**Mathematical
and Computational
Applications**

Mathematical Modelling in Engineering & Human Behaviour 2018

Edited by
Lucas Jódar, Juan Carlos Cortés and Luis Acedo Rodríguez

MDPI

Mathematical Modelling in Engineering & Human Behaviour 2018

Mathematical Modelling in Engineering & Human Behaviour 2018

Special Issue Editors

Lucas Jódar
Juan Carlos Cortés
Luis Acedo Rodríguez

MDPI

Special Issue Editors

Lucas Jódar
Universitat Politècnica de
València
Spain

Juan Carlos Cortés
Universitat Politècnica de
València
Spain

Luis Acedo Rodríguez
Universitat Politècnica de
València
Spain

This is a reprint of articles from the Special Issue published online in the open access journal *Mathematical and Computational Applications* (ISSN 2297-8747) from 2018 to 2019 (available at: https://www.mdpi.com/journal/mca/special_issues/MMEHB18)

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. *Journal Name* **Year**, *Article Number*, Page Range.

Contents

About the Special Issue Editors

Lucas Jódar, Full Professor, Ph.D., graduated and received his doctorate in Mathematical Sciences from the University of Valencia in 1978 and 1982, respectively. He has been a Full University Professor in Applied Mathematics since 1991 and the Head of the Department of Applied Mathematics of Universitat Politècnica de València, during 1992–2002. Since its foundation in 2004, Professor Jódar has been the head of the University Institute of Research in Cross-Disciplinary Mathematics, and holds five six-year periods of research and has directed 30 doctoral theses. He is the co-author of 425 research articles and has been in charge of competitive research projects since 1988. Recently, Professor Jódar was in charge of the Spanish team within the European Project entitled STRIKE, which is based on the use of computational methods in finance. He is an editor in several international journals specialized in such topics as modelling and numerical methods.

Juan Carlos Cortés, Full Professor, Ph.D., received his degree in Mathematics at the University of Valencia and his PhD degree from Universitat Politècnica de València (UPV), where he is currently a Full Professor in Applied Mathematics. At UPV, Professor Cortés is also a Deputy Director at the Institute of Multidisciplinary Mathematics, where he conducts his research, mainly focusing on uncertainty quantification and modelling. He has published more than 130 refereed journal papers and been in charge of competitive research projects on random differential equations and applications. Professor Cortés has been a co-organiser of more than 10 international conferences on mathematical modelling. Currently, he is an associate editor of several international journals in Applied Mathematics and has served as a guest editor of more than 10 Special Issues published in well-recognised and internationally ranked journals in Mathematics.

Luis Acedo Rodríguez, Research Associate, Ph.D., received his graduate degree and his Ph.D. in fundamental physics from the University of Extremadura in Badajoz, Spain. He has been a visiting scientist at the Universities of Utrecht and Mexico. From 1998 to 2006, Dr. Acedo was a teaching assistant in Physics and Mathematics at the Universities of Extremadura and Salamanca. His current research interests are interdisciplinary, including mathematical modelling of infectious diseases, epidemiology, networks and gravitational physics. He has published more than 80 refereed journal papers and book chapters, and been a co-organizer for 10 international conferences on mathematical modelling, as well as a guest editor of several Special Issues by internationally recognised publishers. Dr. Acedo has also participated in research projects modelling meningitis propagation and control, and holds several research contracts with the pharmaceutical industry.

Preface to "Mathematical Modelling in Engineering & Human Behaviour 2018"

This book is devoted to the mathematical models applied in studying problems of engineering and human behaviour. These papers were presented as communications at the conference "Mathematical Modelling in Engineering & Human Behaviour 2018" held in Valencia, Spain, from July 16th to July 18th 2018.

We were particularly interested in the cross-disciplinary applications of mathematics, ranging from medicine, epidemiology, financial mathematics, social addictions and public health issues. Our objective was to develop the most useful techniques for dealing with these problems in a way that could be applied by specialists of the many disciplines in which mathematics has proven to be of critical interest. At the aforementioned conference, we gathered a group of participants within multidisciplinary fields with these common interests and objectives. This allowed us to create the synergy necessary for effectively implementing mathematical modelling methods in the most diverse applications.

In the first three papers of this book, we present three applications within medicine and epidemiology. The first paper is devoted to the modelling of chikungunya virus spread in Colombia. In this work, González-Parra et al. estimate the basic reproduction number of the disease by using bootstrapping and Markov chain Monte Carlo techniques. In the second paper, Trejo et al. discuss an interesting model for assessing the role of the immune system in bone-healing process. This model could be used in the future to monitor bone fracture treatments. The 3rd paper discusses a cellular automata model of the brain that could, in spite of its simplicity, provide some insights into Alzheimer's disease and other neurological disorders.

The 4th to the 6th paper focus on problems connected with the human sciences. In particular, in the 4th paper, Carpitella et al. propose a multicriteria decision-making approach to gauge the organisational risks due to human factors. In the 5th paper, Delgadillo-Alemán et al. identify social factors involved in the problem of intimate partner violence, and develop a model to predict and control this social pandemic. In the 6th paper, Acedo discusses the application of hidden Markov models to linguistics, with the idea of finding some patterns of structure in the ancient Voynich manuscript.

The 7th to the 10th paper are devoted to matrix and numerical analysis. In the 7th paper, Calatayud et al. extend some recent advances in random differential equations to non-autonomous differential equations that are of interest in mathematical modelling. The 8th paper, by Arora et al., provides discussion of a multistep scheme with fast convergence. In the 9th paper, the author, Solis, discusses a family of polynomial discrete dynamical systems with interesting properties, and the difficulties in obtaining points of equilibria and the role of transient behaviour are analysed. The 10th paper, by Carreño et al., discusses the block preconditioning matrix techniques used to obtain the modes of a nuclear reactor.

The book ends with an application to naval defence. In the 11th paper, Carreño et al. analyse the problem of integrating uncertainty to select the best anti-torpedo decoys for frigates.

Lucas Jódar, Juan Carlos Cortés, Luis Acedo Rodríguez
Special Issue Editors

Article

Mathematical Modeling and Characterization of the Spread of Chikungunya in Colombia

Gilberto C. González-Parra [1,*] **, Diego F. Aranda** [2] **, Benito Chen-Charpentier** [3] **, Miguel Díaz-Rodríguez** [4] **and Jaime E. Castellanos** [5]

1   Department of Mathematics, New Mexico Tech, Socorro, NM 87801, USA
2   Facultad de Ciencias, Departamento de Matemáticas, Universidad El Bosque, Edificio F, piso 2, Bogotá, Colombia; arandadiego@unbosque.edu.co
3   Department of Mathematics, University of Texas at Arlington, Arlington, TX 76019, USA; bmchen@uta.edu
4   Facultad de Ing., Centro de Investigacion de Matemáticas Aplicada, Universidad de Los Andes, Mérida, Venezuela; dmiguel@ula.ve
5   Grupo de Virologia, Facultad de Ciencias, Universidad El Bosque, Av. Carrera 9 No. 131-A-02, Bogotá, Colombia; castellanosjaime@unbosque.edu.co
*   Correspondence: gilberto.gonzalezparra@nmt.edu; Tel.: +1-575-835-5758

Abstract: The Chikungunya virus is the cause of an emerging disease in Asia and Africa, and also in America, where the virus was first detected in 2006. In this paper, we present a mathematical model of the Chikungunya epidemic at the population level that incorporates the transmission vector. The epidemic threshold parameter $\mathcal{R}_0$ for the extinction of disease is computed using the method of the next generation matrix, which allows for insights about what are the most relevant model parameters. Using Lyapunov function theory, some sufficient conditions for global stability of the the disease-free equilibrium are obtained. The proposed mathematical model of the Chikungunya epidemic is used to investigate and understand the importance of some specific model parameters and to give some explanation and understanding about the real infected cases with Chikungunya virus in Colombia for data belonging to the year 2015. In this study, we were able to estimate the value of the basic reproduction number $\mathcal{R}_0$. We use bootstrapping and Markov chain Monte Carlo techniques in order to study parameters' identifiability. Finally, important policies and insights are provided that could help government health institutions in reducing the number of cases of Chikungunya in Colombia.

Keywords: Chikungunya disease; mathematical modeling; nonlinear dynamical systems; numerical simulations; parameter estimation; Markov chain Monte Carlo

1. Introduction

The Chikungunya virus is a type of arbovirus, so it is only transmitted by hematopoietic arthropods that become infected after biting some vertebrates. Later, the arthropods can transmit the virus to a susceptible vertebrate through a bite [1]. The arthropod vectors are usually *Aedes aegypti* and *Ae. albopictus* [2–4]. The virus has had mutations, allowing *Aedes albopictus* to transmit the disease [5]. The Chikungunya disease originated in Sub-Saharan Africa, and it has become endemic in Africa, where there is a natural transmission between mosquitoes and primates to humans [3]. The Chikungunya disease is an emerging disease in Asia. In America, it was detected in 2006, and there is an imminent risk of the virus spreading throughout South America. In the last few years, several outbreaks have occurred on the island of Reunion, in Cambodia, Comoros, Mayotte, Madagascar, Mauritius, Italy, Seychelles, and the Maldives [3,6–8]. The first outbreak in Europe was in the warm northeast region of Italy in July 2007. Probably, the virus came from Kerala (India), where

the disease was at its highest peak [7]. The Chikungunya disease causes arthritis, fever, and pain of the joints. Symptoms of chikungunya are generally resolved within 7–10 days, but some patients are plagued with chronic arthralgia, which could persist for months or years [9–11]. There is no vaccine for the Chikungunya virus at this moment that could be used to restrict and control the transmission of the disease [6,12]. Moreover, no effective drug is available for human use for any alphavirus, although analgesics and non-steroidal anti-inflammatory drugs can provide relief from symptoms [12]. The incubation of the virus lasts between 5–12 days and the infectious stage between 5–15 days, in both primates and humans [13,14].

As mentioned above, the Chikungunya virus is spreading around the world, and due to world climate change, it is expected that more regions are going to be reached by it. For instance, in [15], the authors presented results regarding the risk of Zika and Chikungunya virus transmission in human population centers of the eastern United States. In this way, in order to better understand the dynamics of how the Chikungunya virus is transmitted, we propose and analyze a mathematical model given by a system of nonlinear differential equations where the populations of hosts and mosquitoes are homogeneous. This mathematical modeling approach is the standard way to study the dynamical behavior of diseases in populations from an epidemiological point of view [16–19]. In particular, there is a variety of models for vector-borne diseases. For instance, in [20], the authors proposed a mathematical model for vector-borne disease with delay to consider the incubation period. In addition, in [21], the authors used optimal control to minimize the number of infections. Recently, some works extended these types of models using versatile fractional derivatives [22–24].

In [25], the authors proposed a deterministic mathematical model for Chikungunya infection considering that there is transmission of the virus between humans and mosquitoes. The authors used two infected human subpopulations designated as symptomatic and asymptomatic to classify the humans responsible for transmitting the virus. Additionally, they considered the subpopulation of humans carrying the virus, but had no possibility of spreading it. In this paper, the authors demonstrated the influence of humans on the infection of the latency period. They also remarked about the necessity of fitting the model to real data so that it will be useful in controlling the spread of the virus. Another interesting work was presented in [26], where the authors proposed a mathematical model of three age-structured transmissions of Chikungunya virus. The authors divided the human population into juvenile, adult, and senior subpopulations.

In [6], the authors proposed a stochastic mathematical model for a rural region in Cambodia, considering the fact that a stochastic model fits data better in small populations. These authors also used a subgroup of latent human and mosquito individuals and additionally included a larvae subgroup in the mosquito population. An innovation in that work is the introduction of three subclasses of infected humans. Other mathematical models for the Chikungunya virus propagation that include the latency stage were given in [27–29].

In the present work, our aim is to understand and explain some dynamics regarding the prevalence of Chikungunya infection in Colombia. The mathematical model is important since in practice, there are few economic resources to tackle or fight the Chikungunya virus' spread. Thus, the model allows one to decide on the best or most convenient health policies. In the proposed model, we consider a chronic subpopulation that has not been considered previously. The people in this particular class cannot transmit the disease, but have some types of chronic rheumatic symptoms. The main reason to consider this class is that health institutions are interested in the number of chronic cases of Chikungunya and its evolution. In addition, the latency stage for the human and mosquito populations is also introduced in the mathematical model, since this stage is observed in the real world. Although the virus may also spread to other vertebrate populations such as primates [30–33], reservoirs different from humans are not included in the analysis, since in Colombia, known cases are all human. However, future work might include reservoirs different from humans and would increase the complexity of the mathematical model and of the fitting process to the real data.

Numerical simulations are performed to support the theoretical results. In addition, the proposed mathematical model of Chikungunya is used to explain and understand the infected cases with the Chikungunya virus in Colombia. In particular, we use the proposed Chikungunya mathematical model to perform a fitting process to real data of Colombia. Additionally, we used bootstrapping and Markov chain Monte Carlo techniques in order to do analysis of the parameters' identifiability [34–40]. Finally, important policies and insights are provided that could help government health institutions in lowering the infected cases with the Chikungunya virus in Colombia.

The paper is organized as follows. In Section 2, the mathematical model of Chikungunya is presented together with a set of definitions and basic underlying hypotheses. The proposed Chikungunya mathematical model is analyzed in Section 3. The fitting process and numerical simulations are presented in Section 4. Next, the parameter estimation of the model for a Colombia case using Markov chain Monte Carlo, as well as bootstrapping is performed in Section 5. The last part, Section 6, is devoted to the discussion and conclusions.

2. Mathematical Model

In this section, we set out a continuous mathematical model for the transmission and evolution of the Chikungunya infection in the human and mosquito populations. In the proposed mathematical model, vertical transmission is not included because the number of cases is small when compared to the total number of infected cases. However, if the virus prevalence increases in the female population, it may be necessary to include it in the proposed model. Furthermore, we do not consider a vaccinated class since there are no vaccines on the market [12,41].

The proposed model of the Chikungunya virus transmission dynamics incorporates a cross-transmission between the human and vector populations. In particular, it is assumed that the Chikungunya virus spreads by the effective contact between a mosquito infected with a human susceptible, and vice versa. This contact depends on different environmental factors. Some factors to consider are: the weather, the temperature, the altitude, and the mosquito bite rate. Varying these values will generate different degrees of the probability of the transmission of the disease. Here, we will not include seasonal effects or variability in the populations. Using a population-based approach of an epidemiological type, the population of humans is divided into five groups: susceptibles, latents, infected, recovered, and the ones with chronic rheumatic symptoms. It is important to remark that the people with chronic rheumatic symptoms do not have the Chikungunya virus, just some symptoms related to it, since they were infected previously. In addition, the population of the vector is divided into three groups: susceptible vector, latent vector, and infected vector. The resulting mathematical model is a nonlinear system of eight ordinary differential equations, which is analyzed to find the equilibrium points and their stability, including the well-known epidemic threshold parameter known as the basic reproduction number, R_0. We estimate some of the unknown key epidemiological parameters of the model from real data, which allow us to compute R_0, defined as the average number of secondary cases generated by a typical infectious individual in a fully-susceptible population. Moreover, we are able to compute approximately how many individuals are infected during an outbreak. All these estimates can help assist with outbreak planning, assessment of health strategies, and the design of future research regarding Chikungunya infection transmission.

Thus, following the basic ideas and structure of mathematical modeling in epidemiology, the Chikungunya model will be developed under the following basic hypotheses [16,17]:

- The total population of humans $N_h(t)$ is divided into five subpopulations: humans who may become infected (susceptible $S_h(t)$), humans exposed, but still not infected due to the existence of an incubation period of the virus (latent $E_h(t)$), humans infected by the Chikungunya virus and that develop the disease (infected $I_h(t)$), humans who have recovered from the Chikungunya infection (recovered $R_h(t)$), and humans who have the disease chronically (chronic $C_h(t)$).
- The parameter μ_h is the birth rate of humans. The birth rate μ_h is assumed equal to natural death d_h.

- The mortality rate increase due to the disease is a real fact. However, since this rate is small in comparison with other rates and is not going to affect the dynamics, we assume that $\epsilon = 0$.
- The total population of mosquitoes $N_v(t)$ is divided into three subpopulations: mosquitoes who may become infected (susceptible $S_v(t)$), mosquitoes in a latent stage (latent $E_v(t)$), and mosquitoes currently infected or spreading the Chikungunya virus (infected $I_v(t)$).
- The parameter μ_v is the birth rate of the mosquitoes, and it is assumed equal to the death rate d_v.
- A susceptible human can transit to the latent subpopulation $E_h(t)$ because of an effective transmission due to a bite of an infected mosquito at a rate of β_1'.
- A susceptible mosquito can be infected if there exists an effective transmission when it bites an infected human, at a rate β_2.
- A fraction α of the latent humans passes to infected by the virus.
- A fraction γ of the infected humans recovers, i.e., they do not have the disease anymore.
- A fraction ρ of the recovered humans moves to the chronic class.
- A fraction ϕ of the latent mosquitoes goes through to infected mosquitoes.
- Homogeneous mixing is assumed, i.e., all susceptible humans have the same probability of being infected and all susceptible mosquitoes have the same probability of being infected.

It is important to notice that the parameter β_1' depends on two different parameter: the bite rate of an infected mosquito on susceptible people and the probability per bite to transmit the virus from the mosquito to the human. On the other hand, β_2 depends on the bite rate of a susceptible mosquito on an infected human and the probability per bite to transmit the virus from the human to the vector. In reality, these parameters are very difficult to estimate due to different environment conditions. Thus, the values of these parameters are most likely to vary from one region to another. In fact, the values might be very different for rural and urban areas. We assume that the natural birth rate is equal to the natural death rate for the human population since we are interested in fitting the model to the real data of the year 2015 and study the identifiability of the parameters. This is considered a short period where the births can be balanced with deaths. For instance, in [25], the authors did not even consider the births and deaths, even though these might affect the dynamics. For mosquitoes, there are no reliable data; thus, we adopted a conservative approach similar to the one used in [25]. The model could be expanded to consider the variation of mosquito populations. This can be done adding a seasonal forcing function [42], but will introduce additional parameters to the model. Furthermore, it will compromise the identifiability more, which currently is an important issue.

Under the above hypotheses, the following diagram illustrates the interactions of the Chikungunya infection in human and mosquito populations; see Figure 1.
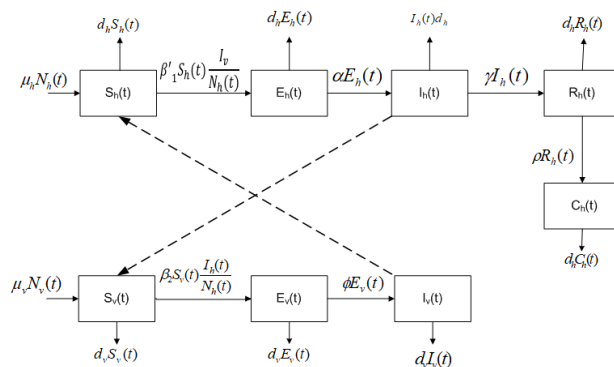

Figure 1. Dynamics of the Chikungunya virus with transmission vector. The boxes represent the subpopulation and the arrows the transition between the subpopulations. Arrows are labeled by their corresponding model parameters.

The corresponding mathematical model is given by the following system of ordinary differential equations:

$$
\begin{aligned}
\dot{S}_h(t) &= \mu_h N_h(t) - S_h(t)(\beta_1' \frac{I_v(t)}{N_h(t)} + d_h), \\
\dot{E}_h(t) &= \beta_1' S_h(t) \frac{I_v(t)}{N_h(t)} - E_h(t)(d_h + \alpha), \\
\dot{I}_h(t) &= \alpha E_h(t) - I_h(t)[d_h + \gamma], \\
\dot{R}_h(t) &= \gamma I_h(t) - R_h(t)(\rho + d_h), \\
\dot{C}_h(t) &= \rho R_h(t) - d_h C_h(t), \\
\dot{S}_v(t) &= \mu_v N_v(t) - S_v(t)(\beta_2 \frac{I_h(t)}{N_h(t)} + d_v), \\
\dot{E}_v(t) &= \beta_2 S_v(t) \frac{I_h(t)}{N_h(t)} - E_v(t)(d_v + \phi), \\
\dot{I}_v(t) &= \phi E_v(t) - d_v I_v(t).
\end{aligned}
\tag{1}
$$

Adding the first five equations, one gets:

$$
\dot{N}_h(t) = \dot{S}_h(t) + \dot{E}_h(t) + \dot{I}_h(t) + \dot{R}_h(t) + \dot{C}_h(t) = 0
$$

and therefore, the human population is constant. Analogously, by adding the last three equations, we have that:

$$
\dot{N}_v(t) = \dot{S}_v(t) + \dot{E}_v(t) + \dot{I}_v(t) = 0
$$

and likewise, the mosquito populations is constant. Thus, we will denote $N_h(t) = N_h$ and $N_v(t) = N_v$, without dependency on time.

All parameters in this model are non-negative. It is easy to prove that the system (1) is well-posed, in the sense that if the initial data ($\dot{S}_h(0), \dot{E}_h(0), \dot{I}_h(0), \dot{R}_h(0), \dot{C}_h(0), \dot{S}_v(0), \dot{E}_v(0), \dot{I}_v(0)$) are in the region $\mathbb{R}_+^8$, then the solutions will be defined for all time $t \geq 0$ and remain in this region.

Normalizing the human and mosquito populations,

$$
s_h(t) = \frac{S_h(t)}{N_h(t)}, e_h(t) = \frac{E_h(t)}{N_h(t)}, i_h(t) = \frac{I_h(t)}{N_h(t)}, r_h(t) = \frac{R_h(t)}{N_h(t)}, c_h(t) = \frac{C_h(t)}{N_h(t)},
$$

and:

$$
s_v(t) = \frac{S_v(t)}{N_v(t)}, e_v(t) = \frac{E_v(t)}{N_v(t)}, i_v(t) = \frac{I_v(t)}{N_v(t)}.
$$

Using the assumptions $\mu_h = d_h$ and $\mu_v = d_v$, one can obtain the following system that describes the dynamics of Chikungunya in each class:

$$
\begin{aligned}
\dot{s}_h(t) &= d_h - s_h(t)(\beta_1 i_v(t) + d_h), \\
\dot{e}_h(t) &= \beta_1 s_h(t) i_v(t) - e_h(t)(d_h + \alpha), \\
\dot{i}_h(t) &= \alpha e_h(t) - i_h(t)[d_h + \gamma], \\
\dot{r}_h(t) &= \gamma i_h(t) - r_h(t)(\rho + d_h), \\
\dot{c}_h(t) &= \rho r_h(t) - d_h c_h(t), \\
\dot{s}_v(t) &= d_v - s_v(t)(\beta_2 i_h(t) + d_v), \\
\dot{e}_v(t) &= \beta_2 s_v(t) i_h(t) - e_v(t)(d_v + \phi), \\
\dot{i}_v(t) &= \phi e_v(t) - d_v i_v(t).
\end{aligned}
\tag{2}
$$

where $\beta_1 = m\beta_1' = \frac{N_v}{N_h}\beta_1'$. Now, we can define a more specific region, $\Omega = \{(s_h(t), e_h(t), i_h(t), c_h(t), s_v(t), e_v(t), i_v(t)) \in [0,1]^8\}$, and the solutions will be defined for all time $t \geq 0$ and remain in this region Ω.

3. Analysis of the Model

3.1. Equilibrium Points and Local Stability of the Chikungunya Mathematical Model

Setting the right-hand side of Equation (2) equal to zero, we can find the equilibrium points of the model. The first point is denoted as the disease-free equilibrium (DFE), and the other is called the endemic equilibrium (EE). Thus, we obtain the equilibrium point $DFE = (1, 0, 0, 0, 1, 0, 0)$. On the other hand, the mathematical expression of the other equilibrium point $EE = (S_h^*, E_h^*, I_h^*, R_h^*, C_h^*, S_v^*, E_v^*, I_v^*)$, is extremely long. Therefore, we will present this endemic equilibrium (EE) in terms of I_h^*. It is important to mention that our main interest is to obtain the conditions under which the population is free of disease. We will discuss later the endemic equilibria and under which conditions they exist.

For infectious diseases, the basic reproduction number is denoted as $\mathcal{R}_0$ and is one of the most useful threshold parameters. It can help determine whether or not an infectious disease will spread throughout a population [43,44]. $\mathcal{R}_0$ also has been used for social epidemics [18,45].

In order to compute the basic reproduction number $\mathcal{R}_0$ for the mathematical model (2), we apply the next generation technique as proposed in [43,44,46]. The infectious classes in this model are $e_h(t)$, $i_h(t)$, $e_v(t)$, and $i_v(t)$. Thus, at the *DFE*, one gets,

$$F = \begin{bmatrix} 0 & 0 & 0 & \beta_1 \\ 0 & 0 & 0 & 0 \\ 0 & \beta_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

and:

$$V = \begin{bmatrix} \alpha + d_h & 0 & 0 & 0 \\ -\alpha & d_h + \gamma & 0 & 0 \\ 0 & 0 & d_v + \phi & 0 \\ 0 & 0 & -\phi & d_v \end{bmatrix},$$

$$K = \begin{bmatrix} 0 & 0 & \frac{\beta_1 \phi}{d_v(d_v + \phi)} & \frac{\beta_1}{d_v} \\ 0 & 0 & 0 & 0 \\ \frac{\beta_2(\alpha d_v^2 + \alpha d_v \phi)}{(\alpha + d_h)d_v(d_h + \gamma)(d_v + \phi)} & \frac{\beta_2}{d_h + \gamma} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

where the matrix F is related to the rate of increase of secondary infections and V to the rate of the disease progression, death, and recovery. The next generation matrix, $K = FV^{-1}$, is non-negative, and therefore, it has a non-negative eigenvalue $\mathcal{R}_0 = \rho(FV^{-1})$ and a non-negative eigenvector ω associated with $\mathcal{R}_0$ [43,46].

There are no other eigenvalues of the matrix K with modulus greater than $\mathcal{R}_0$. Therefore, applying the next generation matrix, we obtain:

$$\mathcal{R}_0 = \sqrt{\frac{\beta_1 \beta_2 \alpha \phi}{(\alpha + d_h)d_v(d_h + \gamma)(d_v + \phi)}}, \tag{3}$$

or in a more compact form:

$$\mathcal{R}_0 = \sqrt{\frac{\beta_1 \beta_2 \alpha \phi}{ABCd_v}},$$

where $A = (\alpha + d_h)$, $B = (d_h + \gamma)$, and $C = (d_v + \phi)$.

Thus, we obtain the following theorem [43,44]:

Theorem 1. *The disease-free equilibrium DFE is locally asymptotically stable for $\mathcal{R}_0 < 1$ and unstable for $\mathcal{R}_0 > 1$.*

This analysis shows that when the threshold parameter R_0 is less than unity, the disease disappears, while the opposite results show an endemic equilibrium in the human population.

Notice that the basic reproduction number $\mathcal{R}_0$ has in the numerator all the factors that contribute to having an epidemic outbreak. For instance, we have the Chikungunya transmissibility β_1 from an infected vector to a susceptible human and the Chikungunya transmissibility β_2 from an infected human to a susceptible vector. In addition, we have the incubation rates of the virus for humans and mosquitoes α and ϕ, respectively. Notice that both chronic and recovered classes have a limited role in the Chikungunya dynamics, since their equations are in some sense uncoupled from the rest of the model and do not play a role in the threshold parameter R_0.

3.2. Endemic Equilibria

In order to find the endemic equilibrium point, we set the right-hand side of Equation (2) equal to zero, and at least one of the infected components of the model (2) is different than zero. Let $EE = (S_h^*, E_h^*, I_h^*, R_h^*, C_h^*, S_v^*, E_v^*, I_v^*)$ represent any arbitrary endemic equilibrium point. Solving the equations for the steady state, we obtain:

$$S_h^* = \frac{ABC(d_v + \beta_2 I_h^*)}{\alpha \phi \mu_v \beta_1 \beta_2}, \qquad E_h^* = \frac{B}{\alpha} I_h^*, \qquad R_h^* = \frac{\gamma}{\rho + d_h} I_h^*,$$

$$C_h^* = \frac{\rho \gamma}{d_h(\rho + d_h)} I_h^*, \qquad S_v^* = \frac{d_v}{d_v + \beta_2 I_h^*}, \qquad E_v^* = \frac{\beta_2 d_v}{C(d_v + \beta_2 I_h^*)}$$

$$I_v^* = \frac{\beta_2 \phi I_h^*}{C(d_v + \beta_2 I_h^*)}.$$

Since we assume that $I_h^* \neq 0$, then we can substitute S_h^* and I_v^* in the first equation of Model (2) at the steady state. Thus, after some calculations, we get,

$$d_h - AB\beta_1 \beta_2 \phi I_h^* - \frac{ABC(d_v + \beta_2 I_h^*)d_h}{\alpha \phi \beta_1 \beta_2} = 0.$$

Then, solving for I_h^*, one gets:

$$I_h^* = \frac{d_h - \left(\dfrac{ABCd_v d_h}{\alpha \phi \beta_1 \beta_2}\right)}{AB\beta_1 \beta_2 \phi + \dfrac{ABCd_h}{\alpha \phi \beta_1}}.$$

Using the threshold parameter $\mathcal{R}_0$, we can rewrite the value of I_h^* as:

$$I_h^* = \frac{d_h \left(1 - \dfrac{1}{\mathcal{R}_0^2}\right)}{AB\beta_1 \beta_2 \phi + \dfrac{ABCd_h}{\alpha \phi \beta_1}}.$$

Clearly, if $\mathcal{R}_0 < 1$, then I_h^* is negative. Thus, in order to have a positive realistic endemic equilibrium point for I_h^*, the threshold parameter $\mathcal{R}_0$ must be greater than one.

3.3. Global Stability Analysis

We can study the global stability of the disease-free equilibrium point $DFE = (1, 0, 0, 0, 0, 1, 0, 0)$ using a Lyapunov function.

Theorem 2. *The disease-free equilibrium DFE is globally asymptotically stable on Ω for $\mathcal{R}_0 < 1$.*

Proof. In order to establish the global stability of the disease-free equilibrium point $DFE = (1, 0, 0, 0, 0, 1, 0, 0)$, we are going to construct a Lyapunov function $V(t)$ [47,48]. The main idea is to find some positive constant weights W_i such that $\dfrac{dV(t)}{dt} < 0$ if $\mathcal{R}_0 < 1$. Let us define the following Lyapunov function:

$$V(t) = W_1\left(s_h - 1 - \log s_h\right) + W_2 e_h + W_3 i_h + W_4\left(s_v - 1 - \log s_v\right) + W_5 e_v + W_6 i_v.$$

Taking the derivative with respect to time along the solutions of the model (2), one gets:

$$\frac{dV(t)}{dt} = W_1\left(1 - \frac{1}{s_h}\right)\left[d_h - s_h\beta_1 i_v - s_h d_h\right] + W_2\left[\beta_1 s_h i_v - e_h(d_h + \alpha)\right] + W_3\left[\alpha e_h - i_h(d_h + \gamma)\right]$$

$$+ W_4\left(1 - \frac{1}{s_v}\right)\left[d_v - s_v\beta_2 i_h - s_v d_v\right] + W_5\left(\beta_2 s_v i_h - e_v(d_v + \phi)\right) + W_6\left(\phi e_v - d_v i_v\right).$$

Rearranging and using $A = (\alpha + d_h)$, $B = (d_h + \gamma)$, and $C = (d_v + \phi)$, one gets:

$$\frac{dV(t)}{dt} = -W_1\beta_1 s_h i_v\left(\frac{(s_h - 1)^2}{s_h}\right) - W_4\beta_2 s_v i_h\left(\frac{(s_v - 1)^2}{s_v}\right) + (W_2 - W_1)\beta_1 s_h i_v$$

$$+ (W_5 - W_4)\beta_2 s_v i_h + (W_3\alpha - W_2 A)e_h + (W_6\phi - W_5 C)e_v$$

$$+ (W_4\beta_2 - W_3 B)i_h + (W_1\beta_1 - W_6 d_v)i_v.$$

Then, we have several options in order to obtain $\dfrac{dV(t)}{dt} < 0$. If we set $W_4 = \dfrac{\alpha\phi\beta_1\beta_2}{ACd_v}$, we can set $W_3 = 1$, and then, the coefficient of i_h will be negative if $\mathcal{R}_0 < 1$. Then, we can set the other weights in order to cancel the other terms. Thus, we can set $W_2 = \dfrac{\alpha}{A}$, $W_5 = W_4$, $W_1 = W_2$, and $W_6 = \dfrac{\alpha\beta_1}{Ad_v}$. Therefore, we have only:

$$\frac{dV(t)}{dt} = -W_1\beta_1 s_h i_v\left(\frac{(s_h - 1)^2}{s_h}\right) - W_4\beta_2 s_v i_h\left(\frac{(s_v - 1)^2}{s_v}\right) + (\mathcal{R}_0 - 1)Bi_h.$$

☐

Theorem 3. *If $\mathcal{R}_0 > 1$, then the epidemiological model (2) is uniformly persistent.*

Proof. Let V be defined as above. When $s_h = 1$ and $s_v = 1$, it follows that:

$$\frac{dV(t)}{dt} > 0,$$

assuming $(\beta_1\beta_2\alpha\phi) - (ABCd_v) > 0$ by hypothesis. Thus, by continuity, $V' > 0$ in a neighborhood of $\{(1, 0, 0, 0, 0, 1, 0, 0)\}$ provided $\mathcal{R}_0 > 1$. Then, the solutions in the positive cone sufficiently close to $\{(1, 0, 0, 0, 0, 1, 0, 0)\}$ move away from $\{(1, 0, 0, 0, 0, 1, 0, 0)\}$, implying that $\{(1, 0, 0, 0, 0, 1, 0, 0)\}$ is unstable and a repeller in Ω. Since $EE = (S_h^*, E_h^*, I_h^*, R_h^*, C_h^*, S_v^*, E_v^*, I_v^*)$ is the only equilibrium that lies

in Ω and is isolated, therefore, by means of Theorem 2.2 [48], we show that the instability of the *DFE* equilibrium point implies the uniform persistence of System (2). $\square$

Notice that the uniform persistence and the positive invariance of the compact set Ω imply the existence of the endemic equilibrium *EE* of (2).

In order to establish the global stability of the endemic equilibrium $EE = (s_h^*, e_h^*, i_h^*, r_h^*, c_h^*, s_v^*, e_v^*, i_v^*)$, we can construct a Lyapunov function. For instance, a general form of Lyapunov functions used in the literature of mathematical biology is $L = \sum_{i=1}^n W_i(x_i - x_i^* - x_i^* \ln \frac{x_i}{x_i^*})$, originally from the first integral of a Lotka–Volterra system [48].

Theorem 4. *If $\mathcal{R}_0 > 1$, then the EE point is globally asymptotically stable on Ω.*

Proof. The main idea is to find some positive constant weights W_i such that $\dfrac{dV(t)}{dt} < 0$ if $\mathcal{R}_0 > 1$. Let us define the following Lyapunov function:

$$V(t) = W_1\left(s_h - s_h^* - s_h^* \log \frac{s_h}{s_h^*}\right) + W_2\left(e_h - e_h^* - e_h^* \log \frac{e_h}{e_h^*}\right) + W_3\left(i_h - i_h^* - i_h^* \log \frac{i_h}{i_h^*}\right)$$

$$+W_4\left(s_v - s_v^* - s_v^* \log \frac{s_v}{s_v^*}\right) + W_5\left(e_v - e_v^* - e_v^* \log \frac{e_v}{e_v^*}\right) + W_6 W_1\left(i_v - i_v^* - i_v^* \log \frac{i_v}{i_v^*}\right).$$

Taking the derivative with respect to time along the solutions of the model (2), one gets:

$$\frac{dV(t)}{dt} = W_1\left(1 - \frac{s_h^*}{s_h}\right)\dot{s}_h(t) + W_2\left(1 - \frac{e_h^*}{e_h}\right)\dot{e}_h(t) + W_3\left(1 - \frac{i_h^*}{i_h}\right)\dot{i}_h(t)$$

$$+W_4\left(1 - \frac{s_v^*}{s_v}\right)\dot{s}_v(t) + W_5\left(1 - \frac{e_v^*}{e_v}\right)\dot{e}_v(t) + W_6\left(1 - \frac{s_v^*}{s_v}\right)\dot{i}_v(t).$$

Using the equations of Model (2), we obtain:

$$\frac{dV(t)}{dt} = W_1\left(1 - \frac{s_h^*}{s_h}\right)\left[d_h - s_h(t)(\beta_1 i_v(t) + d_h)\right] + W_2\left(1 - \frac{e_h^*}{e_h}\right)\left[\beta_1 s_h(t) i_v(t) - e_h(t)(d_h + \alpha)\right]$$

$$+W_3\left(1 - \frac{i_h^*}{i_h}\right)\left[\alpha e_h(t) - i_h(t)[d_h + \gamma]\right] + W_4\left(1 - \frac{s_v^*}{s_v}\right)\left[d_v - s_v(t)(\beta_2 i_h(t) + d_v)\right]$$

$$+W_5\left(1 - \frac{e_v^*}{e_v}\right)\left[\beta_2 s_v(t) i_h(t) - e_v(t)(d_v + \phi)\right] + W_6\left(1 - \frac{s_v^*}{s_v}\right)\left[\phi e_v(t) - d_v i_v(t)\right].$$

Now, setting $W_1 = W_2 = 1$, $W_3 = \dfrac{\beta_1 s_h^* i_v^*}{\alpha e_h^*}$, $W_4 = W_5 = \dfrac{\beta_1 s_h^* i_v^*}{\beta_2 s_v^* i_h^*}$, and $W_6 = \dfrac{\beta_1 s_h^* i_v^*}{\phi e_v^*}$ and using the information regarding the EE point, one gets:

$$\frac{dV(t)}{dt} = d_h s_h^*\left(2 - \frac{s_h}{s_h^*} - \frac{s_h^*}{s_h}\right) + \beta_1 s_h^* i_v^*\left(1 - \frac{s_h^*}{s_h}\right) + \beta_1 s_h^* i_v^*\left(1 - \frac{s_h i_v e_h^*}{s_h^* i_v^* e_h}\right)$$

$$+\beta_1 s_h^* i_v^*\left(1 - \frac{e_h i_h^*}{e_h^* i_h}\right) + \frac{\beta_1 s_h^* i_v^* d_v}{\beta_2 i_h^*}\left(2 - \frac{s_v}{s_v^*} - \frac{s_v^*}{s_v}\right) + \beta_1 s_h^* i_v^*\left(1 - \frac{s_v^*}{s_v}\right)$$

$$+\beta_1 s_h^* i_v^*\left(1 - \frac{s_v i_h e_v^*}{s_v^* i_h^* e_v}\right) + \beta_1 s_h^* i_v^*\left(1 - \frac{e_v i_v^*}{e_v^* i_v}\right).$$

Doing some rearrangement, we obtain:

$$\frac{dV(t)}{dt} = d_h s_h^* \left(2 - \frac{s_h}{s_h^*} - \frac{s_h^*}{s_h} \right) + \frac{\beta_1 s_h^* i_v^* d_v}{\beta_2 i_h^*} \left(2 - \frac{s_v}{s_v^*} - \frac{s_v^*}{s_v} \right)$$

$$+ \beta_1 s_h^* i_v^* \left(6 - \frac{s_h^*}{s_h} - \frac{s_h i_v e_h^*}{s_h^* i_v^* e_h} - \frac{e_h i_h^*}{e_h^* i_h} - \frac{s_v^*}{s_v} - \frac{s_v i_h e_v^*}{s_v^* i_h^* e_v} - \frac{e_v i_v^*}{e_v^* i_v} \right)$$

Since the arithmetic mean is greater than or equal to the geometric mean, we have:

$$\left(2 - \frac{s_h}{s_h^*} - \frac{s_h^*}{s_h} \right) \le 0, \quad \left(2 - \frac{s_v}{s_v^*} - \frac{s_v^*}{s_v} \right) \le 0$$

$$\left(6 - \frac{s_h^*}{s_h} - \frac{s_h i_v e_h^*}{s_h^* i_v^* e_h} - \frac{e_h i_h^*}{e_h^* i_h} - \frac{s_v^*}{s_v} - \frac{s_v i_h e_v^*}{s_v^* i_h^* e_v} - \frac{e_v i_v^*}{e_v^* i_v} \right) \le 0$$

Thus, $\dot{V}(t) \le 0$ for all $(s_h, e_h, i_h, r_h, c_h, s_v, e_v, i_v) \in \Omega$, and the strict equality $\dot{V}(t) = 0$ holds only for $(s_h = s_h^*, e_h = e_h^*, i_h = i_h^*, r_h = r_h^*, c_h = c_h^*, s_v = s_v^*, e_v = e_v^*, i_v = i_v^*)$. Then, the EE point is globally asymptotically stable whenever $\mathcal{R}_0 > 1$. □

4. Numerical Simulation

In this section, a set of numerical simulations using the mathematical model of Chikungunya (2) is performed in order to support the presented theoretical results and validate the importance of the threshold parameter $\mathcal{R}_0$. Those simulations also help us to better understand the relationships among the different groups of the human and vector populations, the parameters, and the dynamics of Chikungunya at the population level. The numerical simulations are presented using the proportions of the subpopulations in Model (2) in order to observe easily the dynamics of Chikungunya. In addition, the simulation results are presented using the parameter values presented in Table 1, which corresponds approximately to the current Colombian scenario. All simulations were done using an adaptive Runge–Kutta–Fehlberg method of order four [49].

Table 1. Parameter values for mathematical model of Chikungunya (2).

Parameter	Symbol	Values	Rate
Average life-span of the human host [50]	$\frac{1}{\mu_h}$	25.000 Days	0.00004
Average life-span of the vector [51]	$\frac{1}{\mu_v}$	14 Days	$0.071\overline{33}$
Average incubation time of the virus (latency in the humans) [52]	$\frac{1}{\alpha}$	5–12 Days	$0.1\overline{33}$
Average incubation time of the virus (latency in the vector) [28,53]	$\frac{1}{\phi}$	3 Days	$0.\overline{33}$
Average infection time (infection in humans) [51]	$\frac{1}{\gamma}$	5–15 Days	0.066
Chronic time [54,55]	$\frac{1}{\rho}$	300 Days	$0.00\overline{33}$

4.1. Numerical Simulations for $\mathcal{R}_0 < 1$

First, we perform numerical simulations of the Chikungunya model (2) when the threshold parameter $\mathcal{R}_0 < 1$ in order to corroborate that the infected populations vanish and to observe the dynamics of the susceptible recovered and chronic populations. We run the simulations for a time long enough so we can observe both the transient dynamics and the steady states of the system.

For this first numerical simulation, the parameter values used are those presented in Table 1, but $\beta_1 = 1/720$. In this particular scenario, the numerical value of β_1 is chosen such that $\mathcal{R}_0 < 1$.

As can deducted directly from the Chikungunya model (2), the units of all these parameters are in days^{-1}, except β_i, which measures the effective contacts per day, i.e., the total number of contacts, effective or not, per unit day, multiplied by the risk of infection with Chikungunya virus.

In Figure 2, it can be observed that for $\mathcal{R}_0 \approx 0.93$, the infected population $I(t)$ dies out, and the susceptible population approaches the disease-free steady state value $S^d = 1$. Moreover, the steady state is as expected the disease-free equilibrium $DFE = (1, 0, 0, 0, 0, 1, 0, 0)$. In this way, if health institutions have to reduce the Chikungunya prevalence in the population, then they need to introduce changes in the health policies that affect the corresponding parameters related to $\mathcal{R}_0$ in order to reduce its value.
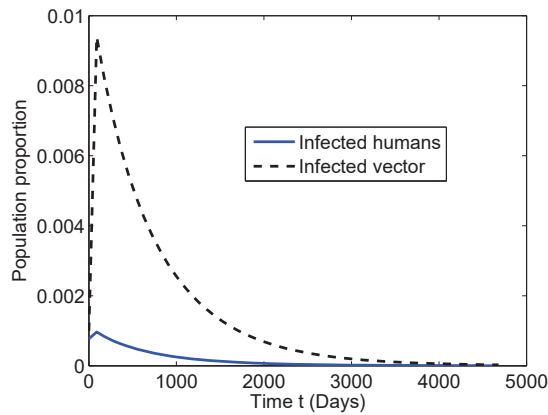
Figure 2. Numerical simulation of the Chikungunya model (2) when the threshold parameter $\mathcal{R}_0 = 0.987$. The parameter values are those presented in Table 1, but $\beta_1 = 1/720$.

4.2. Numerical Simulations for $\mathcal{R}_0 > 1$

The second numerical simulation of the Chikungunya model (2) is when the threshold parameter $\mathcal{R}_0 > 1$. For this particular scenario, the infected populations persist over time, as well as the susceptible and recovered populations. As in the previous scenario, we run the simulations for a time long enough so we can observe both the transient dynamics and the steady states of the system.

In Figure 3, it can be observed that for $\mathcal{R}_0 \approx 1.1977$, the infected subpopulations $i(t)$ and $i_v(t)$ persist and approach an endemic steady state value. However, notice that despite the time-invariant parameter values, the infected subpopulations oscillate at the beginning and then reach a steady state. Thus, some slight changes of the parameter values due to environmental events may change the outcome dynamics. It is important to remark that if health institutions want to reduce the Chikungunya prevalence in the population, then they need to introduce changes in the health policies that affect the corresponding parameters related to $\mathcal{R}_0$ in order to reduce its value.

4.3. Sensitivity Analysis of the Transmission Parameters

Here, we investigate the relevance of each transmission parameter of the Chikungunya model (2) on the dynamics of the subpopulation of humans infected by the virus. To answer this key question, we need to compare the impact of those parameters on the outcome of the infected. As has been mentioned, the threshold parameter $\mathcal{R}_0$ plays a determinant role in the different subpopulations of the model.

In the previous numerical simulation, we can see in Figure 4 that varying the transmission parameter β_2 allows us to observe the effect of this parameter on the dynamics transmission of the disease in the human and mosquito population. In all these simulated scenarios, the human population is exposed to the mosquitoes infected by the Chikungunya virus. Based on the fact that the

human population exposed to the *Aedes aegypti* in Colombia is 19 million [51] and assuming different values of the transmission parameter β_2, we can roughly estimate the number of infected humans by Chikungunya virus, as given in Table 2.
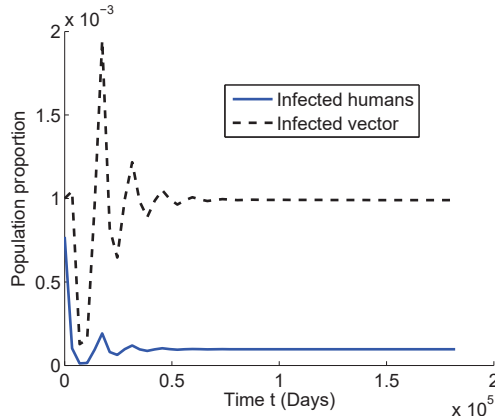
Figure 3. Numerical simulation of the Chikungunya model (2) when the threshold parameter $\mathcal{R}_0 = 1.1977$. The parameter values are those presented in Table 1, but $\beta_1 = 1/600$.
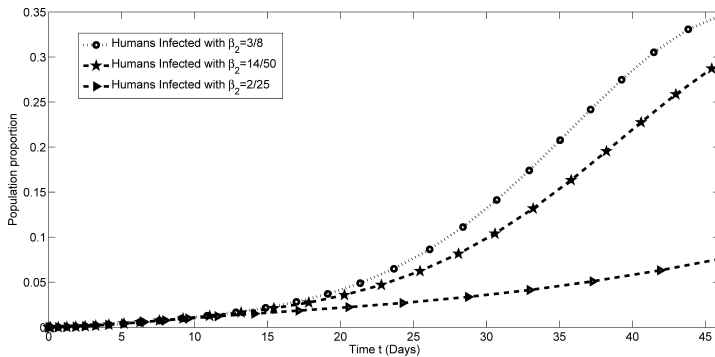
Figure 4. The dynamics of the different subpopulations are sensitive to the changes of the parameter β_2.

Table 2. Variation of the β_2 parameter and its effects on the infected human population.

Value β_2	$\frac{3}{8}$	$\frac{14}{50}$	$\frac{2}{25}$
Value of $\mathcal{R}_0$	4.94	4.27	2.28
Value of Infected Population	7.142 million	6.993 million	5.551 million

5. Estimation of Parameters for the Colombian Scenario

The dynamics of systems are, in general, too complex to allow intuitive predictions and require the support of mathematical modeling for quantitative assessments and a reliable understanding of the system functioning [56]. Moreover, one of the most difficult tasks of mathematical modeling is the estimation of model parameters.

The proposed Chikungunya mathematical model (2) establishes mathematical relationships among the eight different sub-populations and allows for the flux of individuals between

sub-populations. Therefore, several parameters that regulate these relationships and fluxes appear in the model. However, the estimation of these model parameters is not a straightforward task.

5.1. Fitting Algorithm

In order to use the Chikungunya mathematical epidemiological model (2) to simulate the dynamics of Chikungunya virus in the Colombian population, it is necessary to set the parameter values of the model. However, some of the parameter values are not accurately known or belong to different regions. Here, we rely on some parameter values that are available in scientific journals or based on medical considerations. One aim here is to explain the qualitative and quantitative behavior of the Chikungunya infection dynamics at the population level in Colombia for the year 2015.

The Chikungunya disease data were collected using available cumulative data from a national health institution in Colombia. In Table 3, we can see the seroprevalence of Chikungunya infection for different weeks of 2015 in Colombia.

Table 3. Data provided by the National Institute of Health-SIVIGILA, Colombia. The first row corresponds to the number of the week of the year 2015. The second row shows the number of infectious individuals for whom Chikungunya was detected in each week because those people went to see a doctor and it was reported.

Week	1	2	3	4	5	6	7	8
Cases	15,000	16,200	16,100	14,800	13,900	14,000	12,950	12,100
Week	9	10	11	12	13	14	15	16
Cases	13,900	13,000	12,200	11,900	8300	12,200	12,000	11,800
Week	17	18	19	20	21	22	23	24
Cases	13,700	13,600	11,900	11,700	11,500	9800	7900	8000
Week	25	26	27	28	29	30	31	32
Cases	7800	6100	6200	5000	5700	4000	2900	2300
Week	33	34	35	36	37	38	39	40
Cases	2100	2000	1900	1900	1950	1800	1900	1700
Week	41	42	43	44	45	46	47	48
Cases	1750	1750	1650	1600	1650	1500	1600	1000
Week	49	50	51	52				
Cases	1000	1600	1000	900				

In order to adjust the Chikungunya mathematical model (2) to the time-series data of Chikungunya seroprevalence in Colombia, the only parameters to be estimated by a fitting process to real data are the Chikungunya transmissibility β_1 from infected vector to susceptible human, Chikungunya transmissibility β_2 from an infected human to a susceptible vector, and the initial infected human and mosquito proportions. The parameters β depend on the number of bites per unit of time and the transmission probability per bite [57].

In regard to initial conditions for the year 2015, we have several assumptions due to the lack of real data for the vector and human populations. However, we think that the assumptions are biological plausible since we take into account some available data in conjunction with parameter values that are known from the related literature. As we have mentioned before, we used the proportions of the subpopulations in the Chikungunya mathematical model (2). It is important to point out that the data presented in Table 3 are related to individual cases, but we transform these data to proportional values related to the real demographic data of Colombia. We use a total population of 19,471,223, since this is the number of inhabitants that live within the 2200 m where the mosquitoes that transmit the Chikungunya virus live [51,58]. The following initial conditions based on the proportion of infected humans $i_0(i_h(0))$ are assumed: $e_0 = i_0/2$ based on the fact that the latent or exposed stage is half of the infected one; $r_0 = i_0/40$ based on the fact that it takes 300 days for a recovered person to start suffering chronic symptoms and with a probability of approximately 5% [54,55]. Finally, we set the initial condition for the proportion of chronic individuals as $c_0 = 0.05 \times r_0$, and the initial condition for the susceptible population is just found using the relation $s_0 = 1 - e_0 - i_0 - r_0 - c_0$. For the vector

population, we do not have any real data. We will assume that the proportion of vectors at the eclipse phase is a tenth of the infected vectors based on the duration of the latent and infectious stages of the mosquitoes. Here, the simulation interval period of 2015 has been chosen according to the available Chikungunya seroprevalence data in Colombia.

The fitting process to adjust the Chikungunya mathematical model (2) to the time-series data of Chikungunya in Colombia corresponding to the year 2015 is done minimizing the sum of squared errors (SSR). We use two algorithms to find the minimum SSR. We initially used a genetic algorithm [59], which performs a broad search of the parameter space and is less dependent on the initial guess. Once the genetic algorithm had found a good fit, these parameters were used as the initial guess for the trust-region-reflective and interior point algorithms [60,61], which search a more localized region of the parameter space. The use of several different algorithms increases the probability of finding the global minimum for the SSR.

In order to compute the best fitting, we carried out computations, and we implemented the SSR function,

$$\mathbb{F} \quad : \qquad \mathbb{R}^3 \qquad \longrightarrow \qquad \mathbb{R}$$
$$(\beta_1, \beta_2, i_v(0)) \quad \longrightarrow \quad \mathbb{F}(\beta_1, \beta_2, i_v(0))$$

where β_1, β_2, and $i_v(0)$ are variables such that:

1. For a given $(\beta_1, \beta_2, i_v(0))$, numerically solve the system of differential Equation (2) and obtain a solution $\hat{Y}_j(t) = (\hat{s_{hj}}, \hat{e_{hj}}, \hat{i_{hj}}, \hat{r_{hi}}, \hat{c_{hj}}, \hat{s_{vj}}, \hat{e_{vj}}, \hat{i_{vj}})$, which is an approximation of the real data solution $Y(t)$.
2. Set $t_0 = 0$ (the fitting process starts at Week 0), and for $t = 0, 1, 2, \ldots, 51$, corresponding to weeks where data are available, evaluate the computed numerical solution for subpopulation $i_h(t)$; i.e., $\hat{i}_h(0), \hat{i}_h(1), \hat{i}_h(2), \ldots, \hat{i}_h(51)$.
3. Compute the sum of square of the difference between $\hat{i}_h(0), \hat{i}_h(1), \hat{i}_h(2), \ldots, \hat{i}_h(51)$, and infectious data in Table 1. This function $\mathbb{F}$ returns the sum of squared errors (*SSR*), where for the Colombia data are given by:

$$SSR = \sum_{j=0}^{51} (\hat{i}_h(j) - i_h(j))^2$$

4. Find a global minimum for the the the sum of squared errors (*SSR*) using genetic, trust-region-reflective, and interior point algorithms.

The function $\mathbb{F}$ takes values in $\mathbb{R}^3$ and returns a positive real number, the SSR that measures the closeness of the scaled infectious population $(i_h(t))$, provided by the model, to time-series data. In order to ensure that parameter estimates are biologically realistic, we placed bounds on some of the parameters. Thus, for instance, we can obtain values for $i_v(0)$ only from the interval $(0, 1)$ or positive values for β_1 and β_2.

5.2. Numerical Simulation of the Chikungunya Mathematical Model

In this section, numerical results for the solution of the Chikungunya mathematical model (2) are presented. Since no analytic solution to the nonlinear fractional system (2) is available, we use a Runge–Kutta-type method to compute the solution numerically. The data from Colombia related to Chikungunya are from Weeks 1–52. It is important to remark that the data only show a small increasing interval for infected cases and the full decreasing interval. This is due to the fact that at the beginning of the Chikungunya epidemic, the health institutions diagnosed the cases using just symptoms instead of the lab tests that confirm a real infected case. Therefore, health institutions in Colombia disregarded most of the first increasing period in order to avoid inaccurate reports of Chikungunya. Notice that Dengue virus causes similar symptoms to the Chikungunya virus in human populations. Thus, lab tests are necessary to differentiate between both epidemics.

In order to fit the Chikungunya mathematical model (2) to the time-series data of Chikungunya cases in Colombia, we need to estimate the parameters β_1, β_2, and $i_v(0)$. As a first approach, we set

the initial proportion of latent and infected humans based on the real data. Thus, we only need to estimate initially β_1, β_2, and $i_v(0)$. In other works, different assumptions regarding the initial infected and latent populations for humans and vectors have been assumed. For instance, in [62], the authors assumed that the initial numbers of latent and infectious people were equal, and analogously for the mosquito population.

The confirmed cases of Chikungunya from the year 2015 in Colombia and the best model fit along with the best fit parameter values are shown in Figure 5. It can be seen graphically that the Chikungunya mathematical model (2) produces a good adjustment to the real data and predicts that the epidemic will disappear in Colombia due to a $\mathcal{R}_0 = 0.79$. In addition, the epidemic peak of the mathematical model approximates well the peak of the real data. It is important to mention that the Chikungunya mathematical model (2) generates epidemic data that fit well in terms of the SSR. The Chikungunya mathematical model (2) gives a smoother curve due to its deterministic nature, as was expected. In order to catch the natural irregularity of the real data, it would be necessary to introduce stochastic factors to the model, which allows one to obtain a more accurate fitting [63]. However, introducing stochastic or temporal factors would require more detailed information regarding the dynamics of the population in Colombia, and the complexity of the model would increase. It is important to mention that the irregularity of the real data has been observed in many other studies related to other diseases, and can be explained due to different reasons such as weather, under-reporting, the stochastic nature of the virus diffusion, spatial effects, or even the heterogeneity of the human and mosquito population [19]. In Figure 6, the long-term dynamics of the humans with chronic rheumatoid symptoms can be observed. It can be seen that this class population will reach a peak and then start to decrease due to the natural death rate and the decay of new infected humans over the long term.

Additionally, we considered different scenarios fixing the initial conditions for the mosquito infected population $i_v(0)$ and fitting the initial infected human population $i_h(0)$. However, the most important difference with the previous best fit scenario is that the values of the parameters β_2 and β_1 varied in such way that they compensated each other in order to be able to fit the real data level of infected humans. This compensation makes total sense since the transmission parameters β_s are going to depend on the initial infected vector population $i_v(0)$. The smaller $i_v(0)$ is, the greater should be the transmission parameter from vector to humans in order to reproduce the real number of infected cases. However, we notice that the basic reproduction number $\mathcal{R}_0$ stays approximately constant for all the scenarios. Thus, from the results, we obtained a robust numerical value for the basic reproduction number $\mathcal{R}_0$, regardless of the initial proportion of infected mosquitoes.

Finally, we present the numerical simulations for larger values of $i_v(0)$. We vary $i_v(0)$ from 0.2 (20%) to 0.5 (50%) in order to observe other potential realistic scenarios. The idea is to test hypotheses that consider larger values for the initial proportion of infected vectors $i_v(0)$. The best model fits can be seen in Figure 7. It can be observed graphically that the mathematical model (2) does not fit the real data well for large values of $i_v(0)$. Based on these results and the real data available, we can infer that these large value scenarios are unlikely for Colombia in 2015. Moreover, notice that for larger values of the initial proportion of infected vectors $i_v(0)$, the model fits to the real data deteriorate. Thus, we have highlighted some likely scenarios for the parameter values of β_1, β_2, $i_v(0)$, and i_0 that can help to analyze some potential scenarios beyond the data and with some intervention strategies. Moreover, this allows us to compare with different regions where other parameter values have been obtained and then raise some questions regarding the explanations of these differences. In the next section, we will provide further analysis regarding the identifiability of the parameters, the initial conditions, and the basic reproduction number $\mathcal{R}_0$.

Parameter	β_1	β_2	$i_v(0)$	SSR	$\mathcal{R}_0$
Value	0.361	0.012	2.186×10^{-4}	4.701×10^{-7}	0.77

Figure 5. Best fit of the Chikungunya mathematical model (2) to the time-series data of Chikungunya in Colombia corresponding to the year 2015. Red points give the real data, and the blue line shows the best model fit. The best fit parameter values are given in the table. SSR, sum of squared errors.
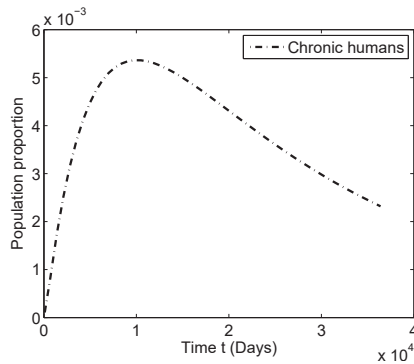
Figure 6. Dynamic of the chronic infected individual humans using the fit of the Chikungunya mathematical model (2) to the time-series data of Chikungunya in Colombia corresponding to the year 2015.
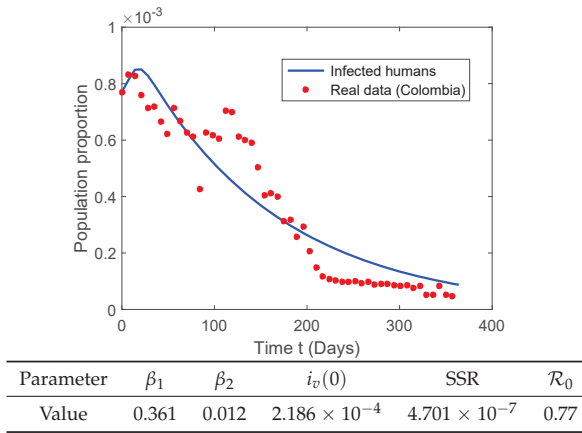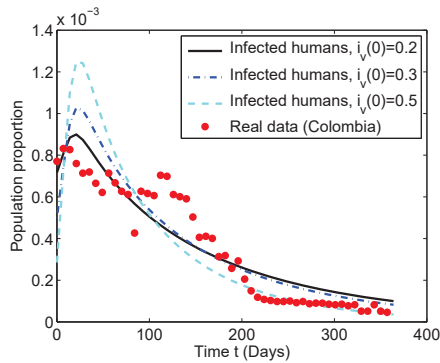

Figure 7. Best fits of the Chikungunya mathematical model (2) to the time-series data of Chikungunya in Colombia corresponding to the year 2015. The initial proportion of infected vectors $i_v(0)$ varies from 0.2 (20%) to 0.5 (50%).

5.3. Identifiability of the Parameters

As we mentioned before, the parameters β_1 and β_2 are not fully identifiable. Moreover, we do not have real data regarding initial conditions for the different subpopulations, but based on our assumptions, all of them can be set in terms of the initial conditions for the infected human and vector populations. Here, we use two approaches: for the first one, we fixed the initial condition of the infected humans to the real data, and consequently, we minimized the function $\mathbb{F}$ in $\mathbb{R}^3$; the second approach was just to leave the initial condition of infected humans as a parameter, and then, we minimized the function $\mathbb{F}$ in $\mathbb{R}^4$.

We now extend our parameters' identifiability analysis, by incorporating two numerical techniques. These techniques are bootstrapping [34,64] and Markov chain Monte Carlo [35,36,39], which allow us to introduce further information regarding the identifiability of the parameters and corroborate the identifiability of the basic reproduction number $\mathcal{R}_0$. It is recommended to ensure that the optimum parameter values of the model can be uniquely determined by the available real data. For non-linear-based models, the issue of identifiability is not straightforward.

An important numerical tool often used to assess the uncertainty in estimated values is bootstrapping [34,64]. The bootstrapping process begins with the generation of artificially-generated datasets, which are created by sampling the best fit curve and adding error such that the SSR is equal to the SSR of the original data. The mathematical model (2) is fitted to each of the surrogated datasets, leading to new parameter estimates. A total of 3000 bootstrap replicates were performed using the real data. In this way, we obtained estimates of the distribution of each parameter, as can be seen in Figure 8. Roughly, it can be seen from the histograms that all the parameter estimates follow a Gaussian distribution. These distributions were used to give the 95% confidence intervals, which are shown in Table 4. This information is useful to have a measure of error in the parameter estimates. There is some small skew present in the histograms of parameters T_h and T_M, which can be explained by their correlation. However, we are more interested in the threshold parameter $\mathcal{R}_0$, which as can be seen in Figure 9, is relatively stable around $\mathcal{R}_0 = 0.78$.

The bootstrapping method allows us to find correlations between estimated parameters. In Figure 9, several two-parameter scatter plots for each of the 3000 bootstrap replicates can be observed. In order to have two uncorrelated parameters, the scatter plot should be roughly a circle, with no clear relationship between the two parameters. For the parameter $i_v(0)$, the plots show a little bit of correlation, and this was in some way expected since $i_v(0)$ is a factor on how strongly the epidemic starts. There is a clear correlation between the transmission parameters β_1 and β_2. This correlation in some way is expected since both transmission parameters are related to the exposure of humans to mosquitoes and vice versa. However, it is important to remark that we are more interested in their product since the epidemic threshold parameter $\mathcal{R}_0$ is proportional to this product. These results suggest that the threshold parameter $\mathcal{R}_0$ of the mathematical model (2)) is identifiable. Thus, we can use the model with caution to describe the dynamics of the spread of Chikungunya in the population of Colombia in the year 2016.

Table 4. Estimated parameters using the mathematical model (2) and the prevalence data of Chikungunya in the population of Colombia in the year 2016.

Parameters	$\mathcal{R}_0$	$i_v(0)$
Values	0.78 (0.74–0.84)	2.18×10^{-4} (1.95×10^{-4}–2.45×10^{-4})
Parameters	β_1	β_2
Values	0.36 (0.34–0.38)	0.012 (0.011–0.13)

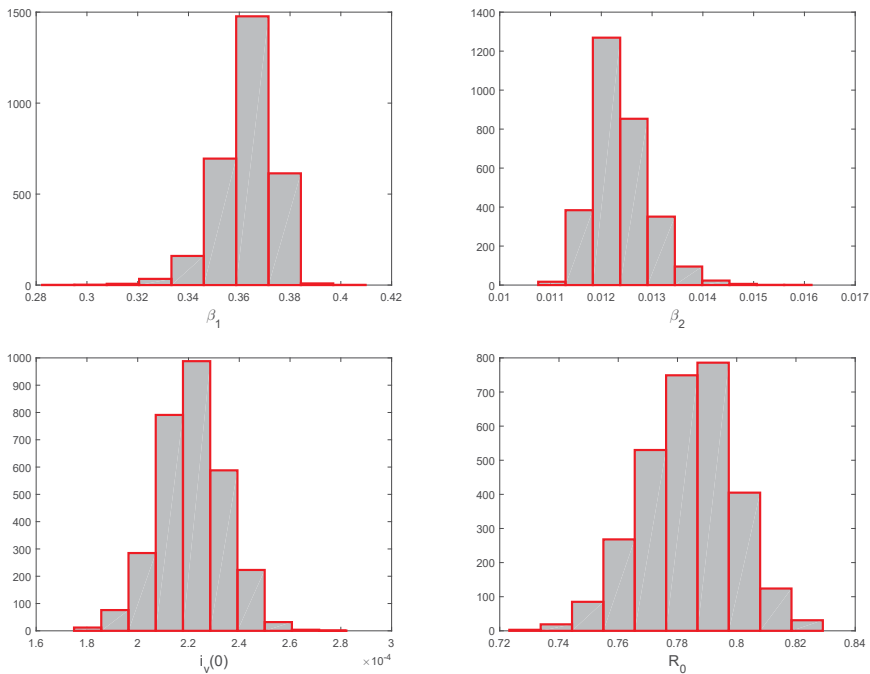The 95% confidence intervals (bootstrap fits) are given in parentheses.

Figure 8. Identifiability assessment of the mathematical model (2) fit to the prevalence data of Chikungunya in the population of Colombia in the year 2016. Correlation plots generated with parameter estimates from bootstrap fits.

The second numerical technique is based on Markov chain Monte Carlo [35,36]. In particular, we used a stable algorithm of the affine-invariant ensemble sampler for Markov chain Monte Carlo (MCMC) proposed in [65]. An algorithm that is affine invariant performs equally well under all linear transformations; it will therefore be insensitive to covariances among parameters [66]. Markov chain Monte Carlo is designed to sample from—and thereby provide sampling approximations to–the posterior PDF efficiently [35,66]. In the algorithm that is used here, we draw samples from a multivariate Gaussian density. It is necessary to set up the specific values of the hyperparameters in three dimensions $(\beta_1, \beta_2, i_v(0))$. In addition, we need to decide how many walkers there are, which are in some way independent paths to reach the maximum of the likelihood function [65]. There are different options to initialize each of the walkers. One of the best techniques is to start in a small ball around the a priori preferred position [66]. It is important to mention that the walkers will spread out and explore the whole space for the parameters. The main goal is to find the maximum of the likelihood function, which is a Gaussian where the variance is underestimated by some fractional amount. We can use Markov Chain Monte Carlo to obtain estimates and confidence intervals for each of the parameters β_1, β_2, and $i_v(0)$. Moreover, we can observe some possible correlations among the parameters.

In Figure 10, all the one- and two-dimensional projections of the posterior probability distributions of the parameters can be seen. This allows us to observe all of the covariances between parameters. In addition, the histograms show the marginalized distribution for a parameter or a set of parameters using the results of Markov chain Monte Carlo. Thus, Figure 10 shows the marginalized distribution for each parameter independently in the histograms along the diagonal and then the marginalized two-dimensional distributions in the other panels [66]. Notice that the maximum likelihood function profile is also presented. We can say that in order to have two uncorrelated parameters, the level

curves' plot should be close to circles, with no clear relationship between the two parameters. For the parameter $i_v(0)$, the plots show less correlation in comparison with the ones with respect to β_1 and β_2, and this was in some way expected since $i_v(0)$ is a factor of how strongly the epidemic starts. There is a clear correlation between the transmission parameters β_1 and β_2, as we also observed with the bootstrapping results. In fact, this relationship has the same pattern. This correlation in some way is expected since both transmission parameters are related to the exposure of humans to mosquitoes and vice versa. However, it is important to remark that we are more interested in their product since the epidemic threshold parameter $\mathcal{R}_0$ is proportional to this product, as we have mentioned before. These results agree with previous results regarding identifiability for epidemic models with vectors [67].

Figure 9. Identifiability assessment of the mathematical model (2) fit to the prevalence data of Chikungunya in the population of Colombia in the year 2016. Correlation plots generated with parameter estimates from bootstrap fits. Correlation between the parameters β_1, β_2, $i_M(0)$, and $\mathcal{R}_0$.
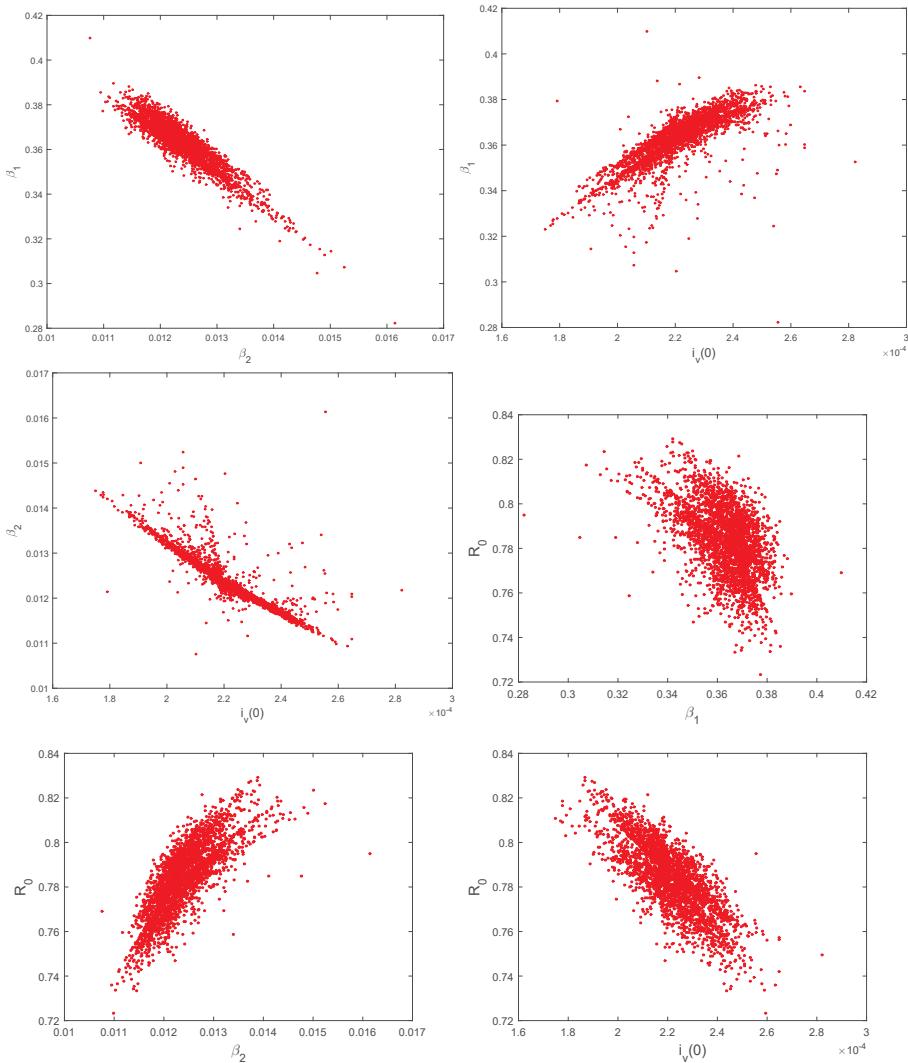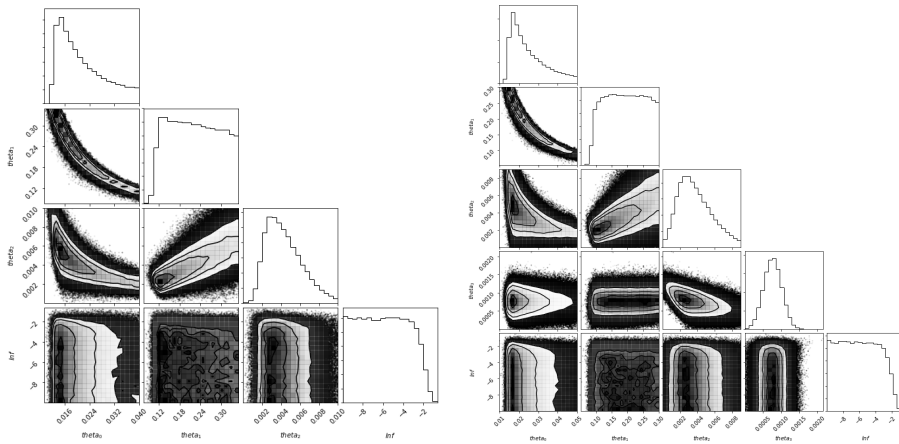
Figure 10. Identifiability assessment of the mathematical model (2) fit to the prevalence data of Chikungunya in the population of Colombia in the year 2016. Correlation plots generated with Markov chain Monte Carlo for three (β_1, β_2, $i_v(0)$) and four parameters (β_1, β_2, $i_v(0)$, $i_h(0)$), respectively.

6. Conclusions

We present a mathematical model of the spread of the Chikungunya disease at the population level that incorporates the transmission vector by including cross-transmission between the human and vector populations. The proposed model includes a chronic subpopulation, which to the best of our knowledge has not been considered in other mathematical models. We determined the epidemic threshold parameter $\mathcal{R}_0$ for the extinction of disease using the method of the next generation matrix. Using Lyapunov function theory, some sufficient conditions for the global stability of the the disease-free equilibrium were obtained. Based on this parameter, we found the parameters that affect the basic reproduction number $\mathcal{R}_0$ and therefore what would be the best policies to control the spread of the Chikungunya disease. We verified that when the threshold parameter $\mathcal{R}_0$ is less than unity, the disease disappears, while when threshold parameter values are larger than one, the disease persists in the population. Numerical simulations were presented to support the established theoretical results.

Using the proposed mathematical model of Chikungunya diffusion, we were able to analyze the dynamics of infection during the 2015 outbreak in Colombia. In particular, we estimated the numerical values of the epidemiological parameters β_1, β_2, and the reproduction number $\mathcal{R}_0$. Based on numerical results of the model of Chikungunya, we were able to better explain and understand the variation of the number of infected cases with the Chikungunya virus in Colombia. We found that the transmission parameters β_1 and β_2 were highly correlated.

Our estimated values for the reproduction number $\mathcal{R}_0$ ranged from 0.74–0.84 across the three different scenarios proposed and also from the results of bootstrapping and Markov chain Monte Carlo methods. However, notice that rain seasons have been changing recently due to climate change, and that could affect some of the parameter values, which is one limitation of this model. The fitting of the model to the observed weekly reports during the 2015 outbreak in Colombia is relatively good, despite the natural irregularity of the real data. This irregularity of the data has been observed in many other studies related to other diseases and can be due to different reasons such as weather, under-reporting, the stochastic nature of the virus diffusion, spatial effects, or even the heterogeneity of the human and mosquito population. All these factors were not considered here explicitly and are other limitations of this work. Other limitations are the fact that the initial proportion of the different subpopulations is unknown. More complex models are necessary to study all the aforementioned factors, and future works might consider these factors or at least some of them. Another important

aspect that we found in this study, based on the previous numerical simulations of the different scenarios, is that the initial infected population of mosquitoes is likely no greater than 20% for the year 2015 in Colombia. To the best of our knowledge, we do not know if there are real data regarding this particular description.

In this paper, we performed identifiability analysis in order to estimate some parameters of the model and found correlations among the parameters of the model. We used the bootstrapping technique and Markov chain Monte Carlo in order to assess the identifiability of the parameters of the model. We found out that the parameters were not fully identifiable with the prevalence of the real data that we had. However, we were able to identify that the reproduction number $\mathcal{R}_0$ ranged from 0.74–0.84. It is important to notice that this result depends on some fixed parameter values such as the duration of the infectious stage and the life-span of the mosquitoes. Changing values for these parameters can change the aforementioned range for the reproduction number $\mathcal{R}_0$. However, regardless of some potential changes, we presented a study with a methodology to deal with these types of epidemic models when real data are available.

Finally, we can conclude that in order to reduce the Chikungunya disease diffusion, it is important first to decrease the transmission parameters. For instance, people could use clothes covering most of the body to avoid mosquito bits and could also use repellents. Any action that could reduce the infected mosquitoes' bites will affect the reproduction number $\mathcal{R}_0$. The aforementioned actions would reduce the transmission of the virus in both directions. Another way to reduce the value of the reproduction number $\mathcal{R}_0$ is to increase the mortality of the mosquito population. Therefore, a health policy that could be implemented is to increase this mortality. Other parameters that are included in $\mathcal{R}_0$ such as incubation or infected period are natural characteristics of the Chikungunya virus and are difficult to modify.

References

1. Zárate, M.L. Arbovirus y Arbovirosis en México. Instituto de Salubridad y Enfermedades Tropicales. Available online: www.fmvz.unam.mx/fmvz/cienciavet/revistas-/CVvol2/CVv2c6.pdf (accessed on 2 January 2019).

