UNIVERSITAT POLITÈCNICA DE VALÈNCIA

# Optimization algorithm based on Novelty Search applied to the treatment of uncertainty in models

Autor: D. David Martínez Rodríguez

Director: Dr. Juan Carlos Cortés López
Dr. Rafael J. Villanueva Micó

Departamento de Matemática Aplicada
Universitat Politècnica de València

**Octubre de 2021**

# Declaration of Authorship

Dr. Juan Carlos Cortés López and Rafael Jacinto Villanueva Micó, full professors at the Universitat Politècnica de València,

CERTIFY that the present PhD thesis entitled: *Optimization algorithm based on Novelty Search applied to the treatment of uncertainty in models* has been performed under our supervision at the *Departamento de Matemática Aplicada* in the Universitat Politècnica de València by David Martínez Rodríguez. It constitutes his thesis dissertation to obtain the PhD degree in Mathematics with international mention.

In compliance with the current legislation, we authorize the presentation of this dissertation signing the present certificate.

**Valencia - October 2021**

**Dr. Juan Carlos Cortés López**

**Dr. Rafael Jacinto Villanueva Micó**

*A todos aquellos que,*
*consciente o inconscientemente,*
*me han hecho llegar hasta aquí.*

# Abstract

Novelty Search is a recent paradigm in evolutionary and bio-inspired optimization algorithms, based on the idea of forcing to look for those unexplored parts of the domain of the function that might be unattractive for the algorithm, with the aim of avoiding stagnation in local optima. Novelty Search has been applied to the Particle Swarm Optimization algorithm, obtaining a new algorithm named Novelty Swarm (NS). NS has been applied to the CEC2005 benchmark, comparing its results with other state of the art algorithms. The results show better behaviour in high nonlinear functions at the cost of increasing the computational complexity. During the rest of the thesis, the NS algorithm has been used in different models, specifically the design of an Internal Combustion Engine, the prediction of energy demand estimation with Grammatical Swarm, the evolution of the bladder cancer of a specific patient and the evolution of COVID-19. It is also remarkable that, in the study of COVID-19 models, uncertainty of the data and the evolution of the disease has been taken in account.

# Resumen

La búsqueda novedosa es un nuevo paradigma de los algoritmos de optimización evolucionarios y bioinspirados, que está basado en la idea de forzar la búsqueda del óptimo global en aquellas partes inexploradas del dominio de la función que no son atractivas para el algoritmo, con la intención de evitar estancamientos en óptimos locales. La búsqueda novedosa se ha aplicado al algoritmo de optimización de enjambre de partículas, obteniendo un nuevo algoritmo denominado algoritmo de enjambre novedoso (NS). NS se ha aplicado al conjunto de pruebas sintéticas CEC2005, comparando los resultados con los obtenidos por otros algoritmos del estado del arte. Los resultados muestran un mejor comportamiento de NS en funciones altamente no lineales, a cambio de un aumento en la complejidad computacional. En lo que resta de trabajo, el algoritmo NS se ha aplicado en diferentes modelos, específicamente en el diseño de un motor de combustión interna, en la estimación de demanda de energía mediante gramáticas de enjambre, en la evolución del cáncer de vejiga de un paciente concreto y en la evolución del COVID-19. Cabe remarcar que, en el estudio de los modelos de COVID-19, se ha tenido en cuenta la incertidumbre, tanto de los datos como de la evolución de la enfermedad.

# Resum

La cerca nova és un nou paradigma dels algoritmes d'optimització evolucionaris i bioinspirats, que està basat en la idea de forçar la cerca de l'òptim global en les parts inexplorades del domini de la funció que no són atractives per a l'algoritme, amb la intenció d'evitar estancaments en òptims locals. La cerca nova s'ha aplicat a l'algoritme d'optimització d'eixam de partícules, obtenint un nou algoritme denominat algoritme d'eixam nou (NS). NS s'ha aplicat al conjunt de proves sintètiques CEC2005, comparant els resultats amb els obtinguts per altres algoritmes de l'estat de l'art. Els resultats mostren un millor comportament de NS en funcions altament no lineals, a canvi d'un augment en la complexitat computacional. En el que resta de treball, l'algoritme NS s'ha aplicat en diferents models, específicament en el disseny d'un motor de combustió interna, en l'estimació de demanda d'energia mitjançant gramàtiques d'eixam, en l'evolució del càncer de bufeta d'un pacient concret i en l'evolució del COVID-19. Cal remarcar que, en l'estudi dels models de COVID-19, s'ha tingut en compte la incertesa, tant de les dades com de l'evolució de la malaltia.

# Contents

**Resumen** **ix**

**Contents** **xv**

**1 General Introduction** **1**

**2 Novelty Swarm algorithm** **5**

2.1 Related Work . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

2.2 Implementation of Novelty Search to the Particle Swarm Optimization Algorithm . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

2.3 CEC2005 Experimental Results . . . . . . . . . . . . . . . . . . . 15

2.4 Conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . 28

**3 Application of Novelty Swarm to the reduction of emissions in combustion systems** **29**

3.1 Related Work . . . . . . . . . . . . . . . . . . . . . . . . . . . 31

3.2 Combustion chamber optimization design . . . . . . . . . . . . . . . . . . 32

3.3 CFD - Novelty Swarm algorithm integration . . . . . . . . . . . . . . . . 37

3.4 Engine Optimization Results . . . . . . . . . . . . . . . . . . . . . . . . 41

3.5 Conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 46

**4 Particle Swarm Grammatical Evolution for Energy Demand Estimation 49**

4.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 49

4.2 Problem Definition . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 53

4.3 Proposed Grammatical Swarm . . . . . . . . . . . . . . . . . . . . . . . 53

4.4 Experiments and Results . . . . . . . . . . . . . . . . . . . . . . . . . . 57

4.5 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 64

4.6 Conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 67

**5 Mathematical modelling of the dynamics of the bladder cancer and the immune response applied to a patient: Evolution and short-term prediction 69**

5.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 69

5.2 Available data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 72

5.3 Model building . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 73

5.4 Model calibration . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 79

5.5 Predictions and patient's follow-up . . . . . . . . . . . . . . . . . . . . 82

5.6 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 86

6 Effect of the early use of antivirals on the COVID-19 pandemic. A computational network modelling approach 89

6.1 Introduction and motivation . . . 89

6.2 Model building . . . 92

6.3 Model validation . . . 101

6.4 Simulation of the effect of the antiviral . . . 102

6.5 Conclusion . . . 109

7 Modelling COVID-19 with uncertainty in Granada, Spain. Intra-hospitality circuit and expectations over next months 111

7.1 Introduction . . . 111

7.2 Materials and Methods . . . 113

7.3 Results . . . 124

7.4 Discussion . . . 134

7.5 Conclusions . . . 135

8 Conclusion 141

Bibliography 143

# Chapter 1

# General Introduction

Mathematical models are a powerful tool used to describe the processes of nature. Different techniques may be used (ie. differential equations, difference equations, graphs, random networks, linear programming, integer programming, etc.), and many of them have in common the need of an optimization algorithm in order to obtain the value of the parameters which describe the behaviour of a specific situation.

The usual way of modelling nature processes is deterministic. However, nature is not deterministic, but random. There are many situations where exact data are not available (contains errors), or the situation has a random behaviour, and in such cases, uncertainty must be taken in account.

Optimization algorithms may be divided into exact and non exact. The first ones ensure to obtain the optimum of the function, but they have the disadvantage of taking a high computational effort. Non exact ones cannot ensure the solution found is optimum, but in return, they have a lower computational cost. This high computational cost of exact optimization algorithms is problematic when trying to study the reality, due to the complexity of mathematical models when they try to be realistic.

This PhD thesis is going to study an optimization paradigm named Novelty Search, which has as purpose the improvement of non exact optimization algorithms through the exploration and exploitation of parts of the domain of the function that have been unvisited. More specifically, we have applied it to the Particle Swarm Optimization algorithm, naming this new algorithm Novelty Swarm.

Usually, non exact optimization algorithms are not designed to optimize stochastic functions. They use to stagnate in local optima in deterministic functions, and this effect use to be higher when uncertainty is included in the modelling. The application of a paradigm such as Novelty Search may improve the behaviour of non exact optimization algorithms in models which include uncertainty.

The models developed through the chapters of this documents are variated, and they are not focused on a single event or a single technique. This has been made in order to study the behaviour of the new algorithm applied to different situations.

Difference equations, grammatical swarm and random networks are used to study combustion engines design, energy consumption prediction, bladder cancer description and COVID-19 spread. All this themes and techniques treated in this thesis show the wide range of applications where Novelty Swarm has a good behaviour, obtaining good enough solutions to the problems proposed.

The PhD thesis is structured in six different chapters, where different issues and themes are developed. All chapter are linked with two main points. The first one is the use of the Novelty Swarm algorithm, which has allowed us to obtain calibrations of the models proposed in such a way that the mathematical description of the studied processes is as close as possible to the reality given by the data. The second one is the use of different methods, trying to study the random nature of the processes modelled. We have use a set of techniques, each one more appropriated to the characteristics of each problem proposed.

Between these techniques, we can mention the evolution of different scenarios in deterministic approaches and the calibration with uncertainty when taking in account the randomness of the evolution of the model. However, even if these techniques are different, all of them pursue the same objective, which is adding the non deterministic behaviour of the reality to the description made of it.

The first chapter is a description of the concept Novelty Search and how it has been applied to the Particle Swarm Optimization algorithm. This new algorithm is called Novelty Swarm. CEC2005 benchmark has been executed in order to compare its behaviour with the original algorithm and other state of the art algorithms. The second chapter is the optimization of an Internal Combustion Engine with the Novelty Swarm algorithm. The design of the optimization process is shown, which leads us to a new engine with lower emissions and a better fuel consumption. The third chapter is the application of Grammatical Swarm to the prediction of energy consumption in Spain and France. Novelty Swarm grammar engine shows good results compared with Genetic Algorithm and Differential Evolution grammar engines. The fourth chapter is the creation of a mathematical model which is able to describe the dynamics of the bladder cancer under treatment. It is applied to a specific patient in order to predict the date of a relapse of the tumour, trying to remove it before metastasis appears. The fifth chapter studies the effect of antivirals to the evolution of COVID-19 pandemics in Spain. A random network which models the spread of COVID-19 through the population, and it is calibrated with Novelty Swarm algorithms. Even though antivirals have not been developed as it was expected at the time this chapter was written, they show a good effect breaking up the transmission chain of the virus, conclusion that might be useful with other infectious diseases. The sixth chapter also studies the evolution of COVID-19, but in this occasion, a difference equation system is used. Instead of studying the total Spanish population, we are focused on the province of Granada thanks to an integrated information system with the Hospital Universitario Virgen de las Nieves, which allows us to have a complete flow of patients through the hospital facilities. This model is used to predict the effect of the restrictive social measures in order to organize the logistics of the hospitals of the province, trying to avoid hospital overloads.

# Chapter 2

# Novelty Swarm algorithm

Particle Swarm Optimization (PSO) [1] has become a powerful search and optimization tool. In the literature we can find many applications for both real world problems [2], [3], [4] and theoretical studies [5], [6], [7]. PSO does not work well for all types of problems and sometimes presents the usual disadvantages of other bio-inspired algorithms, such as premature convergence, dependence on configuration parameters or obtaining local optimum solutions. For this reason, a whole set of variants have been developed to solve the aforementioned problems.

Following the classification of Gou et al. [8], recent studies mainly focused on improving the performance of PSO can be classified into three groups:

1. the study and selection of algorithm parameters (number of particles, evaluations, rates and probabilities, etc.),
2. the modification of the structure and topology of communication between particles, and,

3. the development of hybrid algorithms that combine PSO with other heuristics, seeking to improve the exploration of solution space and the exploitation of promising solutions.

In this chapter, we propose an improvement of PSO, by working on both, the second and the third of the above groups. Specifically, we are investigating the performance of a new hybrid implementation of PSO, in which we propose the use of Novelty Search concepts [9] to improve exploration and exploitation. We call this variant Novelty Swarm (NS) optimization algorithm.

Novelty Search is a recent paradigm in evolutionary and bio-inspired optimization algorithms developed by J. Lehman and K. O. Stanley. It is based on the idea that fitness function-based algorithms "may actively misdirect search towards dead ends" [10], this is, getting stuck in local optima due to the gradient of the objective function. In population based bio-inspired algorithms, individuals are representation of the solutions in the domain. The Novelty Search paradigm proposes that nature has no objective, but just evolves, and any result of this evolution is because of natural selection. It is simply the concatenation of multiple events over time. Evolutionary and bio-inspired algorithms may miss the real solution to the problem because they do not look for novelty solutions, even if they are not always the best fitted [11].

Evolution cells illustrate what has been previously explained. In fact, if we take a fitness function with a parameter that includes reproduction, unicellular organisms are much more efficient in reproduction than multicellular organisms, even if the latter are more evolved and better adapted to the environment [12]. The problem that might arise is that the evolutionary or bio-inspired algorithm does not reach the ideal solutions because it gets stuck in a local optimum (in this case, unicellular organisms). In a mathematical analogy, there are some parts of the domain of the function that are not attractive for the algorithm, even if the global optimum is there [12], so the algorithm is stopped at a local optimum of the problem. Thus, two main approaches emerge from this reasoning trying to solve the aforementioned problem. The first one is that fitness functions must be studied thoroughly in order to avoid stopping in local optimum of the problem. This is, in fact, changing the problem to be solved. However, in real world problems, different modelling approaches can be proposed in order to arrive at the same solution, and these approaches may have different properties during the optimization process. The second one is to make some changes in evolutionary and bio-inspired algorithms in order to find the optimum of the fitness function even if it is not in an attractive region. In this work, we focus on the second approach to the problem. We are going to modify a bio-inspired algorithm with the aim at exploring the parts of the domain of the function that might be unattractive, but where the solution may be allocated. This strategy has already been tried with so-called random mutation applied to PSO (M-PSO) [13]; that is to say, with a certain probability, a particle of the algorithm can mutate somewhere in the domain of the function. However, doing this without any kind of control is like flipping a coin; there is no methodology to follow and everything is left to chance that the particle reaches a better optimum. The Novelty Search paradigm just provides a performance method. Those unexplored parts of the function's domain are explored with a methodical search.

The traditional PSO is the base of the proposed algorithm and the concepts of Novelty Search are applied in the search of unexplored areas of the solution's space. This work evaluates the Novelty Swarm Algorithm on the CEC 2005 benchmark suite [14] researching efficiency on all three kinds of problems: unimodal, multimodal and compositional functions, and then we compare them with previous approaches. The CEC2005 has been chosen because it has been proved by several papers that is a really good test for algorithms in complex functions with a lot of local optima [15], [16], [17].

Therefore, the motivation of this work is to obtain an improvement of the PSO algorithm searching unexplored areas of the search space and avoiding the premature convergence of the algorithm using Novelty Search concepts which characterizes the NS.

The rest of this chapter is structured as follows. Section 2.1 presents some modifications of PSO. Section 2.2 introduces the Novelty Swarm. In Section 2.3, we present the results of applying the proposed NS algorithm to CEC2005 benchmark compared with PSO, M-PSO, CAPSO, LSHADE and jSO optimization algorithms. Finally, Section 2.4 outlines our conclusions.

## 2.1 Related Work

There are a lot of works in the literature focusing on modifications of PSO and their combinations with other metaheuristics or local search algorithms. Ali et al. [18] proposed a hybrid PSO-genetic algorithm implementation with a population partitioning (HPSOGA) for large scale optimization problems. They obtained good solutions at the beginning of the process, but the algorithm stagnated with the increase in the number of evaluations. HPSOGA presents a premature convergence to local optimum and has only been tested on a reduced set of unimodal and multimodal problems. Other approaches combine PSO with differential evolution (DE). In [8], Gou et al. proposed a PSO based on individual difference evolution (IDE-PSO). This work separates the swarm into several subgroups based on particles' performance during evolution processes. IDE-PSO was compared with six PSO variants and outperform the other algorithms especially with multimodal functions. In [17], Epitropakis et al. developed a hybrid approach by combining the PSO and differential evolution algorithms. Although the results were not conclusive, the work is very interesting, since they proposed a framework and made an extensive experimental analysis suggesting that the performance of strategies using DE mutation are better than the canonical PSO. More approaches using DE can be found in [19], where a review and a taxonomy on PSO/DE hybridization is made. The combination of meta-heuristics and local search algorithms is usually classified as memetic algorithms, although in recent years the concept of memetic computing has extended the solutions within this category to more complex multi-agent systems and co-evolution systems among others. An interesting approach is presented in [20], where an aging mechanism is implemented by assigning the leader of the swarm a growing age and lifespan. Surrogate models allow researchers to attack problems whose evaluation of solutions has high computational complexity. In [21], Sun et al. designed the surrogate-assisted social learning-based PSO (SL-PSO) algorithm, where PSO and SL-PSO cooperate in the search for the global optimum. SL-PSO focuses on exploration while the PSO concentrates on the exploitation of local solutions. However, the experimental analysis was only carried out on six 50 dimensional and six 100 dimensional benchmark problems. One of the most successful proposals is CAPSO (centripetal accelerated particle swarm optimization) [22] where a PSO scheme is combined with Newton's laws of motion. The experimental results showed that CAPSO accelerates the convergence of PSO. However, when increasing the number of evaluations, CAPSO does not seem a good option. Another approach, with the same acronym, is presented in [23], which introduced chaos into accelerated PSO.

Novelty Search has been used in the recent years in order to improve some optimization algorithms, such as Differential Evolution [24] or Genetic Programming applied to clustering [25] or automatic generation of game levels [26]. The fact that this paradigm has been used previously with relative success is a good sign of its capabilities.

In summary, the approaches that have been presented so far to improve PSO performance do not seem to be evaluated for all types of functions. Also, the problem of premature convergence is still an inconvenience in some of the results. In this work we propose a modification of the PSO which solves some of the problems discussed above. In order to provide this solution, we include the concepts of Novelty Search in a new version of the PSO algorithm, Novelty Swarm optimization algorithm.

## 2.2 Implementation of Novelty Search to the Particle Swarm Optimization Algorithm

Swarm intelligence is a computational paradigm that was introduced at the end of the decade of the 1980s, and published in the early 1990's [27]. It is based on the assumption that artificial intelligence cannot depend solely on individual behaviour and should take into account society's influence. A swarm is formed by an indeterminate amount of *particles* making elemental actions. All particles interact among themselves and the environment without a central control forcing the particles to do any specific action. Particles are subjected to limited abilities for problem resolution. Nevertheless, the interaction among the different particles which form the swarm allows individual behaviour to be improved, thus obtaining good solutions in many different scenarios [28].

PSO is a bio-inspired meta-heuristic algorithm developed by James Kennedy and Russell Eberhart in 1995 [1]. It is based on the social behaviour of bird flocks, which try to make their movements in the most optimal possible way to find food. Since its proposal, the PSO algorithm has been successfully applied to solve a large amount of optimization problems.

Mathematically, given a function $f$

$$f : \mathbb{R}^D \rightarrow \mathbb{R}, \tag{2.1}$$

the optimization problem consists of finding

$$x_{\mathrm{opt}} | f(x_{opt}) \leq f(x), \qquad \forall x \in \mathbb{R}^D, \tag{2.2}$$

in the minimization case, and

$$x_{\mathrm{opt}} | f(x_{opt}) \geq f(x), \qquad \forall x \in \mathbb{R}^D, \tag{2.3}$$

in the maximization case.

The $D$-dimensional domain of function $f$ in $\mathbb{R}^D$ is called the search space. Every point characterized by the $D$-dimensional vector represents a candidate solution to the problem. These vectors are called particles [29].

Let $X$ be a *swarm*, formed by $N$ particles:

$$X = [x_1, x_2, \ldots, x_i, \ldots, x_N], \tag{2.4}$$

where each particle is a D-dimensional vector in the domain of $f$,

$$x_i = [x_{i1}, x_{i2}, \ldots, x_{iD}]. \tag{2.5}$$

During the search for the optimum of the function $f$, a maximum number of iterations $T$ is set. For each iteration $t$, particles update their position according to the following formula:

$$x_i(t+1) = x_i(t) + v_i(t+1), \tag{2.6}$$

where the velocity vector $v_i(t+1)$ is updated according to:

$$\begin{aligned} v_i(t+1) =& w\ \beta \cdot v_i(t) + c_1\ \tau \cdot (p_i - x_i(t)) \\ &+ c_2\ \gamma \cdot (g - x_i(t)). \end{aligned} \tag{2.7}$$

$w$ is the inertia weight, $c_1$ is the individual weight and $c_2$ is the social weight. This three parameters are scalars and their values are settled independently from each other. $w$ values are usually in the range [0.5, 1.5], $c1$ and $c2$ are usually in the range [1, 3], and these values depend on the problem to be solved. Also, $p_i$ stands for the current best position of $x_i$ (local best position) and $g$ stands for the position with the best value among all the particles which have formed the swarm (the global best position).

$\beta$, $\tau$ and $\gamma$ are random vectors. Every element of each random vector is a different sample of a uniform distribution between 0 and 1. The component-wise product $\cdot$ between vectors is carried out as:

$$a \cdot b = (a_1 b_1, \ldots, a_D b_D),$$

where $a = (a_1, \ldots, a_D)$ and $b = (b_1, \ldots, b_D)$.

### 2.2.1 Development of the Novelty Swarm algorithm

As in many biological systems such as ant colonies or beehives, swarm particles may have different roles in the interaction of the community. Because of this, we deploy two different families of particles, where each family has its own task and is ruled by a different velocity equation.

The first family is formed by "conqueror" particles, ruled by equation (2.7) as in the classical PSO. The name is given because they have to "conquer" the optimum of the problem. The second family is formed by "explorer" particles, where the Novelty Search concept is applied. The name is given because they "explore" the unexplored/less attractive region of the search space.

In order to implement Novelty Search in the algorithm, we store all the particles generated by the algorithm in a repository, and the new explorer particles will avoid the regions close to the particles in the repository. If we call $\mathcal{R}(t)$ the repository in the iteration $t$, we define:

$$MC(t) = \frac{\sum\limits_{x \in \mathcal{R}(t)} x}{card(\mathcal{R}(t))}, \tag{2.8}$$

where $card(\mathcal{R}(t))$ is the number of elements of $\mathcal{R}(t)$. $MC(t)$ is the point in the search space that represents all the points in the system in such a way that the behaviour of $MC(t)$ summarizes the behaviour of the system in the iteration $t$. By analogy on the physics idea of an object's centre of mass, we are going to call $MC(t)$ centre of mass.

At the beginning of the algorithm, the repository of the particles is small and the computational cost of calculating the centre of mass of the system is small. However, when the repository grows as the number of iterations of the algorithm increases, the computational cost becomes expensive. To avoid this problem, the size of the repository that stores the positions of the particles that have formed the swarm must have a maximum value. This value must be big enough to represent the proper dispersion of the particles in the domain of the function, but small enough to avoid a high computational cost in the evaluation of the centre of mass. In order to obtain a statistical significance sample of the population, when $\mathcal{R}(t)$ reaches a prefixed maximum number of particles, say $p$, we select $q < p$ particles randomly as representative of all the particles in the repository. $q$ should be a number that allows a fast calculation of the centre of mass. Also, to prevent the excessive weight of local optimum where the conqueror particles may get stuck, only the explorer particles and the initial conqueror particles are added to the repository.

Thus, the equation that rules the explorer particles is a modification of equation (2.7), adding the interaction with the centre of mass and removing the interaction with the global best position of the swarm,

$$\begin{aligned} v_i(t+1) =& w\ \delta \cdot v_i(t) + c_1\ \phi \cdot (p_i - x_i(t)) + \\ & c_3\ \rho \cdot \exp\left(-\alpha \cdot \left|\frac{x_i(t) - MC(t)}{x_{max} - x_{min}}\right|\right) \cdot (x_i(t) - MC(t)), \end{aligned} \tag{2.9}$$

where $x_{max}, x_{min}$ are vectors of dimension $D$ that represent the boundaries of the search space and $\delta$, $\phi$ and $\rho$ are random vectors like in equation (2.7). The quotient $\frac{x_i(t)-MC(t)}{x_{max}-x_{min}}$ should be carried out componentwise.

Equation (2.9) has three different elements corresponding to each of the terms on the right-hand side:

- Inertia (first term): The particle follows its previous direction.
- Individual best (second term): The particle is attracted by its best position.
- Centre of mass (third term): The particle is repulsed by the position of the centre of mass. Figure 2.1 shows the decrease in repulsion as the particle moves away from the centre of mass, modulated by $\alpha > 0$.

The pseudo-code of Novelty Search implementation in PSO algorithm is shown in Algorithm 1. This new algorithm is named Novelty Swarm (NS)

**Algorithm 1** Novelty Swarm (NS) - Minimization.

```
Require: $w, c_1, c_2, c_3, C, E, p, q$
Ensure: $Min\ of\ f$
for Every particle $i = 1, ..., C$ (conquerors) do
    Initialize the position $x_i(0)$
    Evaluate the best individual position $p_i = f(x_i(0))$
    Add $x_i(0)$ to repository
end for
for Every particle $i = 1, ..., E$ (explorers) do
    Initialize the position $x_i(0)$
    Initialize the best individual position $p_i = f(x_i(0))$
    Add $x_i(0)$ to repository
end for
Initialize the best global position (shared between both families) $g = p_i$ such that $f(p_i)$
  is the minimum.
while The stop criterion is not satisfied do
    Update the centre of mass according to equation (??)
    for Every particle $i = 1, ..., C$ (conquerors) do
        Update velocity according equation (2.7)
        Update position according to equation (2.6)
        Evaluate the function $f(x_i(t+1))$
        if $f(x_i(t+1)) \leq f(p_i)$ then
            $p_i = x_i(t+1)$
        end if
        if $f(x_i(t+1)) \leq f(g)$ then
            $g = x_i(t+1)$
        end if
    end for
    for Every particle $i = 1, ..., E$ (explorers) do
        Update velocity according to equation (2.9)
        Update position according to equation (2.6)
        Evaluate the function $f(x_i(t+1))$
        Add $x_i$ to repository
        if $f(x_i(t+1)) \leq f(p_i)$ then
            $p_i = x_i(t+1)$
        end if
        if $f(x_i(t+1)) \leq f(g)$ then
            $g = x_i(t+1)$
        end if
    end for
    if len(repository) $\geq p$ then
        Select randomly $q$ particles of repository
        Discard the unchosen particles from the repository
    end if
end while
```

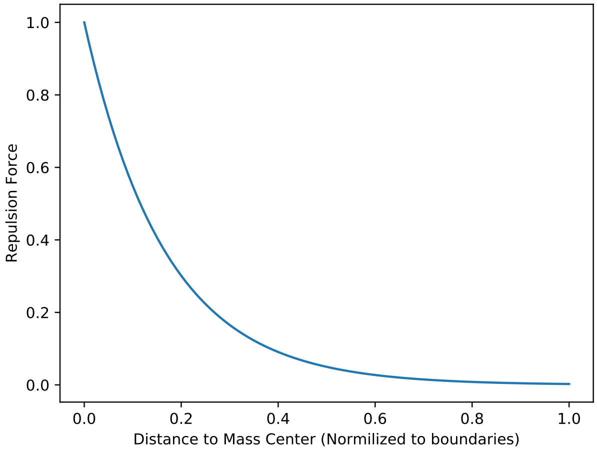


**Figure 2.1:** Evolution of centre of mass repulsion in equation (2.9), $\alpha = 5$. Repulsion decreases with the distance to the centre of mass. The repulsion force is normalized to the boundaries of the domain of the function in the $x$ axis and normalized to the maximum value in the $y$ axis, avoiding to move away from the boundaries.

## 2.3 CEC2005 Experimental Results

In this section, first, we show the CEC2005 benchmark set, a well known and referenced set of problems for evaluation of computational intelligence problems. Then, we describe comparison among the proposed NS with PSO [29], M-PSO [13], [29], CAPSO [22], LSHADE [30] and jSO [30], [31] applied to the CEC2005.

CEC2005 benchmark [14] is a set of functions with different characteristics that has been widely used to compare the performance of optimization algorithms. It is made up of 25 functions that can be divided into:

- Unimodal functions (5). F1 to F5.
- Multimodal functions (20). F5 to F25.
  - Basic functions (7). F6 to F12.

- Expanded functions (2). F13 to F14.
- Hybrid composition functions (11). Formed by unimodal, basic and expanded functions. F15 to F25.

**Table 2.1:** CEC2005 functions execution data [14]. Init. column is the region of the search space where the particles are initialized, Domain column is the search space limit, F(x*) column is the minimum value of the function and Accuracy column is the maximum error allowed to consider the problem solved.

| | **Init.** | **Domain** | **F(x*)** | **Accuracy** |
|---|---|---|---|---|
| F1 | $[-100, 100]^D$ | $[-100, 100]^D$ | -450 | -450 + 1e-6 |
| F2 | $[-100, 100]^D$ | $[-100, 100]^D$ | -450 | -450 + 1e-6 |
| F3 | $[-100, 100]^D$ | $[-100, 100]^D$ | -450 | -450 + 1e-6 |
| F4 | $[-100, 100]^D$ | $[-100, 100]^D$ | -450 | -450 + 1e-6 |
| F5 | $[-100, 100]^D$ | $[-100, 100]^D$ | -310 | -310 + 1e-6 |
| F6 | $[-100, 100]^D$ | $[-100, 100]^D$ | 390 | 390 + 1e-2 |
| F7 | $[0, 600]^D$ | $[-600, 600]^D$ | -180 | -180 + 1e-2 |
| F8 | $[-32, 32]^D$ | $[-32, 32]^D$ | -140 | -140 + 1e-2 |
| F9 | $[-5, 5]^D$ | $[-5, 5]^D$ | -330 | -330 + 1e-2 |
| F10 | $[-5, 5]^D$ | $[-5, 5]^D$ | -330 | -330 + 1e-2 |
| F11 | $[-0.5, 0.5]^D$ | $[-0.5, 0.5]^D$ | 90 | 90 + 1e-2 |
| F12 | $[-\pi, \pi]^D$ | $[-\pi, \pi]^D$ | -460 | -460 + 1e-2 |
| F13 | $[-3, 1]^D$ | $[-3, 1]^D$ | -130 | -130 + 1e-2 |
| F14 | $[-100, 100]^D$ | $[-100, 100]^D$ | -300 | -300 + 1e-2 |
| F15 | $[-5, 5]^D$ | $[-5, 5]^D$ | 120 | 120 + 1e-2 |
| F16 | $[-5, 5]^D$ | $[-5, 5]^D$ | 120 | 120 + 1e-2 |
| F17 | $[-5, 5]^D$ | $[-5, 5]^D$ | 120 | 120 + 1e-1 |
| F18 | $[-5, 5]^D$ | $[-5, 5]^D$ | 10 | 10 + 1e-1 |
| F19 | $[-5, 5]^D$ | $[-5, 5]^D$ | 10 | 10 + 1e-1 |
| F20 | $[-5, 5]^D$ | $[-5, 5]^D$ | 10 | 10 + 1e-1 |
| F21 | $[-5, 5]^D$ | $[-5, 5]^D$ | 360 | 360 + 1e-1 |
| F22 | $[-5, 5]^D$ | $[-5, 5]^D$ | 360 | 360 + 1e-1 |
| F23 | $[-5, 5]^D$ | $[-5, 5]^D$ | 360 | 360 + 1e-1 |
| F24 | $[-5, 5]^D$ | $[-5, 5]^D$ | 260 | 260 + 1e-1 |
| F25 | $[-2, 5]^D$ | $[-5, 5]^D$ | 260 | 260 + 1e-1 |

The optimum is shifted from the origin in all the functions to avoid finding it in the centre of the domain. Table 2.1 shows the general characteristics for each function. For the execution of the test, we implemented the benchmark in Python3 [32] programming language, using Numpy [33] as dependency package. The test was executed on a PC with an Intel Xenon E5-4620 2.20 GHz and

**Table 2.2:** Algorithms parameters setting.

| | **NS** | **PSO** | **M-PSO** | **CAPSO** | **LSHADE** | **jSO** |
|---|---|---|---|---|---|---|
| $w$ | 0.9 | 0.9 | 0.9 | - | - | - |
| $c1$ | 2.0 | 2.0 | 2.0 | - | - | - |
| $c2$ | 2.0 | 2.0 | 2.0 | - | - | - |
| $c3$ | 2.0 | - | - | - | - | - |
| Mutation | 0.05 | - | 0.05 | - | - | - |
| Particles | 60 | 60 | 60 | 60 | - | - |
| Explorer Particles | 30 | - | - | - | - | - |
| $\alpha$ | 5.0 | - | - | - | - | - |
| $NP_0$ | - | - | - | - | 18$\cdot D$ | 18$\cdot D$ |
| $NP_{min}$ | - | - | - | - | 4 | 4 |
| $H$ | - | - | - | - | 6 | 6 |
| $NA_g$ | - | - | - | - | 2.6$\cdot NP_g$ | 2.6$\cdot NP_g$ |
| $p$ | - | - | - | - | 0.11 | 0.11 |
| Reference | - | [29] | [13], [29] | [22] | [30] | [30], [31] |

512 GB RAM. Each of the benchmark functions was executed 25 times for each algorithm for 2, 10 and 30 dimensions. The parameter settings for each algorithm are shown in Table 2.2. In order to select an appropriated parameter $\alpha$, a sensitivity analysis was performed varying the value from 1 to 10. We choose $\alpha = 5$. We also performed the sensitivity analysis in order to obtain the most appropriate number of particles of NS, obtaining 60 conqueror particles and 30 explorer particles.

To set the number of particles $q$ we should take randomly from the repository, we assessed the computational cost of the centre of mass with 3000 to 15000 particles. Looking for a compromise between the costing time and including as much particles as possible, we decided to set the maximum size of the repository $p = 12000$ and a minimum representative number of particles to calculate the mass centre $q = 10000$. The execution stops when the optimum is reached (with a certain error established for each function) or when the maximum function evaluations have been done (even if the optimum has not been reached). The maximum iterations are 10000 for 2 dimensional space and 100000 for 10 and 30 dimensional space. The ending criteria are set by the CEC2005 benchmark, see the 5th column of the Table 2.1.

Tables 2.3, 2.4, 2.5, 2.6, 2.7, 2.8 show, for the 25 runs performed for each function, the average difference between the optimum reached by the algorithms and the real optimum of the function given in Table 2.1 - column F(x*). They also show the standard deviation of the error in the 25 runs of each CEC2005 function. If one of them shows a mean with value 0, it means that the problem has been solved according to the accuracy of Table 2.1. Figures 2.2 and 2.3 show the evolution of the error of the results in Table 2.6 (dimension 10).

Friedman two-way analysis of variances by ranks was performed according to [34], and the results of the calculations are shown in Tables 2.9 and 2.10.

According to Table 2.9, it can be observed that Novelty Swarm has better performance than the other algorithms in 2 and 30 dimensions. Moreover, if we focus our attention on the performance of NS in composition functions, its performance is even better, according to Table 2.10. In 10 dimension, LSHADE and jSO algorithms have better performance in the long run.

In 10 dimensions, Figures 2.2 and 2.3 shows that, regarding to the fitness function when the number of function evaluations is 1000, NS is the best algorithm. NS has better results in 6 of the benchmark functions, followed by PSO, which results are the best in 3 of the benchmark functions. This fact makes NS a good option to optimize functions when the total number of possible evaluations is low.

However, in simpler functions the search space does not need to be explored so thoroughly because the optimum is easier to find. The NS algorithm does not have such good performance in these cases compared with the other algorithms.

Algorithm complexity is calculated according to [14], where $T0$ is the time used to execute a function with basic calculations ($log$, $exp$ and similar), $T1$ is the time used to execute Function 3 of CEC2005 for the specified dimension for 200000 function evaluations and $T2$ is the mean time of 5 optimization processes with the selected algorithm of the same Function 3 with 200000 function evaluations.

