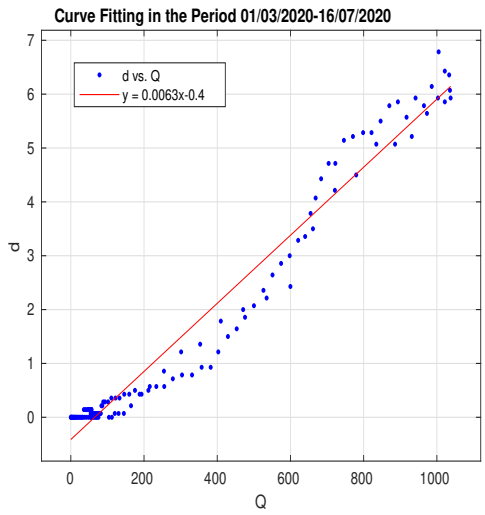
It is assumed that the death rate α_1 is higher than α_2 because in the first time period the percentage of people in Q that were hospitalized was higher, compare the estimated $\hat{Q}(k)$ depicted in Figure 5 with the number of hospitalized people given in Figure 4.

Regarding the recovery time, after the first wave tracing techniques were implemented; when someone with symptoms went to outpatient clinic or hospitals, his/her close contacts were tested to see if someone was infected. So people that were traced and tested positive were introduced in Q, and they stay there during the incubation time and their recovery time. This could cause a decrease of δ_2 with respect δ_1.

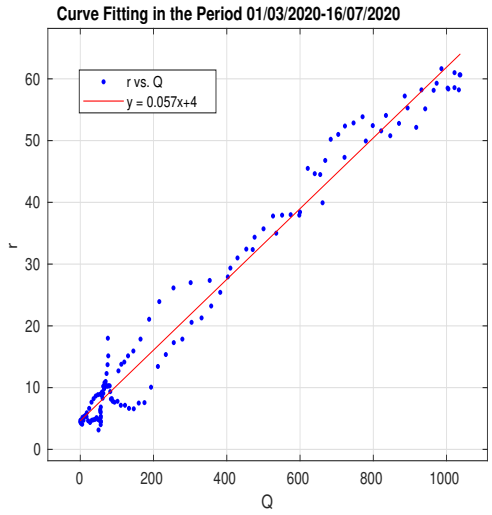
5.4 Estimation of λ, c_r and p_1

Considering that $S/N \approx 1$ the following ordinary differential expression (ODE) of $Q(t)$ can be obtained from the set of equations (3),

$$\dddot{Q}(t) + b_1\ddot{Q}(t) + b_2\dot{Q}(t) - b_3 Q(t) = 0 \tag{41}$$

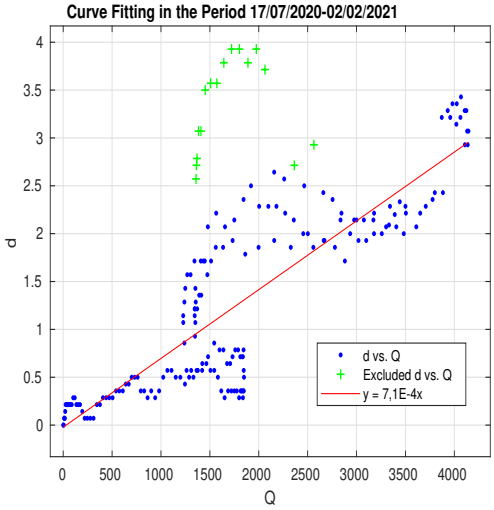


(a) Linear regression (red line) of the set $\{\hat{Q}(k-\tau_2), \hat{d}(k-\tau_4^{(1)}-\tau_6^{(1)})\}$ (blue spots).

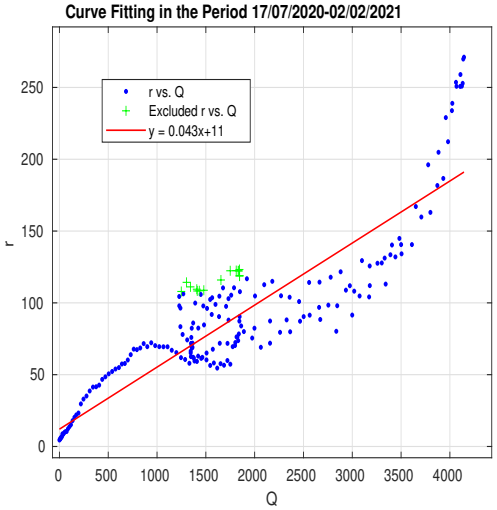


(b) Linear regression (red line) of the set $\{\hat{Q}(k-\tau_2), \hat{r}(k-\tau_6^{(1)})\}$ (blue spots).

Figure 7: Linear regression in the period 01/03/2020-16/07/2020 to estimate the parameters α_1 and δ_1 values.



(a) Linear regression (red line) of the set $\{\hat{Q}(k-\tau_2), \hat{d}(k-\tau_4^{(2)}-\tau_6^{(2)})\}$ (blue spots).



(b) Linear regression (red line) of the set $\{\hat{Q}(k-\tau_2), \hat{r}(k-\tau_6^{(2)})\}$ (blue spots).

Figure 8: Linear regression in the period 17/07/2020-02/02/2021 to estimate the parameters α_2 and δ_2 values. With the aim of obtaining better fitting results, some data has been excluded (green crosses).

where

$$
\begin{aligned}
b_1 &= \lambda(1-p_2)+\gamma+\alpha+\delta+\sigma+3\mu-c_r p_e,\\
b_2 &= (\alpha+\delta+\mu)(\lambda(1-p_2)+\gamma+\mu)+(\sigma+\mu-c_r p_e)(\lambda(1-p_2)+\gamma+\alpha+\delta+2\mu)-\sigma c_r p_s(p_2+(1-p_2)(1-p_1\lambda)),\\
b_3 &= (c_r p_e-\sigma-\mu)(\alpha+\delta+\mu)(\lambda(1-p_2)+\gamma+\mu)+\sigma c_r p_s(p_2+(1-p_2)(1-p_1\lambda))(\alpha+\delta+\mu).
\end{aligned}
\tag{42}
$$

Note that in (3) the variable Q depends on β_s and β_e, and in (42) they have been substitute by $c_r p_s$ and $c_r pe$, respectively, since c_r (the average number of close contacts per day) is a parameter to be

estimated. By applying Euler forward approximation over the expression (41), the following discrete equation is obtained,

$$Q(k) = (3 - b_1)Q(k-1) + (-3 + 2b_1 - b_2)Q(k-2) + (1 - b_1 + b_2 + b_3)Q(k-3). \tag{43}$$

Now, let's suppose that least square method for discrete system $Q(k)$'s parameters estimation is applied for sampling periods in which all parameters maintain approximately constant, so $\hat{\theta} = [\hat{a}_1, \hat{a}_2, \hat{a}_3]^T$ is obtained. Thus, the prediction $\hat{Q}(k)$ will be given by

$$\hat{Q}(k) = -\hat{a}_1 Q(k-1) - \hat{a}_2 Q(k-2) - \hat{a}_3 Q(k-3). \tag{44}$$

When comparing the equations (43) and (44) a nonlinear system is obtained; when expanding the set of equations, in addition to the unknown parameters λ and c_r, nonlinear terms such as $c_r\lambda$ and $c_r p_1 \lambda$ appear. A linear system can be constructed, where the unknown parameters are λ, c_r and λc_r, however another equation must be found out to calculate the value of p_1. Thus, the next system is obtained,

$$\begin{pmatrix} 1-p_2 & -p_e & 0 \\ -\sigma(1-p_2) & p_e\gamma - p_s\sigma & (1-p_2)(pe - p_s p_1 \sigma) \\ (\alpha+\delta+\sigma)(1-p_2) & p_e(\gamma+\alpha+\delta) + p_s\sigma & (1-p_2)(p_s p_1 \sigma - p_e) \end{pmatrix} \begin{pmatrix} \hat{\lambda} \\ \hat{c}_r \\ \hat{\lambda}\hat{c}_r \end{pmatrix} = \begin{pmatrix} \hat{a}_1 + 3 - (\gamma+\alpha+\delta+\sigma+3\mu) \\ -\frac{\hat{a}_3+\hat{a}_2+\hat{a}_3+1}{\alpha+\delta+\mu} + (\sigma+\mu)(\gamma+\mu) \\ 2\hat{a}_1 + \hat{a}_2 + 3 - (\alpha+\delta)\gamma - \sigma(\gamma+\alpha+\delta) \end{pmatrix} \tag{45}$$

when substituting the known parameters (those described in Table 1) into the matrix of the left hand side of (45), it is found that the determinant is of the order 10^{-3} for $\forall p_1 \in [0,1]$, so it has been considered as inconsistent. Since the second and third rows of the system (45) cause the inconsistency, the following cost function has been built up as follows:

$$\begin{aligned} J_1 = &\left[-\sigma(1-p_2)\hat{\lambda} + (p_e\gamma - p_s\sigma)\hat{c}_r + \hat{c}_r\hat{\lambda}(1-p_2)(p_e - p_s p_1\sigma) + \tfrac{\hat{a}_3+\hat{a}_2+\hat{a}_3+1}{\alpha+\delta+\mu} - (\sigma+\mu)(\gamma+\mu)\right]^2 \\ &+ \left[(\alpha+\delta+\sigma)(1-p_2)\hat{\lambda} + \hat{c}_r(p_e(\gamma+\alpha+\delta) + p_s\sigma) + \hat{c}_r\hat{\lambda}(1-p_2)(p_s p_1\sigma - p_e) - (2\hat{a}_1 + \hat{a}_2 + 3) + (\alpha+\delta)\gamma + \sigma(\gamma+\alpha+\delta)\right]^2 \end{aligned} \tag{46}$$

so the solution is given by,

$$\begin{gathered} \lambda(1-p_2) - c_r p_e = \hat{a}_1 + 3 - (\gamma+\alpha+\delta+\sigma+3\mu), \\ \arg\min_{\lambda, c_r} J_1(\lambda, c_r), \\ \forall p_1 \in \mathcal{R}^+. \end{gathered} \tag{47}$$

The system (47) is solved numerically, for a given p_1, combining bisection method[6] and optimization algorithms such as Newton Raphson [6, 4].