2. Jupp, P.; McIntosh, B. Chikungunya virus disease. In *The Arboviruses: Epidemiology and Ecology*; CRC Press: Boca Raton, FL, USA, 1988; Volume 2, pp. 137–157.

3. Pérez Sánchez, G.; Ramírez Alvarez, G.; Pérez Gijón, Y.; Canela Lluch, C. Fiebre de Chikungunya: Enfermedad infrecuente como emergencia médica en Cuba. *Medisan* **2014**, *18*, 848–856.

4. Cook, G.C.; Zumla, A. *Manson's Tropical Diseases*, 13th ed.; Elsevier Health Sciences: Philadelfia, PA, USA, 2008.

5. Benedict, M.; Levine, R.; Hawley, W.; Lounibos, P. Spread of the tiger: Global risk of invasion by the mosquito Aedes albopictus. *Vector Borne Zoonotic Dis.* **2007**, *7*, 76–85. [CrossRef] [PubMed]

6. Robinson, M.; Conan, A.; Duong, V.; Ly, S.; Ngan, C.; Buchy, P.; Tarantola, A.; Rodo, X. A model for a chikungunya outbreak in a rural Cambodian setting: Implications for disease control in uninfected areas. *PLoS Negl. Trop. Dis.* **2014**, *8*, e3120. [CrossRef] [PubMed]

7. Charrel, R.N.; De Lamballerie, X.; Raoult, D. Seasonality of mosquitoes and chikungunya in Italy. *Lancet Infect. Dis.* **2008**, *8*, 5–6. [CrossRef]

8. Halstead, S.B. Reappearance of chikungunya, formerly called dengue, in the Americas. *Emerg. Infect. Dis.* **2015**, *21*, 557–561. [CrossRef]

9. Teng, T.S.; Kam, Y.W.; Lee, B.; Hapuarachchi, H.C.; Wimal, A.; Ng, L.C.; Ng, L.F. A systematic meta-analysis of immune signatures in patients with acute chikungunya virus infection. *J. Infect. Dis.* **2015**, *211*, 1925–1935. [CrossRef] [PubMed]

10. Sissoko, D.; Malvy, D.; Ezzedine, K.; Renault, P.; Moscetti, F.; Ledrans, M.; Pierre, V. Post-epidemic Chikungunya disease on Reunion Island: Course of rheumatic manifestations and associated factors over a 15-month period. *PLoS Negl. Trop. Dis.* **2009**, *3*, e389. [CrossRef]

11. Montero, A. Chikungunya fever—A new global threat. *Med. Clín. (Eng. Ed.)* **2015**, *145*, 118–123. [CrossRef]

12. Poo, Y.S.; Rudd, P.A.; Gardner, J.; Wilson, J.A.; Larcher, T.; Colle, M.A.; Le, T.T.; Nakaya, H.I.; Warrilow, D.; Allcock, R.; et al. Multiple immune factors are involved in controlling acute and chronic chikungunya virus infection. *PLoS Negl. Trop. Dis.* **2014**, *8*, e3354. [CrossRef]

13. Taubitz, W.; Cramer, J.P.; Kapaun, A.; Pfeffer, M.; Drosten, C.; Dobler, G.; Burchard, G.D.; Löscher, T. Chikungunya fever in travelers: Clinical presentation and course. *Clin. Infect. Dis.* **2007**, *45*, e1–e4. [CrossRef]

14. García, G.; Santana, E.; Martínez, J.R.; Álvarez, A.; Rodríguez, R.; Guzmán, M.G. Detección de respuesta linfoproliferativa en monos inoculados con virus dengue 4. *Rev. Cubana Med. Trop.* **2003**, *55*, 27–29. [PubMed]

15. Manore, C.A.; Ostfeld, R.S.; Agusto, F.B.; Gaff, H.; LaDeau, S.L. Defining the risk of Zika and chikungunya virus transmission in human population centers of the eastern United States. *PLoS Negl. Trop. Dis.* **2017**, *11*, e0005255. [CrossRef] [PubMed]

16. Murray, J.D. *Mathematical Biology I: An Introduction*, Vol. 17 of Interdisciplinary Applied Mathematics; Springer: New York, NY, USA, 2002.

17. Castillo-Chavez, C.; Brauer, F. *Mathematical Models in Population Biology and Epidemiology*; Springer: New York, NY, USA, 2012.

18. Hethcote, H.W. Mathematics of infectious diseases. *SIAM Rev.* **2005**, *42*, 599–653. [CrossRef]

19. González-Parra, G.; Arenas, A.J.; Aranda, D.F.; Segovia, L. Modeling the epidemic waves of AH1N1/09 influenza around the world. *Spat. Spatio-Temporal Epidemiol.* **2011**, *2*, 219–226. [CrossRef] [PubMed]

20. Lashari, A.A.; Hattaf, K.; Zaman, G. A delay differential equation model of a vector borne disease with direct transmission. *Int. J. Ecol. Econ. Stat.* **2012**, *27*, 25–35.

21. Lashari, A.A.; Hattaf, K.; Zaman, G.; Li, X.Z. Backward bifurcation and optimal control of a vector borne disease. *Appl. Math. Inf. Sci.* **2013**, *7*, 301–309. [CrossRef]

22. Ullah, S.; Khan, M.A.; Farooq, M. A fractional model for the dynamics of TB virus. *Chaos Solitons Fractals* **2018**, *116*, 63–71. [CrossRef]

23. Sardar, T.; Rana, S.; Bhattacharya, S.; Al-Khaled, K.; Chattopadhyay, J. A generic model for a single strain mosquito-transmitted disease with memory on the host and the vector. *Math. Biosci.* **2015**, *263*, 18–36. [CrossRef]

24. Wojtak, W.; Silva, C.J.; Torres, D.F. Uniform asymptotic stability of a fractional tuberculosis model. *Math. Model. Nat. Phenom.* **2018**, *13*, 9. [CrossRef]

25. Yakob, L.; Clements, A.C. A mathematical model of chikungunya dynamics and control: The major epidemic on Réunion Island. *PLoS ONE* **2013**, *8*, e57448. [CrossRef]

26. Agusto, F.B.; Easley, S.; Freeman, K.; Thomas, M. Mathematical Model of Three Age-Structured Transmission Dynamics of Chikungunya Virus. *Comput. Math. Methods Med.* **2016**, *2016*, 4320514. [CrossRef]

27. Lashari, A.A.; Zaman, G. Global dynamics of vector-borne diseases with horizontal transmission in host population. *Comput. Math. Appl.* **2011**, *61*, 745–754. [CrossRef]

28. Dumont, Y.; Chiroleu, F.; Domerg, C. On a temporal model for the Chikungunya disease: Modeling, theory and numerics. *Math. Biosci.* **2008**, *213*, 80–91. [CrossRef] [PubMed]

29. Moulay, D.; Aziz-Alaoui, M.; Cadivel, M. The chikungunya disease: Modeling, vector and transmission global dynamics. *Math. Biosci.* **2011**, *229*, 50–63. [CrossRef] [PubMed]

30. Sawabe, K.; Isawa, H.; Hoshino, K.; Sasaki, T.; Roychoudhury, S.; Higa, Y.; Kasai, S.; Tsuda, Y.; Nishiumi, I.; Hisai, N.; et al. Host-feeding habits of Culex pipiens and Aedes albopictus (Diptera: Culicidae) collected at the urban and suburban residential areas of Japan. *J. Med. Entomol.* **2010**, *47*, 442–450. [CrossRef]

31. Vourc'h, G.; Halos, L.; Desvars, A.; Boué, F.; Pascal, M.; Lecollinet, S.; Zientara, S.; Duval, T.; Nzonza, A.; Brémont, M. Chikungunya antibodies detected in non-human primates and rats in three Indian Ocean islands after the 2006 ChikV outbreak. *Vet. Res.* **2014**, *45*, 52. [CrossRef] [PubMed]

32. Ligon, B.L. Reemergence of an unusual disease: The chikungunya epidemic. *Semin. Pediatr. Infect. Dis.* **2006**, *17*, 99–104. [CrossRef] [PubMed]

33. Diallo, M.; Thonnon, J.; Traore-Lamizana, M.; Fontenille, D. Vectors of Chikungunya virus in Senegal: Current data and transmission cycles. *Am. J. Trop. Med. Hyg.* **1999**, *60*, 281–286. [CrossRef]

34. Efron, B. Bootstrap Methods: Another Look at the Jackknife. *Ann. Stat.* **1979**, *7*, 1–26. [CrossRef]

35. MacKay, D.J.; Mac Kay, D.J. *Information Theory, Inference and Learning Algorithms*; Cambridge University Press: Cambridge, UK, 2003.

36. Tulu, T.W.; Tian, B.; Wu, Z. Mathematical modeling, analysis and Markov Chain Monte Carlo simulation of Ebola epidemics. *Res. Phys.* **2017**, *7*, 962–968. [CrossRef]

37. Dhamala, J.; Arevalo, H.J.; Sapp, J.; Horácek, B.M.; Wu, K.C.; Trayanova, N.A.; Wang, L. Quantifying the uncertainty in model parameters using Gaussian process-based Markov chain Monte Carlo in cardiac electrophysiology. *Med. Image Anal.* **2018**, *48*, 43–57. [CrossRef] [PubMed]

38. Kroese, D.P.; Brereton, T.; Taimre, T.; Botev, Z.I. Why the Monte Carlo method is so important today. *Wiley Interdisc. Rev. Comput. Stat.* **2014**, *6*, 386–392. [CrossRef]

39. Bretó, C. Modeling and inference for infectious disease dynamics: A likelihood-based approach. *Stat. Sci.* **2018**, *33*, 57–69.

40. Ortiz, G.A.; Alvarez, D.A.; Bedoya-Ruíz, D. Identification of Bouc–Wen type models using the transitional Markov chain Monte Carlo method. *Comput. Struct.* **2015**, *146*, 252–269. [CrossRef]

41. Weber, C. Characterization of the Chikungunya Virus Entry Process and The Development of Novel Antiviral Strategies. Ph.D. Thesis, Univ.-Bibliothek Frankfurt am Main, Frankfurt, Germany, 2015.

42. Arenas, A.J.; González-Parra, G.; Moraño, J.A. Stochastic modeling of the transmission of respiratory syncytial virus (RSV) in the region of Valencia, Spain. *Biosystems* **2009**, *96*, 206–212. [CrossRef] [PubMed]

43. Van den Driessche, P.; Watmough, J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math. Biosci.* **2002**, *180*, 29–48. [CrossRef]

44. Diekmann, O.; Heesterbeek, J.A.P.; Roberts, M. The construction of next-generation matrices for compartmental epidemic models. *J. R. Soc. Interface* **2010**, *7*, 873–885. [CrossRef]

45. Guerrero, F.; González-Parra, G.; Arenas, A.J. A nonstandard finite difference numerical scheme applied to a mathematical model of the prevalence of smoking in Spain: A case study. *Comput. Appl. Math.* **2014**, *33*, 13–25. [CrossRef]

46. Van den Driessche, P.; Watmough, J. Further notes on the basic reproduction number. In *Mathematical Epidemiology*; Springer: New York, NY, USA, 2008; pp. 159–178.

47. Korobeinikov, A.; Maini, P.K. A Lyapunov function and global properties for SIR and SEIR epidemiological models with nonlinear incidence. *Math. Biosci. Eng.* **2004**, *1*, 57–60.

48. Shuai, Z.; van den Driessche, P. Global stability of infectious disease models using Lyapunov functions. *SIAM J. Appl. Math.* **2013**, *73*, 1513–1532. [CrossRef]

49. Kincaid, D.R.; Cheney, E.W. *Numerical Analysis: Mathematics of Scientific Computing*; American Mathematical Soc.: Providence, RI, USA, 2002; Volume 2.

50. Duque, J.E.; Navarro-Silva, M.A. Simulating management of Aedes aegypti (Diptera: Culicidae) and its effects in a dengue epidemic. *Revista Colombiana de Entomología* **2009**, *35*, 66–72.

51. Cardona-Ospina, J.A.; Henao-SanMartin, V.; Paniz-Mondolfi, A.E.; Rodríguez-Morales, A.J. Mortality and fatality due to Chikungunya virus infection in Colombia. *J. Clin. Virol.* **2015**, *70*, 14–15. [CrossRef] [PubMed]

52. Vazeille, M.; Moutailler, S.; Coudrier, D.; Rousseaux, C.; Khun, H.; Huerre, M.; Thiria, J.; Dehecq, J.S.; Fontenille, D.; Schuffenecker, I.; et al. Two Chikungunya isolates from the outbreak of La Reunion (Indian Ocean) exhibit different patterns of infection in the mosquito, Aedes albopictus. *PLoS ONE* **2007**, *2*, e1168. [CrossRef] [PubMed]

53. Chikaki, E.; Ishikawa, H. A dengue transmission model in Thailand considering sequential infections with all four serotypes. *J. Infect. Dev. Countries* **2009**, *3*, 711–722. [CrossRef]

54. Gutierrez, M.F. Personal Communication, 2017.

55. Plan Nacional de respuesta frente a la introducción del virus de chikungunya en Colombia. Available online: https://www.minsalud.gov.co (accessed on 2 January 2019).

56. Chou, I.C.; Voit, E.O. Recent developments in parameter estimation and structure identification of biochemical and genomic systems. *Math. Biosci.* **2009**, *219*, 57–83. [CrossRef] [PubMed]

57. Cruz-Pacheco, G.; Esteva, L.; Vargas, C. Control measures for Chagas disease. *Math. Biosci.* **2012**, *237*, 49–60. [CrossRef] [PubMed]

58. Censo Colombia 2005. Available online: https://www.dane.gov.co/files/censos/libroCenso2005nacional.pdf (accessed on 2 January 2019).

59. Golberg, D.E. *Genetic Algorithms in Search, Optimization, and Machine Learning*; Addion Wesley: Boston, MA, USA, 1989.

60. Coleman, T.F.; Li, Y. An Interior Trust Region Approach for Nonlinear Minimization Subject to Bounds. *SIAM J. Optim.* **1996**, *6*, 418–445. [CrossRef]

61. Press, W.H.; Teukolsky, S.A.; Vetterling, W.T.; Flannery, B.P. *Numerical Recipes: The Art of Scientific Computing*; Cambridge University Press: Cambridge, UK, 1992.

62. Kucharski, A.J.; Funk, S.; Eggo, R.M.; Mallet, H.P.; Edmunds, W.J.; Nilles, E.J. Transmission dynamics of Zika virus in island populations: A modelling analysis of the 2013–14 French Polynesia outbreak. *PLoS Negl. Trop. Dis.* **2016**, *10*, e0004726. [CrossRef]

63. González-Parra, G.; Villanueva, R.J.; Ruiz-Baragaño, J.; Moraño, J.A. Modelling influenza A(H1N1) 2009 epidemics using a random network in a distributed computing environment. *Acta Trop.* **2015**, *143*, 29–35. [CrossRef]

64. Davison, A.C.; Hinkley, D.V. *Bootstrap Methods and Their Application*; Cambridge University Press: Cambridge, UK, 1997; Volume 1.

65. Goodman, J.; Weare, J. Ensemble samplers with affine invariance. *Commun. Appl. Math. Comput. Sci.* **2010**, *5*, 65–80. [CrossRef]

66. Foreman-Mackey, D.; Hogg, D.W.; Lang, D.; Goodman, J. emcee: The MCMC hammer. *Publ. Astron. Soc. Pac.* **2013**, *125*, 306. [CrossRef]

67. Kao, Y.H.; Eisenberg, M.C. Practical unidentifiability of a simple vector-borne disease model: Implications for parameter estimation and intervention assessment. *Epidemics* **2018**, *25*, 89–100. [CrossRef] [PubMed]

Article

Modeling the Macrophage-Mediated Inflammation Involved in the Bone Fracture Healing Process

Imelda Trejo *, **Hristo Kojouharov and Benito Chen-Charpentier**

Department of Mathematics, The University of Texas at Arlington, P.O. Box 19408, Arlington, TX 76019-0408, USA; hristo@uta.edu (H.K.); bmchen@uta.edu (B.C.-C.)
*   Correspondence: imelda.trejo@mavs.uta.edu; Tel.: +1-817-272-3261

Abstract: A new mathematical model is presented to study the effects of macrophages on the bone fracture healing process. The model consists of a system of nonlinear ordinary differential equations that represents the interactions among classically and alternatively activated macrophages, mesenchymal stem cells, osteoblasts, and pro- and anti-inflammatory cytokines. A qualitative analysis of the model is performed to determine the equilibria and their corresponding stability properties. Numerical simulations are also presented to support the theoretical results, and to monitor the evolution of a broken bone for different types of fractures under various medical interventions. The model can be used to guide clinical experiments and to explore possible medical treatments that accelerate the bone fracture healing process, either by surgical interventions or drug administrations.

Keywords: bone repair; macrophages; immune system; cytokines; stem cells

1. Introduction

Bone fractures are becoming a serious worldwide problem, due to their high frequency and surgical complications. Globally, more than 8.9 million fractures occur every year, where 10–15% of them result in nonunion [1–5]. Prolonged healing, disabilities, and high morbidity rates are associated with severe traumas and immune-compromised-fractured people [2,3,6–8]. In addition, medical care costs for bone fractures are expected to be over US$25 billions by 2025; due, in part, to the expensive treatments and prolonged hospitalization and rehabilitation [3,9]. It is essential to have a better understanding of the bone fracture healing process, in order to prevent unsuccessful healing and to develop optimal fracture union treatments.

Recently, experimental and mathematical models have demonstrated that macrophages strongly regulate bone fracture healing [10,11]. Depletion of macrophages results in delayed bone formation [12]. Furthermore, during inflammation, classically activated macrophages attract mesenchymal lineage cells (MSCs) to the injury site and activate the healing process [3,10,12]. In contrast, during the repair phase, alternatively activated macrophages promote MSC proliferation and differentiation, and accelerate bone healing [10,13]. However, the exact mechanisms by which macrophages contribute to bone healing remain unclear [3,10,14]. Also, the interaction between macrophages and tissue cells, as well as the importance of classically and alternatively activated macrophages during the bone healing process, are still not clearly understood [10,12,13].

In [11], a mathematical model, based on the interactions among macrophages, MSCs, and osteoblasts, was developed to study the regulatory effects of two generic pro- and anti-inflammatory cytokines during the early stages of bone fracture healing. To our knowledge, it was the first attempt to incorporate the macrophage interactions in the modeling of the bone fracture healing process. The mathematical model revealed that high concentrations of pro-inflammatory cytokines negatively affect the healing time of a fracture, while the administration of anti-inflammatory cytokines can accelerate the healing time in a

dose-dependent manner. Therefore, it is important to carefully consider and incorporate in the modeling approach all sources of pro- and anti-inflammatory cytokines, such as macrophages [2,3,15], in order to correctly represent the complex progression of the bone fracture healing process.

In this paper, the mathematical model developed in [11] is extended, to separately incorporate the two different phenotypes of macrophages: Classically and alternatively activated macrophages, as they have distinct functions during the healing process [10,15,16]. Classically activated macrophages release high levels of pro-inflammatory cytokines, including TNF-α and IL-1β, which exhibit inhibitory and destructive properties in high concentrations [16,17]. In contrast, alternatively activated macrophages are characterized by the secretion of anti-inflammatory cytokines, such as IL-10, which increase their phagocytic activities, mitigate the inflammatory responses, promote growth, and accelerate fracture healing [2,3,15,16]. This extension leads to a more realistic model, by incorporating the different phagocytic rates and the separate production of the pro- and anti-inflammatory cytokines by the two types of macrophages [15,18]. The model can be used to investigate the macrophage functions during inflammation and their effects during the bone fracture healing process. The model can also be used to investigate potential therapeutic treatments, based on the use of anti-inflammatory cytokines, stem cells, and macrophages, suggesting possible ways to guide clinical experiments and bone tissue engineering strategies [15,16].

The organization of the paper is as follows: Section 2 discusses the cellular and molecular interactions that occur during the bone fracture healing process. The macrophage-mediated inflammation involved in the bone fracture healing process is also described in detail. The simplifying assumptions are presented in Section 3. In Section 4, a system of nonlinear ordinary differential equations is introduced, to mathematically describe the fundamental aspects of the bone fracture healing process during the resolution of inflammation and bone repair. The stability analysis of the system is presented in Section 5. Section 6 demonstrates the functionality of the model, by numerically simulating the progression of the bone fracture healing process under normal and pathological conditions. The discussion and future work are presented in Section 7.

2. Biological Background

Bone fracture healing is a complex biological process, which involves the participation of different cell types (including the immune system and mesenchymal lineage cells [16]) and is strongly regulated by released molecular factors [10,16,19–21]. Particularly, at the beginning of the healing process, cytokines activate and direct both immune and tissue cellular functions.

Cytokines are functionally classified into pro-inflammatory and anti-inflammatory families. Pro-inflammatory cytokines, such as the tumor necrotic factor-α (TNF-α), activate the immune system defense to kill bacteria and fight infections. Anti-inflammatory cytokines block the pro-inflammatory synthesis and activate the mesenchymal lineage cellular functions [2]. Interleukin-10 (IL-10) is one of the most potent anti-inflammatory molecules that inhibits pro-inflammatory production [2,22], and is mainly delivered by macrophages and MSCs [2]. A correct balance between pro- and anti-inflammatory cytokines during fracture healing is necessary for successful fracture repair. High levels of TNF-α induce chronic inflammation and gradual destruction of cartilage and bone tissue [21], while the absence of TNF-α results in nonunion or delayed nonunions [17,23].

Bone fracture healing can be described in three characteristic phases: Inflammatory, repair, and remodelling (see Figure 1) [24]. During inflammation, necroses of cells result in the delivery of pro-inflammatory cytokines which attract inflammatory immune cells, such as neutrophils and monocytes [1,23], to the injury site. In response to their phagocytic activities, these cells magnify the pro-inflammatory production, leading to an acute inflammation [17,23,25]. Subsequently, monocytes differentiate into macrophages to down-regulate the inflammation and resolve it. Once this differentiation begins, the influx of the inflammatory cells ceases, and they die out [26].

During the resolution of inflammation, macrophages increase their population by migration and activate to their classical and alternative phenotypes, accordingly to the cytokine stimuli [16,27].

The two phenotypes can also shift between each other during this process [28,29]. Macrophages have the capability to release both pro- and anti-inflammatory cytokines through their different phenotypes [28]. Classically activated macrophages release high concentration of pro-inflammatory cytokines, including TNF-α, and low levels of anti-inflammatory cytokines [28], in response to their engulfing functions. Alternatively activated macrophages secrete high levels of Il-10 and low levels of TNF-α, as they continue with the clearance of debris and the modulation of inflammation [28].

During the repair phase, migrating MSCs contribute with the delivery of IL-10, and proliferate or differentiate into fibroblasts, chondrocytes, and osteoblasts [2,30]. Fibroblasts and chondrocytes proliferate and release the fibrinous/cartilagenous extracellular matrix, which fills up the fracture gap [10,30,31] while osteoblasts proliferate and deposit the new bone, also called woven bone [30]. Bone deposit results from mineralized collagen and other proteins being delivered by the osteoblasts [10]. After bone mineralization, osteoblasts remain on the bone surface or differentiate into osteocytes, whi
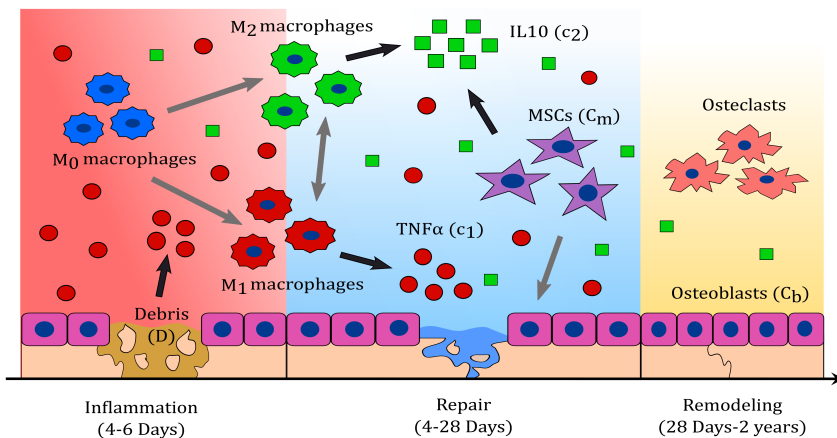

Figure 1. Inflammatory, repair, and remodeling phases of the bone fracture healing process. During the inflammatory phase, debris (D) activates the healing process by attracting macrophages M_0 to the injury site, which subsequently activate into their M_1 or M_2 phenotypes. Activated macrophages remove debris and secrete pro- and anti-inflammatory cytokines, such as tumor necrotic factor-α (TNF-α) (c_1) and interleukin-10 (IL-10) (c_2), which regulate the inflammation and the cellular functions. During the repair phase, migrating mesenchymal stem cells (MSCs) up-regulate IL-10 production, proliferate, and differentiate into osteoblasts (C_b). Mesenchymal and osteoblast cells synthesize the fibro/cartilage and woven bone, which closes the fracture gap. During the bone remodeling phase, osteoblasts and osteoclasts constantly remove and deposit new bone until the fracture is fully repaired.

During the last phase of the bone fracture healing process, the fibrocartilage and the woven bone are constantly removed and replaced by a functional bone [34]. This process is referred to as bone remodeling, and consists of systematic tissue degradation and production by osteoclasts and osteoblasts, respectively. Bone remodeling is a slow process, that can take months to years until the bone recovers to its pre-injury state [14]. In a moderate fracture, acute inflammation is observed 24 h after the injury; it also corresponds to the peak of TNF-α, which returns to baseline levels within 72 h [14,23]. Fibrinous/cartilaginous tissue production is observed in the first 3 days, peaks in about 10 to 12 days, and its removal starts as early as 21 days [30]. The inflammation is considered resolved when the debris is eliminated, activated macrophages emigrate to the lymphatic nodes to die, and inactivated macrophages return to their normal density [26]. These events are observed after two weeks from the beginning of the healing process [34,35]. At approximately 28 to 35 days, osteoclasts populate the injury site and a substantial removal of the fibrocartilage is observed [34]. The fracture healing

outcome is considered a delayed union if the fibrous/cartilaginous tissue is not removed completely in about 3 to 4 months after the injury, while it is considered a nonunion if no functional bone is obtained in 6 months after the trauma [36].

3. Modeling Assumptions

The most important effects of macrophages on bone fracture healing are observed during the inflammatory and repair phases of the healing process [11,16]. During the inflammatory phase, macrophages modulate and resolve the inflammation while, during the repair phase, macrophages provide an optimal environment for the cellular proliferation, differentiation, and tissue production. The primary variables during the inflammatory and repair phases of the bone fracture healing process are debris (D), unactivated macrophages (M_0), classical macrophages (M_1), alternative macrophages (M_2), MSCs (C_m), osteoblasts (C_b), pro-inflammatory cytokines (c_1), anti-inflammatory cytokines (c_2), fibrocartilage (m_c), and woven bone (m_b).

The biological system interactions are depicted in Figure 2. The cells and cellular dynamics are represented by the circular shapes and solid arrows. The molecular concentrations and their production/decay are represented by the octagonal shapes and dashed arrows. The pro- and anti-inflammatory cytokine activation/inhibition effects on the cellular functions are represented by the dotted arrows. Removal of debris and the negative effect among the variables are represented by the dot-ending dotted arrows.
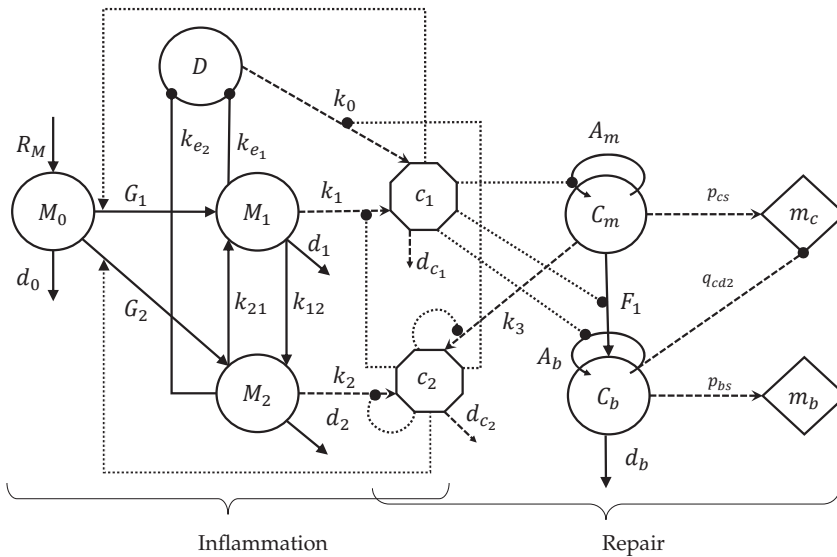
Figure 2. Flow diagram of the cellular and molecular dynamics during the inflammatory and repair phases of the bone fracture healing process.

It is assumed that the cellular functions are regulated by c_1 (such as TNF-α) and c_2 (such as IL-10). It is also assumed that c_1 is delivered through cell necrosis and by the classically activated macrophages, while c_2 is delivered by the alternatively activated macrophages and MSCs. It is further assumed that the repair process is governed by the production of m_c and m_b [30,37], whose final levels are used to classify the outcome of the bone healing process. Additionally, it is assumed that the debris D are proportional to the number of necrotic cells [11]. It is also assumed that unactivated macrophages M_0 do not release cytokines and do not engulf debris. Additionally, the population of M_0 increases proportionally in size to the density of debris, up to a maximal value of M_{max} [27]. The only

source of activated macrophages, M_1 and M_2, is M_0. Even though both phenotypes of activated macrophages have the ability to release both pro- and anti-inflammatory cytokines, it is assumed that only M_1 deliver c_1 and M_2 deliver c_2, as those are the major cytokines for each phenotype [38]. M_0 activate to M_1 under the c_1 stimulus, while they activate to M_2 under the c_2 stimulus. M_1 and M_2 macrophages do not de-differentiate back to the M_0 macrophages [39], and are able to switch phenotypes at a constant rate [29]. The accumulation of macrophages at the injury site is modeled by its recruitment due to inflammation, which is assumed to be proportional to the debris density.

Furthermore, it is assumed that the differentiation rates of MSCs into osteoblasts and osteoblasts into osteocytes are constant. MSCs synthesize the fibrocartilage, while osteoblasts synthesize the woven bone. It is also assumed that only the fibrocartilage is constantly removed by the osteoclasts, with the density of the osteoclasts being assumed proportional to the density of the osteoblasts [30]. In addition, it is assumed that the populations of the two tissue cells, C_m and C_b, experience logistic growth, where the growth rates decrease linearly as the population sizes approach a maximum value, K_{lm} and K_{lb}, respectively, imposed by the limited resources of the environment [30,40]. It is also assumed that there is no recruitment of MSCs and osteoblasts.

4. Model Formulation

The inflammatory and repair phases of the bone fracture healing process are modeled with a mass-action system of nonlinear ordinary differential equations. All variables represent homogeneous quantities in a given volume. Following the outlined biological assumptions and the flow diagram given in Figure 2 yields the resulting system of equations:

$$\frac{dD}{dt} = -R_D(k_{e_1}M_1 + k_{e_2}M_2) \tag{1}$$

$$\frac{dM_0}{dt} = R_M - G_1 M_0 - G_2 M_0 - d_0 M_0 \tag{2}$$

$$\frac{dM_1}{dt} = G_1 M_0 + k_{21} M_2 - k_{12} M_1 - d_1 M_1 \tag{3}$$

$$\frac{dM_2}{dt} = G_2 M_0 + k_{12} M_1 - k_{21} M_2 - d_2 M_2 \tag{4}$$

$$\frac{dc_1}{dt} = H_1(k_0 D + k_1 M_1) - d_{c_1} c_1 \tag{5}$$

$$\frac{dc_2}{dt} = H_2(k_2 M_2 + k_3 C_m) - d_{c_2} c_2 \tag{6}$$

$$\frac{dC_m}{dt} = A_m C_m \left(1 - \frac{C_m}{K_{lm}}\right) - F_1 C_m \tag{7}$$

$$\frac{dC_b}{dt} = A_b C_b \left(1 - \frac{C_b}{K_{lb}}\right) + F_1 C_m - d_b C_b \tag{8}$$

$$\frac{dm_c}{dt} = (p_{cs} - q_{cd1} m_c) C_m - q_{cd2} m_c C_b \tag{9}$$

$$\frac{dm_b}{dt} = (p_{bs} - q_{bd} m_b) C_b \tag{10}$$

Equation (1) describes the rate of change with respect to time of the debris density, which decreases proportionally to M_1 and M_2. The engulfing rate R_D is modeled by a Hill Type II function to represent the saturation of the phagocyte rate of macrophages [38,41]:

$$R_D = \frac{D}{a_{ed} + D}.$$

Equation (2) describes the rate of change with respect to time of the undifferentiated macrophages density. It increases because of migration and decreases by differentiating into M_1 and M_2 or by a

constant emigration rate. It is assumed that M_0 migrate to the injury site proportionally to D, up to a maximal constant rate, k_{max}, [22,28]:

$$R_M = k_{max}\left(1 - \frac{M}{M_{max}}\right)D,$$

where $M = M_0 + M_1 + M_2$. The differentiation rates of M_0 into M_1 and M_2 are stimulated by the cytokines accordingly to Hill Type II equations, respectively [29]:

$$G_1 = k_{01} \times \frac{c_1}{a_{01} + c_1}, \quad G_2 = k_{02} \times \frac{c_2}{a_{02} + c_2}.$$

Equation (3) describes the rate of change with respect to time of M_1, which increases when M_0 activates to M_1, and M_2 shifts phenotype; and decreases by emigration, and when M_1 shift phenotype. Similarly, Equation (4) describes the rate of change with respect to time of M_2. Equations (5) and (6) describes the rate of change with respect to time of c_1 and c_2. Here, k_0, k_1, k_2, and k_3 are the constant rates of the cytokine productions and d_{c_1} and d_{c_2} are the cytokine constant decay rates. The inhibitory effects of the anti-inflammatory cytokines are modeled by the following functions [29]:

$$H_1 = \frac{a_{12}}{a_{12} + c_2}, \quad H_2 = \frac{a_{22}}{a_{22} + c_2}.$$

Equation (7) describes the rate of change with respect to time of C_m, which increases by cellular division up to a constant-maximal carrying capacity, K_{lm}, and decreases by differentiation [30]. The total MSC proliferation rate is modeled by [42]:

$$A_m = k_{pm} \times \frac{a_{pm}^2 + a_{pm_1}c_1}{a_{pm}^2 + c_1^2},$$

where in the absent of inflammation, $c_1 = 0$, MSC proliferate at a constant rate k_{pm}. However, when there is inflammation, $c_1 > 0$, and the proliferation rate of MSCs increases or decreases according to the concentration of c_1 (i.e., high concentration levels of c_1 inhibit C_m proliferation, while low concentration levels of c_1 accelerate C_m proliferation). The differentiation rate of C_m is inhibited by c_1, which is modeled by the following function [11]:

$$F_1 = d_m \times \frac{a_{mb_1}}{a_{mb_1} + c_1}.$$

Equation (8) describes the rate of change with respect to time of C_b. It increases when MSC differentiate into osteoblasts, or when osteoblasts proliferate [30]. It decreases at a constant rate d_b when osteoblasts differentiate into osteocytes. The osteoblast proliferation rate is inhibited by c_1, which is modeled by the following function [11]:

$$A_b = k_{pb} \times \frac{a_{pb}}{a_{pb} + c_1}.$$

Equations (9) and (10) describe the rate of change with respect to time of the fibrocartilage and woven bone, where p_{cs} and p_{bs} are the tissue constant synthesis rates and q_{cd1}, q_{cd2}, and q_{bd} are the tissue degradation rates, respectively [30].

5. Analysis of the Model

The analysis of Model (1)–(10) is done by finding the equilibria and their corresponding stability properties. An equilibrium is a state of the system where the variables do not change over time [40]. Once the equilibria are identified, it is important to determine the behavior of the model near the equilibria by analyzing their local stability properties. An equilibrium is locally stable if the system moves toward it when it is near the equilibrium, otherwise it is unstable [40]. Therefore, the equilibria provide the possible outcomes of the bone fracture healing process, and their corresponding stability properties define the conditions under which a particular healing result occurs.

System (1)–(10) has the following three biologically meaningful equilibria of the vector form $E = (D, M_0, M_1, M_2, c_1, c_2, C_m, C_b, m_c, m_b)$: $E_0 = (0,0,0,0,0,0,0,0, m_{c_0}^*, m_{b_0}^*)$, $E_1 = (0,0,0,0,0,0,0, K_{lb}(1 - d_b/k_{pb}), 0, p_{bs}/q_{bd})$, $E_2 = (0,0,0,0,0, c_2^*, C_m^*, C_b^*, m_c^*, p_{bs}/q_{bd})$, where $C_m^* = K_{lm}(1 - d_m/k_{pm})$, $C_b^* = K_{lb}(k_{pb} - d_b + \sqrt{(k_{pb} - d_b)^2 + 4k_{pb}d_m C_m^*/K_{lb}})/2k_{pb}$, $c_2^* = a_{22}(-1 + \sqrt{1 + 4k_3 C_m^*/a_{22}d_{c_2}})/2$, and $m_c^* = p_{cs}C_m^*/(q_{cd1}C_m^* + q_{cd2}C_b^*)$. The existence conditions for the three equilibria are summarized in Table 1 and their stability conditions are summarized in Table 2, and are proved in Appendix A.

The existence conditions, listed in Table 1, arise from the fact that all biologically meaningful variables are nonnegative. Therefore, the existence condition for E_0 requires the steady state tissue densities to be either zero or any positive number. For E_1, the existence condition arises from the requirement that the steady state density of C_b must be greater than zero, which implies that the proliferation rate of osteoblasts must be greater than their differentiation rate (i.e., $k_{pb} > d_b$).

Similarly for E_2, the existence condition arises from the requirement that the steady state density for C_m must be greater than zero, which implies that the proliferation rate of MSCs must be greater than their differentiation rate (i.e., $k_{pm} > d_m$).

Table 1. Existence conditions for the equilibrium points and their biological meaning.

Equilibrium Points	Existence Conditions	Meaning
$E_0 = (0,0,0,0,0,0,0,0, m_{c_0}^*, m_{b_0}^*)$	$m_{c_0}^* \geq 0, m_{b_0}^* \geq 0$	nonunion
$E_1 = (0,0,0,0,0,0,0, K_{lb}(1 - d_b/k_{pb}), 0, p_{bs}/q_{bd})$	$k_{pb} > d_b$	successful healing
$E_2 = (0,0,0,0,0, c_2^*, C_m^*, C_b^*, m_c^*, p_{bs}/q_{bd})$	$k_{pm} > d_m$	nonunion or delayed union

The stability conditions of each biologically feasible equilibrium are listed in Table 2, and is determined from the eigenvalues of its associated Jacobian matrix (see Appendix A), as follows:

E_0 is stable when $k_{pm} \leq d_m$ and $k_{pb} \leq d_b$ (see Theorem A1), which implies that the differentiation rates of the MSC and osteoblasts are greater than or equal to their proliferation rates, respectively. The steady-state E_0 represents a nonunion. In this case, the inflammation is resolved, since the first five entries of E_0 are zero; however, the repair process has failed since the osteoblasts and osteoclasts have died out before the beginning of the remodeling process. Hence, the tissue densities, $m_{c_0}^*$ and $m_{b_0}^*$, can be any two positive values smaller than their maximal densities, p_{cs}/q_{cd1} and p_{bs}/q_{bd}, respectively (see Theorem A1).

E_1 is stable when $k_{pm} \leq d_m$ and $k_{pb} > d_b$ (see Theorem A2). The steady-state E_1 represents a successful repair of the bone fracture, where the inflammation is resolved, the fibrocartilage is completely removed from the repair site, and the woven bone has achieved its maximal density. In this case, osteoblasts proliferate faster than they differentiate, while MSC have the opposite behavior.

E_2 is stable when $k_{pm} > d_m$ (see Theorem A3). The steady-state E_2 represents a nonunion or delayed union, where the inflammation is resolved, but the osteoclasts have failed to degrade the cartilage in a timely fashion.

Table 2. Stability conditions for the equilibrium points.

Equilibrium Points	Stability Conditions	Stability
E_0	$k_{pm} \leq d_m, k_{pb} \leq d_b$	E_0 belongs to an attracting local set
E_0, E_1	$k_{pm} \leq d_m, k_{pb} > d_b$	E_0 unstable; E_1 locally stable
E_0, E_2	$k_{pm} > d_m, k_{pb} \leq d_b$	E_0 unstable; E_2 locally stable
E_0, E_1, E_2	$k_{pm} > d_m, k_{pb} > d_b$	E_0 and E_1 unstable; E_2 locally stable

6. Numerical Results

The proposed new model (1)–(10) is used to study the importance of macrophages during the inflammatory and repair phases of the bone fracture healing process, which occur within the first 21 days after trauma [11,13]. It is also used to investigate the evolution of a broken bone under normal and pathological conditions. Table 3 summarizes the baseline parameter values and units for the numerical simulations. These values are estimated in a qualitative manner from data in other studies [11,27,30,38,39,43]. Some of those, from [11], were also rescaled to account for the different mathematical expressions of the proliferation and differentiation rates of the tissue cells. All parameter values are based on murine experiments, with healthy mice having a moderate fracture (a broken bone with a gap size less than 3 mm) [30,43]. However, the bone fracture healing process for humans involves the same cells, cytokines, and qualitative dynamics, differing only in the number of cells, concentrations, and the length of time it takes for a full recovery [24].

First, a set of numerical simulation results was presented to compare two mathematical models of the bone fracture healing process that incorporate macrophages: The model developed in [11], and the new model (1)–(10). Next, numerical simulations were performed to support the theoretical stability results (successful and nonunion equilibria) and to numerically monitor the healing progression of a moderate fracture in normal conditions. Another set of numerical simulations was performed to analyze the effects of different debris densities on bone fracture healing. Finally, a set of numerical simulation results was presented, to investigate the effects of different concentrations of anti-inflammatory cytokines and various cellular treatments on the fracture healing under numerous pathological conditions. All simulations were obtained by using the adaptive MATLAB solver ode23s, and were initiated with densities of debris, macrophages, and MSCs set to $D(0) = 5 \times 10^7$, $M_0(0) = 4000$, $C_m(0) = 1000$, respectively, and the pro-inflammatory cytokine concentration set to $c_1(0) = 1$.

6.1. Comparison of Existing Models

The model developed in [11] and the present mathematical model (1)–(10) were compared when $D(0) < a_{ed} = 4.71 \times 10^6$ (i.e., the initial debris concentration was below the half-saturation of debris). In this case, the macrophage digestion rate increased approximately linearly with respect to the debris population, as assumed in model [11]. The same parameter values were used in both models (Table 3), with $k_{e_1} = 11$, $k_{e_2} = 48$, $k_2 = 3.72 \times 10^{-6}$, $k_3 = 8 \times 10^{-6}$, and $q_{bd} = 5 \times 10^{-8}$.

Figure 3 shows the numerical evolutions of the tissues' production when $D(0) = 2 \times 10^6$. In all simulations, we refer to fibrocartilage and woven bone as cartilage and bone, respectively. The production of cartilage m_c and bone m_b given by the present model is much more realistic than the production given by the model developed [11], since, according to the experimental data, the cartilage production peaks to its maximal density of around 1 g/mL about 10–12 days after trauma, and significant bone tissue production is observed after the second week [44].

Table 3. Parameter descriptions and units.

| Parameter	Description	Range of Values	Reference
k_{e_1}	Engulfing debris rate of M_1	3–48/day	[38,41]
k_{e_2}	Engulfing debris rate of M_2	3–48/day	[38,41]
a_{ed}	Half-saturation of debris	4.71×10^6 cells/mL	[38]
k_{max}	Maximal migration rate	0.015–0.1/day	[39,45]
M_{max}	Maximal macrophages density	6×10^5–1×10^6 cells/mL	[27,41]
k_{01}	Activation rate of M_1	0.55–0.611/day	[29,39]
k_{02}	Activation rate of M_0 to M_2	0.0843–0.3/day	[29]
k_{12}	Transition rate from M_1 to M_2	0.083–0.075/day	[29,39]
k_{21}	Transition rate from M_2 to M_1	0.005–0.05/day	[29]
d_0	Emigration rate of M_0	0.156–0.02/day	[29,39]
d_1	Emigration rate of M_1	0.121–0.2/day	[29,38,39]
d_2	Emigration rate of M_2	0.163–0.2/day	[29,38,39]
k_0	Secretion rate of c_1 by debris	5×10^{-7}–8.5×10^{-6} ng/cells/day	[38]
k_1	Secretion rate of c_1 by M_1 macrophages	8.3×10^{-6} ng/cells/day	[38]
k_2	Secretion rate of c_2 by M_2 macrophages	3.72×10^{-6} ng/cells/day	[38]
k_3	Secretion rate of c_2 by MSCs	7×10^{-7}–8×10^{-6} ng/cells/day	[11]
d_{c_1}	Decay rate of c_1	12.79–55/day	[29,38]
d_{c_2}	Decay rate of c_2	2.5–4.632/day	[29,38]
a_{12}	Effectiveness of c_2 inhibition of c_1 synthesis	0.025 ng/mL	[29]
a_{22}	Effectiveness of c_2 inhibition of c_2 synthesis	0.1 ng/mL	[29]
a_{pm}	Effectiveness of c_1 inhibition of C_m proliferation	3.162 ng/mL	[11,46]
a_{mb_1}	Effectiveness of c_1 inhibition of C_m differentiation	0.1 ng/mL	[11,47]
a_{01}	Half-saturation of c_1 to activate M_1	0.01 ng/mL	[29]
a_{02}	Half-saturation of c_2 to activate M_2	0.005 ng/mL	[29]
a_{pb}	Effectiveness of c_1 inhibition of C_b proliferation	10 ng/mL	[11,48]
a_{pm_1}	Constant enhancement of c_1 to C_m proliferation	13 ng/mL	[11,46]
k_{pm}	Proliferation rate of C_m	0.5/day	[11]
d_m	Differentiation rate of C_m	1/day	[11,30]
k_{pb}	Proliferation rate of C_b	0.2202/day	[11,30]
d_b	Differentiation rate of C_b	0.15/day	[11,30]
p_{cs}	Fibrocartilage synthesis rate	3×10^{-6} g/cells/day	[11,30]
q_{cd_1}	Fibrocartilage degradation rate	3×10^{-6} mL/cells/day	[11,30]
q_{cd_2}	Fibrocartilage degradation rate by osteoclasts	0.2×10^{-6} mL/cells/day	[11,30]
p_{bs}	Bone tissue synthesis rate	5×10^{-8} g/cells/day	[11,30]
q_{bd}	Bone tissue degradation rate	5×10^{-8} mL/cells/day	[30]
K_{lb}	Carrying capacity of C_b	1×10^6 cells/mL	[11,30]
K_{lm}	Carrying capacity of C_m	1×10^6 cells/mL	[11,30]
$D(0)$	Density of necrotic cells	1×10^6–2×10^8 cells/mL	[27,38,41]
$C_m(0)$	Initial MSCs density	1000 cells/mL	[11]
$M_0(0)$	Unactivated macrophage density	4000 cell/mL	[45]
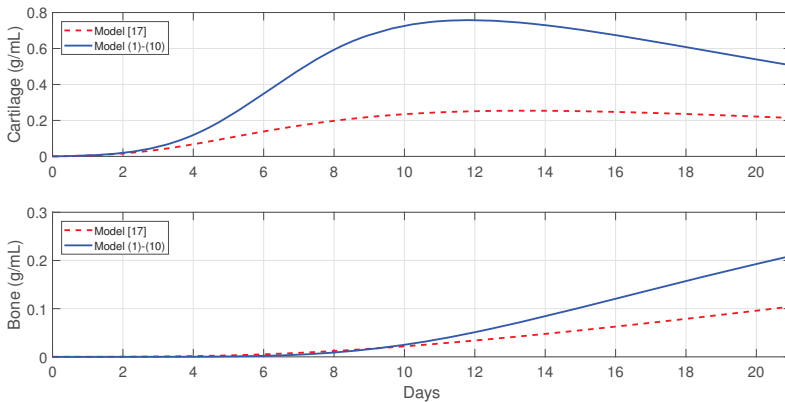
Figure 3. Comparison of tissues evolution in Model [11] and Model (1)–(10).

6.2. Different Outcomes of the Bone Fracture Healing Process

Next, a set of numerical simulations is presented to support the theoretical results. Accordingly to the qualitative analysis of the model there are three equilibria: E_0, E_1, and E_2, where their stability conditions are determined by the tissue cells' proliferation and differentiation rates, k_{pm}, k_{pb}, d_m and d_b, respectively. The following parameter values were used: $k_{pm} = 0.5$, $d_m = 1$, $k_{pb} = 0.2202$, and $d_b = 0.3$, to demonstrate the stability of E_0, since then $k_{pm} < d_m$ and $k_{pb} < d_b$. The stability of E_1 was demonstrated using the following parameter values: $d_m = 1$, $k_{pm} = 0.5$, $k_{pb} = 0.2202$, and $d_b = 0.15$, since then $k_{pm} \leq d_m$ and $k_{pb} > d_b$. Finally, the following parameter values were used: $k_{pm} = 0.5$ and $d_m = 0.1$, to demonstrate the stability of E_2, since then $k_{pm} > d_m$. Different time-periods are used in Figures 4–6 to better demonstrate the qualitative behavior of the system under different stability conditions.

Figure 4 shows the qualitative behavior of E_1 for the macrophages, debris, TNF-α, and IL-10 densities, with the inflammation being resolved in about 40 days. The top-left plot of Figure 4 shows the temporal evolution of M_0 (dashed lines), M_1 (dotted lines), and M_2 (solid lines). It can be observed that M_1 first peaks to its maximum value, which is then followed by M_2. Similar sequences of transitions of first M_1, and then M_2, are commonly observed in normal tissue healing conditions [2,39].
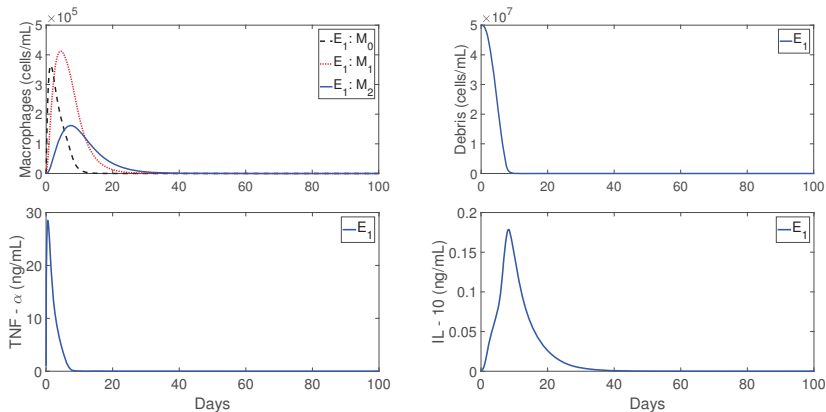

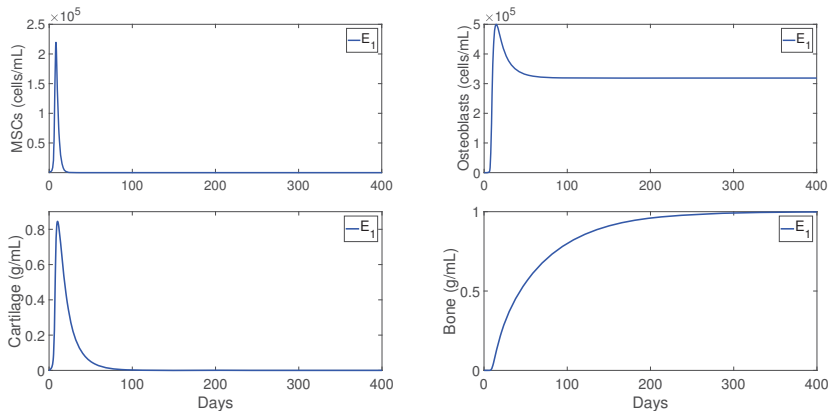**Figure 4.** Cellular and molecular evolution of the resolution of the inflammation in normal conditions.

Figure 5. Cellular and molecular evolution of the repair process in a successful fracture healing.

Figure 5 shows the qualitative behaviors of E_1 for the MSC, osteoblasts, cartilage, and bone densities. Here, the MSC density decayed to zero over time, while the osteoblasts maintained a constant density below their carrying capacity $K_{lb} = 1 \times 10^6$. In addition, the bottom plots of Figure 5 shows that the cartilage was eventually degraded by the osteoclasts and the bone achieved its maximum density of 1 ng/mL. Therefore, E_1 exhibits the temporal progression of a successful bone fracture healing.
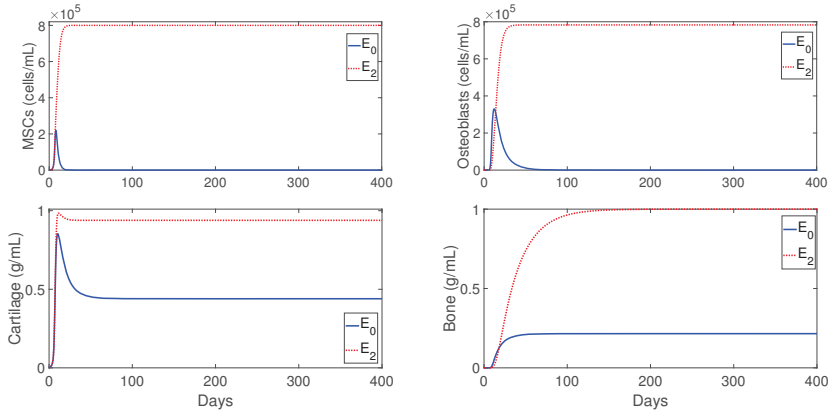

Figure 6. Cellular and molecular evolution of the repair process in a nonunion fracture healing.

Figure 6 shows the qualitative evolution for the MSC, osteoblasts, cartilage, and bone densities for E_0 (solid lines) and E_2 (dotted lines). Since the temporal evolution of macrophages, debris, and cytokines densities in E_0 and E_2 are similar to those for E_1 showed in Figure 4, they are omitted here. It can be observed, in Figure 6, that the two cellular densities in E_0, MSC, and osteoblasts, decayed to zero over time, with the osteoclasts failing to degrade the cartilage; this results in nonunion. Mathematically, this case occurs when osteoblasts proliferate at a rate lower than their differentiation rate, i.e., $k_{pb} < d_b$. In practice, this scenario is commonly observed in advanced-age patients whose MSC and osteoblast cells decrease their capability to proliferate and differentiate [3]. On the other hand, the two cells and the two tissues in E_2 remain at positive constant values (Figure 6), but the final fracture healing outcome is still a nonunion. Here, the osteoclasts again fail to degrade the cartilage [3], even though the bone has achieved its maximum density of 1 ng/mL. Therefore, in this case, migration of osteoclasts must be enhanced through surgical interventions, in order to achieve a successful bone repair [30].

6.3. Importance of Macrophages during the Bone Fracture Healing Process

In this section, the mathematical model is used to investigate the effects of macrophages during the inflammatory and repair phases of the bone fracture healing process. The major contribution of macrophages to fracture healing is through the delivery of pro- and anti-inflammatory cytokines at the repair site. Therefore, the values of the parameters k_i, representing the secretion rates of c_i by M_i, $i = 1, 2$, are varied in the numerical simulations, as compared to their base values from Table 3.

Figure 7 shows that macrophages have a drastic effect on the short-term tissue dynamics during the healing process. In the presence of M_1 and M_2, fibrocartilage formation more than doubles in about 1 week, while woven bone experiences an additional steady increase during the same period and beyond. The simulations presented in Figure 8 demonstrate the individual effects of the different phenotypes of macrophages, and show that the alternatively activated macrophages M_2 have a more dominant contribution to the tissue production, as compared to the classically activated macrophages M_1.
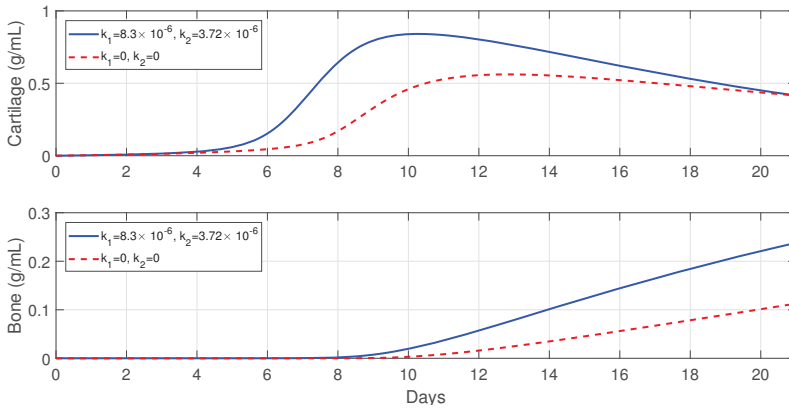
Figure 7. Tissue evolution when macrophages contribute to the healing process (solid line), $k_1, k_2 \neq 0$, and when they do not contribute to the healing process (dashed line), $k_1 = k_2 = 0$.
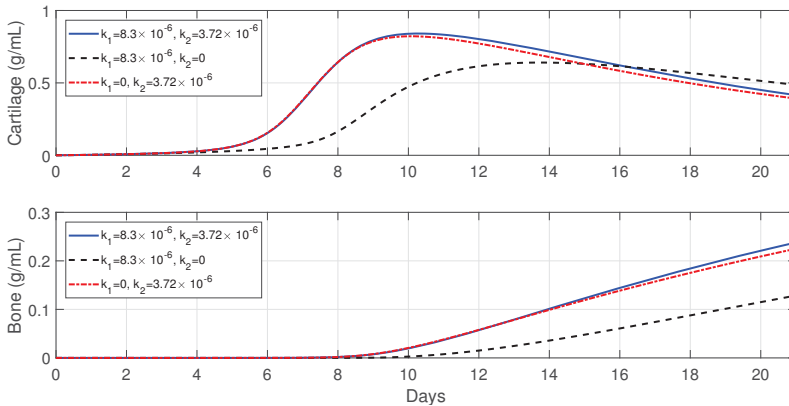

Figure 8. Tissue evolution when the alternatively activated macrophages, M_2, do not contribute to the healing process (dashed line), $k_2 = 0$, and when the classically activated macrophages, M_1, do not contribute to the healing process (dotted line), $k_1 = 0$.

6.4. Evolution of the Healing Process for Different Types of Fractures

In this section, the model is used to monitor the evolution of a successful repair (Table 3) for different types of fractures. In healthy individuals, simple, moderate, and severe fractures are correlated with the debris densities [49,50]. Therefore, the initial debris concentration is set to $D(0) = 3 \times 10^5$, $D(0) = 5 \times 10^7$, and $D(0) = 2 \times 10^8$, for a simple, moderate, and severe fracture, respectively.

Figure 9 shows that the tissue production is a slow process for a simple fracture, since the cartilage and bone densities are less than the corresponding tissue densities for moderate and severe fractures. A slow healing process is commonly observed in micro-crack healing [49]. Furthermore, there is less cartilage formation over time in simple fractures [50]. For a moderate fracture, the maximal production of cartilage is observed around 10 days, followed by a significant degradation, while the bone tissue production occurs after the first week. For a severe fracture, Figure 9 shows that there is a delay in the two tissues' production, compared with those given by moderate fractures, with the peak of the cartilage and bone production observed at around day 16.
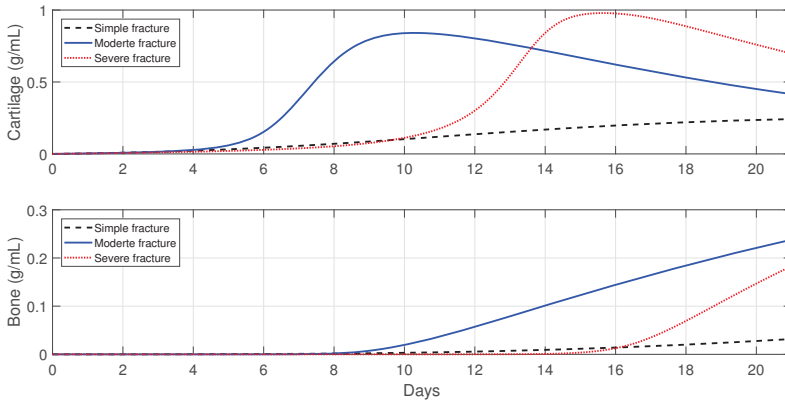
Figure 9. Tissue evolution of a successful repair for different types of fractures.

6.5. Immune-Modulation Therapeutic Treatments of Bone Fractures

The administration of anti-inflammatory drugs and the injection and/or transplantation of MSC and macrophages are two of the clinical trials that have been implemented in orthopedics to stimulate and accelerate bone fracture healing [2,23]. In this section, Model (1)–(10) is used to explore these possible therapeutic treatments to accelerate the healing of a broken bone, under normal and pathological conditions such as severe fractures, advanced age, and senile osteoporosis [3].

6.5.1. Administration of Anti-Inflammatory Drugs at the Beginning of the Healing Process

Treatments based on anti-inflammatory cytokines, such as the cytokine-specific agents that block the pro-inflammatory cytokines productions, have exhibited promising clinical results and have led to intense orthopedic research activities [1,2,17,23,33,51–53]. In this section, a set of numerical simulations is presented to investigate the effect of the administrations of anti-inflammatory cytokines at the beginning of the healing process in healthy individuals and also in immune-compromised patients. In each case of the numerical simulations, $c_2(0) = 0$, 10, and 100 ng/mL.

In healthy individuals, the administration of anti-inflammatory drugs is implemented for a simple fracture and also for two moderate fractures with different debris concentrations: $D(0) = 3 \times 10^5$, $D(0) = 2 \times 10^7$, and $D(0) = 5 \times 10^7$.

Figure 10 shows that the administration of c_2 in the simple fracture slows down both the cartilage and bone productions. Figures 11 and 12 show that the administration of c_2 in the moderate fractures improves the tissue evolution, but in a dose-dependent manner. On one hand, when $D(0) = 2 \times 10^7$, the administration of c_2 has either a positive or negative effect on the two tissue productions. The administration of 10 ng/mL of c_2 enhances the early production of cartilage and increases the bone synthesis, while the administration of 100 ng/mL of c_2 results in the opposite effect. On the other hand, when $D(0) = 5 \times 10^7$, the administration of c_2 enhances the earlier cartilage production and improves the synthesis of the bone for both concentrations, with 10 ng/mL being the optimal of the two doses.

Figure 10. Tissue evolution in a simple fracture under different initial anti-inflammatory cytokines concentrations, $D(0) = 3 \times 10^5$.
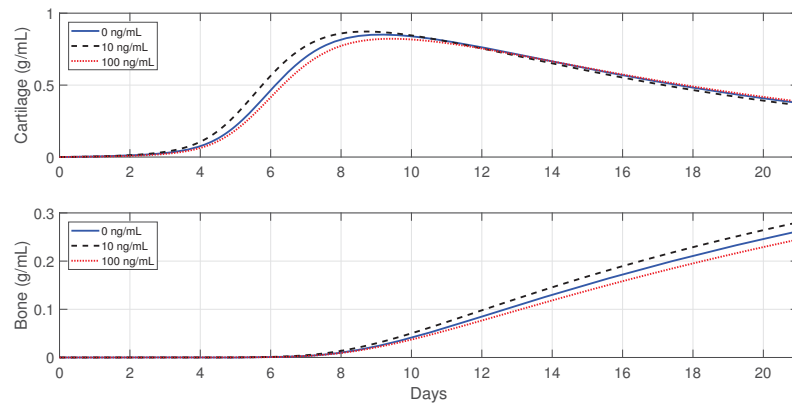
Figure 11. Tissue evolution in a moderate fracture under different initial anti-inflammatory cytokines concentrations, $D(0) = 2 \times 10^7$.

Next, the model is used to implement the administration of anti-inflammatory drugs under different pathological conditions. First, severe fractures in immune-compromised individuals are simulated by using the following parameter values: $D(0) = 2 \times 10^8$ and $k_{max} = 0.0015$, since, in the fractures of such individuals, there is an increase in the accumulation of debris [50] and a decrease in the macrophage migration rate [54]. Second, the following parameter values are used: $k_{e_1} = k_{e_2} = 3$ and $k_1 = 9 \times 10^{-6}$, to simulate bone fracture healing in aging individuals, since, in this case, the macrophage phagocytic rate decreases and there is an increase of pro-inflammatory cytokine synthesis by M_1 [3,25]. Finally, $c_1(0) = 100$, $k_{pm} = 0.2$, $d_m = 0.5$, $k_{pb} = 0.16$, and $d_b = 0.15$ are used to simulate the healing process for an senil osteoporotic fracture, since, in this case, a high level of pro-inflammatory cytokines is observed and the MSC and osteoblast functions decrease [3].
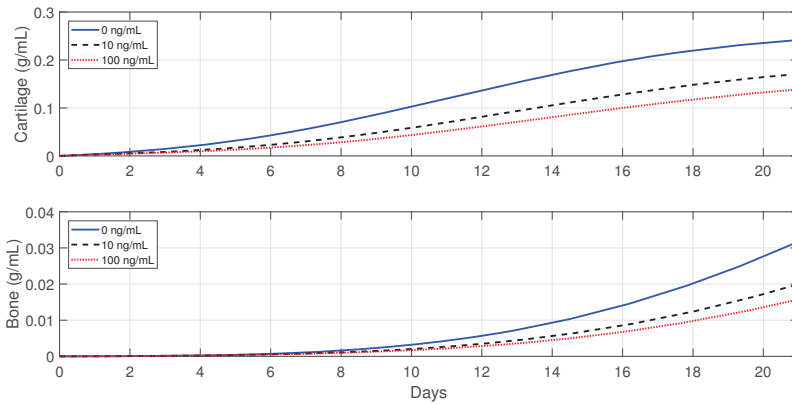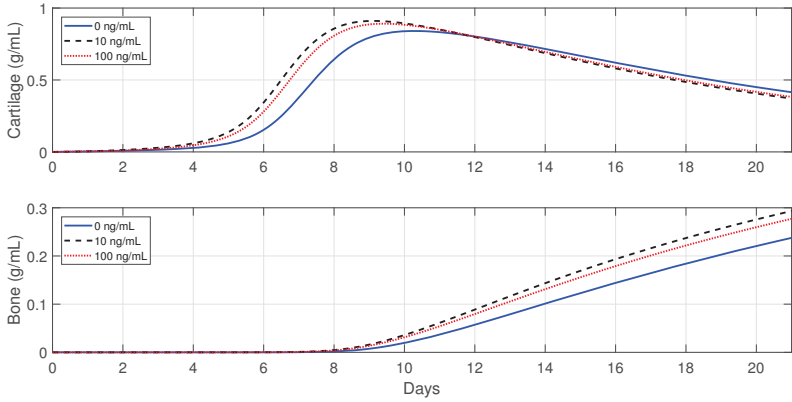
Figure 12. Tissue evolution in a moderate fracture under different initial anti-inflammatory cytokine concentrations, $D(0) = 5 \times 10^7$.

Figures 13–15 show that the administration of anti-inflammatory cytokines under the above three different pathological conditions always improve tissue productions, where the optimal dose of c_2, for both the advanced-age individuals and senile osteoporotic fractures, is 10 ng/mL.
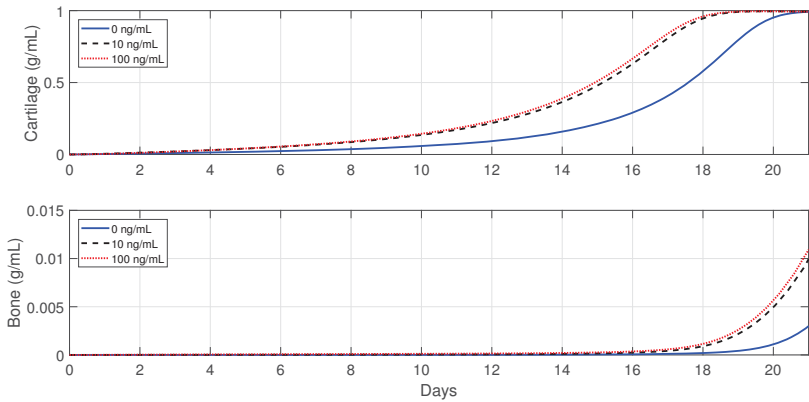
Figure 13. Tissue evolution in a severe fracture under different initial anti-inflammatory cytokines concentrations.
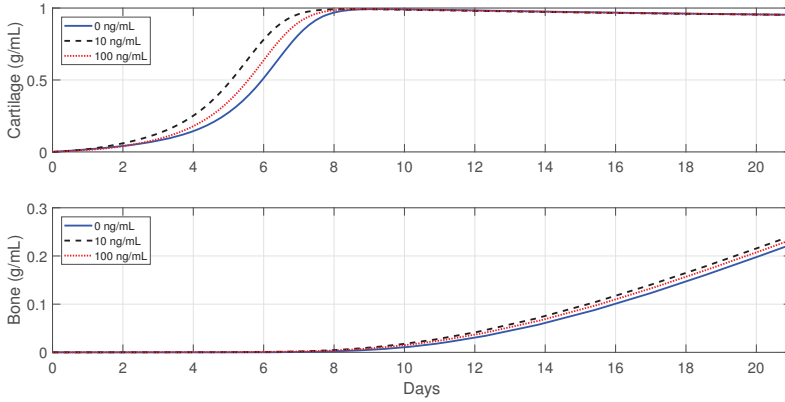
Figure 14. Tissue evolution in an advanced age fracture under different initial anti-inflammatory cytokines concentrations.
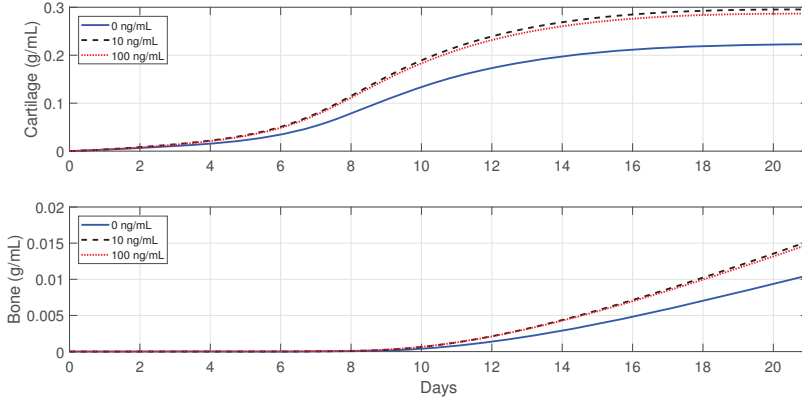
Figure 15. Tissue evolution in a senile osteoporotic fracture under different initial anti-inflammatory cytokines concentrations.

6.5.2. Cellular Therapeutic Interventions under Immune-Compromised Conditions

Additions of MSC to the injury site through injection and/or transplantation have been used in practice to stimulate and augment bone fracture healing [2]. Another cellular intervention is scaffold implants, where undifferentiated macrophages and MSCs are co-cultured together, and cytokines are slowly released to stimulate M_2 activation [3]. The parameter values used in the numerical simulations that explore these possible therapeutic treatments are the same as in Section 6.5.1.

For severe fractures with immune-compromised conditions, the use of scaffold implants is simulated through a fast M_2 activation (i.e., $k_{02} = 0.3$ and $k_{12} = 0.075$), and also an increase in the C_m and M_0 densities (i.e., $M_0(0) = 5000$ and $C_m(0) = 5000$). For fractures in aging individuals and individuals with senile osteoporotic fractures, the MSC injection and the fast M_2 activation are simulated by setting $C_m(0) = 5000$, $k_{02} = 0.3$, and $k_{12} = 0.075$.

Figures 16–18 show that the two cellular interventions increase both tissue productions. Furthermore, those interventions result in larger improvements in severe and senil osteoporotic fractures, when compared to fractures in aging individuals.
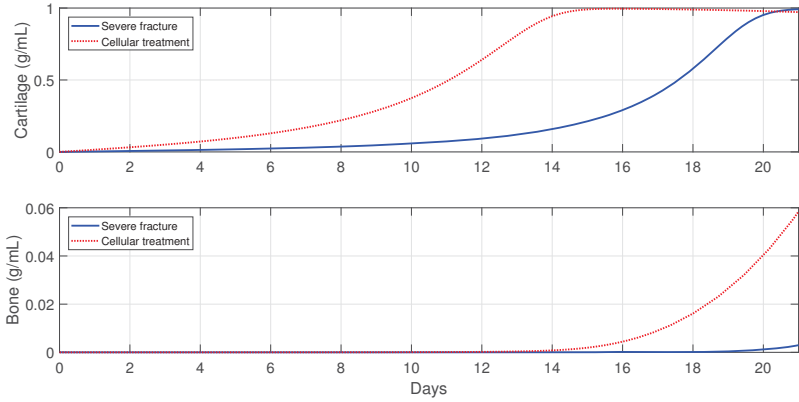
Figure 16. Tissue evolution in a severe fracture without therapeutic innervation (solid line) and with $M_0(0)$ and $C_m(0)$ transplantation (dotted line).
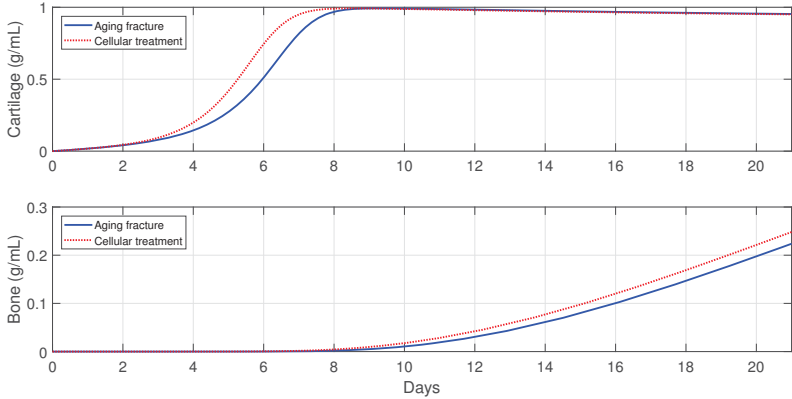
Figure 17. Tissue evolution in an aging fracture without therapeutic innervation (solid line) and with MSCs injection (dotted line).
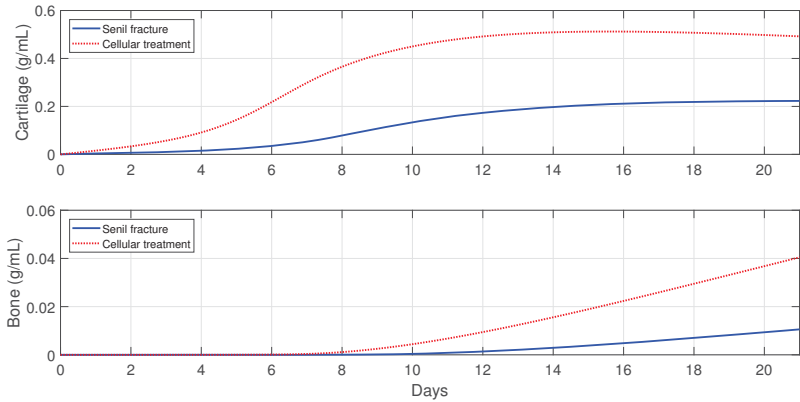
Figure 18. Tissue evolution in a senile osteoporotic fracture without therapeutic innervation (solid line) and with MSCs injection (dotted line).

7. Discussion and Conclusions

A new mathematical model was introduced to mathematically and numerically study the macrophage-mediated inflammation involved in the early stages of the bone fracture healing process: The inflammatory and repair phases. Classically and alternatively activated macrophages were incorporated in the model, to study their capabilities to modulate and resolve the inflammation. The model also included the macrophage abilities to regulate the tissue cellular functions through the delivery of pro- and anti-inflammatory cytokines. The resolution of the inflammation was assumed to be initiated with the activation of the macrophages into their classical phenotype. The classically activated macrophages deliver pro-inflammatory cytokines, such as TNF-α, as they engulf debris. Then, the alternatively activated macrophages and the MSC modulate the inflammation by releasing anti-inflammatory cytokines, such as IL-10. Finally, the classically activated macrophages remove the remaining debris. The model also incorporated different engulfing rates of activated macrophages, saturation rates of phagocytes, and the maximal density of macrophages at the injury site, thus allowing a better understanding of the interplay between macrophages and tissue cells during the bone fracture healing process.

The mathematical analysis revealed that there are three feasible fracture healing outcomes. Two of the outcomes represent a nonunion healing: One is the case when the cells deactivate or die out before the healing process finishes up, and the other is the case when the tissue cells remain constant but the osteoclasts fail to completely remove the cartilage. The third outcome represents a successful healing, where the osteoblasts and osteoclasts are constantly producing and removing the woven bone. The stability conditions of each outcome can be used to biologically explain why the fracture healing fails, as well as to design therapeutic interventions to stimulate or accelerate the healing process.

The new model was used to study the importance of macrophages during the early stages of tissue production. It revealed that macrophages significantly improve the tissue production, with alternatively activated macrophages having the main effect on the process. The flexibility of the model also allowed us to perform a variety of different types of numerical simulations quickly and cost effectively. It was used to monitor the progression of the healing of a broken bone and to predict its final outcome. In particular, the administration of anti-inflammatory drugs to improve the bone fracture healing process was numerically simulated. It was found that the administration of anti-inflammatory cytokines fails to accelerate the healing process in simple fractures, while it accelerates the healing process in moderate fractures, depending on the cytokine concentrations, and always improves the healing process in severe fractures. Such results have been also clinically observed, when corticosteroids and nonsteroidal anti-inflammatory drugs (NSAIDs) are administered in bone fractures [23]. Therefore, based on the model findings, the concentration of debris must be carefully considered when administering anti-inflammatory drugs to enhance the fracture healing process [50]. The model was also used to explore other potential cellular therapeutic approaches, such as MSC injection and transplantation. It was found that such treatments can also improve the healing time of a broken bone, especially in immune-compromised patients. The model can also be easily adapted to other therapeutic approaches, such as the administration of different anti-inflammatory drugs, suggesting a variety of possible ways to guide clinical experiments and bone tissue engineering strategies.