Complexity calculations of Table 2.11 prove that the computational cost of NS is greater than the other PSO-based algorithms, and it increases with the

**Table 2.3:** CEC2005 2 dimensional optimization summary results for functions 1 to 14. In bold, the best result for each function.

| | | NS | PSO | M-PSO | CAPSO | LSHADE | jSO |
|---|---|---|---|---|---|---|---|
| F1 | Mean | **0.0000E+4** | **0.0000E+4** | **0.0000E+4** | 2.5579E+0 | **0.0000E+4** | **0.0000E+4** |
| | Std. | 0.0000E+4 | 0.0000E+4 | 0.0000E+4 | 6.9464E+0 | 0.0000E+4 | 0.0000E+4 |
| | Rank | **3** | **3** | **3** | 6 | **3** | **3** |
| F2 | Mean | **0.0000E+4** | **0.0000E+4** | **0.0000E+4** | 1.8407E+1 | **0.0000E+4** | **0.0000E+4** |
| | Std. | 0.0000E+4 | 0.0000E+4 | 0.0000E+4 | 3.5627E+1 | 0.0000E+4 | 0.0000E+4 |
| | Rank | **3** | **3** | **3** | 6 | **3** | **3** |
| F3 | Mean | **0.0000E+4** | 1.9896E+2 | **0.0000E+4** | 6.0831E+3 | 6.5469E+0 | 2.9705E-7 |
| | Std. | 0.0000E+4 | 3.3626E+2 | 0.0000E+4 | 6.8497E+3 | 3.0857E+1 | 1.4552E-6 |
| | Rank | **1.5** | 5 | **1.5** | 6 | 4 | 3 |
| F4 | Mean | **0.0000E+4** | **0.0000E+4** | **0.0000E+4** | 1.5679E+1 | **0.0000E+4** | **0.0000E+4** |
| | Std. | 0.0000E+4 | 0.0000E+4 | 0.0000E+4 | 2.1033E+1 | 0.0000E+4 | 0.0000E+4 |
| | Rank | **3** | **3** | **3** | 6 | **3** | **3** |
| F5 | Mean | **0.0000E+4** | **0.0000E+4** | **0.0000E+4** | **0.0000E+4** | 3.9825E-1 | 4.0892E-1 |
| | Std. | 0.0000E+4 | 0.0000E+4 | 0.0000E+4 | 0.0000E+4 | 1.4415E+0 | 9.0767E-1 |
| | Rank | **2.5** | **2.5** | **2.5** | **2.5** | 5 | 6 |
| F6 | Mean | 1.5353E-1 | 1.9670E+0 | **1.5905E-3** | 2.5597E+2 | 6.0756E-1 | 6.3876E-2 |
| | Std. | 3.7914E-1 | 6.5648E+0 | 7.7917E-3 | 1.0380E+3 | 1.7654E+0 | 1.1562E-1 |
| | Rank | 3 | 5 | **1** | 6 | 4 | 2 |
| F7 | Mean | 1.1201E-2 | **0.0000E+4** | 5.8820E-3 | 7.2155E-1 | 1.0299E-2 | 2.1723E-3 |
| | Std. | 1.2543E-2 | 0.0000E+4 | 1.1008E-2 | 1.0722E+0 | 3.7010E-2 | 6.9524E-3 |
| | Rank | 5 | **1** | 3 | 6 | 4 | 2 |
| F8 | Mean | **0.0000E+4** | 1.6800E+1 | 1.5201E+1 | 1.3641E+1 | 9.5117E+0 | 8.5634E+0 |
| | Std. | 0.0000E+4 | 7.3322E+0 | 8.5420E+0 | 9.0261E+0 | 8.3903E+0 | 8.4227E+0 |
| | Rank | **1** | 6 | 5 | 4 | 3 | 2 |
| F9 | Mean | **0.0000E+4** | **0.0000E+4** | **0.0000E+4** | 1.0269E+0 | **0.0000E+4** | 1.1673E-3 |
| | Std. | 0.0000E+4 | 0.0000E+4 | 0.0000E+4 | 1.7118E+0 | 0.0000E+4 | 5.7186E-3 |
| | Rank | **2.5** | **2.5** | **2.5** | 6 | **2.5** | 5 |
| F10 | Mean | **0.0000E+4** | **0.0000E+4** | **0.0000E+4** | 8.4527E-1 | 1.2980E-2 | 8.8985E-3 |
| | Std. | 0.0000E+4 | 0.0000E+4 | 0.0000E+4 | 1.1075E+0 | 3.9250E-2 | 2.3352E-2 |
| | Rank | **2** | **2** | **2** | 6 | 5 | 4 |
| F11 | Mean | 1.0246E-2 | **0.0000E+4** | 1.7311E-3 | 3.4177E-1 | 4.2881E-2 | 4.9214E-2 |
| | Std. | 9.1898E-3 | 0.0000E+4 | 4.7977E-3 | 3.4729E-1 | 6.4931E-2 | 6.5874E-2 |
| | Rank | 3 | **1** | 2 | 6 | 4 | 5 |
| F12 | Mean | **0.0000E+4** | **0.0000E+4** | **0.0000E+4** | 3.3646E+1 | 4.1021E-3 | 5.4803E-2 |
| | Std. | 0.0000E+4 | 0.0000E+4 | 0.0000E+4 | 1.2149E+2 | 1.9954E-2 | 1.3780E-1 |
| | Rank | **2** | **2** | **2** | 6 | 4 | 5 |
| F13 | Mean | **0.0000E+4** | 4.7350E-3 | 2.3676E-3 | 1.5768E-2 | 9.6409E-4 | **0.0000E+4** |
| | Std. | 0.0000E+4 | 8.4261E-3 | 6.4116E-3 | 2.3694E-2 | 4.7225E-3 | 0.0000E+4 |
| | Rank | **1.5** | 5 | 4 | 6 | 3 | **1.5** |
| F14 | Mean | 1.6945E-2 | **1.2436E-2** | 1.6907E-2 | 1.3414E-1 | 1.5054E-2 | 1.7313E-2 |
| | Std. | 6.3100E-3 | 9.3273E-3 | 6.2886E-3 | 2.3249E-1 | 6.6922E-3 | 1.8965E-2 |
| | Rank | 4 | **1** | 3 | 6 | 2 | 5 |

**Table 2.4:** CEC2005 2 dimensional optimization summary results for functions 15 to 25. In bold, the best result for each function.

| | | NS | PSO | M-PSO | CAPSO | LSHADE | jSO |
|---|---|---|---|---|---|---|---|
| F15 | Mean | **0.0000E+4** | **0.0000E+4** | 4.0000E+0 | 8.6370E+1 | 1.1218E-4 | 7.0381E-1 |
| | Std. | 0.0000E+4 | 0.0000E+4 | 1.9596E+1 | 9.0751E+1 | 5.4809E-4 | 2.6568E+0 |
| | Rank | **1.5** | **1.5** | 5 | 6 | 3 | 4 |
| F16 | Mean | **0.0000E+4** | 1.2000E+1 | 1.6000E+1 | 1.4856E+2 | 1.6090E+1 | 5.8573E+0 |
| | Std. | 0.0000E+4 | 3.2496E+1 | 3.6661E+1 | 1.1851E+2 | 3.6622E+1 | 2.1000E+1 |
| | Rank | **1** | 3 | 4 | 6 | 5 | 2 |
| F17 | Mean | 1.6326E+0 | 5.6130E+0 | 3.6200E+1 | 1.9686E+2 | **7.6461E-1** | 6.5283E+0 |
| | Std. | 7.9981E+0 | 2.0824E+1 | 4.8267E+1 | 1.5600E+2 | 2.0970E+0 | 2.1277E+1 |
| | Rank | 2 | 3 | 5 | 6 | **1** | 4 |
| F18 | Mean | **2.8153E+1** | 2.6000E+2 | 6.4647E+1 | 3.5923E+2 | 1.7270E+2 | 2.5724E+2 |
| | Std. | 7.2186E+1 | 1.6248E+2 | 9.3566E+1 | 1.6928E+2 | 1.1974E+2 | 1.0756E+2 |
| | Rank | **1** | 5 | 2 | 6 | 3 | 4 |
| F19 | Mean | **2.1116E+2** | 2.8000E+2 | 2.3438E+2 | 3.6917E+2 | 2.3727E+2 | 2.6422E+2 |
| | Std. | 4.1428E+1 | 1.2000E+2 | 7.2576E+1 | 1.1153E+2 | 7.8307E+1 | 7.4131E+1 |
| | Rank | **1** | 5 | 2 | 6 | 3 | 4 |
| F20 | Mean | **0.0000E+4** | 2.4800E+2 | 2.2405E+2 | 4.1788E+2 | 2.4197E+2 | 2.2175E+2 |
| | Std. | 0.0000E+4 | 1.1356E+2 | 7.0804E+1 | 1.7039E+2 | 8.9505E+1 | 1.0374E+2 |
| | Rank | **1** | 5 | 3 | 6 | 4 | 2 |
| F21 | Mean | **5.7223E+1** | 2.7355E+2 | 2.2420E+2 | 3.9920E+2 | 2.3957E+2 | 2.1519E+2 |
| | Std. | 8.9159E+1 | 1.6514E+2 | 1.4079E+2 | 1.5854E+2 | 1.6811E+2 | 1.5009E+2 |
| | Rank | **1** | 5 | 3 | 6 | 4 | 2 |
| F22 | Mean | **1.7321E+2** | 3.1288E+2 | 2.2414E+2 | 3.5004E+2 | 2.3841E+2 | 2.1022E+2 |
| | Std. | 7.1371E+1 | 1.2680E+2 | 8.1017E+1 | 1.1849E+2 | 8.5961E+1 | 3.7995E+1 |
| | Rank | **1** | 5 | 3 | 6 | 4 | 2 |
| F23 | Mean | 1.7269E+2 | 2.6094E+2 | **1.4534E+2** | 3.9930E+2 | 2.6942E+2 | 3.5543E+2 |
| | Std. | 1.1838E+2 | 2.2004E+2 | 2.2679E+2 | 2.1432E+2 | 2.0487E+2 | 1.7943E+2 |
| | Rank | 2 | 3 | **1** | 6 | 5 | 4 |
| F24 | Mean | **1.9631E+2** | 2.0000E+2 | 2.0000E+2 | 2.7823E+2 | 1.9822E+2 | 2.0000E+2 |
| | Std. | 1.0595E+1 | 0.0000E+4 | 1.1441E-6 | 1.4708E+2 | 8.7166E+0 | 9.5344E-4 |
| | Rank | **1** | 4 | 4 | 6 | 2 | 4 |
| F25 | Mean | 1.9391E+2 | 1.9825E+2 | **1.6610E+2** | 3.2487E+2 | 3.4097E+2 | 3.9252E+2 |
| | Std. | 1.8537E+1 | 1.6979E+2 | 3.5262E+1 | 2.3101E+2 | 1.6175E+2 | 1.5745E+2 |
| | Rank | 2 | 3 | **1** | 4 | 5 | 6 |

**Table 2.5:** CEC2005 10 dimensional optimization summary results for functions 1 to 14. In bold, the best result for each function.

| | | NS | PSO | M-PSO | CAPSO | LSHADE | jSO |
|---|---|---|---|---|---|---|---|
| F1 | Mean | **0.0000E+4** | 4.1303E+0 | **0.0000E+4** | 3.1124E+3 | 9.6519E-5 | **0.0000E+4** |
| | Std. | 0.0000E+4 | 2.0234E+1 | 0.0000E+4 | 2.1163E+3 | 5.7582E-5 | 0.0000E+4 |
| | Rank | **2** | 5 | **2** | 6 | 4 | **2** |
| F2 | Mean | **0.0000E+4** | 4.4640E+0 | **0.0000E+4** | 6.3584E+3 | 3.0075E+1 | 1.1866E-1 |
| | Std. | 0.0000E+4 | 1.5138E+1 | 0.0000E+4 | 6.1666E+3 | 1.6061E+1 | 3.4815E-1 |
| | Rank | **1.5** | 4 | **1.5** | 6 | 5 | 3 |
| F3 | Mean | **0.0000E+4** | 1.5614E+5 | **0.0000E+4** | 3.2726E+7 | 6.0690E+5 | 3.2170E+5 |
| | Std. | 0.0000E+4 | 1.4563E+5 | 0.0000E+4 | 4.9593E+7 | 3.0767E+5 | 1.9484E+5 |
| | Rank | **1.5** | 3 | **1.5** | 6 | 5 | 4 |
| F4 | Mean | **0.0000E+4** | 1.2605E+1 | **0.0000E+4** | 7.9207E+3 | 5.9575E+1 | 1.7982E+0 |
| | Std. | 0.0000E+4 | 4.1740E+1 | 0.0000E+4 | 7.8746E+3 | 2.9210E+1 | 2.4437E+0 |
| | Rank | **1.5** | 4 | **1.5** | 6 | 5 | 3 |
| F5 | Mean | 2.7818E-5 | 1.7587E+2 | **8.8682E-6** | 2.6710E+3 | 1.1128E+1 | 1.1213E-4 |
| | Std. | 2.9581E-5 | 8.6157E+2 | 1.3840E-5 | 3.1453E+3 | 4.3352E+0 | 1.8152E-4 |
| | Rank | 2 | 5 | **1** | 6 | 4 | 3 |
| F6 | Mean | 6.9881E+1 | 3.1741E+4 | 3.6373E+1 | 7.9056E+8 | 4.0620E+1 | 2.0497E+1 |
| | Std. | 1.9782E+2 | 1.5538E+5 | 6.5570E+1 | 1.3360E+9 | 2.4400E+1 | **2.3982E+1** |
| | Rank | 4 | 5 | 2 | 6 | 3 | **1** |
| F7 | Mean | 6.9383E-1 | 1.2022E+0 | **6.2975E-1** | 2.9133E+2 | 7.8953E-1 | 1.8334E-1 |
| | Std. | 3.7641E-1 | 2.3816E+0 | 4.1380E-1 | 2.2710E+2 | 1.0782E-1 | **1.3599E-1** |
| | Rank | 3 | 5 | 2 | 6 | 4 | **1** |
| F8 | Mean | 2.0343E+1 | **2.0321E+1** | 2.0366E+1 | 2.0366E+1 | 2.0373E+1 | 2.0307E+1 |
| | Std. | 6.9352E-2 | 7.2808E-2 | 7.5030E-2 | 6.2763E-2 | 6.4728E-2 | **8.8295E-2** |
| | Rank | 3 | 2 | 4.5 | 4.5 | 6 | **1** |
| F9 | Mean | 1.2474E+0 | 3.0645E+0 | 4.8101E-1 | 7.8447E+1 | **3.2076E-2** | 4.6175E-1 |
| | Std. | 1.0964E+0 | 1.3473E+0 | 5.7998E-1 | 2.0866E+1 | 1.2507E-2 | 8.3331E-1 |
| | Rank | 4 | 5 | 3 | 6 | **1** | 2 |
| F10 | Mean | 2.0656E+1 | 1.8708E+1 | 2.0584E+1 | 8.8957E+1 | 1.2226E+1 | **1.0819E+1** |
| | Std. | 1.0733E+1 | 6.9397E+0 | 1.1610E+1 | 2.9111E+1 | 2.8072E+0 | 4.7675E+0 |
| | Rank | 5 | 3 | 4 | 6 | 2 | **1** |
| F11 | Mean | 4.9123E+0 | **4.1064E+0** | 4.2574E+0 | 8.9119E+0 | 5.2369E+0 | 4.6904E+0 |
| | Std. | 1.7328E+0 | 1.4138E+0 | 1.5997E+0 | 1.8131E+0 | 7.8148E-1 | 7.5276E-1 |
| | Rank | 4 | **1** | 2 | 6 | 5 | 3 |
| F12 | Mean | 1.7229E+3 | 1.7145E+3 | 1.4790E+3 | 4.3008E+4 | **2.3047E+2** | 2.4477E+2 |
| | Std. | 3.0218E+3 | 3.6642E+3 | 4.0989E+3 | 2.6943E+4 | 1.0950E+2 | 3.2379E+2 |
| | Rank | 5 | 4 | 3 | 6 | **1** | 2 |
| F13 | Mean | 7.4303E-1 | 6.1004E-1 | 7.1509E-1 | 1.2388E+1 | 4.4500E-1 | **3.8843E-1** |
| | Std. | 2.1967E-1 | 2.4371E-1 | 2.7171E-1 | 6.2551E+0 | 1.0604E-1 | 1.7632E-1 |
| | Rank | 5 | 3 | 4 | 6 | 2 | **1** |
| F14 | Mean | 3.2538E+0 | **2.9641E+0** | 3.0484E+0 | 4.0012E+0 | 3.2106E+0 | 3.0844E+0 |
| | Std. | 3.5841E-1 | 4.9332E-1 | 5.0773E-1 | 3.7659E-1 | 2.1339E-1 | 3.3986E-1 |
| | Rank | 5 | **1** | 3 | 6 | 4 | 2 |

**Table 2.6:** CEC2005 10 dimensional optimization summary results for functions 15 to 25. In bold, the best result for each function.

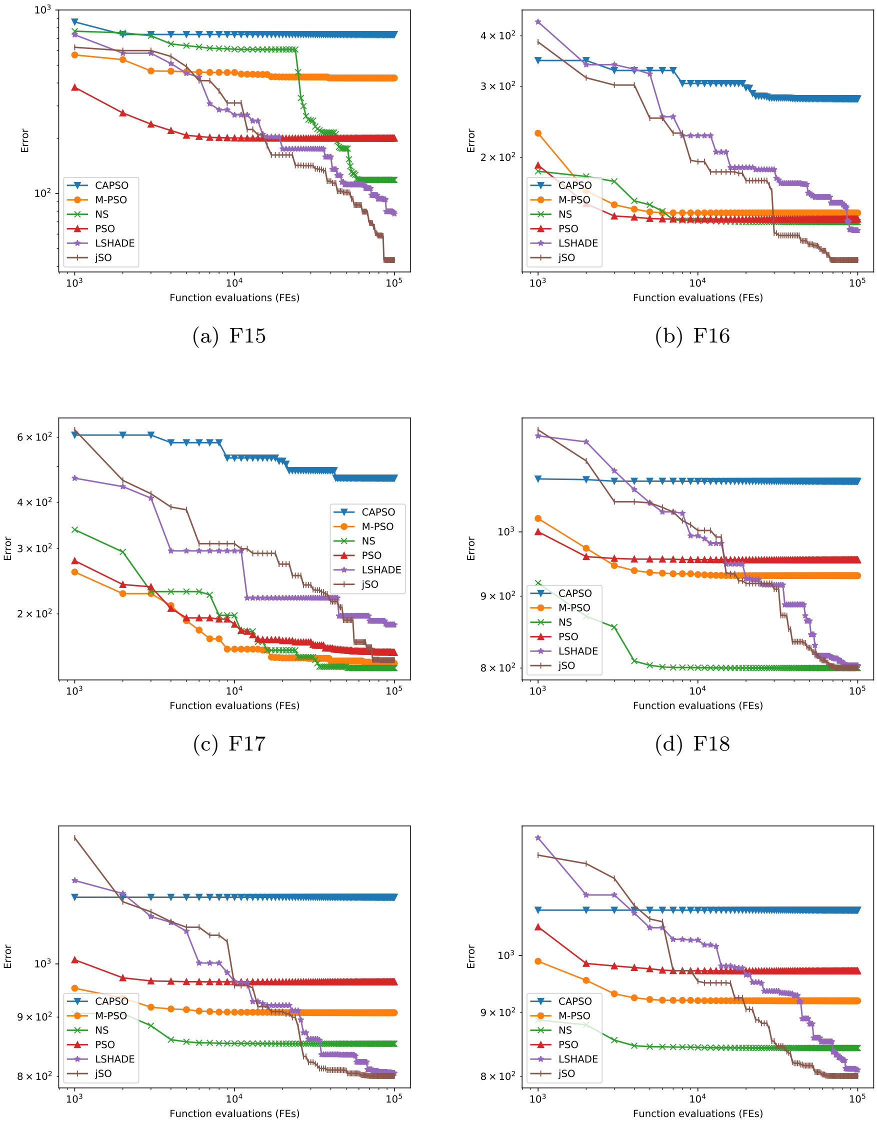
| | | NS | PSO | M-PSO | CAPSO | LSHADE | jSO |
|---|---|---|---|---|---|---|---|
| F15 | Mean | 2.2687E+2 | 2.4113E+2 | 3.3183E+2 | 7.2965E+2 | 7.2330E+1 | **4.4917E+1** |
| | Std. | 2.0275E+2 | 1.6404E+2 | 2.1352E+2 | 9.6639E+1 | 2.2714E+1 | 3.1836E+1 |
| | Rank | 3 | 4 | 5 | 6 | 2 | **1** |
| F16 | Mean | 1.4210E+2 | 1.5240E+2 | 1.4930E+2 | 3.0563E+2 | 1.3508E+2 | **1.1215E+2** |
| | Std. | 2.0791E+1 | 5.6650E+1 | 2.2017E+1 | 9.7846E+1 | 1.6558E+1 | 8.5614E+0 |
| | Rank | 3 | 5 | 4 | 6 | 2 | **1** |
| F17 | Mean | **1.4974E+2** | 1.5656E+2 | 1.5669E+2 | 4.9375E+2 | 1.8725E+2 | 1.5320E+2 |
| | Std. | 2.6083E+1 | 2.9624E+1 | 3.3452E+1 | 2.2330E+2 | 1.2608E+1 | 2.1354E+1 |
| | Rank | **1** | 3 | 4 | 6 | 5 | 2 |
| F18 | Mean | 8.2300E+2 | 9.3370E+2 | 8.3822E+2 | 1.0813E+3 | 7.8654E+2 | **7.5264E+2** |
| | Std. | 1.6717E+2 | 9.8799E+1 | 1.9811E+2 | 8.8321E+1 | 7.0689E+1 | 1.2301E+2 |
| | Rank | 3 | 5 | 4 | 6 | 2 | **1** |
| F19 | Mean | 8.6158E+2 | 9.3526E+2 | 8.3661E+2 | 1.1318E+3 | 7.9574E+2 | **7.6995E+2** |
| | Std. | 1.5852E+2 | 9.8302E+1 | 1.7975E+2 | 7.6381E+1 | 4.5852E+1 | 7.7475E+1 |
| | Rank | 4 | 5 | 3 | 6 | 2 | **1** |
| F20 | Mean | 8.3881E+2 | 9.1932E+2 | 8.6334E+2 | 1.1035E+3 | 8.0170E+2 | **7.7841E+2** |
| | Std. | 1.5090E+2 | 1.2376E+2 | 1.7808E+2 | 9.7235E+1 | 4.9105E+1 | 7.2545E+1 |
| | Rank | 3 | 5 | 4 | 6 | 2 | **1** |
| F21 | Mean | 9.1765E+2 | 8.4462E+2 | 8.2229E+2 | 1.3034E+3 | 4.9655E+2 | **4.8418E+2** |
| | Std. | 3.3248E+2 | 3.7132E+2 | 3.6434E+2 | 1.8055E+2 | 5.3891E+1 | 1.1363E+2 |
| | Rank | 5 | 4 | 3 | 6 | 2 | **1** |
| F22 | Mean | 7.8531E+2 | 8.0734E+2 | 7.9384E+2 | 9.5630E+2 | 7.8631E+2 | **7.6993E+2** |
| | Std. | 4.0272E+1 | 4.9921E+1 | 4.5640E+1 | 6.4496E+1 | 5.9842E+0 | 9.7150E+0 |
| | Rank | 2 | 5 | 4 | 6 | 3 | **1** |
| F23 | Mean | 8.9422E+2 | 1.0661E+3 | 9.2747E+2 | 1.3079E+3 | 5.7309E+2 | **5.4337E+2** |
| | Std. | 2.0851E+2 | 1.8120E+2 | 2.4164E+2 | 5.6407E+1 | 5.3806E+1 | 4.3649E+1 |
| | Rank | 3 | 5 | 4 | 6 | 2 | **1** |
| F24 | Mean | 3.2800E+2 | 6.9416E+2 | 4.0800E+2 | 1.1351E+3 | 2.0003E+2 | **2.0000E+2** |
| | Std. | 2.1075E+2 | 2.9875E+2 | 2.1151E+2 | 1.3570E+2 | 3.5264E-2 | 0.0000E+4 |
| | Rank | 3 | 5 | 4 | 6 | 2 | **1** |
| F25 | Mean | 2.9200E+2 | 8.7564E+2 | 3.4800E+2 | 1.1496E+3 | **2.0000E+2** | **2.0000E+2** |
| | Std. | 2.0380E+2 | 2.3412E+2 | 1.8787E+2 | 1.5780E+2 | 2.2190E-3 | 0.0000E+4 |
| | Rank | 3 | 4 | 5 | 6 | **1.5** | **1.5** |

**Table 2.7:** CEC2005 30 dimensional optimization summary results for functions 1 to 14. In bold, the best result for each function.

| | | NS | PSO | M-PSO | CAPSO | LSHADE | jSO |
|---|---|---|---|---|---|---|---|
| F1 | Mean | **0.0000E+4** | 5.4503E+2 | **0.0000E+4** | 5.9236E+4 | 9.2824E+2 | 5.4566E+2 |
| | Std. | 0.0000E+4 | 5.1217E+2 | 0.0000E+4 | 2.9664E+4 | 1.5182E+2 | 8.4523E+1 |
| | Rank | **1.5** | 3 | **1.5** | 6 | 5 | 4 |
| F2 | Mean | **0.0000E+4** | 7.5932E+2 | **0.0000E+4** | 1.0966E+5 | 2.6828E+4 | 2.0175E+4 |
| | Std. | 0.0000E+4 | 1.7978E+3 | 0.0000E+4 | 4.9057E+4 | 3.9046E+3 | 3.0837E+3 |
| | Rank | **1.5** | 3 | **1.5** | 6 | 5 | 4 |
| F3 | Mean | **0.0000E+4** | 6.8772E+6 | **0.0000E+4** | 4.6978E+8 | 6.1113E+7 | 4.6188E+7 |
| | Std. | 0.0000E+4 | 3.7105E+6 | 0.0000E+4 | 4.1861E+8 | 1.4515E+7 | 9.8415E+6 |
| | Rank | **1.5** | 3 | **1.5** | 6 | 5 | 4 |
| F4 | Mean | **0.0000E+4** | 1.8306E+3 | **0.0000E+4** | 1.0738E+5 | 3.5344E+4 | 3.1080E+4 |
| | Std. | 0.0000E+4 | 1.5752E+3 | 0.0000E+4 | 4.4615E+4 | 4.0673E+3 | 3.6550E+3 |
| | Rank | **1.5** | 3 | **1.5** | 6 | 5 | 4 |
| F5 | Mean | **4.7358E+3** | 5.7386E+3 | 5.0103E+3 | 2.7579E+4 | 8.5086E+3 | 7.0942E+3 |
| | Std. | 1.4503E+3 | 1.6274E+3 | 1.9614E+3 | 5.3745E+3 | 5.1043E+2 | 6.0918E+2 |
| | Rank | **1** | 3 | 2 | 6 | 5 | 4 |
| F6 | Mean | 7.8726E+4 | 2.8159E+7 | **6.8883E+4** | 3.3743E+10 | 1.0424E+7 | 4.1943E+6 |
| | Std. | 2.5561E+5 | 4.1138E+7 | 2.5510E+5 | 3.2102E+10 | 3.2010E+6 | 8.9572E+5 |
| | Rank | 2 | 5 | **1** | 6 | 4 | 3 |
| F7 | Mean | 1.2276E+1 | 6.3930E+2 | **9.7714E+0** | 1.7625E+3 | 1.0267E+2 | 6.2657E+1 |
| | Std. | 3.8350E+0 | 5.1943E+2 | 4.5088E+0 | 7.1474E+2 | 1.7903E+1 | 8.4695E+0 |
| | Rank | 2 | 5 | **1** | 6 | 4 | 3 |
| F8 | Mean | 2.0996E+1 | **2.0962E+1** | 2.0965E+1 | 2.1008E+1 | 2.0975E+1 | 2.0986E+1 |
| | Std. | 5.8077E-2 | 7.0010E-2 | 5.9456E-2 | 4.3315E-2 | 7.2476E-2 | 5.5453E-2 |
| | Rank | 5 | **1** | 2 | 6 | 3 | 4 |
| F9 | Mean | 4.8861E+1 | 3.6908E+1 | **3.2785E+1** | 4.4602E+2 | 7.5165E+1 | 7.1191E+1 |
| | Std. | 1.1648E+1 | 1.2157E+1 | 6.0981E+0 | 6.5390E+1 | 5.1428E+0 | 4.3361E+0 |
| | Rank | 3 | 2 | **1** | 6 | 5 | 4 |
| F10 | Mean | 1.2722E+2 | **9.5479E+1** | 1.2355E+2 | 5.8149E+2 | 2.5411E+2 | 2.4166E+2 |
| | Std. | 3.6738E+1 | 3.8076E+1 | 3.8082E+1 | 1.2665E+2 | 1.6147E+1 | 1.5474E+1 |
| | Rank | 3 | **1** | 2 | 6 | 5 | 4 |
| F11 | Mean | 2.5565E+1 | **2.2611E+1** | 2.2823E+1 | 3.8885E+1 | 3.3810E+1 | 3.2043E+1 |
| | Std. | 3.5215E+0 | 4.0814E+0 | 3.5573E+0 | 3.8102E+0 | 1.5772E+0 | 1.3321E+0 |
| | Rank | 3 | **1** | 2 | 6 | 5 | 4 |
| F12 | Mean | **2.7810E+4** | 3.1153E+4 | 2.8776E+4 | 1.1503E+6 | 1.8528E+5 | 1.4218E+5 |
| | Std. | 1.7805E+4 | 2.3850E+4 | 2.2844E+4 | 4.8001E+5 | 2.1565E+4 | 1.8034E+4 |
| | Rank | **1** | 3 | 2 | 6 | 5 | 4 |
| F13 | Mean | 4.4971E+0 | **3.2139E+0** | 4.5506E+0 | 4.7560E+2 | 1.3153E+1 | 1.1606E+1 |
| | Std. | 1.0547E+0 | 8.2833E-1 | 1.2557E+0 | 3.8171E+2 | 1.2417E+0 | 1.0598E+0 |
| | Rank | 2 | **1** | 3 | 6 | 5 | 4 |
| F14 | Mean | 1.2944E+1 | **1.2734E+1** | 1.2762E+1 | 1.3525E+1 | 1.3350E+1 | 1.3289E+1 |
| | Std. | 2.7132E-1 | 3.7862E-1 | 2.9523E-1 | 2.8549E-1 | 1.4075E-1 | 1.7304E-1 |
| | Rank | 3 | **1** | 2 | 6 | 5 | 4 |

**Table 2.8:** CEC2005 30 dimensional optimization summary results for functions 15 to 25. In bold, the best result for each function.

| | | NS | PSO | M-PSO | CAPSO | LSHADE | jSO |
|---|---|---|---|---|---|---|---|
| F15 | Mean | 4.2850E+2 | 3.9838E+2 | 4.7232E+2 | 9.7835E+2 | 4.0329E+2 | **3.5594E+2** |
| | Std. | 6.3585E+1 | 1.4673E+2 | 1.8866E+2 | 1.6593E+2 | 4.2216E+1 | 4.2931E+1 |
| | Rank | 4 | 2 | 5 | 6 | 3 | **1** |
| F16 | Mean | **2.7335E+2** | 3.7899E+2 | 3.1227E+2 | 7.9408E+2 | 2.9074E+2 | 2.8187E+2 |
| | Std. | 1.4808E+2 | 1.8283E+2 | 1.4777E+2 | 1.7735E+2 | 1.7166E+1 | 1.2618E+1 |
| | Rank | **1** | 5 | 4 | 6 | 3 | 2 |
| F17 | Mean | 4.1152E+2 | **3.3937E+2** | 4.2809E+2 | 8.6913E+2 | 5.1843E+2 | 4.5848E+2 |
| | Std. | 1.1336E+2 | 1.3302E+2 | 1.9242E+2 | 1.8716E+2 | 3.3316E+1 | 3.2996E+1 |
| | Rank | 2 | **1** | 3 | 6 | 5 | 4 |
| F18 | Mean | **9.1202E+2** | 9.3195E+2 | 9.1219E+2 | 1.0997E+3 | 9.2694E+2 | 9.1744E+2 |
| | Std. | 2.3092E+0 | 1.9003E+1 | 3.3989E+0 | 1.0617E+2 | 1.8609E+0 | 1.0614E+0 |
| | Rank | **1** | 5 | 2 | 6 | 4 | 3 |
| F19 | Mean | **9.1346E+2** | 9.4374E+2 | 9.1639E+2 | 1.1074E+3 | 9.2699E+2 | 9.1739E+2 |
| | Std. | 4.4780E+0 | 3.5241E+1 | 1.9272E+1 | 1.0914E+2 | 2.2353E+0 | 1.1603E+0 |
| | Rank | **1** | 5 | 2 | 6 | 4 | 3 |
| F20 | Mean | **9.1137E+2** | 9.3208E+2 | 9.1231E+2 | 1.0822E+3 | 9.2738E+2 | 9.1708E+2 |
| | Std. | 1.6146E+0 | 2.7379E+1 | 2.0168E+0 | 7.7248E+1 | 2.1893E+0 | 1.2174E+0 |
| | Rank | **1** | 5 | 2 | 6 | 4 | 3 |
| F21 | Mean | **7.1840E+2** | 8.1505E+2 | 7.2512E+2 | 1.2791E+3 | 8.5482E+2 | 7.6159E+2 |
| | Std. | 2.4831E+2 | 1.9117E+2 | 2.7918E+2 | 1.2916E+2 | 3.5860E+1 | 2.8277E+1 |
| | Rank | **1** | 4 | 2 | 6 | 5 | 3 |
| F22 | Mean | **9.2496E+2** | 9.4015E+2 | 9.3961E+2 | 1.3126E+3 | 1.0474E+3 | 1.0161E+3 |
| | Std. | 3.8327E+1 | 4.8177E+1 | 4.0714E+1 | 2.1302E+2 | 1.6327E+1 | 1.4372E+1 |
| | Rank | **1** | 3 | 2 | 6 | 5 | 4 |
| F23 | Mean | **6.2683E+2** | 9.1955E+2 | 7.1282E+2 | 1.2446E+3 | 8.5767E+2 | 7.7977E+2 |
| | Std. | 1.7109E+2 | 2.1540E+2 | 2.5594E+2 | 1.0589E+2 | 4.2605E+1 | 2.8693E+1 |
| | Rank | **1** | 5 | 2 | 6 | 4 | 3 |
| F24 | Mean | 8.9105E+2 | 9.6696E+2 | 9.2988E+2 | 1.3029E+3 | 8.2673E+2 | **7.7922E+2** |
| | Std. | 2.4807E+2 | 7.4147E+1 | 2.2214E+2 | 2.2457E+2 | 4.1957E+1 | 4.8687E+1 |
| | Rank | 3 | 5 | 4 | 6 | 2 | **1** |
| F25 | Mean | 9.1076E+2 | 1.0755E+3 | 8.8001E+2 | 1.2403E+3 | 9.0035E+2 | **8.6318E+2** |
| | Std. | 2.3670E+2 | 9.8230E+1 | 2.3884E+2 | 1.5410E+2 | 5.3466E+1 | 5.3058E+1 |
| | Rank | 4 | 5 | 2 | 6 | 3 | **1** |



(a) F15

(b) F16

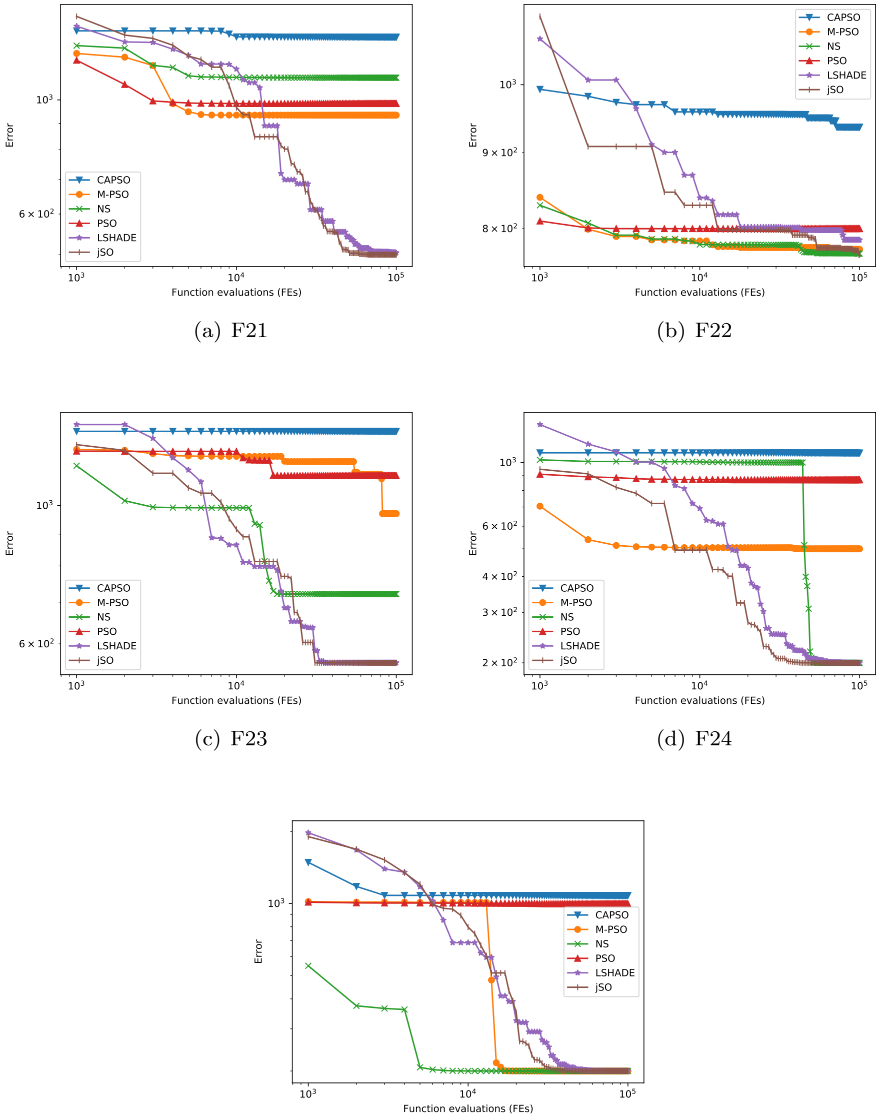
(c) F17

(d) F18

(e) F19

(f) F20

**Figure 2.2:** Convergence of the different algorithms for composition 10 dimensional benchmark from function 15 to 20. The median error run is shown for each algorithm. Figures represent *Error* on the *Y* axis vs number of function evaluations, *FE*, on the *X* axis. Results are in agreement with figures collected in Table 2.6.



(a) F21

(b) F22

(c) F23

(d) F24

(e) F25

**Figure 2.3:** Convergence of the different algorithms for composition 10 dimensional benchmark from function 21 to 25. The median error run is shown for each algorithm. Figures represent *Error* on the $Y$ axis vs number of function evaluations, $FE$, on the $X$ axis. Results are in agreement with figures collected in Table 2.6.

**Table 2.9:** Friedman mean rank for CEC2005 comparison from F1 to F25.

| | NS | PSO | M-PSO | CAPSO | LSHADE | jSO |
|---|---|---|---|---|---|---|
| 2 dimensions | 2.1 | 3.4 | 2.8 | 5.8 | 3.5 | 3.5 |
| 10 dimensions | 3.2 | 4.0 | 3.2 | 6.0 | 3.1 | 1.7 |
| 30 dimensions | 2.0 | 3.2 | 2.2 | 6.0 | 4.3 | 3.3 |

**Table 2.10:** Friedman mean rank for CEC2005 comparison for composition functions (F15 to F25).

| | NS | PSO | M-PSO | CAPSO | LSHADE | jSO |
|---|---|---|---|---|---|---|
| 2 dimensions | 1.3 | 3.9 | 3.0 | 6.0 | 3.5 | 3.5 |
| 10 dimensions | 3.0 | 4.6 | 4.0 | 6.0 | 2.3 | 1.1 |
| 30 dimensions | 1.8 | 4.1 | 2.7 | 6.0 | 3.8 | 2.5 |

**Table 2.11:** Algorithm complexity according to CEC2005 calculations [14].

| T0 | | T1 | | $\hat{T}2$ | $\frac{\hat{T}2-T1}{T0}$ |
|---|---|---|---|---|---|
| 8.5603 | 2 dimensions | 2.6736 | NS | 61.2966 | 6.8482 |
| | | | PSO | 16.5853 | 1.6251 |
| | | | M-PSO | 36.5573 | 3.9582 |
| | | | CAPSO | 39.3964 | 4.2899 |
| | | | LSHADE | 46.6621 | 5.1386 |
| | | | jSO | 42.6005 | 5.0085 |
| | 10 dimensions | 5.1178 | NS | 81.1183 | 8.8782 |
| | | | PSO | 37.6609 | 3.8016 |
| | | | M-PSO | 40.5768 | 4.1422 |
| | | | CAPSO | 42.1109 | 4.3215 |
| | | | LSHADE | 115.6112 | 12.9076 |
| | | | jSO | 118.6915 | 13.2674 |
| | 30 dimensions | 11.1488 | NS | 141.9046 | 15.2746 |
| | | | PSO | 43.9213 | 3.8284 |
| | | | M-PSO | 51.0528 | 4.6615 |
| | | | CAPSO | 48.5369 | 4.3676 |
| | | | LSHADE | 245.6049 | 27.3887 |
| | | | jSO | 251.5818 | 28.0869 |

dimension of the search space, because of the computational cost of the random sampling of particles in the repository. This fact makes NS uncompetitive against other PSO-based algorithms for basic functions. Nevertheless, when the difficulty of the problem to be optimized increases (for example, a problem where the execution time of the evaluation of the function is greater than the execution time of the optimization procedure or a problem where the dimension of the search space is high), NS must be considered as a really good option. NS is also better in complexity than LSHADE and jSO, as its value is lower and it does not increase as fast as the others when increasing the dimensions.

## 2.4 Conclusions

In this work, NS is proposed as an implementation of Novelty Search for the PSO algorithm. The CEC2005 benchmark is used to compare the behavior of NS with PSO, M-PSO, CAPSO, LSHADE and jSO. The experiments have shown that NS has very good results in the optimum search at the expense of increasing the computational cost of the calculations even if few evaluations are required. Furthermore, NS stands out in composition functions, the more complex ones.

# Chapter 3

# Application of Novelty Swarm to the reduction of emissions in combustion systems

The Internal Combustion Engines (ICE) remain as the most used powertrain in transportation [35], as jet engines for air transportation and as reciprocating ICE for marine and land propulsion systems. The interest in these machines is their high fuel efficiency, reasonable durability and moderate pollutant emissions [36]. It is estimated that the transport sector is responsible for around 20% of the total production of the green house gases emissions [37]. In order to attend the regulations of emissions, numerous efforts have been done in the engine industry, specifically in the improvement of combustion systems and in the development of auxiliary devices for aftertreatment. This process of pollutant emission reduction is a challenge since it should also consider the fuel consumption minimization without deteriorating the engine performance, which have an antagonist behavior. The development of the combustion process has been one of the keys for reaching the emission targets, specially for Nitrous Oxides (NOx) and particulate matter (soot).

In the analysis of the combustion performance in compression-ignited engines (CI), computational fluid dynamics (CFD) tools have been used in large scale to simulate the in-cylinder combustion process. In this context, CFD codes are tools commonly used to reproduce, visualize and study the influence of different parameters like the bowl geometry, injection configuration and air management on the combustion behavior [38], [39]. Moreover, these codes are used as support for optimization studies due to their capability for reproducing combustion phenomena, thus generating reliable trends for decision making in design stages. Once the model is validated against experimental data, it is possible to configure different engine settings, finding an optimal design without rebuilding the experimental hardware [40]. Current literature shows a variety of different approaches used for engine optimization, from the traditional design of experiments, to the most innovative as heuristics, meta-heuristics or neural networks [36], [41]–[43].

Therefore, the motivation of this chapter is to obtain an optimized design of a combustion system capable of reducing the pollutant emissions and fuel consumption at the same time using the benefits of the proposed method Novelty Swarm (NS). The goal is to implement the proposed methodology for the obtention of a combustion chamber for a compression-ignited engine based on the reduction of the NOx and soot pollutant emissions, which are a major environmental issue as it was mentioned. For this implementation it is necessary to couple the NS with CFD codes and some other peripheral tools developed to configure the computational cases with their characteristics.

This chapter is structured as follows. Section 3.1 presents a brief summary of other works that have studied optimization in engines. Section 3.2 describes the engine characteristics and the numerical methodology formulation, as well as the CFD model validation and the optimization procedure. Section 3.3 is devoted to the integration of the CFD model and the proposed NS algorithm. Also, fitting functions are defined. In Section 3.4 the engine optimization results are shown, followed by a discussion. Finally, Section 3.5 outlines our conclusions.

## 3.1 Related Work

The optimization of the ICE design is a complex problem since the combustion performance depends on high number of parameters that most of the times are nonlinear, with strong exponential behavior of certain responses and with high cross-effects between variables. The consequence is that the response function would have multiple local optimum values making a challenge to find the global optimum. Another characteristic of the optimization in engines is that the evaluation of the fitness function is highly time consuming, specially when it is coupled with CFD, therefore only few iterations and evaluations can be done. Previous studies proposed a combination of artificial neural networks and ant colony optimization algorithms for reducing soot and NOx emissions in a diesel engine considering 4 inputs variables [41]. Although the work showed potential for application in controlling systems of diesel engines, it is dependent on the algorithm parameters that might cause rapid convergence to local optimum values. Moreover, heuristics algorithms have shown to be effective methods in the optimization of combustion systems. Broatch et al. in [42] combined a CFD modeling with genetic algorithm (GA) technique to optimize the combustion system of a diesel engine, minimizing the indicated specific fuel consumption and combustion noise, while restricting the soot and NOx. The simulations helped to understand the problem, although at a very high computational cost. Another study that implements GA techniques for engine optimization was conducted by Zubel et al. [44], where they combined CFD models and a microgenetic algorithm for the obtention of a piston bowl and injector nozzle geometry design using a fuel with high oxygen content as dimethyl ether. In their work, they tested two possible bowl configurations and they found that the injector design has an impact on the combustion performance. The combination of the spray angle and the piston profiles leaded to a reduction of HC and CO emissions and an improvement of engine efficiency. Recently, Badra et al. [45] presented an optimization of a combustion system using CFD and Machine Learning for reducing the fuel consumption and the pollutant emissions in a compression ignition engine fueled with gasoline. In their procedure a design of experiments (DoE) optimization was performed followed by a Machine Learning – Grid Gradient Ascent approach. Their results show an improvement in the CO and soot emissions. One of the contributions of their exercise is the obtention of acceptable results with relatively low computational cost, compared against traditional CFD-DoE approaches. However, one of the aspects of that study was the narrow range of the parameters tested. In [43] Bertram et al. presented a hybrid method combining PSO and GA for engine performance optimization indicating that the hybrid method offered benefits of both algorithms while preventing the drawbacks of either method alone, such as fast convergence in basic PSO. They concluded that the PSO would provide better results with some modifications or combined with other approaches. In this paper we use the NS algorithm in the optimization problem of reduction of emissions and fuel consumption in combustion systems based on the fast rate of convergence, simple implementation, possibility to use a wide range of inputs parameters and the asynchronous nature of the algorithm what makes possible a time reduction of the optimization process.