In (47) 2 conditions are shown, but however an additional condition is needed since three parameters must be estimated. The previously utilized procedure with Q can be applied with the number of infected population I to gather the last constrain, that is: I is related with the number of positives via the proportionality $\lambda(1-p_2)(1+c_r p_1 p_s)$,

$$p(k) = \lambda(1-p_2)(1+c_r p_1 p_s)I(k) \tag{48}$$

therefore, the condition to determine the value of p_1 can be obtained from the differential equation of I. The ODE expression of I is

$$\ddot{I}(t) = c_1\dot{I}(t) + c_2 I(t) \tag{49}$$

where,

$$\begin{aligned} c_1 &= c_r p_e - \sigma - \lambda(1-p_2) - \gamma - \mu, \\ c_2 &= (\sigma+\mu)\left[c_r p_s(p_2 + (1-p_2)(1-p_1\lambda))\right] + \left[\lambda(1-p_2) + \gamma + \mu\right](c_r p_e + \sigma + \mu) \end{aligned}$$

By applying the Euler Forward approximation method the following discrete equation is obtained,

$$I(k) = (2 + c_1)I(k-1) + (c_2 - c_1 - 1)I(k-2) \tag{50}$$

From the equation (50) it is known that the order of $p(k)$ is two, so let's suppose that from the parameter estimation technique $\hat{\theta} = [\hat{a}_1, \hat{a}_2]$ is calculated, and considering the relation between $I(k)$ and $p(k)$, see expression (48), the conditions

$$\begin{aligned} 2 + c_1 &= -\lambda(1-p2)(1+c_r p_s p_1)\hat{a}_1, \\ c_2 - c_1 - 1 &= -\lambda(1-p2)(1+c_r p_s p_1)\hat{a}_2 \end{aligned} \tag{51}$$

are obtained. As happened previously the system is undetermined, so a new cost function J_2 has been built up

$$J_2 = \left[2 + c_1 + \lambda(1-p2)(1+c_r p_s p_1)\hat{a}_1\right]^2 + \left[c_2 - c_1 - 1 + \lambda(1-p2)(1+c_r p_s p_1)\hat{a}_2\right]^2 \tag{52}$$

and the solution will be given by the condition

$$\arg\min_{p_1} J_2(p_1), \tag{53}$$

which is solved with the Newton Raphson numerical method. In the constructed *Matlab* code used to solve the stated numerical problem, only the parameters values that make $J_{1,2} < 10^{-14}$ have been considered as admissible. See the following repository path to check

CoVid-19-Analysis/code/parameter_estimation/parameter_estimation.m

Before applying the parameter estimation method, the ranges in which the parameters are kept constant must be defined. The time periods for which α and δ are assumed to be constant are described in (40). As far as the parameter λ expresses how fast infected individuals go to the doctor or hospitals due to their symptoms, it will be expected that its value will not vary significantly compared to c_r and p_1:

$$\frac{dc_r}{dt}, \frac{dp_1}{dt} > \frac{d\lambda}{dt}. \tag{54}$$

Regarding c_r and p_1, it has to be considered that, in real life, the performed tests $t(k)$ are a consequence of testing symptomatic people with CoVid-19 ($(1-p_2)\lambda I$), traced close contacts ($c_r\lambda p_1(1-p2)IS/N$) and other people that present symptoms similar to CoVid-19 but are not infected (ϵ). Therefore,

$$t(k) = \left(1 + c_r p_1 \frac{S}{N}\right)\lambda(1-p_2)I + \epsilon. \tag{55}$$

With respect to the total number of positive tests $p(k)$, it is the sum of symptomatic people with CoVid-19 and traced close contacts that have been infected ($c_r p_s \lambda p_1(1-p_2)IS/N$), that is:

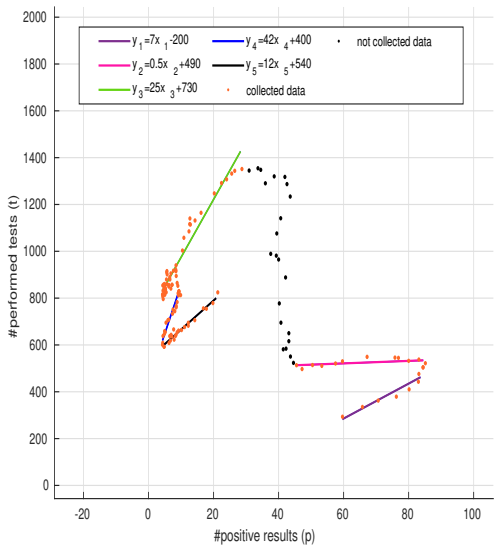
$$p(k) = \left(1 + p_s c_r p_1 \frac{S}{N}\right)\lambda(1-p_2)I, \tag{56}$$

and considering that $S/N \approx 1$, it is possible to rewrite $t(k)$ in terms of $p(k)$,

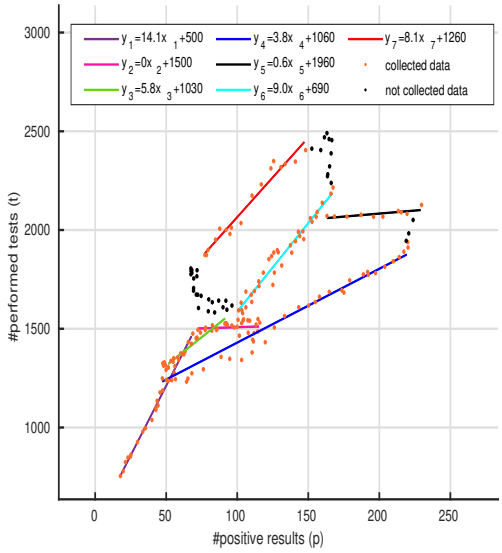
$$t(k) = r_a p(k) + \epsilon \tag{57}$$

where $r_a = (1 + c_r p_1)/(1 + c_r p_1 p_s)$. So, if at certain time period, the relation between $t(k)$ and $p(k)$ is linear, it will be assumed that c_r and p_1 are constant. When plotting $t(k)$ with respect to $p(k)$ it can be seen the existence of subsets $\{t(k)_j, p(k)_j\}$ that fulfill a linear equation. In Figure 9, it is shown $t(k)$ with respect to $p(k)$, plus the identified linear regions. Note that the set of points that could give negative r_a or ϵ values have been rejected, see black points in Figure 9, and they have been interpreted as transition points from one admissible subset to another. Table 4 shows the values of r_a and ϵ with their respective error for each of the identified linear regions.

The time periods shown in Table 4 match quite precisely with the duration of restrictions (e.g in the 13/03/2020 a quarantine was imposed, and the process to recover normality started the 25/05/2020 and finished the 18/06/2020). With the trial and error method the time periods have been adjusted to ensure the best results, that is; first the time periods shown in Table 4 have been used to estimate the values of λ, c_r and p_1 for each interval, then the differential equation (17) has been solved and the results compared with the estimated $\hat{Q}$. If the comparison wasn't suitable the time ranges were



(a) t vs. p (black and orange dots) and the recognized linear patterns (continuous lines) in the time period from the 01/03/2020 to the 16/07/2020.



(b) t vs. p (black and orange dots) and the recognized linear patterns (continuous lines) in the time period from the 17/07/2020/ to the 02/02/2021.

Figure 9: Set of linear regression $\{t(k)_j, p(k)_j\}$ in the periods 01/03/2020-16/07/2020 (left) and 17/07/2020-02/02/2021 (right) to analyse when the parameters c_r and p_1 maintain approximately constant.

dates	r_a	Δr_a	ϵ	$\Delta\epsilon$
[29/02/2020, 07/03/2020]	7	2	−200	200
[08/03/2020, 19/03/2020]	0.5	0.6	490	40
[11/04/2020, 28/05/2020]	25	2	730	20
[29/05/2020, 22/06/2020]	42	7	400	50
[23/06/2020, 12/07/2020]	12	1	540	10
[17/07/2020, 10/08/2020]	14.1	0.4	500	20

dates	r_a	Δr_a	ϵ	$\Delta\epsilon$
[11/08/2020, 05/09/2020]	0	1	1500	100
[06/09/2020, 18/09/2020]	5.8	0.9	1030	60
[19/09/2020, 27/10/2020]	3.8	0.2	1060	30
[31/10/2020, 09/11/2020]	0.6	0.4	1960	90
[10/11/2020,26/11/2020] [10/11/2020, 26/11/2020]	9.0	0.7	690	90
[23/06/2020, 12/07/2020]	8.1	0.9	1260	100

Table 4: Obtained r_a and ϵ values when applying linear regression, and the respective time periods.

slightly modified and the procedure repeated until a satisfactory result was achieved. Thus, the selected intervals are

$$c_r, p_1 = \begin{cases} c_{r(1)}, p_{1(1)} & \text{for } k, \in [29/02/2020, 10/03/2020] \\ c_{r(2)}, p_{1(2)} & \text{for } k \in [11/03/2020, 18/03/2020] \\ c_{r(3)}, p_{1(3)} & \text{for } k \in [19/03/2020, 28/05/2020] \\ c_{r(4)}, p_{1(4)} & \text{for } k \in [28/05/2020, 30/06/2020] \\ c_{r(5)}, p_{1(5)} & \text{for } k \in [17/07/2020, 06/08/2020] \\ c_{r(6)}, p_{1(6)} & \text{for } k \in [10/08/2020, 02/09/2020] \\ c_{r(7)}, p_{1(7)} & \text{for } k \in [04/09/2020, 17/09/2020] \\ c_{r(8)}, p_{1(8)} & \text{for } k \in [17/09/2020, 14/10/2020] \\ c_{r(9)}, p_{1(9)} & \text{for } k \in [15/10/2020, 30/10/2020] \\ c_{r(10)}, p_{1(10)} & \text{for } k \in [30/10/2020, 09/11/2020] \\ c_{r(11)}, p_{1(11)} & \text{for } k \in [09/11/2020, 23/11/2020] \\ c_{r(12)}, p_{1(12)} & \text{for } k \in [25/11/2020, 13/12/2020] \\ c_{r(13)}, p_{1(13)} & \text{for } k \in [19/12/2020, 06/01/2021] \\ c_{r(14)}, p_{1(14)} & \text{for } k \in [07/01/2021, 20/01/2021] \end{cases} \tag{58}$$