Future research directions include modifications of the model by incorporating additional molecular and cellular interactions and processes during the inflammatory and repair phases of fracture healing, such as macrophage migration and MSC differentiation due to cytokine stimulus. Another research direction is the incorporation of the bone remodeling phase of the healing process, which begins at the end of the repair phase and continues long after fracture union. There are different factors that affect the bone remodeling, including other bone cells, such as osteoclasts, osteocytes, progenitor cells, and other sources of cytokines [30,32,43]. This presents a challenging new research direction in the pursuit to better understand the bone fracture healing process and the development of new treatment strategies.

Appendix A

The stability conditions of the equilibria of Model (1)–(10) are stated and proved below. The analysis is conducted using the Jacobian of the system at each equilibrium point and finding its corresponding eigenvalues [40,55].

Theorem A1. *The* $E_0 = \{(0,0,0,0,0,0,0,0,m^*_{c_0},m^*_{b_0})$ *belongs to the set* $B = \{(0,0,0,0,0,0,0,0,m_c,m_b):$ $0 \le m_c \le p_{cs}/q_{cd_1}, 0 \le m_b \le p_{bs}/q_{bd}\}$*, which is a local attractor set of the solution set given by System (1)–(10) if and only if* $k_{pm} \le d_m$ *and* $k_{pb} \le b_d$*.*

Proof of Theorem A1. The right-hand side functions of System (1)–(10) are continuous and bounded, since all model variables and parameters are positive. Hence, for each initial condition of the system, there is a unique solution [55]. Then, as zero is a solution of the System (1)–(10), and by uniqueness of solution, all the solutions of the system with positive initial condition are positive [55].

Next, it will be proved that the hyperplane $A = \{(0,0,0,0,0,0,0,0,m_c,m_b): m_c \ge 0, m_b \ge 0\}$ is an attractor set of the solutions of the system (1)–(10). There are two cases to consider, based on the relation between the cell proliferation and differentiation rates.

First, let us examine the case when $k_{pm} < d_m$ and $k_{pm} < d_b$. The Jacobian matrix $J(E_0)$ is given by the following lower triangular block matrix

$$J(E_0) = \begin{pmatrix} J_1(E_0) & \mathbf{0} & \mathbf{0} \\ * & J_2(E_0) & \mathbf{0} \\ \mathbf{0} & * & J_3(E_0) \end{pmatrix},$$

where

$$J_1(E_0) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ k_{max} & -d_0 & 0 & 0 \\ 0 & 0 & J_{11} & 0 \\ k_0 & 0 & * & -d_{c_1} \end{pmatrix}, \quad J_2(E_0) = \begin{pmatrix} -d_{c_2} & k_3 & 0 \\ 0 & -d_m + k_{pm} & 0 \\ 0 & d_m & -d_b + k_{pb} \end{pmatrix}$$

$$J_{11} = \begin{pmatrix} -d_1 - k_{12} & k_{21} \\ k_{12} & -d_2 - k_{21} \end{pmatrix}, \quad J_3(E_0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Therefore, the corresponding characteristic polynomial associated with $J(E_1)$ is given by the product of the characteristic polynomials associated with each submatrix [56]:

$$p(\lambda) = \lambda^3 (\lambda + d_0) (\lambda + d_{c_1}) (\lambda + d_{c_2}) (\lambda + d_m - k_{pm}) (\lambda + d_b - k_{pm}) (\lambda^2 + a\lambda + b),$$

where $a = d_1 + d_2 + k_{12} + k_{21}$ and $b = k_{12}d_2 + k_{21}d_1 + d_1d_2$. The polynomial factor of order two of $p(\lambda)$ has the following two roots: $(-a \pm \sqrt{a^2 - 4b})/2$, which are negative since $a^2 - 4b = (d_1 - d_2 + k_{12} - k_{21})^2 + 4k_{12}k_{21} > 0$ and $b > 0$. Therefore, the eigenvalues of $J(E_0)$ are negative for the variables M_0, M_1, M_2, c_1, c_2, C_m, and C_b and are equal to zero for D, m_c, and m_b. Since $D'(t) \le 0$ for all the variables in the system (1)–(10) and $(D^*, 0, 0, 0, 0, 0, 0, 0, m_c, m_b)$ with $D^* \ne 0$ is not an equilibrium point, then the solutions of the system (1)–(8) are attracted to the set $A = \{(0,0,0,0,0,0,0,0,m_c,m_b): m_c \ge 0, m_b \ge 0\}$.

Equations (9) and (10) imply that $m_c' \leq 0$ and $m_b' \leq 0$ for all $m_c > p_{cs}/q_{cd_1}$ and $m_b > p_{bs}/q_{bd}$. Therefore, the set B is a local attractor set of A [55].

Next, let us consider the case when $k_{pm} = d_m$ and $d_b = k_{pb}$. Here, the eigenvalues of $J(E_0)$ are the same as above except those associated with C_m and C_b, which are equal to zero. Therefore, in this case, by considering the second order approximations of the right hand sides of Equations (7) and (8), instead of just the first order approximations, and using similar arguments as above, proves that the set B is a local attractor set of A. □

Theorem A2. *The equilibrium* $E_1 = (0, 0, 0, 0, 0, 0, 0, K_{lb}(1 - d_b/k_{pb}), 0, p_{bs}/q_{bd})$ *is locally stable if and only if* $d_m \geq k_{pm}$ *and* $k_{pb} > d_b$.

Proof of Theorem A2. The Jacobian matrix corresponding to the point E_1 is given by the following lower triangular block matrix

$$J(E_1) = \begin{pmatrix} J_1(E_1) & \mathbf{0} & \mathbf{0} \\ * & J_2(E_1) & \mathbf{0} \\ \mathbf{0} & * & J_3(E_1) \end{pmatrix},$$

where $J_1(E_1)$ has the same expression as $J_1(E_0)$, defined in Theorem A1, and

$$J_2(E_1) = \begin{pmatrix} -d_{c_2} & k_3 & 0 \\ 0 & -d_m + k_{pm} & 0 \\ 0 & d_m & d_b - k_{pb} \end{pmatrix}, \quad J_3(E_1) = \begin{pmatrix} -q_{cd_2}K_{lb}(1 - \frac{d_b}{k_{pb}}) & 0 \\ 0 & -q_{bd}K_{lb}(1 - \frac{d_b}{k_{pb}}) \end{pmatrix}.$$

Since $d_m - k_{pm} \geq 0$ and $k_{pb} > d_b$, and all of the eigenvalues of $J_1(E_0)$ are non-positive values, then the eigenvalues of $J(E_1)$ are negative, except the eigenvalues associated with D and C_m when $k_{pm} = d_m$, which are equal to zero. Therefore, E_1 is a locally stable node, since $D' \leq 0$ for all the variables of the system (1)–(10) and $C_m' \leq 0$ when $k_{pm} = d_m$. □

Theorem A3. *The equilibrium* $E_2 = (0, 0, 0, 0, 0, c_2^*, C_m^*, C_b^*, m_c^*, p_{bs}/q_{bd})$ *is locally stable if and only if* $k_{pm} > d_m$, *where* $C_m^* = K_{lm}(1 - d_m/k_{pm})$, $C_b^* = K_{lb}(k_{pb} - d_b + \sqrt{(k_{pb} - d_b)^2 + 4k_{pb}d_mC_m^*/K_{lb}})/2k_{pb}$, $c_2^* = a_{22}(-1 + \sqrt{1 + 4k_3C_m^*/a_{22}d_{c_2}})/2$, *and* $m_c^* = p_{cs}C_m^*/(q_{cd1}C_m^* + q_{cd2}C_b^*)$.

Proof of Theorem A3. The Jacobian matrix corresponding to the point E_2 is given by the following lower triangular block matrix

$$J(E_2) = \begin{pmatrix} J_1(E_2) & \mathbf{0} & \mathbf{0} \\ * & J_2(E_2) & \mathbf{0} \\ \mathbf{0} & * & J_3(E_2) \end{pmatrix},$$

where

$$J_1(E_2) = \begin{pmatrix} 0 & 0 & \mathbf{0} & 0 \\ k_{max} & -d_0 - G_2^* & \mathbf{0} & 0 \\ 0 & * & J_{11} & 0 \\ k_0 H_1^* & 0 & * & -d_{c_1} \end{pmatrix}, \quad J_3(E_2) = \begin{pmatrix} -q_{cd_1}C_m^* - q_{cd_2}C_b^* & 0 \\ 0 & -q_{bd}C_b^* \end{pmatrix},$$

$$J_2(E_2) = \begin{pmatrix} -d_{c_2}\left(1 + \frac{c_2^*}{a_{22}+c_2^*}\right) & k_3 H_2^* & 0 \\ 0 & d_m - k_{mb} & 0 \\ 0 & d_m & -\sqrt{(d_b - k_{pb})^2 + 4\frac{k_{bp}d_mC_m^*}{K_{lb}}} \end{pmatrix},$$

$G_2^* = \frac{c_2^* k_{02}}{a_{02} + c_2^*}$, $H_1^* = \frac{a_{12}}{a_{12} + c_2^*}$, $H_2^* = \frac{a_{22}}{a_{22} + c_2^*}$ and J_{11} is defined as in Theorem A1. Since all of the eigenvalues of J_{11} are negative (Theorem A1) and $k_{pm} > d_m$, and all equilibrium variables and parameter values are positive, then all the eigenvalues of $J_1(E_2)$, $J_2(E_2)$, $J_3(E_2)$ are negative, except for the eigenvalue associated to D which is equal to zero. Therefore, since $D' \leq 0$ for all the variable system, then E_2 is locally stable. □

References

1. Schmidt-Bleek, K.; Kwee, B.J.; Mooney, D.J.; Duda, G.N. Boon and bane of inflammation in bone tissue regeneration and its link with angiogenesis. *Tissue Eng. Part B Rev.* **2015**, *21*, 354–364. [CrossRef] [PubMed]
2. Kovach, T.K.; Dighe, A.S.; Lobo, P.I.; Cui, Q. Interactions between MSCs and immune cells: Implications for bone healing. *J. Immunol. Res.* **2015**, *2015*, 752510. [CrossRef] [PubMed]
3. Loi, F.; Córdova, L.A.; Pajarinen, J.; Lin, T.; Yao, Z.; Goodman, S.B. Inflammation, fracture and bone repair. *Bone* **2016**, *86*, 119–130. [CrossRef] [PubMed]
4. Pisani, P.; Renna, M.D.; Conversano, F.; Casciaro, E.; Di Paola, M.; Quarta, E.; Muratore, M.; Casciaro, S. Major osteoporotic fragility fractures: Risk factor updates and societal impact. *World J. Orthop.* **2016**, *7*, 171. [CrossRef] [PubMed]
5. Teng, G.G. Mortality and osteoporotic fractures: Is the link causal, and is it modifiable? *Clin. Exp. Rheumatol.* **2008**, *26*, 125.
6. Carlier, A.; Geris, L.; Lammens, J.; Van Oosterwyck, H. Bringing computational models of bone regeneration to the clinic. *Wiley Interdisciplin. Rev. Syst. Biol. Med.* **2015**, *7*, 183–194. [CrossRef] [PubMed]
7. Huber-Lang, M.; Kovtun, A.; Ignatius, A. The role of complement in trauma and fracture healing. *Semin. Immunol.* **2013**, *25*, 73–78. [CrossRef]
8. Foundation, I.O. *Osteoporosis Facts and Statistics*; Data and Publication, International Osteoporosis Foundation: Nyon, Switzerland, 2017.
9. Zhou, Z.; Redaelli, A.; Johnell, O.; Willke, R.J.; Massimini, G. A retrospective analysis of health care costs for bone fractures in women with early-stage breast carcinoma. *Cancer* **2004**, *100*, 507–517. [CrossRef]
10. Pajarinen, J.; Lin, T.; Gibon, E.; Kohno, Y.; Maruyama, M.; Nathan, K.; Lu, L.; Yao, Z.; Goodman, S.B. Mesenchymal stem cell-macrophage crosstalk and bone healing. *Biomaterials* **2018**. [CrossRef]
11. Kojouharov, H.V.; Trejo, I.; Chen-Charpentier, B.M. Modeling the effects of inflammation in bone fracture healing. *AIP Conf. Proc.* **2017**, *1895*, 020005.
12. Schlundt, C.; El Khassawna, T.; Serra, A.; Dienelt, A.; Wendler, S.; Schell, H.; van Rooijen, N.; Radbruch, A.; Lucius, R.; Hartmann, S.; et al. Macrophages in bone fracture healing: Their essential role in endochondral ossification. *Bone* **2018**, *106*, 78–89. [CrossRef] [PubMed]
13. Zhang, R.; Liang, Y.; Wei, S. M2 macrophages are closely associated with accelerated clavicle fracture healing in patients with traumatic brain injury: A retrospective cohort study. *J. Orthop. Surg. Res.* **2018**, *13*, 213. [CrossRef]
14. Marsell, R.; Einhorn, T.A. The biology of fracture healing. *Injury* **2011**, *42*, 551–555. [CrossRef] [PubMed]
15. Chung, E.; Son, Y. Crosstalk between mesenchymal stem cells and macrophages in tissue repair. *Tissue Eng. Regen. Med.* **2014**, *11*, 431–438. [CrossRef]
16. Schlundt, C.; Schell, H.; Goodman, S.B.; Vunjak-Novakovic, G.; Duda, G.N.; Schmidt-Bleek, K. Immune modulation as a therapeutic strategy in bone regeneration. *J. Exp. Orthop.* **2015**, *2*, 1. [CrossRef] [PubMed]
17. Chan, J.K.; Glass, G.E.; Ersek, A.; Freidin, A.; Williams, G.A.; Gowers, K.; Santo, A.I.E.; Jeffery, R.; Otto, W.R.; Poulsom, R.; et al. Low-dose TNF augments fracture healing in normal and osteoporotic bone by up-regulating the innate immune response. *EMBO Mol. Med.* **2015**, *7*, 547–561. [CrossRef] [PubMed]
18. Mantovani, A.; Sozzani, S.; Locati, M.; Allavena, P.; Sica, A. Macrophage polarization: Tumor-associated macrophages as a paradigm for polarized M2 mononuclear phagocytes. *Trends Immunol.* **2002**, *23*, 549–555. [CrossRef]
19. Border, W.A.; Noble, N.A. Transforming growth factor β in tissue fibrosis. *N. Engl. J. Med.* **1994**, *331*, 1286–1292.
20. Cavaillon, J.M. Pro-versus anti-inflammatory cytokines: Myth or reality. *Cell. Mol. Biol.* **2001**, *47*, 695–702.
21. Osta, B.; Benedetti, G.; Miossec, P. Classical and paradoxical effects of TNF-α on bone homeostasis. *Front. Immunol.* **2014**, *5*, 48. [CrossRef]

22. Arango Duque, G.; Descoteaux, A. Macrophage cytokines: Involvement in immunity and infectious diseases. *Front. Immunol.* **2014**, *5*, 491. [CrossRef] [PubMed]

23. Mountziaris, P.M.; Mikos, A.G. Modulation of the inflammatory response for enhanced bone tissue regeneration. *Tissue Eng. Part B Rev.* **2008**, *14*, 179–186. [CrossRef] [PubMed]

24. Gibon, E.; Lu, L.Y.; Nathan, K.; Goodman, S.B. Inflammation, ageing, and bone regeneration. *J. Orthop. Transl.* **2017**, *10*, 28–35. [CrossRef] [PubMed]

25. Gibon, E.; Loi, F.; Córdova, L.A.; Pajarinen, J.; Lin, T.; Lu, L.; Nabeshima, A.; Yao, Z.; Goodman, S.B. Aging affects bone marrow macrophage polarization: Relevance to bone healing. *Regen. Eng. Transl. Med.* **2016**, *2*, 98–104. [CrossRef] [PubMed]

26. Serhan, C.N.; Savill, J. Resolution of inflammation: The beginning programs the end. *Nat. Immunol.* **2005**, *6*, 1191–1197. [CrossRef] [PubMed]

27. Newman, S.L.; Henson, J.E.; Henson, P.M. Phagocytosis of senescent neutrophils by human monocyte-derived macrophages and rabbit inflammatory macrophages. *J. Exp. Med.* **1982**, *156*, 430–442. [CrossRef]

28. Italiani, P.; Boraschi, D. From monocytes to M1/M2 macrophages: Phenotypical vs. functional differentiation. *Front. Immunol.* **2014**, *5*, 514. [CrossRef]

29. Wang, Y.; Yang, T.; Ma, Y.; Halade, G.V.; Zhang, J.; Lindsey, M.L.; Jin, Y.F. Mathematical modeling and stability analysis of macrophage activation in left ventricular remodeling post-myocardial infarction. *BMC Genom.* **2012**, *13*, S21. [CrossRef]

30. Bailon-Plaza, A.; Van Der Meulen, M.C. A mathematical framework to study the effects of growth factor influences on fracture healing. *J. Theor. Biol.* **2001**, *212*, 191–209. [CrossRef]

31. Echeverri, L.; Herrero, M.; Lopez, J.; Oleaga, G. Early stages of bone fracture healing: Formation of a fibrin–collagen scaffold in the fracture hematoma. *Bull. Math. Biol.* **2015**, *77*, 156–183. [CrossRef]

32. Doblaré, M.; Garcıa, J.; Gómez, M. Modelling bone tissue fracture and healing: A review. *Eng. Fract. Mech.* **2004**, *71*, 1809–1840. [CrossRef]

33. Ghiasi, M.S.; Chen, J.; Vaziri, A.; Rodriguez, E.K.; Nazarian, A. Bone fracture healing in mechanobiological modeling: A review of principles and methods. *Bone Rep.* **2017**, *6*, 87–100. [CrossRef] [PubMed]

34. Einhorn, T.A. The science of fracture healing. *J. Orthop. Trauma* **2005**, *19*, S4–S6. [CrossRef] [PubMed]

35. Einhorn, T.A.; Gerstenfeld, L.C. Fracture healing: Mechanisms and interventions. *Nat. Rev. Rheumatol.* **2015**, *11*, 45. [CrossRef] [PubMed]

36. Gómez-Barrena, E.; Rosset, P.; Lozano, D.; Stanovici, J.; Ermthaller, C.; Gerbhard, F. Bone fracture healing: Cell therapy in delayed unions and nonunions. *Bone* **2015**, *70*, 93–101. [CrossRef] [PubMed]

37. Carlier, A.; Geris, L.; van Gastel, N.; Carmeliet, G.; Van Oosterwyck, H. Oxygen as a critical determinant of bone fracture healing—A multiscale model. *J. Theor. Biol.* **2015**, *365*, 247–264. [CrossRef] [PubMed]

38. Nagaraja, S.; Wallqvist, A.; Reifman, J.; Mitrophanov, A.Y. Computational approach to characterize causative factors and molecular indicators of chronic wound inflammation. *J. Immunol.* **2014**, *192*, 1824–1834. [CrossRef] [PubMed]

39. Yu, T. Design and Validation of A Mathematical Model to Describe Macrophage Dynamics in Wound Healing. Ph.D. Thesis, Drexel University, Philadelphia, PA, USA, 2014.

40. Otto, S.P.; Day, T. *A Biologist's Guide to Mathematical Modeling in Ecology and Evolution*; Princeton University Press: Princeton, NJ, USA, 2011.

41. Marée, A.F.; Komba, M.; Dyck, C.; Labkecki, M.; Finegood, D.T.; Edelstein-Keshet, L. Quantifying macrophage defects in type 1 diabetes. *J. Theor. Biol.* **2005**, *233*, 533–551. [CrossRef]

42. Sherratt, J.A.; Murray, J. Models of epidermal wound healing. *Proc. R. Soc. Lond. B* **1990**, *241*, 29–36.

43. Isaksson, H. Recent advances in mechanobiological modeling of bone regeneration. *Mech. Res. Commun.* **2012**, *42*, 22–31. [CrossRef]

44. Schell, H.; Duda, G.; Peters, A.; Tsitsilonis, S.; Johnson, K.; Schmidt-Bleek, K. The haematoma and its role in bone healing. *J. Exp. Orthop.* **2017**, *4*, 5. [CrossRef] [PubMed]

45. Van Furth, R.; Diesselhoff-den Dulk, M.M.; Mattie, H. Quantitative study on the production and kinetics of mononuclear phagocytes during an acute inflammatory reaction. *J. Exp. Med.* **1973**, *138*, 1314–1330. [CrossRef] [PubMed]

46. Bastidas-Coral, A.P.; Bakker, A.D.; Zandieh-Doulabi, B.; Kleverlaan, C.J.; Bravenboer, N.; Forouzanfar, T.; Klein-Nulend, J. Cytokines TNF-*α*, IL-6, IL-17F, and IL-4 differentially affect osteogenic differentiation of human adipose stem cells. *Stem Cells Int.* **2016**, *2016*, 1318256. [CrossRef] [PubMed]

47. Gilbert, L.; He, X.; Farmer, P.; Boden, S.; Kozlowski, M.; Rubin, J.; Nanes, M.S. Inhibition of Osteoblast Differentiation by Tumor Necrosis Factor-*α* 1. *Endocrinology* **2000**, *141*, 3956–3964. [CrossRef] [PubMed]

48. Lacey, D.; Simmons, P.; Graves, S.; Hamilton, J. Proinflammatory cytokines inhibit osteogenic differentiation from stem cells: Implications for bone repair during inflammation. *Osteoarthritis Cartilage* **2009**, *17*, 735–742. [CrossRef] [PubMed]

49. Lu, Y.; Lekszycki, T. Modeling of an initial stage of bone fracture healing. *Contin. Mech. Thermodyn.* **2015**, *27*, 851. [CrossRef]

50. Hoegel, F.; Abdulazim, A.; Augat, P.; Buehren, V. Quantification of reaming debris at the fracture gap of diaphyseal A2 and A3 Fractures After reamed intramedullary nailing of the sheep tibia. *Eur. J. Trauma Emerg. Surg.* **2008**, *34*, 587–591. [CrossRef]

51. Gaston, M.; Simpson, A. Inhibition of fracture healing. *Bone Joint J.* **2007**, *89*, 1553–1560. [CrossRef]

52. Roberts, T.T.; Rosenbaum, A.J. Bone grafts, bone substitutes and orthobiologics: The bridge between basic science and clinical advancements in fracture healing. *Organogenesis* **2012**, *8*, 114–124. [CrossRef]

53. Wu, M.; Chen, G.; Li, Y.P. TGF-*β* and BMP signaling in osteoblast, skeletal development, and bone formation, homeostasis and disease. *Bone Res.* **2016**, *4*, 16009. [CrossRef]

54. Recknagel, S.; Bindl, R.; Brochhausen, C.; Göckelmann, M.; Wehner, T.; Schoengraf, P.; Huber-Lang, M.; Claes, L.; Ignatius, A. Systemic inflammation induced by a thoracic trauma alters the cellular composition of the early fracture callus. *J. Trauma Acute Care Surg.* **2013**, *74*, 531–537. [CrossRef] [PubMed]

55. Stuart, A.; Humphries, A.R. *Dynamical Systems and Numerical Analysis*; Cambridge University Press: Cambridge, UK, 1998.

56. Strang, G. *Introduction to Linear Algebra*; Wellesley-Cambridge Press: Wellesley, MA, USA, 1993.

Article

Cellular Automata and Artificial Brain Dynamics

Alberto Fraile [1,2,*] ![ORCID], **Emmanouil Panagiotakis [1], Nicholas Christakis [1] and Luis Acedo [3,*]** ![ORCID]

[1]   Crete Center for Quantum Complexity and Nanotechnology, Department of Physics, University of Crete, 71003 Heraklion, Greece; ph3621@edu.physics.uoc.gr (E.P.); nchrkis@gmail.com (N.C.)

[2]   Department of Control Engineering, Faculty of Electrical Engineering, Czech Technical University in Prague, 16000 Prague, Czech Republic

[3]   Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València, 46022 Valencia, Spain

*   Correspondence: albertofrailegarcia@gmail.com or frailalb@fel.cvut.cz (A.F.); luiacrod@imm.upv.es (L.A.); Tel.: +420-224-35-7687 (A.F.)

Abstract: Brain dynamics, neuron activity, information transfer in brains, etc., are a vast field where a large number of questions remain unsolved. Nowadays, computer simulation is playing a key role in the study of such an immense variety of problems. In this work, we explored the possibility of studying brain dynamics using cellular automata, more precisely the famous Game of Life (GoL). The model has some important features (i.e., pseudo-criticality, $1/f$ noise, universal computing), which represent good reasons for its use in brain dynamics modelling. We have also considered that the model maintains sufficient flexibility. For instance, the timestep is arbitrary, as are the spatial dimensions. As first steps in our study, we used the GoL to simulate the evolution of several neurons (i.e., a statistically significant set, typically a million neurons) and their interactions with the surrounding ones, as well as signal transfer in some simple scenarios. The way that signals (or life) propagate across the grid was described, along with a discussion on how this model could be compared with brain dynamics. Further work and variations of the model were also examined.

Keywords: cellular automata; game of life; brain dynamics

---

1. Introduction

The question "Can cellular automata model brain dynamics?" or in our case "Can the Game of Life model brain dynamics?" is hard to answer for, at least, three reasons. First, despite decades of efforts to understand the Game of Life (GoL hereafter), several questions remain unanswered. Second, although brain dynamics is an exponentially growing field in neuroscience, several questions remain under strong debate. Third, it is still not known which characteristics are fundamental in brain dynamics and which characteristics are derived from (or are dependent upon) a particular biological realization. Hence, the goal of this study was fairly ambitious and only some possible ideas and paths to follow were presented.

Cellular automata (CA) have been studied since the early 1950's [1–3] in connection with many different problems, such as predator–prey, [4,5] chemistry, [6] dendritic growth [7], and many more. For instance, Reference [8] provides a review of the biological applications. Different CA have been used in several problems related to neuroscience [9–13]. Amongst the various CA, the GoL was chosen because it presents several characteristics that are, in our opinion, good reasons to associate brain dynamics with the evolution of life as modelled using the GoL. It is an established idea in science, particularly in chaos theory and complex systems, that very simple rules can often generate great complexity [14], and probably nothing in nature is more complex than the human brain.

The brain was included amongst the abundance of physical and biological systems that exhibit different levels of criticality, as described in References [15–17] and the references therein.

To function optimally, neuronal networks were considered to operate close to a non-equilibrium critical point [18]. However, recent works provide evidence that the human brain operates not in a critical state, but rather it keeps a distance to criticality [19–21], despite the computational advantages of criticality [22]. Therefore, it appears that maximization of information processing is not the only goal function of the brain. According to these results, the brain only tunes itself closer to criticality when it is necessary for information processing. Otherwise, the brain maintains a safety margin, losing processing capacity, but avoiding instabilities (e.g., an epileptic crisis) [19–21]. This hypothesis, however, is still controversial [23–26].

GoL was originally supposed to present self-organized criticality (SOC). Power-law decay of several statistical quantities was interpreted as evidence of SOC [27]. For example, the distribution D(T) of time steps required for the lattice to return to stability following a random single-site perturbation is characterized by D(T) ~$T^{-1.6}$ [27]. More recently, Hemmingsson [28] suggested that the GoL does not converge to a SOC state, but it is in fact subcritical with a long length scale. Additional evidence has also been put forward by Nordfalk et al [29], wherein this capacity of the GoL model suggests a connection to brain dynamics.

A second point is that the GoL presents a $1/f$ noise [30], a feature that is believed to be present in brain dynamics [31,32] in electro- and magneto-encephalogram spectra (EEG and MEG) [33], and in human cognition as well [34]. Results suggest that there is a relationship between computational universality and $1/f$ noise in cellular automata [30]. However, discussion continues regarding whether or not $1/f$ noise is evidence of SOC [35]. Langton asked "Under what conditions will physical systems support the basic operations of information transmission, storage, and modification constituting the capacity to support computation?" [22]. Researching on CA, Langton concluded, "the optimal conditions for the support of information transmission, storage, and modification, are achieved in the vicinity of a phase transition" [22]. It has been shown that the GoL can undergo different phase transitions when topology or synchronicity is modified (see the following section). Wolfram suggested that class IV CA are capable of supporting computation, even universal computation, and that it is this capacity which makes their behavior so complex [36]. Proof that the GoL is computation-universal employs propagating "gliders" (See Figure 2) as signals (i.e., the glider can be considered as a pulse in a digital circuit) and period-2 "blinkers" (See Figure 1) as storage elements in the construction of a general-purpose computer [37–40]. It is possible to construct logic gates such as AND, OR, and NOT using gliders. One may also build a pattern that acts like a finite state machine connected to two counters. This has the same computational power as a universal Turing machine, as such, using the glider, the GoL is theoretically as powerful as any computer with unlimited memory and no time constraints, i.e., it is Turing complete [37,39].

The objective of this paper was to find some characteristics of the statistical behavior of the GoL that made it suitable for comparison with some general features of brain dynamics. Of course, it was not our intention to give an accurate, or even approximate, representation of any of the brain functions and processes with such a simple model. Large scale simulations of a detailed nature for the human brain have already been performed at the so-called Human Brain Project (HBP) [41]. In any case, the fundamental question of how to formulate the most basic mathematical model which features some of the essential characteristics of the brain connected to information processing, dynamics, statistical properties, memory retrieval, and other higher functions, is still open. In this respect, the study of cellular automata could still be a promising path to analyze these characteristics or in answering the question of how the brain should be defined.

The paper is structured as follows: In Section 2, a short introduction is given about the GoL cellular automaton. Then, the idea of the GoL model for basic features of brain dynamics is given in Section 3. In Section 4, the idea of spiking is introduced into the model. Section 5 is devoted to the effect of defects in the GoL lattice and their possible role in degenerative diseases of the brain. Possible extensions of the model are discussed in Section 6. Finally, general conclusions are drawn in Section 7.

2. Game of Life: A Brief Introduction

Conway's GoL is a cellular automaton with an infinite 2D orthogonal grid of square cells, each of which is in one of two possible states, i.e., active or inactive. Every cell interacts with its eight nearest neighbors (NN), which are the cells that are horizontally, vertically, or diagonally adjacent (called Moore neighborhood). At each step in time, the following transitions occur, as in Reference [38]:

- Any active cell with fewer than two active neighbors, becomes inactive, i.e., under-population.
- Any active cell with two or three active neighbors, lives on to the next generation.
- Any active cell with more than three active neighbors becomes inactive, i.e., over-population.
- Any inactive cell with exactly three active neighbors becomes active, i.e., reproduction.

Starting from random initial conditions, "Life" will evolve through complex patterns, eventually settling down to a stationary state composed of different stable configurations, both static and periodic (see Figure 1). Despite its simple algorithm, the GoL simulates the dynamic evolution of a society of living individuals, including processes such as growth, death, survival, self-propagation, and competition. A class IV CA quickly produces apparent randomness, which is consistent with the tendency towards randomness observed in Brownian motion [36].
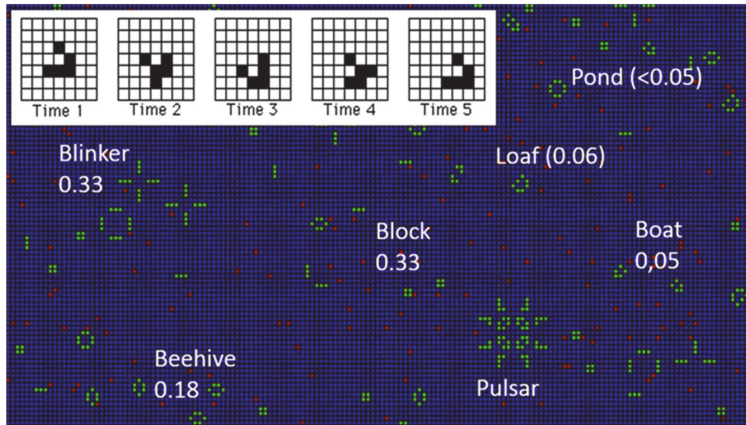
Figure 1. (Color online) Example of the most common stable clusters in the GoL (blinkers, blocks, beehives, etc.) and an example of a not-so-common cluster (pulsar). Green cells are the active cells, blue are inactive, and red ones represent defects (see Section 5). Numbers stand for the probability of appearance [42]. Inset. Glider and its movement. Active cells are represented by a black square, while inactive cells are empty. This glider moves diagonally downward and to the right by one cell every four updates.

Let us examine a few characteristics of the GoL: Simulations can be carried out in open, closed, or periodic boundary conditions. In the present work, open boundary conditions were utilized, unless otherwise stated. Starting from a random initial distribution, life evolves to a final state composed of small stable clusters (see Figure 1 and Video 1 in the Supplementary material), usually blocks, blinkers, beehives, and others less likely to appear (see Table 1). An important entity is the "glider". In the GoL, there also exist configurations which "move" uniformly across the grid, executing a cycle of a few internal states. The simplest example is the glider that contains five live sites (Figure 1, Inset) and undergoes a cycle of length four.

Table 1. The most common clusters found in the GoL. Columns: Name of object, Size (number of live cells per object), Period (0 means no period exists), Heat (number of changing cells), Influence area (all cells which are in the Moore-neighborhood of each live cell of an object).

Name	Size	Period	Heat	Influence
Blinker	3	2	4	21
Block	4	1	-	16
Beehive	6	1	-	26
Loaf	7	1	-	30
Glider	5	4	5	22

Empirical investigations suggest that the density of structures containing L live sites generated from a disordered initial state decreases like e^l/L, where l is the size of the minimal distinct configuration which evolves to the required structure in one time step [43]. The time needed for the system, when N is large, to reach equilibrium is approximately equal to N, N×N being the size of the box. This is not true for special small configurations, as we will discuss later. The final density of life is small, independent of the initial one, unless the initial density is very small or very big (then life disappears completely) [44].

How Life resists noise was studied as early as 1978 [45]. In Schulman's work, the GoL was studied including a stochastic element (that was called "temperature"). The idea was to examine how the introduction of a stochastic element in the local evolution rule would perturb the long-term evolution of the system. Temperature, in their study, was represented by the random perturbation of cell states as the rule was iterated. Then, the statistical properties examined were the density as a function of temperature and the (suitably defined) entropy. A phase transition was observed at a certain "Temperature" and certain "thermodynamic" constant of the motion [45]. A different approach was adopted in Reference [36], wherein by introducing stochastic rules in the GoL, it was found that "Life" exhibited a very rich critical behavior. Discontinuous (first-order) irreversible phase transitions between an extinct phase and a steady state supporting life were found [36]. A statistical analysis of the GoL can be found in Reference [46]. In that study, the number of clusters of *n* live sites at time *t*, the mean cluster size, and the diversity of sizes among other statistical functions was given. The dependence of the statistical functions on the initial density of live sites was examined as well.

Recently, dynamical aspects of CA have been studied by analyzing various characteristic parameters used in network theory [47–49]. The effect of network topology, i.e., the links between cells, on the evolution of the GoL has been studied using a version of the GoL on a 2D small-world [50]. The effect of synchronicity has also been examined in different studies [51–53].

For example, in Reference [52], the authors claim that an asynchronous GoL simulates exactly the behavior of the GoL in terms of universal computation and self-organization, no matter whether the update of cells is simultaneous or independent, according to some updating scheme like a time-driven [54] or step-driven [53] method. An abrupt change of behavior in the asynchronous GoL was presented in the work of Blok and Bergersen [53]. The authors showed that the phenomenon was a second-order phase transition, i.e., the macroscopic functions that describe the global behavior as a function of synchrony rate α (when $\alpha = 1$ we have the classical synchronous updating and the system is deterministic, for $\alpha < 1$ the system becomes stochastic) are continuous but their derivatives are discontinuous at the critical point $\alpha_c \sim 0.9$. The critical threshold α_c separates two well-distinguished macroscopic behaviors or phases. The phase $\alpha > \alpha_c$ is the frozen phase, in which the system evolves with low-density patterns and quickly stabilizes to a fixed point. The second phase ($\alpha < \alpha_c$) is the labyrinth phase, characterized by a steady state with higher density and the absence of stabilization at a fixed point [53]. Further research of the behavior of the GoL under different rules, i.e., different intervals of survival and fertility can be found in Reference [55]. However, all this complexity must not be confused with chaos. The GoL is not a chaotic system, it is

deterministic but extremely complex. Life is not at the "border of chaos," but thrives on the "border of extinction" [56].

The list of mentioned works does not claim to be complete. It only presents some important studies regarding different aspects of the GoL which could be of importance in our model as we will discuss later. In the presented work, to present our ideas, we utilized the original 2D GoL without any modification to either the topology or the synchronicity. To the best of our knowledge, almost nothing has been published connecting the GoL to brain dynamics, except for the work of de Beer [57,58].

3. The Basic Idea of the Model

The human brain comprises of a set of 10^{10}–10^{11} individual neurons and the total number of connections is even larger because every cell projects its synapses to a total of 10^4–10^5 different neurons [59]. With such an immense number of connections, our model is not intended to accurately describe the human brain or any other real nervous system, but rather to examine overall properties of a big connected group of discrete "neuron-like" elements. As a first approximation, we can assume that live cells in the GoL correspond to active neurons and dead cells to inactive neurons. As such, if the cells in the 2D grid are assumed to be single neurons, then the model may be considered unrealistic, since usually, a single neuron is connected on average to 10,000 neurons, and not only eight. This can be modified easily; counting second NN (the second ring around the cell) cells would have 24 connections, counting third NN cells would have 48 and so on, but that would make the model much more expensive, computationally speaking. To count 100 connections, it would be necessary to consider up to 15 NNs, for 1000 connections it would be necessary to consider 49 NNs.

More importantly, the resulting GoL differs significantly from the original Conway's GoL. Another possibility is to assume that a cell in the simulation corresponds to a number of neurons (how big or small is an open question), or the other way around, that a number of cells (again, how many is unknown) correspond to a single neuron. For example, in the model proposed by Hoffman [10] the cells in the simulation represent small groups of neurons in the cortex. A cell in our simulation could be considered to represent, for example, a thousand neurons and that the neurons there, contained in a single cell, are connected with the other 8000 neurons around. This may look artificial but raises the first interesting question: Would a "large" (1000 or more) number of neurons interact, on average, with neighboring groups of the same size in the way that is modeled by the GoL?

In any case, despite the crude oversimplification, our question may be reformulated as: Is there any magnitude or measurable property in brain dynamics that behaves, on average, or presents some similarities with the GoL outputs? It must be noted at this point that all presented simulations (unless otherwise stated) were performed for randomly generated initial configurations in a 1000 × 1000 square box, with open boundary conditions, and the presented results for each different configuration were averaged over 20 simulations. Figure 2 (top panel) shows the time evolution of density (i.e., the number of live cells divided by the total number of cells), starting from different initial densities. As can be seen, in a few steps (100), the density of life decreases very fast (almost exponentially) to a lower density that then decreases to a final value of ~2.7%, independent of the initial one. The exact way in which density decreases is, to the best of our knowledge, unknown. The curves, in the first 100 steps (approximately), can be fitted to an exponential decay.
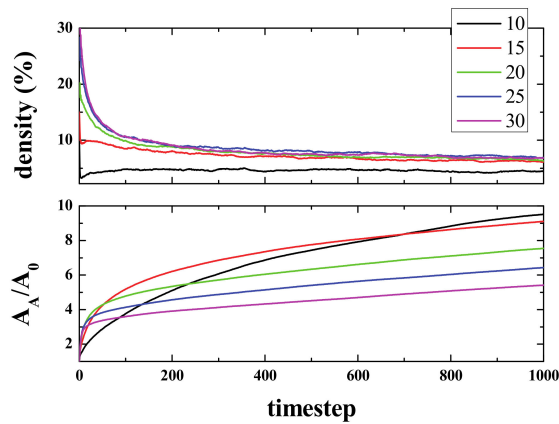
Figure 2. (Color online) Density (top panel) and Activated Area (bottom panel) versus time for five different initial densities, ranging from 10% (black), 15% (red), 20% (green), 25% (blue), and 30% (magenta).

The lower panel shows the normalized Activated Area, A_A/A_0 where A_A is the area in the simulation that has been active at least once, and A_0 is the initial active area (the initial number of live cells). The Activated Area grows faster for high densities (initial density $\rho_i \geq 15\%$) only during the first stages of the simulation (for times ≤ 500). After this first stage, the slope decreases, whilst for $\rho_i = 10\%$, A_A continues growing with a similar slope. Hence, for the activated area, two stages can be differentiated in the evolution of the system. Initially, the area increase occurs at a fast rate, and then, after some time $\tau(\rho_i)$, the rate of increase becomes slower. It is important to note that the ratio of A_A/A_0 is > 1, but that does not mean multi-counting. When the activated area is computed, we count the cells that have been alive at least once. However, when a cell is counted, then it is no longer counted again. That is to say, if a cell is alive at time 0 (or time x), it is not counted again if it goes back to life at time n (> 0 (or > x)). The idea is very much like the random walk study presented in Reference [60]. The reason to have a ratio > 1 is that we start, for example, with an initial density of 10%, i.e., 10% of the initial area is active. If this is the case in a closed simulation, then the ratio can be as high as 10 at a certain time (and unlimited in an open (infinite) simulation). As for the reason to define the activated area in our model, there are studies where the spatial extension of brain activity is studied [61,62].

Clearly, a one-to-one comparison of our model to real brain data would be highly desirable. However, that is not the main goal of this study, but rather it is to create a simple framework (with high complexity at the same time as we will see) capable of being used as a model of brain dynamics. Comparing the activated area, signal transmission, etc., as defined in our model to real experiments (EEG, fMRI, etc.) requires further study and collaboration with neurologists.

At this stage, an interesting conclusion may be drawn: The maximum/optimal spreading of activity in the grid is achieved for initial densities between 10 and 15%. Figure 2 suggests a very likely conjecture: "For large enough sample sizes and long simulation times, on average, the product Density times the Activated Area remains constant for all initial densities". To illustrate this idea, Figure 3 shows the product Density x Active Area versus time. Note that this conjecture cannot be true in open simulations (infinite side), since the area can grow endlessly (due to emitted gliders moving away from the initial live cells), but density will reach a constant value (or equilibrium value).
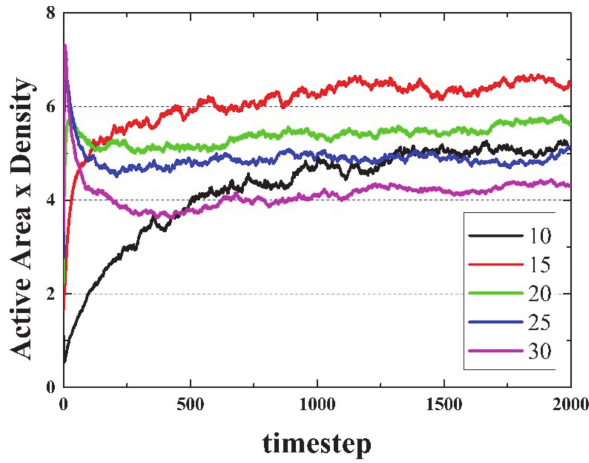
Figure 3. (Color online) Density x Active area vs. time for five different initial densities, from 10 to 30%, as labeled.

The activity can be defined as:

$$X = [\rho(t) - \rho(t-1)]^2 \tag{1}$$

i.e., as the squared difference between consecutive densities. As seen in Figure 4, the activity decays more than three orders of magnitude within the first 100 steps. The figure is noisy, since it has not been averaged over a number of simulations.
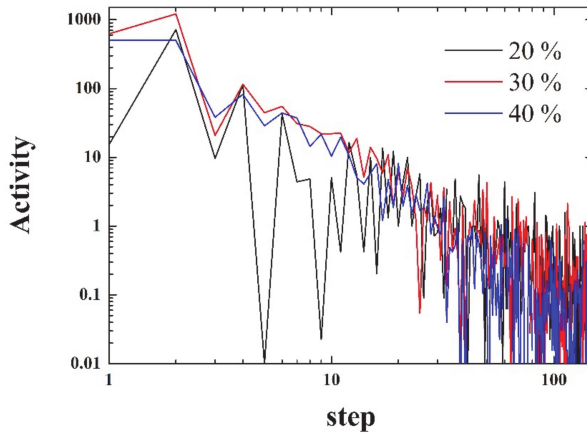
Figure 4. (Color online) Activity (as defined in Equation (1)) vs. timestep for three different initial densities (20, 30, and 40%), as labelled. Note the logarithmic scale on both axes.

Despite the possible answer to the previous questions, a possible first idea is to use a random distribution of neurons, some active and some inactive, in an open grid, and since in the final state some neurons will be active far from the initial ones, associate the simulation to an impulse or stimulus that is originated somewhere (the initial configuration) and evolves temporally. We can study how the "information" or the activity spreads (since is difficult to define information). Figure 5 (top panel) shows an example, in order to make this clearer. It presents the spreading of activity for the same initial density (20%), but different random starting configurations, after 2000 timesteps. The color scale corresponds to the number of times that the cells have been active during the simulation. For the sake

of clarity, the scale has been limited between 0 (dark blue) and 20 (red), meaning that a red cell has been active during at least 20 steps. As can be seen, the activity spreads from the initial emplacement in a "random way". The total number of cells that are active (no matter how many times) for a given time will be called the "activated area" or "sampled area". This area will vary from one simulation to the next, both in size and shape, as can be seen in the three presented examples. It must be noted that there exist spots on the grid that have been active for a total time that can be very close to the total number of steps, which can be the case of a stable configuration formed at the very beginning that did not interact again with the rest of the system. The bottom panel of Figure 5, which refers to spiking activity, will be discussed in a later section.
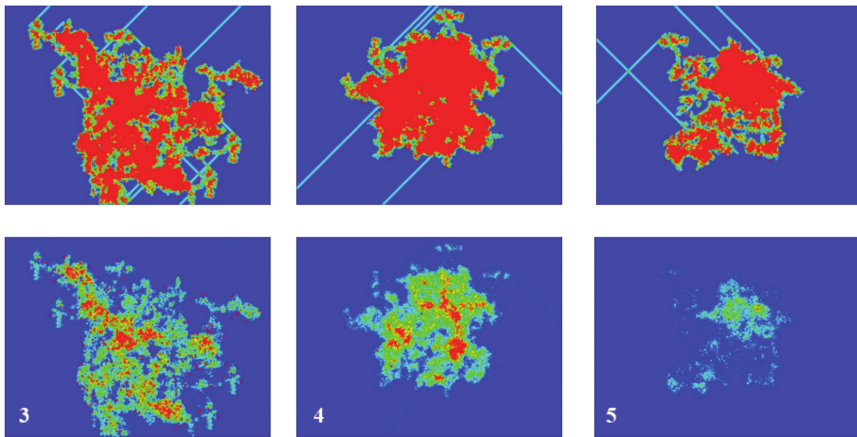
Figure 5. (Color online) Three different simulations with random configuration of equal initial density (20%). Top panel: Active area A_A. The straight lines correspond to the traces left by gliders. Bottom panel: Spiked area A_S, i.e., neurons that have spiked at least once. From left to right, Spiking Rules R3, R4, and R5, respectively (see Section 4). Color scale in both panels is between 0 (dark blue) and 20 (red). All results correspond to time 2000.

To be precise, the simulation may be terminated once the density becomes constant, or when the "sampled area" reaches a final constant value. By choosing the second option, this can lead to an infinite number of steps and a never-ending simulation, since gliders appear quite often, which would lead to a non-stopping increase of the sampled area if they move away from the center of the system. In our analyses, we neglected the fact that the system was still active due to the presence of gliders (or any other "ships"), and we defined equilibrium as the time after which no more changes are observed in the system, except for the periodic ones due to blinkers and similar clusters.

Thus, we can study how far and fast the signal (or activity) spreads. As a first approximation, the activated area (considered a circular one) can be written as $\pi r^2(t)$ (r being the distance from the center of the box), hence we can write the velocity of the spreading signal as:

$$v(t) = \frac{1}{\sqrt{\pi}} \frac{d\sqrt{A_A}}{dt} \qquad (2)$$

Figure 6 shows the velocity calculated from the curves of the activated area vs. time, presented in Figure 2. For comparison, we have also plotted the result considering the spread as a pure random walk (green dashed curve).
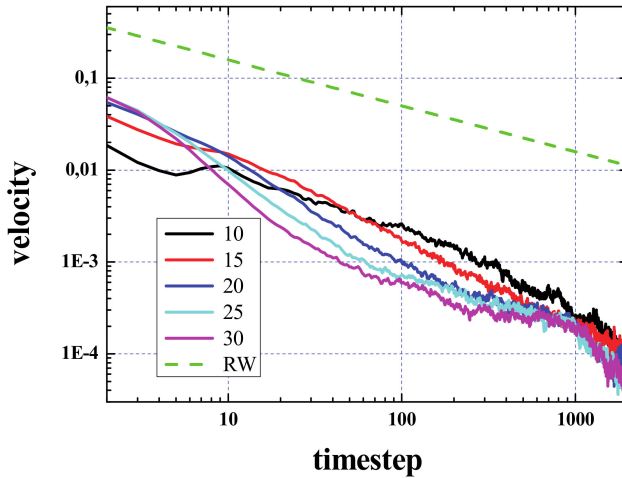
Figure 6. (Color online). Velocity vs. time for five different initial densities as labelled. The green dashed curve shows the result assuming a simple random walk. Note the logarithmic scale on both axes.

It has to be noted that the activated area will increase outwards as well as inwards. The velocity of the signal propagation calculated from the increase of this area as a circle that grows from the center of the initial configuration is only valid after the area of the box containing the initial configuration has been sampled. This occurs for values of time that depend on the density and is about a few hundreds of steps. In the very first steps of the simulation, a big increase of this area is expected, since, grosso modo, most of the dead cells around the living ones will become live in few steps, so that the activated area will increase from A_0 to $2A_0$, $3A_0$, etc., in only a few steps. At the same time, as has already been discussed, in few steps the density decreases, and this process is fast as well (see top panel of Figure 2), so that in a few steps the increase of the initial active area will not be "explosive" anymore.

This simple computational experiment could be the first step towards answering a number of important questions and determining whether the GoL is a suitable model. How does a signal propagate in the brain? Is there experimental data that might relate to this simple model of neuronal activity? Can we define the initial configuration as an impulse or stimulus (in microvolts, with the density determining its range for example), and its subsequent evolution as the spreading of that signal or information?

Moreover, it has been argued that in the brain, volume conduction (or electrical spread) is a passive resistive process, where the amplitude of the extracellular signal decreases as the distance from the active membrane increases, as described in Reference [63] (see Figures 1 and 3 in that paper). This is not an argument to verify our model; however, our model presents similarities, since the signal propagates fast around the initial activated region and slowly far from it. In the CA model proposed by Hofmann, the excitation of a region of space does not propagate to distant regions [10], this is also true of cortical regions under normal conditions [64]. Similar behavior is predicted by our model.

The activated area, A_A increases from A_0 to 10 to 15 times A_0, which means that the initial square is filled (i.e., A_A is multiplied by a factor $100/\rho_i$) and then little spread is observed. In terms of the initial size, N×N, the final activated area (or sampled area) would be less than 2N × 2N, so that the signal (neglecting the propagation due to gliders that will continue growing forever) is not spreading more than N cells from the borders of the initial distribution. However, this is something that may vary from one simulation to the next, and it is certainly not true for special configurations such as the "Methuselahs," as we will discuss later.

The first possible connection of the presented model with brain dynamics is shown at this point. We compare the spreading of activity with the cortical spreading depolarization (CSD) wave [65].

Spreading depolarization waves of the neurons (and neuroglia) propagate across the gray matter at a velocity of 2–5 mm/min, [66,67], i.e., 0.33–0.83 $\times$ 10^{-4} mm/ms. If we calculate the velocity of growth of the active area during the first 100 steps, when A_A grows fast, it is found that V = 10^{-3}–10^{-4} cells/step. If we assume that the timestep is in ms, in accordance to Reference [68], and cell size is of the order of mm^3, then V is predicted to be higher than that of the CSD by one order of magnitude. To make this comparison, we have assumed a one-to-one correspondence between neurons and cells in our model. However, since the number of actual neurons represented by each CA cell is a user-defined parameter, in that sense the CA model is flexible and can achieve better agreement.

It has to be noted that if each computational cell is assumed as a single neuron, there exists some ambiguity, since the size of neurons is not constant but depends on the type of neuron, etc. For instance, pyramidal cells are big (~20μm), whilst stellate cells are smaller [69]. Moreover, CSD is not the only measurable quantity that exists in brain dynamics and other phenomena may be modeled. For instance, during an epileptic crisis, the rate of migration of hypersynchronous activity (hallmark of epilepsy) has been experimentally recorded and the estimated value is about 4 m/ms, [70,71], i.e., approximately 60 times higher than that of the CSD.

4. Spikes in the GoL Model

Inspired by the Wilson-Cowan model [72,73], we defined spikes in a similar manner. Most neurons require a minimum level of input before they fire. That threshold in our model can be the number of adjacent active cells. In this way, a cell (neuron) spikes when it is active and has exactly x number of active neighbors. We will refer to spiking requirements or rules as R3, R4, R5, and R6, with 3, 4, 5, and 6 being the previously defined number x. In each case, a cell is considered as having had spiking activity if it has spiked at least once, obeying either of the previously mentioned spiking rules. In this way, similar to the active area, the spiked area A_S may be defined at each timestep as the total number of cells that have spiked at least once (obviously, A_S as A_A can never decrease). It is a well-known fact in neurophysiology that the same neuron can become inactive and active (firing) many times per second [72,73]. Regardless, real neurons have a period of refractoriness after which they can fire again. This period last about 15 milliseconds and in our model, we assumed that this discrete timestep corresponded to the discrete unit of time.

Please note that in the W-C model, the neighborhood includes only 4 cells (von Neumann neighborhood), whilst in our model it included eight (as in the classical GoL model). Moreover, in the present version of the model, there is no differentiation between excitatory and inhibitory neurons. Obviously, these conditions could be easily modified. A simulation was carried out, where a randomly distributed initial density of 20% (i.e., 200,000 active cells in a 1000 $\times$ 1000 grid) was left to evolve, following spiking rules R3, R4, R5, and R6, and the spiking activity was measured. The averaged results of 20 different simulations for each rule are presented in Figure 7.
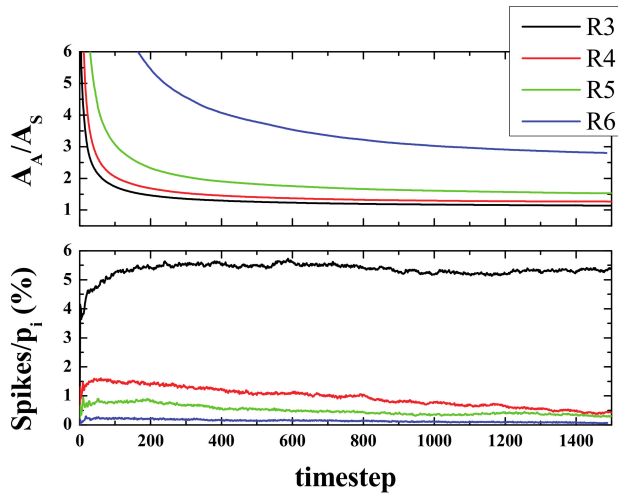
Figure 7. (Color online) Top panel: Ratio between the activated area and the spiked area, A_A/A_S vs. time. Bottom panel: Number of spikes divided by the total initial population (i.e., 200,000), as a percentage of this.

The spike curves prove helpful in understanding the evolution of the system. The number of spikes for a given time are the number of active neurons that have exactly 3, 4, 5, or 6 active neighbors at that step. Since we started from a low-density configuration ($\rho_i = 20\%$), at the very first stages, the number of spikes was found to be small in terms of the initial area (or initial population pi) and it grew, but not by much, within a few steps (less than 10). Then, after this initial reorganization, an almost flat curve was observed, especially for rules R4, R5, and R6. The average value for the first 200 steps after the aforementioned reorganization (from 20 to 220 steps) was as follows: 5.15, 1.48, 0.80, and 0.22 for R3, R4, R5, and R6, respectively (see bottom panel of Figure 7).

The ratio A_A/A_S converged fast to an almost constant value for R3, R4, and R5, but not as fast for R6. For R3 and R4 the ratio was almost 1, i.e., the spiked area becomes almost equal to the active area. However, the number of spikes with the R3 rule was more than 5 times larger than the one with R4. For R5 that ratio was 1.5 and for R6 the ratio was approximately 2.5 (see top panel of Figure 7).

Returning to Figure 5, its bottom panel presents the results of the spiked area A_S at timestep 2000, for three simulations with a random initial configuration of density of 20% in a 1000×1000 grid and spiking rules R3, R4, and R5, respectively. The top panel presents the respective active areas (as has already been discussed) and the colors for the sake of clarity have been limited between 0 and 20. The shape and size of both areas (active and spiked) vary significantly between different simulations, a fact arising from the different initial random configurations. However, the ratio A_A/A_S has not changed and remains almost constant at 1 for R3 and R4 and 1.5 for R5, as discussed in the previous paragraph.

From the above analyses, one could deduce that starting from a small community of live cells, let us say less than 10 or 20, the system would reach equilibrium in a few steps. However, this is not the case for some configurations called "Methuselahs" [74]. Some well-known Methuselahs are the acorn, the r-pentonimo, and the diehard. These apparently simple configurations (and their subsequent evolutions) are good examples of the amazing complexity that can be found in the GoL. Figure 8 shows the (a) acorn, (b) r-pentonimo, (c) diehard.
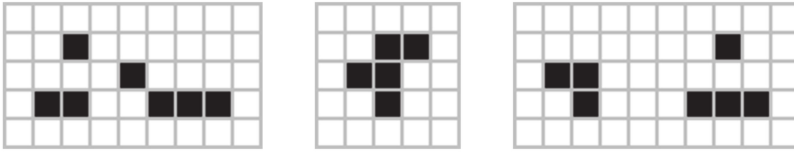
Figure 8. Some well-known Methuselahs; Left to right: (a) acorn, (b) r-pentonimo, (c) diehard, as shown in Reference [74].

See Video 2 in supplementary material for a simulation involving the acorn. The acorn resulted in a stable configuration after approximately 5000 steps. Starting from an initial configuration of the acorn at the center of the box (only 7 cells live), the stable pattern that emerged contained 633 cells, and the total sampled area was around 22,885 cells (a small part of this area accounted for the area sampled by some gliders before the system reached equilibrium). Using the acorn as a starting point, we examined the evolution of the spiking area under different spiking rules. Figure 9 presents the spiked area (for cells that have spiked at least once) at a timestep of 2000, for spiking rules R3, R4, R5, and R6, and as expected, the spiked area is lower when the number of neighbors required for a neuron to spike is higher.
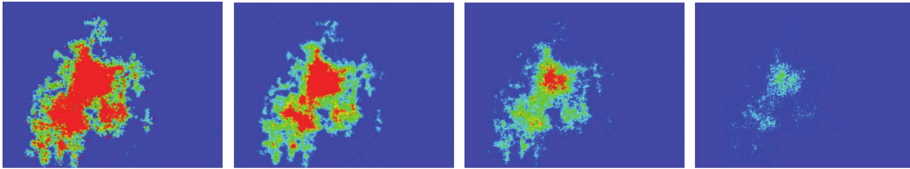
Figure 9. (Color online) Acorn simulation and spikes. (See Video 2). Cells that spiked at least once after 2000 steps. Color scale, as previously discussed, is limited between 0 (dark blue) and 20 (red). From left to right, Rules 3, 4, 5, and 6, respectively.

Figure 10 presents the results of the simulations for the acorn with 4 different spiking rules. As can be seen in the top panel, the spiked area for R3 (continuous black line) was almost equal to the active area (black dashed line), and then it decreased for the other rules in the expected order (i.e., R3, R4, R5, R6 from biggest to smallest). The bottom panel of Figure 10 presents the sum (from timestep 1 up to the current timestep) of spikes that have appeared for each of the 4 different spiking rules. As can be seen, the R3 curve exhibited, after the initial stages of the simulation, more than twice the number of spikes that the R4 curve did.
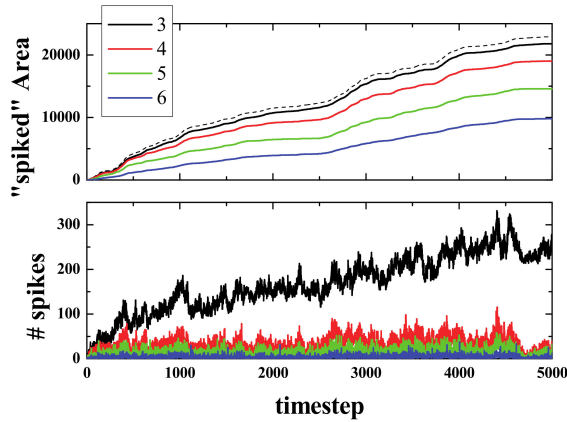
Figure 10. (Color online) The acorn simulation. Spiked area (top panel) and total number of spikes (bottom panel) vs. time (See Video 2) with different spiking rules, as labeled. The dashed black line on top represents the total active area A_A.

Simulating the acorn was interesting to show the expected differences once the spiking rule was modified; however, the results obtained using the acorn may be not valid in general. If the simulations were carried out with various initial random configurations with spiking rule R3, the number of spikes decays fast (see Figure 11, Top panel), and when the number of spikes is divided by the population (that decays exponentially), the resulting behavior of the spikes per neuron versus time is a small decrease in the very first stages of the simulation, only around the first 50 steps, as can be seen in Figure 11's bottom panel, followed by an almost constant value in the next steps.
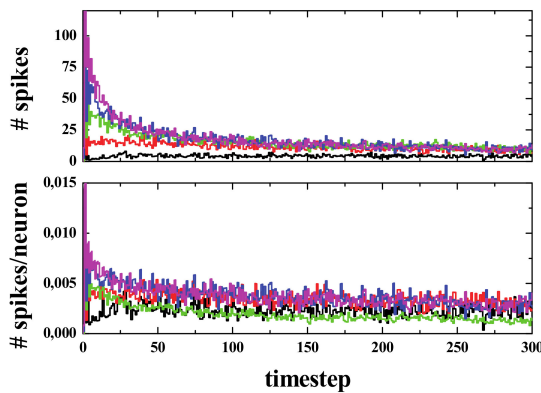
Figure 11. (Color online) Total number of spikes (Top panel) and averaged number of spikes per neuron (i.e., total number of spikes divided by total population, Bottom panel) vs. time for five different initial densities as labelled: 10 (black), 15 (red), 20 (green), 25 (dark blue), and 30% (magenta).

Apart from the very first stage of the simulation, we observed an average of 0.005 spikes (or firing neurons) per active neurons as time progressed. This can be compared to the results of a more realistic model like the one presented in Reference [68]. In that study (See Figure 2), the number of firing excitatory neurons is given. The value was about 400–500 firing neurons within a total of a million neurons, i.e., only 0.04–0.05%. It is also important to note that this result corresponded to a typical test with normal non-periodic behavior. Periodic states with high activity are associated with epileptic episodes [68].

Can this simple model of spiking activity be compared to real spiking activity in a real brain? How many spikes should we have per time and area? We have a certain number of spikes (per time, per area, etc.). It remains to be confirmed if this number is too low or too high compared to the usual number of spikes per neuron in the brain (or some part of it, depending on the activity).

5. Defects and Percolation

A simple ingredient that can be added to the present model is the presence of defects, i.e., cells that are inactive, neither dead nor alive, but are totally inactive all the time. These defects act as boundaries, i.e., they cannot become alive and are not counted when the GoL rules are applied. By doing so, we can study how these defects affect the propagation of life (i.e., a signal) as studied earlier. The basic idea or question behind this is: Is it possible to associate the presence of defects, defined in this way, to dead neurons that are present in neuronal diseases like Alzheimer or autism? [75–78].

We set up two communities separated by an empty space to study how the two communities mixed and/or interacted. In this way, we assumed that the two communities represented two parts of the brain (how big or small is an important question), which were somewhere disconnected (i.e., the part of the brain in between two active areas is in total rest state). We observed how they connected, i.e., we could define some kind of connectivity and studied this connectivity and its dependence to the initial density of active cells and other parameters.

Figure 12 shows the result of one of our simulations of this interaction. The sample size N × M was 100 × 300 (vertical by horizontal), meaning that the empty space between the two communities (i.e., the gap), N/3, was 33 cells. The initial density was 20% in the top and bottom regions (33 × 300 cells each). In the presented case, the number of defects was 2% in the middle region.
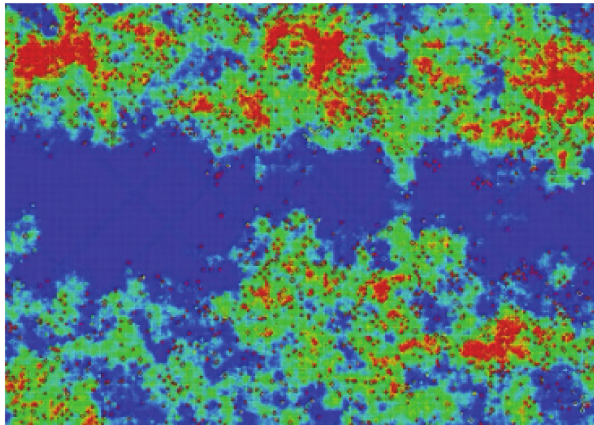
Figure 12. (Color online) Example of the mixing interaction of the two separated communities (up and down). The color scale corresponds to the number of timesteps that the cells have been active. See Video 3 in the supplementary material.

As can be seen, after some steps (350 steps, 35 secs in the video), both communities interact. For clarity's sake, when this occurs, we plotted the resulting "life" in yellow. A simple way to quantify the interaction is to study only the density of "yellow neurons," as we will see later. Henceforth, the density of active neurons after the interaction of the two separated communities will be referred to as "mixed density."

However, to analyze the effect of defects on the transmission of information or in the communication of the two parts of the "brain", let us understand the effect on the transmission or propagation of life. If the simulation was run with an initial density of life plus a percentage

of randomly distributed defects, what is observed is a clear decrease in the time needed to reach equilibrium. This can be observed in the density and active area curves, since both reach the final value in a few steps compared to the 0% defects case (See Figure 13).
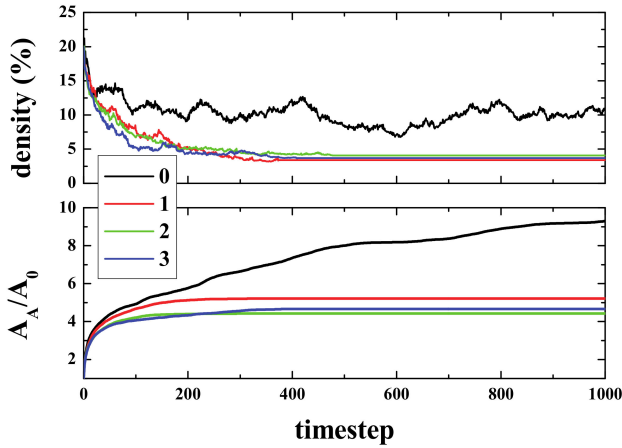
Figure 13. (Color online) Density (top panel) and Active Area, A_A, divided by initial area A_0 (bottom panel) vs. time for four different simulations adding randomly distributed defects to the lattice, from 0% to 3% as labelled.

To show how defects affect the signal propagation, we presented four different simulations with an initial density of 15% in the center of the box, and with defects being randomly distributed across the computational domain. The percentage of defects ranged from 0% to 3% (from left to right, see Figure 14). As can be seen, the higher the number of defects, the smaller the activated area, and the same applies for the spiked area (bottom panel). In the top panel, it can also be seen how the defects block the paths of gliders (for 1% and 3% defects).
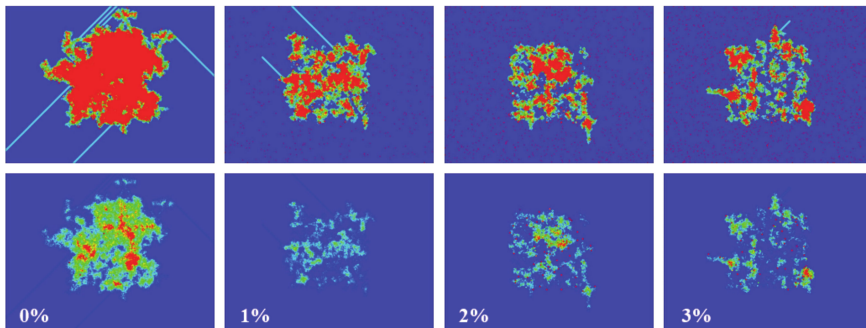
Figure 14. (Color online) Top panel: Activated Area after 1000 steps for initial density = 15%. Bottom panel: Spiked Area applying Rule 4 (scale goes from 0 (dark blue) to 20 (red)). From left to right, the percentage of defects is 0%, 1%, 2%, and 3% as labelled.

Figure 15 shows the "mixed density" for 5 simulations that were performed for two initially separated regions with initial densities, $\rho_i = 20\%$ each, for different densities of defects in the middle region (i.e., the same configuration as described for the simulations presented in Figure 13). The result was not absolute (or was relative) in two ways. First, it presented only single simulations and not

averaged curves. For the same conditions, the results changed significantly from one simulation to another, and in some cases, both communities would not interact at all. The idea was to demonstrate that the lower the number of defects, the easier it becomes for the two regions (upper and lower) to interact (i.e., lesser time is needed for the mixed density to become different from zero). Moreover, the interaction becomes stronger, if we consider the area under the density curve as a measure of strength (See Video 4 in the supplementary material).
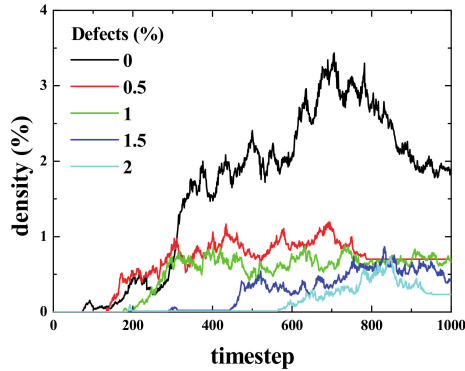
Figure 15. (Color online) Mixed density vs. time for five different percentages of defects as labelled. As can be seen, the lower the number of defects, the easier it is for the two blocks to interact.

The connectivity was achieved in several timesteps that depended on the densities of the upper and lower regions (See Figure 16, right panel). For low densities, the time was larger, and then for densities $\geq$ than 20%, it was approximately 125 steps, the gap distance was 100 cells, meaning that the communities "moved" towards each other with a velocity v = 50/125 = 0.4 grid-cells/timestep. Then for a given density (20% in this case), the time needed to interact will depend on the percentage of defects present in the middle region (Figure 16 left). When the percentage of defects is less than or equal to 0.75%, the effect is almost negligible; however, for bigger percentages, the time clearly increases, and when the defects reach more than 2% of the grid in the middle region, the signal is almost blocked.
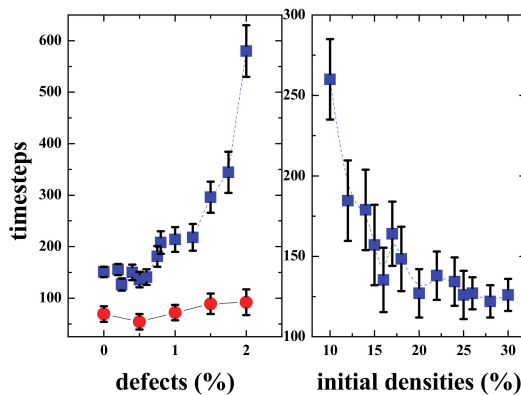
Figure 16. (Color online) Effect of the defects in the "transmission efficiency" of signals. Left figure shows the number of timesteps needed for the two communities to interact depending on the percentage of defects, for two different gaps, N/3, N being either 100 (blue squares) or 60 (red circles). (See videos 3 and 4). Right figure shows the time needed for the "transmission of the signal" between the regions "upper" and "lower," depending on the initial density (in%) of the regions.

It must be noted that the values presented in Figure 16 are obviously geometry-dependent. As already mentioned, the sample size is N×M (vertical by horizontal). If the gap (N/3 in our simulations, though it can be any other fraction of N) is very big compared to N, the two communities are unlikely to interact, except via gliders (or other "ships"), and when the gap is small the interaction will be fast. Note that, as shown in Figure 14, the gliders may be hindered by a single defect. If the size of the two communities (in terms of the fraction of N) is not very big, then it is more unlikely to create "traveling groups" that will end up connecting both sides.

Figure 17 shows a summary of the different properties of Life, depending on the percentage of defects in the grid. The equilibration time decays exponentially with the percentage of defects.
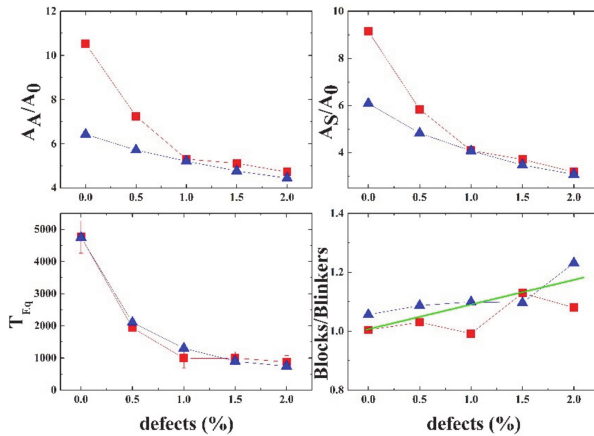
Figure 17. (Color on line) Dependence on the percentage of defects of several properties. For initial densities $\rho_i = 15\%$: Top Left: Active area, A_A, divided by the initial area A_0. Top Right: Spiked area A_S, divided by A_0. Bottom left: Equilibration time, bottom right, ratio #blocks/#blinkers. Straight line represents the linear fit of both results. Red squares and blue triangles are the results using open and closed boundaries, respectively.

In this part, we have established a second idea that deserves further study. This simple model with the addition of defects may be utilized to study neural network dysfunctions in the excitatory and inhibitory circuits that have been proposed as a mechanism in Alzheimer's disease [75–78].

6. Extensions of Game of Life

To conclude, as mentioned earlier, the number of nearest neighbor (NN) cells may be easily modified in the 2-dimensinal GoL. Counting second NN cells would take us to 24 NN, whilst counting third NN cells would take us to 48 NN and so on. However, that would make the model much more expensive computationally, and not only that, but the rules would have to be changed to have "Life," i.e., gliders and other interesting features observed in Conway's GoL [79,80]. Finding the correct rules for different neighborhoods was not trivial, but after some trial and error we found a possible interesting extension: B789/S69 (in the notation used by the Golly/RLE open-source cellular automaton package), where 789 is the interval of fertility and 69 is the interval of survival.

Figure 18 shows the most common clusters in the equilibrium state of the extended GoL B789/S69 CA. Another interesting "animal" found in this CA was the "orbiter" (See Video 5 in the supplementary material), a configuration that turns over itself in 10 steps. At this point, a similar analysis of density, activated area, etc., could be carried out. In Figure 19, the density and activated area versus time are presented. As can be seen, the qualitative behavior was similar but the timescales were different. Further analysis is possible, but it is beyond the scope of this study.
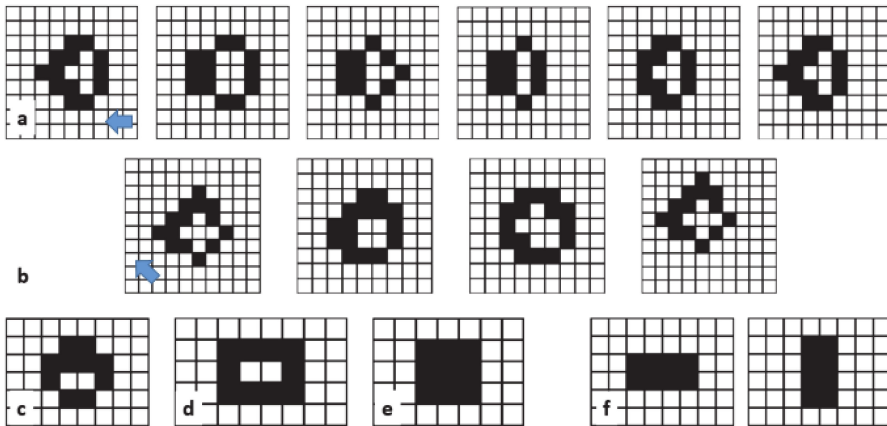
Figure 18. Most common clusters in the equilibrium state of the extended GoL B789/S69. (**a**) Orthogonal glider: Moves a cell forward in 5 steps, parallel or perpendicular to the grid. (**b**) Diagonal glider: Moves a cell in four steps. (**c**), (**d**) and (**e**), three different stable clusters. (**f**) Blinker. Period two.
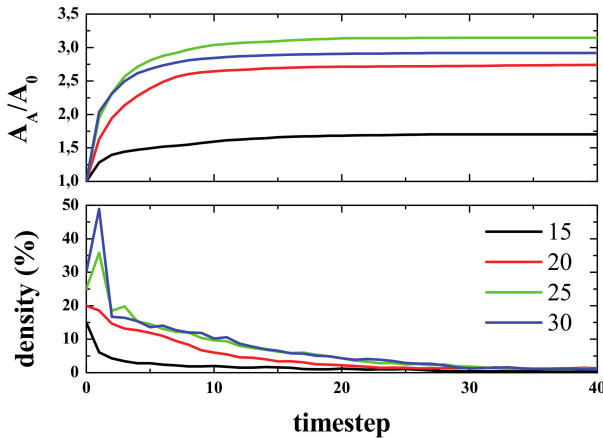
Figure 19. (Color online) Top panel: Activated area vs. time for different initial densities as labelled. Bottom panel: Density vs. time for different initial densities. Final density is approximately 0.9%. Second NN configuration, 69/789 rule applied.

We have shown that we can move from the original rules (B3/S23) to B789/S69 if we move from the first NN to the second NN, (i.e., from r = 1 to r = 2). Hence, it is natural to think that, for r = 3, r = 4, etc., it will be "always" possible to find a new GoL. However, would that model be a more accurate representation of some aspects of brain dynamics? This is a question that is not trivial to answer.

7. Conclusions

The GoL based model seems unrealistic at first glance, to be applied to the modeling of brain dynamics. However, there are some similarities with other well-known models. For instance, the stochastic rate model proposed by Benayoun et al [81] treats neurons as coupled, continuous-time, two-state Markov processes. Briefly, in that model, neurons are 2-state random processes. Each neuron

can exist in either the active state, representing a neuron firing an action potential and its accompanying refractory period, or a quiescent state, representing a neuron at rest [82].

The advantage of our approach lies in the fact that its simplicity makes it computationally cheap, so that one million neurons can be modelled in a single non-expensive computer. Even more, within our model, simulations containing several millions of neurons are feasible without the use of supercomputers. In comparison, as described in Reference [68], for simulations of a million neurons (because the degree of the node, k, in their simulations is k = 300), distributed computing techniques are needed. A model somewhat more realistic than ours, but not as complex as the one proposed by Acedo et al. [68], should be explored.