## 3.2 Combustion chamber optimization design

In this section, the combustion chamber specifications are described. Then, the methodology is presented, along with the model formulation, the CFD model validation and details of the optimization algorithm and the ICE modeling integration.

### *3.2.1 Engine Specification*

A regular production diesel engine has been used as the experimental platform to evaluate the CFD modeling performance and the current combustion conditions. The engine is a four-cylinder configuration equipped with a turbocharger. A ten-nozzle injector with hole diameter of 112 $\mu$m with a spray angle of 154° was used in all experimental tests. The engine has a compression ratio (CR) of 16 and the main specifications of the engine are in Table 3.1

**Table 3.1:** Engine specifications.

| | |
|---|---:|
| Engine type | Direct-injection diesel engine |
| Number of cylinders [-] | 4 |
| Volume [l] | 2.2 |
| Bore – stroke [mm] | 85-96 |
| Connecting rod length [mm] | 152 |
| Compression ratio [-] | 16:1 |
| Injector number of holes | 10 |

### *3.2.2 Methods*

The aim of this study is to conduct a *3-D* CFD-guided combustion system hardware development using efficient optimization tools to obtain an optimum combustion system through the evaluation of emissions and consumption parameters. The following is an overview of the CFD-guided design process:

1. Model formulation and validation, where the setup of the CFD model is done in order to represent the behaviour of the engine that is being modeled. Additionally, the calibration of the spray and emissions models is performed against experimental data (Section 3.2.3).
2. Mesh simplification, where a mesh with fewer cells than the original model is built and validated with experimental results in order to reduce the computational execution time of the model (Section 3.2.4).
3. Integration between the optimization algorithm and the engine model (Section 3.3).
4. Execution of the optimization procedure (Section 3.4).

### *3.2.3 CFD model formulation*

A numerical model of the combustion chamber was developed using the open source OpenFOAM software [46]. Along with OpenFOAM, LibICE, which is a group of libraries and solvers for internal combustion engines [47], has been used. The combustion process is a combination of complex processes involving physical and chemical phenomena. Thus, in order to mimic those processes, the CFD tool couples a variety of sub-models that simulate the different mechanisms involved, for example, gas motion, spray development, chemistry and heat transmission, among others. In this study, the simulations were carried out in closed cycle, that means from the closing of the intake valves until the opening of the exhaust valves. Also, since the combustion chamber volume varies during the engine cycle, LibICE uses the dynamic mesh layering technique [48], [49] to reproduce the piston movement. In order to reduce the computational cost of the model, which is important in this type of studies where hundreds of simulations are carried out, only an axisymmetric sector (1/10) of the geometry is simulated. While n-heptane was used as the diesel surrogate, a Lagragian-typo was used to mimic the spray, assuming a “Blob” injection method [49], [50]. Liquid droplets were grouped into parcels to statistically represent the spray field in the combustion chamber using a specific rate of injection (ROI) profile with a virtual injector model [51]. Parcels evolve into the CFD domain exchanging mass, momentum and energy with the continuous gas phase. Additional submodels are necessary to reproduce the liquid atomization, break-up, heat transfer and evaporation. Both, the Kelvin-Helmholtz (KH) and Rayleigh-Taylor (RT) algorithms, were adopted for the secondary break-up process [52], [53]. To model the in-cylinder turbulence, the Reynolds-Averaged Navier Stokes (RANS) based on re-normalized group (RNG $k - \epsilon$) was used to perform all the simulations [54]. Pressure and velocity equations are coupled by the PIMPLE algorithm. The Angelberger model was used to compute the wall heat transfer coupled with the turbulence model [55]. For fuel chemistry, a reduced chemical kinetic mechanism for primary reference fuel consisting on 162 species and 1543 reactions was used. Combustion simulations were run using the multiple representative interactive flamelets (MRIF) model, that is based on the laminar flamelet concept and assumes that the chemical scales are much smaller than the turbulent time and length scales [50]. The characteristic of this model is that reacting scalars depend on the mixture fraction variable, which is proportional to the local fuel-to-air ratio. For more detailed information about the combustion model the reader is referred to [56]. Although there are more detailed and accurate models available in the literature, the choice of the sub-models in this study was a trade-off between accuracy and computational efficiency for such a large optimization process. A list of the sub-models used is provided in Table 3.2.

**Table 3.2:** Models specifications.

| | |
|---|---|
| Turbulence | RNG $k-\epsilon$ RANS |
| Wall Heat transfer | Angelberger |
| Spray Models | Injection: Blob Injector<br>Break-up: KH-RT<br>Collision: off<br>Evaporation: standard |
| Combustion Model | RIF-based tabulation |
| Chemical Mechanism | NC7Curran |
| Soot Model | Leung Lindstedt Jones |

### *3.2.4 Model Validation and mesh simplification*

A first CFD model was calibrated to reproduce the experimental data. The engine was operated at 3700 rpm and full power. All the boundary conditions used to simulate the engine were obtained from experiments using the methodology explained in [57]. The left-hand side image in Figure 3.1 presents a cross-section view of the mesh used at top dead center (TDC). It consists of 52000 cells at the TDC and 398000 cells at the intake valves closing (IVC). The used mesh was generated automatically using the method developed in [48]. The mesh motion is integrated in the solver and is composed by multi-regions where each region of the mesh motion is accommodated in different ways. Also, the tool combines the use of different topological changes and polyhedral vertex-based motion solver for mesh deformation based on Finite Element Method.

In Figure 3.2, the in-cylinder pressure and the heat release rate (HRR) of the simulation were compared with the experimental data. The black line refers to the experimental data, and the blue line depicts the results using this mesh (fine mesh). Based on Figure 3.2, the predictions are in fairly good agreement with the experimental data and they provide confidence that the comprehensive design optimization is valid. Despite of the good agreement between the simulations and experiments, CFD simulations are highly time consuming. This compromises its suitability in the use of the optimization technique, that requires a large number of model evaluations. Then, the original model was simplified, using a coarser mesh to reduce the number of cells. In this sense, the calculation time was reduced while ensuring enough accuracy-level. The characteristics of the coarse mesh in relation to the fine mesh are presented in Table 3.3 and in the right-hand side image in Figure 3.1.

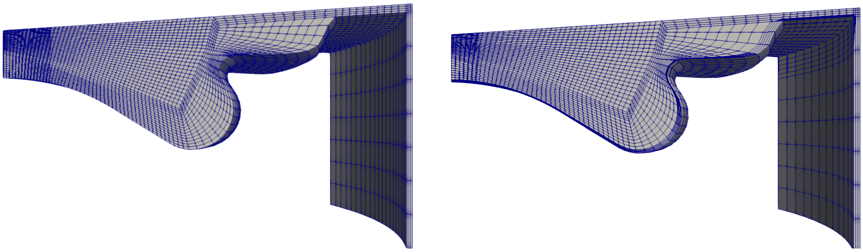


**Figure 3.1:** Fine mesh (left) versus coarse mesh (right).

**Table 3.3:** Comparison between fine and coarse mesh.

| | Fine Mesh | Coarse Mesh |
|---|---:|---:|
| Cell count at TDC | 52000 | 26900 |
| Cell count at IVC | 398000 | 203300 |

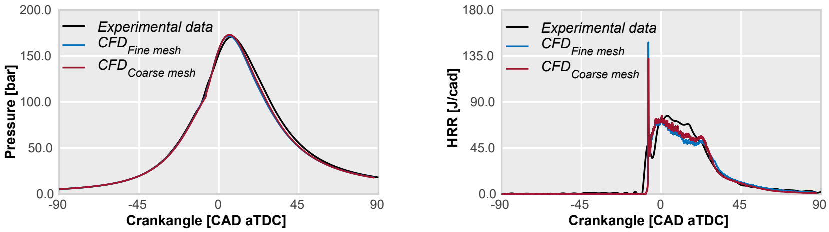


**Figure 3.2:** Comparison between experimental data (black curve) and simulation results for fine and coarse meshes (blue and red curves respectively). Left side: Evolution of the in-cylinder pressure. Right side: Estimated Heat Release Rate.

Using the coarse mesh, the predictions show a good agreement with the experimental data as well. The comparison between experimental data, fine and coarse mesh is shown in Figure 3.2. As expected this configuration provides a simulation time of about 20 hours (on 4 processors), while the fine mesh takes around 30 hours of run-time (on 4 processors). Based on the simulation execution time, the coarse mesh configuration will be employed in the optimization process. Once the CFD simulation is performed, the results are evaluated using a fitness function in order to determine the quality of the solution.

## 3.3 CFD - Novelty Swarm algorithm integration

The procedure for the NS algorithm and the function evaluation using CFD includes the following steps:

1. Candidate solutions are initialized inside the domain of the search space.
2. Candidate solutions are evaluated to obtain the value of the fitness function.
3. Optimization algorithm is updated.
4. New candidate solutions are created.
5. Go to step 2 if stop criteria is not reached.

In order to complete the described steps automatically, some tools were developed for the preparation of the CFD solutions:

- **Geometry generator**: To generate a combustion chamber with plausible design, a piston bowl profile generator was implemented using Bézier polynomial curves with five parameters [58]. All five parameters are dimensionless and have their own ranges and limits providing more freedom in the bowl profile generation. In Figure 3.3, it can be seen the five parameters and two examples of bowl profiles generated by the method. The crevice, which is a small region between piston and cylinder liner, was kept constant in shape during this study. This parametrization approach allowed a flexible variation of the bowl over a wide range. Changing the individual design parameters of the piston bowl obviously has an impact on the compression ratio, which is supposed to stay fixed to a predefined value of 16 in this study. The CR was compensated changing the squish height. Besides, since the spray targeting affects the engine performance, the spray angle was kept constant in this study. The main challenge in the mesh orientation are the re-entrant geometries that can provide negative volumes when the mesh is created. After the bowl profile is defined, every mesh was generated using the DCCmesh tool automatically. The Bézier bowl curve generated is an input and the control points for block definition were updated according the bowl step and re-entrant curvature to avoid negative volumes and skewness issues. The control points and block definition can be seen in Figure 3.4. Additionally, the number of injector orifices was an input for defining the sector mesh.

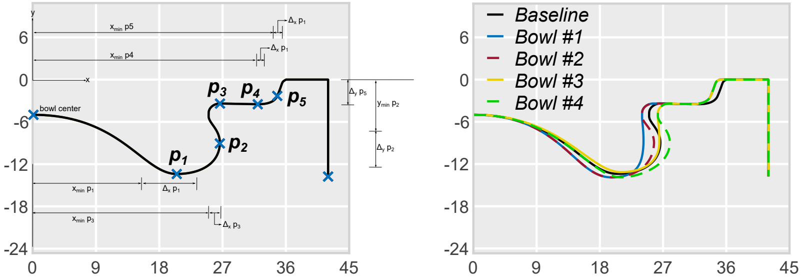


**Figure 3.3:** Parameters definition for Bézier curves (Left-hand side) and example of various generated bowl profiles (Right-hand side).

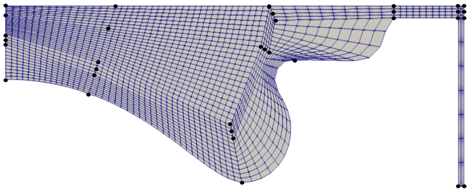


**Figure 3.4:** Block definition and mesh control points location.

- **Virtual Injector Model**: The injection rate profile is a key factor in the combustion of CI engines, since it is able to affect the performance and emissions levels. Considering this, an in-house code was developed to automatically generate the injection profile from the combination of three parameters. They are the injection pressure, the number of injector holes and the total mass fuel injected. The profile of ROI has a trapezoidal form and the code assumes incompressible flow across the nozzle holes and apply the equations of continuity and Bernoulli between inlet and outlet of the orifices [51], [59]. In this study the discharge coefficient was considered equal to 0.88 for all cases (this value was obtained from experimental measurements). Figure 3.5 depicts the good agreement of the curve obtained with the Virtual Injector Model against the experimental data.

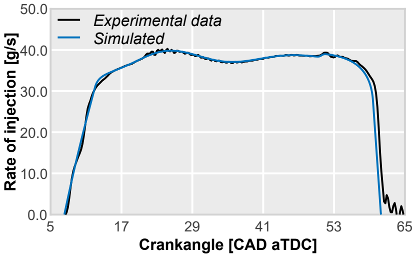


**Figure 3.5:** Rate of injection comparison: experimental data versus in-house code.

### *3.3.1 Optimization parameters and fitness function*

As mentioned before, the target of this optimization is to reduce the specific fuel consumption (ISFC) and the exhaust emissions of NOx and soot. Additionally, the emission values were restricted to a certain level, then, configurations that exceed the emission constrains are accordingly penalized. The operating point used for the CFD model validation was selected as the baseline for the constraint values. This way, the fitness function was formulated to consider the relative importance of ISFC, Soot and NOx against the baseline values. These considerations were expressed in the fitness function as:

$$f_1(NO_x) = \left\{ \begin{array}{ll} \frac{NO_{x,CFD}}{NO_{x,lim}} \quad if\ NO_{x,CFD} < NO_{x,lim} & \\ & \\ \frac{NO_{x,CFD}}{NO_{x,lim}} + 100 \cdot (NO_{x,CFD} - NO_{x,lim})^2 \quad if\ NO_{x,CFD} \geq NO_{x,lim} & \end{array} \right\} \tag{3.1}$$

$$f_2(soot) = \left\{ \begin{array}{ll} \frac{-log(soot_{CFD})}{log(soot_{lim}} \quad if\ soot_{CFD} < soot_{lim} & \\ & \\ \frac{-log(soot_{CFD})}{log(soot_{lim})} + 10^6 \cdot (log(soot_{CFD}) - log(soot_{lim}))^2 \quad if\ soot_{CFD} \geq soot_{lim} & \end{array} \right\} \tag{3.2}$$

$$f_3(ISFC_x) = \left\{ \begin{array}{ll} \frac{ISFC_{CFD}}{ISFC_{lim}} \quad if\ ISFC_{CFD} < ISFC_{lim} & \\ & \\ \frac{ISFC_{CFD}}{ISFC_{lim}} + 100 \cdot (ISFC_{CFD} - ISFC_{lim})^2 \quad if\ ISFC_{CFD} \geq ISFC_{lim} & \end{array} \right\} \tag{3.3}$$

And the total fitness function is

$$FF = f_1(NO_x) \cdot coef_{NO_x} + f_2(soot) \cdot coef_{soot} + f_3(ISFC) \cdot coef_{ISFC} \tag{3.4}$$

where $NO_{x,CFD}$, $soot_{CFD}$ and $ISFC_{CFD}$ are the values obtained in the CFD simulation, whereas $NO_{x,lim}$, $soot_{lim}$ and $ISFC_{lim}$ refers to the emission levels achieved in the baseline configuration. Finally, $coef_{NO_x}$, $coef_{soot}$ and $coef_{ISFC}$ are coefficients used to balance the equation according to the order of magnitude of each term. The value of the objective is the one that feeds back the NS algorithm for the following iteration.

As the main objective is to increase the efficiency of the engine at the same time as it reduces its NOx and soot emissions, the values of the coefficients used are: $coef_{ISFC} = 50$, $coef_{NO_x} = 5$ and $coef_{soot} = 5 \cdot 10^{-5}$.

Nine parameters related to combustion system design were chosen as inputs for the optimization. All the parameters and their ranges (Table 3.4) were selected considering technological limitations in the manufacture of the optimized combustion system. There are five geometrical parameters for defining diverse geometries with a certain degree of freedom. The others are related to the injection system as the number of holes and injection pressure, another associated with the air motion (swirl number), and the last one is the exhaust gas recirculation (EGR). All the parameters were chosen due to their influence on the combustion process and also on the engine emissions.

**Table 3.4:** Range of the input parameters considered in the optimization.

| Parameter | Range |
|---|---:|
| Geometrical parameter 1 [-] | [-0.5, 1.0] |
| Geometrical parameter 2 [-] | [-1.0, 1.25] |
| Geometrical parameter 3 [-] | [-1.0, 1.0] |
| Geometrical parameter 4 [-] | [0.0, 1.0] |
| Geometrical parameter 5 [-] | [-1.4, 0.1] |
| Number of injector nozzles [-] | [4, 12] |
| Swirl number at IVC [-] | [1.0, 3.0] |
| Injection pressure [bar] | [1500, 2000] |
| EGR [%] | [0, 30] |

## 3.4 Engine Optimization Results

In this section, the results of the optimization performed by NS applied to the CI engine are presented and discussed. The main barrier for the optimization process described before is its computational cost. In order to evaluate a single engine configuration (a candidate solution of the algorithm), 4 computational cores are needed, and the time consumed by this process is near 24 hours. For this reason, the maximum number of function evaluations performed in the optimization process is set to 1000. A wide range of the search space must be explored, at the same time that a better optimum than the current engine configuration must be found. Also, the topology of the search space is not known, but it is possible that several local optima exist. The initial step is the evaluation of each constraint parameter separately to verify if each particle satisfy the restrictions. Figure 3.6 shows the distribution of all particles regarding NOx, soot and ISFC sorted by fitness function value (the worst fitness function is sorted on the left side). The graphs in the left and middle represent respectively the NOx and soot emissions with respect to their limit values. Only those configurations which accomplish the limits of all three CFD output limits have a low enough value of the fitness function. This is the reason why, even if the NOx and the soot values are below the imposed limit, if they have bad ISFC value (this is, a higher ISFC than the baseline configuration), they are placed on the left of the figure rather than other configurations with individual worse values.

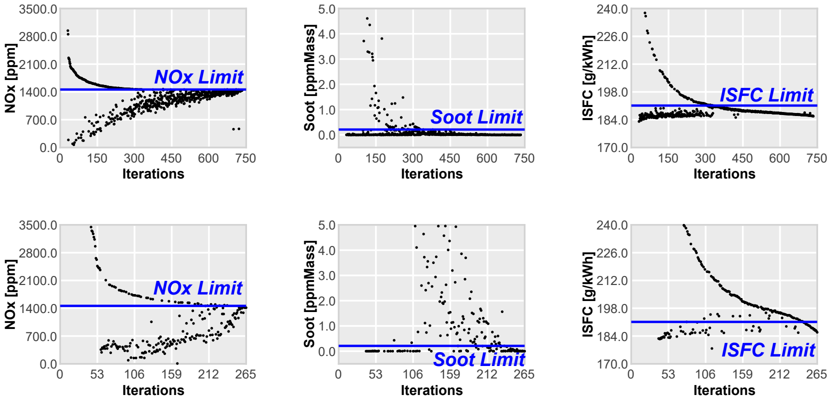


**Figure 3.6:** Fitness results and limits of all particles. Upper row, conqueror particles. Lower row, explorer particles.

In order to locate the particle that provides the best solution within the explored range, NOx, soot and ISFC were contrasted in Figure 3.7. In both figures the characteristic Pareto front of CI engines can be seen, showing the trade-off between NOx and soot, and NOx and ISFC. In those plots, there are several solutions that satisfy the restrictions imposed, improving the optimizing parameters. However, even if there are particles that present better results of NOx, soot or ISFC, separately, the solution of the fitness function could be high, since it depends on all parameters together. For example, if one simulation provides a low value of ISFC, probably the result of NOx is higher, because these parameters have antagonistic behavior in engines and the fitness function presents a higher value because the NOx value penalizes the solution. For this reason, the optimum solution is focused on optimizing the ISFC and soot, while maintaining the NOx within the limits.

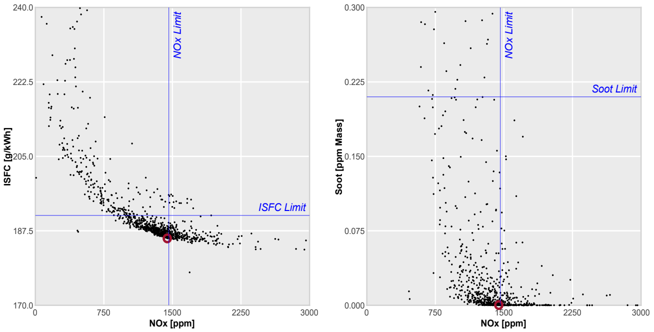


**Figure 3.7:** Pareto's front of ISFC vs. NOx emissions (left) and soot vs. NOx emissions (right).

Based on the best solution of the fitness function and the verification of the reference limits from the Pareto's front, the optimized configuration was compared with the baseline configuration. In Figure 3.8, the differences between the geometries are shown, and Table 3.5 shows the values of the inputs for the optimized case compared to the baseline.

**Table 3.5:** Inputs comparison between baseline and optimized case.

| | Number of holes [-] | Swirl number [-] | Injection pressure (bar) | EGR (%) |
|---|---|---|---|---|
| Baseline | 10 | 2 | 1800 | 0 |
| Optimized | 6 | 1.60 | 1898 | 14.57 |

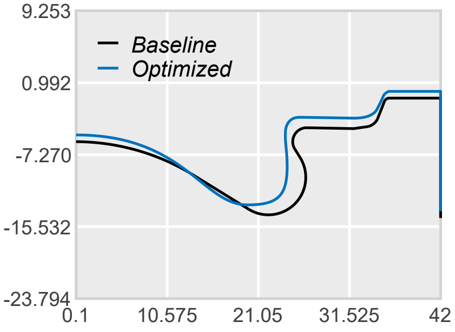


**Figure 3.8:** Difference between baseline and optimized geometries.

Table 3.6 shows the values of the results for the optimized case in comparison with the baseline case, where a reduction in the pollutant emissions and consumption is obtained. From the table, it can be seen that the number of holes decreases, therefore there are more space between sprays, promoting the air entraintment and avoiding the jet-to-jet interaction that enhances the combustion performance. At the same time the injection pressure is slightly increased, which is related to a higher spray momentum and better atomization and evaporation. Moreover, the optimized case uses an EGR ratio of 14.57%, which is a well known and effective practice for reducing NOx emissions. All these trends, corroborates that the methodology is providing reasonable results, delivering a solution that is in agreement with the performance of combustion systems reported in the literature with the added value of finding the solution in less time than other methods [42].

In Figure 3.9, comparison of in-cylinder pressure and rate of heat release traces between the baseline and optimized design is presented. It can be seen that the in-cylinder pressure is similar to the baseline case. However, differences in the HRR are more evident. While the premixed peak is reduced (helping to reduce combustion noise), the burning rate is increased during the non-premixed combustion phase. This could promote a higher temperature of the flame but in the limit to not generate more NOx and providing some thermodynamic advantages.

To better understand these trends, Figure 3.10 shows the temporal evolution of the cylinder mass over three relevant equivalence ratios for the baseline and the optimum case. Specifically, equivalence ratio is into three different bands bounded by 0.55, 1.05 and 1.75. It can be seen that the optimized case configuration increases the mixing rate during the non-premixed combustion, subsequently rising the burning velocity (note that near stoichiometric mixtures completely disappear after 80 CAD aTDC). In contrast, the baseline case is not able to burn all the fuel during the combustion process, keeping some stoichiometric mixture in the cylinder at the exhaust valves opening. This mixing improvement leads to an enhanced combustion, that reduces soot and noise emissions while keeping the NOx emission under control. The improved mixing conditions guaranteed by the optimized bowl correlates well with the shorter combustion duration and the improved performance, as shown previously.

The left-hand side of Figure 3.10 shows the temperature contours of both cases. TDC snapshots show that in the optimized case the jet penetrates faster since the injection pressure is slightly higher and also because it has a bigger nozzle hole diameter. That promotes an improved jet-wall interaction with the piston surface distributing the flame in the combustion chamber. Moreover, the optimized case presents a more homogeneous temperature distribution, thereby lowering the NOx production.

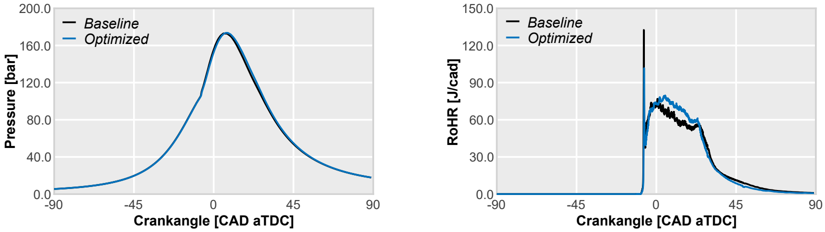


**Figure 3.9:** Comparison between baseline results and optimized case. Left side: In-cylinder pressure evolution versus crank angle. Right side: Estimated Heat Release Rate versus crank angle.

**Table 3.6:** Output comparison between baseline and optimized case.

| | NOx [ppm] | Soot [ppm Mass] | ISFC [g/kWh] |
|---|---|---|---|
| Baseline | 1459.97 | 2.1e-01 | 191.07 |
| Optimized | 1443.35 | 1.9e-04 | 185.63 |

Finally, further studies are required to investigate the implementation to other operating conditions as maximum torque or partial loads in order to analyse if this optimized bowl profile would provide good results in terms of emissions and fuel consumption for those operating conditions. Also, subsequent optimization work is necessary to understand the influence of other input parameters as spray included angle, nozzle tip protrusion, start of injection, among others.

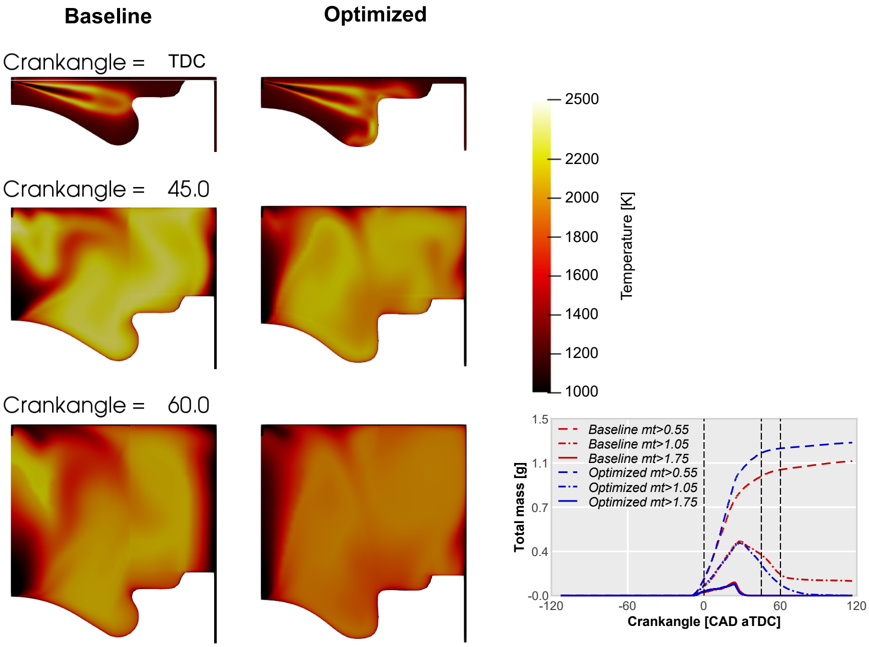


**Figure 3.10:** Comparison between the baseline case and the optimum case. Left side: spatial distribution of temperature in the combustion chamber at three crank angles (Top Dead Center, 45 and 60). Right side: evolution of the normalized mass with equivalence ratio over 0.55 (lean mixture), 1.05 (near to stoichiometric) and 1.75 (rich mixture). The vertical dashed lines correspond to the three instants depicted in the left side.

## 3.5 Conclusions

Several parameters that have a high impact on the combustion system were chosen for the optimization process. These parameters are geometrical variables that controls bowl shape, number of holes, swirl number, injection pressure and EGR rate. Computational tools were linked to obtain an automatic configuration generator allowing to perform a thousand model evaluations automatically. The output targets of this work were pollutant emissions of NOx and soot and specific fuel consumption. These outputs were used to calculate the fitness function that guides the NS algorithm. The obtained results show that coupling CFD with NS can be employed as optimization method to the

CI engines providing a new combustion system that reduces NOx and soot emissions and makes an improvement of 3% of fuel consumption.

Finally, this optimization method can be applied to different engine configurations and combustion concepts requiring just a few changes as the increment of the number of bowl geometrical parameters, the included spray angle or the use of a e-fuel for a full matching, which will be explored in future works.

Chapter 4

# Particle Swarm Grammatical Evolution for Energy Demand Estimation

## 4.1 Introduction

Energy demand estimation from macro-economic variables is one of the most important indicators of a country's economic growth [60]. Financial crises are known to deeply affect energy demand, producing sudden reductions in the demand of the total energy consumed in a country [61], as recently occurred during the hard world-wide crisis of 2008 [62], [63]. In periods of economic bonanza, on the contrary, energy demand usually soars, mainly pushed by the most energy-demanding sectors such as industry and construction [64]. Energy demand is also in close connection with social and environmental issues: currently, about 80% of the global energy demand is covered by non-renewable sources such as coal or petroleum. This dependence on non-renewable resources produces important environmental issues and contributes to global warming and climate change, problems which are currently affecting millions of people, mainly in developing and poor countries [65]. The Paris Agreement, signed in 2015, is a global agreement on the reduction of climate change, which will try to limit the global temperature increase to less than 2 Celsius [66]. Moreover, the effective application of the Paris Agreement should achieve net zero emissions in the second half of this century. This goal will require for sure to deeply reduce emissions by signing countries. None the less, the countries in the Paris Agreement publicly outlined what post-2020 climate actions they intended to take under it, known as Intended Nationally Determined Contributions (INDCs). It is easy to understand that some of these INDCs will be political decisions affecting key macro-economic sectors such as industry and transport, among many others. The idea is to make compatible a correct economic development of a country with the reduction of emissions to fulfil the commitments signed in the Paris Agreement. The correct management of these balance between economic growth and emissions reduction will be, therefore, the guiding line of policy decisions over the next years. In this context, it is clear that obtaining an accurate prediction of medium and long-term energy demand of a country from its macro-economic indicators will be basic for policy and decision makers [67].

Previous works on total energy demand forecasting from macro-economic variables have mainly applied Machine Learning algorithms, including metaheuristics techniques and neural computation approaches. The first work dealing with this problem was [64], where a Genetic Algorithm (GA) was proposed to obtain the parameters of an exponential prediction model for energy demand in Turkey. The model proposed in that work was based on four input macroeconomic variables; Gross Domestic Product or GDP, population, import size and export size. These four predictive variables have been also used in [68], where a multi-layer neural network was applied to solve a problem of total energy demand estimation in Korea. Other approaches such as Particle Swarm Optimization (PSO) have been proposed for problems of total energy demand estimation. In [69] a Particle Swarm approach (PSO) is proposed for this problem in Turkey, including two prediction models, a linear and a quadratic one. Five predictive variables are considered in that work, the ones in [64] plus the country's growth rate in the years of analysis. In [70] a PSO algorithm is also applied, in this case, to a problem of electricity demand in Turkey, considering linear and quadratic models. Ant Colony Optimization (ACO) has been hybridized with PSO in [71], also for a problem of electricity demand estimation in Turkey. Other hybrids based on PSO and GA have been proposed in [72]–[74] applied to problems of energy demand estimation in China. Specifically, in [72] a PSO-GA has been applied to a problem of energy demand forecasting. Three different prediction models, linear, exponential and quadratic are considered, and the predictive variables are economic growth, population, economic structure, urbanization rate, energy structure and energy price. Moreover, there are other techniques in [75]–[78], where different mixtures between clustering and neural networks in order to predict energy consumption have really good performance applied to industry problems. In [74] a problem of energy demand future projection has been considered, and in [73] the energy demand forecasting in the primary sector has been considered. More recently, in [62] a Harmony Search approach with feature selection has been proposed for a problem of energy demand estimation in Spain, and in [63] a Variable Neighbourhood Search (VNS) approach was employed for solving this problem, also comparing the results with that of an Extreme Learning Machine. In [79] new models for energy demand estimation were proposed based on a Grammatical Evolution approach combined with a Differential Evolution algorithm.

In this chapter we explore the performance of a Grammatical Swarm (GS) approach to predict the total energy demand estimation of countries from macroeconomic variables. Grammatical Evolution is a modelling technique based on Genetic Programming. It is divided in two main parts. On one side, there is a set of functions which may explain the process to be modelled. On the other side, there is an optimization engine which finds the best combinations of the set of functions in order to minimize the difference between the real data of the studied process and the data obtained by the model provided by Grammatical Evolution. Grammatical Swarm [80] is an implementation of a Grammatical

Evolution [81], where the optimization engine is a Particle Swarm Optimization (PSO) [1] algorithm. The exploration and intensification phases of the PSO are quite different from the classical grammar evolution based on genetic algorithms, which produces a significant effect in the algorithm's performance. This algorithm has shown to be superior to the classical popular Grammatical Evolution in several problems [80], including neural network training [82].

In our proposal we change the PSO engine algorithm by the NS algorithm developed in Chapter 2. The NS engine searches for the best model according to the training data available. Note that in this algorithm the NS not only selects the best prediction model, including the best predictive variables, but it also tunes its parameters. Thus, the idea of the paper is to use the GS algorithm to explore new models for energy demand estimation. Previous works use existing models, mainly linear and exponential models, though alternative non-linear models have also been considered in the literature [62]. The proposed approach goes a step further, by trying to obtain completely new models using GS algorithm to construct them and tune their parameters, in the same line as [79] did with a classical GE algorithm. In order to show the performance of our proposal, we have tackled two different data sets related to annual energy prediction in Spain and France in the last 30 years. The Grammatical Swarm has shown an excellent performance obtaining good results in terms of mean average relative error for both problems and showing a robust behaviour despite the different distribution of the training data in both problems, escaping from overfitted models.

The remainder of the chapter has been structured as follows: in Section 4.2 we present the problem definition. In Section 4.3, we present the Grammatical Swarm algorithm proposed, including a description of the grammar that has been applied. Section 4.4 details the results obtained by the Grammatical Swarm approach in two real problems of annual energy demand estimation in Spain and France, with comparison of the results to previous algorithms in the literature. In Section 4.5 we discuss the obtained results. Section 4.6 closes the chapter by giving some final conclusions and remarks.

## 4.2 Problem Definition

Let us consider a time series $\mathbf{E} = \{E(t)\}_{t=1}^{n}$ of past energy demands for a given country, with $n$ discrete values corresponding to different years; and a set of $m$ predictive variables $\mathbf{X} = \{X_1(t), \ldots, X_m(t)\}$, with $t = 1, \ldots, n$. A model $\mathcal{M}$ provides an estimation $\hat{\mathbf{E}}$ for $\mathbf{E}$, at a prediction time-horizon $t+1$ respect to the predictive variables. The problem tackled in this paper consists of finding a mathematical model $\mathcal{M}$ for the total energy demand estimation prediction, and tuning its parameters, in such a way that it estimates, as accurately as possible, the real energy demand $\mathbf{E}$ at time $t+1$ (i.e. it obtains $\hat{\mathbf{E}}$), from the set $\mathbf{X} = \{X_1(t), \ldots, X_m(t)\}$.

The quality of a solution $S$ in the GS (a given prediction model $\mathcal{M}$ and its parameters) is evaluated using a given objective function, usually related to the similarity of the model output to the real energy demand values. In this case, it is considered that the root mean square error between the observed values and the predicted ones, which is to be minimized:

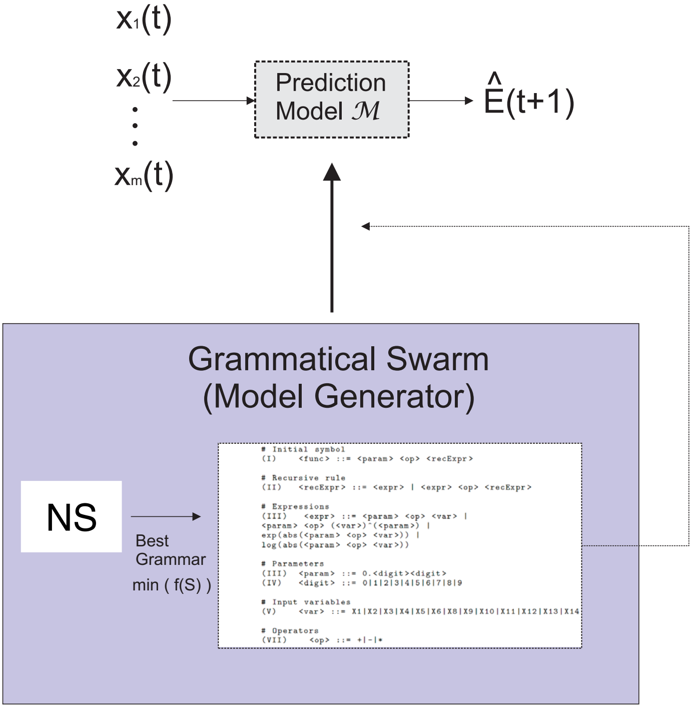
$$f(S) = \sqrt{\frac{1}{n^\star} \sum_{j=1}^{n^\star} \left(E(j) - \hat{E}(j)\right)^2}, \tag{4.1}$$

where $n^*$ is the size of a reduced training sample $(n^\star < n)$.

Figure 4.1 shows an outline of the problem definition, and the role played by the GS in the prediction problem, as generator of the prediction model $\mathcal{M}$.

## 4.3 Proposed Grammatical Swarm

We show the development of a modified GS [80], changing the PSO searching algorithm by the NS and applying it to the problem of annual energy demand estimation from macro-economic variables. In the rest of this section we describe in detail the GS algorithm proposed.


$x_1(t)$
$x_2(t)$
$x_m(t)$
Prediction Model $\mathcal{M}$
$\hat{E}(t+1)$
Grammatical Swarm (Model Generator)
NS
Best Grammar
min ( f(S) )

```
# Initial symbol
(I)    <func> ::= <param> <op> <recExpr>

# Recursive rule
(II)   <recExpr> ::= <expr> | <expr> <op> <recExpr>

# Expressions
(III)   <expr> ::= <param> <op> <var> |
<param> <op> (<var>)^(<param>) |
exp(abs(<param> <op> <var>)) |
log(abs(<param> <op> <var>))

# Parameters
(III)  <param> ::= 0.<digit><digit>
(IV)   <digit> ::= 0|1|2|3|4|5|6|7|8|9

# Input variables
(V)    <var> ::= X1|X2|X3|X4|X5|X6|X8|X9|X10|X11|X12|X13|X14

# Operators
(VII)    <op> ::= +|-|*
```

**Figure 4.1:** Outline of the problem, with GS for evolving the model $\mathcal{M}$.

### 4.3.1 Grammar Production

Grammatical Swarm is a population-based algorithm where each individual of the population is a particle, which represents a solution for a given optimization problem. Specifically, in this case each particle represents a model $\mathcal{M}$ for energy demand estimation, encoded as a vector of integer numbers that will be decoded by means of a given *grammar*, in the same way as performed in

```
# Initial symbol
(I)    <func> ::= <param> <op> <recExpr>

# Recursive rule
(II)   <recExpr> ::= <expr> | <expr> <op> <recExpr>

# Expressions
(III)   <expr> ::= <param> <op> <var> |
<param> <op> (<var>)^(<param>) |
exp(abs(<param> <op> <var>)) |
log(abs(<param> <op> <var>))

# Parameters
(III)  <param> ::= 0.<digit><digit>
(IV)   <digit> ::= 0|1|2|3|4|5|6|7|8|9

# Input variables
(V)    <var> ::= X1|X2|X3|X4|X5|X6|X8|X9|X10|X11|X12|X13|X14

# Operators
(VII)    <op> ::= +|-|*
```

**Figure 4.2:** Grammar applied to the estimation problem for the Spanish data (14 input variables).

Grammatical Evolution (GE) [81]. Figure 4.2 shows the production rules of the grammar that we have considered in this work and Figure 4.3 shows all the possibilities of the grammar expansion. This grammar is taken from [79], in order to make the results of the GS comparable to those described in that paper.

Table 4.1 shows an example of a particle decoding, taking into account the grammar shown in Figure 4.2. Assuming that a particle consists of a vector of integer numbers (named codons) such as (7, 10, 4, 5, 21, 52, 12, 77, 25, 31, 27, 15, 44, 89, 112), shown in the first row of the table, the decoding phase takes the first integer, 7 to decode the `<func>` symbol. The rule in the grammar for that symbol has only one production. Hence, the symbol is translated into `<param>` `<op>` `<recExpr>`. The next integer to be processed is 10, which is used to decode the first non-terminal symbol, `<param>`. Again, the rule for this symbol has only one production, `0.<digit><digit>`, which substitutes the `<param>` symbol. Then, the element 4 is processed to decode the leftmost `<digit>` symbol. The rule for `<digit>` has 10 productions. Given that $4 \, mod \, 10 = 4$, the fifth production of the rule is selected, which corresponds to the value 4.

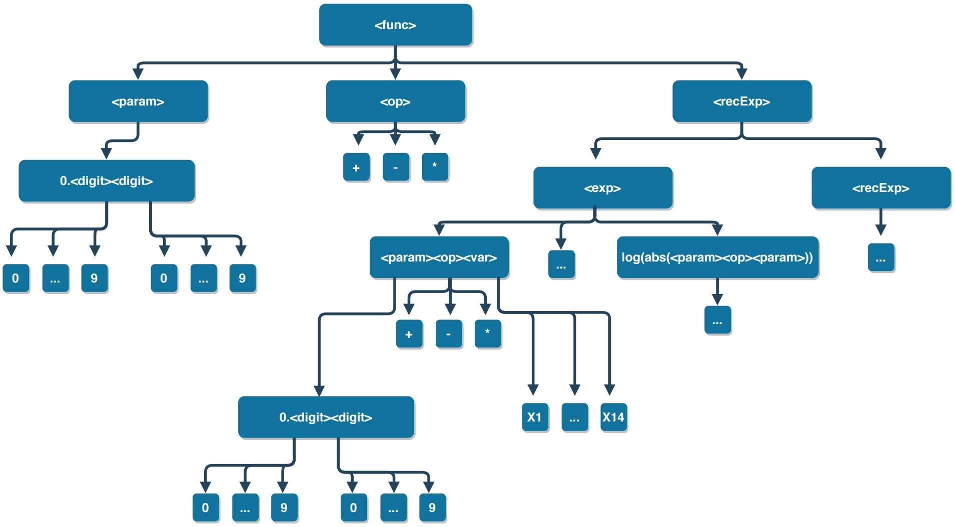


**Figure 4.3:** Possibilities of the grammar expansion.

This process continues using the modulus operation till all the non-terminal symbols are decoded. In the example of the table, three remaining integer values are not used for decoding.