With the bisection and gradient descend methods the set of equations given by (47) and (52) have been evaluated for the time periods specified in (58), and the results have been conferred in Table 5. See the next repository path CoVid-19-Analysis/code/parameter_estimation/data_processing/ to see the Matlab scripts that have been used to estimate the parameter's values shown in Table 5.

dates	Natural positivity rate λ ($days^{-1}$)	Average number of close contacts per day c_r $days^{-1}$	Tracing effectiveness p_1
[29/02/2020, 10/03/2020]	$\hat{\lambda}_1 = 0.685$	$\hat{c}_{r(1)} = 12.88$	$\hat{p}_{1(1)} = 0.06$
[11/03/2020, 18/03/2020]	$\hat{\lambda}_2 = 0.643$	$\hat{c}_{r(2)} = 4.88$	$\hat{p}_{1(2)} = 0.9$
[19/03/2020, 28/05/2020]	$\hat{\lambda}_3 = 0.938$	$\hat{c}_{r(3)} = 2.07$	$\hat{p}_{1(3)} = 0.25$
[28/05/2020, 30/06/2020]	$\hat{\lambda}_4 = 0.859$	$\hat{c}_{r(4)} = 4.15$	$\hat{p}_{1(4)} = 0.9$
[17/07/2020, 06/08/2020]	$\hat{\lambda}_5 = 0.930$	$\hat{c}_{r(5)} = 7.55$	$\hat{p}_{1(5)} = 0.06$
[10/08/2020, 02/09/2020]	$\hat{\lambda}_6 = 0.932$	$\hat{c}_{r(6)} = 4.81$	$\hat{p}_{1(6)} = 0.02$
[04/09/2020, 17/09/2020]	$\hat{\lambda}_7 = 0.737$	$\hat{c}_{r(7)} = 4.63$	$\hat{p}_{1(7)} = 0.42$
[17/09/2020, 14/10/2020]	$\hat{\lambda}_8 = 0.871$	$\hat{c}_{r(8)} = 3.21$	$\hat{p}_{1(8)} = 0.28$
[15/10/2020, 30/10/2020]	$\hat{\lambda}_9 = 0.932$	$\hat{c}_{r(9)} = 6.10$	$\hat{p}_{1(9)} = 0.14$
[30/10/2020, 09/11/2020]	$\hat{\lambda}_{10} = 0.933$	$\hat{c}_{r(10)} = 5.57$	$\hat{p}_{1(10)} = 0.25$
[09/11/2020, 23/11/2020]	$\hat{\lambda}_{11} = 0.888$	$\hat{c}_{r(11)} = 4.68$	$\hat{p}_{1(11)} = 0.36$
[25/11/2020, 13/12/2020]	$\hat{\lambda}_{12} = 0.744$	$\hat{c}_{r(12)} = 3.14$	$\hat{p}_{1(12)} = 1$
[19/12/2020, 06/01/2021]	$\hat{\lambda}_{13} = 0.628$	$\hat{c}_{r(13)} = 6.38$	$\hat{p}_{1(13)} = 0.78$
[07/01/2021, 20/01/2021]	$\hat{\lambda}_{14} = 0.881$	$\hat{c}_{r(14)} = 5.06$	$\hat{p}_{1(14)} = 0.05$

Table 5: Estimated parameters values.

The lowest λ value it is found in the period 19/12/2020-06/01/2021 which matches with winter holidays; people moved to their relative houses, so people with symptoms might have found more complicate go to the outpatient. Regarding the average number of close contacts per day c_r, when tight restrictions have been applied lower values of c_r are observed; the lowest value ($\hat{c}_{r(3)} = 2.07$) it is found from the 19/03/2020 to the 28/05/2020 (approximately when quarantine was imposed), and the next lowest one ($\hat{c}_{r(12)} = 3.14$) from the 25/11/2020 to the 13/12/2020, when the state of alertness was already valid.

5.5 Validation

The estimated values of Q and the obtained parameters hold uncertainties, therefore in the validation process how the uncertainties affect the solution of the differential equation (41) will be analysed and compared with the real data measurements. It is possible to obtain the solution of the expression (41) if three conditions related with the initial state are given (e.g. $Q(t_0)$, $\dot{Q}(t_0)$ and $\ddot{Q}(t_0)$). Considering that the conditions hold a standard deviation, it is possible to calculate the solution's standard deviation.

The solution of equation (41) is given by

$$Q(t) = Ae^{r_1 t} + Be^{r_2 t} + Ce^{r_3 t} \tag{59}$$

where r_1, r_2 and r_3 are the roots of the following equation,

$$r^3 + b_1 r^2 + b_2 r = b_3.$$

With the initial conditions $Q(t_0)$, $\dot{Q}(t_0)$ and $\ddot{Q}(t_0)$ the values of A, B and C are obtained. The expression (59) must accomplish the next system:

$$\begin{pmatrix} e^{r_1 t_0} & e^{r_2 t_0} & e^{r_3 t_0} \\ r_1 e^{r_1 t_0} & r_2 e^{r_2 t_0} & r_3 e^{r_3 t_0} \\ r_1^2 e^{r_1 t_0} & r_2^2 e^{r_2 t_0} & r_3^2 e^{r_3 t_0} \end{pmatrix} \begin{pmatrix} A \\ B \\ C \end{pmatrix} = \begin{pmatrix} Q(t_0) \\ \dot{Q}(t_0) \\ \ddot{Q}(t_0) \end{pmatrix} \tag{60}$$

which gives the following solution,

$$\begin{aligned} A &= \frac{1}{\Delta e^{r_1 t_0}} \left[Q(t_0)(r_2 r_3^2 - r_3 r_2^2) + \dot{Q}(t_0)(r_2^2 - r_3^2) + \ddot{Q}(t_0)(r_3 - r_2) \right] \\ B &= \frac{1}{\Delta e^{r_2 t_0}} \left[Q(t_0)(r_3 r_1^2 - r_1 r_3^2) + \dot{Q}(t_0)(r_3^2 - r_1^2) + \ddot{Q}(t_0)(r_1 - r_3) \right] \\ C &= \frac{1}{\Delta e^{r_3 t_0}} \left[Q(t_0)(r_1 r_2^2 - r_2 r_1^2) + \dot{Q}(t_0)(r_1^2 - r_2^2) + \ddot{Q}(t_0)(r_2 - r_1) \right] \end{aligned} \tag{61}$$

where $\Delta = r_2 r_3^2 + r_3 r_1^2 + r_1 r_2^2 - r_2 r_1^2 - r_1 r_3^2 - r_3 r_2^2$.

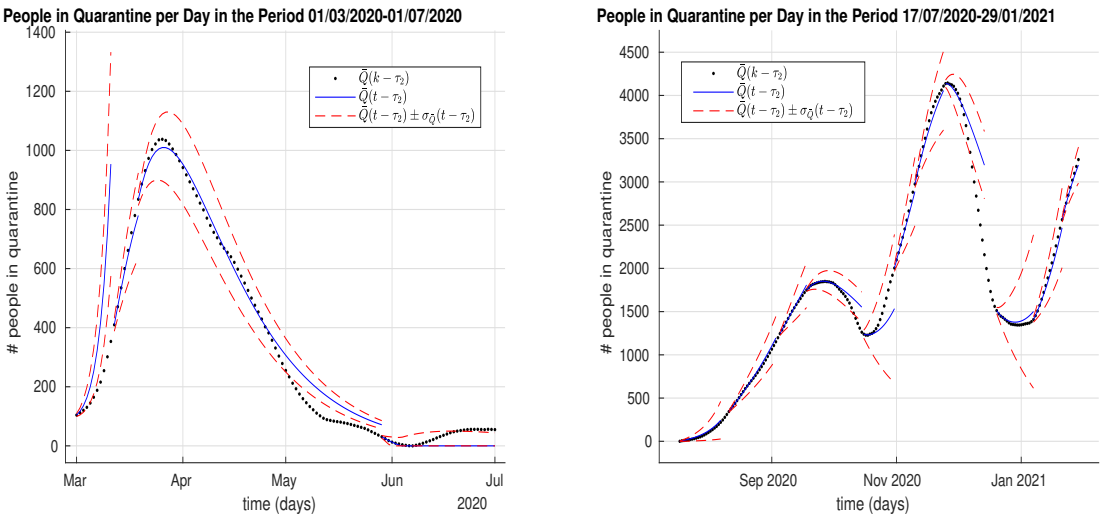
In [2] a set of variables $\{x_1, x_2, \ldots, x_n\}$, their means $\{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n\}$, variances $\{\sigma_{x_1}, \sigma_{x_2}, \ldots, \sigma_{x_n}\}$, and the variances of the mean values (commonly known as the standard error) $\{\sigma_{\bar{x}_1}, \sigma_{\bar{x}_n}, \ldots, \sigma_{\bar{x}_n}\}$ are considered, and it is proved that given a function $f(x_1, x_2, \ldots, x_n)$ its $\bar{f}$, σ_f and $\sigma_{\bar{f}}$ can be calculated with the following set of equations,

$$\begin{aligned} \bar{f} &= f(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n) \\ \sigma_f &= \sqrt{\left(\frac{\partial f}{\partial x_1}\right) \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right) \sigma_{x_2}^2 + \ldots + \left(\frac{\partial f}{\partial x_n}\right) \sigma_{x_n}^2} \\ \sigma_{\bar{f}} &= \sqrt{\frac{1}{n}\left[\left(\frac{\partial f}{\partial x_1}\right) \sigma_{\bar{x}_1}^2 + \left(\frac{\partial f}{\partial x_2}\right) \sigma_{\bar{x}_2}^2 + \ldots + \left(\frac{\partial f}{\partial x_n}\right) \sigma_{\bar{x}_n}^2\right]} \end{aligned} \tag{62}$$

which, combining with the solution (59), it makes possible to calculate the mean $\bar{Q}(t)$ and its variance $\sigma_{\bar{Q}}$.