The GoL can be an interesting model to understand the many problems related to signal or activity propagation in brain dynamics. When the grid is filled with defects (cells that are always inactive), the transmission is clearly damped with the addition of a very small percentage of defects (2%). This could be related to neuronal damage and accumulation of amyloid-β plaques and neurofibrillary tangles, as in Alzheimer's disease [82]. Even a small percentage of damage impairs memory retrieval and other higher functions in patients and this is a feature of the disorder which could be studied with simple models such as cellular automata.

However, eight neighbors only in plane simulations (2D) are likely to be a crude oversimplification of the problem. A more realistic model would be a 3D Game of Life. Here more questions arise. Which 3D game of life? Several possible rules in 3D can be defined as "good life" [83]. In our opinion the GoL, both in 2D and 3D, deserves further attention, as well as the possibility of some kind of connection between the overall behavior of Life (2D or 3D) and brain dynamics. We believe that the GoL can play an important role in theoretical neuroscience. Clearly, besides being a simplification of the problem, it provides a theoretical framework on which more refined and realistic models can be built. In that direction, we have presented some simple analysis regarding the behavior of the GoL and how to use those ideas as tools to better understand the different aspects of brain dynamics, namely, sampling time, signal (or information) spreading or transmission, spikes, and defects. All these ideas can be combined or used in different scenarios (different topologies, for instance small world networks [50] or GoL in 3D [84]), or variants (synchronous–asynchronous) of the GoL [51–53] and combinations.

More interesting questions arise. Can we associate the final clusters (periodic or not, etc.) with some information storage? If so, periodic and stable (period 1) clusters correspond to short and long-term memory (or vice versa)? Recent studies [84] have challenged traditional theories of memory consolidation (i.e., the transfer of information to different parts of the brain for long term storage). They argue that memories form simultaneously in both the prefrontal cortex and the hippocampus (rather than the hippocampus only) and they gradually become stronger in the cortex and weaker in the hippocampus (rather than being transferred from the hippocampus to the neocortex, with only traces remaining in the hippocampus). Our model may be utilized to help determine how cortical and hippocampal regions communicate, to understand how the cortical memory cells maturation process occurs. Moreover, it may help to answer the question of whether memories fade completely from the hippocampal region or if traces remain to be occasionally retrieved.

Finally, another interesting experiment would be to check our results in much bigger samples, 100 times bigger, or more (10^8 or 10^9 neurons). As Anderson said "More is different" [85], so the results regarding the defects can be very size dependent.

Supplementary Materials: The following are available online at http://www.mdpi.com/2297-8747/23/4/75/s1, Video 1: Example of the GoL simulation with color representing activity along the simulation. The scale goes from 1 to 20. Video 2: The acorn time evolution. Number of timesteps is 5500. Color scale shows the number of spikes per cell. Video 3: Interaction of two communities. The color scale represents the number of times a cell has been active. Video 4: Interaction between two communities with defects in the grid. The cells created after the interaction are now colored in yellow. Video 5: The Orbiter; a stable configuration in the extended GoL.

References

1. Turing, A.M. Computing machinery and intelligence. *Mind* **1950**, *59*, 433–460. [CrossRef]
2. Sarkar, P. A brief history of cellular automata. *ACM Comput. Surv.* **2000**, *32*, 80–107. [CrossRef]
3. Ilachinski, A. *Cellular Automata*; World Scientific Publishing: Singapore, 2001.
4. Bard-Ermentrout, G.; Edelstein-Keshet, L. Cellular Automata Approaches to Biological Modeling. *J. Theor. Biol.* **1993**, *160*, 97–133. [CrossRef] [PubMed]
5. Boccara, N.; Roblin, O.; Roger, M. Automata network predator-prey model with pursuit and evasion. *Phys. Rev. E* **1994**, *50*, 4531. [CrossRef]
6. Gerhardt, M.; Schuster, H. A cellular automaton describing the formation of spatially ordered structures in chemical systems. *Phys. D* **1989**, *36*, 209–221. [CrossRef]
7. Zhu, M.F.; Lee, S.Y.; Hong, C.P. Modified cellular automaton model for the prediction of dendritic growth with melt convection. *Phys. Rev. E* **2004**, *69*, 061610. [CrossRef] [PubMed]
8. Bard Ermentrout, G.; Edelstein-Keshet, L. Cellular automata approaches to biological modeling. *J. Theor. Biol.* **1993**, *160*, 97–133. [CrossRef] [PubMed]
9. Kansal, A.R.; Torquato, S. Simulated Brain Tumor Growth Dynamics Using a Three-Dimensional Cellular Automaton. *J. Theor. Biol.* **2000**, *203*, 367–382. [CrossRef] [PubMed]
10. Hoffman, M.I. A Cellular Automaton Model Based on Cortical Physiology. *Complex Syst.* **1987**, *1*, 187–202.
11. Hopfield, J. Biophysics Neural networks and physical systems with emergent collective computational abilities. *PNAS* **1982**, *79*, 2554. [CrossRef] [PubMed]
12. Tsoutsouras, V.; Sirakoulis, G.C.; Pavlos, G.P.; Iliopoulos, A.C. Simulation of healthy and epileptiform brain activity using cellular automata. *Int. J. Bifurc. Chaos.* **2012**, *22*, 9. [CrossRef]
13. Acedo, L.; Lamprianidou, E.; Moraño, J.-A.; Villanueva-Oller, J.; Villanueva, R.-J. Firing patterns in a random network cellular automata model of the brain. *Phys. A* **2015**, *435*, 111–119. [CrossRef]
14. Wolfram, S. *A New Kind of Science*; Wolfram Media Inc.: Champaign, IL, USA, 2002.
15. Chialvo, D.R. Critical brain dynamics at large scale. In *Criticality in Neural Systems*; Niebur, E., Plenz, D., Schuster, H.G., Eds.; John Wiley: Hoboken, NJ, USA, 2014.
16. Chialvo, D.R. Emergent complex neural dynamics. *Nat. Phys.* **2010**, *6*, 744. [CrossRef]
17. Haimovici, A.; Tagliazucchi, A.E.; Balenzuela, P.; Chialvo, D.R. Brain organization into resting state networks emerges at criticality on a model of the human connectome. *Phys. Rev. Lett.* **2013**, *110*, 178101. [CrossRef] [PubMed]
18. Luković, M.; Vanni, F.; Svenkeson, A.; Grigolini, P. Transmission of information at criticality. *Phys. A* **2014**, *416*, 430–438. [CrossRef]
19. Priesemann, V.; Valderrama, M. Neuronal Avalanches Differ from Wakefulness to Deep Sleep–Evidence from Intracranial Depth Recordings in Humans. *PLoS Comput. Biol.* 2013. [CrossRef] [PubMed]
20. Priesemann, V.; Wibral, M.; Valderrama, M.; Pröpper, R.; Le Van Quyen, M.; Geisel, T.; Munk, M.H. Spike avalanches in vivo suggest a driven, slightly subcritical brain state. *Front. Syst. Neurosci.* **2014**, *8*, 108. [CrossRef] [PubMed]
21. Wibral, M.; Lizier, J.; Vögler, S.; Priesemann, V.; Galuske, R. Local active information storage as a tool to understand distributed neural information processing. *Front. Neuroinf.* **2014**, *8*, 1. [CrossRef] [PubMed]
22. Langton, C.G. Computation at the edge of chaos: Phase transitions and emergent computation. *Phys. D* **1990**, *42*, 12–37. [CrossRef]

23. Beggs, J.M.; Timme, N. Being critical of criticality in the brain. *Front. Physiol.* **2012**, *3*, 163. [CrossRef] [PubMed]

24. Friedman, N.; Ito, S.; Brinkman, B.A.; Shimono, M.; DeVille, R.L.; Dahmen, K.A.; Butler, T.C. Universal Critical Dynamics in High Resolution Neuronal Avalanche Data. *Phys. Rev. Lett.* **2012**, *108*, 208102. [CrossRef] [PubMed]

25. Kello, C.T. Critical Branching Neural Networks. *Psychol. Rev.* **2013**, *120*, 230–254. [CrossRef] [PubMed]

26. Werner, G. Metastability, Criticality and Phase Transitions in brain and its models. *Biosystems* **2007**, *90*, 496–508. [CrossRef] [PubMed]

27. Bak, P.; Chen, K.; Creutz, M. Self-organized criticality in the 'Game of Life'. *Nature* **1989**, *342*, 780. [CrossRef]

28. Hemmingsson, J. Consistent results on 'Life'. *Phys. D* **1995**, *80*, 151. [CrossRef]

29. Nordfalk, J.; Alstrøm, P. Phase transitions near the "game of Life". *Phys. Rev. E* **1996**, *54*, R1025. [CrossRef]

30. Ninagawa, S.; Yoneda, M.; Hirose, S. 1/f fluctuation in the 'Game of Life'. *Phys. D* **1998**, *118*, 49–52. [CrossRef]

31. Allegrini, P.; Menicucci, D.; Bedini, R.; Fronzoni, L.; Gemignani, A.; Grigolini, P.; Paradisi, P. Spontaneous brain activity as a source of ideal 1/f noise. *Phys. Rev. E* **2009**, *80*, 061914. [CrossRef] [PubMed]

32. Fox, M.D.; Raichle, M.E. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat. Rev. Neurosci.* **2007**, *8*, 700–711. [CrossRef] [PubMed]

33. Linkenkaer-Hansen, K.; Nikouline, V.V.; Palva, J.M.; Ilmoniemi, R.J. Long-range temporal correlations and scaling behavior in human brain oscillations. *J. Neurosci.* **2001**, *21*, 1370–1377. [CrossRef] [PubMed]

34. Gilden, D.L.; Thornton, T.; Mallon, M.W. 1/f Noise in Human Cognition. *Science* **1995**, *267*, 1837–1839. [CrossRef] [PubMed]

35. Bédard, C.; Kröger, H.; Destexhe, A. Does the 1/f Frequency Scaling of Brain Signals Reflect Self-Organized Critical States? *Phys. Rev. Lett.* **2006**, *97*, 118102. [CrossRef] [PubMed]

36. Wolfram, S. Statistical mechanics of cellular automata. *Rev. Mod. Phys.* **1983**, *55*, 601. [CrossRef]

37. Chapman, P. "Life Universal Computer". Available online: http://www.igblan.free-online.co.uk/igblan/ca/ (accessed on 15 November 2018).

38. Berlekamp, E.R.; Conway, J.H.; Guy, R.K. *Winning Ways for your Mathematical Plays*, 2nd ed.; A. K. Peters Ltd.: Natick, MA, USA, 2004.

39. Rendell, P. Turing universality of the game of life. In *Collision-Based Computation*; Adamatzky, A., Ed.; Springer: London, UK, 2002; p. 513.

40. Rennard, J.P. Implementation of logic functions in the game of life. In *Collision-Based Computation*; Adamatzky, A., Ed.; Springer: London, UK, 2002; p. 491.

41. Koslow, S.H.; Huerta, M.F. *Neuroinformatics: An Overview of the Human Brain Project*; Lawrence Erlbaum Associates, Inc.: Mahwah, NJ, USA, 1997.

42. Bagnoli, F.; Rechtman, R.; Ruffo, S. Some facts of life. *Phys. A* **1991**, *171*, 249–264. [CrossRef]

43. Buckingham, D.J. Some facts of life. *Byte* **1978**, *3*, 54.

44. Monetti, R.A.; Albano, E.V. Critical edge between frozen extinction and chaotic life. *Phys. Rev. E* **1995**, *52*, 5825. [CrossRef]

45. Schulman, L.S.; Seiden, P.E. Statistical Mechanics of a Dynamical System Based on Conway's Game of Life. *J. Stat. Phys.* **1978**, *19*, 3. [CrossRef]

46. Garcia, J.B.C.; Gomes, M.A.F.; Jyh, T.I.; Ren, T.I.; Sales, T.R.M. Nonlinear dynamics of the cellular-automaton "game of Life". *Phys. Rev. E* **1993**, *48*, 3345. [CrossRef]

47. Kayama, Y. Complex networks derived from cellular automata. *arXiv*, 2010; arXiv:1009.4509.

48. Kayama, Y. Network representation of cellular automata. In Proceedings of the 2011 IEEE Symposium on Artificial Life, Paris, France, 11–15 April 2011; pp. 194–202.

49. Kayama, Y.; Imamura, Y. Network representation of the game of life. *J. Artif. Intell. Soft Comput. Res.* **2011**, *1*, 233–240.

50. Huang, S.-Y.; Zou, X.-W.; Tan, Z.J.; Jin, Z.Z. Network-induced nonequilibrium phase transition in the "game of Life". *Phys. Rev. E* **2003**, *67*, 026107. [CrossRef] [PubMed]

51. Fates, N.; Morvan, M. Perturbing the Topology of the Game of Life Increases Its Robustness to Asynchrony. In Proceedings of the International Conference on Cellular Automata, Amsterdam, The Netherlands, 25–28 October 2004; pp. 111–120.

52. Lee, J.; Adachia, S.; Peper, F.; Morita, K. Asynchronous game of life. *Phys. D* **2004**, *194*, 369–384. [CrossRef]

53. Blok, H.J.; Bergersen, B. Synchronous versus asynchronous updating in the "game of Life". *Phys. Rev. E* **1999**, *59*, 3876–3879. [CrossRef]

54. Schönfisch, B.; de Roos, A. Synchronous and asynchronous updating in cellular automata. *BioSystems* **1999**, *51*, 123–143. [CrossRef]

55. Reia, S.M.; Kinouchi, O. Conway's game of life is a near-critical metastable state in the multiverse of cellular automata. *Phys. Rev. E* **2014**, *89*, 052123. [CrossRef] [PubMed]

56. De la Torre, A.C.; Mártin, H.O. A survey of cellular automata like the "game of life". *Phys. A* **1997**, *240*, 560–570. [CrossRef]

57. Beer, R.D. Autopoiesis and Cognition in the Game of Life. *Artif. Life* **2004**, *10*, 309–326. [CrossRef] [PubMed]

58. Beer, R.D. The Cognitive Domain of a Glider in the Game of Life. *Artif. Life* **2014**, *20*, 183–206. [CrossRef] [PubMed]

59. Braitenberg, V.; Schóz, A. *Statistics and Geometry of Neuronal Connectivity*; Springer-Verlag: Berlin, Germany, 1998.

60. Yuste, S.B.; Acedo, L. Number of distinct sites visited by N random walkers on a Euclidean lattice. *Phys. Rev. E* **2000**, *61*, 2340. [CrossRef]

61. Lachaux, J.P.; Pezard, L.; Garnero, L.; Pelte, C.; Renault, B.; Varela, F.J.; Martinerie, J. Spatial extension of brain activity fools the single-channel reconstruction of EEG dynamics. *Hum. Brain Mapp.* **1997**, *5*, 26–47. [CrossRef]

62. McDowell, J.E.; Kissler, J.M.; Berg, P.; Dyckman, K.A.; Gao, Y.; Rockstroh, B.; Clementz, B.A. Electroencephalography/magnetoencephalography study of cortical activities preceding prosaccades and antisaccades. *Neuroreport* **2005**, *16*, 663–668. [CrossRef] [PubMed]

63. Holsheimer, J.; Feenstra, B.W.A. Volume conduction and EEG measurements within the brain: A quantitative approach to the influence of electrical spread on the linear relationship of activity measured at different locations. *Electroencephalogr. Clin. Neurophysiol.* **1977**, *43*, 52–58. [CrossRef]

64. Hodgkin, A.L.; Huxley, A.F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* **1952**, *117*, 500–544. [CrossRef] [PubMed]

65. Leao, A.P.P. Pial circulation and spreading depression of activity in the cerebral cortex. *J. Neurophysiol.* **1944**, *7*, 391–396. [CrossRef]

66. Porooshani, H.; Porooshani, G.H.; Gannon, L.; Kyle, G.M. Speed of progression of migrainous visual aura measured by sequential field assessment. *Neuro-Ophthalmology* **2004**, *28*, 101–105. [CrossRef]

67. Ayata, C.; Lauritzen, M. Spreading Depression, Spreading Depolarizations, and the Cerebral Vasculature. *Physiol. Rev.* **2015**, *95*, 953–993. [CrossRef] [PubMed]

68. Acedo, L.; Moraño, J.-A. Brain oscillations in a random neural network. *Math. Comput. Model.* **2013**, *57*, 1768–1772. [CrossRef]

69. Hutsler, J.J. The specialized structure of human language cortex: Pyramidal cell size asymmetries within auditory and language-associated regions of the temporal lobes. *Brain Lang.* **2003**, *86*, 226–242. [CrossRef]

70. Shusterman, V.; Troy, W.C. From baseline to epileptiform activity: A path to synchronized rhythmicity in large-scale neural networks. *Phys. Rev. E* **2008**, *77*, 061911. [CrossRef] [PubMed]

71. Towle, V.L.; Ahmad, F.; Kohrman, M.; Hecox, K.; Chkhenkeli, S. Electrocorticographic Coherence Patterns of Epileptic Seizures. In *Epilepsy as a Dynamic Disease*; Milton, J., Jung, P., Eds.; Springer: Berlin/Heidelberg, Germany, 2003.

72. Wilson, H.R.; Cowan, J.D. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.* **1972**, *12*, 1–24. [CrossRef]

73. Wilson, H.R.; Cowan, J.D. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik* **1973**, *13*, 55–80. [CrossRef] [PubMed]

74. Conway's Game of Life. Examples of patterns. Available online: https://en.wikipedia.org/wiki/Conway%27s_Game_of_Life#Examples_of_patterns (accessed on 15 November 2018).

75. Wass, S. Distortions and disconnections: Disrupted brain connectivity in autism. *Brain Cognit.* **2011**, *75*, 18–28. [CrossRef] [PubMed]

76. Zhang, H.-Y.; Wang, S.-J.; Liu, B.; Ma, Z.L.; Yang, M.; Zhang, Z.J.; Teng, G.J. Resting Brain Connectivity: Changes during the Progress of Alzheimer Disease. *Radiology* **2010**, *256*, 2. [CrossRef] [PubMed]

77. Supekar, K.; Menon, V.; Rubin, D.; Musen, M.; Greicius, M.D. Network Analysis of Intrinsic Functional Brain Connectivity in Alzheimer's Disease. *PLoS Comput. Biol.* **2011**, *4*, e1000100. [CrossRef] [PubMed]

78. Vissers, M.E.; Cohen, M.X.; Geurts, H.M. Brain connectivity and high functioning autism: A promising path of research that needs refined models, methodological convergence, and stronger behavioral links. *Neurosci. Biobehav. Rev.* **2012**, *36*, 604–625. [CrossRef] [PubMed]
79. Gardner, M. The fantastic combinations of John Conway's new solitaire game 'life'. *Sci. Am.* **1970**, *223*, 120–123. [CrossRef]
80. Packard, N.H.; Wolfram, S. Two-Dimensional Cellular Automata. *J. Stat. Phys.* **1985**, *38*, 901–946. [CrossRef]
81. Benayoun, M.; Cowan, J.D.; van Drongelen, W.; Wallace, E. Avalanches in a stochastic model of spiking neurons. *PLoS Comput. Biol.* **2010**, *6*, e1000846. [CrossRef] [PubMed]
82. Nunomura, A.; Perry, G.; Aliev, G.; Hirai, K.; Takeda, A.; Balraj, E.K.; Jones, P.K.; Ghanbari, H.; Wataya, T.; Shimohama, S.; et al. Oxidative damage is the earliest event of Alzheimer's disease. *J. Neuropathol. Exp. Neurol.* **2001**, *6*, 759–767. [CrossRef]
83. Bays, C. Candidates for the Game of Life in Three Dimensions. *Complex Syst.* **1987**, *1*, 373–400.
84. Kitamura, T.; Ogawa, S.K.; Roy, D.S.; Okuyama, T.; Morrissey, M.D.; Smith, L.M.; Redondo, R.L.; Tonegawa, S. Engrams and circuits crucial for systems consolidation of a memory. *Science* **2017**, *356*, 73–78. [CrossRef] [PubMed]
85. Anderson, P.W. More is Different. *Science* **1972**, *177*, 393–396. [CrossRef] [PubMed]

Article

Managing Human Factors to Reduce Organisational Risk in Industry

Silvia Carpitella [1,2,*], Fortunato Carpitella [3], Antonella Certa [1], Julio Benítez [2] and Joaquín Izquierdo [2]

1   Dipartimento dell'Innovazione Industriale e Digitale (DIID), Università degli Studi di Palermo, 90133 Palermo, Italy; antonella.certa@unipa.it
2   Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València, 46022 Valencia, Spain; jbenitez@mat.upv.es (J.B.); jizquier@upv.es (J.I.)
3   Studio di Ingegneria Carpitella, 91100 Trapani, Italy; fortunato.carpitella@gmail.com
*   Correspondence: silvia.carpitella@unipa.it

Abstract: Human factors are intrinsically involved at virtually any level of most industrial/business activities, and may be responsible for several accidents and incidents, if not correctly identified and managed. Focusing on the significance of human behaviour in industry, this article proposes a multi-criteria decision-making (MCDM)-based approach to support organizational risk assessment in industrial environments. The decision-making trial and evaluation laboratory (DEMATEL) method is proposed as a mathematical framework to evaluate mutual relationships within a set of human factors involved in industrial processes, with the aim of highlighting priorities of intervention. A case study related to a manufacturing process of a real-world winery is presented, and the proposed approach is applied to rank human factors resulting from a previous organisational risk evaluation from which suitable inference engines may be developed to better support risk management.

Keywords: human behaviour; organisational risk; multi-criteria decision-making; DEMATEL; bottling process

---

1. Introduction

Companies are managed by following previously designed strategies, and operate according to processes implemented on the basis of the available resources. These strategies and processes are complex systems that integrate workers, plants and environment. Balancing and mutually adapting these elements make it possible, among others, to implement actions aimed at preventing the occurrence of accidents and occupational disease within workplaces, and also to identify near misses. The concept of human management system (HMS) is important to this issue.

Clerici et al. [1] affirm that an organization is a plurality of "human elements", and risks often depends on organizational criticalities, whose reduction can be undertaken by implementing effective human resource management (HRM). In particular, HRM is defined as a system of structured procedures aimed at optimizing the manpower management in a company [2], with its workers being the most valuable assets of the organisation [3]. As asserted by Cirjaliu and Draghici [4], nowadays companies seek to continuously improve the well-being and satisfaction of their human resources within their own operational environments.

An important aspect to take into account within this context is the integrations of human factors and ergonomics (HF/E), whose optimal management is crucial to achieve central objectives, for instance the transition to sustainable development [5,6].

Indeed, human factors are intrinsically involved at virtually any level of most industrial/business activities [7,8]. They represent the core component of many organisations and may be responsible for

several accidents and incidents if not correctly identified and managed, as asserted by Ergai et al. [9]. However, as authors underline, investigating these aspects depends on the specific features of the workplace of reference, and on the evaluator's background. For this reason, promoting a safe and environmentally responsible manner of working represents one of the most important organisational challenges, currently [10].

The importance of this concept is broadly shared in the literature. Wilson [11] asserts that any understanding of system ergonomics must be related to the idea of system engineering. Hassall et al. [12] stress that analyses based on human factors and ergonomics are commonly used to improve safety and productivity—particularly in complex systems. Sobhani et al. [13] underline that the improvement of workplace ergonomic conditions gives opportunities to better deal with production variations and optimize the performance of system operation.

To address the aspects related to HF/E within industrial workplaces, the decision-making trial and evaluation laboratory (DEMATEL) method, first developed by Fontela and Gabus [14,15], is herein suggested as a mathematical framework to evaluate mutual relationships among some of the most important human factors involved in industrial processes—which, as in many other areas, have usually shown to be deeply intertwined and mutually affected.

Amid various multi-criteria decision-making (MCDM) methods proposed in the literature, DEMATEL is particularly helpful to take into account existing interdependence among the main elements involved in any complex decision-making problem, on the basis of judgments attributed by a team of experts. DEMATEL also finds the central criteria to represent the effectiveness of factors/aspects, and avoids evaluation overfitting. This interdependence is eventually represented by means of a graphical chart, from which causes, and effects, are suitably described.

The present paper is organised as follows. Section 2 discusses human factors and ergonomics in industrial environments. The main investigated areas when leading organisational risk assessment are presented and described. Section 3 deploys the DEMATEL framework, with its various methodological steps, and the obtained outputs. Section 4 presents a real-world case study of a manufacturing process led in a winery: The bottling process. The most critical human factors, emerged from a previous implementation of risk assessment [16], are ranked by means of the DEMATEL to suggest an order of intervention aimed at gradually reducing organisational risk in the analysed operational context. Lastly, conclusions are provided in Section 5 to close the paper.

2. Human Factors and Ergonomics in Industry

Amount and intensity of human interactions with industry processes generally depend on the field in which the organisation operates. Carpitella et al. [17] presents a literature review in this regard, which is herein extended. Saravia-Pinilla et al. [18] analyse the strong bond existing among environmental and human factors. In particular, the authors highlight a gap in the existing literature about this topic, and propose a model combining human and environmental factors with relation to the processes of product/service design and an ad hoc development to potentiate decision-making processes.

A tool that is particularly effective in conducting human factor-based analyses for reducing accidents and incidents is represented by the human factor analysis and classification system (HFACS), developed by Wiegmann and Shappell [19]. This can be applied in a wide variety of contexts, such as, for instance, aviation industry [20] or maritime safety [21]. Chen et al. [22] focus on marine casualties and incidents and deal with human factor management with the aim of reducing accidents and avoiding disasters. The authors implement the framework HFACS for maritime accidents (HFACS-MA), a useful support to increase the level of safety and reduce human errors by identifying possible accident causes. Madigan et al. [23] refer to the rail industry and stress the importance of carefully taking into account also latent factors. They propose HFACS by accomplishing a retrospective analysis to examine causes of minor incidents to prevent future and more severe events.

It is neither possible, nor convenient, to totally eliminating human contribution to processes, even when a high degree of automation is pursued, such as in manufacturing industries [24]. Industries

with high production volumes may consider machines and computers as faster and more reliable than humans in leading automatic operations. In this case, the human contribution given to automated processes would be barely necessary, and this may help reduce possible errors, due to psychological and physical factors, such as health, stress, age, mood, and so on. Moreover, the more customised the manufacturing process, the more crucial the role of human factors.

Furthermore, the aspect of dependence among various phases of a process has to be considered and managed. This kind of dependence strongly impacts on the reliability level, as asserted by Zio et al. [25]. Indeed, considering, for instance, a sequence of two interdependent tasks, a fault on one of them increases the probability of failing on the other. The authors propose a framework based on a fuzzy system for eliciting expert knowledge about those factors mostly influencing dependence between two successive tasks. In particular, relationships between the input factors and the conditional human error probability are represented by means of a set of transparent fuzzy logic rules, and an application, related to two tasks required in response to an accident scenario at a nuclear power plant, is analysed.

Therefore, a current challenge faced by organisations consists in integrating even more machines and workers [26], with the aim of creating a systematic operational environment and optimising all the available resources. In this context, human reliability strongly influences organisations' outcomes, and plays an important role in evaluating risks related to industrial/business activities.

Human reliability refers to the field of human factors and ergonomics, and is defined as the probability to successfully carry out a general human activity [27]. Human performance, as expressed before, can be influenced by such factors as age, physical and psychological health, attitude, etc. For this reason, human reliability is assessed with the aim of supporting risk evaluation and, in particular, of determining the impact of human contribution to the risk of failure or success, especially when humans are directly responsible for system operation, as it is usual today.

Since diverse factors are involved within the operation of systems or processes under analysis, a multidisciplinary approach is necessary to prevent the possible occurrence of human errors. Generally, human errors [28,29] are classified into errors of commission (EOC) and errors of omission (EOO). The first group is related to errors during the phases of identification, interpretation and execution of a specific activity, whereas the second category regards errors, due to forgetfulness or inattention, omitting a step of the task or also the whole task itself.

Triggers for faults are likely represented by human factors, the evolution of the error probability can be understood and approached by modelling human behaviour. Human behaviour can be schematised according to various levels, classified as skill-based, rule-based and knowledge-based [30]. By transiting from the first level to the third, the human error probability (HEP) increases. Specific errors are associated to each kind of behaviour [31]. In particular, the main causes of errors related to skill-based behaviour are the lack of concentration and the presence of stressful situations. Concerning the rule-based level of behaviour, errors derive from wrong approaches to procedures and rules. Lastly, regarding the knowledge-based behaviour, errors are caused by incorrect interpretations of specific situations or also by incomplete knowledge.

On the basis of all the above, organisational risk assessment in industrial environments is conducted with the aim of evaluating, eliminating or at least minimising risks related to ineffective manners of work, in terms of methods and operation management from humans. Such kind of risks derives from psychological and physical conditions that negatively impact on the quality of work and life.

In particular, when leading organisational risk assessment, the main areas presented in Table 1 are analysed with a deep level of detail. The purpose consists in highlighting the presence of possible stressful aspects related to human factors and ergonomics within each area, which could potentially damage the global wellness and health of workers, and therefore, the performance of the whole organisation.

These factors are present in almost all the working environments. Among all organisational aspects, the European agreement on work-related stress held in Brussels in the year 2004 [32] underlines

as managing the problem of stress at work leads to greater efficiency and improvement of health and safety conditions, with consequent economic and social benefits for companies, workers and society. For this reason, the same agreement established to offer models and guidelines for evaluating work-related stress on the basis of two phases, namely preliminary assessment and in-depth evaluation. The first phase is based on the identification of verifiable and quantitative stress indicators. The second phase should be undertaken through surveys, focussing on groups and semi-structured interviews to homogeneous groups of workers.

Table 1. Description of investigated areas related to human factors and ergonomics.

ID	Investigated Area	Description
A₁	Organizational culture and role	Sharing values, upon which the organisation policy is grounded; maintaining relationships among different levels of the same organization; being aware about the own role within the company.
A₂	Career development and job stability	Having clear the possibilities of development in terms of career advances; knowing the path of professional growing; achieving contractual stability.
A₃	Communication, information, consultation and participation of workers	Empowering communication among all the levels of the hierarchy structure of the company; involving workers within decision-making processes to pursue general business objectives.
A₄	Training, awareness and competence	Promoting training paths aimed at increasing specific competencies of workers and at continuously improving the level of safety and security related to industrial processes.
A₅	Operational control: Indication of measures and instruments	Defining scheme, minimum contents and work procedures to lead a safe execution of the main tasks; identifying the main criticalities to be monitored; monitoring and controlling processes and outputs; planning and implementing maintenance interventions on the basis of the policies undertaken by the organisation.
A₆	Extraordinary situations and changes management	Defining criteria, methods and responsibilities to identify possible scenarios of extraordinary situations causing exceptional or unusual results; establishing intervention measures; managing changes to implement corrective measures.
A₇	Outsourcing and interference management	Evaluating direct and indirect impacts of the outsourcing process; implementing a framework of cooperation with external companies to optimise safety both of internal workers and third parties.
A₈	Workload and working hours	Examining the entity of workload; balancing responsibilities related to each group of workers; managing and correctly planning the number of working hours per person; integrating work with life and social contexts of workers.

By analysing the results coming from such evaluations, we propose to focus on the more critical human factors emerged for each target area (Table 1). With this aim, the DEMATEL methodology is suggested to select, within the set of highlighted human factors, those most influencing the others. This approach is useful to suggest an order in planning and implementing mitigation measures of organisational risk.

3. DEMATEL to Increase the Level of Safety in Industrial Processes

In complex systems, many aspects, factor or criteria are, either directly or indirectly, deeply intertwined (sometimes in a hidden way), and mutual interference affects other elements, thus making it difficult to find priorities for action and eventually hindering decision-making. In many

cases, pursuing a specific objective may inadvertently impair several other objectives. Therefore, having a clear vision of the system contributes to the identification of workable solutions. DEMATEL has shown to help confirm interdependence among variables and restrict the relation that reflects the characteristics of a system of management trend [33–35]. DEMATEL's outcome is a visual representation, through which decision-makers may organize better the actions to take. The purpose of the use of DEMATEL in this paper is to discern the direction and intensity of direct and indirect relationships that flow among a number of well-defined elements. Thus, experts' knowledge is used to contribute to better understand the problem components and the way they interrelate.

The DEMATEL technique can be implemented by means of seven steps, in sequence.

1. The problem under analysis has to be clearly expressed in terms of a general goal. The main elements/factors characterising the problem have to be defined by means of the support of a decision-making team composed of experts in the field.
2. Non-negative matrices $X^{(k)}$ have to be produced, where $1 \leq k \leq H$, H being the number of experts, expressing judgments on the mutual influence between pairs of elements. Elements $x_{ij}^{(k)}$ $i,j = 1,\ldots,n$ (n being the number of compared elements) represent the numerical values encoding the judgments. The meanings of those numerical values are defined as follows: 0 (no influence), 1 (very low influence), 2 (low influence), 3 (high influence), 4 (very high influence). The main diagonal values of any of these matrices are zero.
3. The direct-relation matrix A has to be built. This matrix incorporates all the matrices previously filled in by the involved experts. A is a square matrix of order n that averages the opinions of the group of experts:

$$A = \frac{1}{H} \sum_{k=1}^{H} X^{(k)}. \tag{1}$$

4. The normalized direct-relation matrix has to be obtained. From (1), this matrix is calculated as

$$D = sA, \tag{2}$$

where s is a positive number slightly smaller than

$$\min\left[\frac{1}{\max\limits_{1 \leq i \leq n} \sum_{j=1}^{n} a_{ij}}, \frac{1}{\max\limits_{1 \leq j \leq n} \sum_{i=1}^{n} a_{ij}} \right]. \tag{3}$$

Based on matrix D, the initial influence that elements exert on and receive from the others is shown. Then, a continuous decrease of the indirect effects among the considered elements may be obtained along the consecutive powers of matrix D. This enables to obtain the total relation matrix, as explained next.

5. The total relation matrix T has to be calculated. This matrix reflects both direct and indirect effects among elements, and is achievable through the sum of the powers of matrix D. Observe that $\lim\limits_{n \to \infty} D^n = 0$, since the spectral radius of D is smaller than 1, since, by Equation (3), it is bounded by the maximum row and column sum. As a result, see, for example, Example 7.3.1 in [36], the power series of D, $I + D + D^2 + \cdots$, converges to $(I - D)^{-1}$ where I is the identity matrix of size n. Consequently, the total relation matrix may be written as

$$T = D(I - D)^{-1}. \tag{4}$$

As said, this matrix represents the build-up of mutual direct and indirect effects among elements. Observe that the diagonal entries of matrix D (accounting for the direct effects) are zero; however, the diagonal elements of T collect all the non-direct effects associated to their corresponding factors. This fact is crucial in step 7.

6. An influential relation map is obtained through the definition of $\mathbf{r} = (r_i)$ and $\mathbf{c} = (c_j)$ as $n \times 1$ and $1 \times n$ vectors, respectively representing the sum of the rows and the sum of the columns of the total relation matrix T. Particularly, r_i represents both direct and indirect effects of element i on the others, whereas c_i summarizes both direct and indirect effects of the other elements on element i. In such a way, the sum $r_i + c_i$ gives the overall effect (prominence) of element i, and the subtraction $r_i - c_i$ helps in dividing the elements into cause and effect groups (relation). Prominence allows to rank factors according to their global influence, while relation enables to group elements into the cause group—if the subtraction is positive—, and into the effect group—otherwise.

7. Prominence ranking gives crucial information on the impact associated to the factors. However, a cutoff on the factor list is performed through a suitable threshold, bearing in mind that if the threshold is too high important factors may be excluded and if it is too low, too many factors—some of them irrelevant—may be included, which will turn the solution too complex and thus impractical. In the literature, the threshold value is determined in a variety of ways: By experts through discussions [37,38] or brainstorming techniques [39], by following results of literature review, the maximum mean de-entropy (MMDE) [40], the average of all elements in the matrix T [41], among others. In this paper we use this last value. Finally, a causal diagram chart is drawn by mapping the dataset of $(r_i + c_i, r_i - c_i)$, which gives a graphical representation of the main interrelations among factors. Typically, only the interrelations among factors considered within the cutoff are drawn, for the sake of clarity.

The main goal of the DEMATEL application in the present paper consists in identifying key factors based on the causal relationships and the degrees of interrelationship between them, with the aim of providing companies with a structured way of understanding the nature of interdependencies within a set of human factors. As previously asserted, the definition of human factors results from a previous context evaluation carried out in terms of an organizational risk analysis. In other terms, we aim to identify aspects influencing the others and aspects being influenced by others for pursuing a higher level of safety and security in leading industrial processes. To demonstrate the usefulness of our approach, a real-world case study is developed to evaluate interdependencies among critical human factors analysed in a manufacturing process of a Sicilian firm with the aim of reducing organizational criticalities.

4. Real-World Case Study of a Sicilian Winery

The case study refers to a manufacturing firm, a winery located in Trapani, Sicily (Italy). We aim to focus on the wine bottling process carried out in the company. This process is composed of 13 different phases, provided in Figure 1, and takes place in the area dedicated to delivery and production. In the mentioned area, there are three fixed stations and a movable position, respectively occupied by the following operators:

1. W_1, worker dedicated to control that bottles are filled in and plugged;
2. W_2, worker dedicated to control the global quality of bottles;
3. W_3, worker dedicated to wrap final products;
4. W_4, worker dedicated to carry out the following two activities: Raw materials (empty bottles, labels and corks) and packaging supply; handling of wrapped final products.
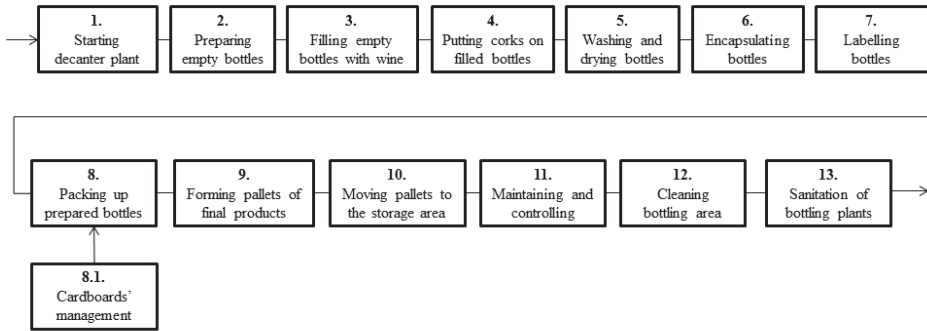
Figure 1. Phases of the bottling process.

The scheme of the production line representing the bottling process is shown in Figure 2. The stations indicated as "DP" and "P", respectively, represents the point in which empty bottles are first taken off from pallets (in which they were originally stocked) for starting the bottling process, and the point in which bottles (after having been filled in, plugged and checked) are finally put in pallets and wrapped to be sent to the storage or final customer areas.
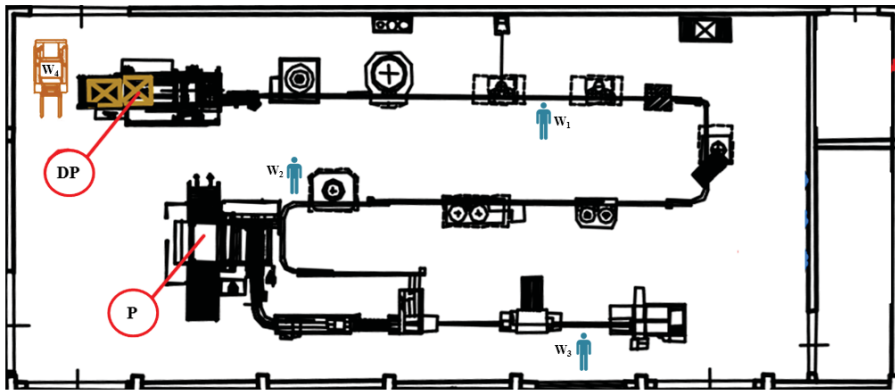
Figure 2. Scheme of the production line representing the bottling process.

With relation to the described process, the firm recently undertook an organisational risk assessment by focusing on the group of workers distributed in the interested zone. In particular, the work-related stress was evaluated by adopting the guidelines provided in 2011 by the National (Italian) Institute for Insurance against Accidents at Work [16]. Within that evaluation, the areas of Table 1 were deeply investigated by means of detailed surveys with the workers. These surveys aimed at highlighting the possible presence of critical human factors for each area, with the final purpose of managing critical aspects and then reducing the organisational risk as much as possible. In particular, the 16 human factors in Table 2 (listed with relation to their related area) emerged as possible sources of problems. The application of the DEMATEL methodology is suggested for establishing an order for implementing mitigating measures.

Table 2. Critical human factors related to each area.

ID	Investigated Area	Description
A_1	Organisational culture and role	• **HF$_1$** System of security and safety management not implemented; • **HF$_2$** Ethical and behavioural code not implemented.
A_2	Career development and job stability	• **HF$_3$** Criteria for career advancement are not defined; • **HF$_4$** Reward systems related to the correct management of human resources are not defined for supervisors; • **HF$_5$** Reward systems related to the achievement of security objectives are not defined.
A_3	Communication, information, consultation and participation of workers	• **HF$_6$** Work may depend on tasks previously accomplished by others; • **HF$_7$** Tools involving workers within decisions and strategies are not implemented; • **HF$_8$** Rigid protocols supervising and controlling workers are implemented.
A_5	Operational control: Indication of measures and instruments	• **HF$_9$** Workers are exposed to noise between the I and the II levels of action; • **HF$_{10}$** Inadequate ventilation and microclimate; • **HF$_{11}$** Inadequate lighting; • **HF$_{12}$** Workers may be exposed to the risk of recurring movements.
A_8	Workload and working hours	• **HF$_{13}$** Unpredictably variations of workload; • **HF$_{14}$** Workers cannot regulate machines' rhythm; • **HF$_{15}$** Workers lead tasks having high level of responsibility for stakeholders, plants and production; • **HF$_{16}$** Shifts may be not well organised.

We apply now the DEMATEL to evaluate existing interdependencies within the set of $n = 16$ human factors detailed in Table 2. Five experts in the field ($H = 5$) were involved to such an aim, whose roles are given in Table 3.

Table 3. Roles of the decision makers.

Decision Maker	Role
H_1	Maintenance responsible
H_2	Quality manager
H_3	Consultant
H_4	Chief of the safety and security system
H_5	Department chief

The experts composing the decision-making group contribute to the process development by playing diverse, but complementary, roles. Indeed, these subjects have been involved with the aim of guaranteeing as complete as possible understanding about the problem under analysis.

Each decision-maker was asked to evaluate the direct influence between any two human factors by means of integer scores from 0 to 4. Five non-negative square matrices $X^{(1)}$, $X^{(2)}$, $X^{(3)}$, $X^{(4)}$ and $X^{(5)}$ (given in the Appendix A) were collected and then aggregated to obtain the direct-relation matrix A of order 16 (Table 4).

Tables 5 and 6 respectively show the normalized direct-relation matrix D and the total relation matrix T. Lastly, Table 7 shows the values of $r_i + c_i$ and $r_i - c_i$ associated to the various factors, and the ranking of factors, obtained on the basis of their prominence, $r_i + c_i$, which collects the direct and indirect effects related to all the other factors.

Table 4. Direct-relation matrix *A*.

A	HF₁	HF₂	HF₃	HF₄	HF₅	HF₆	HF₇	HF₈	HF₉	HF₁₀	HF₁₁	HF₁₂	HF₁₃	HF₁₄	HF₁₅	HF₁₆
HF₁	0.000	3.200	1.800	1.600	4.000	2.000	2.400	2.200	4.000	3.600	4.000	4.000	1.800	2.800	4.000	2.200
HF₂	2.200	0.000	2.800	3.000	2.600	0.000	0.000	3.200	1.200	1.600	1.600	1.200	0.000	0.000	2.600	1.400
HF₃	2.200	2.000	0.000	4.000	3.000	3.000	3.000	3.000	0.000	0.000	0.000	0.000	0.000	0.000	1.800	1.600
HF₄	2.000	3.200	3.200	0.000	4.000	1.200	4.000	3.200	0.000	0.000	0.000	0.000	2.400	1.200	3.600	3.600
HF₅	3.600	2.000	3.200	4.000	0.000	2.800	2.800	2.800	2.400	2.800	2.400	2.800	1.200	4.000	4.000	2.400
HF₆	2.400	1.000	2.200	2.400	3.200	0.000	2.000	3.400	1.400	1.400	1.400	1.400	4.000	2.800	3.000	3.800
HF₇	1.400	0.000	2.200	2.600	3.200	3.000	0.000	4.000	1.600	1.600	1.600	1.600	3.200	0.000	2.200	4.000
HF₈	2.600	2.200	3.000	3.000	3.600	2.400	4.000	0.000	1.800	1.800	1.800	1.800	2.400	2.200	3.200	2.000
HF₉	4.000	1.200	0.000	0.000	2.800	1.400	2.000	1.800	0.000	0.000	0.200	0.200	3.000	1.400	3.200	1.800
HF₁₀	4.000	1.200	0.000	0.000	2.800	1.400	2.000	1.800	0.200	0.000	0.200	0.200	2.600	1.000	2.800	1.800
HF₁₁	4.000	1.200	0.000	0.000	2.800	1.400	2.000	1.800	0.200	0.200	0.000	0.200	2.600	1.400	2.200	1.800
HF₁₂	4.000	1.200	0.000	0.000	1.400	4.000	3.200	2.000	3.000	3.000	3.000	0.000	3.000	1.000	3.200	3.000
HF₁₃	1.800	0.000	0.000	2.200	3.000	2.800	0.000	2.200	1.000	1.000	1.600	1.000	0.000	3.200	3.800	3.800
HF₁₄	4.000	0.000	0.000	1.000	3.000	3.000	2.200	2.200	2.200	3.200	1.800	1.200	3.600	0.000	3.200	2.200
HF₁₅	4.000	2.200	1.400	4.000	3.600	3.000	2.200	3.200	1.800	3.200	1.800	1.200	3.400	3.200	0.000	3.200
HF₁₆	3.600	2.000	1.200	4.000	2.000	3.400	3.000	2.000	1.800	2.800	1.800	2.000	3.600	1.800	2.000	0.000

Table 5. Normalised direct-relation matrix *D*.

D	HF₁	HF₂	HF₃	HF₄	HF₅	HF₆	HF₇	HF₈	HF₉	HF₁₀	HF₁₁	HF₁₂	HF₁₃	HF₁₄	HF₁₅	HF₁₆
HF₁	0.000	0.070	0.039	0.035	0.087	0.044	0.052	0.048	0.087	0.079	0.087	0.087	0.039	0.061	0.087	0.048
HF₂	0.048	0.000	0.061	0.066	0.057	0.000	0.000	0.070	0.026	0.035	0.035	0.026	0.000	0.000	0.057	0.031
HF₃	0.048	0.044	0.000	0.087	0.066	0.066	0.066	0.066	0.000	0.000	0.000	0.000	0.000	0.000	0.039	0.035
HF₄	0.044	0.070	0.070	0.000	0.087	0.026	0.087	0.070	0.000	0.000	0.000	0.000	0.052	0.026	0.079	0.079
HF₅	0.079	0.044	0.070	0.087	0.000	0.061	0.061	0.061	0.052	0.061	0.052	0.061	0.026	0.087	0.087	0.052
HF₆	0.052	0.022	0.048	0.052	0.070	0.000	0.044	0.074	0.031	0.031	0.031	0.031	0.087	0.061	0.066	0.083
HF₇	0.031	0.000	0.048	0.057	0.070	0.066	0.000	0.087	0.035	0.035	0.035	0.035	0.070	0.000	0.048	0.087
HF₈	0.057	0.048	0.066	0.066	0.079	0.052	0.087	0.000	0.039	0.039	0.039	0.039	0.052	0.048	0.070	0.044
HF₉	0.087	0.026	0.000	0.000	0.061	0.031	0.044	0.039	0.000	0.000	0.004	0.004	0.066	0.031	0.070	0.039
HF₁₀	0.087	0.026	0.000	0.000	0.061	0.031	0.044	0.039	0.004	0.000	0.004	0.004	0.057	0.022	0.061	0.039
HF₁₁	0.087	0.026	0.000	0.000	0.061	0.031	0.044	0.039	0.004	0.004	0.000	0.004	0.057	0.031	0.048	0.039
HF₁₂	0.087	0.026	0.000	0.000	0.031	0.087	0.070	0.044	0.066	0.066	0.066	0.000	0.066	0.022	0.070	0.066
HF₁₃	0.039	0.000	0.000	0.048	0.066	0.061	0.000	0.048	0.022	0.022	0.035	0.022	0.000	0.070	0.083	0.083
HF₁₄	0.087	0.000	0.000	0.022	0.066	0.066	0.048	0.048	0.048	0.070	0.039	0.026	0.079	0.000	0.070	0.048
HF₁₅	0.087	0.048	0.031	0.087	0.079	0.066	0.048	0.070	0.039	0.070	0.039	0.026	0.074	0.070	0.000	0.070
HF₁₆	0.079	0.044	0.026	0.087	0.044	0.074	0.066	0.044	0.039	0.061	0.039	0.044	0.079	0.039	0.044	0.000

Table 6. Total direct-relation matrix T.

T	HF$_1$	HF$_2$	HF$_3$	HF$_4$	HF$_5$	HF$_6$	HF$_7$	HF$_8$	HF$_9$	HF$_{10}$	HF$_{11}$	HF$_{12}$	HF$_{13}$	HF$_{14}$	HF$_{15}$	HF$_{16}$
HF$_1$	0.209	0.175	0.140	0.186	0.285	0.198	0.211	0.221	0.190	0.195	0.192	0.187	0.209	0.190	0.287	0.224
HF$_2$	0.162	0.068	0.123	0.155	0.172	0.090	0.098	0.167	0.086	0.102	0.095	0.085	0.094	0.076	0.171	0.129
HF$_3$	0.164	0.113	0.075	0.186	0.188	0.157	0.165	0.174	0.067	0.075	0.068	0.066	0.102	0.081	0.162	0.145
HF$_4$	0.199	0.155	0.158	0.137	0.243	0.155	0.214	0.210	0.090	0.102	0.091	0.088	0.181	0.130	0.235	0.217
HF$_5$	0.281	0.154	0.173	0.240	0.208	0.218	0.223	0.237	0.158	0.179	0.160	0.164	0.199	0.214	0.289	0.232
HF$_6$	0.228	0.117	0.139	0.191	0.243	0.144	0.188	0.224	0.128	0.139	0.129	0.126	0.233	0.177	0.242	0.237
HF$_7$	0.191	0.090	0.132	0.183	0.227	0.192	0.137	0.222	0.122	0.132	0.122	0.120	0.203	0.109	0.209	0.227
HF$_8$	0.236	0.145	0.160	0.207	0.258	0.194	0.231	0.162	0.136	0.148	0.138	0.134	0.202	0.164	0.251	0.206
HF$_9$	0.205	0.092	0.066	0.099	0.181	0.129	0.141	0.145	0.074	0.086	0.079	0.076	0.167	0.116	0.193	0.148
HF$_{10}$	0.199	0.089	0.063	0.095	0.175	0.124	0.136	0.140	0.075	0.078	0.075	0.073	0.153	0.103	0.178	0.142
HF$_{11}$	0.197	0.088	0.062	0.093	0.173	0.122	0.135	0.138	0.074	0.081	0.070	0.072	0.151	0.110	0.164	0.141
HF$_{12}$	0.210	0.095	0.068	0.104	0.185	0.133	0.146	0.149	0.081	0.090	0.082	0.075	0.171	0.110	0.197	0.177
HF$_{13}$	0.223	0.096	0.090	0.180	0.210	0.225	0.211	0.197	0.159	0.171	0.160	0.156	0.161	0.186	0.260	0.240
HF$_{14}$	0.223	0.077	0.072	0.130	0.201	0.170	0.116	0.165	0.103	0.111	0.116	0.101	0.194	0.099	0.210	0.171
HF$_{15}$	0.285	0.157	0.137	0.239	0.278	0.221	0.211	0.242	0.157	0.190	0.151	0.134	0.241	0.201	0.208	0.246
HF$_{16}$	0.252	0.139	0.120	0.219	0.222	0.210	0.208	0.199	0.135	0.166	0.137	0.137	0.226	0.155	0.225	0.162

Table 7. Final ranking.

	$r_i + c_i$	$r_i - c_i$	Ranking	$r_i + c_i \cdots \to$
HF$_1$	6.762	−0.167	HF$_5$	6.779
HF$_2$	3.721	0.024	HF$_{15}$	6.776
HF$_3$	3.765	0.210	HF$_1$	6.762
HF$_4$	5.250	−0.037	HF$_8$	5.964
HF$_5$	6.779	−0.121	HF$_{16}$	5.954
HF$_6$	5.568	0.202	HF$_{13}$	5.809
HF$_7$	5.390	−0.152	HF$_6$	5.568
HF$_8$	5.964	−0.020	HF$_7$	5.390
HF$_9$	3.833	0.162	HF$_4$	5.250
HF$_{10}$	3.943	−0.148	HF$_{14}$	4.477
HF$_{11}$	3.733	0.006	HF$_{10}$	3.943
HF$_{12}$	3.866	0.276	HF$_{12}$	3.866
HF$_{13}$	5.809	0.039	HF$_9$	3.833
HF$_{14}$	4.477	0.040	HF$_3$	3.765
HF$_{15}$	6.776	−0.181	HF$_{11}$	3.733
HF$_{16}$	5.954	−0.133	HF$_2$	3.721

Human factors with higher $r_i + c_i$ value, as explained in step 6 before, give crucial information regarding, in our case, how to reduce organisational risk, since their variations have greater impact on the variations of all the other aspects. As explained in step 7, a threshold has to be established for not taking into account negligible effects. As said, this threshold is here calculated as the average of all the elements in matrix T. In this case the threshold is 0.159. Now, those factors having a value of $T(\mathrm{HF}_i; \mathrm{HF}_i)$ higher than the threshold are selected.

Accordingly, we suggest that the human factors occupying the first six positions of the ranking need to be more carefully monitored during the process of organisational risk management. They are, in order:

- **HF$_5$**: Reward systems related to the achievement of security objectives are not defined;
- **HF$_{15}$**: Workers lead tasks having high level of responsibility for stakeholders, plants and production;
- **HF$_1$**: System of security and safety management not implemented;
- **HF$_8$**: Rigid protocols supervising and controlling workers are implemented;
- **HF$_{16}$**: Shifts may be not well organised;
- **HF$_{13}$**: Unpredictably variations of workload.

Figure 3 presents the four quadrants of the chart derived from the DEMATEL application. From this representation, decision makers can visually identify causal relationships among the considered human factors. The rationale for selecting, Si et al. [42], may be summarized as follows:

- Factors in quadrant I are identified as core factors or intertwined givers since they have high prominence and relation;
- Factors in quadrant II have low prominence but high relation, which are impacted by other factors and cannot be directly improved;
- Factors in quadrant III have low prominence and relation and are relatively disconnected from the system;
- Factors in quadrant IV are identified as driving factors or autonomous givers because they have high prominence but low relation.
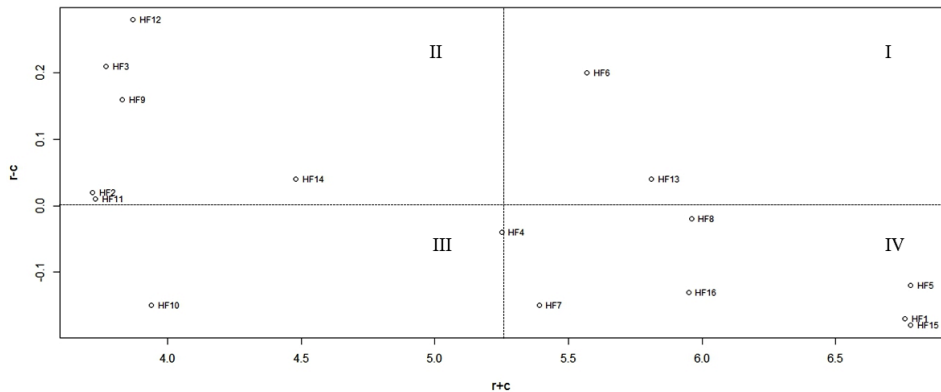
Figure 3. Decision-making trial and evaluation laboratory (DEMATEL) chart with human factors (HFs) spread out into quadrants.

Figure 4 shows the interdependencies among the selected HFs, the casual factors. In this methodology, arrows for the factors with values $T(\mathrm{HF}_i; \mathrm{HF}_i)$ lower than this threshold are not customary indicated in the graph, meaning that the corresponding interdependencies can be neglected [43].

The relations corresponding to the ten unselected HFs are, thus, not represented for the sake of clarity, despite some relation of interdependence between them and the other factors may exist.
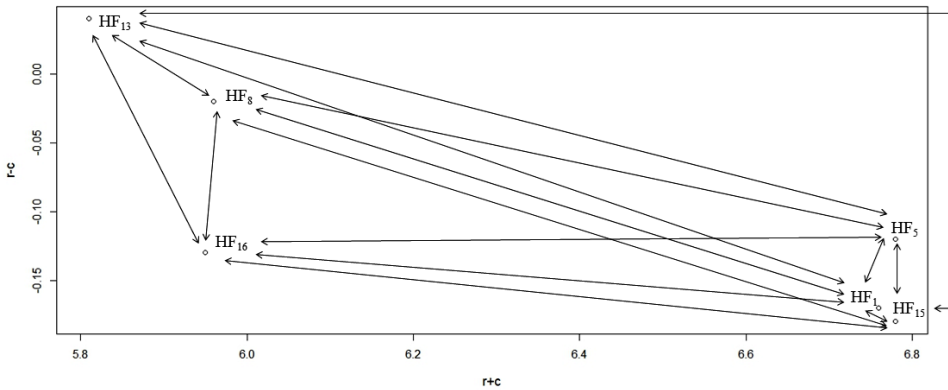
Figure 4. Chart representing interdependencies among the six selected HFs.

By analysing the six selected human factors, we can observe that human factor HF_5, by occupying the first position of the ranking, reveals the need for better defining reward systems related to the achievement of security objectives. This could be pursued by motivating workers in actively participating to the implementation of a system of security and safety management, as suggested also by human factor HF_1. Thus, this implementation may simultaneously enhance these two factors and can be addressed by starting from a clearer definition of procedures related to the planning and execution of preventive maintenance intervention for the bottling plant. Moreover, three of the selected factors (HF_{15}, HF_{16}, HF_{13}) belong to area A_8 (see Table 2), that is, "Workload and working hours". It means that, for example, interventions aimed at rearranging aspects related to the entity of workload and the number of working hours per worker could help improve the entire process under the organisational point of view.

Lastly, let us underline that, among the six selected HFs, the value of the difference $(r_i - c_i)$ is positive just for HF_{13}, what makes this factor a possible cause of bad process organisation and its improvement will produce benefits. The other five factors have associated a negative value of difference $(r_i - c_i)$, so these factors must be interpreted as cause factors of perceived risk.

5. Conclusions

The present paper deals with organisational risk assessment in industry, a field in which the role of human factors is crucial. In particular, a MCDM approach based on the DEMATEL methodology has been proposed to evaluate interdependencies among critical human factors. This method enables to rank human factors so that a framework for prioritising interventions thus reducing risk is suggested. From the influential relation map pictured in the chart, decision makers can visually detect the complex causal relationships among factors and highlight further valuable insights for decision-making.

The proposed approach is applied to a real-world case study related to a winery located in Sicily (Italy). The process of wine bottling has been taken into account, and results coming from a previous organisational risk evaluation have been manipulated. This evaluation highlighted 16 critical human factors with relation to the group of workers distributed in the analysed working area. The DEMATEL has been used to rank human factors on the basis of their interdependencies. The selected human factors (namely HF_5, HF_{15}, HF_1, HF_8, HF_{16}, HF_{13}) give fundamental information, and their variations correspond to variations of all the other aspects. For this reason, these human factors need to be monitored with priority during the process of organisational risk management.

Since the human factor HF_5 occupies the first position in the ranking, the company should consider as primary action the definition of reward systems related to the achievement of security objectives. This may be undertaken by motivating workers in taking part in the implementation of a system of security and safety management starting from a clearer definition of procedures related to the planning and execution of preventive maintenance intervention for the bottling plant. In this way, HF5 and HF1 would be simultaneously taken into account. This aspect may be further investigated, also in terms of management of the maintenance monitoring process, through suitable key performance indices, and the practical validity of our proposal should be carefully tested once the actions we suggest will be implemented in the context of reference.

Moreover, this work may be further extended by means of other MCDM approaches to support, and to the practical implementation of measures aimed at reducing organisational risk. These measures and their planning would directly derive from the ranking achieved in the present research. For example, the fuzzy set theory could be a useful tool to manage uncertainty and vagueness of the involved experts. In particular, with special regard to the critical human factors highlighted in the present article, an inference engine may be developed to support risk management. As another example, the analytic hierarchy process (AHP) method can also be employed to find out the weights of factors/aspects and obtain suitable scores for various actions that could be implemented. Lastly, the analytic hierarchy process (ANP), as an extension of the AHP, can be applied with the aim of understanding more complex dependency relationships among criteria.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Non-negative matrix $X^{(1)}$ filled in by the expert H_1, "maintenance responsible".

X^1	HF_1	HF_2	HF_3	HF_4	HF_5	HF_6	HF_7	HF_8	HF_9	HF_{10}	HF_{11}	HF_{12}	HF_{13}	HF_{14}	HF_{15}	HF_{16}
HF_1	0	3	1	1	4	2	2	2	4	4	4	4	1	3	4	2
HF_2	2	0	3	3	3	0	0	3	1	1	1	1	0	0	2	1
HF_3	2	2	0	4	3	3	3	3	0	0	0	0	0	0	2	2
HF_4	2	3	3	0	4	1	4	3	0	0	0	0	2	1	4	4
HF_5	4	2	3	4	0	3	3	3	3	3	3	3	1	4	4	2
HF_6	2	1	2	2	3	0	2	3	1	1	1	1	4	3	3	4
HF_7	1	0	2	3	3	3	0	4	2	2	2	2	3	0	2	4
HF_8	2	2	3	3	4	2	4	0	2	2	2	2	2	2	3	2
HF_9	4	1	0	0	3	1	2	2	0	0	0	0	3	1	3	2
HF_{10}	4	1	0	0	3	1	2	2	0	0	0	0	3	1	3	2
HF_{11}	4	1	0	0	3	1	2	2	0	0	0	0	3	1	2	2
HF_{12}	4	1	0	0	3	1	2	2	0	0	0	0	3	1	3	3
HF_{13}	1	0	0	2	1	4	3	2	3	3	3	3	0	3	4	4
HF_{14}	4	0	0	1	3	3	0	2	1	1	2	1	4	0	3	2
HF_{15}	4	2	1	4	4	3	2	3	2	3	2	1	4	3	0	3
HF_{16}	3	2	1	4	2	4	3	2	2	3	2	2	4	2	2	0

Table A2. Non-negative matrix $X^{(2)}$ filled in by the expert H_2, "quality manager".

X^2	HF$_1$	HF$_2$	HF$_3$	HF$_4$	HF$_5$	HF$_6$	HF$_7$	HF$_8$	HF$_9$	HF$_{10}$	HF$_{11}$	HF$_{12}$	HF$_{13}$	HF$_{14}$	HF$_{15}$	HF$_{16}$
HF$_1$	0	3	2	2	4	1	2	1	4	3	4	4	2	2	4	2
HF$_2$	2	0	2	3	2	0	0	3	1	2	2	1	0	0	3	2
HF$_3$	2	2	0	4	3	3	3	3	0	0	0	0	0	0	1	1
HF$_4$	1	3	3	0	4	1	4	3	0	0	0	0	2	1	3	3
HF$_5$	3	2	3	4	0	3	3	3	2	3	2	3	1	4	4	3
HF$_6$	2	1	2	3	4	0	2	3	1	1	1	1	4	3	3	4
HF$_7$	1	0	2	2	4	3	0	4	1	1	1	1	3	0	2	4
HF$_8$	2	2	3	2	4	2	4	0	1	1	1	1	2	2	3	2
HF$_9$	4	1	0	0	3	1	2	1	0	0	0	0	3	2	3	2
HF$_{10}$	4	1	0	0	3	1	2	1	0	0	0	0	2	1	2	2
HF$_{11}$	4	1	0	0	3	1	2	1	0	0	0	0	2	2	2	2
HF$_{12}$	4	1	0	0	3	1	2	1	0	0	0	0	3	1	3	3
HF$_{13}$	2	0	0	2	1	4	3	1	3	3	3	3	0	3	4	3
HF$_{14}$	4	0	0	1	3	3	0	2	1	1	1	1	4	0	3	2
HF$_{15}$	4	2	1	4	3	3	2	3	2	3	1	1	4	3	0	3
HF$_{16}$	4	2	1	4	2	3	3	2	2	3	2	2	4	2	2	0

Table A3. Non-negative matrix $X^{(3)}$ filled in by the expert H_3, "consultant".

X^3	HF$_1$	HF$_2$	HF$_3$	HF$_4$	HF$_5$	HF$_6$	HF$_7$	HF$_8$	HF$_9$	HF$_{10}$	HF$_{11}$	HF$_{12}$	HF$_{13}$	HF$_{14}$	HF$_{15}$	HF$_{16}$
HF$_1$	0	3	1	1	4	3	3	3	4	4	4	4	1	3	4	2
HF$_2$	2	0	3	3	3	0	0	3	2	2	2	2	0	0	2	1
HF$_3$	2	2	0	4	3	3	3	3	0	0	0	0	0	0	2	2
HF$_4$	2	3	3	0	4	1	4	3	0	0	0	0	2	1	4	4
HF$_5$	4	2	3	4	0	2	2	2	2	2	2	2	1	4	4	2
HF$_6$	3	1	2	2	2	0	2	3	1	1	1	1	4	3	3	4
HF$_7$	2	0	2	3	2	3	0	4	2	2	2	2	3	0	2	4
HF$_8$	3	2	3	3	3	2	4	0	2	2	2	2	2	2	3	2
HF$_9$	4	2	0	0	2	1	2	2	0	0	0	0	2	1	3	2
HF$_{10}$	4	2	0	0	2	1	2	2	0	0	0	0	2	1	3	2
HF$_{11}$	4	2	0	0	2	1	2	2	0	0	0	0	2	1	2	2
HF$_{12}$	4	2	0	0	2	1	2	2	0	0	0	0	2	1	3	3
HF$_{13}$	1	0	0	2	2	4	3	2	2	2	2	2	0	3	4	4
HF$_{14}$	4	0	0	1	3	3	0	2	1	1	2	1	3	0	3	3
HF$_{15}$	4	2	1	4	4	3	2	3	2	3	2	1	3	3	0	3
HF$_{16}$	3	2	1	4	2	4	3	2	2	3	2	2	3	1	1	0

Table A4. Non-negative matrix $X^{(4)}$ filled in by the expert H_4, "chief of the safety and security system".

X^4	HF$_1$	HF$_2$	HF$_3$	HF$_4$	HF$_5$	HF$_6$	HF$_7$	HF$_8$	HF$_9$	HF$_{10}$	HF$_{11}$	HF$_{12}$	HF$_{13}$	HF$_{14}$	HF$_{15}$	HF$_{16}$
HF$_1$	0	3	2	2	4	1	2	1	4	3	4	4	2	2	4	2
HF$_2$	2	0	2	3	2	0	0	3	1	2	2	1	0	0	3	2
HF$_3$	2	2	0	4	3	3	3	2	0	0	0	0	0	0	1	1
HF$_4$	1	3	3	0	4	1	4	3	0	0	0	0	2	1	3	3
HF$_5$	3	2	3	4	0	3	3	2	2	3	2	3	1	4	4	3
HF$_6$	2	1	2	3	4	0	2	4	3	3	3	3	4	2	2	3
HF$_7$	1	0	2	2	4	3	0	4	1	1	1	1	3	0	2	4
HF$_8$	2	2	2	3	3	3	4	0	1	1	1	1	2	2	3	1
HF$_9$	4	1	0	0	3	3	2	1	0	1	1	1	3	2	3	1
HF$_{10}$	4	1	0	0	3	3	2	1	1	0	1	1	2	1	2	1
HF$_{11}$	4	1	0	0	3	3	2	1	1	1	0	1	2	2	2	1
HF$_{12}$	4	1	0	0	3	3	2	1	1	1	1	0	3	1	3	3
HF$_{13}$	2	0	0	2	1	4	3	1	3	3	3	3	0	3	3	4
HF$_{14}$	4	0	0	1	3	2	0	2	1	1	1	1	3	0	3	2
HF$_{15}$	4	2	1	4	3	2	2	3	2	3	1	1	2	3	0	3
HF$_{16}$	4	2	1	4	2	2	3	1	1	2	1	2	3	2	2	0

Table A5. Non-negative matrix $X^{(5)}$ filled in by the expert H_5, "Department chief".

X^5	HF_1	HF_2	HF_3	HF_4	HF_5	HF_6	HF_7	HF_8	HF_9	HF_{10}	HF_{11}	HF_{12}	HF_{13}	HF_{14}	HF_{15}	HF_{16}
HF_1	0	4	3	2	4	3	3	4	4	4	4	4	3	4	4	3
HF_2	3	0	4	3	3	0	0	4	1	1	1	1	0	0	3	1
HF_3	3	2	0	4	3	3	3	4	0	0	0	0	0	0	3	2
HF_4	4	4	4	0	4	2	4	4	0	0	0	0	4	2	4	4
HF_5	4	2	4	4	0	3	3	4	3	3	3	3	2	4	4	2
HF_6	3	1	3	2	3	0	2	4	1	1	1	1	4	3	4	4
HF_7	2	0	3	3	3	3	0	4	2	2	2	2	4	0	3	4
HF_8	4	3	4	4	4	3	4	0	3	3	3	3	4	3	4	3
HF_9	4	1	0	0	3	1	2	3	0	0	0	0	4	1	4	2
HF_{10}	4	1	0	0	3	1	2	3	0	0	0	0	4	1	4	2
HF_{11}	4	1	0	0	3	1	2	3	0	0	0	0	4	1	3	2
HF_{12}	4	1	0	0	3	1	2	3	0	0	0	0	4	1	4	3
HF_{13}	3	0	0	3	2	4	4	4	4	4	4	4	0	4	4	4
HF_{14}	4	0	0	1	3	3	0	3	1	1	2	1	4	0	4	2
HF_{15}	4	3	3	4	4	4	3	4	3	4	3	2	4	4	0	4
HF_{16}	4	2	2	4	2	4	3	3	2	3	2	2	4	2	3	0

References

1. Clerici, P.; Guercio, A.; Quaranta, L. *Human Management System for Occupational Health and Safety*; Tipolitografia INAIL: Milan, Italy, 2016.

2. Azadeh, A.; Zarrin, M. An intelligent framework for productivity assessment and analysis of human resource from resilience engineering, motivational factors, HSE and ergonomics perspectives. *Saf. Sci.* **2016**, *89*, 55–71. [CrossRef]

3. Boatca, M.E.; Cirjaliu, B. A Proposed Approach for an Efficient Ergonomics Intervention in Organizations. *Procedia Econ. Financ.* **2015**, *23*, 54–62. [CrossRef]

4. Cirjaliu, B.; Draghici, A. Ergonomic Issues in Lean Manufacturing. *Procedia Soc. Behav. Sci.* **2016**, *221*, 105–110. [CrossRef]

5. Radjiyev, A.; Qiu, H.; Xiong, S.; Nam, K. Ergonomics and sustainable development in the past two decades (1992–2011): Research trends and how ergonomics can contribute to sustainable development. *Appl. Ergon.* **2015**, *46*, 67–75. [CrossRef] [PubMed]

6. Thatcher, A.; Yeow, P.H.P. Human factors for a sustainable future. *Appl. Ergon.* **2016**, *57*, 1–7. [CrossRef] [PubMed]

7. Chidambaram, P. Perspectives on human factors in a shifting operational environment. *J. Loss Prevent Proc.* **2016**, *44*, 112–118. [CrossRef]

8. Hinshaw, K. Human factors in obstetrics and gynaecology. *Obstet. Gynaecol. Reprod. Med.* **2016**, *26*, 368–370. [CrossRef]

9. Ergai, A.; Cohen, T.; Sharp, J.; Wiegmann, D.; Gramopadhye, A.; Shappell, S. Assessment of the Human Factors Analysis and Classification System (HFACS): Intra-rater and inter-rater reliability. *Saf. Sci.* **2016**, *82*, 393–398. [CrossRef]

10. Gholami, P.S.; Nassiri, P.; Yarahmadi, R.; Hamidi, A.; Mirkazemi, R. Assessment of Health Safety and Environment Management System function in contracting companies of one of the petro-chemistry industries in Iran, a case study. *Saf. Sci.* **2015**, *77*, 42–47. [CrossRef]

11. Wilson, J.R. Fundamentals of systems ergonomics/human factors. *Appl. Ergon.* **2014**, *45*, 5–13. [CrossRef] [PubMed]

12. Hassall, M.; Xiao, T.; Sanderson, P.; Neal, A. Human Factors and Ergonomics. In *International Encyclopaedia of the Social & Behavioral Sciences*, 2nd ed.; Elsevier: Amsterdam, Netherlands, 2015; pp. 297–305.

13. Sobhani, A.; Wahab, M.I.M.; Neumann, W.P. Incorporating human factors-related performance variation in optimizing a serial system. *Eur. J. Oper. Res.* **2017**, *257*, 69–83. [CrossRef]

14. Fontela, E.; Gabus, A. *DEMATEL, Innovative Methods. Report No. 2 Structural Analysis of the World Problematique*; Battelle Geneva Research Institute: Columbus, OH, USA, 1974.

15. Fontela, E.; Gabus, A. *The DEMATEL Observe*; Battelle Institute, Geneva Research Center: Columbus, OH, USA, 1976.

16. INAIL. Evaluation and Management of Work-Related Stress Risk. Manual for Use by Companies in Implementation of Legislative Decree 81/08 and Subsequent Amendments. 2011. Available online: https://sicurezzaprevenzione.files.wordpress.com/2011/05/manuale-inail-2011-valutazione-e-gestione-rischio-stress-lavoro-correlato.pdf (accessed on 24 October 2018).

17. Carpitella, S.; Certa, A.; Enea, M.; Galante, G.M.; Izquierdo, J.; La Fata, C.M. Human Reliability Analysis to support the development of a software project. In Proceedings of the 23th ISSAT International Conference on Reliability and Quality in Design, Chicago, IL, USA, 4–6 August 2017; pp. 214–218.

18. Saravia-Pinilla, M.H.; Daza-Beltrán, C.; García-Acosta, G. A comprehensive approach to environmental and human factors into product/service design and development. A review from an ergoecological perspective. *Appl. Ergon.* **2016**, *57*, 62–71. [CrossRef] [PubMed]

19. Wiegmann, D.A.; Shappell, S.A. *A Human Error Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System*; Ashgate: Aldershot, UK, 2003.

20. Omole, H.; Walker, G. Offshore Transport Accident Analysis Using HFACS. *Procedia Manuf.* **2015**, *3*, 1264–1272. [CrossRef]

21. Soner, O.; Asan, U.; Celik, M. Use of HFACS–FCM in fire prevention modelling on board ships. *Saf. Sci.* **2015**, *77*, 25–41. [CrossRef]

22. Chen, S.-T.; Wall, A.; Davies, P.; Yang, Z.; Wang, J.; Chou, Y.-H. A Human and Organisational Factors (HOFs) analysis method for marine casualties using HFACS-Maritime Accidents (HFACS-MA). *Saf. Sci.* **2013**, *60*, 105–114. [CrossRef]

23. Madigan, R.; Golightly, D.; Madders, R. Application of Human Factors Analysis and Classification System (HFACS) to UK rail safety of the line incidents. *Accid. Anal. Prev.* **2016**, *97*, 122–131. [CrossRef] [PubMed]

24. Choe, P.; Tew, J.D.; Tong, S. Effect of cognitive automation in a material handling system on manufacturing flexibility. *Int. J. Prod. Econ.* **2015**, *170*, 891–899. [CrossRef]

25. Zio, E.; Baraldi, P.; Librizzi, M.; Podofillini, L.; Dang, V.N. A fuzzy set-based approach for modeling dependence among human errors. *Fuzzy Sets Syst.* **2009**, *160*, 1947–1964. [CrossRef]

26. Evans, D.C.; Fendley, M. A multi-measure approach for connecting cognitive workload and automation. *Int. J. Hum. Comput. Stud.* **2017**, *97*, 182–189. [CrossRef]

27. Swain, A.D. Human Reliability Analysis: Needs, Status, Trends and Limitations. *Reliab. Eng. Syst. Safe* **1990**, *29*, 301–313. [CrossRef]

28. Reer, B. Review of advances in human reliability analysis of errors of commission, Part 1: EOC identification. *Reliab. Eng. Syst. Safe* **2008**, *93*, 1091–1104. [CrossRef]

29. Reer, B. Review of advances in human reliability analysis of errors of commission—Part 2: EOC quantification. *Reliab. Eng. Syst. Safe* **2008**, *93*, 1105–1122. [CrossRef]

30. Drivalou, S.; Marmaras, N. Supporting skill-, rule-, and knowledge-based behaviour through an ecological interface: An industry-scale application. *Int. J. Ind. Ergon.* **2009**, *39*, 947–965. [CrossRef]

31. Reason, J. *Human Error*; Cambridge University Press: Cambridge, UK, 1990.

32. European Social Partners (Business Europe, CEEP, ETUC, UEAPME). Implementation of the European Autonomous Framework Agreement on Work-Related Stress. Brussels, Belgium, 2008. Available online: https://osha.europa.eu/es/legislation/guidelines/implementation-of-the-european-autonomous-framework-agreement-on-work-related-stress (accessed on 24 October 2018).

33. Hori, S.; Shimizu, Y. Designing methods of human interface for supervisory control systems. *Control Eng. Pract.* **1999**, *7*, 1413–1419. [CrossRef]

34. Tamura, M.; Nagata, H.; Akazawa, K. Extraction and systems analysis of factors that prevent safety and security by structural models. In Proceedings of the 41st SICE Annual Conference, Osaka, Japan, 5–7 August 2002.