In the classical GE implementation, it is very common to consider *wrapping*, which is the feature allowing the return to the first codon if the final one is reached and there are non-terminal symbols not yet decoded. In the example shown above, no wrapping is considered.

### *4.3.2 Novelty Swarm as search algorithm*

As previously defined, we have a *Grammar Production*, which is the method that allows the automatic creation of mathematical models from the individuals by means of a given grammar, and also a *NS* algorithm, which allows the optimization of the best grammar by means of swarm intelligence. The combination of both techniques produces *Grammatical Swarm* algorithm, according to [80]. The quality of each particle, i.e., the objective function to be minimized, is the root mean square error (RMSE) between the real energy demand

**Table 4.1:** Mapping procedure that converts an individual (second column) into an expression. Each row represents a derivation step. The grammar applied is represented in Figure 4.2.

| Derivation step | Codons left |
|---|---|
| <func> | $7, 10, 4, 5, 21, 52, 12, 77, 25, 31, 27, 15, 44, 89, 112$ |
| <param><op><recExpr> | $10, 4, 5, 21, 52, 12, 77, 25, 31, 27, 15, 44, 89, 112$ |
| 0. <digit><digit><op><recExpr> | $4, 5, 21, 52, 12, 77, 25, 31, 27, 15, 44, 89, 112$ |
| 0.4 <digit><op><recExpr> | $5, 21, 52, 12, 77, 25, 31, 27, 15, 44, 89, 112$ |
| 0.45 <op><recExpr> | $21, 52, 12, 77, 25, 31, 27, 15, 44, 89, 112$ |
| 0.45+ <recExpr> | $52, 12, 77, 25, 31, 27, 15, 44, 89, 112$ |
| 0.45+ <expr> | $12, 77, 25, 31, 27, 15, 44, 89, 112$ |
| 0.45+ <param><op><var> | $77, 25, 31, 27, 15, 44, 89, 112$ |
| $0.45 + 0.$ <digit><digit><op><var> | $25, 31, 27, 15, 44, 89, 112$ |
| $0.45 + 0.5$ <digit><op><var> | $31, 27, 15, 44, 89, 112$ |
| $0.45 + 0.51$ <op><var> | $27, 15, 44, 89, 112$ |
| $0.45 + 0.51*$ <var> | $15, 44, 89, 112$ |
| $0.45 + 0.51 * X1$ | $44, 89, 112$ |

data and the prediction made by the model generated after the decoding of a particle. Figure 4.4 shows a diagram of this process.

## 4.4 Experiments and Results

This section presents the results obtained applying GS for solving two different problems of energy demand estimation (one year ahead prediction), Spain and France. We have gathered data from both countries that involves several macro-economic variables. The energy prediction has been carried out separately for each country, so we have tackled two different experiments.

The data sets have been randomly divided into two different sets of values: the training set, used for the obtention of the mathematical model; and the test set, used to check the quality of the model found. This division is made in order to avoid the effect of *overfitting*, where the model is only able to describe the data that has been trained with, instead of showing a good generalizing performance [83] and to test that the forecast is correct. We have considered the same values for the parameters of the NS in both experiments. These parameters are summarized in Table 4.2.

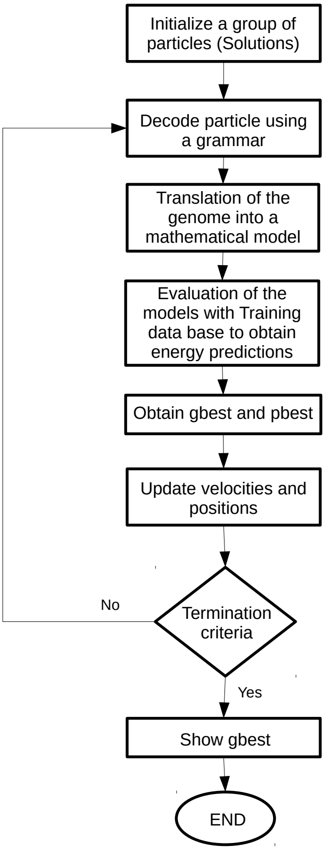

Initialize a group of particles (Solutions)
Decode particle using a grammar
Translation of the genome into a mathematical model
Evaluation of the models with Training data base to obtain energy predictions
Obtain gbest and pbest
Update velocities and positions
Termination criteria
No
Yes
Show gbest
END

**Figure 4.4:** Grammatical Swarm flow diagram.

In order to perform a comparison with the state-of-the-art techniques, we will run two alternative algorithms on the same dataset. These algorithms are a classical implementation of Grammatical Evolution, denoted as *GE*, and a hybridization between GE and Differential Evolution (DE), denoted as *GE+DE*.

| **Parameter** | **Value** |
|---|---:|
| Maximum number of Wraps | 1 |
| Chromosome length | 300 |
| Number of $C$ particles | 60 |
| Number of $E$ particles | 30 |
| $w$ | 0-0.73 |
| $c_1$ | 0-1.5 |
| $c_2$ | 0-1.5 |
| $c_3$ | 0-1.5 |
| Maximum number of iterations | 10000000 |

**Table 4.2:** Parameters of Grammatical Swarm.

Both algorithms follow the implementation described in [79], where *GE+DE* obtained the best results in the data sets studied in that paper.

Table 4.3 shows the parameter configuration for both *GE* and *GE+DE* algorithms in the experiments that we have conducted. Notice that the third column shows the values of the parameters that apply in the DE part of *GE+DE*.

| **Parameter** | **GE** | **GE+DE** | **DE** |
|---|---:|---:|---:|
| Population | 300 | 50 | 75 |
| Generations | 1000 | 40 | 100 |
| Crossover prob. | 0.65 | 0.65 | - |
| Mutation prob. | 0.02 | 0.02 | - |
| Selection | Tournament | Tournament | - |
| Mutation Factor $MF$ | - | - | 0.4717 |
| Recombination Factor $RF$ | - | - | 0.8803 |

**Table 4.3:** Parameters *GE*, *GE+DE* and the DE part of *GE+DE*.

### 4.4.1 Data

*Energy demand estimation in Spain*

For the estimation of the energy demanded in Spain, we have gathered data of fourteen macro-economic variables. The full dataset ranges from 1980 until 2011. The training set is formed by 15 years chosen with a uniform random sampling. Note that the test set is formed by the remaining 16 years. Figure 4.5 shows a graphic representation of the data set of actual demand of energy in Spain, highlighting the training data years (blue points).

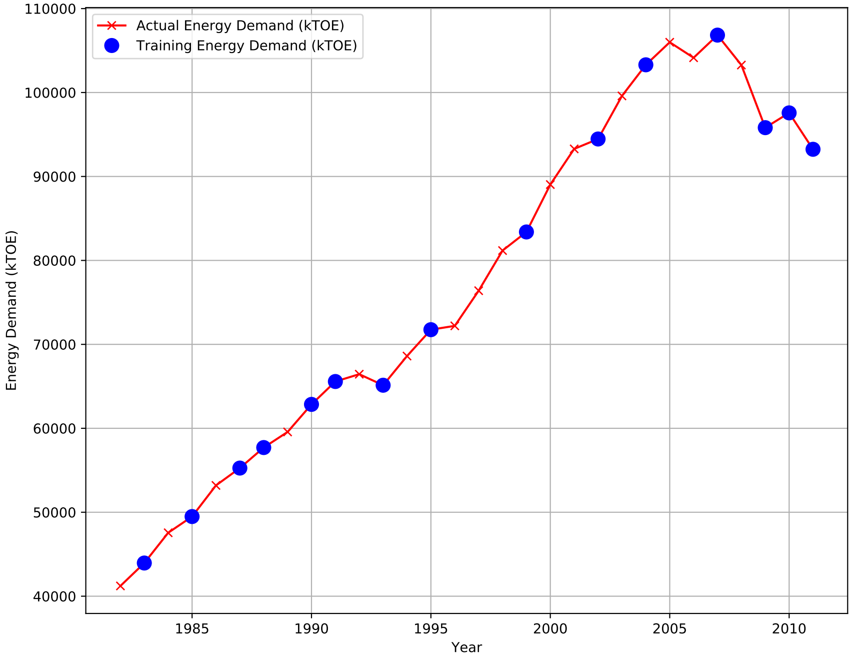


**Figure 4.5:** Spanish data set. Training data are the blue points.

All the variable values have been scaled to the values of 1980, which present the smallest values. The variables that have been taken into account are the following ones:

1. Gross Domestic Product (€)
2. Population
3. Exports (€)
4. Imports (€)
5. Energy production (kTOE)
6. Electricity power transport (kWh)
7. Electricity production (kWh)
8. GDP per unit of energy use (€ per kTOE)
9. Energy imports net (% use)
10. Fossil fuel consumption (% total)
11. Electric power consumption (kWh)
12. $CO_2$ emissions (Mton)
13. Unemployment rate
14. Diesel consumption in road (kTOE)

*Energy demand prediction in France*

Similarly as in the previous works in the literature, we have collected a set of values from nine macro-economic variables from France. In this case, we have data from 1990 until 2016. The training set is formed by 13 years chosen with a uniform random sample. The test set is formed by the remaining 13 years. Figure 4.6 shows a graphic representation of the data set of actual demand of energy in France, highlighting the training data years.

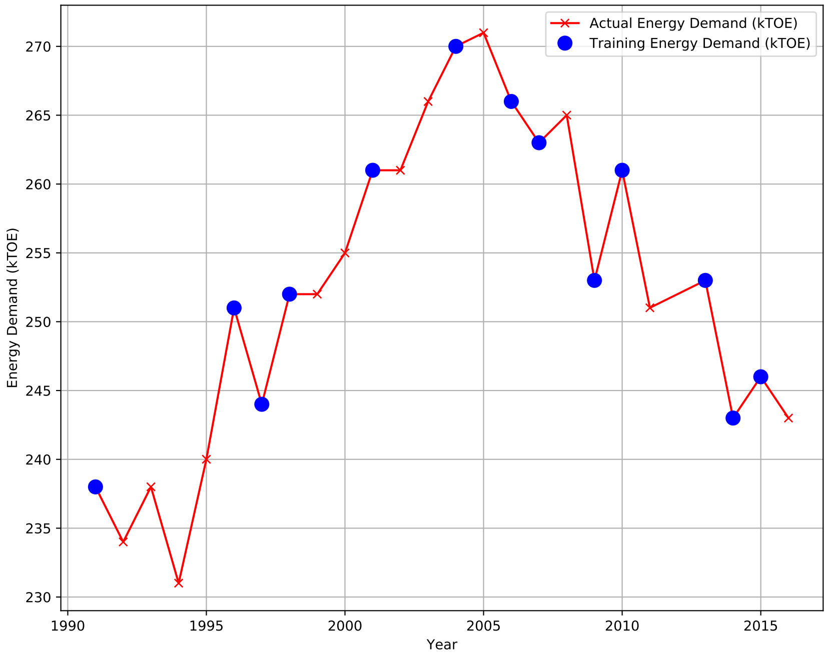


**Figure 4.6:** French data set. Training data are the blue points.

All the variable values have been scaled to the values of 1990, which present the smallest values. The variables that have been taken into account are the following ones:

1. Energy production (kTOE)
2. Population
3. Gross Domestic Product (€)
4. $CO_2$ emissions (Mton)
5. Electricity production (TWh)

6. Electricity consumption (TWh)
7. Unemployment rate
8. Usage of renewable energy (% electicity production)
9. Oil demand (Mtons)

The settings of the grammars are the same as in the experiment for the Spanish data. However, given that the data set of France has only nine variables, the grammar has been modified by removing $X10$ to $X14$ productions from rule `V` (see Figure 4.2).

### *4.4.2 Results*

We have run 32 executions of the optimization process on the training data set, obtaining a set of 32 models for each one of the algorithms and considering as objective function the minimization of the root mean square error (RMSE) between the real data and the data resulting from applying the input parameters to the models obtained during the search process. Each model considers input variables in year $t$ and produces an estimation of energy in the year $t+1$. Both *GE* and *GS* use the grammar shown in Figure 4.2, while *GE+DE* uses the same grammar, but including the parameters to be optimized by the DE part, as described in [79]. For the test set, we have calculated the best average relative error ($\epsilon_R^\star$) and the mean average relative error ($\overline{\epsilon_R}$). In the calculation of the mean average relative error we have removed the worst and the best performing models. Table 4.4 shows the results for the energy demand prediction obtained in the training and test data set by the algorithms under study.

| *Country* | *Spain* | | | | *France* | | | |
|---|---|---|---|---|---|---|---|---|
| | Training | | Test | | Training | | Test | |
| Algorithm | $\epsilon_R^\star$ | $\overline{\epsilon_R}$ | $\epsilon_R^\star$ | $\overline{\epsilon_R}$ | $\epsilon_R^\star$ | $\overline{\epsilon_R}$ | $\epsilon_R^\star$ | $\overline{\epsilon_R}$ |
| GE | 1.46 | 3.48 | 6.53 | 53.25 | 0.78 | 1.44 | 1.57 | 3.48 |
| GE+DE | 1.66 | 2.27 | 11.25 | 59.39 | 0.81 | 1.03 | 1.66 | 2.68 |
| GS | 1.52 | 2.37 | 1.75 | 3.27 | 1.33 | 1.40 | 2.10 | 2.74 |

**Table 4.4:** Results of energy prediction estimations in for training and test data sets in both studied countries. The lower are the error, the better. A high difference between training and test errors means over fitting.

## 4.5 Discussion

As can be seen in Table 4.4, the results obtained by GS on the test data of Spain, outperform both *GE* and *GE+DE* methods, obtaining a reduction of 50% in relation to the mean average relative error of *GE*, which is the second best performer. On the other hand, in the test data of France, the behavior of both GS and *GE+DE* are comparable. However, the results of *GE+DE* are slightly better.

We observe that the behavior of the swarm intelligence is able to find out good models in both datasets. On the other hand, the *GE+DE* and *GE* methods are sensitive to the data that are selected for the test. As seen in Figures 4.5 and 4.6, the training data from Spain are more consecutively selected, which may have caused an overfitting of the *GE+DE* and *GE* method, which obtained good results for the training data. The selected data from France are more sparsely distributed, and *GE+DE* is able to capture the behavior of the data. GS, on the contrary, performs very similarly with both data sets, showing more robustness than the GE methods and obtaining good results, which are also the best ones in the experiment with the data from Spain. This hypothesis can be confirmed by comparing the training data from the same problem in [79], where data is more consecutively distributed and both algorithms (GE, GE+DE) obtain much better results in the test set.

Figure 4.7 shows a graphic representation of the comparison between the estimation and the actual demand of energy in Spain for the test data set, obtained with the best model from the GS executions. As it can be seen, the model is able to successfully capture the trends of the actual data.

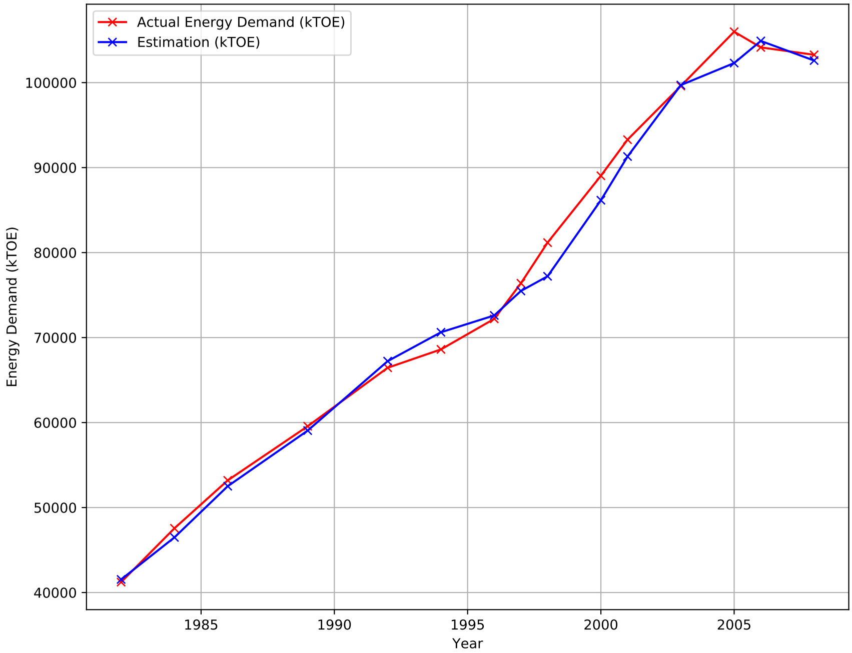


**Figure 4.7:** Actual energy demand compared to the prediction made by GS for the test data set of Spain.

As in the previous experiment, Figure 4.8 shows the comparison between the prediction of the best model obtained by GS and the actual demand of energy in France for the test data set. Again, the behaviour of the GS model is comparable to the actual data.

GS technique has shown to be more robust than DE or GE+DE. The use of this algorithm is recommended when data provided by the system to be modelled (i.e. energy demand) is not uniform.

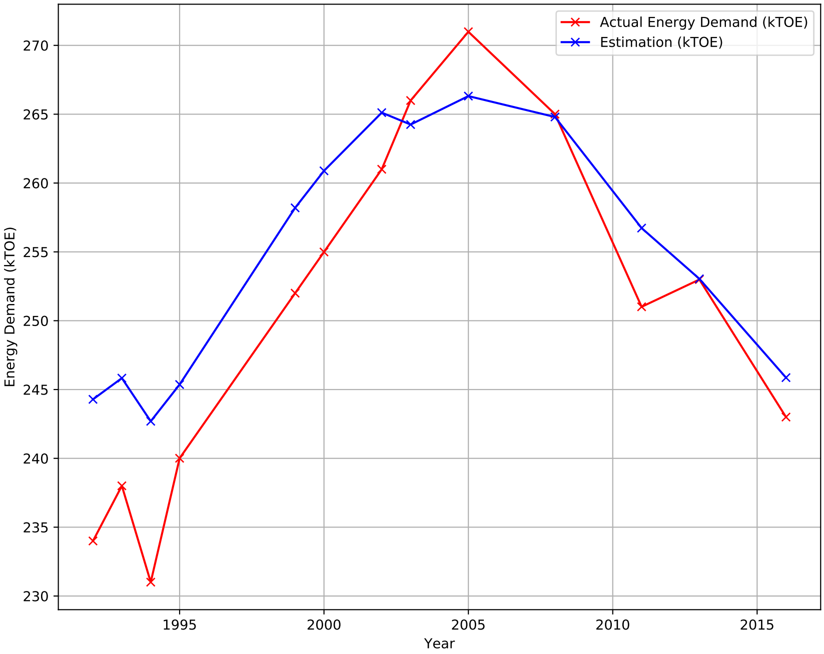


**Figure 4.8:** Actual energy demand compared to the prediction made by GS for the test data set of France.

In future works, different energy fields should be tested. Electricity consumption from the electric network could be an interesting problem to be proposed, since the prediction of future power demand is crucial for a proper system functioning.

## 4.6 Conclusions

Total energy demand estimation from macro-economic variables is an important problem, deeply related to countries' economic growth measurement, and with consequences in policies for energy efficiency and environment protection. As such, research in developing accurate new models for energy demand estimation has been intense in the last years. In this paper, we have evaluated the performance of a Grammatical Swarm algorithm for generating new energy demand estimation models from macro-economic predictors. The proposed algorithm is based on a predefined grammar, able to decode different prediction models evolved with a particle swarm approach. The algorithm is also able to adjust the models to a series of real energy demand data. Once trained, the developed model should be able to reproduce with a good accuracy the series of energy demand from the macro-economic variables, and also estimate new values for the future if needed. We have tested the proposed Grammatical Swarm algorithm in two real problems of one-year ahead energy demand estimation from macro-economic variables, for the case of Spain and France. In both cases the proposed approach is able to obtain models with an excellent performance in terms of prediction error, showing robustness with different distributions of training data, and improving the results of previous algorithms in one of the problems. We can also conclude that the behaviour of NS can improve other well known algorithms of the literature.

### Chapter published

The results of this chapter have been published in [84].

Chapter 5

# Mathematical modelling of the dynamics of the bladder cancer and the immune response applied to a patient: Evolution and short-term prediction

## 5.1 Introduction

Bladder cancer is one of the most common malignant diseases in the urinary system and a highly aggressive neoplasm [85], [86]. Around 80% of patients diagnosed with this cancer present a non invasive carcinoma that can be handled via a Trans Urethral Resection (TUR), a surgical endoscopic procedure to remove the tumour from the inner part of the bladder [87].

In this procedure, a catheter is introduced by the urethra until arriving to the bladder. With a camera attached to the top of the catheter, the tumour is found with a sight inspection and then removed using a scalper. There is no need of cutting into the abdomen, so it is considered a minor surgery.

The bladder cancer is characterized by recursiveness for more than a half of the patients: the tumour appears again after a while and may progress to become a muscle invasive cancer. It is then when the treatments become more aggressive, including the removal of the bladder to avoid the spread of the cancer to other parts of the body [88].

The prognosis of bladder cancer is not generally good and after the TUR, the typical treatment consist of instillations of Bacillus Calmette-Guérin (BCG) into the bladder, with the aim that the BCG stimulates the patient's immune response against the cancer and then, the cancer cells may be eliminated [89].

When BCG instillation is made, the BCG cells attach to the urotherial cells and they get internalized by the bladder cancer cells. Because of this, the bladder cancer cells release substances such as cytokines and antigen presentation, which promotes the immune cells recruitment and suppression activity. This is described visually in figure 5.1.

Despite the advancements in molecular biology techniques [90] and the knowledge of the cancer stages and the treatments [91], [92], it is not well known the mechanisms of bladder cancer evolution in particular patients, why and when some of them heal and others have recurrences. In fact, the treatment protocols have not changed in the last 25 years.

With the aim at providing some light to the problem of customizing the evolution of bladder cancer and its treatment, here, we propose a mathematical model to study the evolution of the bladder cancer of a patient where TURs and the administration of BCG have been considered in the evolution. Also, we consider in the model the immune response caused by the apparition and growth of the tumour and the interaction of cancer cells with the immune system. To our knowledge, only in the paper written by Bunimovich-Mendrazitsky

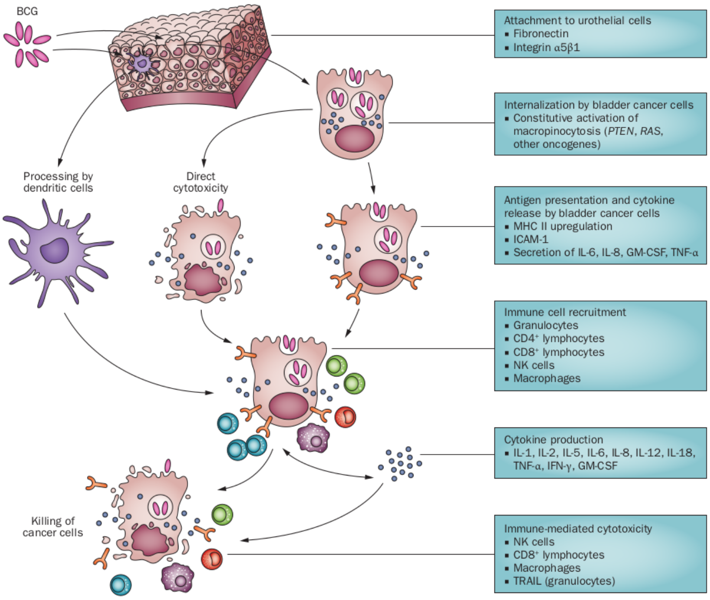

BCG
Processing by dendritic cells
Direct cytotoxicity
Killing of cancer cells
Attachment to urothelial cells
▪ Fibronectin
▪ Integrin α5β1
Internalization by bladder cancer cells
▪ Constitutive activation of macropinocytosis (*PTEN*, *RAS*, other oncogenes)
Antigen presentation and cytokine release by bladder cancer cells
▪ MHC II upregulation
▪ ICAM-1
▪ Secretion of IL-6, IL-8, GM-CSF, TNF-α
Immune cell recruitment
▪ Granulocytes
▪ CD4⁺ lymphocytes
▪ CD8⁺ lymphocytes
▪ NK cells
▪ Macrophages
Cytokine production
▪ IL-1, IL-2, IL-5, IL-6, IL-8, IL-12, IL-18, TNF-α, IFN-γ, GM-CSF
Immune-mediated cytotoxicity
▪ NK cells
▪ CD8⁺ lymphocytes
▪ Macrophages
▪ TRAIL (granulocytes)

**Figure 5.1:** Interaction between urotherial cells, tumoural urotherial cells and BCG [89].

et al. [93], the authors introduce a model of this kind using average parameter values obtained from several in silico patients data.

At this point, we must say that, even though the antecedent of Bunimovich-Mendrazitsky et al. [93] has been very valuable to develop our model, we had to design our own model. This is because our model has to be applied with the available data of our patients, given that the available data are those the pathologists consider relevant to be analysed in order to determine the evolution of the sickness rather than the data we could consider as necessary for the model. Also, our model should describe properly the clinic practice in the Hospital Universitari i Politècnic La Fe [94] in Valencia, Spain, where we collaborate with the doctors of the Urology Department and the Pathology Department. Furthermore, we have to introduce into the model new recent findings.

Thus, we are going to propose a model to describe the evolution of the bladder cancer, we will use data of a patient to determine the model parameters that describe the known evolution and use these calibrated model parameters to predict the evolution of the bladder cancer in the short-term. Due to the lack of data, we will perform the evolution taking into account three possible scenarios: favourable, neutral and unfavourable.

The chapter is organized as follows. In Section 5.2, we will introduce the data available collected from the Hospital Universitari i Politècnic La Fe. In Section 5.3 we build the model describing the evolution of the tumour size and the immune response in stages determined by the medical protocols. In Section 5.4, we find the model parameter values that allow the model to describe the bladder cancer evolution until now and we show the result of the calibration. In Section 5.5, we predict the evolution of the size of the tumour over the next few months in three scenarios: favourable, neutral and unfavourable. Also, we describe the patient's follow-up actions done. Finally, in Section 5.6, we present some conclusions of the work.

## 5.2 Available data

Typically, in the Spanish Region of Valencia, a patient goes to the family doctor or the emergencies of the hospital when he/she starts to urinate blood (haematuria). Then, the patient is addressed to the urologist, who, using ultrasound images or a cystoscopy, diagnoses if there is a tumour inside the bladder. If so, the patient is treated by the doctors in the Urology Department. In our case, we are working with doctors belonging to the Hospital Universitari i Politècnic La Fe. The doctors have collected biological tissues of patients suffering pT1G3 bladder cancer, a non muscle-invasive stage of the bladder cancer. The usual protocol is to perform a TUR to the patient, then instillations of BCG and cytologies and cystoscopies for follow-up control and revisions. In case of recurrence, the doctors repeat the cycle described above until the patient is completely healthy or the stage where the cancer increases, invades the bladder muscle layer and more aggressive treatments have to be applied.

It is a retrospective study, and, among the patients, we have chosen from the Hospital Universitari i Politècnic La Fe database the only one with the complete medical history for our purpose. We are going to call this patient Patient X, and the available data are collected and summarized in Table 5.1.

| Date | Medical procedure | Diameter of the tumour | Inflammatory cells per field |
|---|---|---|---|
| 01/Mar/2012 | Ultrasound | 3-5 mm | - |
| 14/Jun/2012 | TUR | 25 mm | 260 |
| 15/Feb/2015 | Cystoscopy | 1-2 mm | - |
| 28/Apr/2015 | TUR | 5 mm | 515 |
| 30/Jan/2017 | Cystoscopy | 20 mm | - |
| 14/Mar/2017 | TUR | 30-35 mm | 508 |

**Table 5.1:** Data corresponding to Patient X, who was diagnosed in the first of March 2012. Since then, the Patient X has suffered three TURs and has been treated with three BCG instillation sessions, each one after each TUR. The last column shows the number of inflammatory cells per field counted by the pathologist after every TUR.

Apart from the information provided in Table 5.1, instillation dates are also known. BCG instillations were started the days 10/Aug/2012, 26/May/2015 and 27/Apr/2017, after each TUR. Doses introduced in the bladder consisted of 81 ml of BCG colonies instilled once per week during six weeks.

## 5.3 Model building

To design a reliable model, we are going to take into account the available data and their magnitudes, describing properly the clinic practice in the Hospital Universitari i Politècnic La Fe, and including recent discoveries about the interaction between tumour cells and the immune system [89], [95], [96].

The model is going to be built in two stages. It is natural, because from the tumour detection and diagnosis to the first TUR, there are only two interacting actors: the tumour cells and the immune system. However, after the first TUR and the application of the BCG treatment to the patient, tumour cells, infected tumour cells, BCG and the immune system, all of them, interact. This latter interaction will be repeated every time a new TUR is performed and BCG administered.

### 5.3.1 Modelling the first stage

As the tumour size is measured by its diameter, we define $T(t)$ as the diameter of the tumour in the day $t$. Thus, we model the tumour growth using the classical exponential growth difference equation given by

$$T(t+1) = T(t) + kT(t), \tag{5.1}$$

where $k$ is the tumour growth rate and $T(0)$ is the initial diameter (see Table 5.1), being $t = 0$ the day 01/Mar/2012.

Together with the tumour, there are other components which interact in the bladder cancer. The most important one is the immune system. In the study developed by Hanahan and Weinberg [95][p. 660-661], the authors say "... In recent years, however, an increasing body of evidence, both from genetically engineered mice and from clinical epidemiology, suggests that the *immune system operates as a significant barrier to tumour formation and progression* ...". Also, "... it was observed that tumours arose more frequently and/or grew more rapidly in the immunodeficient mice ...". So, we can conclude that the immune system is the mechanism we have to face the tumour cells. The immune system is formed by the inflammatory cells, so this is what we are going to model.

We denote $I(t)$ as the inflammatory cells. However, we are going to use the measurement that the doctors use in their practice. They make the average of five counting from different parts of the microenviroment of the tumour that they observe through a 40x microscopy lens. This is what we call "average inflammatory cells per microscopy field". The general system immune response is described by

- the number $I(t)$ may increase at rate $b_1$ because of the presence of tumour cells, modelled by the term $b_1T(t)$;
- the encounters of inflammatory cells with the tumour cells provokes the inactivation of the inflammatory cells at rate $\lambda_1$, modelled by the term

$\lambda_1 I(t)T(t)$ and the inflammatory cells kill the tumour cells and reduce the tumour size at rate $\lambda_2$, modelled by the term $\lambda_2 I(t)T(t)$.

Parameter $\lambda_2$ measures the effectiveness of the immune system to kill tumour cells and reduce the tumour size [89] (Figure 5.2). Also, the number of inflammatory cells decreases because their natural death rate, modelled by the term $d_1 I(t)$.

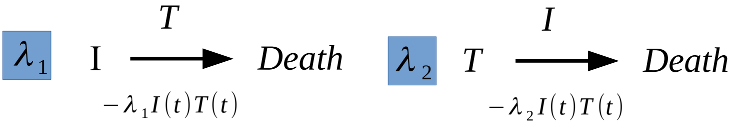


**Figure 5.2:** Interaction between inflammatory cells $I(t)$ and tumour diameter $T(t)$, and their corresponding rates $\lambda_1$ and $\lambda_2$. On the left of the arrows, the affected group, on the top of the arrows, the originator group. On the left of the figure, we see the term that models the death of inflammatory cells because the presence of tumour cells. This term produces a reduction in $I(t)$. On the right of the figure, the term that models the death of tumour cells because the action of the inflammatory cells. This term leads to a reduction in $T(t)$.

Therefore, the dynamics of the tumour growth and the inflammatory cells and their interaction in the first stage is given by the following system of difference equations

$$\begin{aligned} T(t+1) &= T(t) + kT(t) - \lambda_2 I(t)T(t), \\ I(t+1) &= I(t) - d_1 I(t) + b_1 T(t) - \lambda_1 T(t)I(t), \end{aligned} \tag{5.2}$$

where the time $t$ is in days.

### *5.3.2 Modelling the second stage*

This second stage starts when TUR removes the tumours and, after a while, BCG treatment is administered. The tumour is still growing because, although the doctors do their best, the complete removal is almost impossible ("seed and soil" theory) [97]. Apart from the tumour cells and the inflammatory cells, in this stage, BCG appears via instillations and the effect of the BCG is to infect tumour cells to get marked facilitating the attack of the inflammatory cells in a more aggressive way. This way, the action of the inflammatory cells is much more effective than without BCG [89] and this is called specific immune response. Thus, infected tumour cells have to be considered in this stage too.

Then, let us denote by $B(t)$ the millilitres of BCG inside the bladder, $b(t)$ the millilitres of BCG injected in one or several doses and $T_i(t)$ as the diameter of the tumour cells infected by BCG, at the time instant (day) $t$.

Also, the term $d_2B(t)$ describes the natural disappearance of BCG by urination, being $d_2$ its rate. As we explained above, inflammatory cells increase because of tumour cells, and now, also because the infected ones. It is modelled by the term $b_2T_i(t)$.

Now, we model the specific immune response, because the way the inflammatory cells act against the infected and non-infected tumour cells is different [89]. Then,

- the death of the inflammatory cells because the infected tumour cells is modelled by the term $\tau_1T_i(t)I(t)$ [95];
- when a bacillus encounters a tumour cell, the tumour cell becomes infected and the bacillus gets attached to the tumour cell and it is modelled by the term $\tau_2T(t)B(t)$. This term affects decreasing $T(t)$, increasing $T_i(t)$ and decreasing $B(t)$ [89] (Figure 5.3).

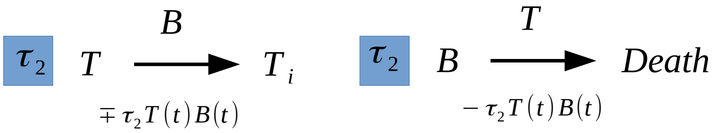


**Figure 5.3:** Interaction between BCG $B(t)$ and tumour cells $T(t)$, with the corresponding rate $\tau_2$. On the left of the arrows, the affected group, on the top of the arrows, the originator group. This interaction increases the tumour infected cells and decreases the tumour cells and the BCG [89].

Now, the remainder interactions are (Figure 5.4):

- the death of BCG because the inflammatory cells attack BCG, and it is modelled by $\tau_3I(t)B(t)$ [96];

- the death of infected tumour cells and the reduction of its size because of inflammatory cells is modelled by $\tau_4 I(t)T_i(t)$ [89];
- the growth of inflammatory cells because of BCG is modelled by $\tau_5 B(t)I(t)$ [96].

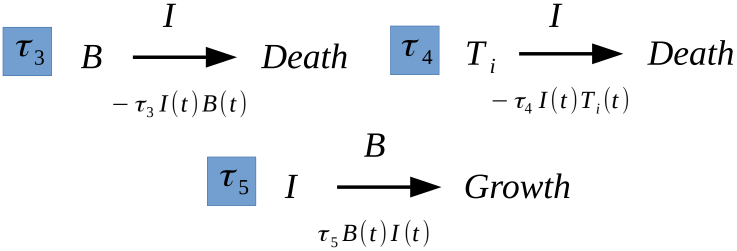


**Figure 5.4:** Interaction between BCG $B(t)$, inflammatory cells $I(t)$, tumour cells $T(t)$ and tumour cells infected by BCG $T_i(t)$, and their corresponding rates $\tau_3$, $\tau_4$, $\tau_5$. On the left of the arrows, the affected group, on the top of the arrows, the originator group.

It is important to remark that the effectiveness of the BCG treatment is based on the fact that parameter $\tau_4$ is much greater than $\lambda_2$ because BCG infects the tumour cells and facilitates the detection and destruction by the inflammatory cells (specific immune response). All the above model parameters and their description have been summarized in Table 5.2.

Then, gathering all the above terms, the evolution of the dynamics of the size of tumour, inflammatory cells and BCG can be modelled using the following system of difference equations (time $t$ in days):

$$\begin{array}{rcl} T(t+1) & = & T(t) + kT(t) - \lambda_2 I(t)T(t) - \tau_2 B(t)T(t), \\ I(t+1) & = & I(t) - d_1 I(t) + b_1 T(t) - \lambda_1 T(t)I(t) + b_2 T_i(t) \\ & & -\tau_1 T_i(t)I(t) + \tau_5 B(t)I(t), \\ T_i(t+1) & = & T_i(t) + \tau_2 T(t)B(t) - \tau_4 I(t)T_i(t), \\ B(t+1) & = & B(t) - d_2 B(t) - \tau_2 T(t)B(t) - \tau_3 I(t)B(t) + b(t). \end{array} \tag{5.3}$$

At this point, we must say that model parameters $\lambda_2$, $\tau_4$ and $\tau_5$ are going to be variable over the stages. They are responsible of the immune response

| Parameter | Description | Term |
|---|---|---|
| $k$ | Tumour size growing rate. | $kT(t)$ |
| $\lambda_1$ | Inflammatory cells death rate because of uninfected tumour cells. | $-\lambda_1 I(t)T(t)$ |
| $\lambda_2$ | Tumour cell death rate because of inflammatory cells (effectiveness). | $-\lambda_2 I(t)T(t)$ |
| $d_1$ | Inflammatory cells natural death rate. | $-d_1 I(t)$ |
| $b_1$ | Production rate of inflammatory cells because of the presence of the tumour cells. | $b_1 T(t)$ |
| $d_2$ | BCG natural dissapearance by urination. | $-d_2 B(t)$ |
| $b_2$ | Production rate of inflammatory cells because of the presence of the infected tumour cells. | $b_2 T_i(t)$ |
| $\tau_1$ | Death of inflammatory cells because the infected tumour cells. | $-\tau_1 T_i(t)I(t)$ |
| $\tau_2$ | Infection of tumour cells and its effect on the reduction of the BCG. | $\tau_2 T(t)B(t)$ |
| $\tau_3$ | BCG death rate because of inflammatory cells. | $-\tau_3 I(t)T_i(t)$ |
| $\tau_4$ | Infected tumour cells death rate because of inflammatory cells (effectiveness of BCG). | $-\tau_4 B(t)I(t)$ |
| $\tau_5$ | Inflammatory cells growth rate because of BCG. | $\tau_5 B(t)IE(t)$ |

**Table 5.2:** Model parameters, description and modelling term. Positive terms mean growth of cells and negative terms mean death or removal of cells. Above the horizontal line, the parameters involved in the first stage. Below the horizontal line, the parameters related to the BCG administration.

before and after the administration of the BCG. Then, after the first BCG administration, the immune system is highly stimulated and the values of the model parameters $\lambda_2$, $\tau_4$ and $\tau_5$ reach their highest values. If more BCG administration is necessary, the stimulation of the immune response is not as well as the first time and the model parameter values $\lambda_2$, $\tau_4$ and $\tau_5$ use to decrease as the stages goes on. This is natural taking into account the behaviour of the immune response (sudden increase because of BCG and slow decrease) and noting that the output returned by the model of the immune response cannot overpass certain values in order to keep the model credibility (around 800).

## 5.4 Model calibration

Model calibration has been made in stages, because it is not a continuous process due to the sudden extraction of the tumour with TURs, and a calibration period is established for each surgery. For Patient X there are three different stages, separated by TURs:

1. the first stage starts the 01/Mar/2012 and lasts until 14/Jun/2012;
2. the second stage starts the 15/Jun/2012 and lasts until 28/Apr/2015;
3. the third stage from the 29/Apr/2015 to 14/Mar/2017.

Predictions will be made for dates after the 14/Mar/2017, considering the instillations administered 27/Apr/2017 and during 6 weeks., once per week.

In the first stage, the model is given by the system (5.2). The system (5.3) will be used for the second and third stage, and also for predictions.

### *5.4.1 Calibration of the first stage*

Taking into account the lack of data because of the nature of its measurement and the large number of model parameters, we have to consider the reduction of the parameters space in order to guarantee reliable calibrations. Looking at the data about the size of the tumour in the Table 5.1, considering only the growth tumour model (5.1) and performing a calibration of the parameter $k$ in the three stages, we can obtain that parameter $k$ cannot be greater that 0.0204. $k$ is the growth tumour rate and it is much greater than the usual growth rates for healthy cells, by definition of cancer. Therefore, the parameters related with the growth and death of cells, $b_1, b_2, d_1, d_2$, should be of magnitudes about $10^{-3}$ or less. Similar reasoning can be done for the remainder parameters insofar they are involved in the growth or death of the cells and BCG.

Furthermore, Table 5.1 shows small changes in the inflammatory cells per field in second and third TURs, where the immune system has been changed because of the instillations of BCG. We are going to assume the same behavior before the first TUR, what means that initially, we expect around 260 inflammatory cells the 01/Mar/2012.

As we mentioned above, although the doctors do their best, the bladder is not completely free of tumour cells after the TUR. Therefore, we are going to consider that the diameter of the tumour after a TUR will be determined by the model after the calibration. This size is small enough to be considered the bladder gets "clean" after a TUR and permits the tumour keep growing, which we know that happens because of recurrences.