First $d(k)$, $r(k)$ and $p(k)$ are gathered and filtered with the move mean technique (it takes 14 samples to compute the mean values), so $\bar{d}(k) = \{\bar{d}(k)\}_{k=1}^n$, $\bar{r}(k) = \{\bar{r}(k)\}_{k=1}^n$ and $\bar{p}(k) = \{\bar{p}(k)\}_{k=1}^n$ are obtained and the variances of the mean values $\sigma_{\bar{d}}$, $\sigma_{\bar{r}}$ and $\sigma_{\bar{p}}$ computed. Note that if the context is clear, it will be used $\bar{d}$, $\bar{r}$ and $\bar{p}$ instead of $\bar{d}(k)$, $\bar{r}(k)$ and $\bar{p}(k)$. Then the delays shown in (21) and (22) are applied to $\{\bar{d}, \sigma_{\bar{d}}\}$ $\{\bar{r}, \sigma_{\bar{r}}\}$ and $\{\bar{p}, \sigma_{\bar{p}}\}$, so $\{\bar{Q}, \sigma_{\bar{Q}}\}$ is determined considering the expressions (19) and (62). Note that $\bar{Q}$ is equal to the previously computed $\hat{Q}$ and its results are depicted in Figure 5. With $\{\bar{Q}, \sigma_{\bar{Q}}\}$ and the Euler forward approximation, the values $\{\dot{\bar{Q}}, \sigma_{\dot{\bar{Q}}}\}$ and $\{\ddot{\bar{Q}}, \sigma_{\ddot{\bar{Q}}}\}$ are computed. The determined $\bar{Q}$, $\dot{\bar{Q}}$ and $\ddot{\bar{Q}}$ will be used as the initial conditions appearing in (60) to compute $\bar{Q}(t)$. Finally, the respective standard error combined with third equation in (62) will give $\sigma_{\bar{Q}(t)}$.

In Figure 10 it is depicted $\bar{Q}(k)$, $\bar{Q}(t)$, and $\bar{Q}(t) \pm \sigma_{\bar{Q}(t)}$ in the case of the first wave (a) and following waves (b), and it can be seen that in most cases $\bar{Q}(k)$ fits inside $\bar{Q}(t) \pm \sigma_{\bar{Q}(t)}$. In the first wave the computed dynamics correspond to the real one, however in the first stage (from the 29/02/2020 to the 10/03/2020) the simulation gives a higher increase, and it cast doubts on the estimated parameters $\{\hat{\lambda}_1, \hat{c}_{r(1)}, \hat{p}_1\}$. In the following waves there are two cases in which the estimated parameters $\{\hat{\lambda}_8, \hat{k}_8, \hat{p}_8\}$ and $\{\hat{\lambda}_{12}, \hat{c}_{r(12)}, \hat{p}_{12}\}$ give a solution that deviates from the real data.



(a) Period from the 01/03/2020 to the 01/07/2020

(b) Period from the 17/07/2020 to the 29/01/2021

Figure 10: Comparison between measured people in quarantine per day (black dots) with respect to the simulated people in quarantine per day (blue solid line) with the estimated parameters. The dashed red lines show the standard error of the simulation.

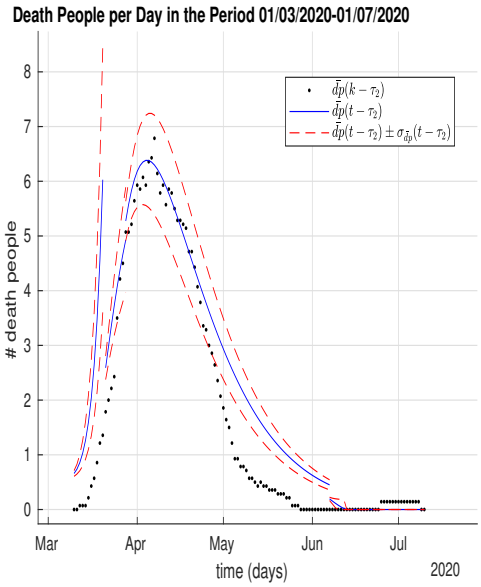
Proceeding in the same manner with $\bar{Q}(t)$, and with the previously calculated α and δ values, see Table 3, it is possible to construct $\bar{d}(t)$ and $\bar{r}(t)$ and their respective standard errors, see Figures 11 and 12. In the first wave, as it happened with the comparison between $\bar{Q}(k)$ and $\bar{Q}(t)$ from the 29/02/2020 to the 10/03/2020, there is a higher growth of $\bar{p}(t)$ and $\bar{r}(t)$ compared to $\bar{p}(k)$ and $\bar{r}(k)$, respectively. In the subsequent waves the simulation gives a good view of the dynamics but fails to predict the exact values, and this could be caused by a bad estimation of α_2 and δ_2: the residuals of the linear regression were less than 0.85.

See the next repository path CoVid-19-Analysis/code/validation/ to see the Matlab scripts used for the validation process.

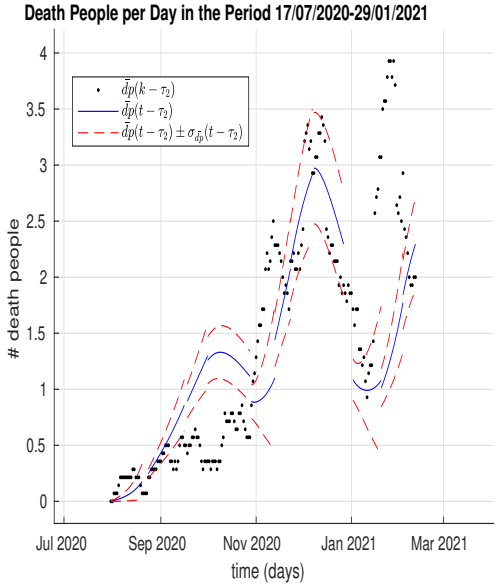
5.6 Discussion

Each of the estimated parameters has a physical meaning, thus they could be used as control inputs. The natural positivity rate λ oscillates approximately from $0.63\,days^{-1}$ to $0.93\,days^{-1}$ which inverse correspond to $1.58\,days$ and $1.07\,days$, respectively. Assuming that at first COVID-19 was underestimated and its symptoms were mixed up with cold/flu ones, infected people with symptoms might have visited the doctor at lower rates. On the other hand, holiday seasons could also be associated to lower λ values since people might be outside their towns and also the health assistance might be reduced. Tracing effectiveness p_1 was never before implemented in Spain on big scale cases such as this pandemic, therefore there has been notably uncertainties regarding how to implement it (e.g on circumstances where the number of positive result were high, people was tested randomly). No relation between the number of positive tests with respect p_1 has been observed, and as far as there is no more information about how tracing techniques have been implemented it is not possible to infer what changes p_1 values. The average number of close contact per day c_r is related with the restrictions on the social interaction limit, mobility, etc. It is notorious that the lower value of c_r was in the first quarantine period, when only on necessary circumstances was allowed going outside. Considering λ, p_1, c_r, and their respective meaning, by now only c_r could be used as a control input. To do it so, the imposed restrictions must be analysed and compared with the estimated c_r, and then seek for a relation between them.

The Autonomous Community of Cantabria publishes in *Boletin Oficial de Cantabria* (BOC) the upcoming decrees and a number of them contain laws to avoid COVID-19 spreading. It must be pointed out that during all pandemic restrictions regarding different matters have been applied, so 5 main top-

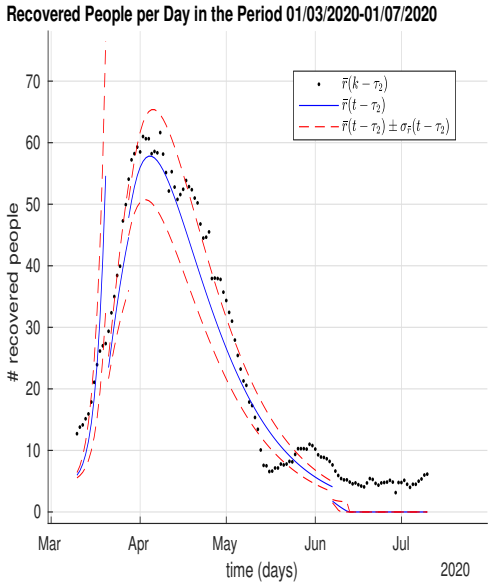


(a) Period from the 01/03/2020 to the 01/07/2020

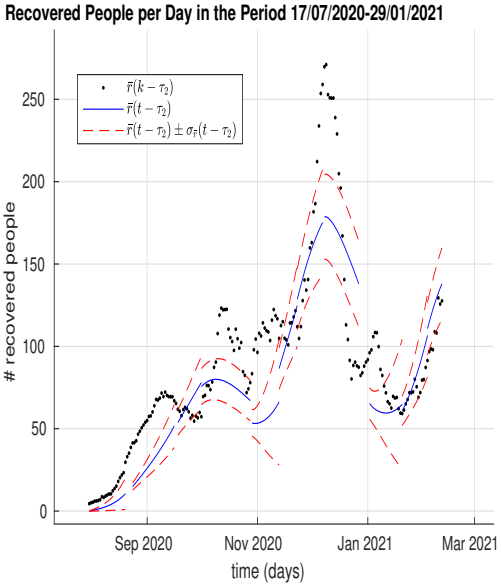


(b) Period from the 17/07/2020 to the 29/01/2021

Figure 11: Comparison between measured death people per day (black dots) with respect to the simulation (blue solid line) carried out with the estimated parameters. The dashed red lines show the standard error of the simulation.



(a) Period from the 01/03/2020 to the 01/07/2020



(b) Period from the 17/07/2020 to the 29/01/2021

Figure 12: Comparison between recovered people per day (black dots) with respect to the simulation (blue solid line) carried out with the estimated parameters. The dashed red lines show the standard error of the solution.

ics (social interaction limit, shops, hopitality, curfew and mobility) have been selected to be analysed through all BOCs, and Table 6 shows a summarize of the decrees related to those topics with respect to time. The shadowed rows in Table 6 are special cases: the restrictions in blue were applied to specific regions of Cantabria, and the red ones hold exceptions. The red shadowed case describes the main restrictions that were imposed during winter holidays, however on specific days more permissions were given (e.g from the 23/12/2020 to the 06/01/2021 it was allowed moving to relative's houses whether it was inside or outside Cantabria). Note that during the strict quarantine and de-escalation periods (from the 25/05/2020 to the 18/06/2020) in the respective BOCs it was recommended only going outside when it was strictly necessary, however government advices and people's fear arrange people at home.