35. Chiu, Y.J.; Chen, H.C.; Tzeng, G.H.; Shyu, J.Z. Marketing strategy based on customer behavior for the LCD-TV. *Int. J. Manag. Decis. Mak.* **2006**, *7*, 143–165. [CrossRef]

36. Meyer, C. *Matrix Analysis and Applied Linear Algebra*; SIAM: Philadelphia, PA, USA, 2000; ISBN 0898714540.

37. Tzeng, G.H.; Chiang, C.H.; Li, C.W. Evaluating intertwined effects in e-learning programs: A novel hybrid MCDM model based on factor analysis and DEMATEL. *Expert Syst. Appl.* **2007**, *32*, 1028–1044. [CrossRef]

38. Lin, C.; Tzeng, G. A value-created system of science (technology) park by using DEMATEL. *Expert Syst. Appl.* **2009**, *36*, 9683–9697. [CrossRef]

39. Azadeh, A.; Zarrin, M.; Abdollahi, M.; Noury, S.; Farahmand, S. Leanness assessment and optimization by fuzzy cognitive map and multivariate analysis. *Expert Syst. Appl.* **2015**, *42*, 6050–6064. [CrossRef]
40. Lee, P.T.-W.; Lin, C.-W. The cognition map of financial ratios of shipping companies using DEMATEL and MMDE. *Marit. Policy Manag.* **2013**, *40*, 133–145. [CrossRef]
41. Sara, J.; Stikkelman, R.M.; Herder, P.M. Assessing relative importance and mutual influence of barriers for CCS deployment of the ROAD project using AHP and DEMATEL methods. *Int. J. Greenh. Gas Control* **2015**, *41*, 336–357. [CrossRef]
42. Si, S.-L.; You, X.-Y.; Liu, H.-C.; Zhang, P. DEMATEL Technique: A Systematic Review of the State-of-the-Art Literature on Methodologies and Applications. *Math. Probl. Eng.* **2018**. [CrossRef]
43. Büyüközkan, G.; Güleryüz, S. An integrated DEMATEL-ANP approach for renewable energy resources selection in Turkey. *Int. J. Prod. Econ.* **2016**, *182*, 435–448. [CrossRef]

Article

A Mathematical Model for Intimate Partner Violence

Sandra Delgadillo-Aleman [1,*] , **Roberto Ku-Carrillo** [1] , **Brenda Perez-Amezcua** [1] **and Benito Chen-Charpentier** [2]

[1] Departamento de Matemáticas y Física, Universidad Autónoma de Aguascalientes, Av. Universidad 940, Cd. Universitaria, Aguascalientes, Mexico; raku@correo.uaa.mx (R.K.-C.); pam_amezcua_61@hotmail.com (B.P.-A.)

[2] Department of Mathematics, University of Texas at Arlington, Arlington, TX 76019, USA; bmchen@uta.edu

[*] Correspondence: sedelgad@correo.uaa.mx; Tel.: +52-449-910-8411

Abstract: In the context of mathematical models applied to social sciences, we present and analyze a model based on differential equations for the intimate partner violence (IPV). Such a model describes the dynamics of a heterosexual romantic couple in which the man perpetrates violence against the woman. We focus on incorporating different key factors reported in the literature as causal or motivational factors to perpetrate IPV. Among the main factors included are the failures in self-regulation, the man's need to control the woman, the social pressure on the woman to remain married, and empowerment programs. Another aspect that we include is periodic alcohol consumption for the man. The discussion of the model includes a stability analysis of its equilibrium points and the asymptotic behavior of its solutions. Also, the interpretation of results is presented in terms of IPV phenomenon. Finally, a brief review is given on different scales to quantify human behavioral traits and numerical simulations for some IPV scenarios.

Keywords: IPV; violence index; independence index; model; ode

1. Introduction

Violence against woman is a social problem that afflicts all societies, without distinction of race, socioeconomic level or academic degree. In fact, it is considered a global public health problem by the World Health Organization (WHO), due to its frequency and alarming consequences of violence against women. In particular, the intimate partner violence (IPV) is one of the most common and includes physical, sexual or emotional abuse by the intimate partner towards the woman, according to WHO [1]. Paradoxically, this type of violence is also usually one of the most forgiven, quiet and ignored, often for being something that occurs in privacy and is perpetrated by the loved one. Although IPV is officially penalized and rejected, it continues to occur with relative impunity, and is sometimes well accepted by some societies. Global statistics estimate that about one out of three women have experienced some type of violence worldwide. In Mexico, six out of ten women have been victims of some kind of aggression, at least once in their lives, and four out of ten have suffered some type of violence caused by her intimate partner. Likewise, 33% of the women in Mexico suffered violence during childhood and seven femicides occur daily according to the statistics of the National Institute of Statistics and Geography (INEGI) [2]. Worldwide, Mexico ranks 71 of 169 countries ranked according to their gender inequality measured with the Gender Inequality Index (GII) [3,4]. From our point of view, these statistics show the importance of studying this problem. However, mathematical modeling involving differential equations is an approach almost not used in social science problems, but it has proven its importance to understand and offer a solution for different problems [5–7].

The classical approach of mathematics in social sciences usually involves the application of statistical techniques. A common situation is the validation of hypothesis through statistical tests.

Another application of statistical models is the measurement and quantification of different factors or variables. As an example of quantification of the level of violence, we can refer to [8] by Sánchez et al. where a study to measure the level of historical violence of a population is described. Our approach is different from these examples. We search for models to describe the evolution in time of violence or related variables. With this approach, we find models such as the one reported by Poza et al. [9] which is an epidemiological type model that describes the evolution of the populations of different types of men, who were classified according to their level of woman abuse in Spain. This model is in terms of difference equations and predicts the evolution of IPV in Spain based on the statistical information reported by the literature. Another, important and pioneer work which is supported by a long and deep research on human behavior for marital couples is reported in [10,11] by Gottman et al. Let us say, that this article is not directly related to violence levels, but it presents and analyzes a model to predict a possible marital dissolution. In fact, it succeeded remarkably well. The model considers two functions of "influence" which aim to model the balance of positive-negative attitudes of a spouse towards the other in the context of an interview on conflict topics. The talk is recorded and coded according to a validated scale to obtain the parameters of the model. Using this information the authors were capable of making predictions on marriage success chances. We borrow some ideas of this article to model the interaction between the intimate partners involved in a violent relationship which we will discuss later.

On the other side, let us describe some sociological and psychological theories related to violence which we will use to construct our model. It is important to mention that sociologists have studied and analyzed this serious problem and developed a number of different theories. For example in [12], L. Walkers points out that violence against women is a cyclic process that consists of an accumulation of stress, acute violence, and reconciliation. Also, that this behavior has its origin in the patriarchal culture, which is enmeshed in the men's and women's thoughts through social learning mechanisms. Moreover, in the context of a patriarchal society, the success of a man in a relationship is based on the use of power and control of the woman, according to Corsi [13]. Another work also in the context of social learning theory which we found useful is the article by Stith and Fairley [14]. There, the authors propose a statistical model to predict violence in the couple, taking external and internal variables of the man. They reported that sex-role egalitarianism, approval, and observation of marital violence have effects on the use of severe marital violence. Moreover, that the observation of marital violence decreases self-esteem which raises marital stress and the level of alcohol consumption, both of which may increase marital violence. However, they remark that egalitarianism has a negative effect on the violence level, meaning, it could decrease the use of severe violence. Otherwise, those violent acts damage the self-esteem of the woman and create codependency toward her intimate partner which could be boosted by social pressure. On the opposite side, we take into account the effect of empowerment programs for the woman which have proven to be an effective strategy to reduce IPV [15], and we will include it as an external factor that could change the violence dynamic in a couple.

Finally, Finkel et al. [16] have reported that self-regulatory failures are important predictors of IPV. Besides that, the self-regulation capacity of a battering man is a trait that can be improved or depleted through training sessions. From the above and from various sociological multiple causal models, we will consider causal and motivational factors in our study. The causal factors are those that set the "perfect scenario" for male violence to occur in a relationship, this is according to the social learning theory. The causal factors in the man that unleash violence are the following: Violence observed during childhood, low self-esteem, acceptance of machismo, and failures in self-regulation. On the other hand, the motivational factors are those that motivate to perpetrate the violence and are more likely to trigger a violent episode in the relationship. The motivational factors that we will consider are the man's need to control woman and the effect of external factors for the woman. Other facts that support our selection of factors that we will include in the model come from the Mexican national surveys on violence [2,17]. There, violence against women is described and fits the situation

that our model intends to simulate. Let us remark that health-related factors such as mental and/or physical illnesses and external support factors such as psychological therapies will not be considered in our proposed model.

Our objective is to model mathematically the dynamics of the IPV problem in one couple in order to contribute to a better understanding of this social phenomenon. Let us remark that while violence affects both men and women, the violence against women is more alarming in some countries, for example in Mexico. For this reason, our model describes violence in one direction, man battering woman. This model was proposed using the theory of differential equations and taking into account causes or relevant factors that may contribute to the triggering of the violent behavior described above. We present the development of a first mathematical model in ordinary differential equations, which allows us to predict the possible violent situations that a woman could face in the relationship. Later, we use the obtained model to simulate computationally probable scenarios of violence.

The article is organized as follows: In Section 2 we present and explain the assumptions made in order to construct the model in differential equations. In particular, in Section 2.1 we present a simple model without considering the interaction between the romantic partners, as we do in Section 2.2. In Section 2.3. we discuss different scales related to the variables and parameters of our model. While in Section 3, we present the asymptotic analysis of the model without considering the man's alcohol consumption and provide its interpretation in terms of the factors considered. In Section 4, we show numerical simulations for some IPV scenarios. Later in Section 5, a discussion about the model and its application is given. Finally, conclusions are presented in the last section.

2. Mathematical Modeling

Our aim is to give a model that describes the IPV dynamics in a romantic couple. In order to explain in detail the model formation, let us present it in two parts. In the first one, we consider equations for each spouse independently of their intimate partner. We will call this interaction non-influenced intimate partner interaction. In the second one, we include the interaction between intimate partners. We call it influenced intimate partner interaction.

2.1. Non-influenced Intimate Partner Interaction

Let us first define two variables, one for the man and the other for the women. The variable $v(t)$ is the man's violent behavior index (positive) at time t which indicates the potential level of the man's hostility or aggression. The variable $i(t)$ is the woman independence behavior index which indicates the woman's state of freedom (positive) or potential of dependency or submission (negative).

- **A0.** In a similar way and following [10], we assume the existence of non-influenced steady states $\tilde{v}$ and $\tilde{i}$ for each individual, man and woman indexes, respectively. Those states are independent of each other intimate partner. Actually, they depend on their own personal history with respect to the IPV problem. For this reason, we assume that the non-influenced steady states are directly proportional to two causal factors: Violence seen in childhood and acceptance of violence. So, we define the non-influenced man's violence steady state as $\tilde{v} = \alpha_1 \beta_1$ and the non-influenced woman's independence steady state as $\tilde{i} = (1 - \alpha_2)(1 - \beta_2)$, where α_1, α_2 are the parameters that quantify the violence watched in childhood for man and woman, respectively, and β_1, β_2 are the parameters that quantify the acceptance of machismo for the man and woman, respectively. Notice that the non-influenced independence index of the woman decreases according to the level of violence watched during her infancy and her acceptance of IPV [18]. We will discuss these parameters in more detail in the next section.
- **A1.** These non-influenced steady states correspond to an equilibrium point $(\tilde{v}, \tilde{i})$ of the following differential equation system:

$$
\begin{array}{ll}
\frac{dv(t)}{dt} = \delta_1(\tilde{v} - v(t)), & v(0) = v_0, \\
\frac{di(t)}{dt} = \delta_2(\tilde{i} - i(t)), & i(0) = i_0,
\end{array}
\tag{1}
$$

where δ_1 and δ_2 are positive constant called "inertias" which are associated with the coping abilities with the problems of each individual of the couple. Mathematically, these coefficients correspond to the speed at which the system solutions tend to the equilibrium point. Note that $(\tilde{v}, \tilde{i})$ is an asymptotically stable equilibrium point of the model, for any initial condition (v_0, i_0). This can be interpreted as, when there is not couple interaction and after an unusual event (violent or not), each individual tends to his or her non-influenced index or state.

2.2. Influenced Intimate Partner Interactions

Now, based on the non-influenced model (1), we add the interactions between the intimate partners by taking into account other causal factors such as the self-esteem and self-regulation. Also, motivational factors such as, the man's need to control woman and the effect of external factors for woman. Then, the proposed system of differential equations that models the interaction between the man and woman in one romantic couple is:

$$
\begin{array}{llll}
\frac{dv}{dt}(t) & = \delta_1(\tilde{v} - v(t)) + k_1(1 - \rho_1)\frac{\beta_1}{\gamma}i(t) + h(t), & v(0) = v_0, \\
\frac{di}{dt}(t) & = \delta_2(\tilde{i} - i(t)) + k_2(1 - \rho_2)\mu v(t), & i(0) = i_0,
\end{array}
\tag{2}
$$

where variables $v(t)$ and $i(t)$ have the same meaning as in Model (1). The same applies for the coefficients α_i, β_i, and δ_i, for $i = 1, 2$. The added coefficients ρ_1 and ρ_2 are the self-regulatory parameters for man and woman, respectively. k_1 and k_2 are proportionality positive constants. γ is the man's self-esteem and μ is an external factor that could be family or social pressure, when μ is positive. Otherwise for μ negative, the woman is exposed to empowerment programs. Let us remark, that all coefficients are positive, but we allow μ to take positive and negative values. A graphical representation of the model is shown in Figure 1.
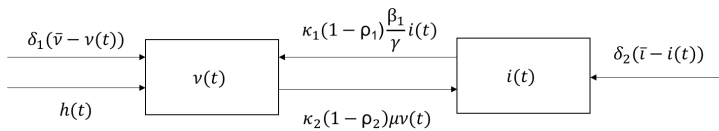
Figure 1. Diagram of model (2) showing the variables and their interaction terms.

In order to explain the terms that we have added, let us summarize the assumptions:

- **A2.** The rate at which the man's violence index increases depends on the man's need to control the woman which is caused by a low man's self-esteem [14]. In particular, we assume that it is inversely proportional to the man self-esteem, $1/\gamma$ in $k_1(1 - \rho_1)\frac{\beta_1}{\gamma}i(t)$. Besides, we assume that the variation rate of $v(t)$ depends linearly on the woman's independence index. This is because the man will have more need to control the woman when woman becomes more independent. This is supported by patriarchal theory as described in [13]. Moreover, the coefficients in the modeling term $k_1(1 - \rho_1)\frac{\beta_1}{\gamma}i(t)$ were set considering the influence of the man's acceptance of machismo β_1, as well as, possible self-regulatory failures ρ_1. It is important to remark, that the self-regulatory factor of a person is something difficult to quantify because depends on several random factors. For these reason and as a first approximation, we assume that these coefficients are constant.
- **A3.** In an analogous way for the modeling term $k_2(1 - \rho_2)\mu v(t)$, we assume that the degree of violence that the man exerts against the woman, in addition with the influence of the social environment, are factors that can change the woman's independence index. This might cause an increase in the level of the woman submission (μ negative) caused by the social pressure of the woman's family or friends to remain in the relationship. Or it can increase the level of the woman's independence index (μ positive) when the woman is enrolled in empowerment

programs. Let us remark the importance of the sign change of μ which implies a turning point in the IPV dynamics of the couple. On the other hand, notice that ρ_2 is included in this term and is the woman's self-regulatory coefficient. So, we propose that $v(t)$ influences linearly the change rate of $i(t)$ and vice versa.

- **A4.** According to the latest National Survey on Violence against Women in Mexico (ENVIM) [2], the frequency of alcohol consumption by a violent man is proportional to the number of aggressions perpetrated by a battering man. Also, that the alcohol consumption events are generally periodic with usually weekly or monthly periodicity. In order to model this behavior, we use the modeling term $h(t)$. We assume that this situation can be properly described by a sine function with parameters that allow us to control its amplitude (ρ) and its frequency (ω): $h(t) = \rho(\sin(\omega t) + \eta)$, where η is a positive constant.

Let us recall, that we do not take into account for our model health-related factors such as mental and/or physical illnesses nor external supporting/therapeutic factors such as psychological therapies or rehabilitation. This is, in fact, worthwhile to study, and we hope to include it in a near future. Let us remark that all these factors involved in the model (2) can be measured through scales developed by social science specialists, which we discuss in the next subsection.

2.3. Variable and Parameter Scales

In this section, we discuss the different scales related to the factors considered in our IPV model and that describes the efforts to measure different human behavioral traits. Let us start with the variables of our model. As previously discussed $v(t)$ is the man's violence index. In the literature, it is reported that there are several scales to measure the violence against women. Several instruments have been designed to quantify the level of violence [19,20]. One of them is known as the "Violentometro". In particular, in Mexico, the National Institute of Women called in Spanish "Inmujeres" and the Institutional Management Program in Gender Perspective of the IPN (Instituto Politecnico Nacional or the National Polytechnic Institute in Mexico) proposed a scale widely applied that takes integer values in the interval $[0, 30]$, see [21]. It assigns the value of 0 (zero) to man's non-violent behavior and 30 corresponds to murder. To compute such a value, a questionnaire is applied with the aim of assessing the degree of violence a woman faces. In an analogous way, we can find studies to measure the woman's independence [22]. For our purposes, this shows how the indexes can be quantified and in order to introduce them to our model we normalize them to the interval $[0, 1]$.

In the case of the parameters, α_1 and α_2 which correspond to the violence observed in childhood for the man and woman, respectively. A revision of the methodology to measure these variables can be found in [23]. There, the main tool to measure the violence exposition during childhood was obtained with conflict tactics scales. The range of this scale varies according to the applied version of the instrument and the number of used items. As an example of application of this type of measurements, we can see the results reported by the ENVIM 2003 [17]. There, the applied index grades the aggression in a scale from 0 to 12 points. In order to be used in our model, we normalize it to the interval $[0, 1]$. A zero value corresponds to a lack of violence childhood, while $\alpha_i = 1$, $i = 1, 2$, indicates a high degree of violence observed in childhood. In the case of parameters that measure the acceptance of machismo β_i, $i = 1, 2$ can be also quantified. As an example, let us cite the work by Arciniega et al. called Toward a Fuller Conception of Machismo: Development of a Traditional Machismo and Caballerismo Scale in [24]. Here, a seven points scale is analyzed to measure this behavior. Also, the work by Bendezu [25] evaluates gender stereotypes, beliefs, and expectations to give a weight in the interval $[0, 1]$ for a romantic couple. A value close to one indicates high acceptance of machismo attitudes. Continuing with the model parameters, γ which measures the man's self-esteem can be measured using the scale of Rosenberg. It takes integer values that range from 0 to 40 points [26]. Again we normalize the scale in such a way that we consider γ values in the interval $[0, 1]$. When $\gamma = 1$ it means that the person has high self-esteem while $\gamma = 0$ is the opposite. For parameters δ_1 and δ_2 which we claim are related to coping styles. We can mention the work reported in [27]. There,

a scale is defined according to different coping styles which define the way an individual returns to its natural level of violence or independence after facing a stressing situation. For a scale applied in Mexico, we refer the reader to [28]. For coefficient μ which is related to the external factors as the empowerment programs for the woman or the social pressure. Let us recall the work by Kim et al. [15] which measure nine indicators of woman empowerment, which are combined in one value called the risk ratio and we standardize it for our purposes in the interval $[-1, 1]$ where one implies a high level of empowerment and minus one means a high level of submission. However, it is important to say there exist in Mexican women's Institute (Inmujeres) local scales and measurement for the Mexican population such as the one reported in [8]. Finally, let discuss the coefficient θ in the function $h(t)$ for the alcohol consumption which corresponds to the man's alcohol consumption frequency. This is given by the number of drinking days per week. This data can be easily found for different cases in the ENVIM [2]. The scale that we consider is $\rho \in [0, 1]$, where 0, means a zero alcohol consumption and 1 implies an excessive one. Let summarize all this information in Table 1.

Table 1. Informational table of the variables and parameters of the model.

Variable or Coefficient	Description	Scale	Reference
$v(t)$	violence index	$[0, 30]$	[20,21]
$i(t)$	independence index	$[0, 1]$	[22]
α_1, α_2	violence in childhood	$[0, 12]$	[17,23]
β_1, β_2	acceptance of machismo	$[0, 1]$	[24,25]
δ_1, δ_2	coping styles	several	[27,28]
ρ_1, ρ_2	self-regulatory coefficients	$[0, 8]$	[16]
γ	man's self-esteem	Rosenberg $[0, 40]$	[26]
μ	external factors	risk ratio	[15]

Once, we have completely described the model we analyze it in the next section.

3. Analysis of the Model without Alcohol Consumption

In this section, we present a classical and numerical analysis of Model (2). In order to accomplish this, we first analyze the model when there is not alcohol consumption. Later, we numerically analyze the complete model by considering alcoholism events.

Let us first assume that the function $h(t) \equiv 0$ because we are considering that the man does not consume alcohol periodically. This leads to an autonomous model. Thus, we start by determining the equilibrium points of the system (2) by making the derivatives equal to zero $\frac{dv}{dt} = 0, \frac{di}{dt} = 0$, and we solve for the model variables,

$$v^* = \frac{\delta_1\delta_2\tilde{v} + \delta_2 k_1(1-\rho_1)(\beta_1/\gamma)\tilde{i}}{\delta_1\delta_2 - k_1 k_2(1-\rho_1)(1-\rho_2)(\beta_1/\gamma)\mu}, \quad i^* = \frac{\delta_1\delta_2\tilde{i} + \delta_1 k_2(1-\rho_2)\mu\tilde{v}}{\delta_1\delta_2 - k_1 k_2(1-\rho_1)(1-\rho_2)(\beta_1/\gamma)\mu}.$$

A simple inspection of v^* and i^* shows us that when $k_1 = k_2 = 0$, we trivially recover the non-influenced steady state. It is more important to note that the equilibrium point of system (2) is now a linear combination of steady states $\tilde{v}$ and $\tilde{i}$. Observe that the denominator is the same for v^* and i^*, and that it is a difference between the terms $\delta_1\delta_2$ and $k_1 k_2(1-\rho_1)(1-\rho_2)(\beta_1/\gamma)\mu$. The first one of these two expressions is in terms of rate of returning to the non-influenced steady states while the second depends on important factors for both of the intimate partners. Now, we determine the stability of the equilibrium point. For this, we compute the characteristic equation $\lambda^2 - \tau\lambda + \Delta = 0$ associated with the system. The roots of this equation are the eigenvalues of the coefficient matrix of the system and are given by, $\lambda_1 = \frac{1}{2}\tau + \frac{1}{2}\sqrt{\tau^2 - 4\Delta}$ y $\lambda_2 = \frac{1}{2}\tau - \frac{1}{2}\sqrt{\tau^2 - 4\Delta}$. where $\tau = -\delta_1 - \delta_2$ and $\Delta = \delta_1\delta_2 - k_1 k_2(1-\rho_1)(1-\rho_2)(\beta_1/\gamma)\mu$. Let us recall, that an equilibrium point is asymptotically stable if $\tau^2 - 4\Delta \geq 0$ and $\Delta \geq 0$ or $\tau^2 - 4\Delta < 0$. In the case of $\tau^2 - 4\Delta > 0$ and $\Delta < 0$, we have that the equilibrium point is unstable. Let us discuss how these conditions apply to our model.

- **Case 1.** If $\tau^2 - 4\Delta < 0$, we have that $(\delta_1 - \delta_2)^2 < -4k_1 k_2(1-\rho_1)(1-\rho_2)(\beta_1/\gamma)\mu$, then the equilibrium point (v^*, i^*) is asymptotically stable, but the solutions are oscillating. A necessary

condition is that $\mu < 0$ which means that external factors make the woman's independence index decrease. This may be caused by family or social pressure on the woman, or by her beliefs that lead her to a highly dependent state. In these circumstances, the IPV for this couple tends to the stable steady state which indicates woman submission and a constant level of IPV. It is important to remark that the left-hand side of the inequality is the difference between δ_1 and δ_2. Then, we can have this case when the inertia coefficients are very similar or equal, and there is social pressure on the woman. Another way in which this condition may hold corresponds to a high social pressure, low self-regulatory conditions, high acceptance of machismo or very low man's self-esteem.

- **Case 2.** If $\tau^2 - 4\Delta \geq 0$ and $\Delta \geq 0$, or equivalently $(\delta_1 - \delta_2)^2 \geq -4k_1k_2(1 - \rho_1)(1 - \rho_2)(\beta_1/\gamma)\mu \geq -4\delta_1\delta_2$, then the equilibrium point is asymptotically stable and the solutions are not oscillatory. Notice that μ could be positive or negative. For this case, we can find a lot of possible parameter combinations which could lead us to this situation where the intimate partner interaction stabilizes rapidly. That means that the man's violence and the woman's independence indexes reach a stationary level and remain there for a long time. Observe that in this case, the inertias for both of the members of the couple could be similar or different, one from each other. However, independently of this, the level of IPV will be high when there exists high acceptance of machismo, low self-esteem or low self-regulation for the man. Otherwise, the model predicts a moderated level of IPV. Another fact is that depending on the μ value, the level of the woman's independence will stabilize on a level of independence or dependency.
- **Case 3.** If $\tau^2 - 4\Delta > 0$ and $\Delta < 0$ or equivalently, $\delta_1\delta_2 < k_1k_2(1 - \rho_1)(1 - \rho_2)(\beta_1/\gamma)\mu$, then the equilibrium point is unstable and it is necessary that $\mu > 0$ for the condition holds. Notice that the condition $\mu > 0$ means that external factors are leading to an increment in the woman's independence index. In particular, this could be achieved through empowerment programs for the woman. Then, our model reproduces the reported fact that this type of programs can lead to break the IPV cycle. But, it is important to remark that other situations must also occur. For example, if the coefficients δ_1 and δ_2 are relatively high, the coefficients ρ_1 and ρ_2 have to be close to zero, and β_1 high, or γ small. In any case, we will have a steep climb on the IPV or the woman's independence.

Let us recall that in the performed analysis we are considering no alcohol consumption. This will be addressed numerically in the next section.

4. Numerical Simulations

Now, we present simulations for the solutions of our model for different scenarios without and with man's alcohol consumption. Let us say that the simulations where obtained using synthetic data for the parameters because we have no found any reported case where all parameters of the our model have been determined for a particular couple. However, we can define different parameter sets that corresponds to different situations or scenarios.

4.1. Model without Alcohol Consumption

In this subsection, we present different scenarios and its simulations of Model (2) for cases without man's alcohol consumption. Before starting the discussion, let us explain the structure used in the figures. In each figure we can observe two graphs as well as the full set of parameters. The left-hand graphs are the phase portrait where only the indexes are shown. Also, we have marked with a star ($*$) the non-influenced steady state point and with a triangle the influenced equilibrium point. For the right-hand side graphs, we have pictured the solutions as functions of time, $v(t)$ and $i(t)$ with time t in weeks. The choice of the time units is with the aim to compare the solutions with the alcohol consumption case, which is reported in weeks.

Scenario 1 Couple with a profile of moderated IPV. In this case, we consider a moderated violent man and a non-independent woman. We are assuming that both of the members of the couple have

observed violence during their childhood, and they accept machismo as a lifestyle. This determines the non-influenced steady state. The rest of the parameters are such that the man's self-esteem is medium as well as the self-regulatory level. With respect to the woman, its self-regulatory level is a medium level, and she is experiencing social pressure to continue in the relationship. Let us recall that for this case $h(t) = 0$ because the man does not consume alcohol regularly. In Figure 2, we can observe that the solutions stabilize after an oscillatory behavior which corresponds to case 1, previously discussed. Also, that under these conditions the man's violence index increases while the woman's independence index reduces when it is compared with non-influenced steady state. This reduction usually is caused by a decrease in the woman's self-esteem which could be explained as a consequence of the woman suffering IPV. On the other hand, let us remark that the maximum value of the violence index is 0.6. If this value is extrapolated to the original scale, it corresponds to aggressions described as shaking and pulling to mistreat to the woman. For the level at which the violence index stabilize, we can say a daily man's behavior corresponds to violent acts such as rough touching or destruction of personal belongings.

Figure 2. Violent man without alcohol consumption and submissive woman. $\alpha_1 = 0.7, \alpha_2 = 0.6, \beta_1 = 0.6, \beta_2 = 0.5, \gamma = 0.5, \delta_1 = 0.2, \delta_2 = 0.25, \rho_1 = 0.5, \rho_2 = 0.5, \mu = -0.2, (v(0), i(0)) = (0.4, 0.2)$.

Scenario 2 Couple with a profile of moderated IPV and empowered woman. For this case, the parameters were set with values similar to the values of Scenario 1. Both of the members of the couple have observed a slightly medium level violence during their childhood, and they have a medium acceptance level of machismo. The man's self-esteem is medium while that his self-regulatory parameter is lesser than the corresponding woman's parameter value. The main change is that the woman empowerment is high as a consequence of being involved in empowerment programs. In Figure 3, we can observe that the solutions for both of the indexes increase rapidly. This is because the man responds violently to the increment in the woman's independence. Let us recall that the man's self-esteem is related to the man's need to control the woman. For this case, violence reaches values so high that they escape of any real situation of IPV. Then, we can interpret this situation as a scenario where the woman faces severe violence perpetration (eventually the woman's death when the violence index exceeds one). A more optimistic interpretation is that the relationship breaks because the woman's independence increases faster than the IPV. It is important to remark that when the machismo acceptance or the man's self-esteem is higher, the IPV grows even faster than the woman's independence. Notice that influenced steady state cannot be reached by the solutions because it is non-stable.

Figure 3. Violent man without alcohol consumption, empowerment woman. $\alpha_1 = 0.5, \alpha_2 = 0.6, \beta_1 = 0.3, \beta_2 = 0.3, \gamma = 0.6, \delta_1 = 0.25, \delta_2 = 0.25, \rho_1 = 0.5, \rho_2 = 0.5, \mu = 0.6, (v(0), i(0)) = (0.4, 0.2)$.

4.2. Model with Alcohol Consumption

Now, we numerically analyze the model including the term that models the effect of alcoholism in the couple.

Scenario 3 Couple with a moderated profile of IPV and man's alcohol consumption. In this case, we take the same parameter set of Scenario 1 and include the effect of the periodic alcohol consumption by the man ($h(t) \neq 0$), i.e., we are considering $\rho = 0.25, \omega = \pi/2, \eta = 0.35$. In Figure 4, we can observe that the model with man's periodic alcohol consumption follows the qualitative model behavior without alcohol consumption. In this case, the violence index increases and the independence index decreases, but the solution oscillates. Let us observe that, there is a time interval where the oscillations are changing in amplitude (transient behavior) to finally reach a periodic solution. Note the that maximum value of the violence index almost reach one, which corresponds to assassination attempt according to the IPN scale. Finally, the solution for $v(t)$ oscillates in an approximate range of $[0.5, 0.8]$. When these values are interpreted according to IPV scale, they correspond to psychological and physical violent acts ranging from controlling the woman action, destroying personal belongings and death threats among others. We also note that there is the displacement of the steady state point, when it is compared with the case without alcohol consumption. This means that the man's violence index and the woman's dependency index increases, confirming that alcohol worsens the IPV perpetration. This agrees with the behavior reported one by the literature [2].
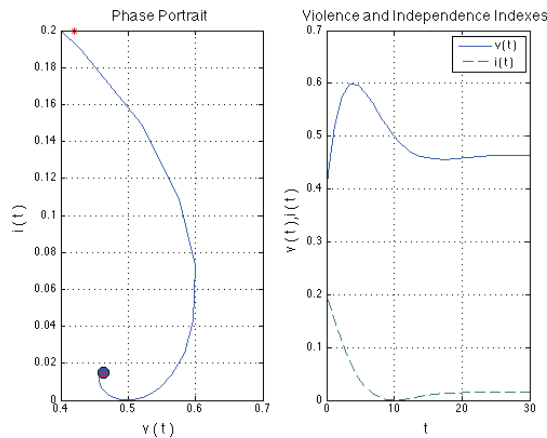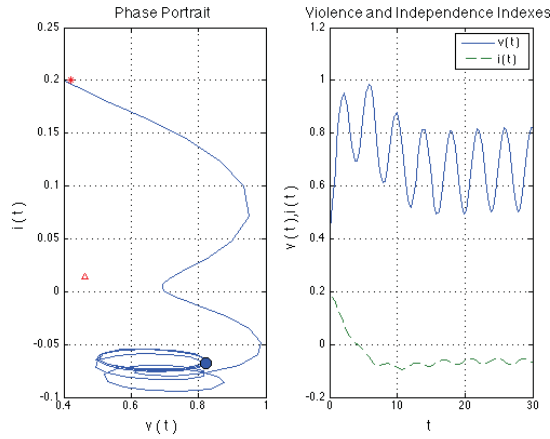
Figure 4. Violent man with alcohol consumption, non-independent woman. $\alpha_1 = 0.7, \alpha_2 = 0.6, \beta_1 = 0.6, \beta_2 = 0.5, \gamma = 0.5, \delta_1 = 0.2, \delta_2 = 0.25, \rho_1 = 0.5, \rho_2 = 0.5, \mu = -0.2, \rho = 0.25, \omega = \pi/2, \eta = 0.35$.

Scenario 4 Couple with a profile of empowered woman and man's alcohol consumption. In this case, we take the parameter set of Scenario 2 and include the effect of the periodic alcohol consumption by the man ($h(t) \neq 0$), i.e., we add the parameters values $\rho = 0.15, \omega = \pi/2, \eta = 0.35$. The results of this simulation are an oscillating increment in both of the indexes which is the same qualitative behavior found in the scenario 2, but with oscillations. See Figure 5. This even occurs more rapidly than in the case with a non-alcoholic man showing the emphasizing effect of periodic alcohol consumption.

Certainly, these all cases are undesirable and intervention strategies to avoid these situations are a must and will be discussed in the next section.
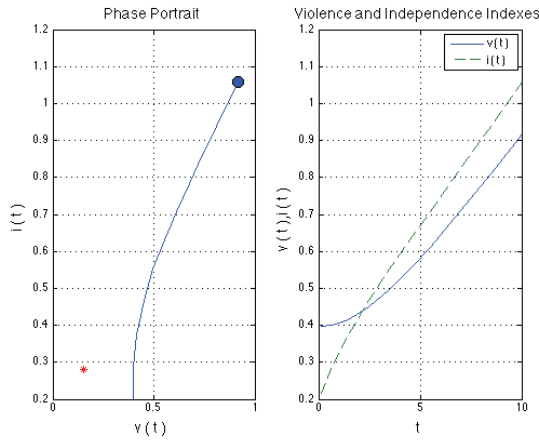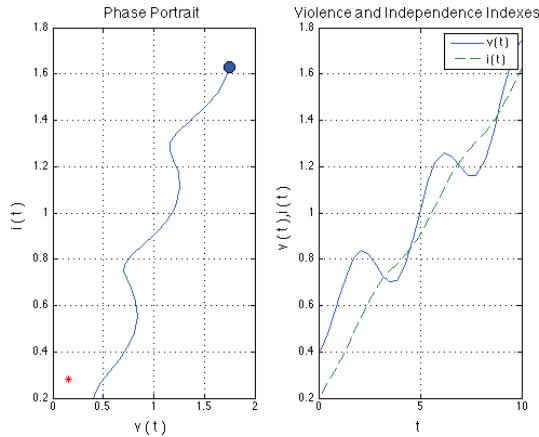

Figure 5. Violent man with alcohol consumption, empowered woman. $\alpha_1 = 0.5, \alpha_2 = 0.6, \beta_1 = 0.3, \beta_2 = 0.3, \gamma = 0.6, \delta_1 = 0.25, \delta_2 = 0.25, \rho_1 = 0.5, \rho_2 = 0.5, \mu = 0.6, \rho = 0.25, \omega = \pi/2, \eta = 0.35$.

5. Discussion of the Model and Its Application

We have constructed a mathematical model based on measures of the human behavioral traits for the IPV dynamics in a romantic couple. Based on this model, it is very important to analyze possible applications. Also as a next step, we may ask about possible intervention strategies to reduce or to

avoid violence. In order to do this, it is important to remark that our model reproduces a small number of reported situations of IPV which are boosted by alcoholism, considering only two variables. Notice, that despite it is very restrictive with respect to the IPV dynamics found in a romantic relationship, we can draw some interesting conclusions.

According to our model, there are two key parameters to change the IPV dynamics in a couple. The first one is the parameter ρ_1, which measures the self-regulatory factor of the IPV perpetrator. Certainly, the reason for introducing this factor is supported by the reported results in [16]. Especially by the fact that it can be used as an IPV predictor. In such a work, the authors argue that violent man impulses do not necessarily end in violent behavior. Moreover, this parameter value or human behavioral trait can be increased by providing the violence perpetrator with techniques to refrain from violent acts. Then, a possible intervention strategy could be to encourage this type of training. However, let us analyze the limit case $\rho_1 = 1$, recall that this parameter is normalized. Our model implies that woman independence does not contribute to increasing the violence in the couple. Then, the model suggests that violent behavior would tend to the non-influenced steady state which is not a zero violence state. Therefore, it is not enough to focus on the education of the perpetrator, but on the factors defining his intrinsic level of violence such as therapies to reduce the trauma caused by the violence watched in childhood. Additionally, educational programs to change man's mind with respect to the machismo are needed.

Another possible strategy of intervention can be understood in terms of the empowering coefficient, μ. By allowing it to take positive or negative values gives the opportunity to change the model dynamics. As a first scenario, with negative μ, there are factors encouraging the woman to remain married, despite the level of violence that the woman is experiencing. Certainly, this simulates a cruel reality that must be avoided. As a first stage to achieve this, the woman must be enrolled in empowerment programs, ie. setting μ positive as a first step to break the violence cycle. violence caused by man's need for control. One may suggest stressing the level of empowerment programs, but according to the model it will lead to increase the IPV. The suggested solution to escape IPV is to set conditions that ensure a more rapidly increment in the independence index with respect to violence increment. This is a consequence of the linearity of the model and points out the need to provide a more balanced model. Such improved model should reproduce more mild IPV scenarios. Also, it would be important to determine conditions where the violence can decrease while the woman independence increase. After that, it would follow the addition of modeling terms for the intervention strategies such as the couple therapy and public policy strategies. Both of them, with a short term impact in the reduction of the IPV problem.

6. Conclusions

In this work, we developed a model for the IPV perpetration in a romantic couple in terms of a system of linear differential equations. This model describes the cycle of violence in one couple by considering important causal and motivational factors. Among those factors, we consider that the most relevant is the inclusion of the need to control woman, the self-regulatory coefficients and the external factors for the woman. From our point of view, our contribution is the identification of important factors that trigger IPV and the corresponding use in setting a first differential equation model. Despite the simplicity of our model, it allowed us to reproduce different IPV scenarios and identify possible intervention strategies. In future work, we would like to explore the inclusion in the model of more realistic assumptions such as a variable self-regulatory coefficients, the inclusion of nonlinear terms to describe interactions. In particular those related with intervention strategies In this way, the model will reproduce more complex IPV scenarios and provide more targeted actions to reduce this social problem.

References

1. WHO Multi-Country Study on Women's Health and Domestic Violence against Women: Summary Report of Initial Results on Prevalence, Health Outcomes and Women's Responses. Available online: https://www.who.int/reproductivehealth/publications/violence/24159358X/en/ (accessed on 1 March 2019).
2. INEGI. Encuesta Nacional Sobre la Dinámica de las Relaciones en los Hogares. 2003. Available online: semujer.zacatecas.gob.mx/pdf/libros/Encuesta%20nacional%20sobre%20la%20dinamica%20de%20las%20relaciones%20en%20los%20hogares%202003%20ENDIREH.pdf (accessed on 1 March 2019).
3. Caprioli, M.; Hudson, V.M.; McDermott, R.; Ballif-Spanvill, B.; Emmett, C.F.; Stearmer, S.M. The Womanstats Project Database: Advancing an Empirical Research Agenda. *J. Peace Res.* **2009**, *46*, 839–851. [CrossRef]
4. Programme, U.N.D.; Malik, K. Human Development Report 2014: Sustaining Human Progress-Reducing Vulnerabilities and Building Resilience. Available online: http://hdr.undp.org/en/content/human-development-report-2014 (accessed on 1 March 2019).
5. Wiley, S.A.; Levy, M.Z.; Branas, C.C. The impact of violence interruption on the diffusion of violence: A mathematical modeling approach. In *Advances in the Mathematical Sciences*; Springer: Berlin, Germany, 2016; pp. 225–249.
6. Bourne, D.A. *Mathematical Modeling of Pharmacokinetic Data*; Routledge: Abingdon, UK, 2018.
7. DeAngelis, D.L. *Individual-Based Models and Approaches in Ecology: Populations, Communities and Ecosystems*; CRC Press: Boca Raton, FL, USA, 2018.
8. Sánchez, J.E.H.; Falconi, R.G. *Instrumento para Medir el Empoderamiento de la Mujer*; Universidad Juárez Autónoma de Tabasco: Villahermosa, Mexico, 2008.
9. Poza, E.; Jódar, L.; Barreda, S. Mathematical modeling of hidden intimate partner violence in Spain: A quantitative and qualitative approach. *Abstr. Appl. Anal.* **2016**, *2016*, 1–8. [CrossRef]
10. Gottman, J.; Swanson, C.; Murray, J. The mathematics of marital conflict: Dynamic mathematical nonlinear modeling of newlywed marital interaction. *J. Fam. Psychol.* **1999**, *13*, 3. [CrossRef]
11. Gottman, J.M. *The Mathematics of Marriage: Dynamic Nonlinear Models*; MIT Press: Cambridge, MA, USA, 2005.
12. Walker, L.E. *The Battered Woman*; Harper & Row: New York, NY, USA, 1980.
13. Corsi, J.; Dohmen, M.L.; Sotes, M.A. *Violencia Masculina en la Pareja: Una Aproximación al Dignóstico y A los Modelos de Intervención*; Paidos Argentina: Buenos Aires, Argentina, 1995.
14. Stith, S.M.; Farley, S.C. A predictive model of male spousal violence. *J. Fam. Violence* **1993**, *8*, 183–201. [CrossRef]
15. Kim, J.C.; Watts, C.H.; Hargreaves, J.R.; Ndhlovu, L.X.; Phetla, G.; Morison, L.A.; Busza, J.; Porter, J.D.; Pronyk, P. Understanding the impact of a microfinance-based intervention on women's empowerment and the reduction of intimate partner violence in South Africa. *Am. J. Public Health* **2007**, *97*, 1794–1802. [CrossRef] [PubMed]
16. Finkel, E.J.; DeWall, C.N.; Slotter, E.B.; Oaten, M.; Foshee, V.A. Self-regulatory failure and intimate partner violence perpetration. *J. Personal. Soc. Psychol.* **2009**, *97*, 483. [CrossRef] [PubMed]
17. Olaiz, G.; Rico, B.; Del Río, A. *Encuesta Nacional sobre Violencia contra las Mujeres 2003*; Instituto Nacional de Salud Pública: Cuernavaca Mor., Mexico, 2004.
18. Carson, A.T.; Baker, R.C. Psychological correlates of codependency in women. *Int. J. Addict.* **1994**, *29*, 395–407. [CrossRef] [PubMed]
19. Thompsn, M.P.; Basile, K.C.; Hertz, M.F.; Sitterle, D. *Measuring Intimate Partner Violence Victimization and Perpetration: A Compendium of Assessment Tools*; National Center for Injury Prevention and Control: Atlanta, GA, USA, 2006.
20. Valdez-Santiago, R.; Híjar-Medina, M.C.; Salgado de Snyder, V.N.; Rivera-Rivera, L.; Avila-Burgos, L.; Rojas, R. Escala de violencia e índice de severidad: una propuesta metodológica para medir la violencia de pareja en mujeres mexicanas. *Salud Pública de México* **2006**, *48*, s221–s231. [CrossRef] [PubMed]

21.   Tronco Rosas, M.A. No sólo Ciencia y Tecnología. Ahora, el IPN a la Vanguardia en Perspectiva de género. El Programa Institucional de Gestión con Perspectiva de Género. Available online: http://www.genero. ipn.mx/Conocenos/Documents/MemoriaPIGPG.pdf (accessed on 1 March 2019).
22.   Garcia, D.; Weber, I.; Garimella, V.R.K. Gender Asymmetries in Reality and Fiction: The Bechdel Test of Social Media. In Proceedings of the 8th International AAAI Conference on Weblogs and Social Media, Ann Arbor, MI, USA, 1–4 June 2014; pp. 131–140.
23.   Gil-González, D.; Vives-Cases, C.; Ruiz, M.T.; Carrasco-Portino, M.; Álvarez-Dardet, C. Childhood experiences of violence in perpetrators as a risk factor of intimate partner violence: A systematic review. *J. Public Health* **2007**, *30*, 14–22. [CrossRef] [PubMed]
24.   Arciniega, G.M.; Anderson, T.C.; Tovar-Blank, Z.G.; Tracey, T.J. Toward a Fuller Conception of Machismo: Development of a Traditional Machismo and Caballerismo Scale. *J. Couns. Psychol.* **2008**, *55*, 19–33. [CrossRef]
25.   Bendezú, A. Los estereotipos de género y el riesgo del embarazo adolescente. Ph.D. Thesis, UNAM, México City, México, 1998.
26.   Robins, R.W.; Hendin, H.M.; Trzesniewski, K.H. Measuring Global Self-Esteem: Construct Validation of a Single-Item Measure and the Rosenberg Self-Esteem Scale. *Personal. Soc. Psychol. Bull.* **2001**, *27*, 151–161. [CrossRef]
27.   Sinclair, V.G.; Wallston, K.A. The Development and Psychometric Evaluation of the Brief Resilient Coping Scale. *Assessment* **2004**, *11*, 94–101. [CrossRef] [PubMed]
28.   Calleja, N. Escalas Psicosociales en México. Available online: http://www.psicologia.unam.mx/ documentos/pdf/repositorio/InventarioEscalasPsicosocialesNaziraCalleja.pdf (accessed on 1 March 2019).

A Hidden Markov Model for the Linguistic Analysis of the Voynich Manuscript

Luis Acedo

Instituto Universitario de Matemática Multidisciplinar, Building 8G, 2º Floor, Camino de Vera,
Universitat Politècnica de València, 46022 Valencia, Spain; luiacrod@imm.upv.es; Tel.: +34-963877007 (ext. 88285)

Abstract: Hidden Markov models are a very useful tool in the modeling of time series and any sequence of data. In particular, they have been successfully applied to the field of mathematical linguistics. In this paper, we apply a hidden Markov model to analyze the underlying structure of an ancient and complex manuscript, known as the Voynich manuscript, which remains undeciphered. By assuming a certain number of internal states representations for the symbols of the manuscripts, we train the network by means of the α and β-pass algorithms to optimize the model. By this procedure, we are able to obtain the so-called transition and observation matrices to compare with known languages concerning the frequency of consonant andvowel sounds. From this analysis, we conclude that transitions occur between the two states with similar frequencies to other languages. Moreover, the identification of the vowel and consonant sounds matches some previous tentative bottom-up approaches to decode the manuscript.

1. Introduction

Hidden Markov models (HMMs) are a particular kind of a Bayesian network obtained by combining a Hidden Markov layer and a second layer of outputs that depends probabilistically on the hidden states of the first layer [1,2]. This model has proven a very versatile and useful tool in many applications of artificial intelligence, among others: (i) modeling of biological sequences of proteins and DNA [3]; (ii) speech recognition systems [4]; (iii) data compression and pattern recognition [5]; and (iv) object tracking in video sequences [6]. Of particular interest to us are the early studies in which HMMs were used to analyze a large body of text, in that case of English (the so-called "Brown Corpus"), considered as a sequence of letters without any previous assumption on the linguistic structure of the text or the meaning of the letters [1,4,7]. Depending on the number of hidden states, thanks to this work, light was shed on the linguistic structure of English in the model. For example, for two hidden states, the basic division among vowels and consonants was recovered as the most natural basic pattern of the English language [7]. As many more states were taken into account, it was discovered a structure including the initial and final letters of a word, vowel followers and preceders, etc. This elucidates the purely statistical nature of a language and it shows that HMMs can be an insightful tool in mathematical and computational linguistics.

Applications of HMMs to the field of Natural Language Processing (NLP) has also flourished in recent years as its has been shown for different layers of NLP such as speech tagging and morphological analysis. By using this approach, successful results for many languages such as Arabic and Persian have been obtained [8,9]. For these reasons, it seems promising to extend these analyses to other sources of text that still cannot be deciphered because they are written in an unknown script and with a unique linguistic structure. Among the candidates to this challenge, the medieval codex known as Voynich manuscript stands out [10].

Discovered by the Polish–Lithuanian book dealer W. Voynich in 1912, it has remained an enigma of historical cryptography since then. For a detailed introduction to the manuscript's history and attempts of decipherment until 1978, the interested reader can find more information in M. d'Imperio's monograph [11] and on Zandbergen's website [10]. It is generally believed by history researchers that this book could have belonged to emperor Rudolph II of Baviera until his death in 1612. This was stated in a letter addressed to the XVIIth century scholar Athanasius Kircher that was found by Voynich himself inside the manuscript. The history of the ownership of the manuscripthas been elucidated throughout the years. It is also known that it was property of the Jesuits and it was kept at the "Collegio Romano" since the last decade of the XVIIth century until the end of the XIXth century when it was moved to Frascati where Voynich acquired it.

Modern physics and chemistry analyses have allowed establishing some rigorous facts. Firstly, in 2009, some samples of ink and paint were taken from the manuscript and analyzed by X-ray spectroscopy and X-ray diffraction techniques showing that these inks and pigments were totally compatible with those used by scribes at the last epoch of the Middle Ages [12]. The same year, a radiocarbon dating of the parchment was carried out by researchers at the University of Arizona [13]. They found that with a 95% probability the parchment corresponds to the period between 1404 and 1438. This places the manuscript in the first half of the fifteen century. It is also clear that the text was added after the drawing of the figures in the manuscript because it usually surrounds these figures very closely.

An image of the first line in the Voynich manuscript is shown in Figure 1. The total set of individual characters depends on some ambiguities in counting but it seems that there are 36 characters in the set as recognized by the Currier alphabet, some of them far more frequent than other. Others alphabets, such as the European Voynich Alphabet (EVA), consider only 25 characters (ignoring the rare ones). Although the symbols seem strange, and not immediately associated with any known alphabet ancient or modern, a closer inspection reveals a similarity with Latin, Arabic numerals and, specially, some Latin abbreviations very common throughout the Middle Ages [14]. Anyway, these clues have helped little in finding an accepted decipherment of the text.
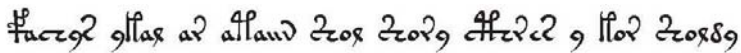
Figure 1. A sample of the first folio of the Voynich manuscript.

Among the possible solutions to this riddle, there have been four main proposals:

1. It is a manuscript written in an extinct natural language with an exotic alphabet [15].
2. It is the encipherment of a known language (possibly Latin, German or other Indo-European language but nobody is sure [11]).
3. It is a hoax consisting of asemic writing with the objective of making the book strange and valuable to collectors of antiquities [16].
4. It is a modern fabrication (perhaps by its discoverer, W. Voynich) [10].

From these hypotheses, the last one seems excluded by modern physicochemical analyses but the other three may still be considered open. The paper is organized as follows. In Section 2, we discuss the basics of Hidden Markov Models and its application to linguistic analysis. Section 3 is devoted to the application of HMMs to the Voynich manuscript and the information we may deduce from this. Finally, the paper ends with a discussion on the meaning of the findings of the paper and guidelines for future work in Section 4.

2. Hidden Markov Models

In this section, we provide a quick summary of the basic concepts and algorithms for HMMs. In Figure 2, we show the structure of a HMM. The Markov process (Figure 2) is characterized by a

sequence $\{X_0, X_1, X_2, \ldots, X_{T-1}\}$ of internal states selected among a total of N. The transition among these states is performed according to the probabilities of a transition matrix A in such a way that the element a_{ij} denotes the probability of performing a transition from the internal i state to the state j. We can also denote the different internal states as q_i, with $i = 0, 1, 2, \ldots, q_{N-1}$.

The second layer we have plotted in Figure 2 corresponds to the observations. The sequence of observations is then denoted by $\{\mathcal{O}_0, \mathcal{O}_1, \ldots, \mathcal{O}_{T-1}\}$ and they can be chosen among a total of M possible observation symbols. The relation between the Markov process layer and the observation layer is also probabilistic because, given an internal state q_j, the probability for observing the symbol k is $b_j(k)$. These elements constitute a row stochastic matrix, B.
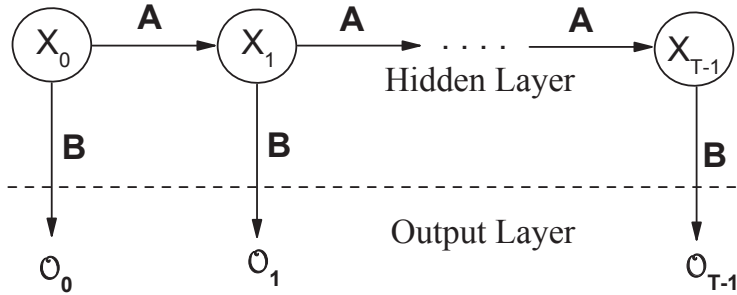
Figure 2. A schematic view of a HMM. See the main text for details.

The problem we want to address is the following: Suppose we have a given sequence of observations, $\mathcal{O}$, consisting of a series of symbols from a total of M. If we assume that there are N internal states in the model, the objective is to find the model (A, B, π) (where π is the distribution of initial states) that provides the best fit of the observation data. The standard technique in HMMs to evaluate this optimum model makes use of a forward algorithm and a backward algorithm described as follows. In the description of the algorithms, we closely follow the pedagogical presentation by Stamp [1].

2.1. The Forward Algorithm

Firstly, we are interested in evaluating the following probability:

$$\alpha_t(i) = P(\mathcal{O}_0, \mathcal{O}_1, \ldots, \mathcal{O}_t, x_t = q_i | \lambda), \tag{1}$$

i.e., the probability that the sequence of observations up to time t is given by $\mathcal{O}_0, \ldots, \mathcal{O}_t$ and the internal state at time t is q_i for a given model λ. Then, for $t = 0$, we have that:

$$\alpha_0(i) = \pi_i\, b_i(\mathcal{O}_0), \tag{2}$$

where $i = 0, 1, \ldots, N - 1$. The reason is that π_i is the probability that the initial internal state is q_i, and $b_i(\mathcal{O}_0)$ is the probability that, given that the internal state is q_i, we have the observation $\mathcal{O}_0$. It is the easy to check that the recursion expression for $\alpha_t(i)$ is given by:

$$\alpha_t(i) = \left[\sum_{j=0}^{N-1} \alpha_{t-1}(j)\, a_{ji}\right] b_i(\mathcal{O}_t). \tag{3}$$

Here, a_{ji} is the transition probability from the inner state j to the inner state i. This algorithm is also called α-pass.

2.2. The Backward Algorithm

The backward algorithm, or β-pass, is proposed for the efficient evaluation of the following probability:

$$\beta_t(i) = P(\mathcal{O}_{t+1}, \mathcal{O}_{t+2}, \dots, \mathcal{O}_{T-1}, x_t = q_i | \lambda) \,. \tag{4}$$

This means that we are interested in finding the probability that the sequence of observations from time $t+1$ to the end is $\mathcal{O}_{t+1}, \dots, \mathcal{O}_{T-1}$ and the inner state at time t is q_i. The algorithm is constructed as follows:

- In the first place, we define $\beta_{T-1}(i) = 1$ for $i = 0, 1, \dots, N-1$.
- Then, for $t = T - 2, T - 3, \dots, 0$ we define the recursive relation:

$$\beta_t(i) = \sum_{j=0}^{N-1} a_{ij} b_j(\mathcal{O}_{t+1}) \beta_{t+1}(j) \,. \tag{5}$$

We use these two algorithms, the forward and the backward, to find the standard algorithm that can be used to reestimate the model and make it approach its optimum.

2.3. Reestimating The Model

Our objective now is to reevaluate the model in such a way that the optimum parameters that fit the observations are found. These parameters are the elements of the matrices A and B as well as those in the vector corresponding to the initial distribution of internal states, π. We need now an algorithm to reestimate the model in such a way that the probability of the observation sequence, $\mathcal{O}_0, \dots, \mathcal{O}_{T-1}$, given the model λ, $P(\mathcal{O}_0, \dots, \mathcal{O}_{T-1} | \lambda)$ is maximized.

The idea of the algorithm begins with the definition of the following probability for a given model and a given observation sequence:

$$\gamma_t(i, j) = P(x_t = q_i, x_{t+1} = q_j | \mathcal{O}, \lambda) \,. \tag{6}$$

This is the probability of finding the internal states q_i and q_j at times t and $t+1$ for the observation sequence $\mathcal{O}$ and the model λ. Using now the standard relations for conditional probabilities and the definitions of the α and β probabilities in Equations (1) and (4), we have:

$$\gamma_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(\mathcal{O}_{t+1}) \beta_{t+1}(j)}{P(\mathcal{O} | \lambda)} \,, \tag{7}$$

for time $t = 0, 1, \dots, T-2$. We also define the sum over the index j, i.e., the probability of finding the inner state q_i at time t for a given model and observation sequence:

$$\gamma_t(i) = \sum_{j=0}^{N-1} \gamma_t(i, j) \,. \tag{8}$$

With these definitions and expressions, we can now propose the evolution algorithm for the reestimation of the parameters:

- We initialize the model $\lambda = (A, B, \pi)$. It is a common practice to choose the elements according to the uniform distribution: $\pi_i \approx 1/N$, $a_{ij} \approx 1/N$, and $b_j(k) \approx 1/M$ but these values must be randomized to avoid that the algorithm becomes stuck at a local maximum.
- We calculate the parameters $\alpha_t(i)$, $\beta_t(i)$, $\gamma_t(i, j)$ and $\gamma_t(i)$ by applying the corresponding expressions in Equations (3), (5), (7) and (8).

- For $i = 0, 1, \ldots, N - 1$ and $j = 0, 1, \ldots, N - 1$ we reestimate the elements of the transition matrix, A, as follows:

$$a_{ij} = \frac{\sum_{t=0}^{T-2} \gamma_t(i,j)}{\sum_{t=0}^{T-2} \gamma_t(i)} \, , \tag{9}$$

- For $i = 0, 1, \ldots, N - 1$ and $j = 0, 1, \ldots, N - 1$ we compute the new values for the elements of the observation probability matrix as follows:

$$b_j(k) = \frac{\sum_{t \in \{0,1,\ldots,T-1\}, \, \mathcal{O}_t = k} \gamma_t(j)}{\sum_{t=0}^{T-1} \gamma_t(j)} \, . \tag{10}$$

Here, the sum in the numerator is restricted to those instants of time in which the observation symbol is the kth.
- Finally, we compute the probability of the given observation sequence, i.e., $P(\mathcal{O}|\lambda)$ (obtained as the sum of $\alpha_{T-1}(i)$ for all the inner state values, i). If this probability increases (with respect to the previous value), the model updating is performed again. However, in practice, the algorithm is run for a given number of steps or until the probability does not increases more than a selected tolerance.

Another issue with this algorithm is that the α-pass and β-pass evaluations may easily lead to underflow. To avoid this problem, a normalization by the sum over j of $\alpha_t(j)$ is performed. For further details on HMMs the interested reader may check the references [1].

2.4. Applications to Linguistics of HMM and Other Network Models

Before the application of this particular model to the examples in the next section, we briefly review some relevant studies concerning the use of networks in linguistics. Although computational linguistics is as old as computers themselves, the field of natural language processing (NLP) started to take off in the 1980s with the development of statistical machine translation [17], machine learning algorithms and neural networks [18]. More recently, NLP has received an important advance within the broader field of Deep Learning [19]. Hidden Markov Models (HMM) also played a role in these developments: Baum and Petrie in their seminal paper laid the foundations of the theory of HMM in 1966 [20], which took form in the late 1960s and early 1970s [21,22]. The forward–backward algorithm described in this section was proposed in this early studies and it is usually known as the Baum–Welch algorithm. As mentioned before, in 1980, Cave and Neuwirth applied HMMs to study the structure of a set of texts in English (known as "Brown Corpus"), which allowed them to derive useful conclusions about the role of individual letters [1,7]. Word alignment was also studied in connection with statistical translation by using a HMM by Vogel et al. [23]. More recent applications of HMM to linguistics include part-of-speech tagging, i.e., the labeling of the different words according to their grammatical category [8,9]. The Baum–Welch algorithm has also been used to identify spoken phrases in VoIP calls [24]. HMM are used in speech recognition since the seminal work of Baker in 1975 [25]. Nowadays, Long Short-Term Deep Neural Networks are the current paradigm in this field due to their capacity tospot long term dependencies [26]. Networks have also proven very useful in the development of the so-called word adjacency model, as shown by Amancio [27] and Nebil et al. [28]. This paradigm has been applied to the classification of artificially generated manuscripts in contrasts to genuine ones by observing their topological properties [29]. The dynamics of word co-occurrence in networks has also allowed identifying the author of a given manuscript [30]. It is known that this adjacency model captures, mainly, syntactical features of texts [31]. The algorithm to find the optimum model for HMM is also oriented to identify the basic syntactical features concerning the probability sequences of individual letters. Nothing can be said about the semantics, although it can certainly help in a Rosetta

stone approach to decode a particular manuscript written in an unknown language, such as the Voynich manuscript [15]. The adjacency model has also been used to analyze the Voynich manuscript [32] with similar conclusions to the ones deduced in this paper: that the manuscript is mostly compatible with a natural language instead of a random text.

3. Results

In this section, we discuss the application of the algorithm discussed in the previous section to several cases. First, we consider the case of a text in English and we implement the model optimization algorithm to classify the letters of the alphabet (after removing all the punctuation signs) into two classes corresponding to the inner states of the HMM. It is shown that these classes are clearly associated with the vowels and the consonants in English and this provides the basic phonemic structure of the language. Testing the algorithm with a known language gives us the necessary confidence to apply it to the Voynich manuscript.

In both applications (to a text in English and to the Voynich manuscript), we only used two hidden states. This raises the question about the advisability of this particular choice, instead of a larger number of hidden states. As the main objective was to show that these texts can be partitioned in different sets of symbols that are different in their statistical properties, selecting $N = 2$ seems the simpler choice. Moreover, in earlier applications to other books (such as the "Brown Corpus", which it is a compilation of roughly one million words with texts ranging from science to literature or religion), this choice was proven to be successful in the identification of the vowels and consonants [7]. More hidden states were considered by Cave and Neuwirth in their seminal application of HMM to language and they even obtained some conclusions for $N = 3$ to $N = 12$. Nevertheless, this was done with the advantage of dealing with a known language. Proving the existence of, at least, two different sets of letters in the Voynich manuscript is already a useful conclusion. If we take into account that the alphabet and the language are completely unknown in this case, this could help linguistics in their research to unveil some meaning in the words of the manuscript. Only after some globally accepted success is achieved in this endeavor the study of the convergence of the HMM model for $N > 2$ could be done with some possibility of interpretation.

3.1. Application to The Quixote

We applied the model evolution algorithm for HMM with $N = 2$ and $M = 27$. Therefore, we considered 26 letters and the space as output symbols. The text of the Quixote in plain ASCII can be freely downloaded from the Gutenberg's project website [33]. This is the English translation of the original Spanish version. As a data pre-processing stage, we transformed all the upper-case letters to lower-case and removed the punctuation signs with the exception of the spaces among subsequent words. This way, we obtained a sequence of 5,693,310 characters but, to our purpose, we restricted ourselves to the first 100,000 characters.

As initial transition matrix, we chose:

$$A = \begin{pmatrix} 0.46 & 0.54 \\ 0.52 & 0.48 \end{pmatrix}, \tag{11}$$

and the distribution of initial hidden states was given at the start of the algorithm by

$$\pi = \begin{pmatrix} 0.52 \\ 0.48 \end{pmatrix}, \tag{12}$$

The observation probability matrix was obtained by randomizing the equal probability assumption: $b_j(\mathcal{O}) = 1/M$ for every j. For example, we could multiply $1/M$ by a random number

in the interval $(0.8, 1.2)$. However, we must verify the condition that the total probability for a given inner state j and all the possible observation outcomes is normalized to one:

$$\sum_{\text{all } \mathcal{O}} b_j(\mathcal{O}) = 1 \, , \tag{13}$$

and this was accomplished by imposing this condition to define the last value $b_j(\mathcal{O})$ for the Mth state, $\mathcal{O}$.

The algorithm was then implemented in Mathematica using lists and the model evolution was run for 200 steps with 100,000 characters from the book, pre-processed to retain only the letters and removing all the punctuation signs (see the supplementary material accompanying this paper for Mathematica code). Results were also checked using other independent implementations of the code in C++ [34].

Firstly, we noticed that the algorithm converged after around 100 iterations, as deduced from the evolution of the logarithm of the observation sequence probability for the given model $P(\mathcal{O}|\lambda)$. This is shown in Figure 3. However, the convergence may seem peculiar because this probability only started raising after iteration 50 and settled to a "plateau" in a few iterations after that. This could be an indication of a small basin of attraction for the fixed point we were looking for. The topography of the landscape for this particular problem would require further investigation.
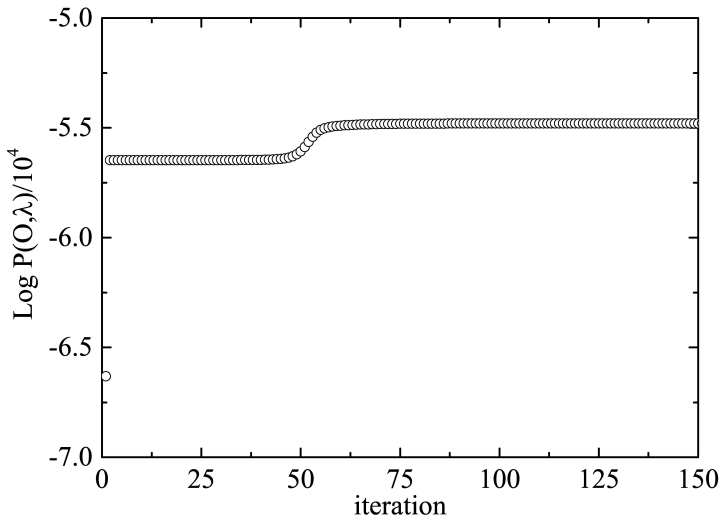
Figure 3. The evolution of the logarithm of the observation sequence's probability as a function of the iteration. Notice the fast convergence to an asymptotic "plateau".

From the results in this figure, we conclude that convergence was achieved after only 60 steps and that further iterations only improve the results very slightly. The final transition matrix we found is given as follows:

$$A = \begin{pmatrix} 0.369 & 0.631 \\ 0.869 & 0.131 \end{pmatrix} , \tag{14}$$

with an initial distribution of initial states given by:

$$\pi = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$ (15)

Transition matrices with large off-diagonal elements are also found using other texts such as the "Brown Corpus" [7]. This compilation contains a million words with texts ranging from religion to science, including also some novels. Consequently, the conclusions we derived from a single book are already supported for a variety of texts in English from the early study of Cave and Neuwirth [7]. The form of the transition matrix in Equation (14) is already pointing towards the existence of two categories of letters (the inner states of the HMM). These categories are, obviously, the vowels and the consonants and this statement is reinforced by the values of the observation probability matrix, B. In our case, the inner state 2 has been identified as the vowels, as clearly shown in Figure 4 where a peak of probability is found for every vowel.
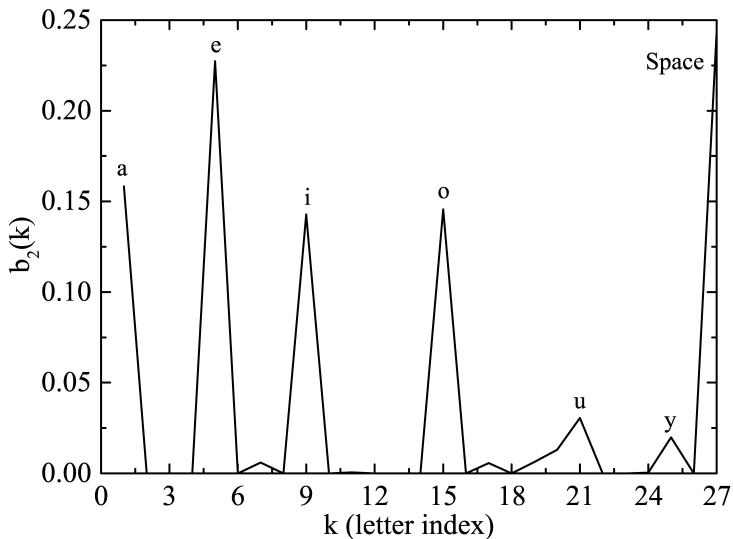
Figure 4. The probability of observation of a given letter for the hidden state 2. Notice the peaks for the vowels as well as "y" and the space.

From the results in Figure 4, we could also derive some interesting conclusions:

1. The most frequent vowel in the English language is "e".
2. The space among words has the structural function of a vowel, although it has no associated sound.
3. The letter "y" is mostly a vowel in the English language. Indeed, the Oxford Dictionary classifies it as a vowel in some cases ("myth"), a semivowel in others ("yes") or forming a diphthong (as in "my") [35].

In rigor, these conclusions would be valid only for the Quixote but they have also been checked for other books (in particular the "Brown Corpus" book compilation). Thus, it is not unjustified to apply them to the English language in general. Similar conclusions can also be deduced for the consonants. It is also remarkable that the results in Figure 4 cannot be deduced from simple frequency analysis. The histogram of the letters obtained from the Quixote's text is shown in Figure 5.
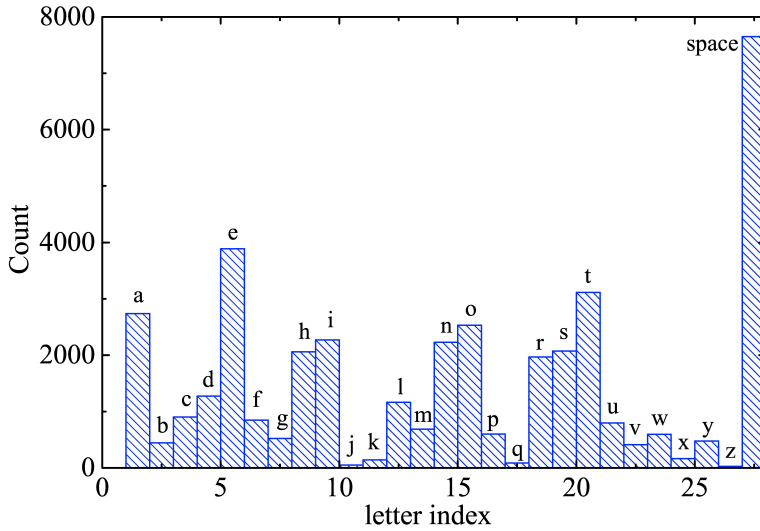
Figure 5. The histogram for the frequency of the totality of letters (and the space) in the English version of the Quixote.

We see that the vowel "e" is still the most frequent letter in the English language (not only the most frequent vowel) but the second most frequent letter is "t". Thus, there is no clear pattern in the histogram to separate vowels from consonants. This ability of HMM make them a very powerful tool in computational linguistics.

Finally, we should emphasize that this experiment was performed by starting with particular forms of the transition matrix, A, and the initial distribution vector, π, as given in Equations (11) and (12). The observation probability matrix was also chosen in a random way as described above. It is possible that, in some cases, the evolution algorithm for the HMM may find a local minima and be sensitive to initial values instead of converging to the global minima we are looking for. Indeed, a solution with A defined as a 2-by-2 identity matrix is always stable and can be approached by certain initial conditions. For these reasons, it is convenient to check the sensitivity of our results to initial conditions by performing several runs of the program. To check the convergence of the algorithm, we chose different initial conditions, in which A and π were randomly selected as follows:

$$A = \begin{pmatrix} x & 1-x \\ y & 1-z \end{pmatrix}, \tag{16}$$

$$\pi = \begin{pmatrix} z \\ 1-z \end{pmatrix}, \tag{17}$$

where x, y, and z are random real numbers in the interval $(0.4, 0.6)$. After performing 10 simulations with 200 runs, we found that the average transition matrix is given by:

$$A = \begin{pmatrix} 0.300(63) & 0.700(63) \\ 0.701(61) & 0.299(61) \end{pmatrix}. \tag{18}$$

Notice that both the matrices in Equations (14) and (18), are unstable as the off-diagonal elements are larger than the diagonal ones. However, this stationary state of the algorithm is coherent with previous results obtained with other texts [7]. This means that two hidden states can be identified in the symbols of the book (corresponding to vowels and consonants) and that the transitions take place

mainly from consonants to vowels and vice versa. This is precisely what happens in natural languages. A transition matrix equal, or approximately equal, to the identity matrix would not separate the states and it is not relevant to this linguistic application.

3.2. Application to the Voynich Manuscript

After the successful implementation of the HMM technique for the Quixote, we thus turned again to our problem of analyzing the Voynich manuscript. Several transcriptions of this manuscript are available but the most popular is the one based upon the so-called European Voynich Alphabet (EVA), as developed by Zandbergen and Landini. Although there is an extended version, which includes the less common symbols in the Voynich manuscript, the basic version uses 26 letters of the English alphabet (excluding "w") to make a correspondence with the most abundant symbols in Voynichese. The correspondence among the EVA code and the Voynich symbol is given in the table of Figure 6.

At this point, it is also necessary to explain why we chose this particular alphabet instead of the other alternatives. The main reasons are its popularity and the fact that many transcriptions are available for it. Otherwise, some specialists would argue that some symbol combinations in this alphabet, such as "ch" and "sh"(corresponding to the so-called "pedestals") should be considered as one single character each [10]. On the other hand, the combinations "in" and "iin" are also candidates for representing letters, although, in some other cases, "i" could be a single character. This is another problem that computational analyses could help to solve [36].

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	x	y	z
ɑ	∂	⌐	8	c	℔	ʂ	ᴢ	﹒	8	℔	℈	℥	℈	o	℔	4	ʔ	ʔ	ℍ	⌀	∧	⊼	9	℞

Figure 6. The correspondence among Voynich's symbols and the associated letter in the EVA transcription systems. Notice that this is merely an arbitrary codification without any relation to the actual phonemes that the Voynich's characters might represent.

This way, a transcription of the whole Voynich manuscript has been performed in such a way that it can be used in computational analysis. These transcriptions are available in several sites, as discussed by Zandbergen [10]. In particular, we used Takahashi's transcription developed in 1999. Of course, some pre-processing was required before applying the HMM algorithm because this file includes some information about each line, including the folium number (recto or verso) and the number of the line within each page of the manuscript. After removing this information, we were left with a set of EVA characters separated by dots. These dots correspond to the spaces between words in the original manuscript. The total number of characters (including spaces) we used for the simulation is 228,836. Convergence of the algorithm was also very fast, as in the case of the English text analyzed in Section 3.1, and, when we used more than 50,000 characters, the results were stable and showed no dependence on the total length of the sequence, T. This is a convincing argument in favor of the consistency of the results.

We started with the same initial conditions as those given in Equations (11) and (12) for the transition matrix and the distribution of the state $t = 0$. The probability matrix for the observation states (the Voynich characters) was randomized in the usual way explained in Section 3.1. In this particular example, we used the first 100,000 characters in the Voynich manuscript and 200 iteration steps. The final transition matrix is similar to those of natural languages (because the off-diagonal elements are larger than the diagonal ones):

$$A = \begin{pmatrix} 0.169 & 0.831 \\ 0.840 & 0.160 \end{pmatrix}, \tag{19}$$

and the distribution of initial states was given by:

$$\pi = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{20}$$

After performing the same randomization of initial conditions as we saw for the example of the Quixote, we found an average transition matrix of the form:

$$A = \begin{pmatrix} 0.182(45) & 0.818(45) \\ 0.831(17) & 0.169(17) \end{pmatrix}, \tag{21}$$

which confirms that the result in Equation (19) is close to the asymptotic fixed point we were looking for.

We can also see that Equation (20) indicates that the first character of the manuscript is associated with a consonant sound. The most interesting results are, however, those obtained with the observation probability matrix, which clearly separate two kinds of characters to be associated with vowel and consonant phonemes, as occurred in the case of the Quixote. In Figure 7, we show the probability of obtaining each character when the hidden state is 1.
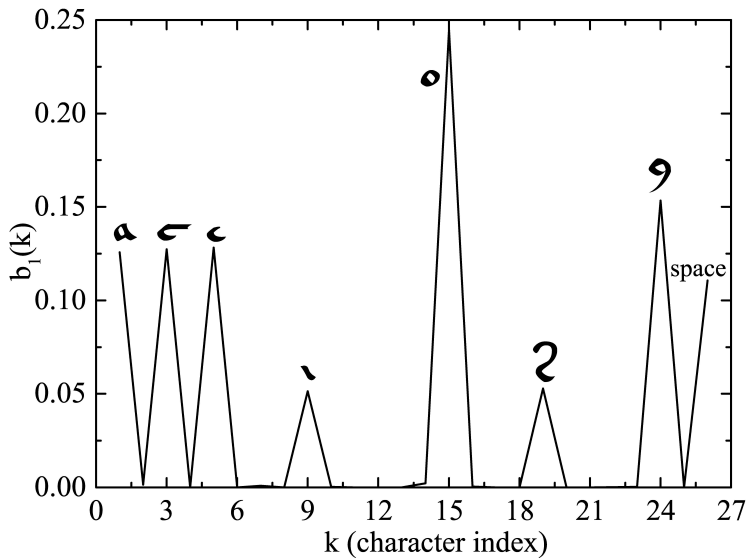
Figure 7. The probability of observation of a given character in the Voynich manuscript for the hidden state 1. The peaks correspond to the symbols: "a", "c", "e", "i", "o", "s" and "y" of the EVA alphabet. There is also a peak for the space among words.

The probabilities for hidden state 2 are given in Figure 8. We see that a set of very conspicuous peaks were obtained in both cases, but there were fewer in the result shown in Figure 7, which could mean that the symbols corresponding to those peaks are associated with vowel's phonemes. On the other hand, this correspondence is not as strong as in the case of the English text of Section 3.1 because there are symbols with noticeable probability that appear in both figures (in particular, the EVA symbols "e", "i", "s" and "y"). Perhaps, the most simple explanation for the absence of a clear separation among vowels and consonants in these four cases is that we are confronted with another example of letters that can function as both vowels and consonants as in the case of "y" in English. However, for the Voynich manuscript, we have four letters with this capacity and this is a peculiarity whose meaning we cannot unravel for the moment. Another possibility is that the Voynich alphabet is some kind of abjad but this hypothesis is not so clear because in abjads the letters that appear in the

text are always a vowel or a consonant (although some vowels are left out). This could have an effect in the transition matrix but it is not evident that it would produce an ambiguity in the classification of those letters in the hidden states, one or two. Further research into HMM applied to languages with abjad's alphabets, such as Arabic, should be necessary to clarify this point.
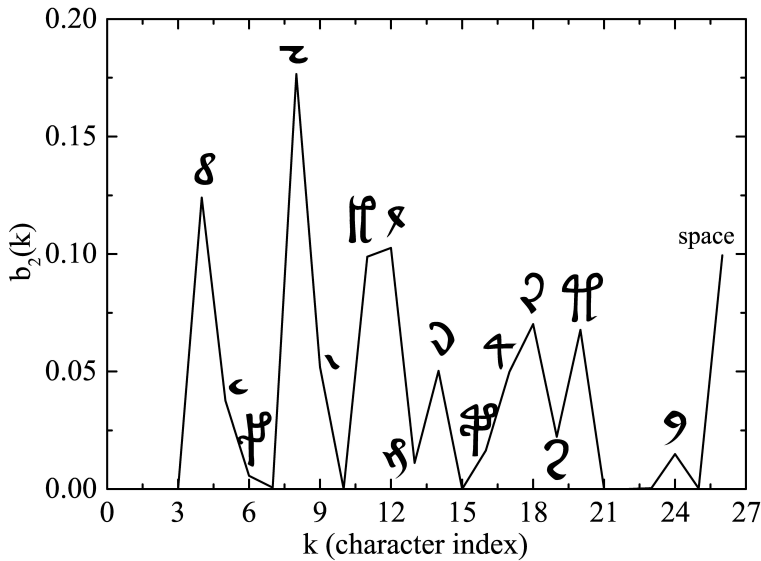
Figure 8. The same as Figure 7 but for the hidden state 2.