Now, in order to find the model parameters that make the model to be as close as possible to the data of Table 5.1 in the corresponding time instants, we define the following fitting function $F_1$:

**INPUT:** Model parameter values $(k, \lambda_1, \lambda_2, d_1, b_1)$;

**Step 1.** Substitute the model parameter values into the model (5.2);

**Step 2.** Run the model and retrieve the model output for tumour size and inflammatory cells per field in the same time instants as those in Table 5.1;

**Step 3.** Calculate the root mean square between the model output retrieved in Step 2 and the data in Table 5.1.

For model calibration, we minimize the function $F_1$ in this stage using the Novelty Swarm algorithm developed in Chapter 2, and the model parameter values are given in Table 5.3.

| Parameters | Value |
|---|---|
| $k$ | 0.0184 |
| $\lambda_1$ | $8.1186 \times 10^{-6}$ |
| $\lambda_2$ | $6.8426 \times 10^{-6}$ |
| $d_1$ | $1.6 \times 10^{-4}$ |
| $b_1$ | $9.08 \times 10^{-5}$ |

**Table 5.3:** Model parameter values calibrated for Patient X in the first stage, until the first TUR.

The calibration procedure allowed us to determine more precisely the initial conditions $T(0) = 4.64$ mm and $I(0) = 265$.

### 5.4.2 Calibration of the remainder stages

Once the first stage has been calibrated, the parameters $k, \lambda_1, d_1, b_1$ are known and only the remainder parameters have to be calibrated in the second stage. As we mentioned above, the model parameters $\lambda_2$, $\tau_4$ and $\tau_5$ are going to be variable over the stages, meanwhile the remainder model parameters will remain constant over the stages. Now, for the calibration in the second stage, we propose the following fitting function $F_2$:

**INPUT:** Model parameter values $(\lambda_2, d_2, b_2, \tau_1, \tau_2, \tau_3, \tau_4, \tau_5)$. The model parameters $k, \lambda_1, d_1$ and $b_1$ are those with the values given in Table 5.3;

**Step 1.** Substitute the model parameter values into the model (5.3);

**Step 2.** Run the model and retrieve the model output for tumour size and inflammatory cells per field in the same time instants as those in Table 5.1;

**Step 3.** Calculate the root mean square between the model output retrieved in Step 2 and the data in Table 5.1.

For model calibration in the second stage, we minimize the function $F_2$ using NS algorithm, and the model parameter values are given in Table 5.4.

Now, for model calibration in the third stage, the model parameter values $k, \lambda_1, d_1, b_1, d_2, b_2, \tau_1, \tau_2, \tau_3$ have been calculated and only $\lambda_2, \tau_4, \tau_5$ have to be calibrated again, using the fitting function $F_2$ where, now, the unknown model parameters are only $\lambda_2, \tau_4, \tau_5$. Then, using NS algorithm, the calibrated model parameters are given in Table 5.5. It can be seen that the model parameter values satisfy the restrictions stated through this section.

In Figure 5.5, we can see the result of the calibration, that is, the evolution of the tumour size and the immune system development. In the lower figure (tumour size) the sudden drops correspond to TURs, when the tumour is

| Parameters | Value |
|---|---|
| $\lambda_2$ | $2.393 \times 10^{-5}$ |
| $d_2$ | $0.8864$ |
| $b_2$ | $5.89 \times 10^{-3}$ |
| $\tau_1$ | $2.8689 \times 10^{-5}$ |
| $\tau_2$ | $2.1 \times 10^{-3}$ |
| $\tau_3$ | $2.207 \times 10^{-4}$ |
| $\tau_4$ | $8.59 \times 10^{-3}$ |
| $\tau_5$ | $1.85 \times 10^{-3}$ |

**Table 5.4:** Model parameter values calibrated for Patient X in the second stage, after the first TUR until the second TUR.

| Parameters | Value |
|---|---|
| $\lambda_2$ | $1.205 \times 10^{-5}$ |
| $\tau_4$ | $3.9132 \times 10^{-4}$ |
| $\tau_5$ | $2.43878 \times 10^{-4}$ |

**Table 5.5:** Model parameter values calibrated for Patient X in the third stage, after the second TUR.

removed. Then, it starts to grow again. Respect to the upper figure (immune response), we can see a quick increasing after TUR corresponding to the BCG instillations and how they influence the immune response. The saw teeth appearing when BCG is administered are due to the weekly doses, once per week, of BCG instilled.

## 5.5 Predictions and patient's follow-up

Due to the uncertainty in the future evolution, in order to predict the dynamics of the bladder cancer for Patient X, three different scenarios have been considered. In the favourable scenario the response of the immune system takes the maximum possible values that allow the reliability of the values corresponding to the immune system response, usually less than the values as after the first TUR (second stage). The neutral scenario consists of the same response of the immune system as in the third stage, that is to say, the immune system response remains constant. The unfavourable scenario is based on the fact

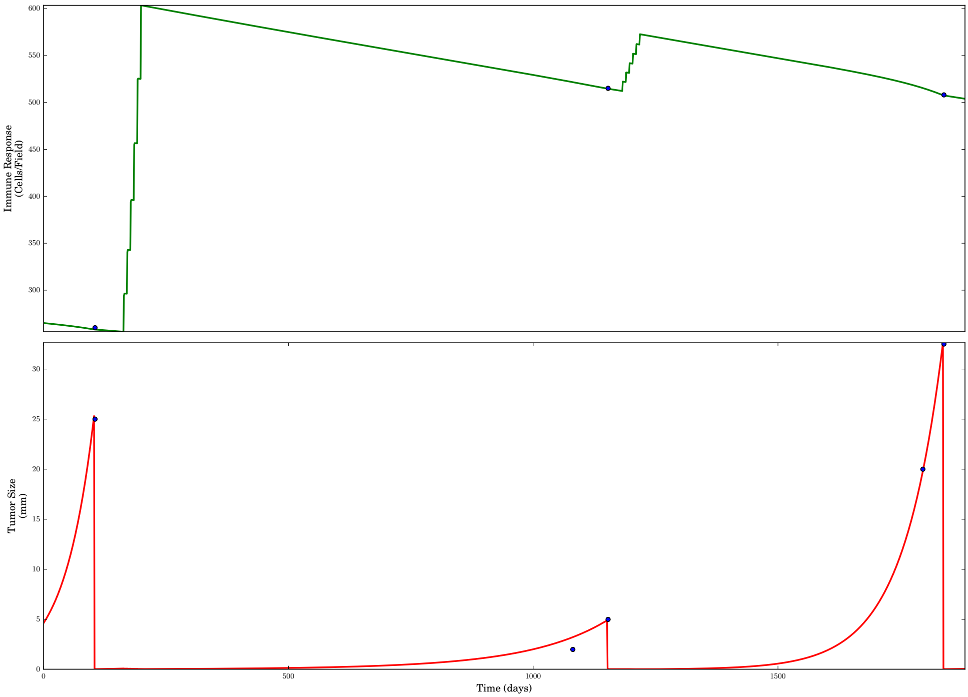


**Figure 5.5:** Here we show the graphs with the evolution of the immune response (upper) and the tumour size (lower) with the parameters obtained in the model calibration. The points represent the available data in Table 5.1. The sudden drops in the lower graph correspond to the TURs, when tumours are removed. The sudden increasings in the upper graph correspond to the administration of the BCG. Time $t = 0$ corresponds to 01/Mar/2012 and time step is a day.

that the response of the immune system decreases in the same proportion as it decreased from the second to the third stage. In Table 5.6 the values of the parameters corresponding to the described scenarios are shown.

In Figures 5.6, 5.7 and 5.8, the prediction in the different scenarios are shown. The prediction starts in the dashed black vertical line. In the favourable scenario, the BCG treatment is successful and the tumour does not seem to grow in the prediction time interval until 15/Sep/2018, after instillation treatment is finished. In the neutral scenario, bladder cancer recurrence with a tumour of size 7-8 mm is predicted for the 15/Sep/2018. In the unfavourable scenario,

| **Parameters** | **Favourable scenario** | **Neutral scenario** | **Unfavourable scenario** |
|---|---|---|---|
| $\lambda_2$ | $2.393 \times 10^{-5}$ | $1.205 \times 10^{-5}$ | $5.9822 \times 10^{-6}$ |
| $\tau_5$ | $8.59 \times 10^{-3}$ | $3.9132 \times 10^{-4}$ | $3.7349 \times 10^{-4}$ |
| $\tau_6$ | $3.692 \times 10^{-4}$ | $2.4387 \times 10^{-4}$ | $2.1166 \times 10^{-4}$ |
| tumour size at 15/Sep/2018 | Non detectable | 7-8 mm | 40-42 mm |

**Table 5.6:** Model parameter values in the three different scenarios: favourable, neutral and unfavourable. Expected tumour size the September 15th, 2018 in the three scenarios.

bladder cancer recurrence with a tumour of size 40-42 mm is predicted for the 15/Sep/2018.

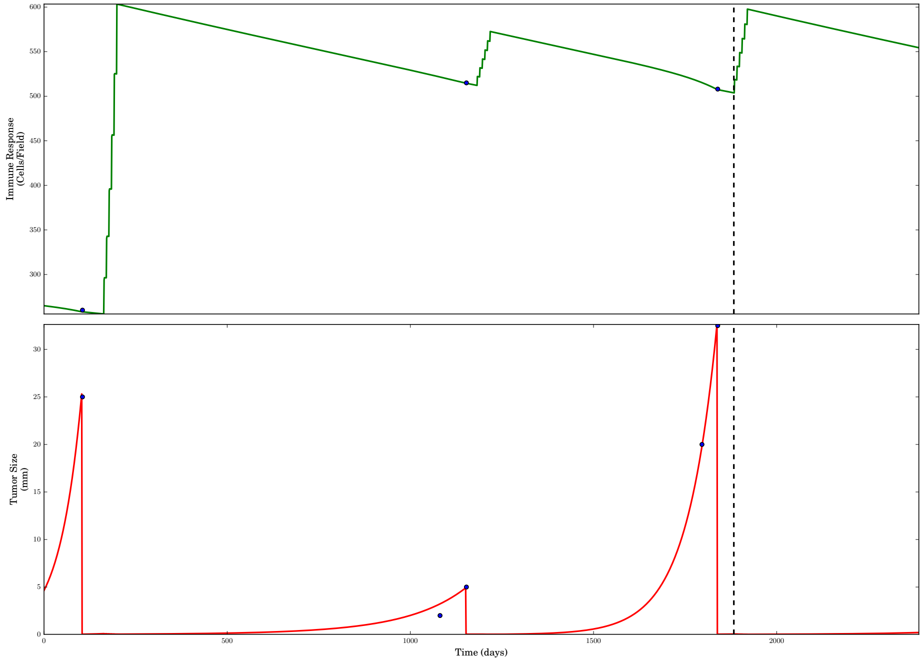


**Figure 5.6:** Favorable scenario. The prediction appears on the right of the dashed vertical black line. It can be seen that the tumour does not seem to grow in the prediction time interval, that is, it seems that the BCG treatment will be able to kill the remainder tumour cells. However, after two years, we must say that the model predicts the apparition of new tumours.

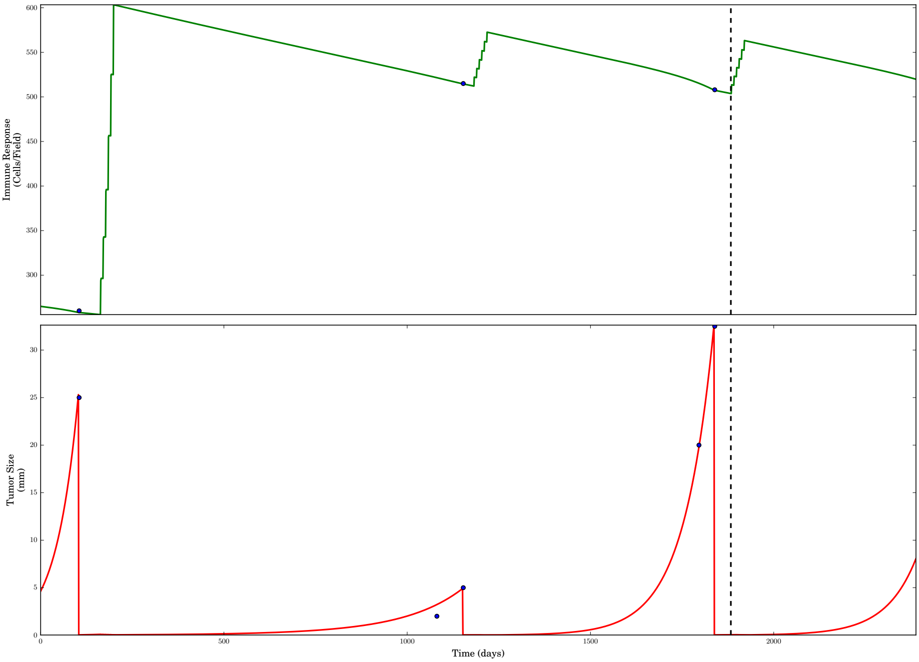


**Figure 5.7:** Neutral scenario. The prediction appears on the right of the dashed vertical black line. After a while, the tumour grows again and a recurrence is expected, reaching the size of 7-8 mm in the middle of September 2018.

### *5.5.1 Patient's follow-up*

Patient X was called for a revision the 15/Mar/2018 and, after a cystoscopy, the doctors did not find any trace of bladder cancer. Then, she was called again the 6/Jun/2018 where the cystoscopy was inconclusive. A week later, a molecular test called XPERT Bladder Cancer Monitor was performed with negative results. Next revision is expected next 10/Sep/2018.

The model in the unfavourable and neutral scenarios, predicts tumour of sizes 9.2 mm and 2.4 mm for 6/Jun/2018. Therefore, the unfavourable scenario is discarded for Patient X. The neutral scenario gives a tumour size not visible in a cystoscopy, although the XPERT Bladder Cancer Monitor test could detect.

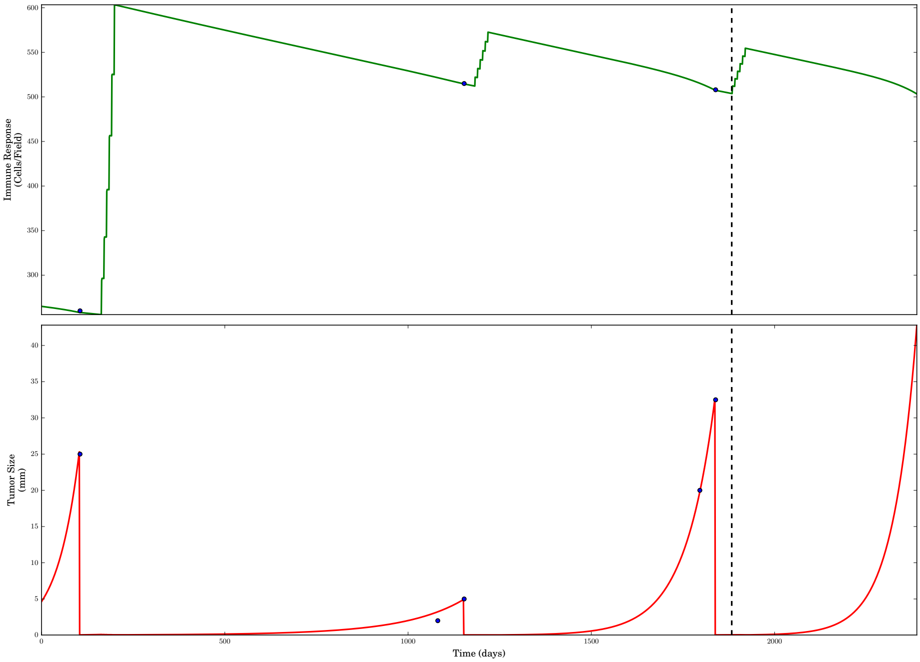


**Figure 5.8:** Unfavourable scenario. The prediction appears on the right of the dashed vertical black line. Here, the effect of BCG is almost inexistent and the tumour grows quickly reaching the size of 40-42 mm in the middle of September 2018. Time $t = 0$ corresponds to 01/Mar/2012.

In the next revision we will check definitively if the neutral scenario is the right one for this patient or we have to wait to see if the favourable scenario is the applicable.

## 5.6 Conclusion

In this chapter, we build a dynamic mathematical model to describe the evolution of the size of a bladder tumour and the immune system response in a given patient. The model had to be calibrated in several stages determined by the TURs included in the hospital protocol and the obtained parameters allowed us to give a prediction about the tumour growth and the immune system response over the next few months in three scenarios: favourable, neutral and unfavourable. To our knowledge, this is the first approach of this type applying the model to a given patient and the results seem to be promising, insofar the doctors of the Hospital Universitari i Politècnic La Fe have scheduled patient's revisions following the predictions given by the model.

Some predictions have been made and they are being checked with programmed patient's revisions. Some scenarios have been discarded, but others have to be examined in the near future. In fact, the Patient X will be called 10/Sep/2018 to measure the size of the tumour, if it exists.

This is a working model, built *ad-hoc* to be adapted to the data provided by the Hospital Universitari i Politècnic La Fe and applied to a particular patient. Thus, we will be able to check the validity of our approach, performing better and more accurate predictions, giving tools to the doctors to administer treatments, to schedule the patient's revisions and, in the future, to determine the best strategies to improve the patient's health avoiding tumour recurrences, if possible, with an important saving of time and resources.

## Chapter published

The results of this chapter have been published in [98].

Chapter 6

# Effect of the early use of antivirals on the COVID-19 pandemic. A computational network modelling approach

## 6.1 Introduction and motivation

This chapter was written in June 2020, when COVID-19 information was still unclear, vaccines has not finished the second phase of study (which means no knowledge about their effectiveness). At the time of reading this thesis, some information might have change, but still, the main conclusions of this work may be useful for future similar situations.

Last December, the Chinese public health authorities informed about several cases of acute respiratory syndrome in Wuhan. By 7 Jan 2020, scientists had isolated a novel coronavirus from these patients not previously identified in humans, and referred as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the agent responsible for the later designated coronavirus disease 2019 (COVID-19) in February 2020, and the World Health Organization (WHO) declared a pandemic [99], [100].

COVID-19 pandemic has already caused more than 430.000 deaths (at the time this lines are written) [101] around the world and it still remains uncontrolled in several continents, mainly in America, Africa and parts of Asia.

Widespread vaccination is probably the only way to restore a fully normal life, as we knew it before the appearance of COVID-19 pandemic, if we do not want to wait until herd immunity is achieved which would also imply a great number of additional human losses. Even though Europe has been one of the continents most affected by COVID till now, recent seroprevalence studies [102] conclude that European countries are probably very far from herd immunity.

Unfortunately, although more than a hundred research groups around the world are developing a vaccine, and healthy volunteers are speeding up the clinical trial process, fabrication of a large quantity for the world population and its widespread distribution, will not be possible at least before the end of Autumn 2020. In the best scenario, some vaccines might be distributed to the most vulnerable population by the end of this year.

Therefore, the research for alternative treatments should be pursued not only for the current situation of the pandemic in America and Africa but also in prevision of a second large wave after the summer season in the northern hemisphere.

One fast way to achieve a successful treatment in the next few months is the repurposing of current approved drugs for its use as antivirals. Unfortunately, there is only one antiviral drug approved for the treatment of COVID that has shown some efficiency against COVID-19 in clinical trials: Remdesivir [103]. However, even though Remdesivir was shown to be superior to placebo in shortening the recovery time in adults hospitalized with COVID-19 and has also shown evidence of lower respiratory tract infection, this drug has not proved to reduce COVID-19 mortality rate as of today. Moreover, this drug needs to be administered at a hospital setting, hampering a more efficient antiviral action if it was delivered at the first symptoms.

This study is based on the hope of the appearance in the clinic of a new antiviral drug approved for the COVID-19 treatment that should be inexpensive, available at the drug stores for a large part of the population, with no significant secondary effects and with some efficacy against the virus. We call these, the *democratic conditions*.

Fortunately, there are many candidates identified by several research groups around the world that might potentially fulfil those democratic conditions. Some of them have been found using an automatic platform to test an *in vitro* library with a large quantity of approved drugs for their action against SARS-2 virus in cell cultures and simultaneously testing its toxicity on those cells [104], [105]. Both quantities, antiviral activity and toxicity, determine a therapeutic window for each drug, and the drugs in the library are ordered according to that window. For instance, Azithromycin has been selected in such a way [106] and clinical trials are being performed with this drug in COVID-19 patients. In a different in vitro screening study [107], Niclosamide and Ciclosonide have been selected as antivirals against SARS-CoV-2 with low toxicity at the required doses. A new formulation of Niclosamide, an anthelminthic drug, to enhance absorption, might be used as a very potent antiviral against SARS-CoV-2. Ciclesonide is a corticosteroid used to treat asthma and allergic rhinitis through inhalation and therefore, even less potent than Niclosamide, is high bio-available at the respiratory tract.

Other drugs have been found through *in silico* (computing programs) studies and then tried its action against the virus in cell cultures. Ivermectin, an anti-parasitic drug, has been identified with that method [108]. Furthermore, a recent clinical trial [109] has shown that Ivermectin might reduce COVID-19 mortality by 40%. Even though, the study had some limitations.

Furthermore, new orally available antivirals have shown potent activity in human epithelial lung cells in vitro (see for instance [110]). Therefore, there are many SARS-CoV-2 antiviral candidates currently under clinical trials.

In this paper we propose a computational network model to assess the effect of the use of an antiviral in democratic conditions and administered early on in the course of the disease, on the COVID-19 transmission dynamics. To do so, we build a random network model, estimate the model parameter values that explain the dynamics of the COVID-19 in Spain, and we simulate possible future scenarios where an accessible and cheap antiviral is available and can be considered as an effective treatment.

The paper is organized as follows. In Section 6.2, the computational random network model is built and the model parameter values estimated. In Section 6.3, the model is validated, that is, we check that reproduces accurately the situation in Spain since the beginning of the pandemics. In Section 6.4, we describe and simulate scenarios to assess the use of antivirals, which may be an additional tool to fight against COVID-19. Finally, in Section 6.5 we present our conclusions.

## 6.2 Model building

### *6.2.1 Random networks*

Networks have become a paradigm of paramount importance in the analysis of many complex systems. In the field of epidemiology, networks have been used to describe the transmission dynamics of several diseases [111]–[114].

A network is a set of nodes representing individuals. Labels or properties may be assigned to each node, such as age, sex, and state respect to the disease (susceptibility, infection, recovery, latency, etc.). Nodes are connected by edges that represent disease transmission paths. Once the network model and the disease evolution rules are stated, it is possible to simulate the evolution of the disease on the network nodes over time and to study its spread on the population.

The already developed of networks provide several standard alternatives for implementing the network substrate. The most traditional one is based upon the pioneering work of Bollobas [115], employing the so-called *random graphs*, where connections among the pairs of subjects are created with the same probability. The most natural sparse generalization of the complete graph in mathematics is provided by the so-called Erdös-Rényi graphs [115], which in the fifties of the past century defined the concept of a random graph. In modern times, this idea has been redefined as a random network and it has become an important paradigm with many applications.

The spread of infectious diseases is determined by random encounters among people around the same location: meeting at the bus stops, crossing in the streets, gathering at shop centers, etc. COVID-19 is known to be transmitted person to person, mainly through respiratory droplets produced when an infected person coughs or sneezes [116]. As other respiratory viruses, it induces coughing and sneezing in infected subjects, which also favours the transmission of the disease.

For these reasons, we have chosen the Erdös-Rényi random network as the most appropriate for the modelling of the transmission of COVID-19 and infectious diseases in general. Random networks are characterized by the number of sites or nodes $N$ and the average number of contacts of every individual $k$. Consequently, the number of links in the network is given by $N \times k/2$. These links are randomly assigned to pairs of nodes with the obvious rule that, at most, only a link can connect two nodes. We will say that two nodes are neighbours if they are connected. An example of a random network can be seen in Figure 6.1.

### *6.2.2 Model compartments*

The network is the substrate that determines how people relate and what the transmission paths are. However, to study the transmission dynamics of diseases, we have to build a compartmental model and determine how individuals transit from one state to another as time goes on. The compartments are the following:

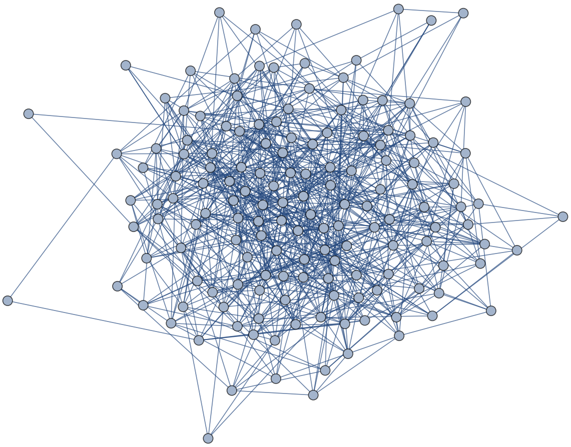


**Figure 6.1:** Random network with $N = 150$ and $k = 10$. The small circles are the nodes representing individuals and the lines between nodes are the links describing the contacts among individuals with possible transmission of the virus. The links describe the transmission paths of the disease.

- Susceptible (S) when the individual is healthy;
- Latent or Exposed (L) when the individual has been infected but he/she still is not infectious;
- Infectious (I) when the individual can infect others;
- Recovered (R) when the individual recovers from the disease being asymptomatic or having mild symptoms;
- Hospitalized (H) when the individual has severe symptoms and needs to be hospitalized;
- Deceased (F) when the individual dies because of the virus;
- Discharged (A) when the individual gets better and is discharged from the hospital.

Also, the nodes have a label determining if they are in Quarantine (Q), that is, when the individual is at home to avoid the spread of the virus. This label can be assigned to all individuals except people in H (hospitalized) and F (deceased).

In the proposed model, we consider that people who decease as a consequence of COVID-19, pass away at the hospital. The data regarding people who die outside the hospital in Spain is scarce and inconclusive. Also, we do not consider re-infection. The total population in Spain will be given by the term $P_T = 47\ 100\ 396$ [117].

### 6.2.3 Transmission dynamics

In this section we are going to define the transmission dynamics of COVID-19 on the network. The time step is set in 1 day.

1. A susceptible (S) node may transit to latent (L) state if it gets infected through a successful contact with an infectious (I) node with a transmission rate $\beta > 0$.

   Within the dynamics of the network, this transition is going to be simulated as follows: if a node is susceptible (S), we collect all the infectious (I) neighbouring (connected) nodes. In turn, for each of the infectious Neighbors, we generate a random number $r \in [0, 1]$ and if $r < \beta$ we consider that this neighbouring node transmits COVID-19 to the node and the node is then labelled as latent (L) and changes its state.

2. A latent (L) node becomes infectious (I) at rate $l_i > 0$ after some days, and it is able to infect other nodes.

   Within the simulation, a latent node needs a number of days to become infectious (3 days as we shall see later).

3. An infectious (I) node may be asymptomatic or have mild symptoms and recover (R) without being admitted in the hospital at rate $i_r > 0$; or it may have severe symptoms and be admitted in the hospital (H) at rate $i_h > 0$.

4. A node in the hospital (H) may get better and finally be discharged (A) at rate $h_a > 0$, or get worse and eventually decease at rate $h_f > 0$.

5. Also, a susceptible (S), latent (L), infectious (I), recovered (R) and discharged (A) node can be labelled as in quarantine, during the quarantine periods, if it remains at home to avoid contagion. In this case, the node cannot infect nor be infected.

Points 3 and 4 correspond to a bifurcation, that is, there are two possible ways to change the state.

In the first case (point 3), the parameters involved in the transition from infectious (I) are $i_h$ (to H) and $i_r$ (to R). These model parameters, as we will see later, have two parts: the first one is the probability $p$ to take the way to H or the probability $1 - p$ to take the way to R, and the second is the time to reach the new state, H or R (may be different). Hence, to simulate the transition of a node from I to H or R, we generate a random number $r \in [0, 1]$ and if $r < p$ its destiny will be H. Otherwise R. Then, after some determined number of time steps, the node becomes hospitalized/recovered.

For the second case (point 4), that is, the transition from H to A or F with the model parameters $h_a$ and $h_f$, we proceed analogously.

Figure 6.2 shows a flow diagram about how an individual may move around respect to the disease, following the paths described above.

### *6.2.4 Model parameters*

The goal of this section is to quantify the model parameters in order to determine the transmission dynamics of the COVID-19. In Figure 6.3 we can see the time line of COVID-19.

1. **Average degree $k$ of the network.** The average degree $k$ of a network is the average number of contacts of the nodes. For infectious diseases as the respiratory syncitial virus (RSV) whose contagion mechanisms are similar to COVID-19, $k$ is in the range $48 - 54$ [113], [118] contacts per day with possible contagion. Recent unpublished network model studies

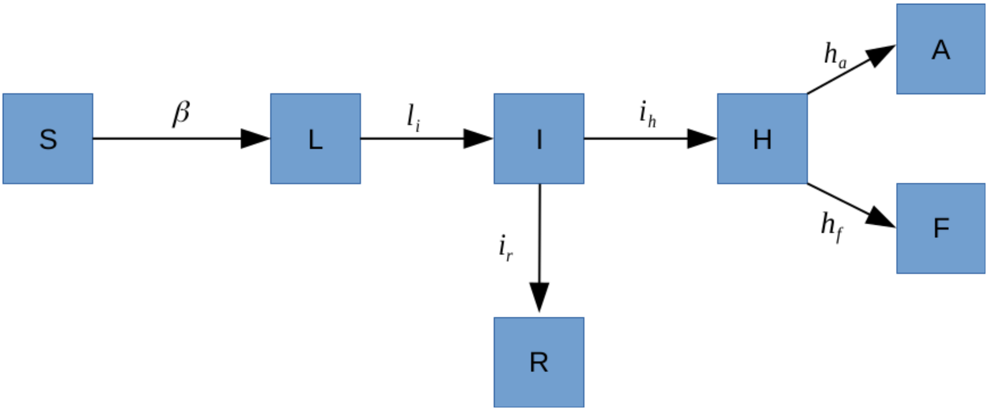


**Figure 6.2:** Flow diagram of the COVID-19 transmission dynamics. The boxes represent the states of the individuals respect to the disease and the letters next to the arrows, the transition rates between the states.

for influenza, returns values for $k$ around $48 - 49$. Hence, we are going to assume in our study

$$k = 50.$$

2. **Quarantine.** In March 14, 2020, 70% of people in Spain were isolated at home (quarantine) [119], [120]. After that, in March 31, 2020 there was a strict quarantine isolating at home 85% of people. The strict quarantine was relaxed April 13, 2020 when the percentage of isolated people was reduced again to 70%.

3. **Transition rate from latent to infectious $l_i$.** The incubation period described in medical literature, refers to the time interval since an individual gets infected until he/she presents symptoms (onset). In [121] the authors affirm that, for COVID-19, the incubation period is 4 days with a range from 2 to 7 days. However, the latency period is the time since one gets infected until he/she is able to infect others, regardless whether symptoms have emerged. Therefore, the incubation period does not correspond exactly to the latency period. There are limited evidences suggesting that the virus may be transmitted one or two days before the onset [122]. Hence, we are going to assume the latency period is going to

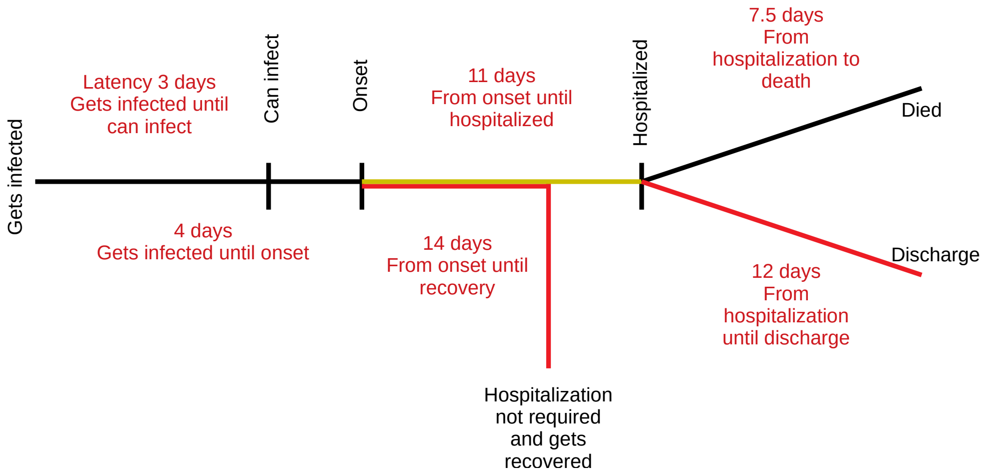


**Figure 6.3:** Time line of COVID-19, from an individual gets infected until gets recovered, dies or is discharged.

be less than 4 days, in particular, 3 days, taking

$$l_i = 1/3.$$

In the following, we are going to add 1 day when we have to take into account the time between the end of the latency period and the onset.

4. **Transition rate from infectious to hospitalized $i_h$.** In Spain, between April 27 to May 11, the preliminary report of the 1st round of the National Sero-epidemiological study for SARS-Cov-2 infection in Spain [123], says that the accumulated infected is 5% of the population. Using data retrieved from [124] in the same time interval as above, the average percentage accumulated hospitalized individuals in Spain is 0.2520%. Then,

$$\frac{100 \times 0.2520}{5} = 5.04\%$$

is an estimation of the percentage of the infected who are hospitalized. Furthermore, the time from onset to get hospitalized is 11 days [100]. We add one day for the end of the latency to the onset. Therefore,

$$i_h = 0.0504/12.$$

5. **Transition rate from infectious to recovered $i_r$.** From the previous point, we have that $100\% - 5.04\% = 94.96\%$ of the infected are not hospitalized and get recovered after 14 days [125, page 14, 1st paragraph] plus a day for the end of the latency to the onset. Thus,

$$i_r = 0.9496/15.$$

6. **Transmission rate $\beta$.** $R_0$ is the basic reproductive number [126] and it can be interpreted as the number of persons an infectious individual may infect during the time he/she is infectious, in a scenario where almost all the people are susceptible, that is, at the very beginning. For our network model, we have that

$$R_0 = \frac{k \times \beta}{i_h + i_r}.$$

Several papers have treated to estimate the value of $R_0$ for COVID-19. We are going to use [127] where $R_0$ is 5.7 CI95% $(3.8 - 8.9)$. This way,

$$5.7 = R_0 = \frac{k \times \beta}{i_h + i_r} = \frac{50 \times \beta}{\frac{0.0504}{12} + \frac{0.9496}{15}},$$

where

$$\beta = 0.00769576.$$

This $\beta$ will be valid at the beginning of the outbreak. Nevertheless, in Spain, we had periods of quarantine and some measures have been taken to avoid social contacts (teleworking, avoiding meetings, preparing children schools, safe distancing, room ventilation, limiting the capacity of bars, restaurants, hotels and public places), measures of protection (face masks, hand washing), the response of the virus to the summer temperatures, the population immunity, etc. Thus, transmission rate has certainly changed.

7. **Transition rate from hospitalized to death $h_f$.** Using data retrieved from [124], the average percentage of accumulated hospitalized individuals who died in Spain between March 11 to May 18 is 19.2348%. Also, the time spent in hospital by people who eventually die is 7.5 days on average [100]. Thus,

$$h_f = 0.192348/7.5.$$

8. **Transition rate from hospitalized to discharged $h_a$.** From the previous point, we have that $100\% - 19.2348\% = 80.7652\%$ of the hospitalized get recovered and are discharged. Also, the time spent in hospital by people who eventually is discharged is 12 days on average [100]. Thus,

$$h_a = 0.807652/12.$$

At this point, we have to establish the initial condition with very little information. First, we take as initial instant January 31 2020. This day was declared the first case in Spain. We are going to consider that this day nobody was in quarantine nor hospitalized nor discharged nor deceased because of COVID-19. However, we have to determine the initial number of latent and infected.

Taking into account that the percentage of infected people between April 27 and May 11 lies in the interval $4.7\% - 5.4\%$ [123] and the intrinsic randomness of the network building, we have performed a calibration of the model, where the parameters to be calibrated are the initial number of latent and infected. The process returned 30 model realizations fulfilling the above infected people restriction. The mean and the 95% confidence band of the 30 calibrated realizations can be seen in Figure 6.4(a). These 30 realizations allow us to say that, initially in January 31, there were in Spain 267 latent IC95% $[135, 393]$ (percentages 0.00057% IC95% $[0.00029\%, 0.00083\%]$) and 497 infectious IC95% $[371, 628]$ (percentages 0.00106% IC95% $[0.00079\%, 0.00133\%]$). The percentages are calculated respect the total Spanish population.

All the above parameters, the initial condition, and the 30 realizations will be used for model validation and further simulation of scenarios. The results and the graphs will be shown in percentages with respect to the total Spanish population in order to facilitate the comparison and the possible extrapolation to other regions or countries.

## 6.3 Model validation

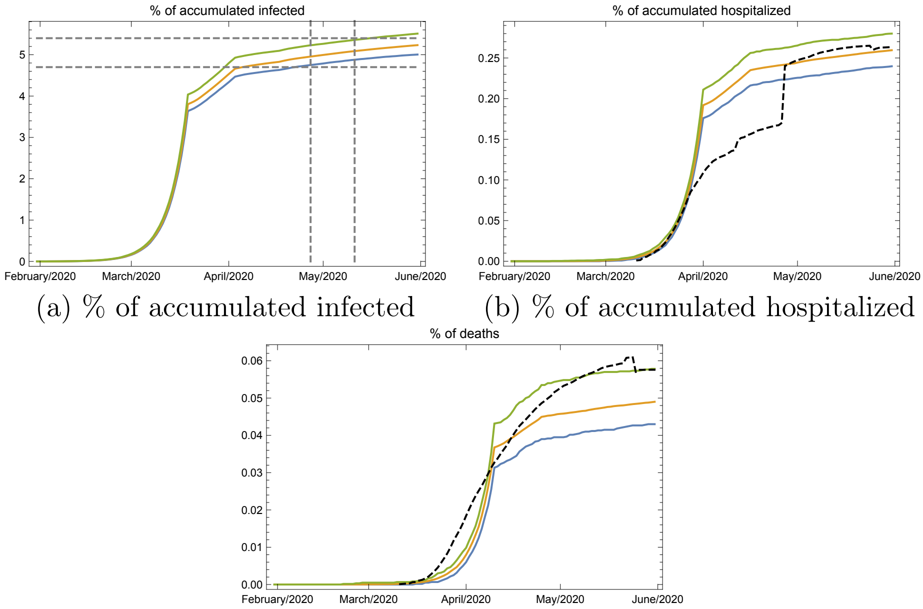
We are going to validate the model comparing the model output with the data [124] from January 31 until May 31. The transmission rate $\beta$ obtained in the previous section, as already mentioned, is valid during the beginning of the outbreak, but later, some measures to avoid the contagion were taken.

Although we initially had data until May 18, during the study new data have been added and they are available until 31 May, which we have included for validation.

In order to simulate these measures we are going to consider that the transmission rate from January 31 (initial condition) to March 14 (starts the quarantine) is $\beta = \beta_1 = 0.00769576$. Then from March 15 until May 31, we consider that, apart from the quarantine, people took measures to prevent the transmission of COVID-19 that we quantify as $\beta_2 = \beta_1 \times 0.5 = 0.00384788$.

Using the model parameter values of Section 6.2.4 and the $\beta$'s defined here, we perform the 30 realizations, calculate the mean and the 95% confidence interval. The similarities between the model output (coloured lines) and the data (black dashed lines) can be seen in Figure 6.4. The black dashed lines in Figure 6.4(b) and (c) correspond to accumulated data. As we can see, in the data time series, there are unexpected jumps and drops improper of accumulated data. This is one of the facts that has generated controversy about the quality of the data.

In any case, given that the model reproduces the data quite well, we can expect to make accurate predictions.



(a) % of accumulated infected

(b) % of accumulated hospitalized

(c) % of deceased

**Figure 6.4:** Evolution of the (a) percentage of accumulated infected, (b) percentage of accumulated hospitalized, (c) percentage of deceased, from January 31, 2020 until May 31, 2020. In Figure (a) we can see that the model output between April 27 and May 11 is inside the interval 4.7% – 5.4%, that is, the grey dashed line rectangle. In Figures (b) and (c) we can see the similarities of the model output (coloured lines: orange represents the mean; blue represents the percentile 2.5; green represents the percentile 97.5, every day) and the data (black dashed lines) for accumulated hospitalized and deceased.

## 6.4 Simulation of the effect of the antiviral

It is clear that COVID-19 is not going to be controlled until we reach herd immunity, and it will be possible when

$$1 - \frac{1}{R_0} = 1 - \frac{1}{5.7} \approx 82.5\%$$

of the population [126] will be immune. This level of immunity can be achieved if most of the people get infected, effective vaccines or effective antivirals are available.

Since March 14, in Spain, we are under a state of emergency, with restrictions in traveling and many activities. Four stages (0, 1, 2 and 3) have been defined to return to normal life and they have been activated from the middle of May, step by step, in regions satisfying certain public health conditions. In mid-June, most part of Spain is in stages $2-3$.

For the simulations, we approximate the stage scenarios assuming that people come out of quarantine in steps of 25% in June 1, 8, 15 and 22. Also, we maintain the typical prevention measures as social distancing, the use of face masks, cleaning the hands regularly, etc. Furthermore, the summer season may difficult the transmission of the virus, at least, because there are a lot of activities performed outdoors. Hence, from June 1 until September 30, the transmission rate considered will be $\beta_3 = \beta_1 \times 0.25 = 0.00192394$. Once the summer ends, during autumn, October 1 2020 until May 31 2021, we increase the transmission rate to $\beta_2 = 0.003847889$. Summarizing:

- from Jan 31 to Mar 14, $\beta_1 = 0.00769576$;
- from March 15 to May 31, $\beta_2 = 0.00384788$;
- from Jun 1 to Sep 31, $\beta_3 = 0.00192394$;
- from Oct 1 to May 31, $\beta_2 = 0.00384788$.