After studying the estimated c_r in Table 5 and the restriction given in Table 6 some correlations with respect to mobility can be observed; when there is mobility freedom and there are partial restrictions in other topics, c_r is expected to be around 6 and 7. On the other hand, when the mobility is reduced to Cantabria region or towns, c_r switches approximately to 5 and 4, respectively.

date	social interaction limit	shops	hospitality	curfew	Mobility	BOC number
13/03/2020	-	closed	closed	-	-	2020-2407
25/05/2020	-	-	50%	-	-	2020-3234
7/06/2020	-	50% inside 75% terrace	50%	-	-	2020-3674
18/06/2020	25	75%	75% inside 100% terrace	-	-	2020-4080
24/07/2020	15	75%	75% inside, 100%terrace, close at 2:00 am	-	-	2020-5390
15/08/2020	10	75%	75% inside, 100%terrace, close at 1:00 am	-	-	2020-6061
02/09/2020 15/09/2020	10	75% in Santoña	closed in Santoña	-	Santoña	2020-6015 2020-6835
11/09/2020 25/09/2020	10	closed in Torrelavega	closed in Torrelavega	-	Torrelavega	2020-6782
23/10/2020	10	50%	50% inside, 75%terrace, close at 23:00 am	-	-	2020-7970
26/10/2020	6	50%	50% inside, 75%terrace, close at 23:00 am	24:00-6:00	-	2020-7994 2020-7995
29/10/2020	6	50%	50% inside, 75%terrace, close at 23:00 am	24:00-6:00	Cantabria	2020-8146 2020-8973
04/11/2020	6	50%	50% inside, 75%terrace, close at 23:00 am	24:00-6:00	municipality	2020-8311 2020-8663 2020-9117
06/11/2020	6	1/3	closed	24:00-6:00	municipality	2020-8382 2020-8778 2020-9280
13/11/2020	6	1/3	closed	22:00-6:00	municipality	2020-8580
17/12/2020	6	1/3	closed inside, 75%terrace	22:00-6:00	Cantabria	2020-9651 2020-9662 2020-10035
13/01/2021	6	1/3	closed	22:00-6:00	Cantabria	2020-9651 2020-9662

Table 6: Summarize of restrictions on 5 main topics (social interaction limit, shops, hospitality, curfew, mobility) with respect to time. The label *date* expresses the respective BOC's publishing date, and exceptionally in the blue shadowed case the second date expresses when the limitations ceased. Regarding the *BOC number* section, the first BOC number collects the implemented restrictions and the following ones their extension with respect to time.

6 CONCLUSION

In the constructed model there is a good trade-off between accuracy, transparency and flexibility. Even if Q gathers people that tested positive and were staying at home and hospitals, reducing so the model's complexity, with the estimated parameters a quantitative and qualitative fit with respect to real data has been obtained. Regarding transparency, the calculated basic reproduction number $\mathcal{R}_0$ shows a simple relation with respect to parameters such as the average number of close contacts per day c_r or the tracing effectiveness p_1. Moreover, assuming that $S/N \approx 1$ the system's complexity lessen and the set of differential equations turns to a set of ordinary differential equations, which solution is straight forward and easy to interpret its dynamics. On the other hand, the model is flexible because it has been built up based on well understood transmission principles. For instance, a delay between the estimated output $\alpha\hat{Q}(k)$ and the measured $d(k)$ was found out, so the model was simply adapted to fit with real measurements by adding a dummy variable D, see expression (30).

The stability threshold for the free disease equilibrium state has been found with two different methods: calculating the Jacobian eigenvalues and analysing their real part, and via the basic reproduction number $\mathcal{R}_0$. It is more difficult to apply the first method as the system's dimension increases, so in such cases it is suggested to calculate $\mathcal{R}_0$. In the cases of the endemic equilibrium, the conditions for its stability have been calculated with the Jacobian matrix and the Routh-Hurwitz criteria. However, as a consequence of the system's high dimension, the analytical condition do not give a clear result.

Before proceeding with the parameter estimation different data sources regarding the epidemiological situation of Spain have been analysed, and it has been found out that the AC of Cantabria offers a file with the respective data including measurements like the number of recovered people $r(k)$ (essential magnitude to estimate the recovery rate δ). Since generally all data sources instead of giving $r(k)$ deliver the number of recovered people from hospitals, it was contemplated spanning the model with another block Ph which will hold hospitalized people. Besides increasing complexity, the recovery rate of infected people not in hospitals was still not possible to infer from data. Therefore, it was opted to continue with the primary model, a less accurate and complex model but parametrizable from real data from Cantabria.

From the data source the number of positive test result per day $p(k)$, the number of death people due to the disease $d(k)$ and recovered people $r(k)$ was gathered. Before estimating the number of people in quarantine $\hat{Q}(k)$, the measurements were previously filtered with the move mean technique and the existence of possible delays was analysed. A noteworthy delay (not expected based on the model) between $\hat{Q}(k)$ with respect $d(k)$ and $r(k)$ was appreciated. The block Q holds two distinguishable population: infected people with symptoms that have tested positive and stay at home (Q'), and hospitalized ones (Ph). Assuming that Ph is a subsequent stage of Q', the Laplace transforms $Q(s)$ and $Ph(s)$ confirm the presence of a delay between the model's predicted deaths $\alpha Q(s)$ and deaths from hospitals $\alpha'' Ph(s)$. It is also concluded that the gain between $Q(s)$ and $Ph(s)$ depends on $Q(\omega)$ power spectrum, so they will hold a linear association in cases in which the distribution of power into frequencies of the signal Q it is mainly in one frequency. This conclusion is reinforced by the relation between $\hat{Q}(k)$ and $d(k)$ on the "first wave" and "following waves": the first case shows almost a perfect linear relation whereas in the second case it worsen.

It is assumed that the variables Q and I of the constructed model can be expressed with an ARX structure. The time ranges in which the number of close contacts c_r, tracing effectiveness p_1 and natural positivity rate λ (parameters of which the ARX parameters depend) maintain constant have been determined, and the parameters of the ARX have been estimated for each case with $\hat{Q}$ and $\hat{p}(k)$ (the last one is proportional to I). With the parameters of the discrete differential expression of Q and I and the ARX models ones, a set of nonlinear equations have been obtained and numerically solved. The validation process has showed that mostly all estimated parameters give an accurate output of Q, and the predicted number of deaths and recovered people display approximately same dynamic as $d(k)$ and $r(k)$, respectively.

The analysis of the imposed restriction with respect to time only gives an overview of how the limitations might affect the value of c_r. However if an optimal control of the disease is wanted an accurate relation between c_r and restriction is needed. With the aim of getting a reliable input-output response, a black box model such as a Neural Network (NN) could be a good option to deal with this type of problem. The input variables of the NN could be the social interaction limit l_{so}, limitation

in shops l_{sh}, in hospitality l_{ho}, curfew l_{cu} and limitation on mobility l_{mo}, and the output will be c_r. To determine the set of admissible values of each input, the restrictions given in Table 6 could be quantified with simple rules (e.g 0 if the limitation is high, and 1 if it is low). Finally, with the set of inputs and outputs, and the Matlab's Deep Learing Toolbox a NN can be obtained.

For the optimal control input a close loop feedback control system is suggested: from the real system it is possible to estimate the current number of people in quarantine $\hat{Q}(k)$, and it could be used as the constructed model initial condition to predict the future number of people in quarantine $\hat{Q}(k+h)$, where h establishes the prediction horizon. The obtained $\hat{Q}(k+h)$ could be subtracted from a reference signal Q_0, providing an error ϵ. The controller will take ϵ as an input, and it will calculate the number of contact per day c_r that people should maintain to achieve Q_0. Then, different combinations of the NN input variables could be tested until its output is approximately equal to the controller's output, and the estimated NN input values can be transformed into the restrictions that must be imposed.

A APPROXIMATED ENDEMIC EQUILIBRIUM POINTS ($\rho^{-1} \approx 0$)

For the analysis of the *Endemic Equilibrium* point it was assumed that the immunity period (ρ^{-1}) was greater compared to other periods used to describe the system, such as the incubation time (σ^{-1}) or recovery time (γ^{-1}). Therefore the system (7) is rewritten as

$$\begin{aligned}
0 &= \Lambda - [\beta_e E^*_{ee} + (p_2\beta_a + (1-p_2)\beta_s) I^*_{ee}] \tfrac{S^*_{ee}}{N^*_{ee}} - \mu S^*_{ee} \\
0 &= (\beta_e E^*_{ee} + (p_2\beta_a + (1-p_2)(1-p_1\lambda)\beta_s) I^*_{ee}) \tfrac{S^*_{ee}}{N^*_{ee}} - (\sigma+\mu) E^*_{ee} \\
0 &= \left(1 + \beta_s p_1 \tfrac{S^*_{ee}}{N^*_{ee}}\right) \lambda(1-p_2) I^*_{ee} - (\mu+\alpha+\delta) Q^*_{ee} \\
0 &= \sigma E^*_{ee} - (\lambda(1-p_2) + \mu + \gamma) I^*_{ee} \\
0 &= \gamma I^*_{ee} + \delta Q^*_{ee} - \mu R^*_{ee}
\end{aligned} \tag{63}$$

which gives the following solutions,

$$\begin{aligned}
S^*_{ee} &= \tfrac{\Lambda}{\mu} - \tfrac{\sigma+\mu}{\mu}\left(\tfrac{\beta_1\beta_e+\sigma\beta_3}{\beta_1\beta_e+\sigma\beta_4}\right) E^*_{ee}, \\
Q^*_{ee} &= \tfrac{1}{(\mu+\alpha+\delta)}\left[(\mu+\sigma)\left(\tfrac{\beta_1\beta_e+\sigma\beta_3}{\beta_1\beta_e+\sigma\beta_4} - 1\right) + \tfrac{\sigma}{\beta_1}(1-p_2)\lambda\right] E^*_{ee}, \\
I^*_{ee} &= \tfrac{\sigma}{\beta_1} E^*_{ee}, \\
R^*_{ee} &= \tfrac{1}{\mu}\left[\tfrac{\delta(\mu+\sigma)}{\mu+\alpha+\delta}\left(\tfrac{\beta_1\beta_e+\sigma\beta_3}{\beta_1\beta_e+\sigma\beta_4} - 1\right) + \tfrac{\sigma}{\beta_1}\left(\tfrac{\delta(1-p_1)\sigma}{\mu+\sigma+\delta} + \delta\right)\right] E^*_{ee}, \\
N^*_{ee} &= \tfrac{1}{\mu\beta_1}\left[\tfrac{\beta_e\beta_1+\sigma\beta_4}{(\sigma+\mu)}\Lambda - (\beta_1\beta_e + \sigma\beta_3) E^*_{ee}\right]
\end{aligned} \tag{64}$$

where,

$$\beta_1 = \lambda(1-p_2) + \mu + \gamma, \quad \beta_2 = \tfrac{1}{(\mu+\alpha+\delta)}\left[\tfrac{\lambda(1-p_2)}{\beta_1} + \tfrac{(\sigma+\mu)\beta_s(1-p_2)p_1\lambda}{\beta_1\beta_e+\sigma[\beta_a p_2+(\beta_s(1-p_2)(1-p_1\lambda))]}\right],$$
$$\beta_3 = p_2\beta_a + (1-p_2)\beta_s \quad \text{and,} \quad \beta_4 = p_2\beta_a + (1-p_2)(1-p_1\lambda)\beta_s.$$

Note that β_4 can be rewritten in terms of β_3,

$$\beta_4 = \beta_3 - \beta_5$$

where $\beta_5 = (1-p_2)p_1\lambda\beta_s$.