Nevertheless, the appearance of high peaks in Figure 7 also suggests the existence of vowel symbols in the manuscript so this abjad would be impure. Bax also suggested in 2014 [15] that this is the case after proposing a provisional decoding of some words in the manuscripts associated with plants and a constellation. In the next section, we discuss this intriguing possibility.

4. Discussion

Although a lot of research effort has been devoted to understanding the Voynich manuscript, we still lack a consensual opinion about the nature of the writing it contains. Early efforts by Friedman, Tiltman, Currier and Bennett, among others [10,11] helped to elucidate some features of the manuscript and exclude the hypothesis of a simple substitution cipher. The interest in the application of mathematical and computational methods to the Voynich manuscript is maintained nowadays, as shown by recent publications [32,37].

The main result of this paper is the evidence of the existence of two states associated with the symbols in the manuscript that could correspond to consonants and vowels. This reaffirms to the conclusions of Amancio et al. [32] and Montemurro et al. [37] of the existence of word co-occurrences and keywords in the same proportion than natural languages.

The recent proof of the old age of the "vellum" by carbon-dating places this manuscript in the first half of the XVth century and all other tests are very compatible with an origin in the Middle Ages [12,13]. Anyway, although a forgery by Voynich or other recent author is excluded, some authors support the idea that this object was fabricated in the Middle Ages with a pecuniary intention by making it similar, in appearance, to a real, but enciphered, text [16]. As said before, statistical measurements carried out by Amancio et al. in 2013 [32] show that the Voynich manuscript is incompatible with shuffled texts and, moreover, that certain keywords appear throughout the manuscript. These keywords organize in patterns of semantic networks, as shown by Montemurro et al. [37]. These conclusions, and those

of this paper, make the hypothesis of a fake more unlikely but a clever fake designed to simulate a human language with the adequate transition among letters cannot completely be excluded.

Other authors, as most researchers in the past, think that the text is enciphered in some way. Hauer and Kondrak assumed that the manuscript was written in some abjad's alphabet using some transposition of letters or anagramming [38]. These assumptions are very strong and their conclusions of a relation to Hebrew have been widely dismissed.

In this paper, we use the conservative approach of applying the standard technique of HMM for the linguistic analysis of the manuscript. We have shown that a division among vowels and consonant phonemes is very clear in the resulting observation probability matrix but that some characters (such as the EVA symbols "e", "i", "s" and "y") could participate in the vowel and consonant natures, either because they are semivowels or because there is an implicit vowelling in them as in impure abjad's alphabets. A positional dependence of the vowelling as in abugida's alphabets cannot be excluded. In abugida's alphabets, some vowels are written but, sometimes, they are attached as part of the consonant, which makes the analysis more difficult. Further research with HMM into languages with these alphabets would be necessary to obtain a reliable conclusion.

In any case, we found some characters in the script that very possibly represent vowel's phonemes: "a", "c", "e", "o" and "y" in the EVA notation (see Figure 6). There are also some exclusively consonant phonemes, such as "d", "h", "k", "l" or "t" in the EVA's notation.

The idea of using HMM to analyze the Voynich manuscript has already been proposed by other authors but their results have not been published officially in scientific journals. For example, Zandbergen discussed his application of HMM to the manuscript on a webpage [36]. His conclusions are similar to those found in this present paper. Another study was performed by Reddy and Knight [39]. However, they indicated that their algorithm separates vowels and consonants in an odd fashion because one of the states always corresponds to the last symbol in a letter. Anyway, their explanations are cursory and they did not describe their algorithm in enough detail to allow for a reproduction of their results. Much earlier, D'Imperio wrote a paper about the use of an algorithm called PTAH at the National Security Agency [40]. Although we can infer from her discussion that the algorithm is some kind of HMM and that the results obtained from a five-state model are compatible with our results, it is difficult to compare her algorithm with the one discussed here because PTAH is still classified. On the other hand, implementations with a higher-state HMM could provide useful information if the linguistics provide some agreement about the possible translation of some words.

It also seems interesting that Prof. Bax in 2014 [15] identified some names of plants and the constellation Taurus in the manuscript and that his associations with phonemes are similar to the one discussed in this paper. In Figure 9, we show some of these associations for the words "Taurus", "Coriander" and "Juniper" (in Arabic).

Voynich's Word	Phonetics' transcription
o ? o ?	/a/ r /a/ r
8 o a ? 9	T /a/ /ᵊ/ /r/ (plus vowel) N
ℍ c c ? o 8 a ꭓ	K O O R A T /ᵊ/ ?

Figure 9. Some words in the Voynich manuscript and the phonetics transcriptions proposed by Bax [15].

Although these associations are considered very preliminary, even by its author, the similarity with our identification of vowels and consonants is striking. Thus, we can gain some confidence that a serious scholarship effort could enhance these identifications and provide a sure path for further research. In any case, we must stress that the identifications performed by Bax are not generally accepted but, at least, they show that our mapping of vowels and consonants allows for some meaningful words to appear.

Although other possibilities might still be open, we have increasing support of the view that the text in the manuscript is neither a hoax nor an intentional cipher but a genuine language written in an unknown script. Notwithstanding this progress, we are still far from identifying the language because it could even be a dead tongue, for which the script was devised. Further analysis including more states could help if the decoding of the manuscript adds further evidence to these preliminary conclusions. We hope that this work stimulates further research by expert linguists that could shed additional light into this ancient enigma.

Abbreviations

The following abbreviations are used in this manuscript:

EVA   European Voynich Alphabet
HMM   Hidden Markov Models
NLP   Natural Language Processing

References

1. Stamp, M. A Revealing Introduction to Hidden Markov Models. Available online: http://www.cs.sjsu.edu/faculty/stamp/RUA/HMM.pdf (accessed on 19 January 2019).
2. Ghahramani, Z. An Introduction to Hidden Markov Models and Bayesian Networks. *Int. J. Pattern Recognit. Artif. Intell.* **2001**, *15*, 9–42. [CrossRef]
3. Yoon, B.J. Hidden Markov Models and their Applications in Biological Sequence Analysis. *Curr. Genom.* **2009**, *10*, 402–415. [CrossRef] [PubMed]
4. Juang, B.H.; Rabiner, L.R. Hidden Markov Models for Speech Recognition. *Technometris* **1991**, *33*, 251–272. [CrossRef]
5. Bicego, M.; Castellani, U.; Murino, V. Using Hidden Markov models and wavelets for face recognition. In Proceedings of the 12th International Conference on Image Analysis and Processing, Mantova, Italy, 17–19 September 2003.
6. Lefèvre, S.; Bouton, E.; Brouard, T.; Vincent, N. A new way to use Hidden Markov Models for object tracking in video sequences. In Proceedings of the 2003 International Conference on Image Processing, Barcelona, Spain, 14–18 September 2003.
7. Cave, R.L.; Neuwirth, L.P. Hidden Markov Models for English. In *Hidden Markov Models for Speech*; IDA-CRD: Princeton, NJ, USA, 1980. Available online: https://www.cs.sjsu.edu/~stamp/RUA/CaveNeuwirth/index.html (accessed on 19 January 2019).
8. Suleiman, D.; Awajan, A.; Al Etaiwi, W. The Use of Hidden Markov Model in Natural ARABIC Language Processing: A Survey. *Proc. Comput. Sci.* **2017**, *113*, 240–247. [CrossRef]
9. Okhovvat, M.; Bidgoli, B.M. A Hidden Markov Model for Persian Part-of-Speech Tagging. *Proc. Comput. Sci.* **2011**, *3*, 977–981. [CrossRef]
10. Zandbergen, R. The Voynich Manuscript. Available online: http://www.voynich.nu (accessed on 19 January 2019).
11. D'Imperio, M.E. *The Voynich Manuscript: An Elegant Enigma*; National Security Agency, Central Security Service: Maryland, MD, USA, 1978.
12. Repp, K. Materials Analysis of the Voynich Manuscript. Available online: https://beinecke.library.yale.edu/sites/default/files/voynich_analysis.pdf (accessed on 19 January 2019).

13. Zandbergen, R. The Radio-Carbon Dating of the Voynich MS. Available online: http://www.voynich.nu/extra/carbon.html (accessed on 19 January 2019).

14. Capelli, A. The Elements of Abbreviation in Medieval Latin Paleography. Translated by Heimann, D. and Kay, R. University of Kansas Libraries, 1982. (Translation of the original, Lexicon abbreviaturarum, published in 1899). Available online: https://kuscholarworks.ku.edu/bitstream/handle/1808/1821/47cappelli.pdf (accessed on 19 January 2019).

15. Bax, S. A Proposed Partial Decoding of the Voynich Script. Available online: https://stephenbax.net/wp-content/uploads/2014/01/Voynich-a-provisional-partial-decoding-BAX.pdf (accessed on 19 January 2019).

16. Rugg, G.; Taylor, G. Hoaxing statistical features of the Voynich Manuscript. *Cryptologia* **2017**, *41*, 247–268. [CrossRef]

17. Koehn, P. *Statistical Machine Translation*; Cambridge University Press: Cambridge, UK, 2009; p. 27.

18. Goldberg, Y. A Primer on Neural Network Models for Natural Language Processing. *J. Artif. Intell. Res.* **2016**, *57*. [CrossRef]

19. Deng, L.; Liu, Y. *Deep Learning in Natural Language Processing*; Springer: Singapore, 2018.

20. Baum, L.E.; Petrie, T. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *Ann. Math. Stat.* **1966**, *37*, 1554–1563. [CrossRef]

21. Baum, L.E.; Eagon, J.A. An Inequality with Applications to Statistical Estimation for Probabilistic Functions of a Markov Process and to a Model for Ecology. *Bull. Am. Math. Soc.* **1967**, *73*, 360–363. [CrossRef]

22. Baum, L.E. A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *Ann. Math. Stat.* **1970**, *41*, 164–171. [CrossRef]

23. Vogel, S.; Ney, H.; Tillmann, C. HMM-Based Word Alignment in Statistical Translation. Available online: http://aclweb.org/anthology/C96-2141 (accessed on 19 January 2019).

24. Wright, C.; Ballard, L.; Coull, S.; Monrose, F.; Masson, G. Spot Me If You Can: Uncovering Spoken Phrases in Encrypted VoIP Conversations. Available online: https://ieeexplore.ieee.org/document/4531143/authors#authors (accessed on 19 January 2019).

25. Baker, J.K. The DRAGON system—An overview. *IEEE Trans. Acoust. Speech Signal Process.* **1975**, *23*, 24–29. [CrossRef]

26. Graves, A.; Mohamed, A.-R.; Hinton, G. Speech Recognition with Deep Recurrent Neural Networks. Available online: https://arxiv.org/abs/1303.5778 (accessed on 19 January 2019).

27. Amancio, D.R. Probing the Topological Properties of Complex Networks Modeling Short Written Texts. *PLoS ONE* **2015**, *10*, e0118394. [CrossRef] [PubMed]

28. Nebil, O.; Malod-Dognin, N.; Davis, D.; Levnajic, Z.; Janjic, V.; Karapandza, R.; Stojmirovic, A.; Pržulj, N. Revealing the Hidden Language of Complex Networks. *Sci. Rep.* **2014**, *4*, 4547.

29. Amancio, D.R. Comparing the topological properties of real and artificially generated scientific manuscripts. *Scientometrics* **2015**, *105*, 1763–1779. [CrossRef]

30. Akimushkin, C.; Amancio, D.R.; Oliveira, O.N., Jr. Text Authorship Identified Using the Dynamics of Word Co-Occurrence Networks. *PLoS ONE* **2017**, *12*, e0170527. [CrossRef] [PubMed]

31. De Arruda, H.F.; Marinho, V.Q.; Costa, L.d.F.; Amancio, D.R. Paragraph-Based Complex Networks: Application to Document Classification and Authenticity Verification. Available online: https://arxiv.org/abs/1806.08467 (accessed on 19 January 2019).

32. Amancio, D.R.; Altmann, E.G.; Rybski, D.; Oliveira, O.N., Jr.; Costa, L.D.F. Probing the Statistical Properties of Unknown Texts: Application to the Voynich Manuscript. *PLOS ONE* **2013**, *8*, e67310. [CrossRef] [PubMed]

33. Gutenberg Project. The Quixote by Miguel de Cervantes Saavedra. Available online: http://www.gutenberg.org/ebooks/996 (accessed on 19 January 2019).

34. An implementation in C++ of the HMM algorithm developed by M. Stamp. Available online: http://www.cs.sjsu.edu/faculty/stamp/RUA/HMM_ref_fast.zip (accessed on 19 January 2019).

35. Is the Letter "Y" a Vowel or a Consonant ? Available online: https://en.oxforddictionaries.com/explore/is-the-letter-y-a-vowel-or-a-consonant/ (accessed on 19 January 2019).

36. Zandbergen, R. What We May Learn from the MS Text Entropy. Available online: http://www.voynich.nu/extra/sol_ent.html (accessed on 19 January 2019).

37. Montemurro, M.A.; Zanette, D.H. Keywords and Co-Occurrence Patterns in the Voynich Manuscript: An Information-Theoretic Analysis. *PLoS ONE* **2013**, *8*, e66344. [CrossRef] [PubMed]

38. Hauer, B.; Kondrak, G. Decoding Anagrammed Texts Written in an Unknown Language and Script. *Trans. Assoc. Comput. Linguist.* **2016**, *4*, 75–86. [CrossRef]
39. Reddy, S.; Knight, K. What we know about the Voynich Manuscript. In Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, Portland, OR, USA, 24 June 2011; pp. 78–86.
40. D'Imperio, M. An Application of PTAH to the Voynich Manuscript (U). *Natl. Secur. Agency Tech. J.* **1979**, *24*, 65–91. Available online: https://www.nsa.gov/Portals/70/documents/news-features/declassified-documents/tech-journals/application-of-ptah.pdf (accessed on 19 January 2019).

Some Notes to Extend the Study on Random Non-Autonomous Second Order Linear Differential Equations Appearing in Mathematical Modeling

Julia Calatayud Gregori [ID]**, Juan Carlos Cortés López** [ID] **and Marc Jornet Sanz *** [ID]

Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València, 46022 Valencia, Spain; jucagre@doctor.upv.es (J.C.G.); jccortes@imm.upv.es (J.C.C.L.)
* Correspondence: marjorsa@doctor.upv.es

Abstract: The objective of this paper is to complete certain issues from our recent contribution (Calatayud, J.; Cortés, J.-C.; Jornet, M.; Villafuerte, L. Random non-autonomous second order linear differential equations: mean square analytic solutions and their statistical properties. *Adv. Differ. Equ.* **2018**, *392*, 1–29, doi:10.1186/s13662-018-1848-8). We restate the main theorem therein that deals with the homogeneous case, so that the hypotheses are clearer and also easier to check in applications. Another novelty is that we tackle the non-homogeneous equation with a theorem of existence of mean square analytic solution and a numerical example. We also prove the uniqueness of mean square solution via a habitual Lipschitz condition that extends the classical Picard theorem to mean square calculus. In this manner, the study on general random non-autonomous second order linear differential equations with analytic data processes is completely resolved. Finally, we relate our exposition based on random power series with polynomial chaos expansions and the random differential transform method, the latter being a reformulation of our random Fröbenius method.

Keywords: random non-autonomous second order linear differential equation; mean square analytic solution; random power series; uncertainty quantification

MSC: 34F05; 60H10; 60H35; 65C20; 65C30

1. Introduction

The important role played by differential equations in dealing with mathematical modeling is beyond discussion. They are powerful tools to describe the dynamics of phenomena appearing in a variety of distinct realms, such as engineering, biomedicine, chemistry, social behavior, etc. [1–3]. In this paper, we concentrate on a class of differential equations that have played a distinguished role in a variety of applications in science, in particular in physics and engineering, namely second order linear differential equations. Indeed, these equations have been successfully applied to describe, for example, vibrations in springs (free undamped or simple harmonic motion, damped vibrations subject to a frictional force, or forced vibrations affected by an external force), the analysis of electric circuits made up of an electromotive force supplied by a battery or a generator, a resistor, an inductor, and a capacitor. In the former type of problems, second order linear differential equations are formulated by applying Newton's second law, while in the latter case, this class of equations appears via the application of Kirchhoff's voltage law. In these examples, the formulation of second order linear differential equations to describe the aforementioned physical problems appears as direct applications of important laws of physics. However, this class of equations also arises indirectly when solving significant partial differential equations in physics. In this regard, Airy, Hermite, Laguerre, Legendre, or Bessel differential equations are non-autonomous second order linear differential equations that

emerge in this way. For example, the Airy equation appears when solving Schrödinger's equation with triangular potential and for a particle subject to a one-dimensional constant force field [4]; the Hermite equation emerges in dealing with the analysis of Schrödinger's equation for a harmonic oscillator in quantum mechanics [5]; the Laguerre equation plays a main role in quantum mechanics for the study of the hydrogen atom via Schrödinger's equation using radial functions [6] (Ch. 10); the Legendre equation appears when solving the Laplace equation to compute the potential of a conservative field such as the space gravitational potential using spherical coordinates [7]; and finally, the Bessel equation is encountered, for example, when solving the Helmholtz equation in cylindrical or spherical coordinates by using the method of the separation of variables [6] (Ch. 9).

In all the previous examples, two important features can be highlighted to motivate our subsequent analysis. First, the coefficients of the differential equations are analytic (specifically polynomials). Second, these coefficients depend on physical parameters that, in practice, need to be fixed after measurements; therefore, they involve uncertainty. Both facts motivate the study of random non-autonomous second order linear differential equations, whose coefficients and initial conditions are analytic stochastic processes and random initial conditions, respectively. The aim of our contribution is to advance the analysis (both theoretical and practical) of this important class of equations. In our subsequent development, we mainly focus on the theoretical aspects of such an analysis with the conviction that it can become really useful in future applications where randomness is considered in that class of differential equations. In this sense, some numerical experiments illustrating and demonstrating the potentiality of our main findings are also included. The study of random non-autonomous second order linear differential equations has been carried out for particular cases, such as Airy, Hermite, Legendre, Laguerre, and Bessel equations (see [8–13], respectively), and the general case [14–17]. Alternative approaches to study this class of random/stochastic differential equations include the so-called probabilistic transformation method [18] and stochastic numerical schemes [19,20], for example.

For the sake of clarity in the presentation, the layout of the paper is as follows: In Section 2, we study the homogeneous random non-autonomous second order linear differential. The analysis includes a result (Theorem 2) that simplifies the application of a recent finding by the authors that is particularly useful in practical cases. This issue is illustrated via several numerical examples where both the random Airy and Hermite differential equations are treated. In Section 3, the analysis is extended to a random non-autonomous second order linear differential equation with a forcing term. In this study, we have included conditions under which the solution is unique in the mean square sense. This theoretical study is supported with a numerical example as well. Section 4 is addressed to enrich our contribution by comparing the random Fröbenius method proposed in this paper against other alternative approaches widely used in the extant literature, specifically generalized polynomial chaos (gPC) expansions, Monte Carlo simulations, and the random differential transform method. Conclusions and future research lines are drawn in Section 5.

2. Homogeneous Case

We consider the general form of a homogeneous random non-autonomous second order linear differential equation in an underlying complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$:

$$\begin{cases} \ddot{X}(t) + A(t)\dot{X}(t) + B(t)X(t) = 0, \ t \in \mathbb{R}, \\ X(t_0) = Y_0, \\ \dot{X}(t_0) = Y_1. \end{cases} \tag{1}$$

It is assumed that the stochastic processes $A(t)$ and $B(t)$ are analytic at t_0 in the mean square sense [21] (p. 99):

$$A(t) = \sum_{n=0}^{\infty} A_n (t - t_0)^n, \quad B(t) = \sum_{n=0}^{\infty} B_n (t - t_0)^n, \quad t \in (t_0 - r, r_0 + r),$$

with convergence in $L^2(\Omega)$. The terms $A_0, A_1, \ldots$ and $B_0, B_1, \ldots$ are random variables. In fact, they are related to $A(t)$ and $B(t)$, respectively, via Taylor expansions:

$$A_n = \frac{A^{(n)}(t_0)}{n!}, \quad B_n = \frac{B^{(n)}(t_0)}{n!},$$

where the derivatives are considered in the mean square sense. According to the Fröbenius method, we look for a solution stochastic process $X(t)$ also expressible as a mean square convergent random power series on $(t_0 - r, t_0 + r)$:

$$X(t) = \sum_{n=0}^{\infty} X_n (t - t_0)^n.$$

mean square convergence is important, as it allows approximating the expectation (average) and variance (dispersion) statistics of $X(t)$ at each t [21] (Th. 4.2.1, Th. 4.3.1). This is one of the primary goals of uncertainty quantification [22].

Proposition 1 ([21] (Th. 4.2.1, Th. 4.3.1)). *Let $\{Z_n\}_{n=1}^{\infty}$ and Z be second order random variables. If Z_n converges to Z as $n \to \infty$ in $L^2(\Omega)$ (i.e., in the mean square sense), then the expectation and variance of Z can be approximated as follows:*

$$\mathbb{E}[Z(t)] = \lim_{n \to \infty} \mathbb{E}[Z_n(t)], \quad \mathbb{V}[Z(t)] = \lim_{n \to \infty} \mathbb{V}[Z_n(t)].$$

In [17], some auxiliary theorems on random power series were stated and proven: differentiation of random power series in the $L^p(\Omega)$ sense [17] (Th. 3.1) and Mertens' theorem for random series in the mean square sense [17] (Th. 3.2), which generalize their deterministic counterparts.

Proposition 2 (Differentiation of a random power series in the $L^p(\Omega)$ sense [17] (Th. 3.1)). *Let $A(t) = \sum_{n=0}^{\infty} A_n (t - t_0)^n$ be a random power series in the $L^p(\Omega)$ setting ($p \geq 1$), for $t \in (t_0 - r, t_0 + r)$, $r > 0$. Then, the random power series $\sum_{n=1}^{\infty} n A_n (t - t_0)^{n-1}$ exists in $L^p(\Omega)$ for $t \in (t_0 - r, t_0 + r)$, and moreover, the $L^p(\Omega)$ derivative of $A(t)$ is equal to it: $\dot{A}(t) = \sum_{n=1}^{\infty} n A_n (t - t_0)^{n-1}$, for all $t \in (t_0 - r, t_0 + r)$.*

Proposition 3 (Mertens' theorem for random series in the mean square sense [17] (Th. 3.2)). *Let $U = \sum_{n=0}^{\infty} U_n$ and $V = \sum_{n=0}^{\infty} V_n$ be two random series that converge in $L^2(\Omega)$. Suppose that one of the series converges absolutely, say $\sum_{n=0}^{\infty} \|V_n\|_{L^2(\Omega)} < \infty$. Then:*

$$\left(\sum_{n=0}^{\infty} U_n \right) \left(\sum_{n=0}^{\infty} V_n \right) = \sum_{n=0}^{\infty} W_n,$$

where:

$$W_n = \sum_{m=0}^{n} U_{n-m} V_m,$$

and $\sum_{n=0}^{\infty} W_n$ is understood in $L^1(\Omega)$. The series $\sum_{n=0}^{\infty} W_n$ is known as the Cauchy product of the series $\sum_{n=0}^{\infty} U_n$ and $\sum_{n=0}^{\infty} V_n$.

With these two auxiliary results, the main theorem of [17] was stated as follows:

Theorem 1 ([17] (Th. 3.3)). *Let $A(t) = \sum_{n=0}^{\infty} A_n (t - t_0)^n$ and $B(t) = \sum_{n=0}^{\infty} B_n (t - t_0)^n$ be two random series in the $L^2(\Omega)$ setting, for $t \in (t_0 - r, t_0 + r)$, $r > 0$ being finite and fixed. Assume that the initial conditions Y_0 and Y_1 belong to $L^2(\Omega)$. Suppose that there is a constant $C_r > 0$, maybe dependent on r, such that*

$\|A_n\|_{L^\infty(\Omega)} \le C_r/r^n$ and $\|B_n\|_{L^\infty(\Omega)} \le C_r/r^n$, $n \ge 0$. Then, the stochastic process $X(t) = \sum_{n=0}^{\infty} X_n(t - t_0)^n$, $t \in (t_0 - r, t_0 + r)$, where:

$$X_0 = Y_0, \quad X_1 = Y_1, \tag{2}$$

$$X_{n+2} = \frac{-1}{(n+2)(n+1)} \sum_{m=0}^{n} \left[(m+1)A_{n-m}X_{m+1} + B_{n-m}X_m\right], \ n \ge 0, \tag{3}$$

is the unique analytic solution to the random initial value problem (1) *in the mean square sense.*

This theorem is a generalization of the deterministic Fröbenius method to a random framework. As was demonstrated in [17], Theorem 1 has many applications in practice. It supposes a unified approach to study the most well-known second order linear random differential equations: Airy [8], Hermite [9], Legendre [10,11], Laguerre [12], and Bessel [13]. The results established in these articles [8–13] are particular cases of Theorem 1. The main reason why this fact occurs is explained in [17] (Section 3.3): given a random variable Z, the fact that its centered absolute moments grow at most exponentially, $\mathbb{E}[|Z|^n] \le HR^n$ for certain $H > 0$ and $R > 0$, is equivalent to Z being essentially bounded, $\|Z\|_{L^\infty(\Omega)} \le R$.

Notice that Theorem 1 does not require any independence assumption about the random input parameters. Moreover, from Theorem 1, [17] obtained error estimates for the approximation of the solution stochastic process, its mean, and its variance.

Let us see that Theorem 1 may be put in an easier to handle form. We substitute the growth condition on the coefficients $A_0, A_1, \ldots$ and $B_0, B_1, \ldots$ by the $L^\infty(\Omega)$ convergence of the random power series that define $A(t)$ and $B(t)$. In this manner, in practical applications, one does not need to find any constant C_r; see the forthcoming Examples 1–4.

Theorem 2. *Let* $A(t) = \sum_{n=0}^{\infty} A_n(t - t_0)^n$ *and* $B(t) = \sum_{n=0}^{\infty} B_n(t - t_0)^n$ *be two random series in the* $L^\infty(\Omega)$ *setting, for* $t \in (t_0 - r, t_0 + r)$, $r > 0$ *being finite and fixed. Assume that the initial conditions* Y_0 *and* Y_1 *belong to* $L^2(\Omega)$. *Then, the stochastic process* $X(t) = \sum_{n=0}^{\infty} X_n(t - t_0)^n$, $t \in (t_0 - r, t_0 + r)$, *whose coefficients are defined by* (2) *and* (3), *is the unique analytic solution to the random initial value problem* (1) *in the mean square sense.*

Proof. By [11] (Lemma 2.3),

$$\sum_{n=0}^{\infty} \|A_n\|_{L^\infty(\Omega)}|t - t_0|^n < \infty, \quad \sum_{n=0}^{\infty} \|B_n\|_{L^\infty(\Omega)}|t - t_0|^n < \infty,$$

for $t \in (t_0 - r, t_0 + r)$. Thus, for each $0 \le r_1 < r$,

$$\sum_{n=0}^{\infty} \|A_n\|_{L^\infty(\Omega)}r_1^n < \infty, \quad \sum_{n=0}^{\infty} \|B_n\|_{L^\infty(\Omega)}r_1^n < \infty.$$

Since the sequences $\{\|A_n\|_{L^\infty(\Omega)}r_1^n\}_{n=0}^{\infty}$ and $\{\|B_n\|_{L^\infty(\Omega)}r_1^n\}_{n=0}^{\infty}$ tend to zero, they are both bounded by a number $C_{r_1} > 0$:

$$\|A_n\|_{L^\infty(\Omega)} \le \frac{C_{r_1}}{r_1^n}, \quad \|B_n\|_{L^\infty(\Omega)} \le \frac{C_{r_1}}{r_1^n}, \quad n \ge 0.$$

Then, Theorem 1 is applicable with r_1: the stochastic process $X(t) = \sum_{n=0}^{\infty} X_n(t - t_0)^n$ whose coefficients are given by (2) and (3) is a mean square solution to (1) on $(t_0 - r_1, t_0 + r_1)$. Now, since r_1 is arbitrary, we can extend this result to the whole interval $(t_0 - r, r_0 + r)$. $\square$

Notice that we have proven that Theorem 1 from [17] entails Theorem 2. However, the other way around also holds: Theorem 2 implies Theorem 1. Thus, both theorems are equivalent and offer the same information. Indeed, if we assume the hypotheses from Theorem 1, then:

$$\|A_n\|_{L^\infty(\Omega)} r_1^n \le C_r \left(\frac{r_1}{r}\right)^n, \quad \|B_n\|_{L^\infty(\Omega)} \le C_r \left(\frac{r_1}{r}\right)^n,$$

for any $0 \le r_1 < r$, and since $\sum_{n=0}^\infty (\frac{r_1}{r})^n < \infty$, by comparison, we derive that:

$$\sum_{n=0}^\infty \|A_n\|_{L^\infty(\Omega)} r_1^n < \infty, \quad \sum_{n=0}^\infty \|B_n\|_{L^\infty(\Omega)} r_1^n < \infty,$$

which entails that the series of $A(t) = \sum_{n=0}^\infty A_n (t - t_0)^n$ and $B(t) = \sum_{n=0}^\infty B_n (t - t_0)^n$ converge $L^\infty(\Omega)$, by [11] (Lemma 2.3), for $t \in (t_0 - r_1, t_0 + r_1)$. As r_1 is arbitrary, the $L^\infty(\Omega)$ convergence holds for $t \in (t_0 - r, t_0 + r)$. This is exactly the hypothesis used in Theorem 2.

Let us see that Theorem 2 has an easier to handle form by checking the hypotheses in the examples from [17]. We refer the reader to [17] (Section 4) for approximations of the expectation and variance statistics of the solution stochastic process to each one of the examples.

Example 1. *Airy's random differential equation is defined as follows:*

$$\begin{cases} \ddot{X}(t) + At X(t) = 0, \ t \in \mathbb{R}, \\ X(0) = Y_0, \\ \dot{X}(0) = Y_1, \end{cases} \tag{4}$$

where A, Y_0, and Y_1 are random variables. We suppose that Y_0 and Y_1 have centered second order absolute moments. In [8], the hypothesis used in order to obtain a mean square analytic solution $X(t)$ was $\mathbb{E}[|A|^n] \le HR^n$, $n \ge n_0$. See [8] (expr. (18)–(19)) for the explicit expression of the solution process $X(t)$. Notice that this growth assumption is equivalent to $\|A\|_{L^\infty(\Omega)} \le R$, by [17] (Section 3.3). In our general notation, $A(t) = 0$ and $B(t) = At$. Due to the boundedness of the random variable A, the $L^\infty(\Omega)$ convergence of the series that define $A(t)$ and $B(t)$ holds, so Theorem 2 (and Theorem 1) is applicable: there is an analytic solution stochastic process $X(t)$ to (4) on $\mathbb{R}$, whose coefficients are defined by (2) and (3).

Example 2. *Hermite's random differential equation is given as follows:*

$$\begin{cases} \ddot{X}(t) - 2t\dot{X}(t) + A X(t) = 0, \ t \in \mathbb{R}, \\ X(0) = Y_0, \\ \dot{X}(0) = Y_1, \end{cases} \tag{5}$$

where A, Y_0, and Y_1 are random variables. We suppose that $Y_0, Y_1 \in L^2(\Omega)$. In [9], the hypothesis utilized to derive a mean square analytic solution $X(t)$ was $\mathbb{E}[|A|^n] \le HR^n$, $n \ge n_0$. See [9] (expr. (26)–(27)) for the explicit expression of the solution process $X(t)$. This growth hypothesis is equivalent to $\|A\|_{L^\infty(\Omega)} \le R$, by [17] (Section 3.3). Under the boundedness of the random variable A, the input stochastic processes $A(t) = -2t$ and $B(t) = A$ are expressible as $L^\infty(\Omega)$ convergent random power series. Hence, both Theorem 2 and Theorem 1 are applicable and guarantee the existence of a mean square solution process $X(t)$ on $\mathbb{R}$, whose coefficients are defined by (2) and (3).

Example 3. *We consider the following random linear differential equation with polynomial data processes:*

$$\begin{cases} \ddot{X}(t) + (A_0 + A_1 t)\dot{X}(t) + (B_0 + B_1 t)X(t) = 0, \ t \in \mathbb{R}, \\ X(0) = Y_0, \\ \dot{X}(0) = Y_1. \end{cases} \tag{6}$$

If the initial conditions Y_0 and Y_1 belong to $L^2(\Omega)$ and the random input parameters A_0, A_1, B_0, and B_1 are bounded random variables, then the hypotheses of Theorem 2 (and Theorem 1) are fulfilled, and we derive

that there is a mean square solution process $X(t)$ on $\mathbb{R}$, with coefficients defined by (2) and (3). In contrast to Example 1 and Example 2, the partial sums of the series $X(t)$ are not obtained explicitly. One computes the partial sums computationally via the recursion (2) and (3); see [17] (Example 4.3) for further details.

Example 4. *We consider (1) with the non-polynomial analytic stochastic process:*

$$\begin{cases} \ddot{X}(t) + A(t)\dot{X}(t) + B(t)X(t) = 0, \ t \in \mathbb{R}, \\ X(0) = Y_0, \\ \dot{X}(0) = Y_1, \end{cases} \tag{7}$$

where $A_n \sim Beta(11, 15)$, for $n \geq 0$, $B_n = 1/n^2$, for $n \geq 1$, and $Y_0, Y_1 \in L^2(\Omega)$. Since $\sum_{n=0}^{\infty} \|A_n\|_{L^{\infty}(\Omega)} |t|^n = \sum_{n=0}^{\infty} |t|^n$ is finite for $t \in (-1, 1)$, and analogously for $B(t)$, Theorem 2 (and consequently, Theorem 1) implies that there is a mean square solution $X(t)$ to (7) on $(-1, 1)$, with coefficients expressed by (2) and (3). Unlike Example 1 and Example 2, the partial sums of the series $X(t)$ are not obtained explicitly, and one acts computationally by means of the recursion (2) and (3); see [17] (Example 4.4) for details.

We raise the following open problem, which would imply that the hypotheses used in Theorem 2 are necessary: "If there exists a point $t_1 \in (t_0 - r, t_0 + r)$ such that $A(t_1) \notin L^{\infty}(\Omega)$ or $B(t_1) \notin L^{\infty}(\Omega)$, then there exist two initial conditions $Y_0, Y_1 \in L^2(\Omega)$ such that (1) has no mean square solution on $(t_0 - r, t_0 + r)$". Although we have not been able to prove this statement (which might be false), we think that the proof might be based on the reasoning used in [23] (Example, pp. 4–5).

This open problem, despite being of theoretical interest, does not contribute in practical applications. In numerical experiments, one usually truncates the stochastic processes $A(t)$ and $B(t)$ (that is, works with a partial sum instead of the whole Taylor series). This is not uncommon when dealing with stochastic systems computationally, as one requires a dimensionality reduction of the problem. If the coefficients of $A(t)$ and/or $B(t)$ have unbounded support, one may truncate them so that the hypotheses of Theorems 1 and 2 are fulfilled, and the probabilistic behavior of the data processes does not change much.

3. Non-Homogeneous Case

In this section, we generalize (1) by adding a stochastic source term:

$$\begin{cases} \ddot{X}(t) + A(t)\dot{X}(t) + B(t)X(t) = C(t), \ t \in \mathbb{R}, \\ X(t_0) = Y_0, \\ \dot{X}(t_0) = Y_1. \end{cases} \tag{8}$$

This new term $C(t)$ is analytic at t_0 in the mean square sense [21] (p. 99), with Taylor series:

$$C(t) = \sum_{n=0}^{\infty} C_n (t - t_0)^n, \quad t \in (t_0 - r, t_0 + r).$$

The coefficients $C_0, C_1, \ldots$ are random variables. For this new model (8), we want to find conditions under which $X(t)$ is an analytic mean square solution on $(t_0 - r, t_0 + r)$. This work was not done in [17], and it completes the study on the random non-autonomous second order linear differential equation with analytic input processes.

The following theorem is a generalization of Theorem 2:

Theorem 3. *Let $A(t) = \sum_{n=0}^{\infty} A_n (t - t_0)^n$ and $B(t) = \sum_{n=0}^{\infty} B_n (t - t_0)^n$ be two random series in the $L^{\infty}(\Omega)$ setting, for $t \in (t_0 - r, t_0 + r)$, $r > 0$ being finite and fixed. Let $C(t) = \sum_{n=0}^{\infty} C_n (t - t_0)^n$ be a random series*

in the mean square sense on $(t_0 - r, t_0 + r)$. Assume that the initial conditions Y_0 and Y_1 belong to $L^2(\Omega)$. Then, the stochastic process $X(t) = \sum_{n=0}^{\infty} X_n(t - t_0)^n$, $t \in (t_0 - r, t_0 + r)$, whose coefficients are defined by:

$$X_0 = Y_0, \quad X_1 = Y_1, \tag{9}$$

$$X_{n+2} = \frac{1}{(n+2)(n+1)} \left\{ -\sum_{m=0}^{n} [(m+1)A_{n-m}X_{m+1} + B_{n-m}X_m] + C_n \right\}, \; n \geq 0, \tag{10}$$

is the unique analytic solution to the random initial value problem (8) in the mean square sense.

Proof. Suppose that $X(t) = \sum_{n=0}^{\infty} X_n(t - t_0)^n$ is a solution to (8) in the $L^2(\Omega)$ sense, for $t \in (t_0 - r, t_0 + r)$. By Proposition 2 with $p = 2$, the mean square derivatives of $X(t)$ are given by:

$$\dot{X}(t) = \sum_{n=1}^{\infty} n X_n(t - t_0)^{n-1} = \sum_{n=0}^{\infty} (n+1)X_{n+1}(t - t_0)^n,$$

$$\ddot{X}(t) = \sum_{n=2}^{\infty} n(n-1)X_n(t - t_0)^{n-2} = \sum_{n=0}^{\infty} (n+2)(n+1)X_{n+2}(t - t_0)^n.$$

By Proposition 3,

$$A(t)\dot{X}(t) = \sum_{n=0}^{\infty} \left(\sum_{m=0}^{n} A_{n-m}(m+1)X_{m+1} \right) (t - t_0)^n,$$

$$B(t)X(t) = \sum_{n=0}^{\infty} \left(\sum_{m=0}^{n} B_{n-m}X_m \right) (t - t_0)^n,$$

where these two random series converge in $L^1(\Omega)$. From $\ddot{X}(t) + A(t)\dot{X}(t) + B(t)X(t) = C(t)$,

$$\sum_{n=0}^{\infty} \left[(n+2)(n+1)X_{n+2} + \sum_{m=0}^{n} (A_{n-m}(m+1)X_{m+1} + B_{n-m}X_m) \right] (t - t_0)^n = \sum_{n=0}^{\infty} C_n(t - t_0)^n,$$

where the infinite series converge in $L^1(\Omega)$. By Proposition 2 with $p = 1$, differentiating over and over again in the $L^1(\Omega)$ sense and evaluating at $t = t_0$ yield:

$$(n+2)(n+1)X_{n+2} + \sum_{m=0}^{n} (A_{n-m}(m+1)X_{m+1} + B_{n-m}X_m) = C_n.$$

Isolating X_{n+2}, we obtain the recursive expression (10).

Thus, it only remains to prove that the random power series $X(t) = \sum_{n=0}^{\infty} X_n(t - t_0)^n$, whose coefficients are defined by (9) and (10), converges in the mean square sense.

From the hypothesis $Y_0, Y_1 \in L^2(\Omega)$ and by induction on n in Expression (10), we obtain that $X_n \in L^2(\Omega)$ for all $n \geq 0$. On the other hand, by [11] (Lemma 2.3),

$$\sum_{n=0}^{\infty} \|A_n\|_{L^\infty(\Omega)} s^n < \infty, \quad \sum_{n=0}^{\infty} \|B_n\|_{L^\infty(\Omega)} s^n < \infty, \quad \sum_{n=0}^{\infty} \|C_n\|_{L^2(\Omega)} s^n < \infty,$$

for $0 < s < r$. As the general term of a convergent series tends to zero, we have the following bounds:

$$\|A_n\|_{L^\infty(\Omega)} \leq \frac{D_s}{s^n}, \quad \|B_n\|_{L^\infty(\Omega)} \leq \frac{D_s}{s^n}, \quad \|C_n\|_{L^2(\Omega)} \leq \frac{D_s}{s^n}, \; n \geq 0, \tag{11}$$

for a certain constant $D_s > 0$ that depends on s. Then, from (10), if we apply $L^2(\Omega)$ norms and (11), we obtain:

$$\|X_{n+2}\|_{L^2(\Omega)} \leq \frac{1}{(n+2)(n+1)} \left\{ \sum_{m=0}^{n} \left[(m+1)\|A_{n-m}X_{m+1}\|_{L^2(\Omega)} + \|B_{n-m}X_m\|_{L^2(\Omega)} \right] + \|C_n\|_{L^2(\Omega)} \right\}$$

$$\leq \frac{1}{(n+2)(n+1)} \frac{D_s}{s^n} \left\{ \sum_{m=0}^{n} s^m \left((m+1)\|X_{m+1}\|_{L^2(\Omega)} + \|X_m\|_{L^2(\Omega)} \right) + 1 \right\}. \tag{12}$$

Define $H_0 := \|Y_0\|_{L^2(\Omega)}$, $H_1 := \|Y_1\|_{L^2(\Omega)}$, and:

$$H_{n+2} := \frac{1}{(n+2)(n+1)} \frac{D_s}{s^n} \left\{ \sum_{m=0}^{n} s^m \left((m+1)H_{m+1} + H_m \right) + 1 \right\}, \quad n \geq 0. \tag{13}$$

From (12) and (13), by induction on n, it is trivially seen that $\|X_n\|_{L^2(\Omega)} \leq H_n$, for $n \geq 0$. If we check that $\sum_{n=0}^{\infty} H_n \rho_1^n < \infty$, for all $0 < \rho_1 < s < r$, then the random series that defines $X(t)$ converges in the mean square sense on $(t_0 - r, t_0 + r)$, as wanted.

We rewrite (13) so that H_{n+2} is expressed as a function of H_{n+1} and H_n (second order recurrence equation). By assuming $n \geq 1$, we perform the following operations:

$$H_{n+2} = \frac{1}{(n+2)(n+1)} \frac{D_s}{s^n} \left(\sum_{m=0}^{n-1} s^m \left((m+1)H_{m+1} + H_m \right) + 1 + s^n \left((n+1)H_{n+1} + H_n \right) \right)$$

$$= \frac{1}{(n+2)(n+1)} \frac{D_s}{s^n} \frac{(n+1)n}{D_s} s^{n-1} \left(\underbrace{\frac{1}{(n+1)n} \frac{D_s}{s^{n-1}} \left\{ \sum_{m=0}^{n-1} s^m \left((m+1)H_{m+1} + H_m \right) + 1 \right\}}_{=H_{n+1}} \right) \tag{14}$$

$$+ \frac{D_s}{n+2} H_{n+1} + \frac{D_s}{(n+2)(n+1)} H_n$$

$$= \left(\frac{n}{(n+2)s} + \frac{D_s}{n+2} \right) H_{n+1} + \frac{D_s}{(n+2)(n+1)} H_n.$$

This difference equation of order two has as initial conditions:

$$H_2 = \frac{D_s}{2} (H_1 + H_0 + 1), \quad H_1 = \|Y_1\|_{L^2(\Omega)}, \quad H_0 = \|Y_0\|_{L^2(\Omega)}.$$

Notice that H_2 is obtained from (13). Expression (14) coincides with [17] (expr. (12)) (although with different initial conditions). Then, the method of proof for $\sum_{n=0}^{\infty} H_n s^n < \infty$ is identical to the last part of the proof of [17] (Th. 3.3). Indeed, fixing $0 < \rho_1 < \rho < s$, we have:

$$H_{n+2}\rho^{n+2} = \left(\frac{n\rho}{(n+2)s} + \frac{D_s\rho}{n+2} \right) H_{n+1}\rho^{n+1} + \frac{D_s\rho^2}{(n+2)(n+1)} H_n\rho^n.$$

Let $M_n = \max_{0 \leq m \leq n} H_m\rho^m$. We have:

$$H_{n+2}\rho^{n+2} \leq \left(\frac{n\rho}{(n+2)s} + \frac{D_s\rho}{n+2} + \frac{D_s\rho^2}{(n+2)(n+1)} \right) M_{n+1}.$$

Since:

$$\lim_{n \to \infty} \frac{n\rho}{(n+2)s} + \frac{D_s\rho}{n+2} + \frac{D_s\rho^2}{(n+2)(n+1)} = \frac{\rho}{s} < 1,$$

it holds $M_{n+2} = M_{n+1}$ for all large n, and call the common value M. Hence, $H_n\rho^n \leq M$ for all large n, therefore, $H_n\rho_1^n \leq M(\rho_1/\rho)^n$. Since $\sum_{n=0}^{\infty}(\rho_1/\rho)^n < \infty$, by comparison, the series $\sum_{n=0}^{\infty} H_n\rho_1^n$ converges, and we are done. □

Example 5. *Let us consider Hermite's random differential equation with a stochastic source term:*

$$\begin{cases} \ddot{X}(t) - 2t\dot{X}(t) + AX(t) = Ct^2, \ t \in \mathbb{R}, \\ X(0) = Y_0, \\ \dot{X}(0) = Y_1, \end{cases} \tag{15}$$

where A, C, Y_0, and Y_1 are random variables. Due to the non-homogeneity of the equation, this example cannot be addressed with [17]. We have set the following probability distributions:

$$A \sim \text{Bernoulli}(0.35), \quad Y_0 \sim \text{Gamma}(2,2), \quad (Y_1, C) \sim \text{Multinomial}(3, \{0.2, 0.8\})$$

(for the Gamma distribution, we use the shape-rate notation) where A, Y_0 and (Y_1, C) are independent. Notice that we are considering both discrete and absolutely continuous random variables/vectors and also both independent and non-independent random variables/vectors. Thus, the Fröbenius method covers a wide variety of situations in practice. Since A is bounded and $Y_0, Y_1, C \in L^2(\Omega)$, Theorem 3 ensures that the random power series $X(t) = \sum_{n=0}^{\infty} X_n (t - t_0)^n$ defined recursively by (9) and (10) is a mean square solution to (15) on $\mathbb{R}$. By considering the partial sums $X^N(t) = \sum_{n=0}^{N} X_n (t - t_0)^n$, we approximate the expectation and variance of $X(t)$ as:

$$\mathbb{E}[X(t)] = \lim_{N \to \infty} \mathbb{E}[X_N(t)], \quad \mathbb{V}[X(t)] = \lim_{N \to \infty} \mathbb{V}[X_N(t)],$$

see Proposition 1. The computations have been performed in the software Mathematica®. Our code to build the partial sum $X^N(t)$ was the following one:

```
X[n_?NonPositive] := Y0;
X[1] = Y1;
X[n_] := 1/(n*(n - 1))*(-Sum[(m + 1)*A[n - 2 - m]*X[m + 1] +
B[n - 2 - m]*X[m], {m, 0, n - 2}] + CC[n - 2]);
seriesX[t_, t0_, N_] := X[0] + Sum[X[n]*(t - t0)^n, {n, 1, N}];
```

This implementation in the computer is necessary, as no closed-form expression for $X^N(t)$ is available due to the complexity of (15). For each numeric value of N, the functions $t \mapsto \mathbb{E}[X^N(t)]$ and $t \mapsto \mathbb{V}[X^N(t)]$ have been calculated with the built-in function Expectation applied to seriesX[t, 0, N] (with symbolic t), by setting the desired probability distributions to A[n], B[n], and CC[n]. In Tables 1 and 2, we show $\mathbb{E}[X^N(t)]$ and $\mathbb{V}[X^N(t)]$ for $N = 19$, $N = 20$, and $0 \le t \le 1.5$. Both orders of truncation produce similar results, which agrees with the theoretical convergence. Observe that, as we move away from the initial condition $t_0 = 0$, larger orders of truncation are needed. This indicates that the Fröbenius method might be computationally inviable for large t. The results have been compared with Monte Carlo simulation (with 100,000 and 200,000 realizations).

t	$\mathbb{E}[X^{19}(t)]$	$\mathbb{E}[X^{20}(t)]$	MC 100,000	MC 200,000
0.00	1	1	0.995893	1.00266
0.25	1.14231	1.14231	1.13899	1.14544
0.50	1.28890	1.28890	1.28672	1.29236
0.75	1.49183	1.49183	1.49130	1.49547
1.00	1.85892	1.85892	1.86087	1.86246
1.25	2.62573	2.62574	2.63173	2.62863
1.50	4.34772	4.34784	4.36111	4.34892

Table 1. Approximation of $\mathbb{E}[X(t)]$ with $N = 19$, $N = 20$, and Monte Carlo simulations. Example 5.

t	$\mathbb{V}[X^{19}(t)]$	$\mathbb{V}[X^{20}(t)]$	MC 100,000	MC 200,000
0.00	0.5	0.5	0.493124	0.504501
0.25	0.520298	0.520298	0.514702	0.524803
0.50	0.597008	0.597008	0.593603	0.601376
0.75	0.790556	0.790556	0.790161	0.794549
1.00	1.27425	1.27425	1.27702	1.27759
1.25	2.60694	2.60694	2.60987	2.60982
1.50	6.94095	6.94100	6.92663	6.94787

Table 2. Approximation of $\mathbb{V}[X(t)]$ with $N = 19$, $N = 20$, and Monte Carlo simulations. Example 5.

Notice that the theoretical error estimates from [17] (Section 3.6) apply in this case as well, since all estimates rely on the majorization $\|X_n\|_{L^2(\Omega)} \leq H_n$ and the recursive Equation (14), which also hold in [17].

An important issue that was not treated in the recent contribution [17] is the uniqueness of the mean square solution. To deal with uniqueness, we use a habitual extension of the classical Picard theorem to mean square calculus [21] (Th. 5.1.2); see Theorem 4. Notice that, in our setting of analyticity for $A(t)$ and $B(t)$ in the $L^\infty(\Omega)$ sense, one has that $A(t)$ and $B(t)$ are continuous in $L^\infty(\Omega)$, so the uniqueness from Theorem 4 is applicable.

Theorem 4. *If $A(t)$ and $B(t)$ are continuous stochastic processes in the $L^\infty(\Omega)$ sense, then the mean square solution to (8) is unique.*

Proof. We write (1) as a first order random differential equation, which is the setting under study in [21]:

$$\underbrace{\begin{pmatrix} \dot{X}(t) \\ \ddot{X}(t) \end{pmatrix}}_{\dot{Z}(t)} = \underbrace{\begin{pmatrix} 0 & 1 \\ -B(t) & -A(t) \end{pmatrix}}_{M(t)} \underbrace{\begin{pmatrix} X(t) \\ \dot{X}(t) \end{pmatrix}}_{Z(t)} + \underbrace{\begin{pmatrix} 0 \\ C(t) \end{pmatrix}}_{q(t)}.$$

We work in the space $L_2^2(\Omega)$ of two-dimensional random vectors whose components belong to $L^2(\Omega)$. Given $Z = (Z_1, Z_2) \in L_2^2(\Omega)$, its norm is defined as:

$$\|Z\|_{L_2^2(\Omega)} = \max\{\|Z_1\|_{L^2(\Omega)}, \|Z_2\|_{L^2(\Omega)}\}.$$

On the other hand, given a random matrix $B = (B_{ij})$, we define the following norm:

$$|||B||| = \max_i \sum_j \|B_{ij}\|_{L^\infty(\Omega)}.$$

In the case of the random matrix $M(t)$, it holds:

$$|||M(t)||| = \max\{1, \|A(t)\|_{L^\infty(\Omega)} + \|B(t)\|_{L^\infty(\Omega)}\}. \tag{16}$$

Given $Z, Z' \in L_2^2(\Omega)$, we have:

$$\|(M(t)Z + q(t)) - (M(t)Z' + q(t))\|_{L_2^2(\Omega)} = \|M(t)(Z - Z')\|_{L_2^2(\Omega)} \leq \underbrace{|||M(t)|||}_{k(t)} \cdot \|Z - Z'\|_{L_2^2(\Omega)}.$$

Since $A(t)$ and $B(t)$ are continuous stochastic processes in the $L^\infty(\Omega)$ sense, the real maps:

$$t \in (t_0 - r, t_0 + r) \mapsto \|A(t)\|_{L^\infty(\Omega)}, \quad t \in (t_0 - r, t_0 + r) \mapsto \|B(t)\|_{L^\infty(\Omega)}$$

are continuous. By (16), the deterministic function $k(t)$ is continuous on $(t_0 - r, t_0 + r)$. This implies that $k \in L^1([t_0 - r_1, t_0 + r_1])$ for each $0 < r_1 < r$. By [21] (Th. 5.1.2), there is the uniqueness of the mean square solution for (1) on $[t_0 - r_1, t_0 + r_1]$. Since r_1 is arbitrary, there is the uniqueness of the solution on $(t_0 - r, t_0 + r)$. □

4. Comparison with Other Methods

A final objective of this paper is to relate our method based on [17] (which is based on the deterministic Fröbenius method) to other well-known techniques to tackle (8). In [17], the random power series method was compared, both theoretically and in numerical experiments, with Monte Carlo simulations and the dishonest method [24]. It was demonstrated that Monte Carlo simulations imply a more expensive computational cost to calculate accurately the expectation and variance

statistics for t near t_0, due to the slow rate of convergence. However, Monte Carlo simulations usually allow validating the numerical results obtained, as they always present convergence with a similar rate for every stochastic system [25] (p. 53).

The article [17] does not compare the Fröbenius method with generalized polynomial chaos (gPC) expansions [25–30], although it has been proven to be a powerful technique to deal with general continuous and discrete stochastic systems with absolutely continuous random input coefficients. Due to the spectral mean square convergence of the Galerkin projections, the expectation and variance statistics of the response process can be approximated with small orders of truncation. In the particular setting of random second order linear differential equations, only [31,32] analyzed the application of gPC expansions to Airy's random differential equation, by assuming independence between the random input parameters. Recently, we have also studied the application of gPC expansions to the Legendre random differential equation with statistically-dependent inputs in an arXiv preprint [11]. The application of gPC expansions to general random second order linear differential equations (8) could be part of a future work. We believe that this is important because both the Fröbenius method and gPC expansions may validate each other in applications, since they provide good approximations of the expectation and variance statistics rapidly. Moreover, we believe that the gPC approach may provide better approximations of the statistics in the case of large times; see for example [33], where for the classical continuous epidemic models (SIS, SIR, etc.) uncertainty quantification is performed via gPC up to Time 60 with chaoses bases of order just two and three, producing very similar results; or [30], where an analogous study was performed for the corresponding discrete epidemiological models up to Time 30. Nonetheless, an excessively large number of input parameters may pose problems to the gPC-based method: if the chaos order is p and the degree of uncertainty is s, then the length of the basis for the gPC expansions is $(p+s)!/(p!s!)$ [25], which may make the method computationally inviable. Another drawback of the gPC technique is that catastrophic numerical errors usually appear for large chaos orders, specially when dealing with truncated distributions ([11] (Example 4.3) and [34]).

In [17], we did not compare our methodology with the random differential transform method proposed in [35]. Given a stochastic process $U(t)$, its random differential transform is defined as:

$$\hat{U}(k) = \frac{U^{(k)}(t_0)}{k!}.$$

Its inverse transform is defined as:

$$U(t) = \sum_{k=0}^{\infty} \hat{U}(k)(t-t_0)^k.$$

Notice that we are actually considering Taylor series in a random calculus setting. It is formally assumed that the series $\sum_{k=0}^{\infty} \hat{U}(k)(t-t_0)^k$ is mean square convergent on an interval $(t_0 - r, t_0 + r)$, $r > 0$. The computations with the random differential transform method were analyzed in [35] (Th. 2.1).

Proposition 4 ([35] (Th. 2.1)). *Let $F(t)$ and $G(t)$ be two second order stochastic processes, with mean square derivatives of k order $F^{(k)}(t)$ and $G^{(k)}(t)$. Then, the following results hold:*

(i) If $U(t) = F(t) \pm G(t)$, then $\hat{U}(k) = \hat{F}(k) \pm \hat{G}(k)$.
(ii) If $U(t) = \lambda F(t)$, where λ is a bounded random variable, then $\hat{U}(k) = \lambda \hat{F}(k)$.
(iii) If $U(t) = G^{(m)}(t)$, then $\hat{U}(k) = (k+1) \cdots (k+m) \hat{G}(k+m)$ (here, m is a nonnegative integer).
(iv) If $U(t) = F(t)G(t)$, then $\hat{U}(k) = \sum_{n=0}^{k} \hat{F}(n) \hat{G}(k-n)$.

Notice that (iii) and (iv) can be seen as consequences of differentiating random power series [17] (Th. 3.1) (Proposition 2) and multiplying random power series [17] (Th. 3.2) (Proposition 3), respectively.

Thereby, the random transform method is actually the random Fröbenius method. The recursive equations found for $\hat{X}(k)$ are as in (3). Our Theorems 1–3 give the conditions under which the inverse transform $\sum_{k=0}^{\infty} \hat{X}(k)(t-t_0)^k$ converges.

Thus, we believe that our recent contribution [17] together with the notes presented in this paper give an excellent approach to tackle (1) and/or (8) with analytic random input processes. Apart from obtaining a mean square analytic solution to (1) and/or (8), the expectation and variance of it can be calculated for uncertainty quantification.

5. Summary, Conclusions, and Future Lines of Research

In this paper, we have written some notes and comments to complete our recent contribution [17] on the random non-autonomous second order linear differential equation. The main theorem from [17], which deals with the homogeneous case, has been restated in a more convenient form to deal with practical applications. We addressed the non-homogeneous case, by proving an existence theorem of the mean square solution and performing a numerical example. On the other hand, the uniqueness of the solution has been established by using the Picard theorem for mean square calculus. A comparison of the extant techniques for uncertainty quantification (Monte Carlo, gPC expansions, random differential transform method) with respect to the random Fröbenius method was studied.

This paper is a contribution to the field of random differential equations, as it completely generalizes to a random framework the deterministic theory on second order linear differential equations with analytic input data. To carry out the study, mean square calculus and, in general, L^p random calculus become powerful tools to establish the theoretical results and perform uncertainty quantification.

Some future research lines related to the contents of this paper are the following:

- Solve the open problem raised in this paper at the end of Section 2, concerning the necessity of the hypotheses of Theorem 2.
- Apply the technique of gPC expansions and stochastic Galerkin projections to general random second order linear differential equations.
- Extend Theorem 3 to higher order random linear differential equations. Probably, one would need to require all input stochastic processes to be random power series in an L^{∞} sense, in analogy with the hypotheses of Theorem 3.
- Apply the random Fröbenius method to the random Riccati differential equation with the analytic input processes. In [35] (Section 3), the authors applied the random differential transform method (which is equivalent to a formal random Fröbenius method) to a particular case of the random Riccati differential equation with a random autonomous coefficient term. It would be interesting to apply the random Fröbenius method in the situation in which all input coefficients are analytic stochastic processes, by proving theoretical results and performing numerical experiments.

Author Contributions: Investigation, J.C.G. and M.J.S.; methodology, J.C.G. and M.J.S.; software, J.C.G. and M.J.S.; supervision, J.C.C.L.; validation, J.C.C.L.; visualization, J.C.G., J.C.C.L., and M.J.S.; writing—original draft, J.C.G. and M.J.S.; writing—review and editing, J.C.G., J.C.C.L., and M.J.S.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Goodwine, B. *Engineering Differential Equations. Theory and Applications*; Springer: New York, NY, USA, 2011.
2. Schiesser, W.E. *Differential Equation Analysis in Biomedical Science and Engineering: Ordinary Differential Equation Applications with R*; John Wiley & Sons Inc.: New York, NY, USA, 2014.

3. Brown, C.M. *Differential Equations: A Modeling Approach (Quantitative Applications in the Social Sciences)*; SAGE, Inc.: New York, NY, USA, 2007.

4. Valée, O.; Soares, M. *Airy Functions and Applications to Physics*; Imperial College Press: London, UK, 2004.

5. Fedoryuk, M.V. *Encyclopedia of Mathematics*; Springer Science+Business Media B.V./Kluwer Academic Publisher: New York, NY, USA, 2001.

6. Spain, B.; Smith, M.G. *Functions of Mathematical Physics*; Van Nostrand Reinhold Company: London, UK, 1970.

7. Johnson, C.S; Pedersen, L.G. *Problems and Solutions in Quantum Chemistry and Physics*; Dover Publ.: New York, NY, USA, 1986.

8. Cortés, J.-C.; Jódar, L.; Camacho, J.; Villafuerte, L. Random Airy type differential equations: Mean square exact and numerical solutions. *Comput. Math. Appl.* **2010**, *60*, 1237–1244. [CrossRef]

9. Calbo, G.; Cortés, J.-C.; Jódar, L. Random Hermite differential equations: Mean square power series solutions and statistical properties. *Appl. Math. Comput.* **2011**, *218*, 3654–3666. [CrossRef]

10. Calbo, G.; Cortés, J.-C.; Jódar, L.; Villafuerte, L. Solving the random Legendre differential equation: Mean square power series solution and its statistical functions. *Comput. Math. Appl.* **2011**, *61*, 2782–2792. [CrossRef]

11. Calatayud, J.; Cortés, J.-C.; Jornet, M. Improving the approximation of the first and second order statistics of the response process to the random Legendre differential equation. *arXiv* **2018**, arXiv:1807.03141.

12. Cortés, J.-C.; Jódar, L.; Company, R.; Villafuerte, L. Laguerre random polynomials: Definition, differential and statistical properties. *Util. Math.* **2015**, *98*, 283–295.

13. Cortés, J.-C.; Jódar, L.; Villafuerte, L. Mean square solution of Bessel differential equation with uncertainties. *J. Comput. Appl. Math.* **2017**, *309*, 383–395. [CrossRef]

14. Golmankhaneh, A.A.; Porghoveh, N.A.; Baleanu, D. Mean square solutions of second-order random differential equations by using homotopy analysis method. *Rom. Rep. Phys.* **2013**, *65*, 350–362.

15. Khudair, A.R.; Ameen, A.A.; Khalaf, S.L. Mean square solutions of second-order random differential equations by using variational iteration method. *Appl. Math. Sci.* **2011**, *5*, 2505–2519.

16. Khudair, A.R.; Ameen, A.A.; Khalaf, S.L. Mean square solutions of second-order random differential equations by using Adomian decomposition method. *Appl. Math. Sci.* **2011**, *5*, 2521–2535.

17. Calatayud, J.; Cortés, J.-C.; Jornet, M.; Villafuerte, L. Random non-autonomous second order linear differential equations: Mean square analytic solutions and their statistical properties. *Adv. Differ. Equ.* **2018**, *2018*, 392. [CrossRef]

18. Kadry, S. On the generalization of probabilistic transformation method. *Appl. Math. Comput.* **2007**, *190*, 1284–1289. [CrossRef]

19. Kloeden, P.E.; Platen, E. *Numerical Solution of Stochastic Differential Equations*; Springer: New York, NY, USA, 1992.

20. Díaz-Infante, S.; Jerez, S. The linear Steklov method for SDEs with non-globally Lipschitz coefficients: Strong convergence and simulation. *J. Comp. Appl. Math.* **2017**, *309*, 408–423. [CrossRef]

21. Soong, T.T. *Random Differential Equations in Science and Engineering*; Academic Press: New York, NY, USA, 1973.

22. Smith, R.C. *Uncertainty Quantification. Theory, Implementation, and Applications*; SIAM in the Computational Science and Engineering Series; CS12; SIAM: New York, NY, USA, 2014.

23. Strand, J.L. Random ordinary differential equations. *J. Differ. Equ.* **1970**, *7*, 538–553. [CrossRef]

24. Henderson, D.; Plaschko, P. *Stochastic Differential Equations in Science And Engineering*; World Scientific Pub Co. Inc.: Singapore, 2006.

25. Xiu, D. *Numerical Methods for Stochastic Computations. A Spectral Method Approach*; Princeton University Press: Princeton, NJ, USA, 2010.

26. Xiu, D.; Karniadakis, G.E. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.* **2002**, *24*, 619–644. [CrossRef]

27. Chen-Charpentier, B.-M.; Cortés, J.-C.; Licea, J.-A.; Romero, J.-V.; Roselló, M.-D.; Santonja, F.-J.; Villanueva, R.-J. Constructing adaptive generalized polynomial chaos method to measure the uncertainty in continuous models: A computational approach. *Math. Comput. Simul.* **2015**, *109*, 113–129. [CrossRef]

28. Cortés, J.-C.; Romero, J.-V.; Roselló, M.-D.; Villanueva, R.-J. Improving adaptive generalized polynomial chaos method to solve nonlinear random differential equations by the random variable transformation technique. *Commun. Nonlinear Sci. Numer. Simul.* **2017**, *50*, 1–15. [CrossRef]

29. Cortés, J.-C.; Romero, J.-V.; Roselló, M.-D.; Santonja, F.-J.; Villanueva, R.-J. Solving continuous models with dependent uncertainty: A computational approach. *Abstr. Appl. Anal.* **2013**, *2013*, 1–10. [CrossRef]

30. Calatayud, J.; Cortés, J.-C.; Jornet, M.; Villanueva, R.-J. Computational uncertainty quantification for random time-discrete epidemiological models using adaptive gPC. *Math. Meth. Appl. Sci.* **2018**, 1–10. [CrossRef]
31. Chen-Charpentier, B.-M.; Cortés, J.-C.; Romero, J.-V.; Roselló, M.-D. Some recommendations for applying gPC (generalized polynomial chaos) to modeling: An analysis through the Airy random differential equation. *Appl. Math. Comput.* **2013**, *219*, 4208–4218. [CrossRef]
32. Chen-Charpentier, B.-M.; Cortés, J.-C.; Romero, J.-V.; Roselló, M.-D. Do the generalized polynomial chaos and Fröbenius methods retain the statistical moments of random differential equations? *Appl. Math. Lett.* **2013**, *26*, 553–558. [CrossRef]
33. Chen-Charpentier, B.-M.; Stanescu, D. Epidemic models with random coefficients. *Math. Comput. Mod.* **2010**, *52*, 1004–1010. [CrossRef]
34. Calatayud, J.; Cortés, J.-C.; Jornet, M. On the convergence of adaptive gPC for non-linear random difference equations: Theoretical analysis and some practical recommendations. *J. Nonlinear Sci. Appl.* **2018**, *11*, 1077–1084. [CrossRef]
35. Villafuerte, L.; Chen-Charpentier, B.-M. A random differential transform method: Theory and applications. *Appl. Math. Lett.* **2012**, *25*, 1490–1494. [CrossRef]

Article

Modified Potra–Pták Multi-step Schemes with Accelerated Order of Convergence for Solving Systems of Nonlinear Equations

Himani Arora [1,*], Juan R. Torregrosa [2] and Alicia Cordero [2]

[1] Department of Mathematics, D.A.V. University, Sarmastpur, 144012 Jalandhar, India
[2] Instituto de Matemática Multidisciplinar, Universitat Politècnica de València, 46022 Valencia, Spain; jrtorre@mat.upv.es (J.R.T.); acordero@mat.upv.es (A.C.)
[*] Correspondence: arorahimani362@gmail.com

Abstract: In this study, an iterative scheme of sixth order of convergence for solving systems of nonlinear equations is presented. The scheme is composed of three steps, of which the first two steps are that of third order Potra–Pták method and last is weighted-Newton step. Furthermore, we generalize our work to derive a family of multi-step iterative methods with order of convergence $3r + 6, r = 0, 1, 2, \ldots$. The sixth order method is the special case of this multi-step scheme for $r = 0$. The family gives a four-step ninth order method for $r = 1$. As much higher order methods are not used in practice, so we study sixth and ninth order methods in detail. Numerical examples are included to confirm theoretical results and to compare the methods with some existing ones. Different numerical tests, containing academical functions and systems resulting from the discretization of boundary problems, are introduced to show the efficiency and reliability of the proposed methods.

Keywords: systems of nonlinear equations; iterative methods; Newton's method; order of convergence; computational efficiency; basin of attraction

1. Introduction

Many applied problems in Science and Engineering [1–3] are reduced to solve nonlinear systems $F(x) = 0$ numerically, that is, for a given nonlinear function $F(x) : D \subset R^m \longrightarrow R^m$, where $F(x) = (f_1(x), f_2(x), ..., f_m(x))^T$ and $x = (x_1, x_2, ..., x_m)^T$, to find a vector $\alpha = (\alpha_1, \alpha_2, ..., \alpha_m)^T$ such that $F(\alpha) = 0$. The most widely used method for this purpose is the classical Newton's method [3,4], which converges quadratically under the conditions that the function F is continuously differentiable and a good initial approximation $x^{(0)}$ is given. It is defined by

$$x^{(k+1)} = x^{(k)} - F'(x^{(k)})^{-1} F(x^{(k)}), \quad k = 0, 1, \ldots, \tag{1}$$

where $F'(x^{(k)})^{-1}$ is the inverse of Fréchet derivative $F'(x^{(k)})$ of the function $F(x)$ at $x^{(k)}$. In order to improve the order of convergence of Newton's method, several methods have been proposed in literature, see, for example [5–11] and references therein. In particular, the third order method by Potra–Pták [11] for systems of nonlinear equations is given by

$$\begin{aligned} y^{(k)} &= x^{(k)} - F'(x^{(k)})^{-1} F(x^{(k)}), \\ x^{(k+1)} &= y^{(k)} - F'(x^{(k)})^{-1} F(y^{(k)}) \quad k = 0, 1, \ldots \end{aligned} \tag{2}$$

It is quite clear that this scheme requires the evaluation of two functions, one derivative and one matrix inversion per iteration, that is usually avoided by solving a linear system. This algorithm is illustrious not only for its simplicity but also for its efficient character.

In this paper, based on Potra–Pták method (2), we develop a three-step scheme with increased order of convergence and still maintaining the efficient character. With these considerations, we propose a three-step iterative method with accelerated sixth order of convergence; of the three steps, the first two are those of Potra–Pták method whereas the third is a weighted Newton-step. Then, based on this scheme, a multi-step family with increasing order of convergence $3r + 6, r = 0, 1, 2, \ldots$, is developed. The sixth order method is the special case of this multi-step scheme for $r = 0$. The family gives a four-step ninth order scheme for $r = 1$. As much higher order methods are not used in practice, so we study sixth and ninth order methods in particular.

The rest of the paper is organized as follows: In Section 2, we present the new three-step scheme for solving nonlinear systems and we analyze its order of convergence. In Section 2.1, the order of this scheme is improved in three units, by adding another step that needs a new functional evaluation and to solve a linear system with the same matrix of coefficients as before. This idea can be generalized for designing an iterative method with arbitrary order of convergence. The computational efficiency of the proposed schemes is studied in Section 3, doing a comparative analysis with the efficiency of other known methods. Section 4 is devoted to the numerical experiments with academical multivariate functions and nonlinear systems resulted of the discretization of boundary problems. For some cases, the basins of attraction of the methods are also showed. The paper finishes with some conclusions and the references used in it.

2. The Method and Analysis of Convergence

We start with the following iterative scheme:

$$
\begin{aligned}
y^{(k)} &= x^{(k)} - F'(x^{(k)})^{-1}F(x^{(k)}), \\
z^{(k)} &= y^{(k)} - F'(x^{(k)})^{-1}F(y^{(k)}), \\
x^{(k+1)} &= z^{(k)} - (aI + F'(x^{(k)})^{-1}[z^{(k)}, y^{(k)} ; F](bI + cF'(x^{(k)})^{-1}[z^{(k)}, y^{(k)} ; F]))F'(x^{(k)})^{-1}F(z^{(k)}),
\end{aligned}
\tag{3}
$$

where the first two-steps are those of Potra–Pták scheme [11] for nonlinear systems and $[z^{(k)}, y^{(k)} ; F]$ is the first order divided difference of F.

In order to discuss the behavior of scheme (3), we consider the following expression of divided difference operator $[\cdot, \cdot ; F] : D \times D \subset R^m \times R^m \longrightarrow L(R^m)$, see for example, [2,10],

$$
[x + h, x ; F] = \int_0^1 F'(x + th)\, dt, \ \forall\, x, h \in R^m.
\tag{4}
$$

By expanding $F'(x + th)$ in Taylor series at the point x and integrating, we have

$$
[x + h, x ; F] = \int_0^1 F'(x + th)\, dt = F'(x) + \frac{1}{2}F''(x)h + \frac{1}{6}F'''(x)h^2 + O(h^3).
\tag{5}
$$

where $h^i = (h, h, \cdot\,\overset{i}{\cdots}, h)$, $h \in R^m$.
Let $e^{(k)} = x^{(k)} - \alpha$. Developing $F(x^{(k)})$ in a neighborhood of α and assuming that $\Gamma = [F'(\alpha)]^{-1}$ exists, we have

$$
F(x^{(k)}) = F'(\alpha)(e^{(k)} + A_2(e^{(k)})^2 + A_3(e^{(k)})^3 + A_4(e^{(k)})^4 + O((e^{(k)})^5)),
\tag{6}
$$

where $A_i = \frac{1}{i!}\Gamma F^{(i)}(\alpha) \in L_i(R^m, R^m)$ and $(e^{(k)})^i = (e^{(k)}, e^{(k)}, \cdot\,\overset{i}{\cdots}, e^{(k)})$, $e^{(k)} \in R^m$.
Also,

$$
F'(x^{(k)}) = F'(\alpha)(I + 2A_2 e^{(k)} + 3A_3(e^{(k)})^2 + 4A_4(e^{(k)})^3 + O((e^{(k)})^4)),
\tag{7}
$$

$$
F''(x^{(k)}) = F'(\alpha)(2A_2 + 6A_3 e^{(k)} + 12A_4(e^{(k)})^2 + O((e^{(k)})^3)),
\tag{8}
$$

$$
F'''(x^{(k)}) = F'(\alpha)(6A_3 + 24A_4 e^{(k)} + O((e^{(k)})^2)).
\tag{9}
$$

Inversion of $F'(x^{(k)})$ yields,

$$F'(x^{(k)})^{-1} = \left[I - 2A_2 e^{(k)} + (4A_2^2 - 3A_3)(e^{(k)})^2 - (8A_2^3 - 6A_2 A_3 - 6A_3 A_2 + 4A_4)(e^{(k)})^3 + O((e^{(k)})^4)\right]\Gamma. \quad (10)$$

We are in a position to analyze the behavior of scheme (3). Thus, the following result is proven.

Theorem 1. *Let $F : D \subset R^m \to D$ be sufficiently differentiable function in an open neighborhood D of its zero α. Let us suppose that the Jacobian matrix $F'(x)$ is continuous and nonsingular at α. If an initial approximation $x^{(0)}$ is sufficiently close to α, then the local order of convergence of method (3) is at least 6, provided $a = 13/4, b = -7/2$ and $c = 5/4$.*

Proof. Let $e_y^{(k)} = y^{(k)} - \alpha$ is the local error of Newton's method given by

$$e_y^{(k)} = A_2(e^{(k)})^2 - 2(A_2^2 - A_3)(e^{(k)})^3 + O((e^{(k)})^4). \quad (11)$$

Employing Equations (10) and (11) in the second step of (3), we get

$$e_z^{(k)} = z^{(k)} - \alpha = 2A_2 e^{(k)} e_y^{(k)} + O((e^{(k)})^4). \quad (12)$$

Using Equations (7)–(9) in (5) for $x + h = z^{(k)}$, $x = y^{(k)}$ and $h = e_z^{(k)} - e_y^{(k)}$, it follows that

$$[z^{(k)}, y^{(k)} ; F] = F'(\alpha)(I + A_2(e_y^{(k)} + e_z^{(k)}) + O((e^{(k)})^4)). \quad (13)$$

With the help of Equations (10) and (13), we can write

$$
\begin{aligned}
aI + &F'(x^{(k)})^{-1}[z^{(k)}, y^{(k)} ; F](bI + cF'(x^{(k)})^{-1}[z^{(k)}, y^{(k)} ; F]) \\
&= \Big[(a + b + c)I - 2(b + 2c)A_2 e^{(k)} + ((5b + 14c)A_2^2 - 3(b + 2c)A_3)(e^{(k)})^2 - (10(b + 4c)A_2^3 \\
&\quad - (8b + 22c)A_2 A_3 - (6b + 18c)A_3 A_2 + 4(b + 2c)A_4)(e^{(k)})^3 + O((e^{(k)})^4)\Big]\Gamma.
\end{aligned}
\quad (14)
$$

Expanding $F(z^{(k)})$ about α, we obtain

$$F(z^{(k)}) = F'(\alpha)\left(e_z^{(k)} + A_2(e_z^{(k)})^2 + O((e_z^{(k)})^3)\right). \quad (15)$$

Using (10), (14) and (15) in the last step of (3), we get

$$
\begin{aligned}
e^{(k+1)} = x^{(k+1)} - \alpha = {}& 2(1 - a - b - c)A_2^2(e^{(k)})^3 + 4(a + 2b + 3c)A_2^3(e^{(k)})^4 \\
&- 2((4a + 13(b + 2c))A_2^4 - 3A_3(a + 2b + 3c))A_2^2(e^{(k)})^5 + 4((3a + 17b + 45c)A_2^5 \\
&- (3a + 10(b + 2c))A_2 A_3 A_2^2 + (3a + 6(b + 3c))A_3 A_2^3 + 2(a + 2b + 3c)A_4 A_2^2)(e^{(k)})^6 \\
&+ O((e^{(k)})^7).
\end{aligned}
\quad (16)
$$

In order to achieve sixth order of convergence it is clear that terms $1 - a - b - c$, $a + 2b + 3c$ and $4a + 13(b + 2c)$ must vanish for some values of a, b, and c. This happens when $a = \frac{13}{4}$, $b = -\frac{7}{2}$ and $c = \frac{5}{4}$.