Let us now focus on the antivirals. The effective antiviral works as follows: when a person has symptoms, he/she can take the treatment with antivirals in such a way that mitigates the symptoms and consequences of the disease, but at the same time, allows him/her to generate antibodies and get protected. If, in addition, this person informs his/her contacts (nodes connected or neighbors) that he/she is infected, his/her contacts can also follow the same treatment and the transmission cut-off effect would be increased, preventing, at least in part, the appearance of possible new outbreaks.

Now, we define the base case simulation. Let us assume that the antiviral is available from October 1, 2020. When a person has symptoms of COVID-19 disease, about 15% of those infected [125, page 12, 3rd paragraph][1], can take the treatment of this antiviral for 15 days with an effectiveness of $p_{ef}\%$.

When this person takes the treatment, he/she informs the $p_c\%$ of his/her contacts who also take the treatment, unless they are recovered, discharged or in hospital.

If one of the contacts is susceptible (S), the treatment will protect him/her for 30 days and afterwards he/she will return to the susceptible state. If one of his/her contacts is latent (L) or infectious (maybe asymptomatic), after 3 days of treatment he/she will move to recovered because the antiviral has relieved him/her (with a probability of $p_{ef}\%$) of the hardest part of the illness and his/her body has had antibodies to successfully defend against possible re-infections.

Under the above conditions and the possibility of having available antivirals, we are going to simulate the following scenarios:

- Base case: no antivirals are available;
- Case 1: antivirals are available from October 1 with several percentages of effectiveness and with several percentages of contacts who are informed and take the treatment;
- Case 2: the same as Case 1, where the transmission rate from Oct 1 2020 to May 31 2021 is $\beta_4 = 2 \times \beta_1 = 0.00513051$, higher than the Base Case $\beta_2 = 0.00384788$. This case simulates the scenario where suppression

[1] In [125, page 12, 3rd paragraph] the authors say that *Most people infected with COVID-19 virus have mild disease and recover. Approximately 80% of laboratory confirmed patients have had mild to moderate disease, which includes non-pneumonia and pneumonia cases, 13.8% have severe disease (dyspnea, respiratory frequency $\geq 30$/minute, blood oxygen saturation $\leq 93\%$, PaO2/FiO2 ratio $< 300$, and/or lung infiltrates $> 50\%$ of the lung field within 24-48 hours) and 6.1% are critical (respiratory failure, septic shock, and/or multiple organ dysfunction/failure). Asymptomatic infection has been reported, but the majority of the relatively rare cases who are asymptomatic on the date of identification/report went on to develop disease. The proportion of truly asymptomatic infections is unclear but appears to be relatively rare and does not appear to be a major driver of transmission.* In view of this information, taking into account that it is not clear what can be determined as "having symptoms", we decide to consider that 15% of those infected have recognizable symptoms.

and/or mitigation of the measures may happen and, consequently, the transmission rate may be higher;

- Case 3: the same as Case 1, where antivirals are available later, from December 1.

### *6.4.1 Base case*

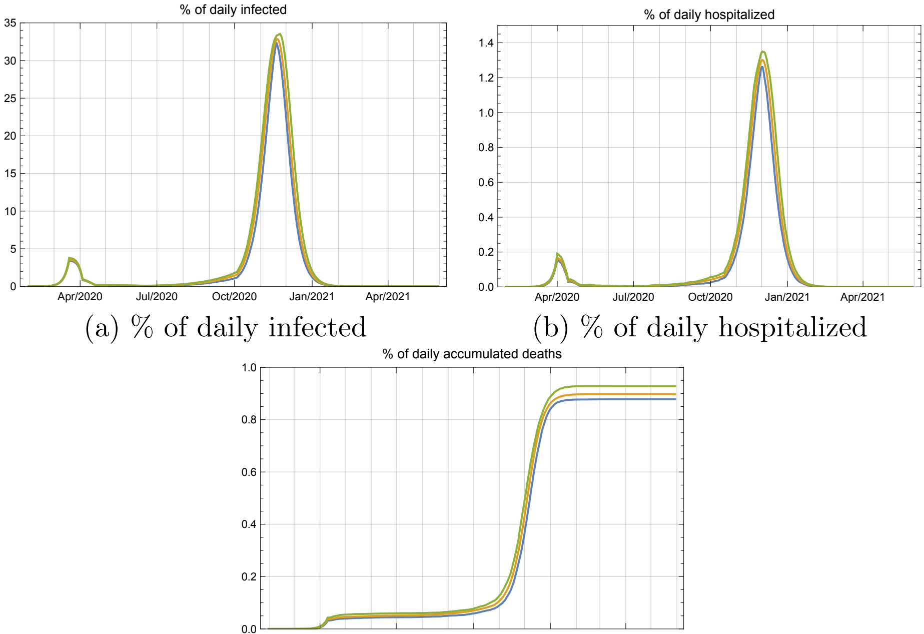
This base case is a situation that would allow us to get out of quarantine while maintaining due precautions in time to avoid contagion that could lead to a new outbreak.

The result of this simulation can be seen in the Figure 6.5. Under the simulated conditions, the control over the epidemic is maintained until September, when starts a new outbreak that becomes more acute when the transmission rate increases in October (children in schools, colder weather, shorter days, fewer outdoor activities). Note that if the outbreak is not controlled, it could reach a much higher peak in November than the one last March, which corresponds to the peak on the left of the graph.

### *6.4.2 Simulation with several percentages of effectiveness*

In Cases 1, 2 and 3, we perform simulations to estimate the sensitivity of the antiviral effectiveness with respect to the base case. To do this, we consider that the percentage of contacts to which we inform that we are taking the antiviral is $p_c = 75\%$ and the effectiveness of the antiviral $p_{ef}$ varies by 35%, 50%, 65% and 80%. The comparison with the base case for the average percentage of hospitalized and deaths can be seen in Figure 6.6. We used hospitalized patients to assess the possibility of the health system becomes overcrowded again.

In Figure 6.6(a) (Case 1), it is interesting to note that, in the scenario where the percentage of people who have symptoms is low (15%) and effectiveness of the antiviral is also low (35%), there is a significant reduction of more than 50% of hospitalized people at the peak. Also, while the effectiveness of the antiviral



(a) % of daily infected

(b) % of daily hospitalized

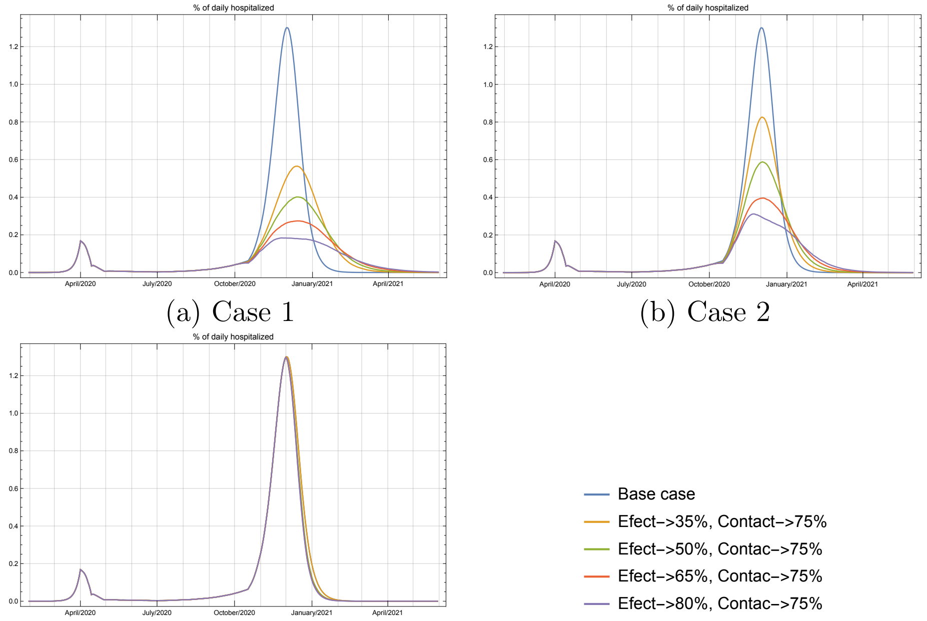
(c) % of daily accumulated deceased

**Figure 6.5:** Base case. Evolution of (a) percentage of infected per day, (b) percentage of hospitalized per day and (c) percentage of deaths, from 31 January 2020 to 31 May 2021. The figures show that the pandemic has been under control until September, when there is again an increase in the number of people infected, which grows by the increase in the transmission rate in October. The peak on the left in graphs (a) and (b) corresponds to the peak of March 2020. In all three graphs, the ordinate axis represents a percentage of the total population and the abscissa axis represents time.

significantly reduces the number of hospitalizations, it also delays around 15 days the peak and, consequently, the saturation of the health system.

In Figure 6.6(b) (Case 2), we can see that the increasing of the transmission rate $\beta_4 = 0.00513051$ reduces the effect of the antivirals and the peak delay effect we mentioned before, does not appear.

In Figure 6.6(c) (Case 3), the availability of the antivirals is delayed 2 months until December 1. Here, we can see the importance of having accessible antivi-



(a) Case 1

(b) Case 2

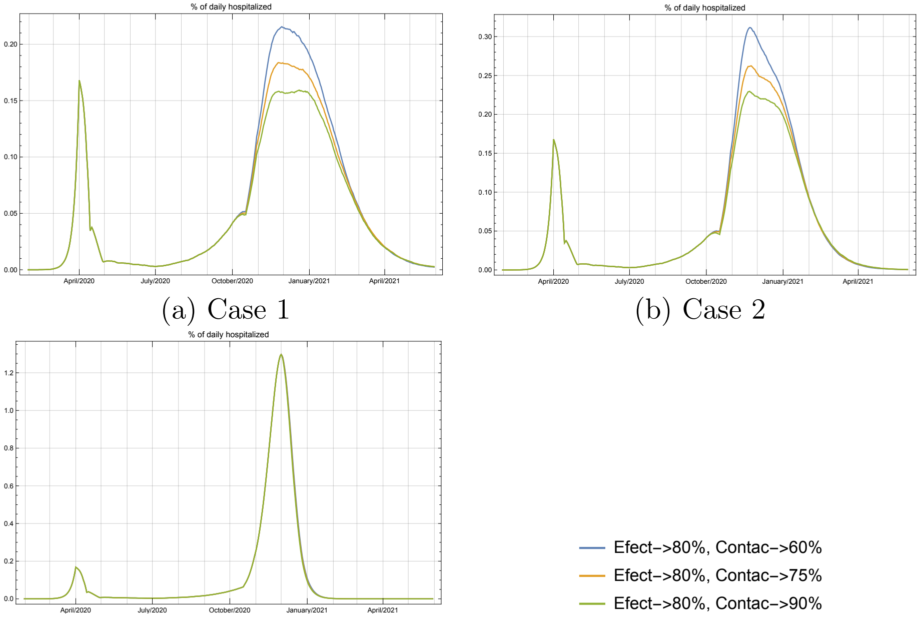
(c) Case 3

**Figure 6.6:** Evolution of the percentage of hospitalized per day, in Case 1 (a), Case 2 (b) and Case 3 (c) from 31 January 2020 to 31 May 2021, compared with the Base Case (blue line). In (a) we can see the reduction effect of the antivirals even with low effectiveness. In (b), the effect of the antivirals is reduced compared to (a) because the increasing of the transmission rate from Oct 1 to May 31. In (c), we can see that the antiviral do not have any perceptible effect. This shows the importance of the time where the antivirals should be available. The peak on the left in graphs (a), (b) and (c) corresponds to the peak of March 2020. In all three graphs, the ordinate axis represents a percentage of the total population and the abscissa axis represents time.

rals as soon as possible, because any delay may reduce its effect to the point of making it useless.

### *6.4.3 Simulation with several percentages of contacts who are informed and take the treatment*

In Cases 1, 2 and 3 we assess the effect of informing to a higher or lower percentage of our contacts that we are in treatment for COVID-19 and they take it. We assume $p_{ef} = 80\%$ constant and $p_c$ varying in the 60%, 75% and 90%.



(a) Case 1

(b) Case 2

(c) Case 3

**Figure 6.7:** Evolution of the percentage of hospitalized per day, in Case 1 (a), Case 2 (b) and Case 3 (c) from 31 January 2020 to 31 May 2021. In (a) we can see how antivirals may reduce the hospitalized people to levels similar to the last March 2020. It does not happen in (b) because of the increasing of the transmission rate from Oct 1 to May 31. In (c), where the availability of the antiviral is delayed, we do not have any perceptible effect of the antivirals. The peak on the left in graphs (a), (b) and (c) corresponds to the peak of March 2020. In all three graphs, the ordinate axis represents a percentage of the total population and the abscissa axis represents time.

In this case, Figure 6.7, we can see how the communication to our contacts that we have symptoms may produce a reduction of the hospitalized preventing the collapse of the health system. It is worth mentioning that the simulated scenarios can be considered conservative, if we take into account that our contacts may inform their contacts to take the treatment, expanding the subnetwork of people under treatment and cutting the disease transmission paths.

In the simulated case Figure 6.7(a) and (b) (Cases 1 and 2), the peak is flatter and appears in mid-November - December. Comparison of graphs in Figure 6.7 shows again the importance of the accessibility of antivirals in the right moment, that is, before the main outbreak starts.

## 6.5 Conclusion

In this chapter, we present a computational network model to assess the effect of the antivirals on the COVID-19 pandemic under several scenarios.

As we are far from having COVID-19 controlled and while an effective vaccine is found, antivirals may be a good option that we study in this work. The simulations carried out show us that the use of effective antivirals would help to obtain herd immunity without the cost in health resources and human lives that has taken place until now. In fact, it shows that even using low-effective antivirals and communicating our disease to a low percentage of our contacts, we can achieve a significant reduction in hospitalizations and avoid further saturation of the public health system if the antivirals are available before the starting of the new main outbreak.

The scenarios presented here may be considered as conservative in the sense that we do not consider the likely possibility that, if we have symptoms and communicate it to our contacts, in turn they may also communicate to their contacts and so on, expanding the network of people who can take the antiviral treatment cutting the virus transmission ways and blocking the contagion.

On the other hand, we are assuming that infected people acquire permanent immunity, which it is still under discussion.

Finally, one of the most important conclusion of this simulation is that any action we take against COVID-19 adds up significantly when it comes to fighting it. As already indicated for the masks [128], whatever percentage reduction an antiviral may have on the number of infected/hospitalized/dead, it must be applied.

## Chapter published

The results of this chapter have been published in [129].

Chapter 7

# Modelling COVID-19 with uncertainty in Granada, Spain. Intra-hospitality circuit and expectations over next months

## 7.1 Introduction

Regarding COVID-19, we have been buying time looking for a balance between the overload of hospitals and the economy until antivirals and vaccines can be available. In this sense, mathematical models have been remarkable tools to know in advance the appropriate time to enforce the population restrictions and distribute the hospital resources to face the new waves.

Several approaches have tried to model mathematically the transmission dynamics of COVID-19 using SIR and SEIR-type models [130]–[136]. In other references as [137]–[140], moreover, consider asymptomatic infected individuals, most of the infected population, who play an important role in spreading the virus. One of the main advantages of these mathematical models is that different simulations can be done, allowing the study of different variety of complex scenarios, although with limitations due to the many uncertainties regarding to the disease.

One of the main parameters in the epidemiological models is the reproductive number ($R_0$), which quantifies the average of secondary cases generated from each infected person in a susceptible environment during an outbreak. Different studies have estimated that the $R_0$ value in China at the beginning of the pandemic was between 2 and 7.1. This means that, following [126], the pandemics is not going to be controlled until we reach the heard immunity, and it will be possible when the percentage $H = 100(1 - \frac{1}{R_0})$ of the population will be immune, that is, recovered plus vaccinated. Thus, $H = 50\%$ for $R_0 = 2$ and $H = 86\%$ for $R_0 = 7.1$. Also, it has been observed that $R_0$ value drops sharply after the implementation of control measures against COVID-19 [141] reducing the percentage to reach the herd immunity.

Under a demographic point of view, Granada is a province in southern Spain with a population of 921 511 inhabitants, around 30% of the population live in the capital and the province has a population density of 73.5 people per $km^2$ [142]. Compared with the province of Madrid with 829.6 people per $km^2$, the province of Granada has a low density of population. Also, the percentage of accumulated infected between 1 May 2020 and 1 June 2020 was $[2.3\% - 4.8\%]$ in Granada and $[10.3\% - 13.3\%]$ in Madrid [143], [144]. Therefore, the province of Granada had a low prevalence at the end of the first wave.

In this chapter, we propose a mathematical model to study the transmission dynamics of SARS-CoV-2. It is a SEIR model in which the circuit of patients moving throughout the hospital dependencies is also considered, returning a more precise portrait of the use of hospital resources with the aim at predicting hospital's overloads. To our knowledge, this is a novelty that our model provides, thanks to the data retrieved from the Granada's hospitals. Also, we consider the asymptomatic infectious individuals who transmit the disease but are not reported by the health system. Furthermore, two main seasons September-April (virulence season) and May-August (non-virulence season) are considered. In these seasons the hospital pressure is significantly different, and consequently, admissions and deaths. Once the model is implemented and calibrated, we validate it, analyse the results and perform calculations simulating possible future scenarios to evaluate the effect of the COVID-19 vaccination, that allow us to provide some public health recommendations.

In the whole process, calibration-validation-prediction, we consider the data uncertainty. Then, we are able to obtain model parameter values with mean and 95% confidence intervals (CI95%) and, consequently, the predictions will also be represented using means and CI95%.

The chapter is structured as follows. In Section 7.2, we propose the model and calibrate it with data provided. In section 7.3, results are shown, and Sections 7.4 and 7.5 discuss the results and outline the conclusions, respectively.

## 7.2 Materials and Methods

### *7.2.1 Model building*

With respect to COVID-19, an individual can be considered as:

- **(S)** susceptible, when the individual is healthy;
- **(Q)** in lockdown, when the individual is at home to avoid the spread of the virus. In our model, lockdown only considers the March's lockdown time;
- **(L)** latent or exposed, when the individual has been infected but it is not infectious yet;
- **(I)** infectious, when the individual is capable of spreading SARS-CoV-2;

- **(R)** recovered, when the individual recovers from the disease being asymptomatic or having mild symptoms;
- **(H)** hospitalized at ward, when the individual has severe symptoms and needs to be hospitalized;
- **(U)** in intensive care unit (ICU), when the individual has severe symptoms and needs to be treated in the intensive care unit;
- **(F)** deceased, when the individual dies because of the disease;
- **(HU)** after ICU, when an individual is transferred from ICU to other non-ICU department due to improvement in the evolution but still requires hospitalization;
- **(A)** discharged, when the individual gets better and is discharged from hospital.

We are going to use data of patients hospitalized, in ICU, discharges and deaths of the Spanish province of Granada provided by the hospital H.U. Virgen de las Nieves and we scale these data in order to generate an inclusive model which represents the whole province. The province of Granada has $P_T = 921\ 511$ inhabitants, and we consider the $P_T$ value as constant during the simulation time. Our simulations begin on 1 March 2020 and the time step $t$ is one day.

1. When the alarm state is decreed and most of the people have to be in lockdown, that is, move from $S$ to $Q$, it is modelled by the term $s_q(t)$, where the model parameter $s_q(t)$ determines the transit of people from $S$ to $Q$. $s_q(t)$ takes the value 0 except for 16 March 2020 and 31 March 2020, when the lockdown and the strict lockdown began in Spain, and 700 000 and 150 000 more individuals in Granada, respectively, move from S to Q [145].
2. When the lockdown finishes, the transit of individuals from $Q$ to $S$ is modelled by the term $q_s(t)$ where the model parameter $q_s(t)$ is 0 except for 13 April 2020 when the strict lockdown finishes and 150 000 individuals

move from Q to S, and from 5 May to 21 June 2020, leaving the lockdown 8 750 people every day due to the gradual end of the confinement.

3. An individual moves to latent state (L) if he/she gets infected by contact with an infectious individual. People in the hospital are isolated and controlled and, therefore, discarded for contagions. The transit is modeled by the non-linear term $\beta S(t)\frac{I(t)}{P_T}$, where the transmission rate parameter $\beta$ has to be calibrated. Furthermore, this parameter will change over time due to the global public health interventions.

4. A latent individual transits to infectious state after a while and this is modelled by the linear term $l_i L(t)$, where the latency period $t_0 = 1/l_i$ is the time from the moment in which an individual is infected until the moment in which is able to transmit the virus, $t_0$. This period is different to the typical incubation period time from infection to onset, and according to [121], it takes from 2 to 7 days. Even though there is limited evidence about the possibility of infection one or two days before onset [122], let us consider that $t_0$ may take values from 1 to 6 days, and $l_i = 1/t_0$.

5. An infectious individual may become hospitalized (H), admitted to Intensive Care Unit (U) or get recovered (R), and these transits are modeled by the linear terms $i_h I(t)$, $i_u I(t)$ and $i_r I(t)$, respectively. Here we have 3 possible ways for infectious individuals. Every one takes its time and has its probability, that is, $i_h = p_1/t_1$, $i_u = p_2/t_2$ and $i_r = (1 - p_1 - p_2)/t_3$.

   - $p_1$ is the percentage of infected who become hospitalized and $t_1$ days the time it takes; $p_2$ is the percentage of infected people admitted directly at ICU and $t_2$ days the time it takes. Then, the parameters of the transition from $I$ to $H$ and from $I$ to $U$ are $i_h = p_1/t_1$ and $i_u = p_2/t_2$, respectively.

   - Those infected not hospitalized require around $t_3 = 14$ days to recover [146]. Thus, the transition from $I$ to $R$ is governed by the parameter $i_r = (1 - p_1 - p_2)/14$.

6. People hospitalized may move to ICU (U) if they get worse, may decease (F) or may be discharged (A). These transits are modelled by the linear terms $h_u H(t)$, $h_f H(t)$ and $h_a H(t)$, respectively. As before, here we have 3 possible ways for hospitalized individuals and each one takes its time and its probability.

   - $p_4$ is the percentage of hospitalized people who need to be admitted to ICU and $t_4$ days the time it takes. Hence, the people in $H$ move to $U$ governed by the parameter $h_u = p_4/t_4$.

   - $p_5$ is the percentage of hospitalized people who die after an average of $t_5$ days. Then, the transition parameter from $H$ to $F$ is $h_f = p_5/t_5$.

   - $(1 - p_4 - p_5)$ is the percentage of hospitalized who are discharged after an average of $t_6$ days in the hospital. Hence, the transition parameter from $H$ to $A$ is $h_a = (1 - p_4 - p_5)/t_6$.

7. People in ICU (U) may decease (F) or may get better and be transferred to other non-ICU department (HU). These transits are modeled by the linear terms $u_f U(t)$ and $u_{hu} U(t)$, respectively. Here we have 2 possible ways for individuals in ICU, die or get better and each one takes its time.

   - The transition parameter from $U$ to $F$ is $u_f = p_7/t_7$, where $p_7$ is the probability to die after $t_7$ days (in average) if the individual is in the ICU.

   - The parameter that governs the transition from $U$ to $HU$ is $u_{hu} = (1 - p_7)/t_8$, where $t_8$ days is the average time an individual needs to leave the ICU because he/she gets better.

8. Finally, an individual in $HU$ may get better and be discharged. This transit is modeled by the linear term $hu_a HU(t)$, where $h_{ua} = 1/t_9$, where $t_9$ is the average time to be discharged after leaving ICU.

Taking into account the above description of the populations and the model parameters, the following system of difference equations (7.1) describes the transmission dynamics of COVID-19 in the province of Granada (Spain) over time.

$$
\begin{array}{lcl}
S(t+1) & = & S(t)+q_s(t)-s_q(t)-\beta(t)S(t)\frac{I(t)}{P_T}, \\
Q(t+1) & = & Q(t)+s_q(t)-q_s(t), \\
L(t+1) & = & L(t)+\beta(t)S(t)\frac{I(t)}{P_T}-l_iL(t), \\
I(t+1) & = & I(t)+l_iL(t)-(i_r(t)+i_h(t)+i_u(t))I(t), \\
H(t+1) & = & H(t)+i_h(t)I(t)-(h_u(t)+h_f(t)+h_a(t))H(t), \\
U(t+1) & = & U(t)+i_u(t)I(t)+h_u(t)H(t)-(u_f(t)+u_{hu}(t))U(t), \\
HU(t+1) & = & HU(t)+u_{hu}(t)U(t)-hu_a(t)HU(t), \\
A(t+1) & = & A(t)+h_a(t)H(t)+hu_a(t)HU(t), \\
F(t+1) & = & F(t)+h_f(t)H(t)+u_f(t)U(t), \\
R(t+1) & = & R(t)+i_r(t)I(t).
\end{array}
\tag{7.1}
$$

Fig. 7.1 shows a flow diagram about how individuals may move throughout the different states with respect to the disease. Note that the right part of Fig. 7.1 (H, U, F, A, HU) represents the flow of patients within hospital departments. There is not movement between infectious (I) to deceased (F) as there are not data available and because, in Spain, most of the people who die by COVID do it in the hospital. Also, even though it seems that there are evidences of COVID re-infections [147], [148], they are very few and we are not going to consider it in our model.

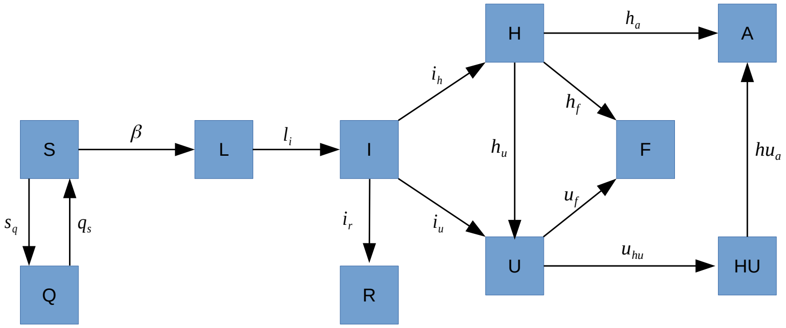


**Figure 7.1:** Flow diagram of the COVID-19 transmission dynamics. The boxes represent the states of individuals respect to the disease, and the letters next to the arrows represent the transition rates between states.

Since other known coronavirus have a seasonal behaviour [149], as a hypothesis, two main periods have been considered: virulence season (September to April) and non-virulence season (May to August). This division is supported by the available data related to number of hospitalizations and deaths. Due to this, some model parameter values will vary depending on the season.

### *7.2.2 Model calibration with uncertainty*

As we have mentioned before, we get hospitalization data from the Spanish province of Granada with the aim at calibrating the model parameters, in such a way that the model may describe the evolution of the pandemics in this area.

Hence, the model parameters $\beta$, $t_0$, $p_1$, $t_1$, $p_2$, $t_2$, $p_4$, $t_4$, $p_5$, $t_5$, $t_6$, $p_7$, $t_7$, $t_8$ and $t_9$ have to be calibrated in the intervals described in Table 7.2. Most of these intervals are established using the available data. Also, the initial number of latent and infectious to determine an appropriate initial condition the 1 March 2020. According to the data, at 1 March there were 5 people in the hospital circuit and none at ICU or recovered.

To calibrate the model, we consider four different transmission rates $\beta$: 1) before the lockdown; 2) after the lockdown; once the confinement has finished, 3) non-virulence season and 4) virulence season. Also, we take into account that the basic reproductive number of our model $R_0 = \beta/(i_r + i_u + i_h)$ is consistent with real values as $t$ goes on, time $t$ in days.

Our main source data (hospital's data) collects the number of daily hospitalizations, daily people in ICU, accumulated deaths and accumulated discharges in the hospitals of the province of Granada from 1 March to 22 September 2020. These data allows us to determine intervals where to search the model parameter values during the calibration.

In order to obtain suitable numerical values of the model parameters, the 10 and 90 percentiles of the stay time in each state of the hospital circuit (H, U, F, A, HU) are obtained and settled as the calibration range where looking for the proper values of the parameters. The intervals to determine the parameter

| Parameter | Transition | Time | Value |
|---|---|---|---|
| $s_q$ | $S \longrightarrow Q$ | 16 March<br>31 March | 700 000<br>150 000 |
| $q_s$ | $Q \longrightarrow S$ | 13 April<br>5 May to 21 June | 150 000<br>8 750 |
| $t_3$ | $I \longrightarrow R$ | All simulation | 14 |

**Table 7.1:** Summary of the known model parameters.

values, those given by the hospital's data and the others, can be seen in Table 7.2 under the column *Calibration range*.

Also, we are going to use the seroprevalence study [143], [144], which determines the percentage of accumulated infected people from the whole population. According to the prevalence study of COVID-19 in Spain, carried out between May and June 2020, the accumulated infected population in the province of Granada from the 27 April to 11 May is in the range $[1.7\%, 4.4\%]$, the range from 18 May to 1 June is $[2.3\%, 4.8\%]$ and the range from 8 June to 22 June is $[2.3\%, 5.5\%]$.

The model calibration with uncertainty is done applying the bootstrapping technique described in [150]. Mainly, this technique consists of performing a deterministic calibration, using the optimization algorithm NS developed in Chapter 2, to find model parameter values, in their corresponding calibration ranges in Table 7.2, in such a way that the model output is as close as possible to the data of hospitalized at ward, ICU, deceases and discharges, and the model output for recovered are inside the prevalence confidence interval provided by [143], [144]. The closeness is measured using the following fitting function:

| **Parameter** | **Transition** | **Time** | **Calibration range** |
|---|---|---|---|
| $I(0)$ | – | 1 March | $100 - 5000$ |
| $L(0)$ | – | 1 March | $100 - 5000$ |
| $\beta$ | $S \longrightarrow L$ | 1 March to 15 March<br>15 March to 4 May<br>virulence season<br>non-virulence season | $0.07 - 0.7$<br>$0.03 - 0.4$<br>$0.03 - 0.4$<br>$0.03 - 0.4$ |
| $t_0$ | $L \longrightarrow I$ | All simulation | $1 - 6$ |
| $p_1$ | $I \longrightarrow H$ | virulence season<br>non-virulence season | $0.01 - 0.04$<br>$0.01 - 0.04$ |
| $t_1$ | $I \longrightarrow H$ | All simulation | $9 - 12$ |
| $p_2$ | $I \longrightarrow U$ | virulence season<br>non-virulence season | $0.001 - 0.004$<br>$0.001 - 0.004$ |
| $t_2$ | $I \longrightarrow U$ | All simulation | $9 - 12$ |
| $p_4$ | $H \longrightarrow U$ | virulence season<br>non-virulence season | $0.000 - 0.139$<br>$0.000 - 0.119$ |
| $t_4$ | $H \longrightarrow U$ | virulence season<br>non-virulence season | $1 - 8.6$<br>$3.2 - 13.6$ |
| $p_5$ | $H \longrightarrow F$ | virulence season<br>non-virulence season | $0.097 - 0.297$<br>$0.018 - 0.218$ |
| $t_5$ | $H \longrightarrow F$ | virulence season<br>non-virulence season | $1 - 15$<br>$1 - 15.2$ |
| $t_6$ | $H \longrightarrow A$ | virulence season<br>non-virulence season | $3 - 15$<br>$4 - 29$ |
| $p_7$ | $U \longrightarrow F$ | virulence season<br>non-virulence season | $0.245 - 0.445$<br>No deaths, $p_7 = 0$ |
| $t_7$ | $U \longrightarrow F$ | virulence season<br>non-virulence season | $1 - 23.3$<br>– |
| $t_8$ | $U \longrightarrow HU$ | virulence season<br>non-virulence season | $6 - 36.2$<br>$4.7 - 80.1$ |
| $t_9$ | $HU \longrightarrow A$ | virulence season<br>non-virulence season | $4 - 25$<br>$2.7 - 18.9$ |

**Table 7.2:** Summary of the model parameters to be calibrate and their calibration ranges. The calibration ranges are obtained from the hospital's data.

$$\begin{aligned}
&10^4 \times d(\tfrac{R(t=\text{11 May})}{P_T}, [1.7\%, 4.4\%])+\\
&10^4 \times d(\tfrac{R(t=\text{1 Jun})}{P_T}, [2.3\%, 4.8\%])+\\
&10^4 \times d(\tfrac{R(t=\text{22 Jun})}{P_T}, [2.3\%, 5.5\%])+\\
&\textstyle\sum_{t=\text{1 Mar}}^{\text{22 Sep}} |H(t) - dH_t| + |U(t) - dU_t| + |A(t) - dA_t| + |F(t) - dF_t|,
\end{aligned} \tag{7.2}$$

where

$$d(p, [a, b]) = \begin{cases} 0 & a \leq p \leq b, \\ min(|p - a|, |p - b|) & \text{otherwise,} \end{cases}$$

measures the distance from a point to an interval, $R(t)$, $H(t)$, $U(t)$, $A(t)$, $F(t)$ are the model output corresponding to the number of recovered, hospitalized at ward, at ICU, discharges and deaths, respectively, at the day $t$, and $dH_t$, $dU_t$, $dA_t$, $dF_t$ are the data of hospitalized at ward, at ICU, discharges and deaths, respectively, at the day $t$.

The first three terms of the fitting function (7.2) guarantee that the model prevalence fits the data of the prevalence study [143], [144]. The last term of (7.2) guarantees that the number of hospitalized at ward, at ICU, discharges and deaths returned by the model are close to their corresponding data.

Once the model is calibrated, we calculate the difference between the model output and the data for each time instant (residuals). Then, we estimate a probability distribution of the residuals. Thus, we sample 1 000 times the residuals' probability distribution, we sum up them to the data generating 1 000 sets of perturbed data, describing the data uncertainty. Then, the model is deterministically calibrated, using NS for PSO, for the 1 000 sets of perturbed data. For all the model parameter values obtained we calculate the mean and the percentiles 2.5 and 97.5 (95% confidence interval) describing the uncertainty of the model parameters. Also, for the model outputs, calculating the mean and the 95% confidence intervals for each time instant, we obtain the 95% confidence bands of the evolution of the subpopulations over time.

### *7.2.3 Vaccination*

In Spain, the vaccination started 27 December 2020. In 23 February 2021, 3 million doses have been administered and 1.2 million people have received the two doses [151]. In Granada, these figures correspond to an average of 526 people vaccinated every day from the beginning of the vaccination campaign. The Spanish government expect to vaccinate 70% of people before the end of summer [152]. This imply to increase by six the number of daily vaccinated.

In order to simulate the effect of the vaccine in the population of the Spanish region of Granada, a new state has to be considered in the mathematical model:

- **(V)** vaccinated, when the individual is vaccinated and the vaccine is effective, protecting the individual.

We suppose that only susceptible people are vaccinated. We assume that the vaccine protects against the most dangerous symptoms of the COVID-19 and blocks the spread of the disease. The transit from $S$ to $V$ is modelled by the term $s_v(t)$ and the transit from $V$ to $S$, it is modelled by the term $v_s(t)$. Both parameters describe the number of people who are moving from one state to other, at time $t$.

The system of difference equations, which includes the vaccination of the population is shown in (7.3). In black, the new terms.

$$\begin{array}{lcl}
S(t+1) & = & S(t)+q_s(t)-s_q(t)+\mathbf{v_s(t)}-\mathbf{s_v(t)}-\beta(t)S(t)\frac{I(t)}{P_T}, \\
Q(t+1) & = & Q(t)+s_q(t)-q_s(t), \\
\mathbf{V(t+1)} & = & \mathbf{V(t)}+\mathbf{s_v(t)}-\mathbf{v_s(t)}, \\
L(t+1) & = & L(t)+\beta(t)S(t)\frac{I(t)}{P_T}-l_iL(t), \\
I(t+1) & = & I(t)+l_iL(t)-(i_r(t)+i_h(t)+i_u(t))I(t), \\
H(t+1) & = & H(t)+i_h(t)I(t)-(h_u(t)+h_f(t)+h_a(t))H(t), \\
U(t+1) & = & U(t)+i_u(t)I(t)+h_u(t)H(t)-(u_f(t)+u_{hu}(t))U(t), \\
HU(t+1) & = & HU(t)+u_{hu}(t)U(t)-hu_a(t)HU(t), \\
A(t+1) & = & A(t)+h_a(t)H(t)+hu_a(t)HU(t), \\
F(t+1) & = & F(t)+h_f(t)H(t)+u_f(t)U(t), \\
R(t+1) & = & R(t)+i_r(t)I(t).
\end{array} \tag{7.3}$$

Not every vaccinated person acquires protection against the disease. We consider the vaccine effectiveness as the percentage of vaccinated people who acquire immunity. Also, the duration effect of the vaccine is unknown, and we will assume a permanent immunity for effective vaccinations, then $v_s(t) = 0$ for all $t$. In [153], the authors say that the vaccine effectiveness is 95%. However, in order to take into account possible mutations that may affect the effectiveness, we will consider 75% vaccine effectiveness. Furthermore, the first vaccination dose provides 33% effectiveness [154].

Here, we are going to perform three simulations. The simulations have been done 10 March 2021. In the three, we vaccinate at the current pace, with 75% effectiveness, from 27 December 2020 until 31 March. From 1 April,

1) we keep vaccinating at the same pace and no new population restrictions are applied;

2) the same as 1) applying stronger restrictions from 1 to 15 April;

3) the same as 2) and the vaccination pace increases to fulfill 70% coverage the 31 August 2021.

## 7.3 Results

### *7.3.1 Calibration*

For calibration with uncertainty, bootstrapping technique [150] is applied. Technical details can be seen in the Appendix of this chapter. This way, we are able to capture and quantify the uncertainty in the evolution of the disease. The results of the calibration are good enough to represent the effects of the pandemics in Granada during 1 March - 22 September 2020 period (Fig. 7.2).

The mean and confidence interval of all the obtained calibrated parameters are shown in Table 7.3. The availability of reliable hospital's data allows us to provide realistic intervals to calibrate the model parameters and, once the model is calibrated, meaningful model parameter values describing the transitions among states of the disease.

In 1 March 2020 (initial condition) the calibrated number of infectious is 1192, CI95% $1166 - 1226$, (0.13%, CI95% 0.125% – 0.133%) and latent 574, CI95% $559 - 589$, (0.062%, CI95% 0.060% – 0.064%). Data provided by the hospitals in the province of Granada, from 1 March to 15 March, compared with the registry of the previous years did not show an increase of respiratory disease cases that could be attributable to COVID-19, what it is in accordance with our estimation of the low number of COVID-19 patients at the beginning of the pandemic (1 March 2020). Later, the number of infected patients experienced a great increase, what supports that the effective reproductive number $R_0$ was initially very high, (5.44, CI95% $5.34 - 5.61$), value which is in accordance with Sanche et al [127].

After the beginning of the lockdown declared by the Spanish government in 16 March, the number of infected kept growing as a consequence of the contagions produced in the previous weeks with high transmission rate $\beta$. Nevertheless, the basic reproductive number $R_0$ decreased to 0.55, CI95% $0.50 - 0.60$, not only because of lockdown but also the social distancing, the use of face-masks and other population measures, reducing the $\beta$, as it can be seen in Fig. 7.2(F).

| **Parameter** | **Transition** | **Time** | **Calibrated value**<br>**Mean (2.5%CI - 97.5%CI)** |
|---|---|---|---|
| $I(0)$ | – | 1 March | 1192 (1166 – 1226) |
| $L(0)$ | – | 1 March | 574 (559 – 589) |
| $\beta$ | $S \longrightarrow L$ | 1 March to 15 March<br>15 March to 4 May<br>virulence season<br>non-virulence season | 0.391 (0.384 – 0.403)<br>0.039 (0.036 – 0.043)<br>0.125 (0.114 – 0.136)<br>0.100 (0.094 – 0.104) |
| $t_0$ | $L \longrightarrow I$ | All simulation | 3.8 (3.5 – 4.0) days |
| $p_1$ | $I \longrightarrow H$ | virulence season<br>non-virulence season | 0.027 (0.026 – 0.029)<br>0.009 (0.006 – 0.013) |
| $t_1$ | $I \longrightarrow H$ | All simulation | 11.5 (10.8 – 11.9) days |
| $p_2$ | $I \longrightarrow U$ | virulence season<br>non-virulence season | 0.00239 (0.00206 – 0.00278)<br>0.00060 (0.00045 – 0.00075) |
| $t_2$ | $I \longrightarrow U$ | All simulation | 10.5 (9.5 – 11.6) days |
| $p_4$ | $H \longrightarrow U$ | virulence season<br>non-virulence season | 0.019 (0.018 – 0.021)<br>0.018 (0.017 – 0.020) |
| $t_4$ | $H \longrightarrow U$ | virulence season<br>non-virulence season | 4.3 (4.3 – 4.4) days<br>8.3 (8.2 – 8.4) days |
| $p_5$ | $H \longrightarrow F$ | virulence season<br>non-virulence season | 0.136 (0.135 – 0.137)<br>0.075 (0.071 – 0.084) |
| $t_5$ | $H \longrightarrow F$ | virulence season<br>non-virulence season | 8.6 (8.5 – 8.7) days<br>7.5 (7.4 – 7.6) days |
| $t_6$ | $H \longrightarrow A$ | virulence season<br>non-virulence season | 13.6 (13.5 – 13.7) days<br>13.2 (13.2 – 13.3) days |
| $p_7$ | $U \longrightarrow F$ | virulence season<br>non-virulence season | 0.248 (0.243 – 0.254)<br>0 |
| $t_7$ | $U \longrightarrow F$ | virulence season<br>non-virulence season | 13.7 (13.6 – 13.8) days<br>– |
| $t_8$ | $U \longrightarrow HU$ | virulence season<br>non-virulence season | 15.0 (14.8 – 15.1) days<br>19.9 (19.3 – 20.6) days |
| $t_9$ | $HU \longrightarrow A$ | virulence season<br>non-virulence season | 13.7 (13.6 – 13.7) days<br>11.0 (10.9 – 11.1) days |

**Table 7.3:** Summary of the calibrated model parameter values and their CI95%.

During the non-virulence season after the lockdown is finished, May - August, the incidence of the disease kept low, with a stabilization of the number of hospitalizations and deceases. The $R_0$ value was low, 1.40, CI95% 1.31 – 1.46, but it increases until slightly higher values than at the beginning of the virulence season ($R_0 = 1.74$, CI95% 1.59 – 1.89). This can be observed with the increase of hospitalized and intensive care unit people, together with the end of the stagnation in the number of deceased and discharged in September on Figure 7.2 (leftmost part of the figure).