The expressions of a_0 and a_2, which are defined in the subsection 3.2.2, can be handled so a fancier expression is obtained, and with the new calculated *Endemic Equilibrium* points, see (64), it can be proved that a_0 and a_2 will remain negative. If the context is clear, the *Endemic Equilibrium* points will be defined as $(S^*, E^*, Q^* I^*, R^*)$ instead of $(S^*_{ee}, E^*_{ee}, Q^*_{ee} I^*_{ee}, R^*_{ee})$ to simplify the notation. Thus,

$$\begin{aligned}
a_0 &= \left(\beta_e\tfrac{E^*}{N^*} + \beta_3\tfrac{I^*}{N^*} + \mu\right)\left[\left(\beta_e\tfrac{S^*}{N^*} - \sigma - \mu\right)\beta_1 + \sigma\beta_4\tfrac{S^*}{N^*}\right] - \left(\beta_e\tfrac{E^*}{N^*} + \beta_4\tfrac{I^*}{N^*}\right)[\beta_e\beta_1 + \sigma\beta_3]\tfrac{S^*}{N^*} \\
&= \left(\beta_e\tfrac{E^*}{N^*} + \beta_3\tfrac{I^*}{N^*} + \mu\right)\left[(\beta_e\beta_1 + \sigma\beta_3)\tfrac{S^*}{N^*} - \sigma\beta_5\tfrac{S^*}{N^*} - (\sigma+\mu)\beta_1\right] - \left(\beta_e\tfrac{E^*}{N^*} + \beta_3\tfrac{I^*}{N^*} - \beta_5\tfrac{I^*}{N^*}\right)(\beta_e\beta_1 + \sigma\beta_3)\tfrac{S^*}{N^*} \\
&= (\beta_e\beta_1 + \sigma\beta_3)\left(\mu + \beta_5\tfrac{I^*}{N^*}\right)\tfrac{S^*}{N^*} - \left(\beta_e\tfrac{E^*}{N^*} + \beta_3\tfrac{I^*}{N^*} + \mu\right)\left[\sigma\beta_5\tfrac{S^*}{N^*} + (\sigma+\mu)\beta_1\right] \\
&= (\beta_e\beta_1 + \sigma\beta_3)\left(\mu + \tfrac{\sigma\beta_5}{\beta_1}\tfrac{E^*}{N^*}\right)\tfrac{S^*}{N^*} - \left[(\beta_e\beta_1 + \sigma\beta_3)\tfrac{1}{\beta_1}\tfrac{E^*}{N^*} + \mu\right]\left[\sigma\beta_5\tfrac{S^*}{N^*} + (\sigma+\mu)\beta_1\right] \\
&= (\beta_e\beta_1 + \sigma\beta_3)\left[\mu\tfrac{S^*}{N^*} - (\sigma+\mu)\tfrac{E^*}{N^*}\right] - \mu\left[\sigma\beta_5\tfrac{S^*}{N^*} + (\sigma+\mu)\beta_1\right] \\
&= \mu(\beta_e\beta_1 + \sigma\beta_4)\tfrac{S^*}{N^*} - (\sigma+\mu)\left[(\beta_e\beta_1 + \sigma\beta_3)\tfrac{E^*}{N^*} + \mu\beta_1\right]
\end{aligned}$$

By substituting $S^*/N^* = \beta_1(\sigma+\mu)/(\beta_e\beta_1 + \sigma\beta_4)$ in a_0, it is proved that

$$a_0 = -(\sigma+\mu)(\beta_e\beta_1 + \sigma\beta_3)\frac{E^*}{N^*} \leqslant 0.$$

Regarding a_2,

$$\begin{aligned}
a_2 &= \beta_e\left(\tfrac{S^*}{N^*} - \tfrac{E^*}{N^*}\right) - \beta_3\tfrac{I^*}{N^*} - \sigma - \mu - \beta_1 = -(\beta_e\beta_1 + \sigma\beta_3)\tfrac{1}{\beta_1}\tfrac{E}{N} + \tfrac{\beta_e\beta_1(\sigma+\mu)}{\beta_e\beta_1+\sigma\beta_4} - (\sigma+\mu+\beta_1) \\
&= -(\beta_e\beta_1 + \sigma\beta_3)\tfrac{1}{\beta_1}\tfrac{E}{N} - \tfrac{[\beta_1^2+\sigma\beta_4(\sigma+\mu+\beta_1)]}{\beta_e\beta_1+\sigma\beta_4} \leqslant 0
\end{aligned}$$

With respect to a_1, its expression can be reduced to

$$\begin{aligned}
a_1 &= \left(\beta_e \tfrac{E^*}{N^*} + \beta_3 \tfrac{I^*}{N^*} + \mu\right)\left(\beta_e \tfrac{S^*}{N^*} - \sigma - \mu - \beta_1\right) + \beta_1\left(\beta_e \tfrac{S^*}{N^*} - \sigma - \mu\right) - \left(\beta_e \tfrac{E^*}{N^*} + \beta_4 \tfrac{I^*}{N^*}\right)\beta_e \tfrac{S^*}{N^*} + \sigma\beta_4 \tfrac{S^*}{N^*} \\
&= \beta_e \tfrac{S^*}{N^*}\left(\beta_5 \tfrac{I^*}{N^*} + \mu + \beta_1\right) - \left(\beta_e \tfrac{E^*}{N^*} + \beta_3 \tfrac{I^*}{N^*} + \mu\right)(\sigma + \mu + \beta_1) - \beta_1(\sigma + \mu) + \sigma\beta_4 \tfrac{S^*}{N^*} \\
&= \beta_e \tfrac{S^*}{N^*}\left(\tfrac{\beta_5\sigma}{\beta_1} \tfrac{E^*}{N^*} + \mu + \beta_1\right) - \left[(\beta_e\beta_1 + \sigma\beta_3)\tfrac{1}{\beta_1}\tfrac{E^*}{N^*} + \mu\right](\sigma + \mu + \beta_1) - \beta_1(\sigma + \mu) + \sigma\beta_4 \tfrac{S^*}{N^*} \\
&= \beta_e \tfrac{S^*}{N^*}\left(\tfrac{\beta_5\sigma}{\beta_1} \tfrac{E^*}{N^*} + \mu\right) - \left[(\beta_e\beta_1 + \sigma\beta_3)\tfrac{1}{\beta_1}\tfrac{E^*}{N^*} + \mu\right](\sigma + \mu + \beta_1).
\end{aligned}$$

B SINGULAR AND NON-SINGULAR M MATRIX

In cases in which a system is defined by a set of differential equations, it is often seen that its matrix representation has the following form,

$$A = \begin{bmatrix} a_{11} & -a_{12} & \cdots \\ -a_{21} & a_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \quad : \; a_{ij} \geqslant 0, \; \forall i \neq j \in m$$

in these cases the matrix has Z sign pattern ($A \in Z^{m \times m}$).

Definition B.1. [29, 3, 33] If $A \in Z^{m \times m}$ and it can be represented as $A = sI - P$, where $P \geqslant 0$ and $s \geqslant \rho(P)$, then A is called M matrix ($\rho(\cdot)$ represents the spectral radius, that is, its eigenvalue maximum modulus).

Depending on s two different M matrix can be distinguish; *non-singular* and *singular*.

Theorem B.1. *[3, 33] A matrix A is a non-singular M matrix if s is greater than P's spectral radius ($s > \rho(P)$). The implications are the following ones*[4]*,*

- *The real part of A's eigenvalues is positive ($\mathrm{Re}(\lambda(A)) > 0$).*
- *The principal minors of A are positive.*
- *The inverse of A (A^{-1}) exists, and it is positive ($A^{-1} \geqslant 0$).*
- *Exist a vector $X > 0$ such that $AX > 0$.*
- *A is monotone on V_A,*
$$Ax > 0 \Rightarrow x > 0 \quad \textit{and},$$
$$Ax = 0 \Rightarrow x = 0 \quad \forall x \in V_A$$
- *A has a convergent regular splitting,*
$$A = M - N, \quad M^{-1} \geqslant 0, \quad N \geqslant 0$$
where MN^{-1} is convergent ($\rho(M^{-1}N) < 1$).

Theorem B.2. *[3, 33] A matrix A is a singular M matrix if s is equal to P's spectral radius ($s = \rho(P)$). The implications are the following ones,*

- *Exists an eigenvalue of A which real part is equal to zero.*
- *There exists a vector $X > 0$ such that $AX = 0$.*
- *all principal minors are nonnegative.*
- *A is monotone in V_A,*
$$Ax \geqslant 0 \Rightarrow x \geqslant 0 \quad \forall x \in V_A.$$

4 when it is about non-singular M matrices many implication can bee seen (e.g in [3] 50 implications are defined), however in this case only the implications that will be used in this work are itemized.