Thus, the error equation (16) becomes

$$e^{(k+1)} = x^{(k+1)} - \alpha = (26A_2^3 + A_2 A_3 - 3A_3 A_2)A_2^2(e^{(k)})^6 + O((e^{(k)})^7). \quad (17)$$

This proves the sixth order of convergence. □

Thus, the proposed scheme (3) is given by

$$
\begin{aligned}
y^{(k)} &= x^{(k)} - F'(x^{(k)})^{-1} F(x^{(k)}), \\
z^{(k)} &= y^{(k)} - F'(x^{(k)})^{-1} F(y^{(k)}), \\
x^{(k+1)} &= z^{(k)} - \left(\frac{13}{4} I - F'(x^{(k)})^{-1} [z^{(k)}, y^{(k)} ; F] \left(\frac{7}{2} I - \frac{5}{4} F'(x^{(k)})^{-1} [z^{(k)}, y^{(k)} ; F] \right) \right) \\
&\quad \times F'(x^{(k)})^{-1} F(z^{(k)}).
\end{aligned}
\tag{18}
$$

Clearly this formula uses three functional evaluations, one evaluation of the Jacobian matrix, one of a divided difference, and one matrix inversion per iteration. We denote this scheme by $H_{6,1}$.

2.1. Multi-step Method with Order $3r + 6$

In this section, we improve the $H_{6,1}$ by adding a functional evaluation per each new step to get the multi-step version called $H_{3r+6,1}$ method. The method is defined as

$$
\begin{aligned}
x^{(k+1)} &= H_{3r+6,1}(x^{(k)}) = v_r(x^{(k)}), \\
v_j(x^{(k)}) &= v_{j-1}(x^{(k)}) - \theta(x^{(k)}) F'(x^{(k)})^{-1} F(v_{j-1}(x^{(k)})), \\
\theta(x^{(k)}) &= \frac{13}{4} I - F'(x^{(k)})^{-1} [z^{(k)}, y^{(k)} ; F] \left(\frac{7}{2} I - \frac{5}{4} F'(x^{(k)})^{-1} [z^{(k)}, y^{(k)} ; F] \right), \\
v_0 &= H_{6,1}(x^{(k)}), \quad j = 1, 2, \ldots, r ; \ r \geqslant 1.
\end{aligned}
\tag{19}
$$

Let us note that case $r = 0$ is $H_{6,1}$ method given by (18). This multi-step version has the order of convergence $3r + 6$, $r \geqslant 1$, which we shall prove through the following result.

Theorem 2. *Let us Assume that $F : D \subset \mathbb{R}^m \longrightarrow D$ is a sufficiently Fréchet differentiable function in an open convex set D containing the zero α of $F(x)$ and $F'(x)$ is continuous and nonsingular at α. Then, sequence $\{x^{(k)}\}_{k \geqslant 0}$, $x^{(0)} \in D$ obtained by using method (19) converges to α with convergence order $3r + 6$.*

Proof. Let $e_{v_j}^{(k)} = v_j^{(k)} - \alpha$, for all $j = 1, 2, \ldots, r$. Taylor's expansion of $F(v_{j-1}^{(k)}(x^{(k)}))$ about α yields

$$
F(v_{j-1}(x^{(k)})) = F'(\alpha) \left(e_{v_{j-1}}^{(k)} + A_2 (e_{v_{j-1}}^{(k)})^2 + O((e_{v_{j-1}}^{(k)})^3) \right).
\tag{20}
$$

Using (10) and (14) for $a = 13/4, b = -7/2$ and $c = 5/4$, we have

$$
\begin{aligned}
\theta(x^{(k)}) F'(x^{(k)})^{-1} &= \left(\frac{13}{4} I - F'(x^{(k)})^{-1} [z^{(k)}, y^{(k)} ; F] \left(\frac{7}{2} I - \frac{5}{4} F'(x^{(k)})^{-1} [z^{(k)}, y^{(k)} ; F] \right) \right) F'(x^{(k)})^{-1} \\
&= \left[I - S_3 (e^{(k)})^3 + O((e^{(k)})^4) \right] \Gamma,
\end{aligned}
\tag{21}
$$

where $S_3 = 15 A_2^3 + \frac{1}{2} A_2 A_3 - \frac{3}{2} A_3 A_2$.

With the help of (20) and (21), we can write

$$
\begin{aligned}
\theta(x^{(k)}) F'(x^{(k)})^{-1} F(v_{j-1}(x^{(k)})) &= (I - S_3 (e^{(k)})^2 + \cdots)(e_{v_{j-1}}^{(k)} + A_2 (e_{v_{j-1}}^{(k)})^2 + \cdots) \\
&= e_{v_{j-1}}^{(k)} - S_3 (e^{(k)})^3 e_{v_{j-1}}^{(k)} + \cdots.
\end{aligned}
\tag{22}
$$

By substituting (22) in (19), we get

$$
\begin{aligned}
e_{v_j}^{(k)} = v_j^{(k)} - \alpha &= e_{v_{j-1}}^{(k)} - (e_{v_{j-1}}^{(k)} - S_3 (e^{(k)})^3 e_{v_{j-1}}^{(k)} + \cdots) \\
&= S_3 (e^{(k)})^3 e_{v_{j-1}}^{(k)} + \cdots.
\end{aligned}
\tag{23}
$$

As we know that $e_{v_0}^{(k)} = (26A_2^3 + A_2A_3 - 3A_3A_2)A_2^2(e^{(k)})^6$, so from (23), for $j = 1, 2, \ldots$, we have that

$$
\begin{aligned}
e_{v_1}^{(k)} &= S_3(e^{(k)})^3 e_{v_0}^{(k)} \\
&= S_3(e^{(k)})^3 (26A_2^3 + A_2A_3 - 3A_3A_2)A_2^2(e^{(k)})^6 + \cdots, \\
e_{v_2}^{(k)} &= S_3(e^{(k)})^2 e_{v_1}^{(k)} \\
&= S_3^2(e^{(k)})^6 (26A_2^3 + A_2A_3 - 3A_3A_2)A_2^2(e^{(k)})^6 + \cdots.
\end{aligned}
\tag{24}
$$

Proceeding by induction, we have

$$
e_{v_r}^{(k)} = S_3^r(26A_2^3 + A_2A_3 - 3A_3A_2)A_2^2(e^{(k)})^{3r+6} + O((e^{(k)})^{3r+7}), \quad r \geqslant 0.
\tag{25}
$$

Hence, the result of the theorem is proven. □

As a special case, this family when $r = 1$ gives a ninth order method:

$$
\begin{aligned}
y^{(k)} &= x^{(k)} - F'(x^{(k)})^{-1}F(x^{(k)}), \\
z^{(k)} &= x^{(k)} - F'(x^{(k)})^{-1}F(y^{(k)}), \\
v_0^{(k)} &= H_{6,1}(x^{(k)}, y^{(k)}), \\
x^{(k+1)} &= v_0^{(k)} - \theta(x^{(k)})F'(x^{(k)})^{-1}F(v_0^{(k)}) \quad k = 0, 1, 2, \ldots
\end{aligned}
\tag{26}
$$

It is clear that this scheme requires four functional evaluations, one Jacobian matrix, one divided difference, and one matrix inversion per full iteration. We denote this scheme by $H_{9,1}$.

Remark 1. *Multi-step version* H_{3r+6} *($r \geqslant 0$), utilizes $r + 3$ functional evaluations of F, one evaluation of F', and one divided difference. Also, the scheme (19) requires only one matrix inversion per iteration.*

3. Computational Efficiency

To obtain an estimation of the efficiency of the proposed methods we shall make use of efficiency index, according to which the efficiency of an iterative method is given by $E = \rho^{\frac{1}{C}}$, where ρ is the order of convergence and C is the computational cost per iteration. For a system of m nonlinear equations and m unknowns, the computational cost per iteration is given by (see [9,10])

$$
C(\mu_0, \mu_1, m) = P_0(m)\mu_0 + P_1(m)\mu_1 + P(m),
\tag{27}
$$

where $P_0(m)$ represents the number of evaluations of scalar functions $(f_1, f_2, \ldots, f_m)$ used in the evaluations of F and $[y, x; F]$. The divided difference $[y, x; F]$ of F is an $m \times m$ matrix with elements (see [11,12])

$$
\begin{aligned}
[y, x; F]_{ij} =&(f_i(y_1, \ldots, y_{j-1}, y_j, x_{j+1}, \ldots, x_m) - f_i(y_1, \ldots, y_{j-1}, x_j, x_{j+1}, \ldots, x_m) \\
&+ f_i(x_1, \ldots, x_{j-1}, y_j, y_{j+1}, \ldots, y_m) \\
&- f_i(x_1, \ldots, x_{j-1}, x_j, y_{j+1}, \ldots, y_m)) / (2(y_j - x_j)), \quad 1 \leqslant i, j \leqslant m.
\end{aligned}
$$

The number of evaluations of scalar functions of F', i.e. $\frac{\partial f_i}{\partial x_j}, 1 \leqslant i, j \leqslant m$, is $P_1(m)$. $P(m)$ represents the number of products or quotients needed per iteration, and μ_0 and μ_1 are ratios between products and evaluations required to express the value of $C(\mu_0, \mu_1, m)$ in terms of products.

To compute F in any iterative method we calculate m scalar functions and if we compute the divided difference $[y, x; F]$ then we evaluate $2m(m - 1)$ scalar functions, where $F(x)$ and $F(y)$ are computed separately. We must add m^2 quotients from any divided difference. The number of scalar evaluations is m^2 for any new derivative F'. In order to compute an inverse linear operator we solve a linear system, where we have $m(m - 1)(2m - 1)/6$ products and $m(m - 1)/2$ quotients in the LU

decomposition and $m(m-1)$ products and m quotients in the resolution of two triangular linear systems. We suppose that a quotient is equivalent to l products. Moreover, we add m^2 products for the multiplication of a matrix with a vector or of a matrix by a scalar and m products for the multiplication of a vector by a scalar.

Denoting the efficiency indices of $H_{\rho,i}$ ($\rho = 6, 9$ and $i = 1, 2, 3, 4$) by $E_{\rho,i}$ and computational cost by $C_{\rho,i}$, then taking into account the above considerations, we obtain

$$C_{6,1} = (2m^2 + m)\mu_0 + m^2\mu_1 + \frac{m}{6}(2m^2 + 39m - 11 + 9l(m+3)) \quad \text{and} \quad E_{6,1} = 6^{1/C_{6,1}}. \tag{28}$$

$$C_{9,1} = (2m^2 + 2m)\mu_0 + m^2\mu_1 + \frac{m}{6}(2m^2 + 69m - 11 + 9l(m+5)) \quad \text{and} \quad E_{9,1} = 9^{1/C_{9,1}}. \tag{29}$$

To check the performance of the new sixth order method, we compare it with some existing sixth order which belongs to the same class. So, we choose the sixth order methods presented in [10,13]. The sixth order methods presented in [10] are given by

$$
\begin{aligned}
y^{(k)} &= x^{(k)} - F'(x^{(k)})^{-1}F(x^{(k)}), \ k = 0, 1, 2, \ldots \\
z^{(k)} &= y^{(k)} - (2[y^{(k)}, x^{(k)}; F] - F'(x^{(k)}))^{-1}F(y^{(k)}), \\
x^{(k+1)} &= H_{6,2}(x^{(k)}, y^{(k)}, z^{(k)}) = z^{(k)} - (2[y^{(k)}, x^{(k)}; F] - F'(x^{(k)}))^{-1}F(z^{(k)})
\end{aligned} \tag{30}
$$

and

$$
\begin{aligned}
y^{(k)} &= x^{(k)} - F'(x^{(k)})^{-1}F(x^{(k)}), \ k = 0, 1, 2, \ldots \\
z^{(k)} &= y^{(k)} - (2[y^{(k)}, x^{(k)}; F]^{-1} - F'(x^{(k)})^{-1})F(y^{(k)}), \\
x^{(k+1)} &= H_{6,3}(x^{(k)}, y^{(k)}, z^{(k)}) = z^{(k)} - (2[y^{(k)}, x^{(k)}; F]^{-1} - F'(x^{(k)})^{-1})F(z^{(k)}).
\end{aligned} \tag{31}
$$

The sixth order method proposed in [13] is given by

$$
\begin{aligned}
y^{(k)} &= x^{(k)} - F'(x^{(k)})^{-1}F(x^{(k)}), \ k = 0, 1, 2, \ldots \\
z^{(k)} &= y^{(k)} - (3I - 2F'(x^{(k)})^{-1}[y^{(k)}, x^{(k)}; F])F'(x^{(k)})^{-1}F(y^{(k)}), \\
x^{(k+1)} &= H_{6,4}(x^{(k)}, y^{(k)}, z^{(k)}) = z^{(k)} - (3I - 2F'(x^{(k)})^{-1}[y^{(k)}, x^{(k)}; F])F'(x^{(k)})^{-1}F(z^{(k)}).
\end{aligned} \tag{32}
$$

Per iteration these methods utilize the same number of function evaluations as that of $H_{6,1}$. The computational cost and efficiency for the methods $H_{6,2}$, $H_{6,3}$ and $H_{6,4}$ is given as below:

$$C_{6,2} = (2m^2 + m)\mu_0 + m^2\mu_1 + \frac{m}{3}(2m^2 + 9m - 8 + 6l(m+1)) \quad \text{and} \quad E_{6,2} = 6^{1/C_{6,2}}, \tag{33}$$

$$C_{6,3} = (2m^2 + m)\mu_0 + m^2\mu_1 + \frac{m}{3}(2m^2 + 12m - 8 + 6l(m+2)) \quad \text{and} \quad E_{6,3} = 6^{1/C_{6,3}}, \tag{34}$$

$$C_{6,4} = (2m^2 + m)\mu_0 + m^2\mu_1 + \frac{m}{3}(2m^2 + 39m - 5 + 9l(m+3)) \quad \text{and} \quad E_{6,4} = 6^{1/C_{6,4}}. \tag{35}$$

3.1. Comparison among the Efficiencies

To compare the iterative methods $H_{\rho,i}$, we consider the ratio

$$R_{p,i;q,j} = \frac{\log E_{p,i}}{\log E_{q,j}} = \frac{\log(p)\, C_{q,j}}{\log(q)\, C_{p,i}}. \tag{36}$$

It is clear that if $R_{p,i;q,j} > 1$, the iterative method $H_{p,i}$ is more efficient than $H_{q,j}$. Moreover, if we need to compare the methods having the same order then using (36) we can say that the iterative method $H_{p,i}$

is more efficient than $H_{q,j}$, if $C_{q,j} > C_{p,i}$. This means that the comparison of the methods possessing same order can be done by just comparing their computational costs.

$H_{6,1}$ *versus* $H_{6,2}$ *case:*

In this case comparison of the corresponding values of $C_{6,1}$ and $C_{6,2}$ in (28) and (33) gives $E_{6,1} > E_{6,2}$ for all $m \geqslant 4$ and $l \geqslant 1$.

$H_{6,1}$ *versus* $H_{6,3}$ *case:*

Comparing the corresponding values of $C_{6,1}$ and $C_{6,3}$ in (28) and (34) we obtain $E_{6,1} > E_{6,3}$ for all $m \geqslant 2$ and $l \geqslant 1$.

$H_{6,1}$ *versus* $H_{6,4}$ *case:*

Comparison of computational costs $C_{6,1}$ and $C_{6,4}$ in (28) and (34) gives $E_{6,1} > E_{6,4}$ for all $m \geqslant 2$ and $l \geqslant 1$.

In addition to the above comparisons we compare the proposed methods $H_{6,1}$ and $H_{9,1}$ with each other.

$H_{9,1}$ *versus* $H_{6,1}$ *case:*

The particular boundary $R_{9,1;6,1} = 1$ expressed by μ_0 written as a function of μ_1 and m is

$$\mu_0 = \frac{1}{12} \frac{2(r-s)m^2 + 3m(2\mu_1(r-s) + 3l(r-s) + 3r - 23s) + 3l(r-5s) - 22s}{m(s-r) + s}. \tag{37}$$

where $r = \ln(3)$ and $s = \ln(2)$. This function has a vertical asymptote for $m = s/(r-s) = 1.7095$.

Note that for $m \geqslant 44$, the numerator of (37) is positive and the denominator is negative, which shows that μ_0 is always negative for $m \geqslant 44$. That is, the boundary is out of admissible region for $m \geqslant 44$ and we have $E_{9,1} > E_{6,1} \ \forall \, (\mu_1, \mu_0) \in (0, +\infty) \times (0, +\infty)$ and $l \geqslant 1$.

We summarize the above results in following theorem:

Theorem 3. *For all $\mu_0 > 0$, $\mu_1 > 0$ and $l \geqslant 1$ we have:*
(a) $E_{6,1} > E_{6,2} \ \forall \, m \geqslant 4$.
(b) $E_{6,1} > E_{6,3}$ *and* $E_{6,1} > E_{6,4} \ \forall \, m \geqslant 2$.
(c) $E_{9,1} > E_{6,1} \ \forall \, m \geqslant 44$.
Otherwise, the efficiency comparison depends on m, μ_0, μ_1 and l.

4. Numerical Results

This section is devoted to check the effectiveness and efficiency of some of our proposed methods on different types of applications. In all cases, we apply our proposed schemes $H_{6,1}$ and $H_{9,1}$ and compare the results with those obtained by known methods $H_{6,2}$, $H_{6,3}$, and $H_{6,4}$. We consider academic examples, a special case of a nonlinear conservative problem which is transformed in a nonlinear system by approximating the derivatives by divided differences and, also the approximation of the solution of an elliptic partial differential equation that model a chemical problem.

All the experiments have been carried out in Matlab 2017 with variable precision arithmetics with 1000 digits of mantissa. These calculations have been made with an Intel Core processor i7-4700HQ with a CPU of 2.40GHz and 16.0 GB al RAM memory. In the tables we include the number of iterations (iter), the residual error of the corresponding function ($\|F(x^{(k+1)})\|$) in the last iterated and the difference between the two last iterates ($\|x^{(k+1)} - x^{(k)}\|$). We also present the approximated computational order of convergence (ACOC) defined in [14] with the expression

$$ACOC \approx \frac{\log\left(\|x^{(k+1)} - x^{(k)}\| / \|x^{(k)} - x^{(k-1)}\|\right)}{\log\left(\|x^{(k)} - x^{(k-1)}\| / \|x^{(k-1)} - x^{(k-2)}\|\right)}.$$

When the components of this vector are stable, it is an approximation of the theoretical order of convergence. In other case, it does not give us any information and we will denote it by $-$. Also the execution time (in seconds) has been calculated (by means of "cputime" Matlab command) with the mean value of 100 consecutive executions, for big-sized systems corresponding to Examples 1 to 3.

4.1. Example 1

We consider the case of a nonlinear conservative problem described by the differential equation

$$y''(t) + \Phi(y(t)) = 0, \ t \in [0, 1],$$

with the boundary conditions

$$y(0) = y(1) = 0.$$

We transform this boundary problem into a system of nonlinear equations by approximating the second derivative by a divided difference of second order. We introduce points $t_i = 0 + ih$, $i = 0, 1, \ldots, m+1$, where $h = \frac{1}{m+1}$ and m is an appropriate positive integer. A scheme is then designed for the determination of numbers y_i as approximations of the solution $y(t)$ at point t_i. By using divided differences of second order

$$y''(t_i) \approx \frac{y_{i+1} - 2y_i + y_{i-1}}{h^2},$$

we transform the boundary problem into the nonlinear system

$$y_{i+1} - 2y_i + y_{i-1} + h^2\Phi(y_i) = 0, \ i = 1, 2, \ldots, m. \tag{38}$$

Introducing vectors

$$y = (y_1, y_2, \ldots, y_m)^T, \ \Phi_y = (\Phi(y_1), \Phi(y_2), \ldots, \Phi(y_m))^T$$

and the matrix of size $m \times m$

$$A = \begin{pmatrix} -2 & 1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & \cdots & 0 \\ 0 & 1 & -2 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & -2 \end{pmatrix},$$

system (38) can be written in the form

$$F(y) \equiv Ay + h^2\Phi_y = 0.$$

In this case, we choose the law $\Phi(y(t)) = 1 + y(t)^3$ for the heat generation in the boundary problem and we solve system (38) by using iterative methods $H_{6,1}$, $H_{6,2}$, $H_{6,3}$, $H_{6,4}$, and $H_{9,1}$. In all cases, we use as initial estimation $x^{(0)} = (0.5, 0.5, \ldots, 0.5)^T$ and the stoping criterium $\|x^{(k+1)} - x^{(k)}\| < 10^{-100}$ or $\|F(x^{(k+1)})\| < 10^{-100}$. We can see in Table 1 the results obtained for $m = 20$ and $m = 50$.

There are no significant differences among the results obtained for different values of step h, i.e. for different sizes of the nonlinear system resulting from discretization. Let us remark that, for the lowest number of iterations, the best results in terms of lowest residuals have been obtained by methods $H_{9,1}$ and $H_{6,1}$.

Table 1. Numerical results for conservative boundary problem.

Method		$H_{6,1}$	$H_{6,2}$	$H_{6,3}$	$H_{6,4}$	$H_{9,1}$
	ACOC	5.5833	6.0210	-	-	6.2081
	iter	3	3	4	4	3
$m = 20$	$\|x^{(k+1)} - x^{(k)}\|$	2.78×10^{-35}	1.37×10^{-34}	2.69×10^{-98}	3.59×10^{-96}	8.63×10^{-59}
	$\|F(x^{(k+1)})\|$	6.10×10^{-125}	1.17×10^{-101}	4.55×10^{-229}	8.10×10^{-225}	1.87×10^{-210}
	CPU time (seconds)	0.025	0.024	0.028	0.026	0.022
	ACOC	3.1024	2.1869	-	-	6.0462
	iter	3	3	4	4	3
$m = 50$	$\|x^{(k+1)} - x^{(k)}\|$	3.81×10^{-34}	2.08×10^{-34}	6.93×10^{-97}	9.18×10^{-95}	2.24×10^{-57}
	$\|F(x^{(k+1)})\|$	2.76×10^{-121}	8.62×10^{-101}	1.03×10^{-225}	1.80×10^{-221}	7.16×10^{-206}
	CPU time (seconds)	0.042	0.042	0.043	0.042	0.044

4.2. Example 2

Let us consider the system of nonlinear equations

$$\sum_{j=1, j \neq i}^{m} x_j - e^{-x_i} = 0, \quad i = 1, 2, \ldots, m,$$

for an arbitrary positive integer m. We solve this system whit the same schemes as before, using as initial guess $x^{(0)} = (1, 1, \ldots, 1)^T$ and two values for the size, $m = 20$ and $m = 50$, being the solution $\alpha = (0.05, 0.05, \ldots, 0.05)^T$ and $\alpha = (0.02, 0.02, \ldots, 0.02)^T$, respectively. The stopping criterium used is again $\|x^{(k+1)} - x^{(k)}\| < 10^{-100}$ or $\|F(x^{(k+1)})\| < 10^{-100}$, that is, the process finishes when one of them is satisfied. The obtained results are shown in Table 2.

Table 2. Numerical results for Example 2.

Method		$H_{6,1}$	$H_{6,2}$	$H_{6,3}$	$H_{6,4}$	$H_{9,1}$
	ACOC	5.9898	5.9078	5.9248	5.9442	8.4359
	iter	3	3	3	3	3
$m = 20$	$\|x^{(k+1)} - x^{(k)}\|$	3.10×10^{-45}	6.62×10^{-46}	1.67×10^{-46}	3.55×10^{-47}	8.19×10^{-78}
	$\|F(x^{(k+1)})\|$	3.45×10^{-155}	1.94×10^{-127}	1.24×10^{-128}	5.59×10^{-130}	6.49×10^{-271}
	CPU time (seconds)	0.031	0.028	0.030	0.028	0.029
	ACOC	4.3931	2.0177	5.8962	5.9425	7.0463
	iter	3	3	3	3	3
$m = 50$	$\|x^{(k+1)} - x^{(k)}\|$	1.04×10^{-49}	1.10×10^{-52}	2.90×10^{-53}	8.37×10^{-54}	2.50×10^{-83}
	$\|F(x^{(k+1)})\|$	9.16×10^{-170}	6.01×10^{-142}	4.15×10^{-143}	3.46×10^{-144}	5.37×10^{-289}
	CPU time (seconds)	0.060	0.058	0.058	0.061	0.059

In this case, all the methods use the same number of iterations to achieve the solution, but the lowest residual are those of proposed schemes $H_{9,1}$ and $H_{6,1}$.

4.3. Example 3

Gas dynamics can be modeled by a boundary problem described by the following elliptic partial differential equation and the boundary conditions:

$$
\begin{aligned}
u_{xx} + u_{yy} &= u^3, \quad (x, y) \in [0, 1] \times [0, 1], \\
u(x, 0) &= 2x^2 - x + 1, \qquad u(x, 1) = 2, \\
u(0, y) &= 2y^2 - y + 1, \qquad u(1, y) = 2.
\end{aligned}
\tag{39}
$$

By using central divided differences and step $h = 1/5$ in both variables, this problem is discretized in the nonlinear system

$$F(x) = Ax + h^2 \phi(x) - b = 0,$$

where

$$A = \begin{pmatrix} B & -I & 0 & 0 \\ -I & B & -I & 0 \\ 0 & -I & B & -I \\ 0 & 0 & -I & B \end{pmatrix}, B = \begin{pmatrix} 4 & -1 & 0 & 0 \\ -1 & 4 & -1 & 0 \\ 0 & -1 & 4 & -1 \\ 0 & 0 & -1 & 4 \end{pmatrix},$$

being I the identity matrix of size 4×4, $\phi(x) = (x_1^3, x_2^3, \ldots, x_{16}^3)^T$ and

$$b = \left(\frac{44}{25}, \frac{23}{25}, \frac{28}{25}, \frac{87}{25}, \frac{23}{25}, 0, 0, 2, \frac{28}{25}, 0, 0, 2, \frac{87}{25}, 2, 2, 4 \right)^T.$$

To solve problem (39) we have used as initial guess $x^{(0)} = (1, 1, ..., 1)^T$. Also the stopping criteria $\|F(x^{(k+1)})\| < 10^{-100}$ or $\|x^{(k+1)} - x^{(k)}\| < 10^{-100}$ have been used and the process finishes when one of them is satisfied or the number of iterations reachs to 50.

The results are shown in Table 3. The first column shows the numerical aspects (approximated computational order of convergence ACOC, last difference between consecutive iterates $\|x^{(k+1)} - x^{(k)}\|$ and residual $\|F(x^{(k+1)})\|$) analyzed for the schemes used to solve the problem. In the rest of the columns we show the numerical results obtained by the methods $H_{6,1}$, $H_{6,2}$, $H_{6,3}$, $H_{6,4}$, and $H_{9,1}$.

Table 3. Numerical results for elliptic partial differential equation.

Method	$H_{6,1}$	$H_{6,2}$	$H_{6,3}$	$H_{6,4}$	$H_{9,1}$
ACOC	3.0100	2.0107	5.6132	5.8724	5.2651
iter	3	3	3	3	3
$\|x^{(k+1)} - x^{(k)}\|$	4.51×10^{-40}	1.82×10^{-48}	3.74×10^{-47}	3.56×10^{-46}	6.95×10^{-67}
$\|F(x^{(k+1)})\|$	6.27×10^{-138}	4.19×10^{-129}	1.67×10^{-126}	1.52×10^{-124}	2.45×10^{-234}
CPU time (seconds)	0.035	0.036	0.035	0.035	0.035

The low number of iterations needed justify that the ACOC does not approximate properly the theoretical order of convergence. The methods giving lowest exact error at the last iteration are again $H_{9,1}$ and $H_{6,1}$.

4.4. Example 4

Let us consider the nonlinear two-dimensional system $F(x_1, x_2) = 0$ with coordinate functions

$$\begin{aligned} f_1(x_1, x_2) &= \log(x_1^2) - 2\log(\cos(x_2)), \\ f_2(x_1, x_2) &= x_1 \tan\left(\frac{x_1}{\sqrt{2}} + x_2\right) - \sqrt{2}. \end{aligned} \tag{40}$$

This system has two real solutions at points, approximately, $(0.954811, 0.301815)^T$ and $(-0.954811, -0.301815)^T$. By using $x^{(0)} = (1, 0.5)^T$ as initial estimation, we have obtained the results appearing in Table 4, in which the residuals $\|x^{(k+1)} - x^{(k)}\|$ and $\|F(x^{(k+1)})\|$ appear, for the three first iterations.

Table 4. Numerical results for Example 4.

	$H_{6,1}$	$H_{6,2}$	$H_{6,3}$	$H_{6,4}$	$H_{9,1}$
$\|x^{(1)} - x^{(0)}\|$	1.90×10^{-1}	1.55×10^{-1}	1.45×10^{-1}	1.27×10^{-1}	2.01×10^{-1}
$\|x^{(2)} - x^{(1)}\|$	1.44×10^{-2}	5.69×10^{-2}	7.36×10^{-2}	1.18×10^{-1}	2.62×10^{-3}
$\|x^{(3)} - x^{(2)}\|$	1.07×10^{-9}	$5.15{-}7$	1.25×10^{-5}	1.74×10^{-4}	1.69×10^{-18}
$\|F(x^{(1)})\|$	4.12×10^{-2}	1.21×10^{-1}	1.41×10^{-1}	1.85×10^{-1}	7.56×10^{-3}
$\|F(x^{(2)})\|$	2.41×10^{-9}	2.65×10^{-6}	3.05×10^{-5}	2.99×10^{-4}	6.77×10^{-18}
$\|F(x^{(3)})\|$	3.21×10^{-44}	1.54×10^{-23}	6.19×10^{-20}	6.93×10^{-16}	5.39×10^{-86}

It can be observed that, although the difference between two consecutive iterations is not very small, the precision in the estimation of the root is very high, being the best for proposed schemes $H_{6,1}$ and $H_{9,1}$.

Moreover, dynamical planes help us to get global information about the convergence process. In Figure 1 we can see the dynamical planes of the proposed methods $H_{6,1}$ and $H_{9,1}$ and known schemes $H_{6,2}$, $H_{6,3}$, and $H_{6,4}$ when they are applied on system $F(x_1, x_2) = 0$. This figures are obtained by using the routines described in [15]. To draw these images, a mesh of 400×400 initial points has been used, 80 was the maximum number of iterations involved and 10^{-3} the tolerance used as the stopping criterium. In this paper, we have used a white star to show the roots of the nonlinear system. A color is assigned to each initial estimation (each point of the mesh) depending on where they converge to: Blue and orange correspond to the basins of the roots of the system (40) (brighter as lower is the number of iterations needed to converge) and black color is adopted when the maximum number of iterations is reached or the process diverges.
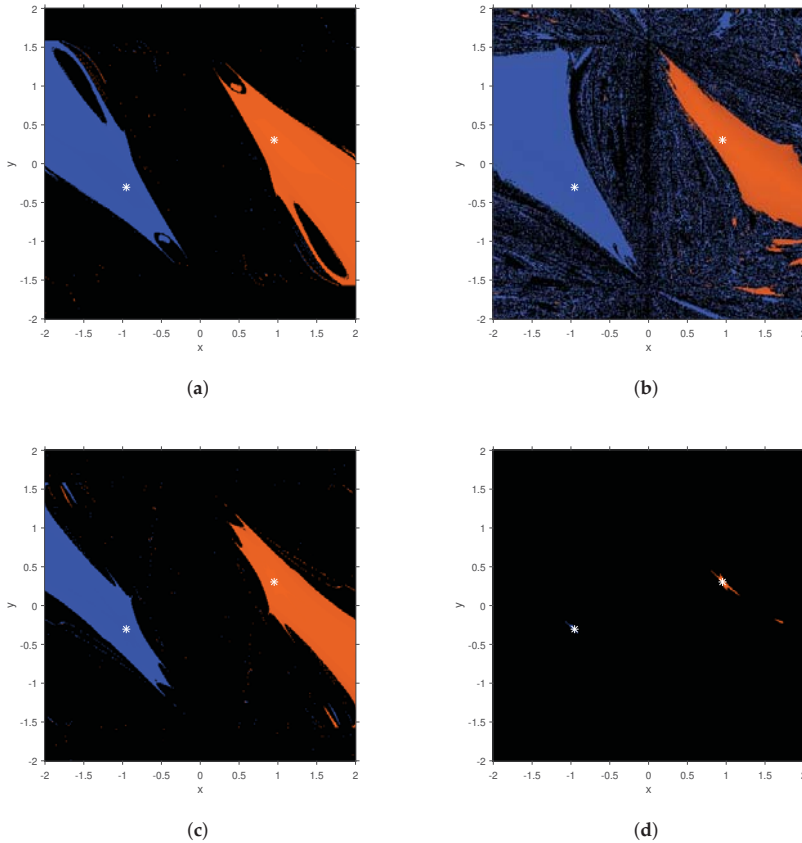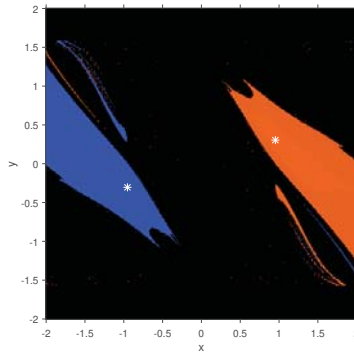

(a)

(b)

(c)

(d)

Figure 1. *Cont.*

(e)

Figure 1. Basins of attraction of known and proposed methods on $F(x_1, x_2) = 0$. (**a**) $H_{6,1}$; (**b**) $H_{6,2}$; (**c**) $H_{6,3}$; (**d**) $H_{6,4}$; (**e**) $H_{9,1}$.

In Figure 1, the shape and wideness of the basins of attraction show that $H_{6,1}$, $H_{6,2}$, $H_{6,3}$, and $H_{6,4}$ can found any of both roots by using a great amount of initial estimations, some of them far from the roots. However, $H_{6,4}$ is hardly able to find the roots and their basins are very small.

4.5. Example 5

Let us consider the nonlinear two-dimensional system $G(x_1, x_2) = 0$ with coordinate functions

$$g_1(x_1, x_2) = x_1^2 + x_2^2 - 1, \tag{41}$$

$$g_2(x_1, x_2) = x_1^2 - x_2^2 + \frac{1}{2}. \tag{42}$$

This system has four real solutions at points $\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)^T$, $\left(\frac{1}{2}, -\frac{\sqrt{3}}{2}\right)^T$, $\left(-\frac{1}{2}, \frac{\sqrt{3}}{2}\right)^T$, and $\left(-\frac{1}{2}, -\frac{\sqrt{3}}{2}\right)^T$. Let us remark that they are symmetric as the system shows the intersection between two conical curves, an ellipse and a circumference. By using $x^{(0)} = (1, 1)^T$ as initial estimation, we have obtained the results appearing in Table 5, in which the residuals $\|x^{(k+1)} - x^{(k)}\|$ and $\|F(x^{(k+1)})\|$ also appear, for the three first iterations.

Table 5. Numerical results for Example 5.

	$H_{6,1}$	$H_{6,2}$	$H_{6,3}$	$H_{6,4}$	$H_{9,1}$
$\|x^{(1)} - x^{(0)}\|$	5.10×10^{-1}	5.15×10^{-1}	5.125×10^{-1}	5.10×10^{-1}	5.16×10^{-1}
$\|x^{(2)} - x^{(1)}\|$	7.96×10^{-3}	2.38×10^{-3}	5.63×10^{-3}	8.30×10^{-3}	1.46×10^{-3}
$\|x^{(3)} - x^{(2)}\|$	6.03×10^{-12}	3.54×10^{-16}	3.60×10^{-13}	8.89×10^{-12}	1.14×10^{-23}
$\|F(x^{(1)})\|$	1.13×10^{-2}	3.37×10^{-3}	8.00×10^{-3}	1.18×10^{-2}	2.07×10^{-3}
$\|F(x^{(2)})\|$	8.53×10^{-12}	5.00×10^{-16}	5.10×10^{-13}	1.26×10^{-11}	1.61×10^{-23}
$\|F(x^{(3)})\|$	2.56×10^{-56}	8.87×10^{-62}	8.99×10^{-57}	5.02×10^{-54}	6.87×10^{-161}

The numerical results appearing in Table 5 show as the estimation of the roots has similar errors in all sixth order methods, although $H_{6,2}$ and $H_{6,3}$ are slightly better than $H_{6,1}$. Indeed, the ninth order scheme has the best precision in the approximation of the roots of $G(x_1, x_2)$, as corresponds to its high order of convergence. In terms of computational time, the mean of one hundred consecutive iterations

have been used in order to get good estimations of the real time and, in case of proposed methods, they use similar times as the existing schemes getting better accuracy.

In Figure 2 the colors blue, orange, green and purple correspond to the basins of the roots of the system (41). We can see the dynamical planes of the proposed method methods $H_{6,1}$ and $H_{9,1}$ and known schemes $H_{6,2}$, $H_{6,3}$, and $H_{6,4}$ when they are applied on system $G(x_1, x_2) = 0$.
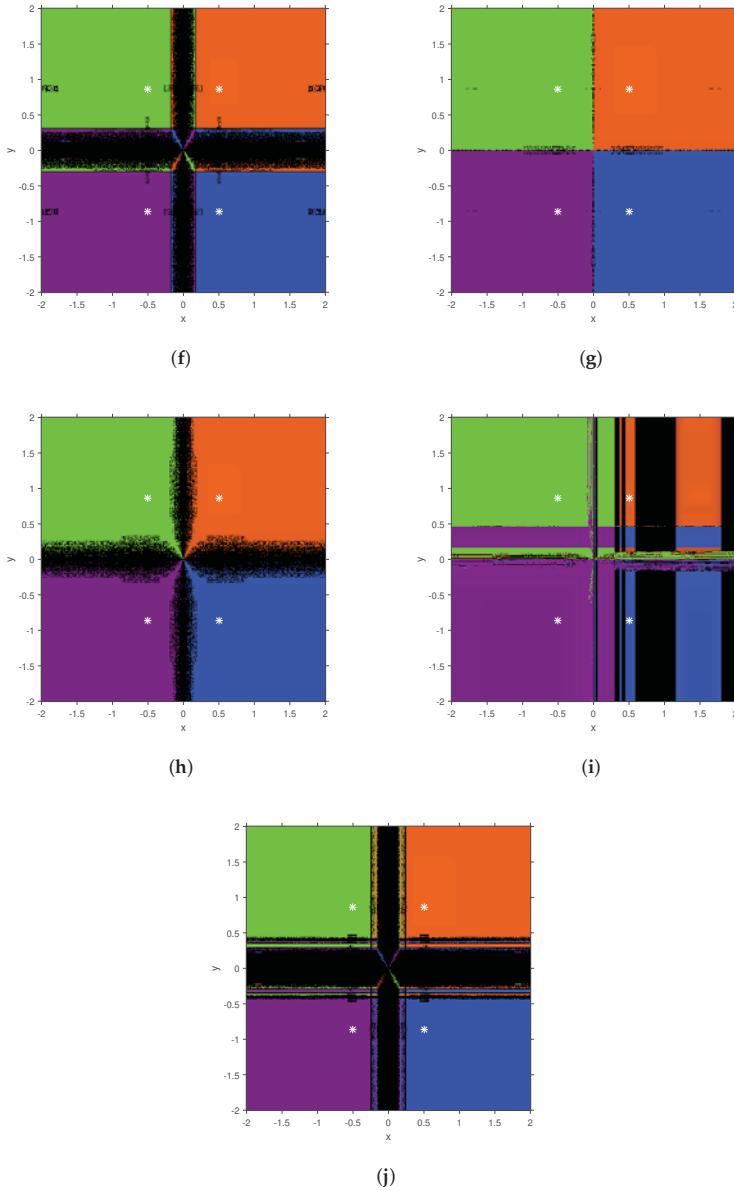
(f)

(g)

(h)

(i)

(j)

Figure 2. Basins of attraction of known and proposed methods on $G(x_1, x_2) = 0$. (a) $H_{6,1}$; (b) $H_{6,2}$; (c) $H_{6,3}$; (d) $H_{6,4}$; (e) $H_{9,1}$.

Regarding the stability of proposed and known iterative processes on $G(x_1, x_2)$, it can be observed in Figure 2 that, in all cases except $H_{6,4}$, the connected component of the basins of attraction that holds the root (usually known as immediate basin of attraction) is very wide. For all iterative methods black areas of low convergence or divergence appear, being more diffuse in case of $H_{6,2}$ and wider in $H_{6,4}$. Methods $H_{6,1}$, $H_{6,3}$ and $H_{9,1}$ present black areas that are mainly regions of slower convergence, that would be colored if a higher number of iterations is fixed as a maximum; meanwhile, the biggest black area of the dynamical plane associated to method $H_{6,4}$ corresponds to divergence.

5. Concluding Remarks

From Potra–Pták third order scheme, an efficient sixth order method for solving nonlinear system is proposed. Moreover, it has been extended by adding subsequent steps with the same structure that adds a new functional evaluation per new step. Then the order of the resulting procedure increases in three units per step. The proposed methods have been shown to be more efficient that several known methods of the same order. Some numerical tests with academic and real-life problems have been made to check the efficiency and applicability of the designed procedures. The numerical performance have been opposed to the stability shown by the dynamical planes of new and known methods on two-dimensional systems.

Author Contributions: Methodology, H.A.; Writing—original draft, J.R.T.; Writing—review & editing, A.C.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ostrowski, A.M. *Solution of Equations and Systems of Equations*; Academic Press: New York, NY, USA, 1966.
2. Ortega, J.M.; Rheinboldt, W.C. *Iterative Solution of Nonlinear Equations in Several Variables*; Academic Press: New York, NY, USA, 1970.
3. Kelley, C.T. *Solving Nonlinear Equations with Newton's Method*; SIAM: Philadelphia, PA, USA, 2003.
4. Traub, J.F. *Iterative Methods for the Solution of Equations*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1964.
5. Homeier, H.H.H. A modified Newton method with cubic convergence: The multivariable case. *J. Comput. Appl. Math.* **2004**, *169*, 161–169. [CrossRef]
6. Darvishi, M.T.; Barati, A. A fourth-order method from quadrature formulae to solve systems of nonlinear equations. *Appl. Math. Comput.* **2007**, *188*, 257–261. [CrossRef]
7. Cordero, A.; Hueso, J.L.; Martínez, E.; Torregrosa, J.R. A modified Newton-Jarratt's composition. *Numer. Algor.* **2010**, *55*, 87–99. [CrossRef]
8. Cordero, A.; Hueso, J.L.; Martínez, E.; Torregrosa, J.R. Efficient high-order methods based on golden ratio for nonlinear systems. *Appl. Math. Comput.* **2011**, *217*, 4548–4556. [CrossRef]
9. Grau-Sánchez, M.; Grau, À.; Noguera, M. On the computational efficiency index and some iterative methods for solving systems of nonlinear equations. *J. Comput. Appl. Math.* **2011**, *236*, 1259–1266. [CrossRef]
10. Grau-Sánchez, M.; Grau, À.; Noguera, M. Ostrowski type methods for solving systems of nonlinear equations. *Appl. Math. Comput.* **2011**, *218*, 2377–2385. [CrossRef]
11. Potra, F-A.; Pták, V. *Nondiscrete Induction and Iterarive Processes*; Pitman Publishing: Boston, MA, USA, 1984.
12. Grau-Sánchez, M.; Noguera, M.; Amat, S. On the approximation of derivatives using divided difference operators preserving the local convergence order of iterative methods. *J. Comput. Appl. Math.* **2013**, *237*, 363–372. [CrossRef]
13. Sharma, J.R.; Arora, H. On efficient weighted-Newton methods for solving systems of nonlinear equations. *Appl. Math. Comput.* **2013**, *222*, 497–506. [CrossRef]

14. Cordero, A.; Torregrosa, J. R. Variants of Newton's method using fifth order quadrature formulas. *Appl. Math. Comput.* **2007**, *190*, 686–698. [CrossRef]

15. Chicharro, F.I.; Cordero A.; Torregrosa, J.R. Drawing dynamical and parameter planes of iterative families and methods. *Sci. World J.* **2013**, *2013*, 780153. [CrossRef] [PubMed]

Evolution of an Exponential Polynomial Family of Discrete Dynamical Systems

Francisco Solis![ORCID]

Center of Research in Mathematics (CIMAT), Guanajuato 36023, Mexico; solis@cimat.mx; Tel.: +52-473-7327155

Abstract: In this paper, we introduce and analyze a family of exponential polynomial discrete dynamical systems that can be considered as functional perturbations of a linear dynamical system. The stability analysis of equilibria of this family is performed by considering three different parametric scenarios, from which we show the intricate and complex dynamical behavior of their orbits.

Keywords: exponential polynomial; discrete dynamical systems; convergence

1. Introduction

One of the most important properties of a discrete dynamical system (DDS) is its asymptotic behavior. Most systems exhibit diverse types of transient behavior, followed by convergence to an invariant attracting region in space. It is interesting to note that the transient behavior is defined as the behavior of the orbits before they settle in the attracting set, but no one seems to explain or give conditions when this phenomenon has expired [1,2]. The study of discrete systems requires appropriate analysis methods to extract and characterize their transient behaviors. Such analysis should allow both the classification of the transient motion and the parametric variation. Different efforts to develop analytical methods have been made in fields such as biology [3–6], physics [7–11], engineering [12,13], and ecology [14,15]. Even in medicine, epilepsy may be considered as a transient periodic phenomenon [16,17]. In economics, there are conditions for practical asymptotic stability and approximately optimal transient performance [18]. However, the methods developed are usually applied to describe specific behaviors. In consequence, these methods are rather particular and can hardly be applied to other different systems. Moreover, some methods are not well-developed or may fail to notice important factors, such as the slow convergence of the orbits at points of slow convergence that includes bifurcation points (see References [19,20]). Thus, in addition to practical implications, the present study aims to gain insights by analyzing a particular but novel family of discrete dynamical systems.

Our goal in this work was to introduce and study a family of discrete systems that exhibits a large variety of transient behaviors with slow convergence not at specific points but for most parametric values. Such a family is given by exponential polynomials that act as iteration functions for such systems [21,22]. It is known that real exponential polynomials often appear as solutions of differential equations and also in theoretical applications, as in transcendental number theory, where they appear as auxiliary functions [23]. In our context, the inclusion of these functions generates a complicated behavior of the system orbits and leads to controversial dilemmas between theory and practice. The analysis of these exponential polynomial dynamical systems also shows how relationships between different terms of the iteration function give rise to a complex collective behavior of the system. Moreover, even when the analysis of a nonlinear discrete system is done correctly, the numerical evaluation of its orbits needs to be consistent with the theory. Here, we show that transient behavior does not correspond to asymptotic theoretical behavior and also contradicts theory showing false equilibria and also displaying unstable features of the system. Additionally, the usual

local linearization around equilibria makes sense only when the fixed points can be calculated and needs to reflect the correctness and meaningful of the numerical evaluations, which, in practice, is far from true for exponential polynomials. We explain this new paradox and give implications of paramount importance in the study of nonlinear discrete dynamical systems. It is important to remark that the equations from which models of exponential polynomial systems are developed generally derive from diverse areas, such as population models, models for infectious diseases, and so forth.

This paper is organized as follows: In Section 2, we introduce exponential polynomials in order to define the family of discrete dynamical systems of our interest, and we also provide its relationship with known discrete systems. In Section 3, the stability analysis of equilibria for this family is performed by dividing it into three scenarios: A perturbed linear case, a basic exponential polynomial case defined a posteriori, and, finally, a mixed case. Next, in Section 4 we discuss the transient behavior of the basic exponential polynomial case, since in practical situations, it presents a fictitious asymptotic behavior with geometric characteristics similar to the unstable equilibria. The general conclusions of this work are summarized in Section 5. All numerical calculations in this work were done using our own codes written in the high-level programming language C with an Intel Xeon(R) CPU X5670 @ 2.93GHz × 6 processor.

2. Exponential Polynomial Family of DDS

The goal of this section is to introduce exponential polynomial discrete dynamical systems. In order to achieve this goal, our first step is to define the corresponding iteration functions for those systems. Thus, we have the following definition:

Definition 1. *An exponential polynomial function with polynomial coefficients and exponents is a function* $f : \mathbb{R} \to \mathbb{R}$ *of the form:*

$$f(x) = p_1(x)e^{q_1(x)} + p_2(x)e^{q_2(x)} + \cdots + p_k(x)e^{q_k(x)}, \tag{1}$$

where k is a fixed natural number and p_j *and* q_j *are polynomials* $j = 1, 2, \ldots, k$, *with at least one pair* p_j, q_j, *for some j, of nonconstant polynomials.*

An exponential polynomial will be called a basic exponential polynomial when $k = 1$. Functions of Equation (1) are typical solutions of linear differential equations. Moreover, they arise in different areas of applied and theoretical mathematics, for example, in neural network models with delays, in statistical mechanics, and in approximation theory (see References [24,25]). A review of exponential polynomials with a list of their properties can be found in Reference [26]. Leaving aside these specific applications, we focused specifically on the basic properties of the asymptotic behavior of parametric families of exponential polynomial discrete dynamical systems, which we define as follows:

Definition 2. *An exponential polynomial discrete dynamical system is a system of the form:*

$$x_{n+1} = f(\lambda, x_n),$$

where f is an exponential polynomial function and λ *is a parameter family.*

A treatment of these general dynamical systems is important and so far has not yet been fully implemented. Therefore, as a first step in this direction, we direct our efforts to analyze a very particular exponential family in order to show their diverse and intricate dynamical behavior.

Consider the family of exponential polynomial discrete dynamical systems given by:

$$x_{n+1} = f(\alpha, \beta, x_n) = \beta x_n + (2 - x_n)e^{-\alpha(x_n-1)^2}, \tag{2}$$

where the parameters satisfy $\beta \in [0,1]$, $\alpha \geq 0$ and also $x_n \geq 0$. System (2) generalizes typical linear discrete dynamical systems, known in several applications as exponential growth models with constant growth rates. The new system incorporates a new term that for small values of α can be considered as a small perturbation term, which biologically models a period of abundance of resources for the population. Notice how the function f is basically linear on x, except that it has a localized bump at $x \approx 1$. Additionally, the function f can be considered as a modification of a unimodal function, see Figure 1, where a linear growth replaces the typical downhill behavior. The importance of unimodal systems resides in their very similar dynamical behavior (see References [27–29]).
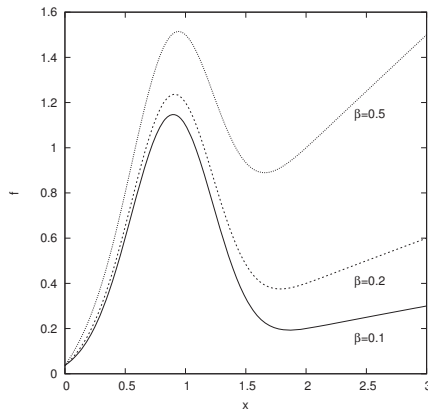

Figure 1. Graphs of the iteration function $f(\alpha, \beta, x)$ with $\alpha = 4$ and with different values of β.

3. Analysis of the Family

Let us analyze the family of systems (Equation (2)) by classifying its equilibria along with their stability properties. The fixed points of Equation (2) satisfy the equation:

$$(\beta - 1)x + (2 - x)e^{-\alpha(x-1)^2} = 0. \tag{3}$$

In order to focus on the intricacies of the number of roots of Equation (3) and their stability properties, we divided its analysis into three cases, namely, $\beta = 1$, which corresponds to a perturbed linear case, a basic exponential polynomial case given by $\beta = 0$, and, finally, a mixed case with $\beta \in (0,1)$. Let us start with the first case.

3.1. Perturbed Linear Case, $\beta = 1$

This particular case allows us to analyze a more general discrete dynamical system than System (2), namely:

$$x_{n+1} = x_n + P(x_n)e^{-\alpha Q(x_n)},$$

where P and Q are polynomials. The fixed points of this new system are given by the roots of $P(x)$. Assume that Q is nonnegative and P has different roots:

$$r_0 < r_1 < r_2 < \cdots < r_m,$$

satisfying $\left. \frac{|P(r_j)|}{|r-r_j|} \right|_{r=r_j} < 1$ for $j = 0, 1, \ldots, m$. With these assumptions, we have the following result:

Theorem 1. *If* $(-1)^m \frac{d^m P(x)}{dx^m} < 0$, *then the system* $x_{n+1} = F(x_n) = x_n + P(x_n)e^{-\alpha Q(x_n)}$ *has stable fixed points at roots of P with even subindexes* $x = r_j$, $j = 0, 2, 4, \ldots$ *and unstable fixed points at roots of P with odd*

subindexes $x = r_j$, $j = 1, 3, 5, \ldots$ for all positive values of α. Otherwise, the fixed points change stability, that is, roots with even (odd) subindexes are unstable (stable).

Proof. Notice that $F'(r_j) = 1 + P'(r_j)e^{-\alpha Q(r_j)} = 1 + \left. \frac{P(r_j)}{(r-r_j)} \right|_{r=r_j} e^{-\alpha Q(r_j)}$. □

In particular, System (2) has only one fixed point, given by $x = 2$, for every value of α. That is, a branch of fixed points that is constant for every value of α. A subregion of its basin of attraction is shown on Figure 2; notice that the system for $\alpha > 3.8$ requires more iterations to converge to the fixed point for initial conditions satisfying $x_0 < 2$. Since $\frac{\partial f}{\partial x}(\alpha, 1, 2) = 1 - e^{-\alpha} < 1$, the fixed point is always stable for all nonnegative values of α, which means that the local asymptotic behavior of System (2) is trivial; its orbits converge to this fixed point or diverge to infinity. What about in practical applications, though?
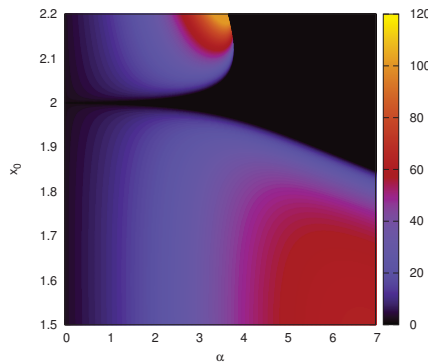

Figure 2. Basin of attraction of the fixed point $x = 2$ for $\alpha \in (0, 7)$ and initial conditions $x_0 \in [1.5, 2.2]$. Colors denotes the number of iterations within a tolerance of 10^{-3}.

In order to answer this question, we made use of bifurcation diagrams which display properties of the asymptotic solution of a dynamical system as a function of α, allowing one to see at a glance where qualitative changes in the asymptotic solution occur. Let us observe the bifurcation diagram for System (2) with $\beta = 1$ given in Figure 3a, where the initial condition is $x_0 = 0.05$ and the first 3000 iterations are discarded for each value of α. Here, the fixed point $x = 2$ seems to be stable only for values of α in (0, 5.56). Let us observe that after the value $\alpha = 5.66$, the branch of fixed points decreases down to the value of 1.982. We also obtained the same qualitative behavior when we varied the initial conditions, see Figure 3. For large α, the fixed point is only weakly attracting. In this respect, Figure 2 seems irrelevant, but the novelty of this figure is that it shows unreal fixed points, namely, those appearing after the value $\alpha = 5.66$.
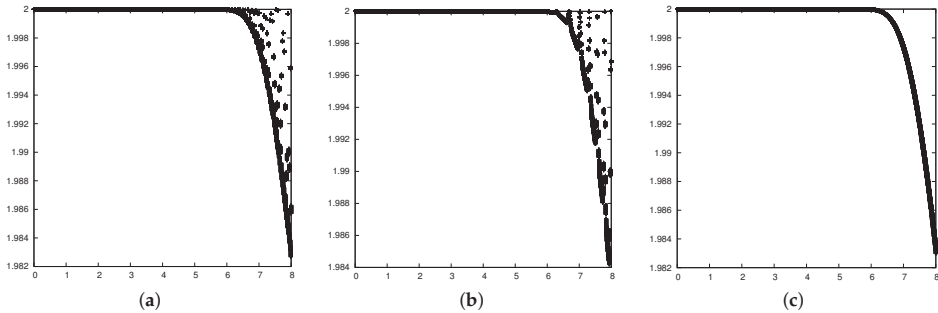
Figure 3. Bifurcation diagram of System (2) with initial conditions (**a**) 0.05, (**b**) 0.19, (**c**) 1.7. Horizontal axis is α and vertical axis is x_n. $n > 3000$ for (**a**) and $n > 8000$ for (**b**,**c**).

How can we explain this controversial phenomenon? Here is where the fact that we are dealing with an exponential polynomial as iteration function plays a significant role, meaning that the phenomenon of slowing down takes place [30]. Notice that for $\alpha \gg 1$, the system behaves as $x_{n+1} = x_n + \epsilon(x_n)$, $\epsilon(x_n)$ being a small perturbation depending on x_n. Thus, a large number of iterations is required before transient responses have died out. Also notice that the correct asymptotic value depends on the initial condition. In Figure 3b, we show the previous behavior, but we increased the number of iterations up to 8000 and then plotted the next one hundred iterations. Clearly, the numerical asymptotic behavior is correct up to $\alpha = 6.55$, whereas with 3000 iterations, it was correct only up to $\alpha = 5.56$. We carried out different simulations in order to obtain how the correct behavior depends on α and on initial conditions, from our results, we conclude that:

Theorem 2. *The system* $x_{n+1} = x_n + (2 - x_n)e^{-\alpha(x_n-1)^2}$ *has a stable fixed point at* $x = 2$ *for all positive values of* α.

This previous result is straightforward to verify, but, as we found out previously, in numerical calculations, the fixed point seems to vary with α. Therefore, special care must be taken to make sure that the orbits settle in the attracting fixed point. We have the following heuristic, which was found by bounding the number of iterations for $\alpha \in (0, 8]$ by numerical means:

Remark 1. *In order to have "practical" convergence to the fixed point, that is, convergence with a tolerance of* 10^{-7} *for* $\alpha \in (0, 8]$ *at least* $\left\lceil 1000e^{0.3(4+\epsilon)^2} \right\rceil$ *iterations must be performed with an initial value of* $2 - \epsilon$.

3.2. Basic Exponential Polynomial Case, $\beta = 0$.

Let us consider the basic exponential polynomial case, namely, the case when $\beta = 0$. For this case, the number of fixed points of System (2) depends on α. For all values of α, $x = 1$ is a branch of fixed points. Since $\frac{\partial f}{\partial x}(\alpha, 0, 1) = -1$, and computing the Schwarzian derivative of f (see References [31,32]), which is given in this case by $-\frac{\partial^3 f}{\partial x^3}(\alpha, 0, 1) - 1.5 \left[\frac{\partial^2 f}{\partial x^2}(\alpha, 0, 1)\right]^2 = -6\alpha(\alpha + 1)$, it follows that $x = 1$ is locally stable for $\alpha > 0$. For $\alpha \gtrsim 3.432987771$, there is another branch of fixed points which can not be obtained in an explicit analytical form. In Figure 5a, we plot the locus of fixed points that satisfy Equation (3), which consists of a horizontal line at $x = 1$ and a c-shaped curve with one branch that approaches asymptotically to zero, and the other branch approaches one as α goes to infinity. Stability is shown with black circles and nonstability with white circles. Therefore, there are three fixed points for every value of $\alpha \geq 3.432987771$. Before we continue analyzing the fixed points of the system, we state the following two definitions from Reference [32]:

Definition 3. *Let $x_{n+1} = f(\alpha, x_n)$ be a discrete dynamical system with a fixed point $x_f = 0$ that is attracting from the left (or from the right). The point (α_0, x_f) is r-slow if there exists a real number $r > 1$ such that:*

$$|f(\alpha_0, x) - x| = \mathcal{O}(x^r), \quad as \; x \to 0.$$

This definition is closely related to the concept of asymptotic expansions: A function $g(x)$ is asymptotic to a power series if, for fixed N and sufficiently small x:

$$g(x) - \sum_{j=0}^{N} a_j x^j = \mathcal{O}(x^{N+1}).$$

In our case, $g(x) = f(\alpha_0, x)$, $N = 1$, $a_0 = 0$ and $a_1 = 1$, and we are replacing the natural number $N + 1$ by a real number r. However, Definition 3 can not be used as a precise classification of slow discrete dynamical systems, since the order of a function usually only provides an upper bound on its growth rate. To avoid this inconvenience and provide a more precise classification of slowness, the following new definition is used:

Definition 4. *Let $x_{n+1} = f(\alpha, x_n)$ be a discrete dynamical system with a r-slow point (α_0, x_f). The slow point belongs to the η class of slowness if:*

$$\eta = \max\{r \in \mathbb{R} \mid (\alpha_0, x_f(\alpha_0)) \; is \; r - slow\}.$$

In our case, the fixed point $x = 1$ with arbitrary α is characterized as a slow fixed point belonging to the $\eta = 2$ class of slowness class, meaning that its rate of convergence is similar to the oscillating system $x_{n+1} = -x_n + x_n^2$. Thus, we expect that convergence for this system to the fixed point will require at least 10^5 iterations for all initial points in the range $[0, 0.99]$. More concretely, as an example for a tolerance of 10^{-3} (a number relatively small), the minimum number of required iterations is 249,733 for $\alpha \in (0, 1]$ with initial contition $x_0 = 0.5$, see Figure 4. Let us recall that for the case $\beta = 1$, the required number of iterations used in the examples was in the order of 10^4 iterations within a tolerance of 10^{-6}.
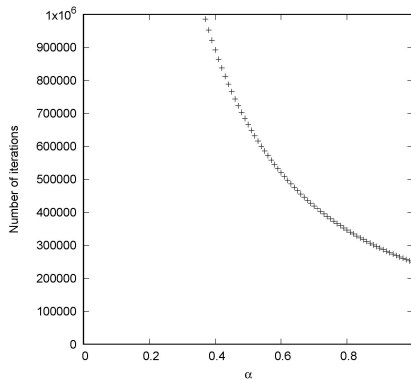

Figure 4. Number of iterations required to obtain convergence to the fixed point given a tolerance of 10^{-3} for $\alpha \in (0, 1)$ and initial condition $x_0 = 0.5$.

In Figure 5b, we show a bifurcation diagram, where the lower branch of the curve is stable, whereas the upper branch is unstable. Notice that in the same figure, for $1 < \alpha < 3.43$, there is a transient behavior (a "ghost" solution) , a behavior that will be discussed in the next section. Thus, we obtain the following theoretical result:

Theorem 3. *The system $x_{n+1} = (2 - x_n)e^{-\alpha(x_n-1)^2}$ has a fixed point at $x = 1$ for all values of α, which is locally stable for $\alpha > 0$. For $\alpha \gtrsim 3.432987771$, the system has two equilibria, the larger one unstable and the smaller one stable. This last equilibrium point approaches zero as α approaches ∞.*

By numerical evaluation of the orbits, we obtain (see Figure 5b)):

Remark 2. *The system $x_{n+1} = (2 - x_n)e^{-\alpha(x_n-1)^2}$ has oscillatory behavior for all initial conditions close to $x = 1$ and $\alpha \lesssim 3.432987771$. The oscillation consists of a 2 periodic orbit.*
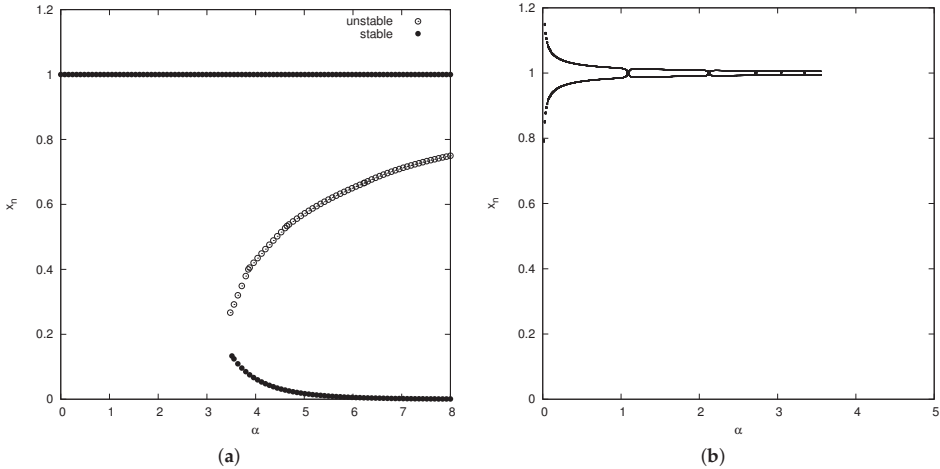

Figure 5. (**a**) Fixed points for System (2) when $\beta = 0$. (**b**) Bifurcation diagram with initial condition $x_0 = 0.3$ and 1000 initial iterations discarded.

3.3. Mixed Case $\beta \in (0, 1)$

So far, we have analyzed the two boundary values of β. For $\beta \in (0, 1)$, $x = 1$ is not longer a fixed point; instead, there are two branches of fixed points which are similar to those given in Figure 5a. The first branch consists of a decreasing curve that starts at the point $(0, \frac{2}{2-\beta})$ in the plane $x - \alpha$ and approaches decreasingly asymptotically to $x_f = 1$ as α goes to infinity. The second branch is similar to the c-shaped one for the case of $\beta = 0$, meaning that for $\alpha \in [0, b(\beta)]$, there is a unique fixed point, where the values of the function $b(\beta)$ are given in Figure 6a.

Along the first branch, the equilibrium remains stable for $\alpha \in (0, \gamma(\beta))$, where the values of function $\gamma(\beta)$ are also shown in Figure 6a. After that, the branch of equilibrium points becomes unstable, giving rise to a two periodic stable branch, which loses its stability at $b(\beta)$ for small initial conditions; for larger initial conditions, the two periodic branch bifurcates again to a fourth periodic branch and so on. For $\alpha \in (b(\beta), \infty)$, there are three fixed points, and only the smallest one is locally stable.

Notice that the smallest fixed point approaches zero when α goes to infinity, making the discrete system approach a unimodal system, which means that for larger values of α, the asymptotic behavior of the system is independent of initial conditions. It is important to remark that this scenario is the only case where stable periodicity is present. Moreover, the equilibria can not be shown explicitly.

Let us now vary parameter β in search of chaotic behavior. When parameter β is small (less than 0.26), the system basically behaves as in the previous scenario, meaning that the system is robust under such perturbation. However, as β increases, the system behaves totally different. Such a parameter increase triggers in the system a period doubling sequence to chaos similar to the behavior of

the classical logistic discrete dynamical system. This chaotic behavior is present for all values of $\beta \in (\beta_0, 1)$, where $\beta_0 \approx 0.26$. The values of parameter α where chaos is present depends on parameter β, as increasing β also increases the values of α (see Figure 6b,c).
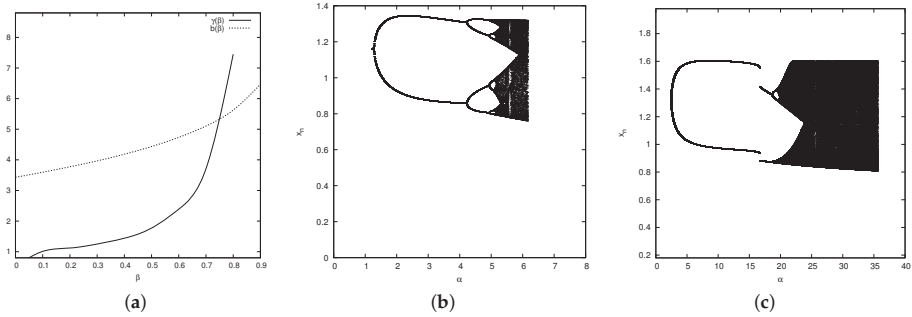

Figure 6. (**a**) Graph of $b(\beta)$. (**b**) Bifurcation diagram for $\beta = 0.3$. (**c**) Bifurcation diagram for $\beta = 0.6$.

4. Transient Behavior for the Case $\beta = 0$

From now on, we will center our attention to the case when $\beta = 0$. Figure 7a is a bifurcation diagram, which is a zoom of Figure 5b, obtained by choosing the initial condition $x_0 = 0.3$ and plotting two hundred iterates after the first one thousand iterations have been discarded. Now, if instead of discarding the first one thousand iterations, we wait longer and discard the first five thousand iterations, we obtain the bifurcation diagram given in Figure 7b. From these two diagrams, we notice that there are some distinctive features with complicated behavior. First, the discrete system is very slow in the sense of having a slow fixed point belonging to the $\eta = 2$ class of slowness class, as was defined previously. Thus, we would like to analyze the characteristic exponents associated with this system at all values of α, especially those where we have that the "links" of the chain in the bifurcation diagrams are connected.
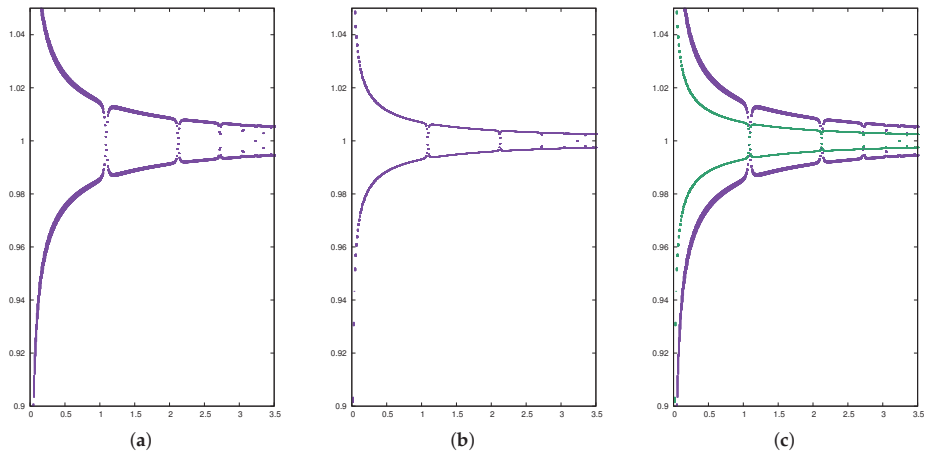

Figure 7. Bifurcation diagrams for $\beta = 0$ and initial condition $x_0 = 0.3$. Horizontal axis is α and vertical axis is x_n. (**a**) First 1000 iterations discarded. (**b**) First 5000 iterations discarded. (**c**) Both diagrams (**a**) and (**b**).

Let us define characteristic exponents to our system as defined by Hao [30]. Let x_f be a fixed point of the system $x_{n+1} = F(\alpha, x_n)$. Let $x_n = x_f + \epsilon_n$ and assume that $|\epsilon_n|$ decays as $exp(-n/\tau)$,

that is, we used the exponential function as a function of comparison. Then, we may define the time of convergence as:

$$\tau = - \left(\ln \left| \frac{\partial F}{\partial x}(\alpha, x_f) \right| \right)^{-1},\tag{4}$$

and define the critical exponent, Δ, as the smallest power of the nonzero term in the Taylor series of $\ln \left| \frac{\partial F}{\partial x}(\alpha, x_f) \right|$ around the point $|\alpha - \alpha_0|$, where α_0 satisfies $|F'(\alpha_0, xf)| = 1$. Hao discovered the exponent with value $\Delta = 1$ for discrete dynamical systems. In our particular case, $\ln \left| \frac{\partial F}{\partial x}(\alpha, x_f) \right| \equiv 0$. Therefore, the value of Δ is not even defined, which shows that our system is special and thus, it is neccesary to introduce new comparision functions in order to define a more general concept of exponent characteristics.

We already know that the behavior presented in the bifurcation diagram is not asymptotically correct, since the orbits must approach $x = 1$. Thus, our goal now is to discover why such a behavior is present in the numerical simulations. Let us consider System (2) again with $\beta = 0$:

$$x_{n+1} = (2 - x_n)e^{-\alpha(x_n-1)^2} = (2 - x_n)\sum_{k=0}^{\infty} \frac{(-1)^k \alpha^k}{k!}(x_n - 1)^{2k}.$$

If we truncated the previous series just to get only two terms from the series and substituting $x_n = y_n + 1$, then we get the following cubic system:

$$y_{n+1} = -y_n + \alpha y_n^2(y_n - 1).$$

which has zero as an equilibrium point (same as the nontruncated case) and two extra fixed points at $y = \frac{\alpha \pm \sqrt{\alpha(\alpha+8)}}{2\alpha}$. Those are spurious equilibria for the nontruncated case. The Schwarzian derivative evaluated at zero is given by $-6\alpha(1 + \alpha)$, which is negative for all nonnegative values of α. Therefore, zero is locally stable for all $\alpha > 0$. However, the two other equilibrium points are unstable. The fact that zero is also a slow equilibrium point belonging to the $\eta = 2$ class of slowness class explains why the system does not show signs of convergence to the stable equilibrium point. To give a better understanding of this slow behavior, one may compare this system to $x_{n+1} = x_n \cos(\sqrt{x_n})$. Both systems have the same value of slowness class. Using asymptotic analysis and telescopic cancellation, we can show that $x_n = \frac{1}{2n} + \mathcal{O}(n^2 \log(n))$, which explains the slow behavior. Notice also that the shape of the spurious equilibria corresponds to the shape of the behavior illustrated on Figure 8a. Bifurcation diagrams for the cubic system and for a quartic system are given in Figure 8b,c.

The distinctive points that simulate links (points of breakdown in the pace of convergence) correspond to eventually periodic points (see Reference [19]), and they generate a converging sequence $\{\alpha_m\}_{m=1}^{\infty}$ to the value 3.43298771..., which is the value where the second branch of fixed points arises. Each value α_m is characterized for a fixed initial condition x_0 by $\alpha_m = f^{-1}(\alpha_{m-1}, x_o)$ with $\alpha_1 = 2 - x_0/(\ln|2 - x_0|)$. This sequence is well defined because f is a decreasing function on α and convergent because it is increasing and bounded. The limit is independent of initial conditions for appropriate connected sets that include the fixed point, and therefore, it inherits universal properties. Therefore, the system also shows universal properties that are characteristic of unimodal maps, and they also appear in applications with chaotic oscillations in mechanical systems.
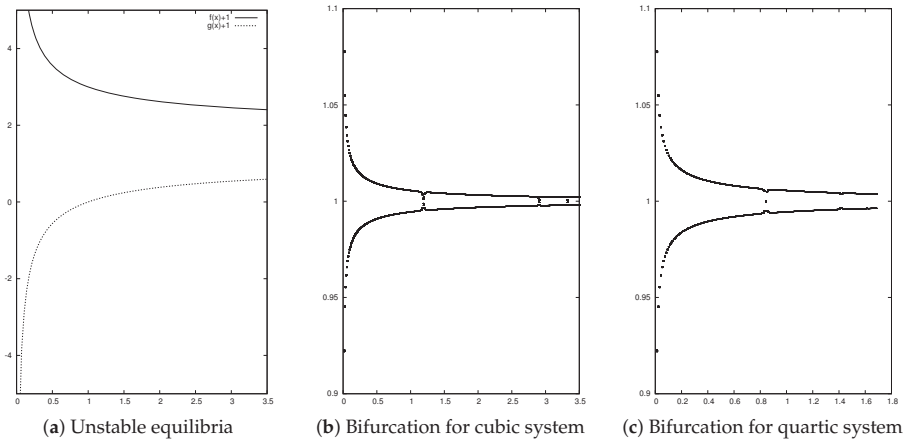
(**a**) Unstable equilibria (**b**) Bifurcation for cubic system (**c**) Bifurcation for quartic system

Figure 8. Bifurcation diagrams for $\beta = 0$ and initial condition $x_0 = 0.3$. Horizontal axis is α and vertical axis is x_n.

5. Conclusions

We have discovered and established that polynomial discrete systems are complicated to analyze and most of the time, even the equilibria are impossible to express explicitly. These systems may exhibit transient behavior that does not correspond to the asymptotic behavior, a characteristic that may lead to wrong results if one is using only numerical calculations due to the impossibility of carry out an analytical process.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Arenas, A.J.; González-Parra, G.; Chen-Charpentier, B.M. A nonstandard numerical scheme of predictor–corrector type for epidemic models. *Comput. Math. Appl.* **2010**, *59*, 3740–3749. [CrossRef]
2. Kojouharov, H.V.; Dimitrov, D.T. Compatible Discretizations for Continuous Dynamical Systems. *AIP Conf. Proc.* **2008**, *1067*, 28–37.
3. Shreiber, D.I.; Barocas, V.H.; Tranquillo, R.T. Temporal variations in cell migration and traction during fibroblast-mediated gel compaction. *Biophys. J.* **2003**, *84*, 4102–4114. [CrossRef]
4. Zaliapin, I.; Semenova, I.; Kashina, A.; Rodionov, V. Multiscale trend analysis of microtubule transport in melanophores. *Biophys. J.* **2005**, *88*, 4008–4016. [CrossRef] [PubMed]
5. Simson, R.; Sheets, E.D.; Jacobson, K. Detection of temporary lateral confinement of membrane proteins using single-particle tracking analysis. *Biophys. J.* **1995**, *69*, 989–993. [CrossRef]
6. Fujiwara, T.; Ritchie, K.; Murakoshi, H.; Jacobson, K.; Kusumi, A. Phospholipids undergo hop diffusion in compartmentalized cell membrane. *J. Cell Biol.* **2002**, *157*, 1071–1081. [CrossRef]
7. Fiutak, J.; Mizerski, J.Z. Transient behaviour of laser. *Phys. B* **1980**, *39*, 347–352. [CrossRef]
8. Tang, C.; Telle, J.; Ghizoni, C. Transient effects in wavelength-modulated dye lasers. *Appl. Phys. Lett.* **1975**, *26*, 534–537. [CrossRef]
9. Krapivsky, P.L.; Redner, S.; Ben-Naim, E. *A Kinetic View of Statistical Physics*; Cambridge University Press: Cambridge, UK, 2010.
10. Castellano, C.; Fortunato, S.; Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* **2009**, *81*, 591. [CrossRef]
11. Chowdhury, D.; Santen, L.; Schadschneider, A. Statistical Physics of Vehicular Traffic and Some Related Systems. *Phys. Rep.* **2000**, *329*, 199–329. [CrossRef]

12. Hinton, E. The dynamic transient analysis of axisymmetric circular plates by finite element method. *J. Sound Vib.* **1976**, *46*, 465–472. [CrossRef]

13. Kunow-Baumhauer, A. The response of a beam subjected to a transient pressure wave load. *J. Sound Vib.* **1984**, *92*, 491–506. [CrossRef]

14. Hastings, A. Transients: The key to long-term ecological understanding? *Trends. Ecol. Evol.* **2004**, *19*, 39–45. [CrossRef] [PubMed]

15. Van Geest, G.; Coops, H.; Scheffer, M.; van Nes, E. Long Transients Near the Ghost of a Stable State in Eutrophic Shallow Lakes with Fluctuating Water Levels. *Ecosystems* **2007**, *10*, 37–47. [CrossRef]

16. Schaffer, W.M.; Kendall, B.; Tidd, C.W.; Olsen, L.F. Transient periodicity and episodic predictability in biological dynamics. *IMA J. Math. Appl. Med. Biol.* **1993**, *10*, 227–247. [CrossRef]

17. Fisher, R.S.; Boas, W.E.; Blume, W.; Elger, C.; Genton, P.; Lee, P.; Engel, J. Epileptic Seizures and Epilepsy: Definitions Proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). *Epilepsia* **2005**, *46*, 470–472. [CrossRef]

18. Grüne, L.; Stieler, M. Asymptotic stability and transient optimality of economic MPC without terminal conditions. *J. Process Control* **2014**, *24*, 1187–1196. [CrossRef]

19. Solis, F.; Jódar, L. Nonisolated slow convergence in discrete dynamical systems. *Appl. Math. Lett.* **2004**, *17*, 597–599. [CrossRef]

20. Solis, F.; Felipe, R. Slow convergence of Maps. *Nonlinear Stud.* **2001**, *8*, 389–392.

21. Weidner, P. The Durand-Kerner method for trigonometric and exponential polynomials. *Computing* **1988**, *40*, 175–179. [CrossRef]

22. Schock, E. Approximate solution of ill-posed equations: Arbitrarily slow convergence vs. superconvergence. In *Constructive Methods for the Practical Treatment of Integral Equations*; Birkhäuser: Basel, Switzerland, 1985; pp. 234–243.