Different calibrated model parameter values can be seen in Table 7.3 for virulence and non-virulence seasons. More precisely, the percentage of infected people who needs hospitalization (hospitalization at ward $p_1$, or ICU treatment $p_2$). During the virulence season, about 2.7% of the infected people need hospitalization at ward, and 0.24% of the infected people enter the ICU directly without passing through the hospitalization ward. During the non-virulence season, about 0.9% of the infected people need hospitalization at ward, and 0.06% of the infected people enter the ICU directly without passing through the hospitalization ward. Analogous differences can be seen in probability parameters $p_5$ and $p_7$ where in the non-virulence season the probability to get worse is significantly lower.

Furthermore, in Figure 7.3, we can see the evolution of the infection fatality ratio (IFR), that is, the proportion of infected people who die (accumulated). This is a measure of the disease severity [155]. After April 2020, the IFR is decreasing until October, corresponding to the non-virulence season.

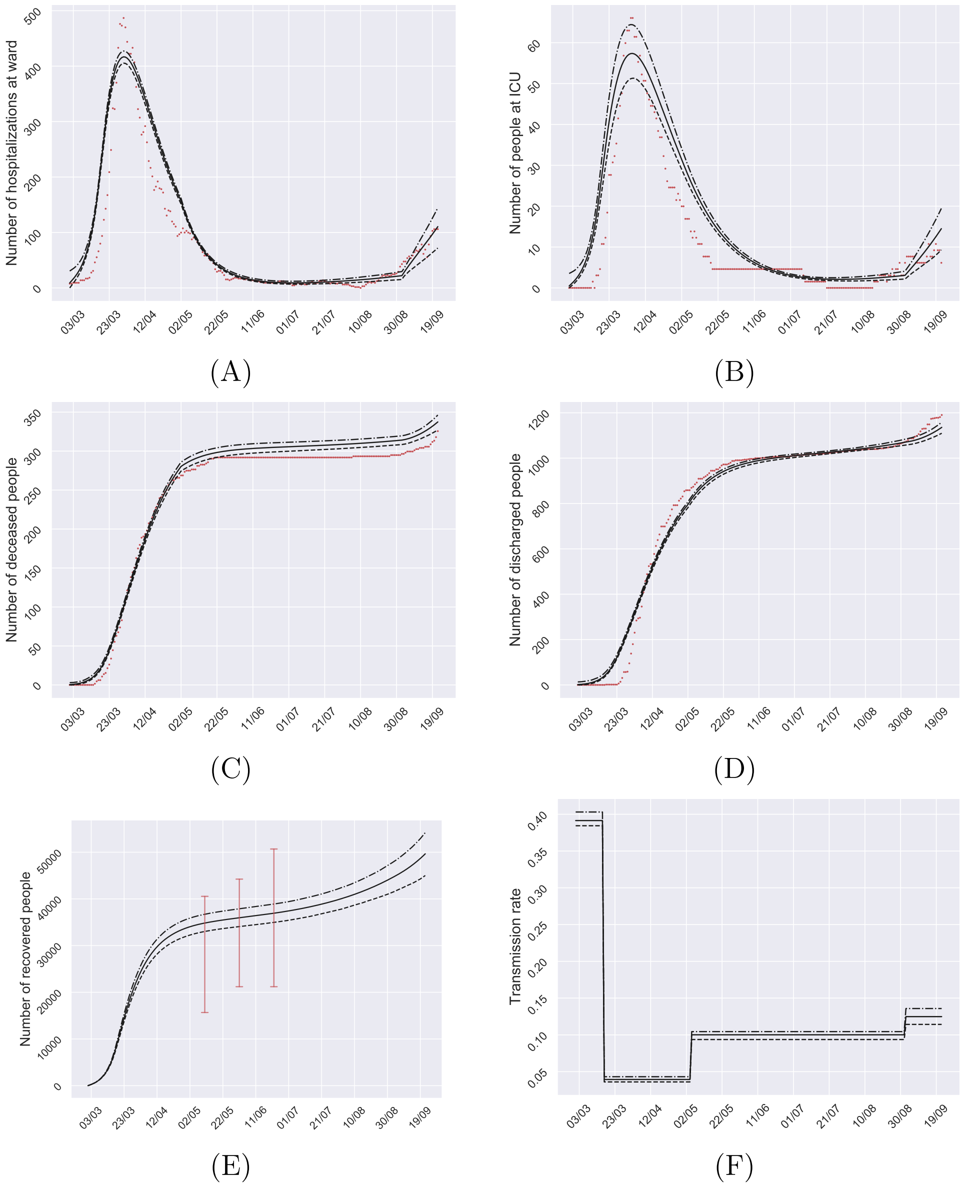


**Figure 7.2:** Model calibration. (A) Number of daily hospitalized at ward people. (B) Number of daily people in ICU. (C) Number of accumulated deceased people. (D) Number of accumulated discharged people. (E) Number of accumulated recovered people. (F) Transmission rate $\beta$. The black bands are the mean and the CI95%. The red points are the data from the hospitals. Vertical red lines in (E) show the confidence intervals of the prevalence in Granada given by [143], [144].

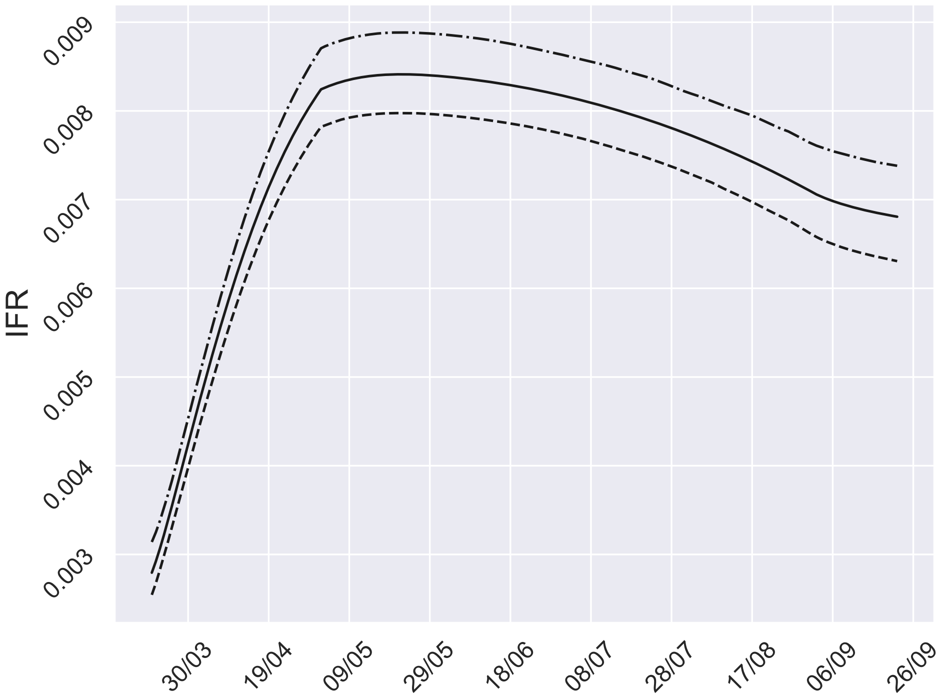


**Figure 7.3:** Infection Fatality Ratio (IFR). Evolution of the IFR over time in Granada. Observe that IFR is decreasing in the non-virulence season (May-September).

### 7.3.2 Model validation

In order to validate the calibration of the model, data from 23 September 2020 until 26 January 2021 have been collected with the aim at comparing the error between the model evolution and the real data not used for calibration. The results can be seen in Figure 7.4. The model 95% confidence bands capture most of the data after 23 September (red vertical line) taking into account that this is the infectious disease most intervened in history with interventions whose actual effect is unknown or contains a lot of uncertainty. In fact, the model is able to predict the evolution of the second and the beginning of the third wave.

### 7.3.3 Vaccination simulation

Once the model has been calibrated and validated, we can simulate with reliability possible scenarios for the future. In Figure 7.5 we can see the result of the simulation scenarios: 1) the current situation does not change; 2) scenario 1) with stronger restrictions from $1-15$ April; 3) scenario 2) and 70% of population is vaccinated at the end of Summer with an effectiveness of 75%.

The Figure 7.5 shows that the 3rd wave is finishing and the 4th wave will start soon because at the beginning of March the restrictions have been relaxed. However, the peak of this 4th wave may be high if we are not careful satisfying the population measures (solid black line in Figure 7.5). The application of strong restrictions again or the increase of the pace of vaccination are the strategies to avoid a new high wave (dotted and dashed lines in Figure 7.5) and quickly starts to decline the number of infected. Similar effects can be seen in hospitalizations and ICU graphs. Also, in the Figure 7.5(D), a stabilization in the deaths can be seen around May.

As we can see, in the three cases, when next May starts, we enter in the non-virulence season and the number of infected people declines. In the scenarios 1) and 2) where the vaccination pace does not increase, the decline may have new waves for hospitalized at ward and ICU next October. Although this wave is very low for infected, as we are in the virulence season, the transition parameters to subpopulations $H$ and $U$ are significantly higher, explaining why a very small wave in infected people produce noticeable waves in hospitalized at ward and ICU. In the scenario 3) the decline is very fast, keeping it low without new waves.

This general decline happens because the percentage of immunized people (recovered+vaccinated), appearing in the last row (% immunized) of each scenario in Table 7.4, is close or greater than the percentage of immunized people required for the herd immunity, given by $H = 100(1 - 1/R_0)$, rows $H$ in Table 7.4.

It is noticeable that in September, starting the next virulence season, the percentage of immunized people is high enough to avoid a new infected wave (see Figure 7.5) if we are careful and we satisfy non-pharmaceutical measures as the use of face-masks, social distancing, etc. Especial mention deserves the scenario 3, because in 1 September the percentage of immunized people is so high that the return to normal life may be very likely.

Although these results seem to be optimistic, the social distancing, face-masks and other non-pharmaceutical measures cannot be removed suddenly. Otherwise, the transmission rate increases, also $R_0$, and a new outbreak (wave) may arise when virulence season comes again, except, maybe, in the scenario 3 where most of the people is immunized, mainly, because the increase of the vaccinated pace.

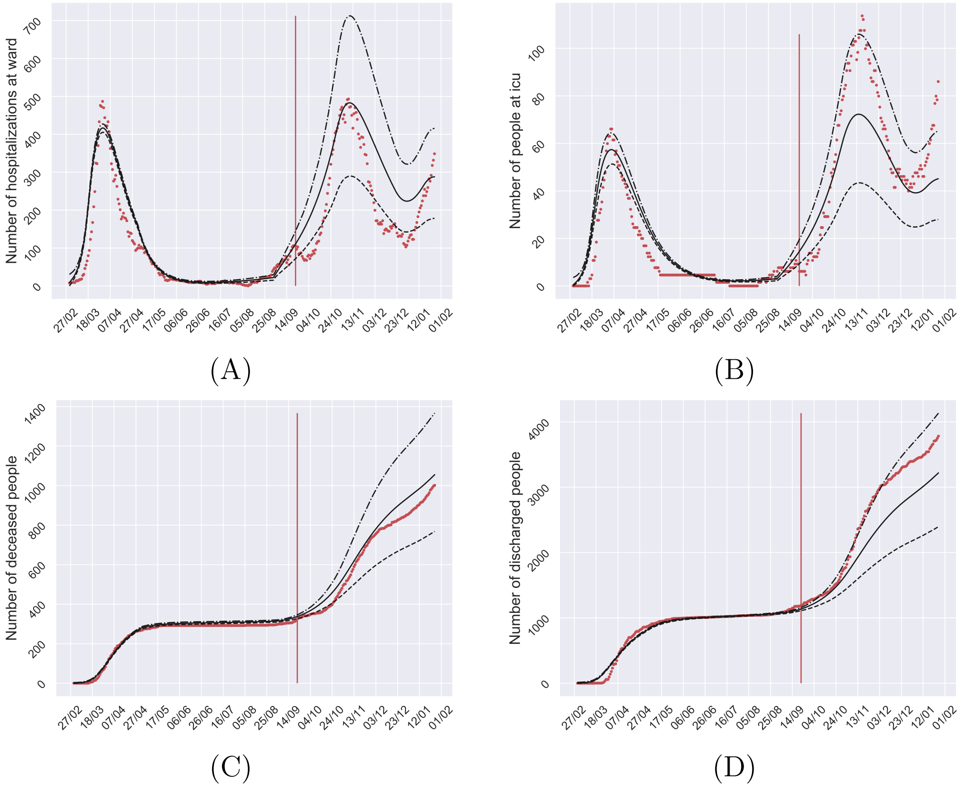
The simulations shown in Figure 7.5 may help policy makers in hospitals and public health to plan the use and distribution of resources to face the coming months.

| **Scenario 1** | | |
|---|---|---|
| | 1 May 2021 | 1 Sep 2021 |
| $R_0$ | 1.40, (1.31 – 1.46) | 1.74, (1.59 – 1.89) |
| $H$ | **28.47**%, (23.66% – 31.51%) | **42.53**%, (37.11% – 47.09%) |
| Recovered | 164902, (121916 – 212431) | 223362, (168933 – 275884) |
| Vaccinated | 77948 | 170135 |
| Immunized | 242850, (199864 – 290379) | 393497, (339068 – 446019) |
| % immunized | **26.35**%, (21.69% – 31.51%) | **42.70**%, (36.79% – 48.40%) |

| **Scenario 2** | | |
|---|---|---|
| | 1 May 2021 | 1 Sep 2021 |
| $R_0$ | 1.40, (1.31 – 1.46) | 1.74, (1.59 – 1.89) |
| $H$ | **28.47**%, (23.66% – 31.51%) | **42.53**%, (37.11% – 47.09%) |
| Recovered | 162738, (120647 – 209646) | 204643, (154719 – 253861) |
| Vaccinated | 77948 | 170135 |
| Immunized | 240685, (198594 – 287593) | 374779, (324855 – 423997) |
| % immunized | **26.12**%, (21.55 % – 31.21%) | **40.67**%, (35.25 % – 46.01%) |

| **Scenario 3** | | |
|---|---|---|
| | 1 May 2021 | 1 Sep 2021 |
| $R_0$ | 1.40, (1.31 – 1.46) | 1.74, (1.59 – 1.89) |
| $H$ | **28.47**%, (23.66% – 31.51%) | **42.53**%, (37.11% – 47.09%) |
| Recovered | 162737, (120641 – 209647) | 185675, (138932 – 235429) |
| Vaccinated | 113917 | 648532 |
| Immunized | 276654, (234558 – 323564) | 834207, (787464 – 883961) |
| % immunized | **30.02**%, (25.45 % – 35.11 %) | **90.53**%, (85.45 % – 95.93%) |

**Table 7.4:** Comparison between the required percentage to reach the herd immunity ($H$) with the percentage of immunized people the 1 May and the 1 September (% immunized). Observe that the percentage of immunized people is very close to the required in Scenarios 1 and 2, and greater in Scenario 3. This explains why, after 1 May, there is a decline in the number of infected, more significant in scenario 3 where the vaccination pace increases.



(A)

(B)

(C)

(D)

**Figure 7.4:** Model validation. (A) Number of daily hospitalized at ward people. (B) Number of daily people in ICU. (C) Number of accumulated deceased people. (D) Number of accumulated discharged people. The vertical red lines (23 September 2020) divide the calibration from the validation. The black bands are the mean and the CI95% of the model outputs. The red points are the data collected from the hospitals. The model captures most of the data red points after 23 September, predicting the evolution of the second and the third wave.

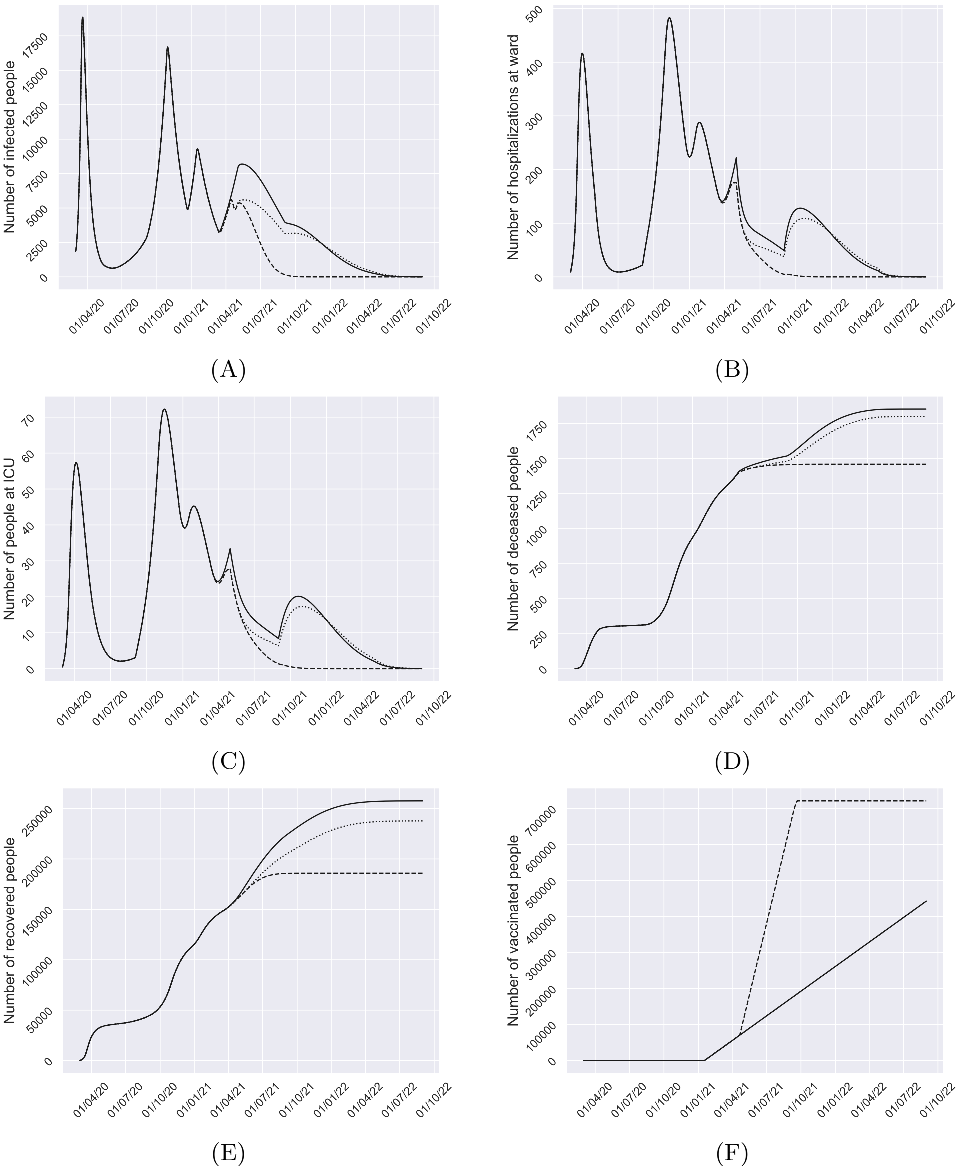


**Figure 7.5:** Average COVID-19 evolution in the three proposed scenarios with 70% of vaccine effectiveness. Scenario 1, continuous line. Scenario 2, dotted line. Scenario 3, dashed line. (A) Number of infected people. (B) Number of daily hospitalized at ward people. (C) Number of daily people in ICU. (D) Number of accumulated deaths. (E) Number of recovered people. (F) Number of vaccinated people. The wave in October 2021 in hospitalized and ICU appears because we are in the virulence season and the probability to be hospitalized or at ICU increases and the time to enter decreases.

## 7.4 Discussion

Explanatory mathematical models of infectious diseases are intended to allow to predict the evolution of the disease considering a wide range of possible scenarios. They should be able to estimate the healthcare demands that will have to be covered by hospitals early enough as to implement management and operative adaptations regarding multiple aspects, including the extension of ICU to other places that could be equipped and be used for the management of the COVID-19 patients, as well as the reservation of out-of-hospital stays, such as hotels, for health personnel or even for patients with low severity who must be lockdown in a controlled environment [156].

A direct and permanent line of communication is important for the transmission of data in real time between the heads of hospital centres and the teams in charge of analysing the epidemiological evolution of the disease, especially through tools such as the one presented in this chapter. This will ensure the optimization of future predictions on the basis of the augmentation of the knowledge database, contributing to face the fluctuations that may appear in the transmission of the virus depending on variables that are not sufficiently clarified at the moment.

One of the strong aspects of our model is that it is possible to calibrate and validate from data collected and reported through an informative protocol agreed at the province of Granada, which ensures homogeneity as well as prevents biases that may appear on the count of each type of subject, consideration of infected by diagnostic, etc. Also, the model has been fed regularly with new data to update the predictions.

Good planning supported by models like the one we propose permits better distribution of the hospital resources, preparing a COVID-19 area without reducing the care attention to patients with other diseases, preventing the collapse of the health services, as happened last March-April 2020 and January-February 2021.

Our simulations show that we have to keep population measures and/or increase the vaccination pace in order to control the pandemic and avoid new concerning outbreaks. However, a decline in the number of infectious is expected to start in May without a new increase in September when the virulence season arrives if we are able to maintain population non-pharmaceutical measures as the use of face-masks, social distancing, etc.

The transmission capacity of the virus at the beginning of the pandemic was very high, with $R_0$ greater than 5. Numerous studies have stated the effectiveness of non-pharmaceutical measures in reducing the transmission of respiratory virus [157]–[166]. Then, $R_0$ decreased significantly, reaching an average of 1.4 and making it possible to control the pandemic by approaching to 1. The return to normal life as it was before the pandemics, is not to be expected in the short term, because our simulations assume low transmission rates due to the social distancing, face-masks, restricted schedules, curfew and other non-pharmaceutical measures. However, we expect these non-pharmaceutical measures will be relaxed as the immunized people increases. Nevertheless, if the rate of vaccination grows sufficiently, it is possible that the return to normal life becomes very likely.

## 7.5 Conclusions

Mathematical models allow us to make predictions that may provide enough time to the health systems to establish the appropriate measures to face outbreaks that could collapse human and material resources. Moreover, future pharmaceutical and non-pharmaceutical interventions, as well as aspects related to SARS-CoV-2 biology and behaviour, can be represented through the model parameters to act as modulator of the transmission rate $\beta$ what provide the required flexibility to adapt the model throughout the time.

The modified SEIR model presented in this paper can be a useful tool for providing insight into the transmission dynamics of SARS-CoV-2. It considers the circuit of patients moving inside the hospital dependencies, returning a more precise portrait of the use of hospital resources to foresee future needs.

Once the model is calibrated and validated, three scenarios are simulated to evaluate the evolution of COVID-19 in Granada. The simulations show that we are finishing the 3rd wave but the 4th wave is around the corner reaching the peak in May 2021. In the worst case, the 4th wave seems to be a bit lower than the previous ones.

In any case, a decline in the number of infected people is expected in all the simulated scenarios, starting in May and maintained when we change to the virulence season next September. This is possible if we keep lower transmission rates, and consequently low $R_0$, due to the social distancing and other non-pharmaceutical measures. That is, we will not be out of danger yet. Nevertheless, we will be able to resume the life as we knew it before the pandemic relaxing measures of social situation and face-masks in a gradual and controlled way to avoid new outbreaks as the percentage of immunized people increases.

### *Limitations of the model*

The proposed model is a classical system of difference equations. For its building, we assume usual hypotheses as the homogeneous mixing of the population (any individual may infect any individual). No age groups are considered, and the hospitalization and decease rate may vary depending on the age of the individuals. Also, we assume that all lockdown people are susceptible, when latent, asymptomatic infectious and recovered may also be in lockdown. Furthermore, mobility and other spatial aspects are not considered.

We assume that there is no re-infection or is not significant and the vaccine blocks the transmission of COVID-19. Furthermore, we assume a permanent protection for those who the vaccine is effective. If these hypotheses were eventually not true, an increase in the number of cases could be seen next autumn.

It is important to consider that the model must be particularized for each population, either provincial or larger, since the transmission rate must respond to the effect of health policy measures and the response of the population to the recommendations, such as opening educational centres, social distancing, use of face-masks, adaptation of public and private services to establish physical barriers that hinder the spread of the virus, etc.

## Appendix

Taking into account the general procedure presented in [150], we study error terms for the estimated parameters and resample these error terms using bootstrapping. Then, we obtain new data by adding the resampled error to the output model. For each new data, we estimate the parameters.

This is based on the bootstrapping method, more specifically in residual bootstrapping, which is a commonly used and robust method, that can deal dynamical systems, which in turn often violate the standard assumptions for residuals of traditional methods such as normality, independence, homoscedasticity and not autocorrelation.

To do this, we implement the following steps. First we calibrate the model to the data, thus obtaining a first estimate of the parameters. Then, residuals are obtained from the difference of the outputs of the model fitted to the data and the real ones. In Figure 7.6 we can see the residuals for ICU (solid line).

Now, we apply the Pearson's correlation coefficient in order to analyse the correlation between the error terms. The results appear in Table 7.5. We reject the null hypothesis if the p-value is less than $\alpha = 0.05$.

| | Coefficient | P-value | | Coefficient | P-value |
|---|---|---|---|---|---|
| $\rho_{12}$ | 0.693 | 0 | $\rho_{23}$ | 0.019 | 0.782 |
| $\rho_{13}$ | -0.213 | 0.00215 | $\rho_{24}$ | -0.122 | 0.08 |
| $\rho_{14}$ | -0.115 | 0.00769 | $\rho_{34}$ | 0.3164 | 0 |

**Table 7.5:** Pearson correlation coefficient and the corresponding p-values.

We would like to comment about the correlation coefficient $\rho_{12} = 0.693$ in Table 7.5. It denotes a moderate correlation, however, in order to simplify the analysis and assuming the corresponding errors, we are going to ignore this moderate correlation.

In the next step, we check the assumptions for residuals of asymptotic methods by computing a hypothesis test, where the purpose of it is whether the autocorrelation values are significantly different from zero or not. In other words, this non-parametric test is used to check the hypothesis that the elements of a sequence are mutually independent. For that, we figure out the autocorrelation function values and their variances in the residual terms for different lags $k$, according to the formulas in [150]. In all the cases (people hospitalized at ward, people at ICU, discharged people and deceased people) there is significant autocorrelation for lags $k = 6$, $k = 9$, $k = 11$ and $k = 10$, respectively.

Accordingly, the auto-dependency of residuals can be described using autoregressive model (AR) for each residual term. Hence, we can estimate the underlying white noise by subtracting the autocorrelation from the errors. In Figure 7.6 (dashed line) we can see the AR for ICU.

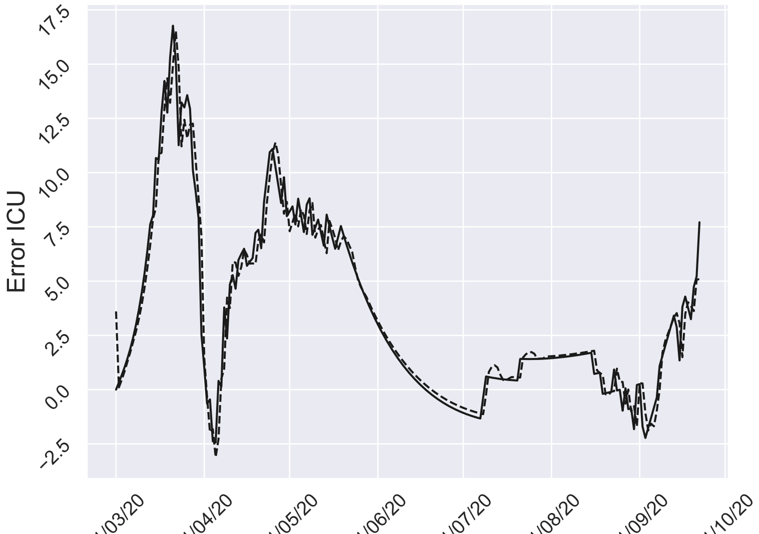


**Figure 7.6:** Residuals of ICU. Solid line are the actual residuals and the dashed line is the AR approximation.

In the following step, we assess the independence of the underlying white noise by autocorrelation and normality test. With regard to the former, we implement the same procedure as before and the test statistic is not significant for all lags in the four cases. This means that there is no significant autocorrelation in the underlying noise, that is, independence. Regarding the normality test, Kolmogorov-Smirnov test is used to make a determination as to whether the distribution of underlying white noise matches the characteristics of a normal distribution. In all cases, the underlying error terms are not normally distributed. This fact is reflected in the Figure 7.7.

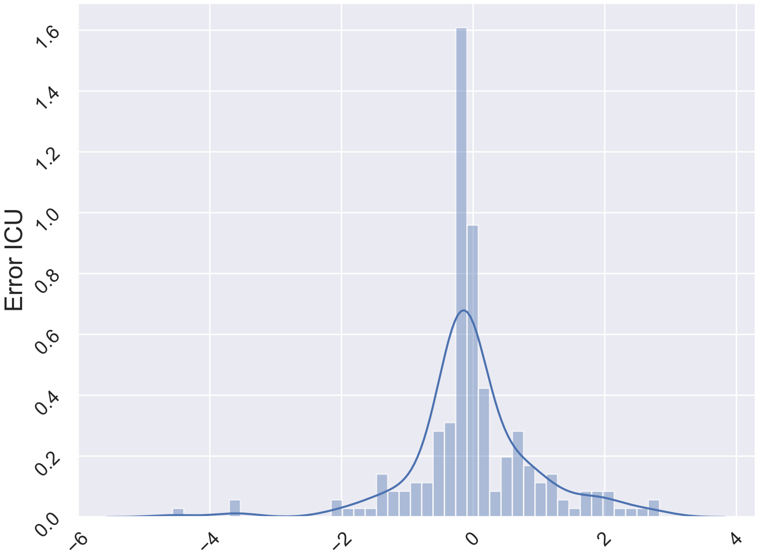


**Figure 7.7:** White noise histogram of AR model for ICU residual terms.

In the next step we resample this white noise using a nonparametric bootstrapping, since it is not normally distributed, to obtain new residual values for each time instant. In this technique, each new set of residuals is generated by resampling randomly, with replacement, from the original white noise dataset. In total, we use 1 000 bootstraps. These sets are added to the data, thereby creating 1 000 new data sets. Then, new model parameter values are obtained by calibrating the model to each one of the 1 000 new data sets. At the end of the process, we have 1 000 estimates for the parameters, which can be used to construct the empirical distribution.

## Chapter published

The results of this chapter have been published in [167].

# Chapter 8

# Conclusion

This thesis has explored the performance of applying Novelty Search to the Particle Swarm Optimization algorithm (PSO), showing that this paradigm is able to improve metaheuristic approaches when functions are complex, this is, with a high nonlinear behaviour. The new algorithm, Novelty Swarm (NS), has been used to calibrate the value of the parameters of several mathematical models which include a wide number of techniques, showing its ability to capture the reality of the data. NS not only has been able to solve deterministic problems, but also has managed uncertainty outstandingly.

Chapter 2 has proved that NS has a better performance than original PSO, PSO variations, jSO and LSHADE in functions with a high nonlinear behaviour at expenses of higher computational cost.

Chapter 3 shows that NS is able to solve the design of an internal combustion engine, improving its fuel consumption at the same time that pollutant emissions are reduced. This is remarkable, taking in account the computational effort of the engine model, which only allowed to perform 1000 model evaluations during the optimization process.

In chapter 4, we substitute the optimization engine of Grammatical Evolution with the NS algorithm, and we test this new procedure for the prediction of energy demand estimation in Spain and France. The results show that the modified algorithm has a better performance in terms of error and robustness.

In chapter 5, we build a model which describes the dynamics of bladder cancer. We show the ability of describing in a proper way the evolution of the illness in a specific patient after calibrating the parameters of the model with NS. This could improve the treatment of the cancer thanks to the schedule of the revisions.

Chapter 6 describes the dynamics of COVID-19 in Spain with a random network, and after estimating the parameters of the network based on published data, we conclude that democratic antivirals, with low effectiveness, could reduce the propagation of the pandemics. Although we have not used the NS algorithm in this chapter, the study made of the COVID-19 disease has allowed us to develop the model described in chapter 7.

In chapter 7, we also describe the dynamics of COVID-19, but instead of using a random network describing the Spanish population, we have used a system of difference equations including the hospital patients circuit, and we have calibrated it to the population of Granada. During the calibration, the uncertainty of the evolution of the disease is taken in account mixing the bootstrapping technique with the NS algorithm. After including a flux of constant information with the Hospital Universitario Virgen de las Nieves, we have been able to predict the saturation of the hospital resources of the province, which has helped us to improve the logistics of the sanitary resources.

Summing up, this thesis has shown the ability of the mathematical models and of the optimization algorithms to describe in a proper level of detail, different natural phenomena. Uncertainty has proved to increase the amount of information, and the use of NS with models which include uncertainty is a good approach to study several problems related with engineering and health care.

# Bibliography

[1] J. Kennedy and R. Eberhart, “Particle Swarm Optimization”, in *Proceedings of ICNN 95 - International Conference on Neural Networks*, IEEE, 1995. DOI: 10.1109/icnn.1995.488968 (cit. on pp. 5, 10, 52).

[2] W. Liu, Z. Wang, X. Liu, N. Zeng, and D. Bell, “A Novel Particle Swarm Optimization Approach for Patient Clustering from Emergency Departments”, *IEEE Transactions on Evolutionary Computation*, 2018, ISSN: 1089-778X. DOI: 10.1109/TEVC.2018.2878536 (cit. on p. 5).

[3] M. R. AlRashidi and M. E. El-Hawary, “A Survey of Particle Swarm Optimization Applications in Electric Power Systems”, *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 4, pp. 913–918, Aug. 2009, ISSN: 1089-778X. DOI: 10.1109/TEVC.2006.880326 (cit. on p. 5).

[4] H. B. Nguyen, B. Xue, and P. Andreae, “PSO with surrogate models for feature selection: static and dynamic clustering-based methods”, *Memetic Computing*, vol. 10, no. 3, pp. 291–300, 2018. DOI: 10.1007/s12293-018-0254-9 (cit. on p. 5).

[5] M. R. Bonyadi and Z. Michalewicz, "Impacts of coefficients on movement patterns in the particle swarm optimization algorithm", *IEEE Transactions on Evolutionary Computation*, vol. 21, no. 3, pp. 378–390, Jun. 2017, ISSN: 1089-778X. DOI: `10.1109/TEVC.2016.2605668` (cit. on p. 5).

[6] V. Kadirkamanathan, K. Selvarajah, and P. J. Fleming, "Stability analysis of the particle dynamics in particle swarm optimizer", *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 3, pp. 245–255, Jun. 2006, ISSN: 1089-778X. DOI: `10.1109/TEVC.2005.857077` (cit. on p. 5).

[7] P. Novoa-Hernández, C. C. Corona, and D. A. Pelta, "Efficient multi-swarm PSO algorithms for dynamic environments", *Memetic Computing*, vol. 3, no. 3, p. 163, 2011. DOI: `10.1007/s12293-011-0066-7` (cit. on p. 5).

[8] J. Gou, Y.-X. Lei, W.-P. Guo, C. Wang, Y.-Q. Cai, and W. Luo, "A novel improved particle swarm optimization algorithm based on individual difference evolution", *Applied Soft Computing*, vol. 57, pp. 468–481, 2017. DOI: `10.1016/j.asoc.2017.04.025` (cit. on pp. 5, 8).

[9] J. Lehman and K. O. Stanley, "Exploiting open-endedness to solve problems through the search for novelty", *Artificial Life - ALIFE*, Jan. 2008 (cit. on p. 6).

[10] ——, "Abandoning Objectives: Evolution Through the Search for Novelty Alone", *Evolutionary Computation*, vol. 19, pp. 189–223, Jun. 2011. DOI: `10.1162/EVCO_a_00025` (cit. on p. 6).

[11] ——, "Efficiently evolving programs through the search for novelty", in *Proceedings of the 12th annual conference on Genetic and Evolutionary Computation*, ACM Press, 2010. DOI: `10.1145/1830483.1830638` (cit. on p. 6).

[12] ——, "Novelty search and the problem with objectives", in *Genetic Programming Theory and Practice IX*. Springer New York, 2011, pp. 37–56, ISBN: 978-1-4614-1770-5. DOI: `10.1007/978-1-4614-1770-5_3` (cit. on p. 6).

[13] J. Chen, Z. Ren, and X. Fan, "Particle swarm optimization with adaptive mutation and its application research in tuning of PID parameters", in *2006 1st International Symposium on Systems and Control in Aerospace and Astronautics*, IEEE, 2006. DOI: `10.1109/isscaa.2006.1627490` (cit. on pp. 7, 15, 17).

[14] P. N. Suganthan, N. Hansen, J. J. Liang, *et al.*, "Problem definitions and evaluation criteria", in *CEC 2005 Special Session on Real Parameter Optimization*, May 2005. [Online]. Available: `http://www.ntu.edu.sg/home/epnsugan/index%5C_files/cec-05/cec05.htm` (cit. on pp. 7, 15, 16, 18, 27).

[15] H. Wang, L. L. Zuo, J. Liu, W. J. Yi, and B. Niu, "Ensemble Particle Swarm Optimization and differential evolution with alternative mutation method", *Natural Computing*, Sep. 2018. DOI: `10.1007/s11047-018-9712-z` (cit. on p. 7).

[16] R. A. Sarker, S. M. Elsayed, and T. Ray, "Differential Evolution With Dynamic Parameters Selection for Optimization Problems", *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 5, pp. 689–707, Oct. 2014. DOI: `10.1109/tevc.2013.2281528` (cit. on p. 7).

[17] M. G. Epitropakis, V. P. Plagianakos, and M. N. Vrahatis, "Evolving cognitive and social experience in particle swarm optimization through differential evolution: a hybrid approach", *Information Sciences*, vol. 216, pp. 50–92, 2012. DOI: `10.1016/j.ins.2012.05.017` (cit. on pp. 7, 8).

[18] A. F. Ali and M. A. Tawhid, "A hybrid Particle Swarm Optimization and Genetic Algorithm with population partitioning for large scale op-

timization problems", *Ain Shams Engineering Journal*, vol. 8, no. 2, pp. 191–206, 2017. DOI: 10.1016/j.asej.2016.07.008 (cit. on p. 8).

[19] B. Xin, J. Chen, J. Zhang, H. Fang, and Z.-H. Peng, "Hybridizing Differential Evolution and Particle Swarm Sptimization to design powerful optimizers: a review and taxonomy", *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 5, pp. 744–767, 2012. DOI: 10.1109/TSMCC.2011.2160941 (cit. on p. 8).

[20] W.-N. Chen, J. Zhang, Y. Lin, *et al.*, "Particle Swarm Optimization with an aging leader and challengers", *IEEE Transactions on Evolutionary Computation*, vol. 17, no. 2, pp. 241–258, 2013. DOI: 10.1109/TEVC.2011.2173577 (cit. on p. 8).

[21] C. Sun, Y. Jin, R. Cheng, J. Ding, and J. Zeng, "Surrogate-assisted cooperative swarm optimization of high-dimensional expensive problems", *IEEE Transactions on Evolutionary Computation*, vol. 21, no. 4, pp. 644–660, 2017. DOI: 10.1109/TEVC.2017.2675628 (cit. on p. 8).

[22] Z. Beheshti and S. M. H. Shamsuddin, "CAPSO: Centripetal accelerated particle swarm optimization", *Information Sciences*, vol. 258, pp. 54–79, 2014. DOI: 10.1016/j.ins.2013.08.015 (cit. on pp. 9, 15, 17).

[23] A. H. Gandomi, G. J. Yun, X.-S. Yang, and S. Talatahari, "Chaos-enhanced accelerated Particle Swarm Optimization", *Communications in Nonlinear Science and Numerical Simulation*, vol. 18, no. 2, pp. 327–340, 2013. DOI: 10.1016/j.cnsns.2012.07.017 (cit. on p. 9).

[24] I. Fister, A. Iglesias, A. Galvez, *et al.*, "Novelty search for global optimization", *Applied Mathematics and Computation*, vol. 347, pp. 865–881, Apr. 2019. DOI: 10.1016/j.amc.2018.11.052 (cit. on p. 9).

[25] E. Naredo and L. Trujillo, "Searching for novel clustering programs", in *Proceeding of the fifteenth annual conference on Genetic and evolu-*

*tionary computation conference - GECCO '13*, ACM Press, 2013. DOI: 10.1145/2463372.2463505 (cit. on p. 9).

[26] A. Liapis, G. N. Yannakakis, and J. Togelius, "Constrained Novelty Search: A Study on Game Content Generation", *Evolutionary Computation*, vol. 23, no. 1, pp. 101–129, Mar. 2015. DOI: 10.1162/evco_a_00123 (cit. on p. 9).