Theorem B.3. *[3] A M matrix A has a regular splitting,*

$$A = M - N, \quad M^{-1} \geqslant 0, \quad N \geqslant 0,$$

where $V_{M^{-1}N} = V_A$ *and* $\rho(M^{-1}N) \leqslant 1$.

Corollary B.1. *Considering that a M matrix A has a regular splitting such that* $A = M - N$, *see Theorem B.3, then* $M^{-1}A$ *is nonsingular if and only if A is nonsingular, and* $M^{-1}A$ *is singular if and only if A is singular.*

Proof: singular M matrix $A \implies M^{-1}A$ singular M matrix.

- There exists $X_1 > 0$ such that $AX_1 = 0$. Then, $(M^{-1}A)X_1 = M^{-1}0 = 0$
- Considering $M^{-1}A$ as nonsingular M matrix, then $(M^{-1}A)X_1 > 0$ (against the first implication). Considering $M^{-1}A$ as a singular M matrix, there exist $X_2 > 0$ such that $(M^{-1}A)X_2 = 0$. Since $V_{M^{-1}N} = V_A$, then $X_2 = X_1$.

The same procedure can be followed to prove the backward implication (singular M matrix $A \Leftarrow M^{-1}A$ singular M matrix), and also in the case of nonsingular M matrices.

Corollary B.2. *If A is a singular M matrix, it has a regular splitting* $A = M - N$, *see Theorem B.3, and* $\rho(M^{-1}N) = 1$.

Proof: if A is a singular M matrix $\iff M^{-1}A$ is a singular M matrix. So, $M^{-1}A = I - M^{-1}N$ must fulfill with $M^{-1}A = \rho(M^{-1}A)I - M^{-1}N$. Thus, $\rho(M^{-1}N) = 1$.

C THE BASIC REPRODUCTION NUMBER

The disease transmission model is written as follows,

$$\dot{x}_i = \mathcal{F}_i(x) - \mathcal{V}_i(x) = \mathcal{F}_i(x) - (\mathcal{V}_i^-(x) - \mathcal{V}_i^+(x)) \quad i = 1, \dots n \tag{65}$$

where $x = (x_1, \dots, x_n)^t : x_i \geqslant 0$ represents the population size in the block i (the first m blocks will represent the population that firstly suffer from the disease). $\mathcal{F}_i(x)$ is the new infection appearance rate in the block i, and $\mathcal{V}_i^-(x)$ ($\mathcal{V}_i^+(x)$) express how the population is going out (in) into the block i. Let's define $\mathbf{X}_s$ as the set of free disease states,

$$\mathbf{X}_s = \{x \geqslant 0 | \, x_i = 0, \, i = 1, \dots, m\},$$

so if the system remains in a free disease state, it is not possible an infected population growth because the first m compartments are empty.
Conditions C.1: the vectors $\mathcal{F}(x)$, $\mathcal{V}_i^+(x)$, $\mathcal{V}_i^-(x)$ must accomplish with,

- $x \geqslant 0 \Rightarrow \mathcal{F}(x), \mathcal{V}_i^+(x), \mathcal{V}_i^-(x) \geqslant 0$.
- If the block i is empty, there is no leaving population, that is: $x_i = 0 \Rightarrow V_i^-(x) = 0$.
- $\mathcal{F}_i(x) = 0$ if $i > m$
- $x \in \mathbf{X}_s \Rightarrow \mathcal{F}_i(x), V_i^+(x), V_i^-(0) = 0 : i = 1, \dots, m$
- if $\mathcal{F} = 0$, then the system's eigenvalues will have negative real parts.

Considering that the system stands in the DFE x_0, the system (65) can be linearized around x_0 by using the Taylor approximation method,

$$\dot{x}_i \approx \left(\frac{d\mathcal{F}_i(x)}{dx_j} - \frac{d\mathcal{V}_i(x)}{dx_j} \right) \Big|_{x_0} (x - x_0) \tag{66}$$

and the equation 66 can be rewritten as,

$$\dot{x} \approx \begin{pmatrix} F & 0 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} V & 0 \\ J_3 & J_4 \end{pmatrix} \tag{67}$$

where F and V are matrices with $m \times m$ dimension and they represent the derivatives, evaluated in x_0, of $\mathcal{F}_i$ and $\mathcal{V}_i$, respectively, when $1 \leqslant i, j \leqslant m$.

It can be easily proven that F is non-negative by applying Euler approximation,

$$F = \frac{d\mathcal{F}_i}{dx_j}\Big|_{x_0} \approx \lim_{h \to 0^+} \frac{\mathcal{F}_i(x_0 + he_j) - \mathcal{F}_i(x_0)}{h} = \lim_{h \to 0^+} \frac{\mathcal{F}_i(x_0 + he_j)}{h} \geqslant 0$$

where $e_j = \{e_i | 1 \leqslant i \leqslant m\}$ is a vector with a dimension m which value is equal to 1 when $i = j$, otherwise it is equal to 0. Regarding the V matrix it is a non-singular M matrix (see appendix B to know more about this type of matrices), that is,

$$V = \frac{d\mathcal{V}_i}{dx_j}\Big|_{x_0} \approx \lim_{h \to 0^+} \frac{\mathcal{V}_i(x_0 + he_j) - \mathcal{V}_i(x_0)}{h} = \lim_{h \to 0^+} \frac{\mathcal{V}_i^-(x_0 + he_j) - \mathcal{V}_i^+(x_0 + he_j)}{h}$$

Considering the first m compartments and the second condition that matrices from (65) must fulfill ($x_i = 0 \Rightarrow \mathcal{V}_i^-(x) = 0$),

$$V = \lim_{h \to 0^+} \frac{1}{h} \begin{pmatrix} \mathcal{V}_1(h, 0, \dots) & -\mathcal{V}_1^+(0, h, \dots) & \dots \\ -\mathcal{V}_2^+(h, 0, \dots) & \mathcal{V}_2(0, h, \dots) & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

so V has the Z sign pattern. As consequence of the 5th condition given in **Conditions C.1**, V's eigenvalues will have a positive real part, and therefore V is a non-singular M matrix.

The DFE x_0 is asymptotically stable if the real part of the eigenvalues of the system, given by (67), are negative. In contrast, if the eigenvalues have a positive real part x_0 will be unstable. The subsystem defined by $-J_3 - J_4$ has eigenvalues with negative real part, so the system's stability depends on the subsystem defined by $V - F$. The following theorem defines the basic reproduction number $\mathcal{R}_0$ which is a threshold for the system stability, and it is consequence of $V - F$ stability analysis.

Theorem C.1. *[29] The basic reproduction number $\mathcal{R}_0$ is equal to the spectral radius of $V^{-1}F$ ($\mathcal{R}_0 = \rho(V^{-1}F)$). The system will be locally asymptotically stable around x_0 if $\mathcal{R}_0 < 1$, otherwise if $\mathcal{R}_0 > 1$ it will be unstable.*

Proof: To create the nexus between $\mathcal{R}_0$ and $F - V$ stability, the subsystem will be rewritten as $J = V - F$, which has a Z sing pattern, and definitions and theorems from appendix B will be used. Moreover, the proof will be divided in two steps; they will prove why $\mathcal{R}_0 < 1$ and $\mathcal{R}_0 > 0$ implies stability and instability, respectively.

- Let's consider that $F - V$ is asymptotically stable, so J has eigenvalues with positive real parts. Since J has Z sing pattern and eigenvalues with positive real part, it is a nonsingular M matrix, see Theorem B.1; J can be rewritten as $J = sI - P$ where $s > \rho(P)$.
  By multiplying J with V^{-1} (the inverse of V exists and $V^{-1} \geqslant 0$ because it is a nonsingular M matrix), $V^{-1}J = I - V^{-1}F$ is obtained. Since J is nonsingular, $V^{-1}J$ is nonsingular, see Corollary B.1, and $\rho(V^{-1}F) > 1$.

- Let's consider that $F - V$ has an eigenvalue which real part is equal to zero. Since J has a Z sing pattern and eigenvalues with zero real part, it is a singular M matrix, see Theorem B.2; J can be rewritten as $J = \rho(P)I - P$.
  By multiplying J with V^{-1}, $V^{-1}J = I - V^{-1}F$ is obtained. Since J is singular, $V^{-1}J$ is singular, see Corollary B.1, and $\rho(V^{-1}F) = 1$. Therefore, when $\rho(V^{-1}F) = 1$ the system might be unstable (it depends of the multiplicity of eigenvalues with zero real part), and instable when $\rho(V^{-1}F) > 1$.

D $V^{-}F$'S SPECTRAL RADIUS

This section will focus on explaining the steps followed to obtain the spectral radius of $V^{-1}F$, since V and F matrices dimensions are 3x3 and they contain considerable variables and the procedure might be tricky. Firstly, the inverse matrix V^{-1} of V and the product $V^{-1}F$ will be calculated. Then, $V^{-1}F$ eigenvalues will be obtained and determined which of them will have the maximum absolute value.