[27] G. Beni and J. Wang, "Swarm intelligence in cellular robotic systems", in *Robots and Biological Systems: Towards a New Bionics?*, P. Dario, G. Sandini, and P. Aebischer, Eds., Springer Berlin Heidelberg, 1993, pp. 703–712, ISBN: 978-3-642-58069-7 (cit. on p. 9).

[28] J. Krause, G. D. Ruxton, and S. Krause, "Swarm intelligence in animals and humans", *Trends in Ecology & Evolution*, vol. 25, no. 1, pp. 28–34, Jan. 2010. DOI: 10.1016/j.tree.2009.06.016 (cit. on p. 10).

[29] F. Marini and B. Walczak, "Particle swarm optimization (PSO). A tutorial", *Chemometrics and Intelligent Laboratory Systems*, vol. 149, pp. 153–165, 2015, ISSN: 0169-7439. DOI: 10.1016/j.chemolab.2015.08.020 (cit. on pp. 10, 15, 17).

[30] A. P. Piotrowski, "L-SHADE optimization algorithms with population-wide inertia", *Information Sciences*, vol. 468, pp. 117–141, Nov. 2018. DOI: 10.1016/j.ins.2018.08.030 (cit. on pp. 15, 17).

[31] J. Brest, M. S. Maucec, and B. Boskovic, "Single objective real-parameter optimization: Algorithm jSO", in *2017 IEEE Congress on Evolutionary Computation (CEC)*, IEEE, Jun. 2017. DOI: 10.1109/cec.2017.7969456 (cit. on pp. 15, 17).

[32] *Python programming language - https://www.python.org*, [Online; accessed 31 January 2020]. [Online]. Available: https://www.python.org (cit. on p. 16).

[33] *Package for scientific computing with Python - https://numpy.org*, [Online; accessed 31 January 2020]. [Online]. Available: `https://numpy.org` (cit. on p. 16).

[34] J. Derrac, S. García, D. Molina, and F. Herrera, "A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms", *Swarm and Evolutionary Computation*, vol. 1, no. 1, pp. 3–18, 2011, ISSN: 2210-6502. DOI: `10.1016/j.swevo.2011.02.002` (cit. on p. 18).

[35] G. Kalghatgi, "Is it really the end of internal combustion engines and petroleum in transport?", *Applied Energy*, vol. 225, pp. 965–974, 2018, ISSN: 0306-2619. DOI: `10.1016/j.apenergy.2018.05.076` (cit. on p. 29).

[36] K. I. Wong, P. K. Wong, C. S. Cheung, and C. M. Vong, "Modelling of diesel engine performance using advanced machine learning methods under scarce and exponential data set", *Applied Soft Computing*, vol. 13, no. 11, pp. 4428–4441, 2013, ISSN: 1568-4946. DOI: `10.1016/j.asoc.2013.06.006` (cit. on pp. 29, 30).

[37] T. Johnson and A. Joshi, "Review of Vehicle Engine Efficiency and Emissions", *SAE Technical Paper Series*, vol. 1, pp. 1–23, 2018. DOI: `10.4271/2018-01-0329` (cit. on p. 29).

[38] A. J. Torregrosa, A. Broatch, X. Margot, and J. Gomez-Soriano, *Understanding the unsteady pressure field inside combustion chambers of compression-ignited engines using a computational fluid dynamics approach*, 2018. DOI: `10.1177/1468087418803030` (cit. on p. 30).

[39] B. Wahono, A. Setiawan, and O. Lim, "Experimental study and numerical simulation on in-cylinder flow of small motorcycle engine", *Applied Energy*, vol. 255, p. 113 863, Dec. 2019, ISSN: 0306-2619. DOI: `10.1016/J.APENERGY.2019.113863` (cit. on p. 30).

[40] G. Decan, S. Broekaert, T. Lucchini, G. D'Errico, J. Vierendeels, and S. Verhelst, "Evaluation of wall heat flux calculation methods for CFD simulations of an internal combustion engine under both motored and HCCI operation", *Applied Energy*, vol. 232, pp. 451–461, Dec. 2018, ISSN: 0306-2619. DOI: 10.1016/J.APENERGY.2018.09.214 (cit. on p. 30).

[41] J. Mohammadhassani, A. Dadvand, S. Khalilarya, and M. Solimanpur, "Prediction and reduction of diesel engine emissions using a combined ANN–ACO method", *Applied Soft Computing*, vol. 34, pp. 139–150, 2015, ISSN: 1568-4946. DOI: 10.1016/j.asoc.2015.04.059 (cit. on pp. 30, 31).

[42] A. Broatch, R. Novella, J. Gomez-Soriano, P. Pal, and S. Som, *Numerical methodology for optimization of compression-ignited engines considering combustion noise control*, Apr. 2018. DOI: 10.4271/2018-01-0193 (cit. on pp. 30, 31, 44).

[43] A. M. Bertram, Q. Zhang, and S.-C. Kong, "A novel Particle Swarm and Genetic Algorithm hybrid method for diesel engine performance optimization", *International Journal of Engine Research*, vol. 17, no. 7, pp. 732–747, 2016. DOI: 10.1177/1468087415611031 (cit. on pp. 30, 32).

[44] M. Zubel, T. Ottenwälder, B. Heuser, and S. Pischinger, "Combustion system optimization for dimethyl ether using a genetic algorithm", *International Journal of Engine Research*, vol. 22, no. 1, pp. 22–38, May 2019. DOI: 10.1177/1468087419851577 (cit. on p. 31).

[45] J. Badra, F. khaled, J. Sim, *et al.*, "Combustion System Optimization of a Light-Duty GCI Engine Using CFD and Machine Learning", in *SAE Technical Paper Series*, SAE International, Apr. 2020. DOI: 10.4271/2020-01-1313 (cit. on p. 31).

[46] H. G. Weller, G. Tabor, H. Jasak, and C. Fureby, “A tensorial approach to computational continuum mechanics using object-oriented techniques”, *Comput. Phys.*, vol. 12, no. 6, pp. 620–631, Nov. 1998, ISSN: 0894-1866. DOI: 10.1063/1.168744 (cit. on p. 33).

[47] G. Montenegro, A. Onorati, F. Piscaglia, and G. D’Errico, “Integrated 1D-MultiD fluid dynamic models for the simulation of ICE Intake and exhaust systems”, *SAE Technical Paper Series*, vol. 1, pp. 776–790, 2007. DOI: 10.4271/2007-01-0495 (cit. on p. 33).

[48] T. Lucchini, G. D’Errico, H. Jasak, and Z. Tukovic, “Automatic mesh motion with topological changes for engine simulation”, *SAE Technical Paper Series*, vol. 1, pp. 1–20, 2007. DOI: 10.4271/2007-01-0170 (cit. on pp. 34, 35).

[49] H. Ismail, H. Ng, S. Gan, and T. Lucchini, “Approach for the Modeling of Reacting Biodiesel Fuel Spray using OpenFOAM”, *SAE Technical Paper Series*, vol. 1, pp. 1–9, 2014. DOI: 10.4271/2014-01-2565 (cit. on p. 34).

[50] G. D’Errico, T. Lucchini, G. Hardy, F. Tap, and G. Ramaekers, “Combustion Modeling in Heavy Duty Diesel Engines Using Detailed Chemistry and Turbulence-Chemistry Interaction”, *SAE Technical Paper Series*, vol. 1, pp. 1–14, 2015. DOI: 10.4271/2015-01-0375 (cit. on p. 34).

[51] R. Payri, J. Gimeno, R. Novella, and G. Bracho, “On the rate of injection modeling applied to direct injection compression ignition engines”, *International Journal of Engine Research*, vol. 17, no. 10, pp. 1015–1030, 2016. DOI: 10.1177/1468087416636281 (cit. on pp. 34, 39).

[52] H. [ Ismail], H. K. Ng, S. Gan, T. Lucchini, and A. Onorati, “Development of a reduced biodiesel combustion kinetics mechanism for CFD modelling of a light-duty diesel engine”, *Fuel*, vol. 106, pp. 388–400,

2013, ISSN: 0016-2361. DOI: `10.1016/j.fuel.2012.10.015` (cit. on p. 34).

[53] H. M. Ismail, H. K. Ng, X. Cheng, S. Gan, T. Lucchini, and G. D'Errico, "Development of Thermophysical and Transport Properties for the CFD Simulations of In-Cylinder Biodiesel Spray Combustion", *Energy & Fuels*, vol. 26, no. 8, pp. 4857–4870, 2012. DOI: `10.1021/ef300862u` (cit. on p. 34).

[54] V. Yakhot and S. A. Orszag, "Renormalization-Group Analysis of Turbulence", vol. 57, no. 14, pp. 1722–1724, 1986. DOI: `10.1007/BF01061452` (cit. on p. 34).

[55] C. Angelberger, T. Poinsot, and B. Delhay, "Improving near-wall combustion and wall heat transfer modeling in si engine computations", *SAE Technical Paper Series*, vol. 1, 1997. DOI: `10.4271/972881` (cit. on p. 34).

[56] G. D'Errico, T. Lucchini, F. Contino, M. Jangi, and X.-S. Bai, "Comparison of well-mixed and multiple representative interactive flamelet approaches for diesel spray combustion modelling", *Combustion Theory and Modelling*, vol. 18, no. 1, pp. 65–88, 2014. DOI: `10.1080/13647830.2013.860238` (cit. on p. 34).

[57] J. Benajes, P. Olmeda, J. Martín, and R. Carreño, "A new methodology for uncertainties characterization in combustion diagnosis and thermodynamic modelling", *Applied Thermal Engineering*, vol. 71, pp. 389–399, 2014. DOI: `10.1016/j.applthermaleng.2014.07.010` (cit. on p. 35).

[58] J. Benajes, R. Novella, J. M. Pastor, *et al.*, "Optimization of the combustion system of a medium duty direct injection diesel engine by combining CFD modeling with experimental validation", *Energy Conversion and Management*, vol. 110, pp. 212–229, 2016. DOI: `10.1016/j.enconman.2015.12.010` (cit. on p. 37).

[59] R. Payri, G. Bracho, J. Gimeno, and A. Bautista, “Rate of injection modelling for gasoline direct injectors”, *Energy Conversion and Management*, vol. 166, pp. 424–432, 2018. DOI: `10.1016/j.enconman.2018.04.041` (cit. on p. 39).

[60] K. Safarzynska and J. C. J. M. van den Bergh, “Integrated crisis-energy policy: Macro-evolutionary modelling of technology, finance and energy interactions”, *Technological Forecasting and Social Change*, vol. 114, pp. 119–137, 2017. DOI: `10.1016/j.techfore.2016.07.033` (cit. on p. 49).

[61] F. Li, Z. Song, and W. Liu, “China’s energy consumption under the global economic crisis: Decomposition and sectoral analysis”, *Energy Policy*, vol. 64, pp. 193–202, 2014. DOI: `10.1016/j.enpol.2013.09.014` (cit. on p. 49).

[62] S. Salcedo-Sanz, J. Muñoz-Bulnes, J. Portilla-Figueras, and J. D. Ser, “One-year-ahead energy demand estimation from macroeconomic variables using computational intelligence algorithms”, *Energy Conversion and Management*, vol. 99, pp. 62–71, 2015. DOI: `10.1016/j.enconman.2015.03.109` (cit. on pp. 49, 51, 52).

[63] J. Sánchez-Oro, A. Duarte, and S. Salcedo-Sanz, “Robust total energy demand estimation with a hybrid Variable Neighborhood Search – Extreme Learning Machine algorithm”, *Energy Conversion and Management*, vol. 123, pp. 445–452, 2016. DOI: `10.1016/j.enconman.2016.06.050` (cit. on pp. 49, 51).

[64] H. Ceylan and H. K. Ozturk, “Estimating energy demand of Turkey based on economic indicators using genetic algorithm approach”, *Energy Conversion and Management*, vol. 45, pp. 2525–2537, 2004. DOI: `10.1016/j.enconman.2003.11.010` (cit. on pp. 49–51).

[65] S. Shaik and O.-A. Yeboah, "Does climate influence energy demand? A regional analysis", *Applied Energy*, vol. 212, pp. 691–703, 2018. DOI: `10.1016/j.apenergy.2017.11.109` (cit. on p. 50).

[66] U. N. C. C. Conference, *The Paris Agreement*, UNTC XXVII 7.d, Dec. 12, 2015. [Online]. Available: `https://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf` (cit. on p. 50).

[67] L. Suganthi and A. A. Samuel, "Energy models for demand forecasting–a review", *Renewable and Sustainable Energy Reviews*, vol. 16, pp. 1223–1240, 2012. DOI: `10.1016/j.rser.2011.08.014` (cit. on p. 50).

[68] Z. W. Geem and W. E. Roper, "Energy demand estimation of south korea using artificial neural network", *Energy Policy*, vol. 37, pp. 4049–4054, 2009. DOI: `10.1016/j.enpol.2009.04.049` (cit. on p. 50).

[69] A. Ünler, "Improvement of energy demand forecasts using swarm intelligence: The case of Turkey with projections to 2025", *Energy Policy*, vol. 36, pp. 1937–1944, 2008. DOI: `10.1016/j.enpol.2008.02.018` (cit. on p. 50).

[70] M. S. Kiran, E. Özceylan, M. Gündüz, and T. Paksoy, "Swarm intelligence approaches to estimate electricity energy demand in Turkey", *Knowledge-Based Systems*, vol. 36, pp. 93–103, 2012. DOI: `10.1016/j.knosys.2012.06.009` (cit. on p. 51).

[71] ——, "A novel hybrid approach based on particle swarm optimization and ant colony optimization to forecast energy demand of Turkey", *Energy Conversion and Management*, vol. 53, pp. 75–83, 2012. DOI: `10.1016/j.enconman.2011.08.004` (cit. on p. 51).

[72] S. Yu and K. J. Zhu, "A hybrid procedure for energy demand forecasting in China", *Energy*, vol. 37, pp. 396–404, 2012. DOI: `10.1016/j.energy.2011.11.015` (cit. on p. 51).

[73] S. Yu, Y. M. Wei, and K. Wang, "A PSO-GA optimal model to estimate primary energy demand of china", *Energy Policy*, vol. 42, pp. 329–340, 2012. DOI: `10.1016/j.enpol.2011.11.090` (cit. on p. 51).

[74] S. Yu, K. Zhu, and X. Zhang, "Energy demand projection of China using a path-coefficient analysis and PSO-GA approach", *Energy Conversion and Management*, vol. 53, no. 1, pp. 142–153, 2012. DOI: `10.1016/j.enconman.2011.08.015` (cit. on p. 51).

[75] Y. Han, C. Long, Z. Geng, Q. Zhu, and Y. Zhong, "A novel DEACM integrating affinity propagation for performance evaluation and energy optimization modeling: Application to complex petrochemical industries", *Energy Conversion and Management*, vol. 183, pp. 349–359, Mar. 2019. DOI: `10.1016/j.enconman.2018.12.120` (cit. on p. 51).

[76] Y. Han, H. Wu, M. Jia, Z. Geng, and Y. Zhong, "Production capacity analysis and energy optimization of complex petrochemical industries using novel extreme learning machine integrating affinity propagation", *Energy Conversion and Management*, vol. 180, pp. 240–249, Jan. 2019. DOI: `10.1016/j.enconman.2018.11.001` (cit. on p. 51).

[77] Z. Geng, R. Zeng, Y. Han, Y. Zhong, and H. Fu, "Energy efficiency evaluation and energy saving based on DEA integrated affinity propagation clustering: Case study of complex petrochemical industries", *Energy*, vol. 179, pp. 863–875, Jul. 2019. DOI: `10.1016/j.energy.2019.05.042` (cit. on p. 51).

[78] Y. Han, H. Wu, M. Jia, Z. Geng, and Y. Zhong, "Production capacity analysis and energy optimization of complex petrochemical industries using novel extreme learning machine integrating affinity propagation", *Energy Conversion and Management*, vol. 180, pp. 240–249, Jan. 2019. DOI: `10.1016/j.enconman.2018.11.001` (cit. on p. 51).

[79] J. Colmenar, J. Hidalgo, and S. Salcedo-Sanz, “Automatic generation of models for energy demand estimation using grammatical evolution”, *Energy*, vol. 164, pp. 183–193, 2018, ISSN: 0360-5442. DOI: `10.1016/j.energy.2018.08.199` (cit. on pp. 51, 52, 55, 59, 63, 64).

[80] M. O’Neill and A. Brabazon, “Grammatical swarm: The generation of programs by social programming”, *Natural Computing*, vol. 5, no. 4, pp. 443–462, Aug. 2006. DOI: `10.1007/s11047-006-9007-7` (cit. on pp. 51–53, 56).

[81] M. O’Neill and C. Ryan, “Grammatical evolution”, *IEEE Trans. Evolutionary Computation*, vol. 5, no. 4, pp. 349–358, 2001. DOI: `10.1109/4235.942529` (cit. on pp. 52, 55).

[82] I. G. Tsoulos, D. Gavrilis, and E. Glavas, “Neural network construction using grammatical evolution”, in *Proceedings of the Fifth IEEE International Symposium on Signal Processing and Information Technology, 2005.*, IEEE, 2005, pp. 827–831. DOI: `10.1016/j.neucom.2008.01.017` (cit. on p. 52).

[83] C. X. Ling, “Overfitting and generalization in learning discrete patterns”, *Neurocomputing*, vol. 8, no. 3, pp. 341–347, Aug. 1995. DOI: `10.1016/0925-2312(95)00050-g` (cit. on p. 57).

[84] D. Martínez-Rodríguez, J. M. Colmenar, J. I. Hidalgo, R.-J. Villanueva Micó, and S. Salcedo-Sanz, “Particle swarm grammatical evolution for energy demand estimation”, *Energy Science & Engineering*, vol. 8, no. 4, pp. 1068–1079, 2020. DOI: `10.1002/ese3.568` (cit. on p. 67).

[85] *Official Site for Spanish Medic Oncology Society*, `https://www.seom.org`, Accessed: 08/05/2018 (cit. on p. 69).

[86] R. T. Greenlee, M. B. Hill-Harmon, T. Murray, and M. Thun, “Cancer statistics, 2001”, *CA: A Cancer Journal for Clinicians*, vol. 51, no. 1, pp. 15–36, Jan. 2001. DOI: 10.3322/canjclin.51.1.15 (cit. on p. 69).

[87] R. Devere White and E. Stapp, “Predicting prognosis in patients with superficial bladder cancer”, *ONCOLOGY*, vol. 12, no. 12, pp. 1717–1723, 1998 (cit. on p. 69).

[88] S. Holmang, H. Hedelin, C. Anderstrom, and S. L. Johansson, “The relationship among multiple recurrences, progression and prognosis of patients with stages TA and t1 transitional cell cancer of the bladder followed for at least 20 years”, *The Journal of Urology*, vol. 153, no. 6, pp. 1823–1827, Jun. 1995. DOI: 10.1016/s0022-5347(01)67321-x (cit. on p. 70).

[89] G. Redelman-Sidi, M. S. Glickman, and B. H. Bochner, “The mechanism of action of BCG therapy for bladder cancer—a current perspective”, *Nature Reviews Urology*, vol. 11, no. 3, pp. 153–162, Feb. 2014. DOI: 10.1038/nrurol.2014.15 (cit. on pp. 70, 71, 73, 75–77).

[90] F. Audenet, K. Attalla, and J. P. Sfakianos, “The evolution of bladder cancer genomics: What have we learned and how can we use it?”, *Urologic Oncology: Seminars and Original Investigations*, Mar. 2018. DOI: 10.1016/j.urolonc.2018.02.017 (cit. on p. 70).

[91] *Bladder Cancer Treatment (PDQ)-Health Professional Version*, https://www.cancer.gov/types/bladder/hp/bladder-treatment-pdq, Accessed: 07/05/2018 (cit. on p. 70).

[92] *Bladder Cancer Treatment (PDQ)-Patient Version*, https://www.cancer.gov/types/bladder/patient/bladder-treatment-pdq, Accessed: 07/05/2018 (cit. on p. 70).

[93] S. Bunimovich-Mendrazitsky, E. Shochat, and L. Stone, "Mathematical model of BCG immunotherapy in superficial bladder cancer", *Bulletin of Mathematical Biology*, vol. 69, no. 6, pp. 1847–1870, Apr. 2007. DOI: 10.1007/s11538-007-9195-z (cit. on p. 71).

[94] *Official Site for Hospital Universitari i Politècnic La Fe, Valencia, Spain*, http://www.hospital-lafe.com, Accessed: 07/05/2018 (cit. on p. 71).

[95] D. Hanahan and R. Weinberg, "Hallmarks of cancer: The next generation", *Cell*, vol. 144, no. 5, pp. 646–674, Mar. 2011. DOI: 10.1016/j.cell.2011.02.013 (cit. on pp. 73, 74, 76).

[96] S. Choi, H.-G. Choi, K.-W. Shin, *et al.*, "Mycobacterium tuberculosis protein rv3841 activates dendritic cells and contributes to a t helper 1 immune response", *Journal of Immunology Research*, vol. 2018, pp. 1–13, 2018. DOI: 10.1155/2018/3525302 (cit. on pp. 73, 76, 77).

[97] A. Wein, L. Kavoussi, A. Novick, A. Partin, and C. Peters, *Campbell-Walsh Urology*, ser. Campbell-Walsh Urology. Elsevier Health Sciences, 2011, ISBN: 9781455722983 (cit. on p. 75).

[98] C. Burgos-Simón, N. García-Medina, D. Martínez-Rodríguez, and R.-J. Villanueva, "Mathematical modeling of the dynamics of the bladder cancer and the immune response applied to a patient: Evolution and short-term prediction", *Mathematical Methods in the Applied Sciences*, Feb. 2019. DOI: 10.1002/mma.5536 (cit. on p. 87).

[99] T. Zhang, Q. Wu, and Z. Zhang, "Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak", *Current Biology*, vol. 30, no. 7, 1346–1351.e2, 2020, ISSN: 0960-9822. DOI: 10.1016/j.cub.2020.03.022 (cit. on p. 90).

[100] F. Zhou, T. Yu, R. Du, *et al.*, "Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospec-

tive cohort study", *The Lancet*, vol. 395, no. 10229, pp. 1054–1062, Mar. 2020. DOI: 10.1016/s0140-6736(20)30566-3 (cit. on pp. 90, 98, 100).

[101] Worldometer, *COVID-19 Coronavirus Pandemic*. [Online]. Available: https://www.worldometers.info/coronavirus/ (cit. on p. 90).

[102] L. C. Okell, R. Verity, O. J. Watson, *et al.*, "Have deaths from COVID-19 in Europe plateaued due to herd immunity?", *The Lancet*, Jun. 2020. DOI: 10.1016/s0140-6736(20)31357-x (cit. on p. 90).

[103] J. H. Beigel, K. M. Tomashek, L. E. Dodd, *et al.*, "Remdesivir for the Treatment of Covid-19 — Preliminary Report", *New England Journal of Medicine*, May 2020. DOI: 10.1056/nejmoa2007764 (cit. on p. 90).

[104] L. Riva, S. Yuan, X. Yin, *et al.*, "A Large-scale Drug Repositioning Survey for SARS-CoV-2 Antivirals", *bioRxiv*, 2020. DOI: 10.1101/2020.04.16.044016 (cit. on p. 91).

[105] F. Touret, M. Gilles, K. Barral, *et al.*, "In vitro screening of a FDA approved chemical library reveals potential inhibitors of SARS-CoV-2 replication", *bioRxiv*, 2020. DOI: 10.1101/2020.04.03.023846 (cit. on p. 91).

[106] B. Damle, M. Vourvahis, E. Wang, J. Leaney, and B. Corrigan, "Clinical Pharmacology Perspectives on the Antiviral Activity of Azithromycin and Use in COVID-19", *Clinical Pharmacology & Therapeutics*, May 2020. DOI: 10.1002/cpt.1857 (cit. on p. 91).

[107] S. Jeon, M. Ko, J. Lee, *et al.*, "Identification of antiviral drug candidates against SARS-CoV-2 from FDA-approved drugs", Mar. 2020. DOI: 10.1101/2020.03.20.999730 (cit. on p. 91).

[108] L. Caly, J. D. Druce, M. G. Catton, D. A. Jans, and K. M. Wagstaff, "The FDA-approved drug ivermectin inhibits the replication of SARS-

CoV-2 in vitro", *Antiviral Research*, vol. 178, p. 104 787, Jun. 2020. DOI: 10.1016/j.antiviral.2020.104787 (cit. on p. 91).

[109] J. C. Rajter, M. Sherman, N. Fatteh, F. Vogel, J. Sacks, and J.-J. Rajter, "ICON (Ivermectin in COvid Nineteen) study: Use of Ivermectin is Associated with Lower Mortality in Hospitalized Patients with COVID19", *medRxiv*, 2020. DOI: 10.1101/2020.06.06.20124461 (cit. on p. 91).

[110] T. P. Sheahan, A. C. Sims, S. Zhou, *et al.*, "An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cell cultures and multiple coronaviruses in mice", *Science Translational Medicine*, vol. 12, no. 541, eabb5883, Apr. 2020. DOI: 10.1126/scitranslmed.abb5883 (cit. on p. 92).

[111] S. Dorogovtsev and J. Mendes, *Evolution of Networks*. Oxford University Press, Jan. 2003. DOI: 10.1093/acprof:oso/9780198515906.001.0001 (cit. on p. 92).

[112] S. Helleringer and H.-P. Kohler, "Sexual network structure and the spread of HIV in Africa: evidence from Likoma Island, Malawi", *AIDS*, vol. 21, no. 17, pp. 2323–2332, Nov. 2007. DOI: 10.1097/qad.0b013e328285df98 (cit. on p. 92).

[113] L. Acedo, J.-A. Moraño, R.-J. Villanueva, J. Villanueva-Oller, and J. Díez-Domingo, "Using random networks to study the dynamics of respiratory syncytial virus (RSV) in the Spanish region of Valencia", *Mathematical and Computer Modelling*, vol. 54, no. 7-8, pp. 1650–1654, Oct. 2011. DOI: 10.1016/j.mcm.2010.11.068 (cit. on pp. 92, 96).

[114] J. Díez-Domingo, V. Sánchez-Alonso, R.-J. Villanueva, L. Acedo, J.-A. Moraño, and J. Villanueva-Oller, "Random Network Models to Predict the Long-Term Impact of HPV Vaccination on Genital Warts", *Viruses*, vol. 9, no. 10, p. 300, Oct. 2017. DOI: 10.3390/v9100300 (cit. on p. 92).

[115] B. Bollobás and B. Béla, *Random graphs*, 73. Cambridge University Press, 2001 (cit. on p. 93).

[116] Centers for Disease Control and Prevention (CDC), *How COVID-19 Spreads.* [Online]. Available: `https://www.cdc.gov/coronavirus/2019-ncov/faq.html#How-COVID-19-Spreads` (cit. on p. 93).

[117] National Institute of Statistics of Spain (INE), *Number of Spanish population.* [Online]. Available: `https://www.ine.es` (cit. on p. 95).

[118] J. Villanueva-Oller, L. Acedo, J. A. Moraño, and A. Sánchez-Sánchez, "Epidemic Random Network Simulations in a Distributed Computing Environment", *Abstract and Applied Analysis*, vol. 2013, pp. 1–10, 2013. DOI: `10.1155/2013/462801` (cit. on p. 96).

[119] J. H. Tanne, E. Hayasaki, M. Zastrow, P. Pulla, P. Smith, and A. G. Rada, "Covid-19: How doctors and healthcare systems are tackling coronavirus worldwide", *BMJ*, vol. 368, 2020. DOI: `10.1136/bmj.m1090` (cit. on p. 97).

[120] Google, *Google Local Mobility Reports About COVID-19.* [Online]. Available: `https://www.google.com/covid19/mobility/` (cit. on p. 97).

[121] W.-j. Guan, Z.-y. Ni, Y. Hu, *et al.*, "Clinical Characteristics of Coronavirus Disease 2019 in China", *New England Journal of Medicine*, vol. 382, no. 18, pp. 1708–1720, Apr. 2020. DOI: `10.1056/nejmoa2002032` (cit. on pp. 97, 115).

[122] R. M. Anderson, H. Heesterbeek, D. Klinkenberg, and T. D. Hollingsworth, "How will country-based mitigation measures influence the course of the COVID-19 epidemic?", *The Lancet*, vol. 395, no. 10228, pp. 931–934, Mar. 2020. DOI: `10.1016/s0140-6736(20)30567-5` (cit. on pp. 97, 115).

[123] Instituto de Salud Carlos III, *Estudio ENE-COVID19: Primera ronda. Estudio Nacional de Sero-Epidemiología de la Infección Por SARS-Cov-2 en España. Informe Preliminar 13 de Mayo de 2020.* [Online]. Available: `https://s03.s3c.es/imag/doc/2020-05-13/ENE_COVID_Informe_preliminar_cierre_de_la_primera_ronda_13Mayo2020.pdf` (cit. on pp. 98, 100).

[124] DATADISTA, *Datasets relacionados con la incidencia de la COVID-19 en España (Datasets related to the incidence of COVID-19 in Spain).* [Online]. Available: `https://github.com/datadista/datasets/tree/master/COVID%2019` (cit. on pp. 98, 100, 101).

[125] World Health Organization (WHO), *Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19), 16-24 February 2020.* [Online]. Available: `https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf` (cit. on pp. 99, 104).

[126] F. Brauer and C. Castillo-Chavez, *Mathematical Models in Population Biology and Epidemiology.* Springer New York, 2012. DOI: `10.1007/978-1-4614-1686-9` (cit. on pp. 99, 103, 112).

[127] S. Sanche, Y. T. Lin, C. Xu, E. Romero-Severson, N. Hengartner, and R. Ke, "High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2", *Emerging Infectious Diseases*, vol. 26, no. 7, Jul. 2020. DOI: `10.3201/eid2607.200282` (cit. on pp. 99, 124).

[128] T. Greenhalgh, M. B. Schmid, T. Czypionka, D. Bassler, and L. Gruer, "Face masks for the public during the COVID-19 crisis", *BMJ*, p. m1435, Apr. 2020. DOI: `10.1136/bmj.m1435` (cit. on p. 109).

[129] J.-M. Benlloch, J.-C. Cortés, D. Martínez-Rodríguez, R.-S. Julián, and R.-J. Villanueva, "Effect of the early use of antivirals on the COVID-19 pandemic. a computational network modeling approach", *Chaos, Soli-*

*tons & Fractals*, vol. 140, p. 110 168, Nov. 2020. DOI: 10.1016/j.chaos.2020.110168 (cit. on p. 110).

[130] G. D. Barmparis and G. Tsironis, "Estimating the infection horizon of covid-19 in eight countries with a data-driven approach", *Chaos, Solitons & Fractals*, p. 109 842, 2020. DOI: 10.1016/j.chaos.2020.109842 (cit. on p. 112).

[131] M. Gatto, E. Bertuzzo, L. Mari, *et al.*, "Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures", *Proceedings of the National Academy of Sciences*, vol. 117, no. 19, pp. 10 484–10 491, 2020, ISSN: 0027-8424. DOI: 10.1073/pnas.2004978117 (cit. on p. 112).

[132] B. Tang, X. Wang, Q. Li, *et al.*, "Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions", *Journal of Clinical Medicine*, vol. 9, no. 2, p. 462, Feb. 2020. DOI: 10.3390/jcm9020462 (cit. on p. 112).

[133] Z. Liu, P. Magal, O. Seydi, and G. Webb, "A COVID-19 epidemic model with latency period", *Infectious Disease Modelling*, vol. 5, pp. 323–337, 2020. DOI: 10.1016/j.idm.2020.03.003 (cit. on p. 112).

[134] K. Prem, Y. Liu, T. W. Russell, *et al.*, "The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in wuhan, china: A modelling study", *The Lancet Public Health*, vol. 5, no. 5, e261–e270, May 2020. DOI: 10.1016/s2468-2667(20)30073-6 (cit. on p. 112).

[135] J. Grauer, H. Löwen, and B. Liebchen, "Strategic spatiotemporal vaccine distribution increases the survival rate in an infectious disease like covid-19", *Scientific reports*, vol. 10, no. 1, pp. 1–10, 2020. DOI: 10.1038/s41598-020-78447-3 (cit. on p. 112).

[136] C. Hou, J. Chen, Y. Zhou, *et al.*, "The effectiveness of quarantine of wuhan city against the corona virus disease 2019 (covid-19): A well-mixed seir model analysis", *Journal of medical virology*, 2020. DOI: `10.1002/jmv.25827` (cit. on p. 112).

[137] M. Serhani and H. Labbardi, "Mathematical modeling of COVID-19 spreading with asymptomatic infected and interacting peoples", *Journal of Applied Mathematics and Computing*, Aug. 2020. DOI: `10.1007/s12190-020-01421-9` (cit. on p. 112).

[138] U. Avila-Ponce de León, A. G. Pérez, and E. Avila-Vales, "An SEIARD epidemic model for COVID-19 in mexico: Mathematical analysis and state-level forecast", *Chaos, Solitons & Fractals*, vol. 140, p. 110 165, 2020, ISSN: 0960-0779. DOI: `10.1016/j.chaos.2020.110165` (cit. on p. 112).

[139] K. Sarkar, S. Khajanchi, and J. J. Nieto, "Modeling and forecasting the COVID-19 pandemic in India", *Chaos, Solitons & Fractals*, vol. 139, p. 110 049, 2020, ISSN: 0960-0779. DOI: `10.1016/j.chaos.2020.110049` (cit. on p. 112).

[140] G. Giordano, F. Blanchini, R. Bruno, *et al.*, "Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy", *Nature Medicine*, vol. 26, no. 6, pp. 855–860, Apr. 2020. DOI: `10.1038/s41591-020-0883-7` (cit. on p. 112).

[141] E. Shim, A. Tariq, W. Choi, Y. Lee, and G. Chowell, "Transmission potential and severity of COVID-19 in South Korea", *International Journal of Infectious Diseases*, vol. 93, pp. 339–344, 2020, ISSN: 1201-9712. DOI: `10.1016/j.ijid.2020.03.031` (cit. on p. 112).

[142] J. de Andalucía, *Instituto de estadistica y cartografia de Andalucia*, `http://www.juntadeandalucia.es/institutodeestadisticaycartografia` (cit. on p. 112).

[143] M. Pollán, B. Pérez-Gómez, R. Pastor-Barriuso, *et al.*, “Prevalence of SARS-CoV-2 in Spain (ENE-COVID): A nationwide, population-based seroepidemiological study”, *The Lancet*, vol. 396, no. 10250, pp. 535–544, Aug. 2020. DOI: `10.1016/s0140-6736(20)31483-5` (cit. on pp. 112, 119, 121, 127).

[144] *Informe final del estudio nacional de sero-epidemiología de la infección por SARS-COV-2 en España (Sero-epidemiology national study of the infection for SARS-COV-2 in Spain. Final report (in Spanish))*, `https://www.mscbs.gob.es/ciudadanos/ene-covid/docs/ESTUDIO_ENE-COVID19_INFORME_FINAL.pdf`, Jul. 2020 (cit. on pp. 112, 119, 121, 127).

[145] Google, *Google Local Mobility Reports About COVID-19*, `https://www.google.com/covid19/mobility/` (cit. on p. 114).

[146] World Health Organization (WHO), *Q&A on coronaviruses (COVID-19)*, `https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-coronaviruses` (cit. on p. 115).

[147] A. Iwasaki, “What reinfections mean for COVID-19”, *The Lancet Infectious Diseases*, vol. 21, no. 1, pp. 3–5, Jan. 2021. DOI: `10.1016/s1473-3099(20)30783-0` (cit. on p. 117).

[148] Centers for Disease Control and Prevention, *Reinfection with COVID-19*, `https://www.cdc.gov/coronavirus/2019-ncov/your-health/reinfection.html` (cit. on p. 117).

[149] A. W. D. Edridge, J. Kaczorowska, A. C. R. Hoste, *et al.*, “Seasonal coronavirus protective immunity is short-lasting”, *Nature Medicine*, Sep. 2020. DOI: `10.1038/s41591-020-1083-1` (cit. on p. 118).

[150] G. Dogan, “Bootstrapping for confidence interval estimation and hypothesis testing for parameters of system dynamics models”, *System Dy-*

*namics Review*, vol. 23, no. 4, pp. 415–436, 2007. DOI: 10.1002/sdr.362 (cit. on pp. 119, 124, 137, 138).

[151] Ministerio de Sanidad. Telegram, *COVID-19 Vaccination in Spain*, https://t.me/sanidadgob/1047, Feb. 2021 (cit. on p. 122).

[152] El País, *The pace expected by the Health Ministry: a million vaccinated per week*, https://elpais.com/sociedad/2021-01-07/el-ritmo-que-preve-sanidad-un-millon-de-vacunados-por-semana.html, Jan. 2021 (cit. on p. 122).

[153] F. P. Polack, S. J. Thomas, N. Kitchin, *et al.*, "Safety and efficacy of the BNT162b2 mRNA covid-19 vaccine", *New England Journal of Medicine*, vol. 383, no. 27, pp. 2603–2615, Dec. 2020. DOI: 10.1056/nejmoa2034577 (cit. on p. 123).

[154] S. Mallapaty, "Are COVID vaccination programmes working? scientists seek first clues", *Nature*, vol. 589, no. 7843, pp. 504–505, Jan. 2021. DOI: 10.1038/d41586-021-00140-w (cit. on p. 123).

[155] Centers for Disease Control and Prevention, *COVID-19 Pandemic Planning Scenarios*, https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html (cit. on p. 126).

[156] G. E. Weissman, A. Crane-Droesch, C. Chivers, *et al.*, "Locally informed simulation to predict hospital capacity needs during the COVID-19 pandemic", *Annals of Internal Medicine*, vol. 173, no. 1, pp. 21–28, Jul. 2020. DOI: 10.7326/m20-1260 (cit. on p. 134).

[157] S.-C. Chen and C.-M. Liao, "Modelling control measures to reduce the impact of pandemic influenza among schoolchildren", *Epidemiology and Infection*, vol. 136, no. 8, pp. 1035–1045, Sep. 2007. DOI: 10.1017/s0950268807009284 (cit. on p. 135).

[158] J. Cui, Y. Zhang, Z. Feng, S. Guo, and Y. Zhang, "Influence of asymptomatic infections for the effectiveness of facemasks during pandemic influenza", *Mathematical Biosciences and Engineering*, vol. 16, no. mbe-16-05-194, p. 3936, 2019, ISSN: 1551-0018. DOI: 10.3934/mbe.2019194 (cit. on p. 135).

[159] S.-C. Chen and C.-M. Liao, "Cost-effectiveness of influenza control measures: A dynamic transmission model-based analysis", *Epidemiology and Infection*, vol. 141, no. 12, pp. 2581–2594, 2013. DOI: 10.1017/S0950268813000423 (cit. on p. 135).

[160] N. Arinaminpathy, N. Raphaely, L. Saldana, *et al.*, "Transmission and control in an institutional pandemic influenza A(H1N1) 2009 outbreak", *Epidemiology and Infection*, vol. 140, no. 6, pp. 1102–1110, Aug. 2011. DOI: 10.1017/s0950268811001518 (cit. on p. 135).

[161] B. J. Cowling, S. T. Ali, T. W. Y. Ng, *et al.*, "Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: An observational study", *The Lancet Public Health*, vol. 5, no. 5, e279–e288, 2020, ISSN: 2468-2667. DOI: 10.1016/S2468-2667(20)30090-6 (cit. on p. 135).

[162] S. M. Tracht, S. Y. Del Valle, and J. M. Hyman, "Mathematical Modeling of the Effectiveness of Facemasks in Reducing the Spread of Novel Influenza A (H1N1)", *PLOS ONE*, vol. 5, no. 2, pp. 1–12, Feb. 2010. DOI: 10.1371/journal.pone.0009018 (cit. on p. 135).

[163] A. E. Aiello, G. F. Murray, V. Perez, *et al.*, "Mask use, hand hygiene, and seasonal influenza-like illness among young adults: A randomized intervention trial", *The Journal of Infectious Diseases*, vol. 201, no. 4, pp. 491–498, Feb. 2010. DOI: 10.1086/650396 (cit. on p. 135).

[164] S.-A. Lee, S. A. Grinshpun, and T. Reponen, "Respiratory performance offered by n95 respirators and surgical masks: Human subject evaluation

with NaCl aerosol representing bacterial and viral particle size range", *The Annals of Occupational Hygiene*, vol. 52, no. 3, pp. 177–185, Mar. 2008. DOI: `10.1093/annhyg/men005` (cit. on p. 135).

[165] N. C. J. Brienen, A. Timen, J. Wallinga, J. E. Van Steenbergen, and P. F. M. Teunis, "The effect of mask use on the spread of influenza during a pandemic", *Risk Analysis*, vol. 30, no. 8, pp. 1210–1218, 2010. DOI: `10.1111/j.1539-6924.2010.01428.x` (cit. on p. 135).

[166] S. Sim, K. Moey, and N. Tan, "The use of facemasks to prevent respiratory infection: A literature review in the context of the health belief model", *Singapore Medical Journal*, vol. 55, no. 3, Mar. 2014. DOI: `10.11622/smedj.2014037` (cit. on p. 135).

[167] J. M. Garrido, D. Martínez-Rodríguez, F. Rodríguez-Serrano, S.-M. Sferle, and R.-J. Villanueva, "Modeling COVID-19 with uncertainty in granada, spain. intra-hospitalary circuit and expectations over the next months", *Mathematics*, vol. 9, no. 10, p. 1132, May 2021. DOI: `10.3390/math9101132` (cit. on p. 140).