The matrix F and V are the following ones,

$$F = \left(\begin{array}{ccc} \beta_e S/N & 0 & (\beta_a p_2 + (1-p_2)(1-p_1\lambda)\beta_s)S/N \\ 0 & 0 & (1-p_2)\beta_s p_1 \lambda S/N \\ 0 & 0 & 0 \end{array} \right) \Bigg|_{P_{dfe}},$$

$$V = \left(\begin{array}{ccc} \mu+\sigma & 0 & 0 \\ 0 & (\mu+\delta+\alpha) & 0 \\ -\sigma & 0 & (\lambda(1-p_2)+\mu+\gamma) \end{array} \right) \Bigg|_{P_{dfe}}$$

Considering that the inverse matrix is defined as

$$V = \frac{V^T}{\det(V)},$$

let's calculate the determinant of V;

$$\det(V) = (\mu+\sigma)(\mu+\delta+\alpha)(\lambda(1-p_1)+\mu+\gamma)$$

and the adjugate of the matrix V,

$$V^T = \left(\begin{array}{ccc} (\mu+\delta+\alpha) & 0 & 0 \\ 0 & (\mu+\sigma)(\lambda(1-p_2)+\mu+\gamma) & 0 \\ \sigma(\mu+\delta+\alpha) & 0 & (\mu+\sigma)(\mu+\delta+\alpha) \end{array} \right)$$

so the inverse matrix is the following one,

$$V^{-1} = \left(\begin{array}{ccc} (\mu+\sigma)^{-1} & 0 & 0 \\ 0 & (\mu+\delta+\alpha)^{-1} & 0 \\ \sigma\left[(\sigma+\mu)(\lambda(1-p_2)+\mu+\gamma)\right]^{-1} & 0 & (\lambda(1-p_2)+\mu+\gamma)^{-1} \end{array} \right)$$

Now, let's calculate the product $M = V^{-1}F$,

$$M = \left(\begin{array}{ccc} \frac{\beta_e S}{(\mu+\sigma)N} & 0 & \frac{(\beta_a p_2+(1-p_2)(1-p_1\lambda)\beta_s)S}{(\mu+\sigma)N} \\ 0 & 0 & \frac{(1-p_2)p_1\lambda S}{(\mu+\delta+\alpha)} \\ \frac{\sigma\beta_e S}{(\lambda(1-p_2)+\mu+\gamma)(\sigma+\mu)N} & 0 & \frac{\sigma(\beta_a p_2+(1-p_2)(1-p_1\lambda)\beta_s)S}{(\lambda(1-p_2)+\mu+\gamma)(\sigma+\mu)} \end{array} \right)$$

so the eigenvalues are obtained by solving the next equation,

$$-\lambda' \left| \begin{array}{cc} \frac{\beta_e S}{(\mu+\sigma)N} - \lambda' & \frac{(\beta_a p_2+(1-p_2)(1-p_1\lambda)\beta_s)S}{(\mu+\sigma)N} \\ \frac{\sigma\beta_e S}{(\lambda(1-p_2)+\mu+\gamma)(\sigma+\mu)N} & \frac{\sigma(\beta_a p_2+(1-p_2)(1-p_1\lambda)\beta_s)S}{(\lambda(1-p_2)+\mu+\gamma)(\sigma+\mu)} - \lambda' \end{array} \right| = 0$$

$$\Rightarrow \lambda'_{1,2,3} = 0,\, 0,\, \left[\frac{\beta_e}{(\mu+\sigma)} + \frac{\sigma(\beta_a p_2+(1-p_2)(1-p_1\lambda)\beta_s)}{(\lambda(1-p_2)+\mu+\gamma)(\sigma+\mu)}\right] \frac{S}{N}\Big|_{S=\Lambda/\mu}$$

and the spectral radius,

$$\rho(V^{-1}F) = \left[\frac{\beta_e}{(\mu+\sigma)} + \frac{\sigma(\beta_a p_2+(1-p_2)(1-p_1\lambda)\beta_s)}{(\lambda(1-p_2)+\mu+\gamma)}\right] \frac{S}{N}\Big|_{S=\Lambda/\mu}$$

REFERENCES

[1] Z. Allam. Chapter 1 - the first 50 days of covid-19: A detailed chronological timeline and extensive review of literature documenting the pandemic. In Z. Allam, editor, *Surveying the Covid-19 Pandemic and its Implications*, pages 1–7. Elsevier, 2020.

[2] G. B. Arfken, H. J. Weber, and F. E. Harris. Chapter 23 - probability and statistics. In G. B. Arfken, H. J. Weber, and F. E. Harris, editors, *Mathematical Methods for Physicists (Seventh Edition)*, pages 1125–1179. Academic Press, Boston, seventh edition edition, 2013.

[3] A. Berman and R. J. Plemmons. Chapter 6 - m-matrices. In A. Berman and R. J. Plemmons, editors, *Nonnegative Matrices in the Mathematical Sciences*, pages 132 – 164. Academic Press, 1979.

[4] S. L. Brunton and J. N. Kutz. *Regression and Model Selection*, page 117–153. Cambridge University Press, 2019.

[5] O. Byambasuren, O. Byambasuren, M. Cardona, K. Bell, J. Clark, M. Mclaws, and P. Glasziou. Estimating the extent of asymptomatic covid-19 and its potential for community transmission: systematic review and meta-analysis. 06 2020.

[6] S. Chapra, R. Canale, J. Brito, and M. Hano. *Métodos numéricos para ingenieros*. Elibro Catedra. McGraw-Hill, 2006.

[7] G. Chen. *Stability of Nonlinear Systems*. American Cancer Society, 2005.

[8] W. Chiu, R. Fischer, and M. Ndeffo-Mbah. State-level needs for social distancing and contact tracing to contain covid-19 in the united states. *Nature Human Behaviour*, 4:1080, 10 2020.

[9] S. C. de Salud. Evolución del coronavirus covid-19 en cantabria [database]. https://www.scsalud.es/coronavirus, 2021.

[10] I. de Salud Carlos III. Informe covid-19 nº 9. 13 de marzo de 2020. https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/EnfermedadesTransmisibles/Documents/INFORMES/Informes%20COVID-19/Informe%20COVID-19.%20N%c2%ba%209_13marzo2020_ISCIII.pdf, 13 March 2020.

[11] I. de Salud Carlos III. Covid-19 [database]. https://cnecovid.isciii.es/covid19/#ccaa, 25/06/2021.

[12] J. del Estado. Real decreto-ley 21/2020, de 9 de junio, de medidas urgentes de prevención, contención y coordinación para hacer frente a la crisis sanitaria ocasionada por el covid-19. https://www.boe.es/boe/dias/2020/06/10/pdfs/BOE-A-2020-5895.pdf, 9 June 2020.

[13] C. for Disease Control and Prevention. Cdc confirms person-to-person spread of new coronavirus in the united states. https://www.cdc.gov/media/releases/2020/p0130-coronavirus-spread.html, 30 January 2020.

[14] INE. Indicadores demográficos básicos [database]. https://www.ine.es/dyngs/INEbase/operacion.htm?c=Estadistica_C&cid=1254736177003&menu=resultados&secc=1254736195380&idp=1254735573002#!tabs-1254736195380, 2019.

[15] T. Jie and etal. Transmission and clinical characteristics of asymptomatic patients with sars-cov-2 infection. *Future Virology*, 2020 May.

[16] M. J. Keeling and P. Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2008.

[17] F. Konané and A. Traoré. Modeling the effects of contact tracing on covid-19 transmission. *PubMed*, 2020, 09 2020.

[18] S. Lee, T. Kim, E. Lee, C. Lee, H. Kim, H. Rhee, S. Y. Park, H.-J. Son, S. Yu, J. W. Park, E. J. Choo, S. Park, M. Loeb, and T. H. Kim. Clinical Course and Molecular Viral Shedding Among Asymptomatic and Symptomatic Patients With SARS-CoV-2 Infection in a Community Treatment Center in the Republic of Korea. *JAMA Internal Medicine*, 08 2020.

[19] T. G. Lennart Ljung. Modelling of dynamic systems. chapter 9, pages 227–253. PTR Prettince Hall, Englewood Cliffs, New Jersey 07632, 1994.

[20] W.-C. Liu and S. Zhang. Variable gradient approach to construct lyapunov function for judging stability of non-linear systems. *DEStech Transactions on Engineering and Technology Research*, 06 2017.

[21] L. Markus and H. Yamabe. Global stability criteria for differential systems. *Osaka Math. J.*, 12(2):305–317, 1960.

[22] C. McAloon, A. Collins, K. Hunt, and a. et. Incubation period of covid-19: a rapid systematic review and meta-analysis of observational research. *BMJ Open*, 10(8): e039652, 2020 08 16, 2020.

[23] W. H. Organization. Coronavirus disease 2019 (covid-19) situation report – 51. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57_10, 11 March 2020.

[24] W. H. Organization. Who director-general's opening remarks at the media briefing on covid-19 - 11 march 2020. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020, 11 March 2020.

[25] W. H. Organization. Coronavirus disease 2019. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200402-sitrep-73-covid-19.pdf, 2 April 2020.

[26] W. H. Organization. Novel coronavirus(2019-ncov) situation report - 10. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200130-sitrep-10-ncov.pdf?sfvrsn=d0b2e480_2, 30 January 2020.

[27] J. Rawlings, S. Pantula, and D. Dickey. *Applied regression analysis*. Springer texts in statistics. Springer, New York, NY [u.a.], 2. ed edition, 1998.

[28] A. Traore. Modeling the effects of contact tracing on covid-19 transmission. *Advances in Difference Equations*, 2020, 09 2020.

[29] P. van den Driessche and J. Watmough. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences*, 180(1):29 – 48, 2002.

[30] L. Wang and etal. Inference of person-to-person transmission of covid-19 reveals hidden superspreading events during the early outbreak phase. *Nature Communications*, 11, 06 2020.

[31] W. Wei, Z. Li, C. Chiew, S. Yong, M. Toh, and V. Lee. Presymptomatic transmission of sars-cov-2 - singapore, january 23-march 16, 2020. *MMWR Morb Mortal Wkly Rep*, 69(14):411-415, 2020 Apr 10.

[32] S. Weisberg. *Applied Linear Regression*. Wiley, Hoboken NJ, fourth edition, 2014.

[33] I. wen Kuo. The moore-penrose inverses of singular m-matrices. *Linear Algebra and its Applications*, 17(1):1 – 14, 1977.

[34] S. Zhang, W. Liu, Y. Tang, and C. Luo. Crasovskii approach to construct lyapunov function and its derivative function for analyzing stability of non-linear systems. *IOP Conference Series: Earth and Environmental Science*, 69:012093, jun 2017.

[35] Y. Zhang, D. Muscatello, Y. Tian, Y. Chen, S. Li, W. Duan, C. Ma, Y. Sun, S. Wu, L. Ge, P. Yang, L. Jia, Q. Wang, and C. R. MacIntyre. Role of presymptomatic transmission of covid-19: evidence from beijing, china. *Journal of Epidemiology & Community Health*, 2020.

[36] L. Zou, F. Ruan, M. Huang, L. Liang, H. Huang, Z. Hong, J. Yu, M. Kang, Y. Song, J. Xia, Q. Guo, T. Song, J. He, H.-L. Yen, M. Peiris, and J. Wu. Sars-cov-2 viral load in upper respiratory specimens of infected patients. *New England Journal of Medicine*, 382(12):1177–1179, 2020. PMID: 32074444.