23. Braun, M. *Differential Equations and Their Applications: An Introduction to Applied Mathematics*; Springer: New York, NY, USA, 2013; Volume 15.

24. Dunham, C.B. Chebyshev approximation by exponential-polynomial sums. *J. Comput. Appl. Math.* **1979**, *5*, 53–57. [CrossRef]

25. Viswanathan, K.S. Statistical mechanics of a one-dimensional lattice gas with exponential-polynomial interactions. *Comm. Math. Phys.* **1976**, *47*, 131–141. [CrossRef]

26. Boyadzhiev, K.N. Exponential Polynomials, Stirling Numbers, and Evaluation of Some Gamma Integrals. *Abstr. Appl. Anal.* **2009**, *2009*, 168672. [CrossRef]

27. Mayoral, E.; Robledo, A. Multifractality and nonextensivity at the edge of chaos of unimodal maps. *Phys. A* **2004**, *340*, 219–226. [CrossRef]

28. Andrecut, M.; Ali, M.K. Robust chaos in smooth unimodal maps. *Phys. Rev. E* **2001**, *64*, 025203–025205. [CrossRef] [PubMed]

29. Jensen, R.V.; Ma, L.K.H. Nonuniversal behavior of asymmetric unimodal maps. *Phys. Rev. A* **1985**, *31*, 3993–3995. [CrossRef]

30. Hao, B.L. Universal Slowing-Down Exponent Near Periodic-Doubling Bifurcation Points. *Phys. Lett. A* **1981**, *86*, 267–268. [CrossRef]

31. San Martín, J.; Porter, M.A. Convergence Time towards Periodic Orbits in Discrete Dynamical Systems. *PLoS ONE* **2014**, *9*, e92652. [CrossRef] [PubMed]

32. Solis, F.; Chen, B.; Kojouharov, H. A classification of slow convergence near parametric periodic points of discrete dynamical systems. *Int. J. Comput. Math.* **2016**, *93*, 1011–1021. [CrossRef]

Block Preconditioning Matrices for the Newton Method to Compute the Dominant λ-Modes Associated with the Neutron Diffusion Equation

Amanda Carreño [1,*], Luca Bergamaschi [2], Angeles Martinez [3], Antoni Vidal-Ferrándiz [1], Damian Ginestar [4] and Gumersindo Verdú [1]

[1] Instituto Universitario de Seguridad Industrial, Radiofísica y Medioambiental, Universitat Politècnica de València, Camino de Vera, s/n, 46022 Valencia, Spain; anvifer2@upvnet.upv.es (A.V.-F.); gverdu@iqn.upv.es (G.V.)

[2] Department of Civil Environmental and Architectural Engineering, University of Padua, Via 8 Febbraio, 2, 35122 Padua, Italy; luca.bergamaschi@unipd.it

[3] Department of Mathematics "Tullio Levi-Civita", University of Padua, Via 8 Febbraio, 2, 35122 Padua, Italy; acalomar@math.unipd.it

[4] Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València, Camino de Vera, s/n, 46022 Valencia, Spain; dginesta@mat.upv.es

* Correspondence: amcarsan@iqn.upv.es

Abstract: In nuclear engineering, the λ-modes associated with the neutron diffusion equation are applied to study the criticality of reactors and to develop modal methods for the transient analysis. The differential eigenvalue problem that needs to be solved is discretized using a finite element method, obtaining a generalized algebraic eigenvalue problem whose associated matrices are large and sparse. Then, efficient methods are needed to solve this problem. In this work, we used a block generalized Newton method implemented with a matrix-free technique that does not store all matrices explicitly. This technique reduces mainly the computational memory and, in some cases, when the assembly of the matrices is an expensive task, the computational time. The main problem is that the block Newton method requires solving linear systems, which need to be preconditioned. The construction of preconditioners such as ILU or ICC based on a fully-assembled matrix is not efficient in terms of the memory with the matrix-free implementation. As an alternative, several block preconditioners are studied that only save a few block matrices in comparison with the full problem. To test the performance of these methodologies, different reactor problems are studied.

1. Introduction

The neutron transport equation is a balance equation that describes the behavior of the neutrons inside the reactor core. This equation for three-dimensional problems is an equation defined in a phase space of dimension seven, and this makes the problem very difficult to solve. Thus, some approximations are considered such as the multigroup neutron diffusion equation by relying on the assumption that the neutron current is proportional to the gradient of the neutron flux by means of a diffusion coefficient.

Given a configuration of a nuclear reactor core, its criticality can be forced by dividing the production operator in the neutron diffusion equation by a positive number, λ, obtaining a neutron

balance equation: the λ-modes problem. For the two energy groups approximation and without considering up-scattering, this equation can be written as [1]:

$$
\begin{pmatrix} -\vec{\nabla}(D_1\vec{\nabla}) + \Sigma_{a_1} + \Sigma_{12} & 0 \\ -\Sigma_{12} & -\vec{\nabla}(D_2\vec{\nabla}) + \Sigma_{a_2} \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \frac{1}{\lambda} \begin{pmatrix} \nu\Sigma_{f1} & \nu\Sigma_{f2} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix},
\tag{1}
$$

where ϕ_1 and ϕ_2 denote the fast and thermal flux, respectively. The macroscopic cross-sections D_g, Σ_{ag}, and $\nu\Sigma_{fg}$, with $g = 1, 2$, and $\Sigma_{1,2}$, are values that depend on the position.

The largest eigenvalue in magnitude, called the k-effective (or multiplication factor), indicates a measure of the criticality of the reactor, and its corresponding eigenfunction describes the steady-state neutron distribution in the reactor core. Next, dominant eigenvalues and their corresponding eigenfunctions are useful to develop modal methods for the transient analysis.

To make a spatial discretization of Problem (1), a high order continuous Galerkin finite element method is used, leading to a generalized algebraic eigenvalue problem of the form:

$$
\begin{pmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} \tilde{\phi}_1 \\ \tilde{\phi}_2 \end{pmatrix} = \frac{1}{\lambda} \begin{pmatrix} M_{11} & M_{12} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \tilde{\phi}_1 \\ \tilde{\phi}_2 \end{pmatrix},
\tag{2}
$$

where the matrix elements are given by:

$$
\begin{aligned}
L_{ij} = \sum_{e=1}^{N_t} &\left(D_1 \int_{\Omega_e} \vec{\nabla}N_{1i} \cdot \vec{\nabla}N_{1j}\, dV - D_1 \int_{\Gamma_e} N_{1i}\vec{\nabla}N_{1j}\, d\vec{S} + D_2 \int_{\Omega_e} \vec{\nabla}N_{2i} \cdot \vec{\nabla}N_{2j}\, dV \right. \\
&- D_2 \int_{\Gamma_e} N_{2i}\vec{\nabla}N_{2j}\, dV + (\Sigma_{a1} + \Sigma_{12}) \int_{\Omega_e} N_{1i}N_{1j}\, dV + \Sigma_{a2} \int_{\Omega_e} N_{2i}N_{2j}\, dV \\
&\left. - \Sigma_{12} \int_{\Omega_e} N_{2i}N_{1j}\, dV \right), \\
M_{ij} = \sum_{e=1}^{N_t} &\left(\nu\Sigma_{f1} \int_{\Omega_e} N_{1i}N_{1j}\, dV + \nu\Sigma_{f2} \int_{\Omega_e} N_{1i}N_{2j}\, dV \right),
\end{aligned}
$$

where N_i is the prescribed shape function for the ith node. The vector $\tilde{\phi} = (\tilde{\phi}_1, \tilde{\phi}_2)^{\mathsf{T}}$ is the algebraic vector of the finite weights corresponding to the neutron flux in terms of the shape functions. The shape functions used in this work are Lagrange polynomials. The subdomains Ω_e ($e = 1, \ldots, N_t$) denote the cells in which the reactor domain is divided and where the cross-sections are assumed to be constant. Similarly, Γ_e is the corresponding subdomain surface, which is part of the reactor boundary. More details on the finite element discretization can be found in [2]. For the implementation of the finite element method, the open source finite elements library Deal.II [3] has been used.

In this work, a matrix-free strategy for the blocks of the matrix M and for the non-diagonal blocks of L is developed. In this way, matrix-vector products are computed on the fly in a cell-based interface. For instance, we can consider that a finite element Galerkin approximation that leads to the matrix $M_{1,1}$ takes a vector u as input and computes the integrals of the operator multiplied by trial functions, and the output vector is v. The operation can be expressed as a sum of N_t cell-based operations,

$$
v = M_{1,1}u = \sum_{e=1}^{N_t} P_e^T M_{1,1}^e P_e u,
\tag{3}
$$

where P_e denotes the matrix that defines the location of cell-related degrees of freedom in the global vector and $M_{1,1}^e$ denotes the submatrix of $M_{1,1}$ on finite element e. This sum is optimized through sum-factorization. Details about the implementation are explained in [4]. This strategy greatly reduces the memory used by the matrix elements.

Calculation of the dominant lambda mode has traditionally utilized the classical power iteration method, which although robust, converges slowly for dominance ratios near one, as occurs in some practical problems. Thus, acceleration techniques are needed to improve the convergence of the power iteration method. Some approaches in diffusion theory are, for instance, Chebyshev iteration [5] and Wielandt shift [6]. Alternative approaches to the power iteration method have been studied in an attempt to improve upon the performance of accelerated power iteration methods [7,8]. The subspace iteration method [9], the Implicit Restarted Arnoldi method (IRAM) [10], the Jacobi–Davidson [11], and the Krylov–Schur method [2] implemented in the SLEPclibrary [12] have been used to compute the largest or several dominant eigenvalues for the neutron diffusion equation and their corresponding eigenfunctions. More recently, other Krylov methods have been used to compute these modes for other approximations of the neutron transport equation [7,13]. Usually, applying these kinds of methods requires either transforming the generalized problem (2) into an ordinary eigenvalue problem or applying a shift and invert technique. In both cases, in the solution process, it is necessary to solve numerous linear systems. These systems are not well-conditioned, and they need to be preconditioned. Thus, the time and computational memory needed to compute several eigenvalues become very high.

One alternative is to use a method that does not require solving any linear system, such as the generalized Davidson, used for neutron transport calculations in [14]. Other methods are the block Newton methods that have been shown to be very efficient in the computation of several eigenvalues in neutron diffusion theory. These methods either do not need to solve as many linear systems as the Krylov methods or avoid solving any linear system with some hybridization. One of these Newton methods is the modified block Newton method, which has been considered for the ordinary eigenvalue problem associated with Problem (2) [15] or directly for the generalized eigenproblem (2) [16]. One advantage of these block methods is that several eigenvectors can be approximated simultaneously, and as a consequence, the convergence behavior improves. The convergence of the eigensolvers usually depends on the eigenvalue separation, and if there are clustered or multiple eigenvalues, the methods may have problems finding all the eigenvalues. In practical situations of reactor analysis, the dominance ratio corresponding to the dominant eigenvalues is often near unity, resulting in a slow convergence. In the block methods, this convergence only depends on the separation of the group of target eigenvalues from the rest of the spectrum. Another advantage is that these methods do not require solving as many linear systems as the previous methods. However, these linear systems still need to be preconditioned. Another of this kind of Newton method is the Jacobian-free Newton–Krylov methods that have been studied with traditional methods such as the power iteration used as the preconditioner [17,18] or with a more sophisticated Schwarz preconditioner [19]. In this work, we use the Modified Generalized Block Newton Method (MGBNM) presented in [16], and we propose several ways to precondition the linear systems that need to be solved in this method in an efficient way.

The structure of the rest of the paper is as follows. In Section 2, the modified generalized block Newton method is described. In Section 3, the different preconditioners for the MGBNM are presented. The performance of the preconditioners is presented in Section 4 for two different benchmark problems. Finally, Section 5 synthesizes the main conclusions of this work.

2. The Modified Generalized Block Newton Method

This method was presented by Lösche in 1998 [20] for ordinary eigenvalue problems, and an extension to generalized eigenvalue problems was studied in [16]. Given the partial generalized eigenvalue problem (2) written as:

$$MX = LX\Lambda, \tag{4}$$

where $X \in \mathbb{R}^{n \times q}$ is a matrix with q eigenvectors and $\Lambda \in \mathbb{R}^{q \times q}$ is a diagonal matrix with the q eigenvalues associated, we suppose that the eigenvectors can be factorized as $X = ZS$, where Z is

an orthogonal matrix. Moreover, the biorthogonality condition $W^\mathrm{T}Z = I$ is introduced, where W is a fixed matrix. Thus, if we denote $K = S\Lambda S^{-1}$, the problem (4) can be rewritten as:

$$MX = LX\Lambda \Leftrightarrow MZ = LZS\Lambda S^{-1} \Leftrightarrow MZ = LZK.$$

We construct this projection to ensure that the method converges to independent eigenvectors.

Then, the solution of Problem (4) is obtained by solving the non-linear problem:

$$F(Z, \Lambda) := \begin{pmatrix} MZ - LZK \\ W^\mathrm{T}Z - I_q \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{5}$$

By applying Newton's method, a new iterated solution arises as:

$$Z^{(k+1)} = Z^{(k)} - \Delta Z^{(k)}, \qquad K^{(k+1)} = K^{(k)} - \Delta K^{(k)}, \tag{6}$$

where $\Delta Z^{(k)}$ and $\Delta K^{(k)}$ are solutions of the system obtained when the equations (6) are substituted into the equations (5), and these are truncated at the first order terms.

The matrix $K^{(k)}$ is not necessarily a diagonal matrix, and as a result, the system is coupled. To avoid this problem, the modified generalized block Newton method (MGBNM) needs to apply the previous two steps. Firstly, the modified Gram–Schmidt process is used to orthonormalize the matrix $Z^{(k)}$. Then, the Rayleigh–Ritz projection method for the generalized eigenvalue problem [21] is applied. Thus, $\Delta Z^{(k)} = (\Delta z_1^{(k)}, \dots, \Delta z_q^{(k)})$, where $\Delta z_i^{(k)} \in \mathbb{R}^n$ and $\Delta K^{(k)} = (\Delta k_1^{(k)}, \dots, \Delta k_q^{(k)})$ where $\Delta k_i^{(k)} \in \mathbb{R}$ are obtained from the solutions of the linear systems:

$$\begin{pmatrix} M - \lambda_i^{(k)} L & LZ^{(k)} \\ Z^{(k)^\mathrm{T}} & 0 \end{pmatrix} \begin{pmatrix} \Delta z_i^{(k)} \\ -\Delta k_i^{(k)} \end{pmatrix} = \begin{pmatrix} Mz_i^{(k)} - Lz_i^{(k)}\lambda_i^{(k)} \\ 0 \end{pmatrix}, \quad i = 1, \dots, q.$$

The solution of these systems is computed by using the Generalized Minimal Residual method (GMRES) computing the matrix vector products with block matrix multiplications. However, these systems need to be preconditioned (in each iteration and for each eigenvalue) to reduce the condition number of the matrix.

3. Preconditioning

The first choice for a preconditioner is assembling the matrix:

$$A = \begin{pmatrix} M - \lambda_i^{(k)} L & LZ^{(k)} \\ Z^{(k)^\mathrm{T}} & 0 \end{pmatrix},$$

and constructing the full preconditioner associated with the matrix. We use the ILUT(0) preconditioner since A is a non-symmetric matrix. There are no significant differences if the preconditioner obtained for the matrix associated with the first eigenvalue is used for all eigenvalues in the same iteration because in the matrix, A only changes the value of $\lambda_i^{(k)}$, and usually, the eigenvalues in reactor problems are clustered. This preconditioner is denoted by P.

To devise an alternative preconditioner without the necessity of assembling the matrix A, we write the explicit inverse of A, by using its block structure,

$$A^{-1} = \begin{pmatrix} J^{-1}(I - C_1(C_2^\mathrm{T}C_1)^{-1}C_2^\mathrm{T}) & J^{-1}C_1(C_2^\mathrm{T}C_1) \\ (C_2^\mathrm{T}C_1)^{-1}C_2^\mathrm{T} & -(C_2^\mathrm{T}C_1)^{-1} \end{pmatrix},$$

where:

$$J = M - \lambda_i L, \quad C_1 = LZ, \quad C_2^\mathsf{T} = Z^\mathsf{T} J^{-1}.$$

We desire a preconditioner for A by suitably approximating A^{-1}. Let us call P_J a preconditioner for J. For instance, $P_J = (LU)^{-1}$, where L, U are the incomplete L and U factors of J. Thus, we can define, after setting $C_2^\mathsf{T} = Z^\mathsf{T} P_J$, the preconditioner of A as:

$$\hat{P}_J = \begin{pmatrix} P_J(I - C_1(C_2^\mathsf{T} C_1)^{-1} C_2^\mathsf{T}) & P_J C_1 (C_2^\mathsf{T} C_1) \\ (C_2^\mathsf{T} C_1)^{-1} C_2^\mathsf{T} & -(C_2^\mathsf{T} C_1)^{-1} \end{pmatrix}.$$

The previous preconditioner does not need to assemble the entire matrix A, but it needs to assemble the matrix J to build its ILUpreconditioner. Therefore, the next alternative that we propose is using a preconditioner of $-L$ instead of $J = M - \lambda_1 L$. This preconditioner works well because in the discretization process, the L matrix comes from the discretization of the differential matrix that has the gradient operators and the diffusion terms. In addition, in nuclear calculations, λ_1 is near 1.0. Thus, we can build a preconditioner of $-L$ instead of the matrix J. We denote by $\hat{P}_L$ the preconditioner $\hat{P}_J$ where the preconditioner of $-L$ is used to precondition the block J.

Finally, the last alternative is avoiding assembling the matrix L taking advantage of its block structure. For that purpose, we carry out a similar process as the one used for matrix A. We write the explicit form of the inverse of L as:

$$L^{-1} = \begin{pmatrix} L_{11}^{-1} & 0 \\ -L_{22}^{-1} L_{21} L_{11}^{-1} & L_{22}^{-1} \end{pmatrix}, \tag{7}$$

and substitute the inverses of the blocks by preconditioners. Thus, the preconditioner of L has the following structure:

$$Q_L = \begin{pmatrix} P_{11} & 0 \\ -P_{22} L_{21} P_{11} & P_{22} \end{pmatrix},$$

where P_{11}, P_{22} denote a preconditioner of L_{11} and L_{22}, respectively. The block matrices L_{11}, L_{22} are symmetric and positive definite. Then, we can use as preconditioner the Incomplete Cholesky decomposition (IC(0)). However, the main advantage of this preconditioner is that it permits using a matrix-free implementation that does not require allocating all matrices. We only need to assemble the blocks L_{11} and L_{22} to construct the associated IC(0) preconditioners. The application of $\hat{P}_J$ with $-Q_L$ to precondition J is called $\hat{P}_Q$.

4. Numerical Results

In this section, the performance of the proposed preconditioners has been tested on two different problems: a version of the 3D NEACRPreactor [22] and a configuration of the Ringhals reactor [23]. The neutron diffusion equation in both problems has been discretized using the finite element method presented in Section 1 using Lagrange polynomials of degree three because it is shown in previous works that this degree is necessary to obtain accurate results in similar reactor problems [2]. The number of eigenvalues computed was four for each reactor.

The incomplete lower-upper preconditioner with Level 0 of fill (ILU) has been provided by the PETScpackage [24].

As the modified generalized block Newton method needs an initial approximation of a set of eigenvectors, a multilevel initialization with two meshes was used to obtain this approximation (for more details, see [25]).

The stopping criteria for all solvers has been set equal to 10^{-6} in the global residual error,

$$\varepsilon_{\text{res}} = \max_{i=1,\dots,q} \| M x_i - \lambda_i L x_i \|^2 \, ,$$

where λ_i is the ith eigenvalue and x_i its associated eigenvector such that $\|x_i\| = 1$.

The modified block Newton method has been implemented using a dynamic tolerance in the residual error of the solution in the linear systems. The tolerance values have been set to $\{10^{-2}, 10^{-3}, 10^{-5}, 10^{-8}, 10^{-8}, \dots\}$.

The methods have been implemented in C++ based on the data structures provided by the library Deal.ii [3] and PETSc [24]. The computer used for the computations was an Intel® Core™i7-4790 @3.60 GHz with 32 Gb of RAM running on Ubuntu GNU/Linux 16.04 LTS.

4.1. NEACRP Reactor

The NEACRP benchmark in a near-critical configuration [22] is chosen to compare the proposed methodology. The reactor core has a radial dimension of 21.606 cm × 21.606 cm per assembly. Axially, the reactor, with a total height of 427.3 cm, is divided into 18 layers with height (from bottom to top): 30.0 cm, 7.7 cm, 11.0 cm, 15.0 cm, 30.0 cm (10 layers), 12.8 cm (two layers), 8.0 cm, and 30.0 cm. Figure 1 shows the reactor geometry and the distribution of the different materials. The cross-sections of materials are displayed in Table 1. The total number of cells of the reactor domain is 3978. Zero flux boundary conditions were set. The spatial discretization of the neutron diffusion equation, by using polynomials of degree three, gave a number of 230,120 degrees of freedom. The mesh built to obtain an initial guess had 1308 cells, and the computational time needed to obtain this approximation was 24 s. The four dominant eigenvalues computed are collected in Table 2. This table shows that the spectrum associated with the problem is clustered with two degenerated eigenvalues. A representation of the fast flux distribution for each mode is displayed in Figure 2.
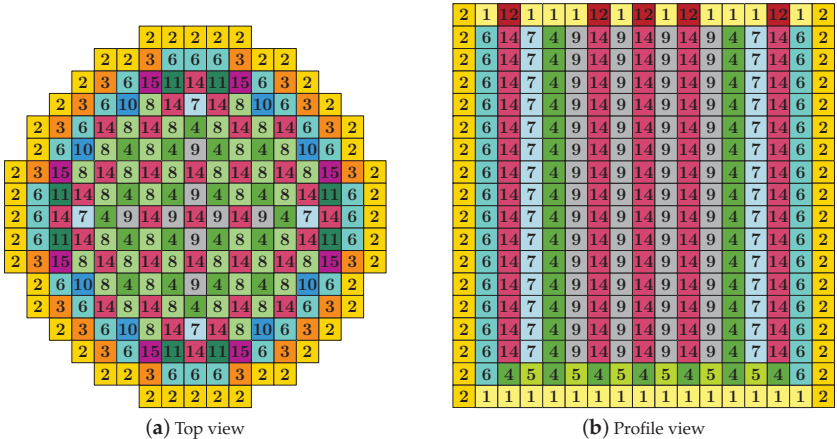

(a) Top view   **(b)** Profile view

Figure 1. Geometry and distribution of the materials of the NEACRPreactor.

Table 1. Macroscopic cross-section of the NEACRP reactor.

Mat.	D_1 (cm)	D_2 (cm)	Σ_{a1} (cm^{-1})	Σ_{a2} (cm^{-1})	Σ_{12} (cm^{-1})	$\nu\Sigma_{f1}$ (cm^{-1})	$\nu\Sigma_{f2}$ (cm^{-1})
1	5.9264	8.2289×10^{-1}	2.5979×10^{-4}	1.7085×10^{-1}	2.7988×10^{-2}	0.0000	0.0000
2	1.1276	1.7053×10^{-1}	1.1878×10^{-3}	1.9770×10^{-1}	2.3161×10^{-2}	0.0000	0.0000
3	1.1276	1.7053×10^{-1}	1.1878×10^{-3}	1.9770×10^{-1}	2.0081×10^{-2}	0.0000	0.0000
4	1.4624	3.9052×10^{-1}	8.4767×10^{-3}	6.2569×10^{-2}	1.9686×10^{-2}	5.0150×10^{-3}	8.7712×10^{-2}
5	1.4637	3.9485×10^{-1}	8.8225×10^{-3}	6.9978×10^{-2}	1.9436×10^{-2}	5.6085×10^{-3}	1.0424×10^{-1}
6	1.4650	3.9851×10^{-1}	9.1484×10^{-3}	7.6850×10^{-2}	1.9196×10^{-2}	6.1819×10^{-3}	1.1954×10^{-1}
7	1.4641	4.0579×10^{-1}	9.0869×10^{-3}	7.7687×10^{-2}	1.8526×10^{-2}	5.5830×10^{-3}	1.0289×10^{-1}
8	1.4642	4.0946×10^{-1}	9.1738×10^{-3}	8.0302×10^{-2}	1.8223×10^{-2}	5.5741×10^{-3}	1.0232×10^{-1}
9	1.4642	4.1314×10^{-1}	9.2596×10^{-3}	8.2924×10^{-2}	1.7920×10^{-2}	5.5650×10^{-3}	1.0169×10^{-1}
10	1.4653	4.0919×10^{-1}	9.4097×10^{-3}	8.4462×10^{-2}	1.8288×10^{-2}	6.1564×10^{-3}	1.1807×10^{-1}
11	1.4655	4.1277×10^{-1}	9.4956×10^{-3}	8.7030×10^{-2}	1.7986×10^{-2}	6.1474×10^{-3}	1.1744×10^{-1}
12	5.5576	8.7013×10^{-1}	2.7375×10^{-3}	1.9644×10^{-1}	2.4796×10^{-2}	0.0000	0.0000
13	5.6027	8.6371×10^{-1}	2.4169×10^{-3}	1.9313×10^{-1}	2.5209×10^{-2}	0.0000	0.0000
14	1.4389	4.0085×10^{-1}	1.0954×10^{-2}	8.8157×10^{-2}	1.6493×10^{-2}	4.9122×10^{-3}	8.4889×10^{-2}
15	1.4413	4.0665×10^{-1}	1.1578×10^{-2}	1.0250×10^{-1}	1.6054×10^{-2}	6.0593×10^{-3}	1.1626×10^{-1}

Table 2. Eigenvalues for the NEACRP reactor.

Eigenvalue	Value
1	1.002
2	0.98862
3	0.985406
4	0.985406

(**a**) 1st mode

(**b**) 2nd mode

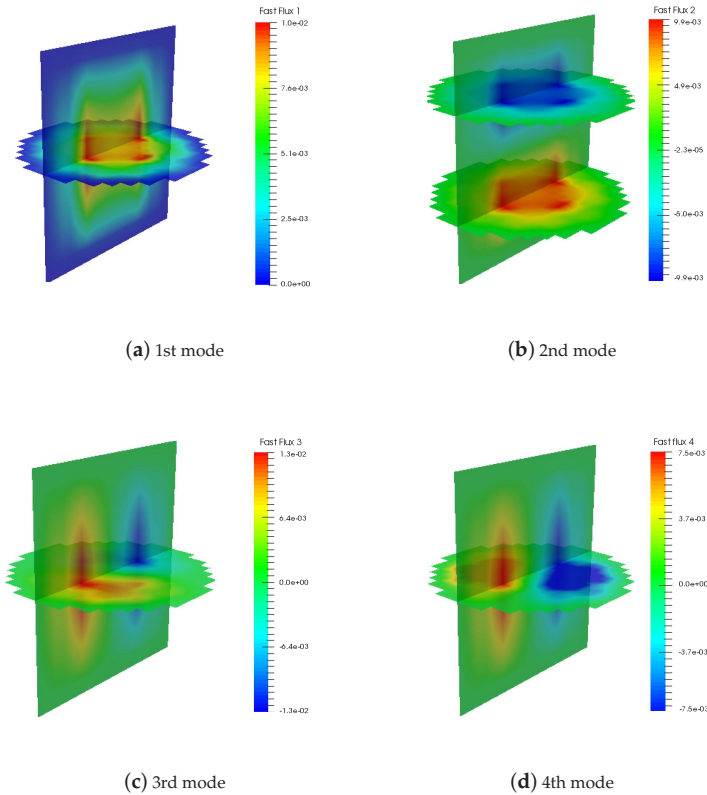(**c**) 3rd mode

(**d**) 4th mode

Figure 2. Fast fluxes' distribution of the NEACRP reactor corresponding to the first four modes.

First, we show the convergence history of the MGBNM to obtain the solution of the eigenvalue problem. Figure 3 shows the number of iterations against the residual error for the NEACRP reactor. It is observed that the MGBNM only needed four iterations to reach a residual error equal to 10^{-6}.
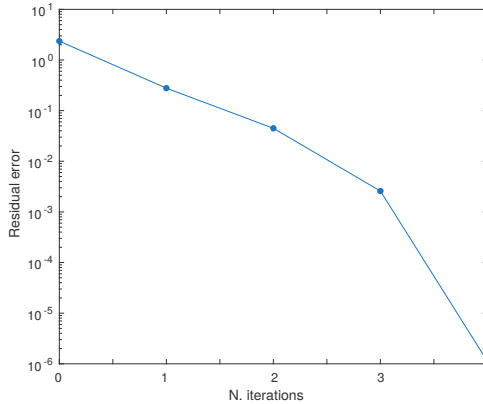

Figure 3. Convergence history of the Modified Generalized Block Newton Method (MGBNM) for the NEACRP reactor.

Table 3 collects the average number of iterations obtained by directly applying the ILU preconditioner of A and the total time that the GMRES method needs to reach the residual error in the linear systems given in $\text{Tol}(\|b - Ax\|)$. The time spent to assemble the matrices and to build the preconditioner (setup time (s)) is also displayed. These data are presented for each iteration and in a total sum. This table shows that the number of iterations is not very high, but the time spent to assemble the matrix and to construct the preconditioner increases the total CPU time considerably. It is necessary to build in each iteration a new preconditioner for A because the columns related to the block Z change considerably in each updating.

Table 3. Data for the preconditioner P for the NEACRP reactor. GMRES, Generalized Minimal Residual.

No. It. MGBNM	Tol ($\|b - Ax\|$)	Mean Its. GMRES	Setup Time (s)	Total Time (s)
1	1×10^{-2}	4.5	12.0	18.0
2	1×10^{-3}	9.75	12.0	20.4
3	1×10^{-5}	20.75	12.0	25.2
4	1×10^{-8}	37.5	12.0	33.2
Total		72.5	48.0	96.8

Table 4 displays these data related to the block preconditioner proposed $\hat{P}_J$ that uses the ILU preconditioner for approximating the inverse of $M - \lambda_1 L$. It is observed that we only needed to assemble the matrix $M - \lambda_1 L$ once in the first iteration to build the preconditioner. This is because we only needed a preconditioner of $M - \lambda_1 L$, and the value of λ_1 was very similar for all iterations. The mean of the number of iterations of the GMRES preconditioned with $\hat{P}_J$ was larger than in the previous case, but the total CPU time of using this block preconditioner was reduced by 26 s with respect to the full preconditioner.

Table 4. Data for the preconditioner $\hat{P}_J$ with ILU for the NEACRP reactor.

No. It. MGBNM	Tol ($\|b - Ax\|$)	Mean Its GMRES	Setup Time (s)	Total Time (s)
1	1×10^{-2}	8.25	6.6	12.9
2	1×10^{-3}	13.25	-	9.5
3	1×10^{-5}	23.25	-	16.6
4	1×10^{-8}	41.25	-	30.0
Total		86.0	6.6	70.0

Table 5 shows the data related to the block preconditioner $\hat{P}_J$, but in this case, we have used the Geometric Multigrid (GMG) preconditioner to approximate the inverse of $M - \lambda_1 L$. The results show, in comparison with the results of Table 4, that in spite of the total number of iterations and the setup time being much lower for the GMG, the total computational time is much higher. This is due to the application of the GMG preconditioner being more expensive than the application of the ILU preconditioner.

Table 5. Data for the preconditioner $\hat{P}_J$ with the Geometric Multigrid (GMG) for the NEACRP reactor.

No. It. MGBNM	Tol ($\|b - Ax\|$)	Mean Its GMRES	Setup Time (s)	Total Time (s)
1	1×10^{-2}	6.00	2.5	19.8
2	1×10^{-3}	9.75	-	30.8
3	1×10^{-5}	12.75	-	39.4
4	1×10^{-8}	20.50	-	61.3
Total		49.00	2.5	151.3

The next results were obtained by using the block preconditioner, $\hat{P}$, but in these cases, approximating the $(M - \lambda_1 L)^{-1}$ by the ILU preconditioner of $-L$ ($\hat{P}_L$) and by a block preconditioner of $-L$ ($Q_{\hat{L}}$). The most relevant data to compare the preconditioners considered in this work are exposed in Table 6. They were the total iterations of the GMRES, the total setup time, the total time to compute the solution, and the maximum computational memory spent by the matrices. We observe that the number of iterations increased when worse approximations of the inverse of A were considered, but the setup time that each preconditioner needs became smaller. Moreover, the maximum CPU memory was also reduced significantly. In the total CPU times, we observed that the block preconditioner ($\hat{P}$), in all of its versions, improved the times obtained by applying the ILU preconditioner of A directly. Between the possibilities for obtaining a preconditioner of $M - \lambda_1 L$, there were no big differences in the computational times, but there was an important savings of the computational memory. The best results were obtained by $\hat{P}_{\hat{L}}$ if the computational memory consumption was taken into account.

Table 6. Data obtained by using different preconditioners for the NEACRP reactor.

Prec.	Its GMRES	Time Setup	Total Time	Max. CPU mem.
P^{ILU}	72.5	48.0 s	96.8 s	2062 Mb
$\hat{P}_J^{\text{ILU}}$	86.0	6.6 s	70.0 s	1418 Mb
$\hat{P}_L^{\text{ILU}}$	98.0	4.4 s	73.2 s	787 Mb
$\hat{P}_Q^{\text{ILU}}$	100.25	1.8 s	74.4 s	787 Mb

Table 7 shows the timings and the memory spent in the matrix allocation by using the matrix-free technique or without using this strategy. The results show that not only the matrix memory consumption and the time to assemble were reduced, but also the time spent to compute the matrix-vector products. That implies that the matrix-free strategy reduced the total CPU time by about 30%.

Table 7. Data obtained using different matrix implementations for the NEACRP reactor.

	Matrix Memory	Time Matvec Products	Time Assembly	Time Newton	Total Time
Sparse Matrix	787 Mb	27 s	7 s	51 s	74 s
Matrix Free	319 Mb	10 s	4 s	33 s	52 s

Finally, we compared the MGBNM with this methodology against other eigenvalue solvers commonly used in the neutron diffusion computations (Table 8). We show the results by using a different number of computed eigenvalues (No. eigs). In particular, we have chosen for this comparison the generalized Davidson preconditioned with the block Gauss–Seidel preconditioner and the Krylov–Schur method by previously reducing the generalized eigenvalue problem to an ordinary eigenvalue problem as in [2]. To use both methods, the library SLEPc has been used [12]. From the computational times, we can deduce that the MGBNM was twice as fast as the rest of solvers for the computation of one and two eigenvalues, and it was very competitive at computing four eigenvalues.

Table 8. Computational times for the MGBNM with $\hat{P}_Q$, the generalized Davidson method, and the Krylov–Schur method for the NEACRP reactor. eigs, eigenvalues.

No. Eigs	MGBNM	Generalized Davidson	Krylov–Schur
1	14 s	28 s	27 s
2	23 s	39 s	37 s
4	53 s	48 s	52 s

4.2. Ringhals Reactor

For a practical application of the preconditioners in a real reactor, we have chosen the configuration of the Ringhals rector. Particularly, we have chosen the C9 point of the BWRreactor Ringhals I stability benchmark, which corresponds to a point of operation that degenerated in an out-of-phase oscillation [23]. It is composed of 27 planes with 728 cells in each plane. A representation with more detail of its geometry can be observed in Figure 4. The spatial discretization using finite elements of degree three gave 1,106,180 degrees of freedom. The coarse mesh considered to obtain an initial guess for the MGBNM had 6709 cells and the problem associated with this mesh a size of 386,768 degrees of freedom. The computed dominant eigenvalues were 1.00191, 0.995034, 0.992827, and 0.991401. The corresponding fast fluxes are represented in Figure 5.
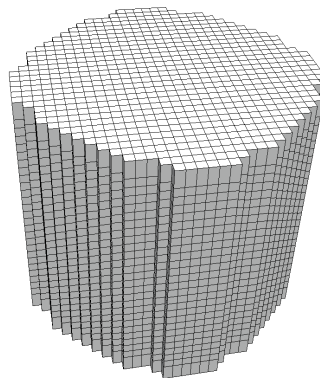

Figure 4. Geometry of the Ringhals reactor.

(**a**) 1st mode

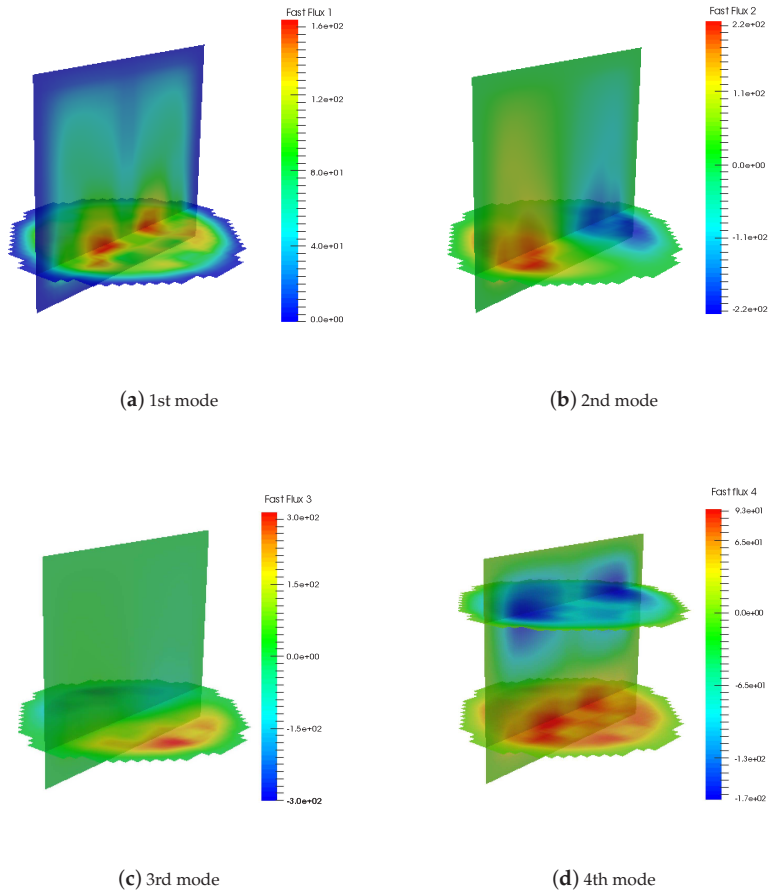(**b**) 2nd mode

(**c**) 3rd mode

(**d**) 4th mode

Figure 5. Fast fluxes' distribution of the Ringhals reactor corresponding to the first four modes.

The convergence history of the MGBNM associated with the Ringhals reactor is represented in Figure 6. For this reactor, the number of iterations needed to reach the tolerance (10^{-6}) was also equal to four.
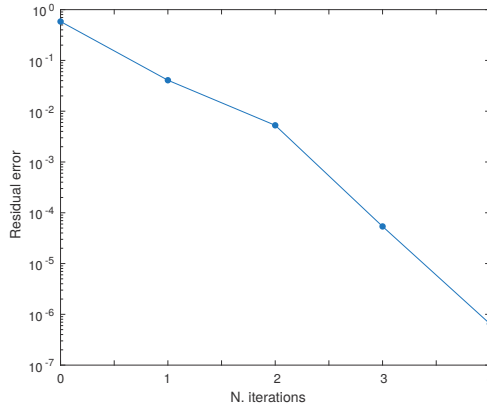
Figure 6. Convergence history of the MGBNM for the Ringhals reactor.

Table 9 collects the average number of iterations for the GMRES method for each iteration of MGBNM, the time to assemble the matrices and build the preconditioners, the total time of the MGBNM to reach the tolerance, and the maximum computational memory requested to assemble the matrices. From this table, similar conclusions as the ones obtained for the previous reactor are deduced. The number of iterations was not reduced, but the total CPU time and the maximum memory decreased considerably. For the Ringhals reactor, the most efficient option, in terms of computational memory, was also to apply the block preconditioner $\hat{P}_Q$. However, as the size of this reactor is much larger, the differences between the preconditioners for computational memory were much higher.

Table 9. Data obtained using different preconditioners for the Ringhals reactor.

Prec.	Its GMRES	Time Setup	Total Time	Max. CPU mem.
P	71.5	155 s	408 s	12.5 Gb
$\hat{P}_J$	81.0	39 s	331 s	9.3 Gb
$\hat{P}_L$	85.2	36 s	348 s	6.2 Gb
$\hat{P}_Q$	88.2	8 s	308 s	3.7 Gb

Finally, in Table 10, we compare the MGBNM with the generalized Davidson method and the Krylov–Schur method from the SLEPc library as in the previous reactor. The results show that the MGBNM was more efficient in terms of the computational time to compute one or a set of the lambda modes than the generalized Davidson and the Krylov–Schur methods.

Table 10. Computational times for the MGBNM with $\hat{P}_Q$, the generalized Davidson method, and the Krylov–Schur method for the Ringhals reactor.

No. Eigs	MGBNM	Generalized Davidson	Krylov–Schur
1	100 s	264 s	324 s
2	207 s	294 s	471 s
4	308 s	317 s	528 s

5. Conclusions

The modified generalized block Newton method (MGBNM) is an efficient eigenvalue solver that has been used to compute the dominant λ-modes associated with the neutron diffusion equation. This problem has been previously discretized by using a high order finite element method. This method

requires solving many linear systems that need to be previously preconditioned. Different block preconditioners have been studied as an alternative to assemble the full matrix and to construct a preconditioner in each iteration. The different preconditioners have been tested in a benchmark reactor problem (NEACRP) and in a realistic reactor problem (Ringhals). The preconditioners proposed in this work break down the setup cost at the price of a slight increase of the number of iterations. The result is a significant reduction of the total CPU time needed to reach convergence and the memory occupancy. Among the implementations studied, it is shown that the best option is the one that uses the block structure of the *L* matrix. Moreover, this implementation permits implementing the MGBNM with a matrix-free technique, thus greatly reducing the memory consumption. The differences increase when the size of the problem is larger. In comparison with other eigenvalue solvers, such as the generalized Davidson and the Krylov–Schur methods, the numerical results conclude that the MGBNM with this strategy of preconditioning is more efficient in some cases and very competitive in the rest. In future works, the MGBNM with these strategies for the preconditioning will be applied to other approximations of the neutron transport equations as the SP_N equations.

References

1. Stacey, W.M. *Nuclear Reactor Physics*; John Wiley & Sons: Hoboken, NJ, USA, 2018.
2. Vidal-Ferrandiz, A.; Fayez, R.; Ginestar, D.; Verdú, G. Solution of the Lambda modes problem of a nuclear power reactor using an h–p finite element method. *Ann. Nucl. Energy* **2014**, *72*, 338–349. [CrossRef]
3. Bangerth, W.; Hartmann, R.; Kanschat, G. deal.II—A general-purpose object-oriented finite element library. *ACM Trans. Math. Softw.* **2007**, *33*, 24. [CrossRef]
4. Kronbichler, M.; Kormann, K. A generic interface for parallel cell-based finite element operator application. *Comput. Fluids* **2012**, *63*, 135–147. [CrossRef]
5. Hageman, L.A.; Young, D.M. *Applied Iterative Methods*; Courier Corporation: North Chelmsford, MA, USA, 2012.
6. Sutton, T.M. Wielandt Iteration as Applied to the Nodal Expansion Method. *Nucl. Sci. Eng.* **1988**, *98*, 169–173. [CrossRef]
7. Warsa, J.S.; Wareing, T.A.; Morel, J.E.; McGhee, J.M.; Lehoucq, R.B. Krylov subspace iterations for deterministic k-eigenvalue calculations. *Nucl. Sci. Eng.* **2004**, *147*, 26–42. [CrossRef]
8. Allen, E.; Berry, R. The inverse power method for calculation of multiplication factors. *Ann. Nucl. Energy* **2002**, *29*, 929–935. [CrossRef]
9. Verdú, G.; Ginestar, D.; Vidal, V.; Muñoz-Cobo, J. 3D λ-modes of the neutron-diffusion equation. *Ann. Nucl. Energy* **1994**, *21*, 405–421. [CrossRef]
10. Verdú, G.; Miró, R.; Ginestar, D.; Vidal, V. The implicit restarted Arnoldi method, an efficient alternative to solve the neutron diffusion equation. *Ann. Nucl. Energy* **1999**, *26*, 579–593. [CrossRef]
11. Verdú, G.; Ginestar, D.; Miró, R.; Vidal, V. Using the Jacobi–Davidson method to obtain the dominant Lambda modes of a nuclear power reactor. *Ann. Nucl. Energy* **2005**, *32*, 1274–1296. [CrossRef]
12. Hernandez, V.; Roman, J.E.; Vidal, V. SLEPc: A scalable and flexible toolkit for the solution of eigenvalue problems. *ACM Trans. Math. Softw.* **2005**, *31*, 351–362. [CrossRef]
13. Evans, T.M.; Stafford, A.S.; Slaybaugh, R.N.; Clarno, K.T. Denovo: A new three-dimensional parallel discrete ordinates code in SCALE. *Nucl. Technol.* **2010**, *171*, 171–200. [CrossRef]
14. Hamilton, S.P.; Evans, T.M. Efficient solution of the simplified PN equations. *J. Comput. Phys.* **2015**, *284*, 155–170. [CrossRef]
15. González-Pintor, S.; Ginestar, D.; Verdú, G. Updating the Lambda Modes of a nuclear power reactor. *Math. Comput. Model.* **2011**, *54*, 1796–1801. [CrossRef]

16. Carreño, A.; Vidal-Ferràndiz, A.; Ginestar, D.; Verdú, G. Spatial modes for the neutron diffusion equation and their computation. *Ann. Nucl. Energy* **2017**, *110*, 1010–1022. [CrossRef]
17. Gill, D.; Azmy, Y. Newton's method for solving k-eigenvalue problems in neutron diffusion theory. *Nucl. Sci. Eng.* **2011**, *167*, 141–153. [CrossRef]
18. Knoll, D.; Park, H.; Newman, C. Acceleration of k-eigenvalue/criticality calculations using the Jacobian-free Newton-Krylov method. *Nucl. Sci. Eng.* **2011**, *167*, 133–140. [CrossRef]
19. Kong, F.; Wang, Y.; Schunert, S.; Peterson, J.W.; Permann, C.J.; Andrš, D.; Martineau, R.C. A fully coupled two-level Schwarz preconditioner based on smoothed aggregation for the transient multigroup neutron diffusion equations. *Numer. Linear Algebra Appl.* **2018**, *25*, e2162. [CrossRef]
20. Lösche, R.; Schwetlick, H.; Timmermann, G. A modified block Newton iteration for approximating an invariant subspace of a symmetric matrix. *Linear Algebra Appl.* **1998**, *275*, 381–400. [CrossRef]
21. Saad, Y. *Iterative Methods for Sparse Linear Systems*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2003.
22. Finnemann, H.; Galati, A. NEACRP 3-D LWR Core Transient Benchmark: Final Specifications. Available online: https://www.oecd-nea.org/science/docs/1991/neacrp-l-1991-335.pdf (accessed on 13 January 2019).
23. Lefvert, T. OECD/NEA Ringhals 1 Stability Benchmark. Available online: http://www.nea.fr/science/docs/1996/nsc-doc96-22.pdf (accessed on 13 January 2019).
24. Balay, S.; Abhyankar, S.; Adams, M.; Brune, P.; Buschelman, K.; Dalcin, L.; Gropp, W.; Smith, B.; Karpeyev, D.; Kaushik, D.; et al. *PETSc Users Manual Revision 3.7*; Technical Report for Argonne National Lab: Argonne, IL, USA, 2016.
25. Carreño, A.; Vidal-Ferràndiz, A.; Ginestar, D.; Verdú, G. Block hybrid multilevel method to compute the dominant λ-modes of the neutron diffusion equation. *Ann. Nucl. Energy* **2018**, *121*, 513–524. [CrossRef]

Article

An Approach for Integrating Uncertainty When Selecting an Anti-Torpedo Decoy in Brand New Warships

Rafael M. Carreño [1,*] **, Javier Martínez** [2] **and José Benito Bouza** [3]

1   Centro Universitario de la Defensa, Universidad de Vigo, Plaza de España s/n, 36920 Marín, Spain
2   Escuela Superior de Ingeniería y Tecnología, Universidad Internacional de La Rioja, 26006 Logroño, Spain;
    javier.martineztorres@unir.net
3   Departamento de Diseño en la Ingeniería, Universidad de Vigo, Escuela de Ingeniería Industrial,
    C/Torrecedeira, 86-36208 Vigo, Spain; jbouza@uvigo.es
*   Correspondence: rafaelm@cud.uvigo.es

Abstract: The Spanish Navy has planned that the F-80 frigates will be replaced by the brand new F-110 frigates in 2022. The F-110 program is in the conceptual design phase and one of the objectives is to provide the new F-110 frigate with a salient anti-submarine capability. Therefore, it is necessary to choose what anti-torpedo decoy should be installed in the warship. The Joint Chiefs of Navy Staff (EMA) established some guidelines and, considering the Navy guidance's, the Analytic Hierarchy Process (AHP) method was applied. After applying the AHP method, none of the decoys obtained a better score to the other one to make a decision. This paper addresses the problem of the selection of the best anti-torpedo decoy to be installed in the new frigates. This allowed implementing a new approach, the Graphic Method of Measurement of Uncertainty Beyond Objectivity (GMUBO). This approach considers different scenarios from the AHP, quantifies the uncertainty, and evaluates which is the best alternative. The method integrates the uncertainty in the AHP and allows measuring the robustness of the selected alternative, also providing a useful graphical tool. Furthermore, GMUBO has a great ease of use and it is helpful to make decisions under uncertainty conditions.

Keywords: F-110 frigate; decision-making; ASW; anti-torpedo decoy; AHP; uncertainty modelling

1. Introduction

Currently, the Spanish Navy has two types of frigates in service, the F-80 class and the F-100 one. F-80 frigates will be replaced in 2022, after 35 years of service, due to obsolescence. It is planned that F-80 units will be replaced by brand new F-110 frigates. At present, the F-110 program is in the definition and decision phase, specifically in the determination stage of obtaining alternatives [1]. This stage will end when the feasibility document (DDV) is signed. DDV establishes that a certain technical proposal, with known costs and deadlines, is feasible and can be hired. It is necessary that Spain's government take the necessary steps to guarantee a financing of the program that makes the hiring possible [2].

The new F-110 frigates will be complex vessels properly equipped so that the Spanish Navy carries out operational and strategic missions for the defense and security of Spain and its allies. The F-110 frigates must be designed to be balanced ships in all areas of naval warfare, but with a remarkable anti-submarine warfare (ASW) capability. In warfare, there are two categories of protection in order to defeat the threat: destruction of the threat (hard kill) and distraction of the threat (soft kill). Hard kill includes both kinetic and directed energy weapons. Soft kill is the prevention of a successful attack by the threat after the attack has commenced. Distraction could be accomplished through decoys, chaff,

jammers, flares, signature management, and electronic, infrared, or acoustic countermeasures [3]. ASW is one of the most complex concerns in a surface warship. The concept of uncertainty is deeply linked to ASW due to the lack of dominance of warships over the underwater environment.

The presence of an enemy submarine close to the operations area of a surface warship is a dangerous threat against the ship. Furthermore, the presence of submarines not only represents a direct threat against the physical integrity of a ship. Gathering intelligence data is considered one of the main objectives of a submarine, and a major threat to the national defense of the warship's country. Therefore, one of the important decisions that must be taken is the choice of the best anti-torpedo decoy that will be implemented in the F-110.

The ASW capability possessed by some surface warships poses an important threat to submarines. These warships can use ASW countermeasures to prevent submarines from accessing certain areas due to the risk of detection and to avoid torpedo damage [4]. When a submarine fires a torpedo, it goes straight to the target because it is fitted with an electronic device that enables it to find and hit the target. An anti-torpedo decoy is an acoustic device used as a countermeasure to avoid the attack of torpedoes. A decoy works by transmitting the emulated ship's signature to confuse the torpedo [5].

Both the conception and design of a warship are part of the process of developing a product of great complexity, i.e., the warship. This process is essentially a continuous decision-making process for which a number of techniques and methods could be used to help decision-making. This implies the consideration of multiple criteria and the possibility of conflict between these criteria. Therefore, the use of methods multiple-criteria decision-making (MCDM) methods is required to solve the decision-making problem. It can be said that all techniques and methods have two important objectives for these techniques to be successful and applicable in the industry: ease of use and exactness of results. However, both objectives can easily conflict. Exactness and uncertainty are closely related since uncertainty affects the variables involved in the decisions. A method that does not consider uncertainty will not provide correct results. Thus, the lower the uncertainty when applying the method, the greater the exactness of the results and, therefore, the greater the utility and applicability of the method. Methods that do not model uncertainty will fail in the exactness of their results. However, many methods that model uncertainty are not easy to apply.

One of the most used methods is the well-known Analytic Hierarchy Process (AHP), which is able to obtain a hierarchical classification of the different alternatives by comparing pairs of alternatives and criteria. A disadvantage of AHP is that it does not model uncertainty. Therefore, it is useful to have methods that consider the uncertainty inherent in MCDM problems. In recent years, MCDM methods have been developed to help decision-making in different fields [6–14].

Using MCDM methods to solve problems implies a complex analysis due to the number of factors that must be considered. Technical, political, institutional, standards, and economic factors take part of the problem [15]. Some methods that we want to mention in this work are Weighted Sum Method (WSM), ELimination Et Choix Traduisent la REalité (ELECTRE), Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), Multi-Attribute Utility Theory (MAUT), Preference Ranking Organization Method For Enrichment Evaluation (PROMETHEE), and AHP. However, WSM and AHP stand out for their simplicity of calculation and their adaptability to multiple situations.

This work seeks to provide an approach that helps the EMA to decide on the best anti-torpedo decoy for the F-110 frigates. Two criteria, (1) logistics, and (2) operational capabilities were the monitored and chosen criteria by the EMA. Keeping in mind the Navy guidelines, two possible decoy alternatives were chosen: towed device (N) and expendable device (L) [16]. The towed device can send acoustic signals to deceive the torpedo away from the ship. This device is a passive and electro-acoustic system that provides deceptive countermeasures, by simulating a variety of machinery noises. The expendable device is a system to deceive torpedoes as the towed device does. The difference (Figure 1) lies in that the expendable device is launched into the sea from the warship and the towed device is dragged. A group of experts from the Navy established the criteria weights and the AHP method [17] was applied to determine the best anti-torpedo decoy. The AHP result

is not enough to make a clear decision since sensitivity analysis identified that the result depends on criteria weights. Consequently, none of the decoys obtained a better score to make a correct decision. This brought uncertainty due to the criteria weights could come from subjective evaluations. However, by integrating uncertainty in the AHP, through the medium of the Graphic Method of Measurement of Uncertainty Beyond Objectivity (GMUBO) [18], it is possible to help the EMA to make a decision. GMUBO considers different scenarios and results of the AHP, which is the starting point. Moreover, it allows measuring the robustness of the selected alternative by means of a mathematical and graphical tool.
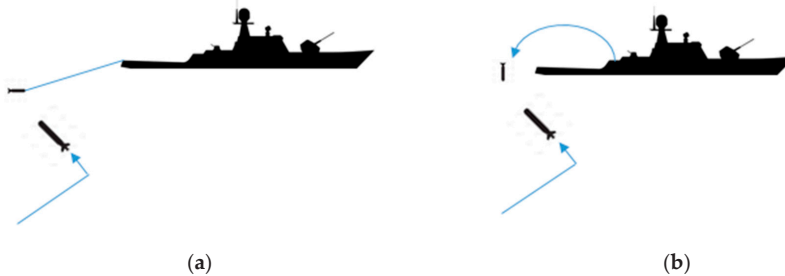

(**a**) (**b**)

Figure 1. Difference between towed and expendable devices: (**a**) a towed device is dragged by the warship; and (**b**) an expendable device is launched from the warship.

The capability of managing the uncertainty is one advantage of GMUBO, just as it enables alternatives to be assessed graphically and determine the best one. An additional benefit of GMUBO is that it considers changes in the process conditions themselves, when assessing and selecting alternatives. A case study is presented to demonstrate the approach by comparing AHP (initial method) with GMUBO. Managing the selection of alternatives, under uncertainty, in a graphic way, and considering the susceptibility of the selection to changes, are the novelties of the approach.

2. Methodology

A frigate is a surface warship and the new F-110 frigates should be balanced ships in all areas or scenarios of naval warfare. Keeping in mind that one of the main threats to surface ships is the presence of submarines, the F-110 should have a great capability of survival against possible attacks. This is the reason why one of the objectives is to provide the new frigate F-110 with an important anti-submarine capability [19]. Considering the aforementioned statements, the EMA decided to preselect two possible decoy alternatives: towed device (N) and expendable device (L), as indicated in the previous section. In the same way, the EMA established two criteria: logistics issues and operational capacities. The meaning of the criteria is detailed below.

1. Logistics issues. This allows evaluating different aspects: storage of decoys, volume they occupy, existence of available means to train personnel in the use of the type of decoy.
2. Operational capacities or effectiveness issues. This is an essential criterion to make an adequate decision to the tactical and strategic demands of a surface ship in the field of ASW.

Initially, the methodology that was followed is provided on the left of the Figure 2. According to the Navy guidelines, alternatives, criteria, and sub-criteria are the starting point to manage the problem. Next, a survey was conducted by expert officers of the Navy to determine the weights of the criteria. After applying the AHP, a sensitivity analysis was carried out by varying the weights assigned to the criteria. Since none of the alternative decoys could be selected in a suitable way, it was necessary to change the methodology. Hence, the GMUBO methodology was implemented, which allows integrating uncertainty into the AHP results.
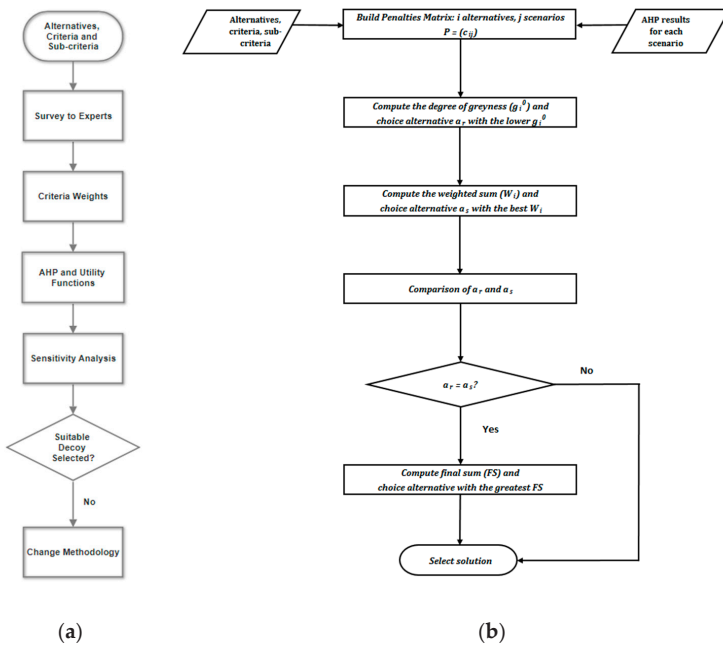
(a) **(b)**

Figure 2. Initial methodology versus the GMUBO methodology: (**a**) Initially, neither the uncertainty nor different scenarios were considered; (**b**) GMUBO considers the uncertainty since it foresees different scenarios that could occur.

GMUBO methodology is applied considering two data sets: (1) alternatives, criteria and sub-criteria and, (2) AHP results from different scenarios. The application of GMUBO is shown on the right of the Figure 2.

Since the initial methodology was modified, it must be indicated that it was improved by integrating the uncertainty. Therefore, GMUBO starts from the data provided by the EMA, that is, decoy alternatives, criteria, and sub-criteria. Likewise, the data of the survey made by the experts were used, obtaining the weightings of criteria and sub-criteria. From here, the scenarios are defined and the AHP method is applied. Therefore, the EMA guidelines and the AHP results allow starting GMUBO. The following paragraphs give a breakdown of the whole process.

First, the chosen criteria were monitored by the EMA and the sub-criteria that emanate from each of the two criteria are provided in Figure 3. All values are fictitious due to real ones being classified information.
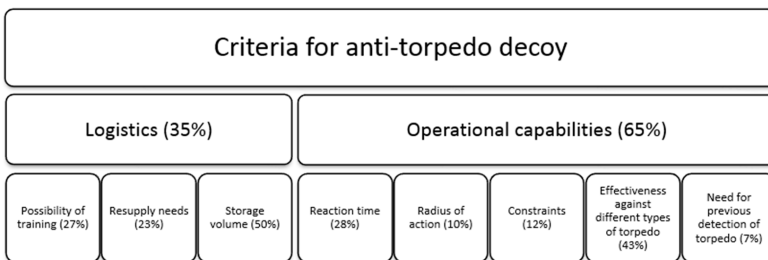

Figure 3. Criteria and sub-criteria established by the EMA.

Second, the survey was conducted as follows (Figure 4): A scale of numbers proposed by Saaty indicates how many times more important one element is over another element, considering the criterion with respect to which they are compared [20]. Let x_{ij} denote the intensity of importance of the criteria being compared, that is, logistics, and operational capabilities. Let y_{ij} denote the intensity of importance of the sub-criteria belonging logistics criterion. Finally, z_{ij} denotes the intensity of importance given to the sub-criteria related to operational capabilities criterion. Pairwise comparison matrices for criteria and sub-criteria were given to experts of the Navy in order to fill in the x_{ij}, y_{ij}, and z_{ij} values. The numerical values to denote the intensity of the importance of criteria and sub-criteria are the integers from 1 to 9. The results were treated statistically to obtain the weights of the criteria, which were used in the AHP method.

Criteria evaluation

	Logistics	Operational Capabilities
Logistics	1,00	x_{12}
Operational Capabilities	$1/x_{12}$	1,00

Sub-criteria evaluation

Logistics	Possibility of training	Resupply needs	Storage volume
Possibility of training	1,00	y_{12}	y_{13}
Resupply needs	$1/y_{12}$	1,00	y_{23}
Storage volume	$1/y_{13}$	$1/y_{23}$	1,00

Operational Capabilities	Reaction time	Radius of action	Constraints	Effectiveness against different types of torpedo	Need for previous detection of torpedo
Reaction time	1,00	z_{12}	z_{13}	z_{14}	z_{15}
Radius of action	$1/z_{12}$	1,00	z_{23}	z_{24}	z_{25}
Constraints	$1/z_{13}$	$1/z_{23}$	1,00	z_{34}	z_{35}
Effectiveness against different types of torpedo	$1/z_{14}$	$1/z_{24}$	$1/z_{34}$	1,00	z_{45}
Need for previous detection of torpedo	$1/z_{15}$	$1/z_{25}$	$1/z_{35}$	$1/z_{45}$	1,00

Figure 4. Survey conducted to experts of the Navy.

Next, the AHP method was applied to determine the best anti-torpedo decoy. The decision process based on the AHP considers a finite number of alternatives x_i, for i from 1 to n. A score is assigned to each alternative (w_i is the score of alternative x_i), providing a weight vector. A square

matrix of pairwise comparison is used to solve the multi-criteria decision-making (MCDM). Let A be the pairwise comparison matrix:

$$A = (a_{ij}) \tag{1}$$

where $a_{ij} = 1/a_{ji}$ and $a_{ii} = 1$ for all i and j from 1 to n [17]. Saaty proposed a consistency index (CI) to evaluate the consistency of the pairwise comparison matrix:

$$CI = (\lambda_{max} - n)/(n - 1) \tag{2}$$

To deduce the weight vector, is used the eigenvector theory. The method consists of finding the Perron-Frobenius eigenvector [21], which corresponds to the maximum eigenvalue of the pairwise comparison matrix, that is:

$$A \cdot w = \lambda_{max} \cdot w \tag{3}$$

The AHP decomposes a problem into a hierarchy of smaller sub-problems, which can more easily be evaluated. Thus, the AHP provides a hierarchy of goal, criteria (c_i), sub-criteria (sc_{ij}), and alternatives (a_i), as shown in Table 1.

Table 1. Hierarchy of the AHP.

Goal			
c_1	c_2	...	c_n
sc_{11} ...	sc_{21} ...	...	sc_{n1} ...
a_1	a_2	...	a_m

Two sub-criteria needed to be analyzed differently from the rest. For the logistics sub-criteria "storage volume" (SV), a utility function was used. This allows adding objective data of the real volume occupied by each decoy in a warship. Similarly, a utility function is used to evaluate the sub-criteria "reaction time" (RT). This adds objectivity since the real values of the time it takes to make effective use of the decoy of each of the alternatives are known [22]. Thus, the linear utility functions shown in Equations (4) and (6) were considered:

$$y_{SV} = C_1 \cdot x_{SV} + C_2 \tag{4}$$

$$\text{where } x_{SV} = 5 \text{ m}^3 \text{ if } y_{SV} = 0 \text{ and } x_{SV} = 1 \text{ m}^3 \text{ if } y_{SV} = 1 \tag{5}$$

$$y_{RT} = C_3 \cdot x_{RT} + C_4 \tag{6}$$

$$\text{where } x_{RT} = 30 \text{ s if } y_{RT} = 0 \text{ and } x_{RT} = 0 \text{ s if } y_{RT} = 1 \tag{7}$$

For a volume of 5 m^3 the minimum utility y_{SV} is assigned, i.e., a value of 0%, given that this volume is too large because it reduces space to other logistic needs of the vessel. Then, for a volume of 1 m^3 the maximum utility y_{SV} is assigned, i.e., a value of 100%, since it is not possible to store decoys in a smaller space.

In a like manner, the minimum utility y_{RT} (0%) happens when x_{RT} is 30 s. Half a minute is considered as the maximum time established to make effective use of decoy capabilities. That is, once that time has passed, it is considered that the vessel is no longer able to make effective use of any of the decoys. Likewise, the maximum utility y_{RT} (100%) happens when the reaction time is 0 s. That is to say, the optimal value of the reaction time is a null value, which would represent the ideal situation. However, it is only a reference, because in practice reaction time can never be zero. Hence, the utility functions are:

$$y_{SV} = -25 \cdot x_{SV} + 125 \text{ and } y_{RT} = -(10/3) \cdot x_{RT} + 100 \tag{8}$$

Next, the GMUBO method considers the uncertainty of the process. In order to achieve a robustness in the results, the uncertainty in the process must be integrated. To do this, different

scenarios were considered. These scenarios are changes to the weightings of the objective. Then, uncertainty of the alternatives, considered as grey numbers [23,24], was calculated. Subsequently, the best alternative was determined, taking into account that the scenarios are not controllable by the decision-maker.

In the AHP method, a criteria comparison matrix is multiplied by a priority vector and an overall priority vector (*OPV*) is obtained. The *OPV* determines a hierarchy on the selection of alternatives and provides a first selection that does not consider uncertainty. A decision maker does not know which scenario is going to arise. Then, it is necessary to repeat the process considering a number of scenarios. Each scenario allows to obtain a corresponding *OPV*. Keeping this in mind a Penalties Matrix ($m \times n$) is built:

$$P = (c_{ij}) \tag{9}$$

Consider S_A the set of alternatives and S_S the set of scenarios. Then:

$$S_A = \{a_i\} \text{ for } i = 1 \text{ to } m \tag{10}$$

$$S_S = \{s_j\} \text{ for } j = 1 \text{ to } n \tag{11}$$

In *P* matrix, each c_{ij} is the penalty obtained after choosing the alternative a_i when the given scenario is s_j.

To choose the best alternative two calculations are performed. On the one hand, a measure of the uncertainty is needed. On the other hand, weighted sums calculation is carried out. The best alternative should have the highest value of weighted sum and the lowest value for uncertainty.

Since the penalties are considered as grey numbers, the following expression is a measure of the uncertainty for each alternative [18]:

$$g_i^0 = (c_{ik} - c_{im})/(1 \cdot c_{im} + \Sigma \mu_{ij} \cdot c_{ij} + 0 \cdot c_{ik}) \tag{12}$$

where:

$$c_{ik} = \max c_{ij} \text{ and } c_{im} = \min c_{ij} \text{ with } \mu_{ij} \in [0,1] \tag{13}$$

Now the weighted sum is calculated for each alternative *i*:

$$W_i = \lambda_{i1} \cdot p_{i1} + \ldots + \lambda_{nn} \cdot p_{nn} \tag{14}$$

Finally, the best alternative should have the highest value of weighted sum and the lowest value for uncertainty. If both values lead to more than one alternative, then the alternative with the greatest final sum *FS* must be selected [18].

$$FS = W_i + (\max g_i^0 - g_i^0) + 1/\Sigma c_{ij} \tag{15}$$

3. Results and Discussion

The GMUBO methodology is fed from the results of applying the AHP method for different scenarios. The scenarios represent the way in which uncertainty is modeled since the decision-maker does not know what scenario will occur in the future. The changes in the scenario produce variations in the comparison matrices, all of which allows the implementation of GMUBO. One advantage of GMUBO is that it takes into account the imprecision that may be present in the minds of the experts when making the pairwise comparison.

The AHP method has wide applicability and allows dealing with complex problems by synthesizing them. Moreover, it establishes a ratio scale that makes easy the measurement [25]. The AHP has an axiomatic foundation and uses a clearly defined mathematical structure [26].

The AHP is a MCDM method that uses a hierarchy to represent a decision problem. The AHP assumes the hypothesis that each element in the hierarchy is independent. An advantage of the AHP is that it enables to include intangible qualitative criteria along with tangible quantitative criteria [27].

Although the criteria, sub-criteria and their weights were provided by a survey conducted with EMA experts, a utility functions had to be used for two of the sub-criteria. Specifically, the sub-criteria of SV and RT, since the real values of the storage volume and the reaction times were available. This was possible because the technical features of the different alternatives were available, all of which allowed the process to be more objective.

Table 2 provide the final decision matrix after applying the AHP. Then, as is shown in column 1, there is no clear preference for one of the alternatives. As shown in Table 2, the results are clearly different if only one of the criteria is taken into account. Specifically, alternative N acquires a clear advantage over alternative L if only the "logistics" criterion is considered in the evaluation. Similarly, if only the criterion "operational capabilities" is taken into account, it is observed that alternative L is better than alternative N.

Table 2. Final decision matrix after applying the AHP.

Alternatives	Logistics (35%)	Operational Capabilities (65%)
N (49.5%)	84%	41%
L (50.5%)	16%	59%

It follows that, once the AHP method is applied, practically equal results are obtained for both N and L alternatives. To be exact, it is obtained that the alternative L is slightly preferable to the alternative N, but with only a 1% difference between L and N. Since this difference is very small, it can be concluded that, given the available information, it is not advisable to establish which alternative is better.

Figures 5 and 6 show a sensitivity analysis obtained by varying the weights assigned to the criteria. The sensitivity analysis allows observing that a small variation in the weights assigned to the criteria produces that the chosen alternative changes. Therefore, the result obtained in the problem is very sensitive to variations in the influence or weight of the two criteria "logistics" and "operational capabilities".
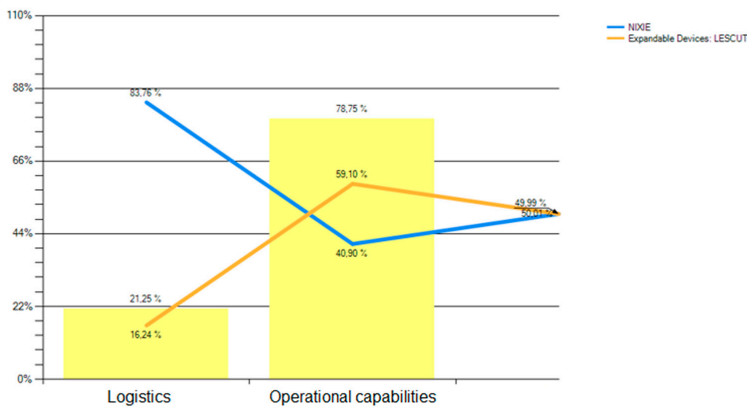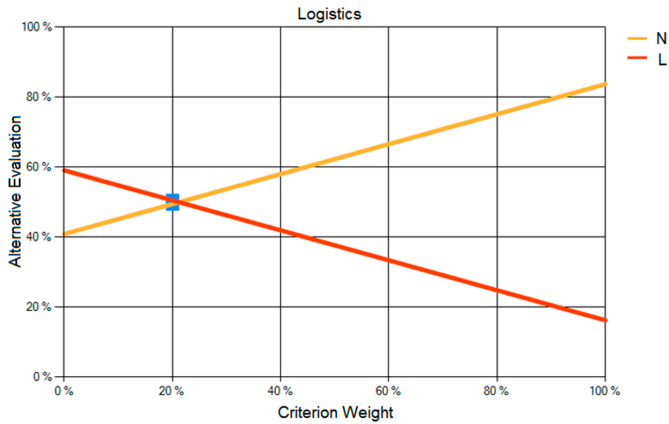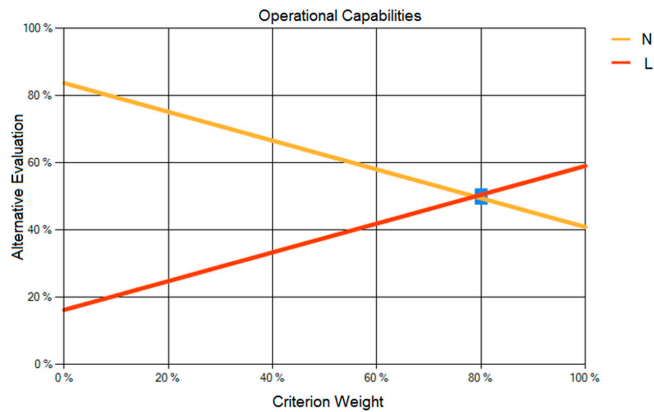

Figure 5. Sensitivity analysis.

(a)

(b)

Figure 6. Sensitivity analysis by considering different criteria: (**a**) Sensitivity of the weight assigned to the logistics criterion; and (**b**) sensitivity of the weight assigned to the operational capabilities criterion.

This means that, according to this result, it is not advisable to choose any of the two alternative solutions if you want to be certain of success. This is because the method does not clearly decide which alternative is preferable to the other. This is one source of uncertainty in the decision-making process. The AHP is a method that does not handle uncertainty although it is a good starting point to implement a better solution.

Different options were evaluated: re-evaluation and addition of more criteria, the use of complementary calculations, simulations or sophisticated software, and the integration of uncertainty in the AHP method. These options are discussed below.

Initially, by adding new criteria to the hierarchy to proceed to a re-evaluation, applying again the AHP. This option was discarded because it required more time and did not take into account the uncertainty inherent in the process. It should also be noted that this solution would imply that the experts would not have done their job correctly. However, the EMA took special care to define the most important criteria for decision-making. Moreover, the experts assigned the scores according to their experience and current environment.

Then, some mathematical techniques were considered as the theory of probability and fuzzy logic, among others, which do consider uncertainty. The implementation of these techniques involved

complex calculations, simulations and sophisticated software. Actually, decision-makers do not have time to develop techniques that require excessive amounts of knowledge. Hence, ease of use and exactness are basic features to encourage decision-makers in using of useful techniques.

Next, let us see an example of application of the GMUBO approach. With the proposed new approach, it is assumed that the result obtained through the application of the AHP is due to the consideration of a deterministic situation, that is, without uncertainty. This situation is called Scenario S_0. Therefore, to apply GMUBO, three additional scenarios are going to be assumed in order to show how this new approach works.

Then, scenarios S_1, S_2, and S_3 are considered, taking into account that these scenarios present different conditions that will modify the valuations given by the experts of the Navy. In the following points the characteristics and implications of each scenario are explained.

- Scenario S_0. This is the initial scenario, due to a deterministic and known situation, where AHP is applied. The final decision matrix was given in Table 2 above. In addition, in Figure 3 of Section 2, the assessments of the criteria and sub-criteria were provided.
- Scenario S_1. In this new scenario, the implementation of important logistical improvements in the Navy is assumed. This has produced a change in the valuations of the criteria and sub-criteria, which is presented in Table 3. The values given for "possibility of training" and "resupply needs" have changed.
- Scenario S_2. The existence of this scenario is based on the subjectivity of the values given to two of the sub-criteria of the operational capabilities criterion. This means that the sub-criteria "radius of action" and "constraints" are very variable. Therefore, the assessments given to them have been reviewed. The values are presented in Table 4.
- Scenario S_3. In this case, it is simply assumed that the weight of the logistics criterion is reduced by up to 15% and the operational capabilities criterion increases up to 85% (Table 5).

Table 3. Final decision matrix for S_1 after applying the AHP.

Alternatives	Logistics (40%)	Operational Capabilities (60%)
N (55.4%)	73.9%	43.1%
L (44.6%)	26.1%	56.9%

Table 4. Final decision matrix for S_2 after applying the AHP.

Alternatives	Logistics (30%)	Operational Capabilities (70%)
N (59.6%)	83.5%	49.3%
L (40.4%)	16.5%	50.7%

Table 5. Final decision matrix for S_3 after applying the AHP.

Alternatives	Logistics (15%)	Operational Capabilities (85%)
N (42.0%)	68.6%	37.9%
L (58.0%)	31.4%	62.1%

Now, following the steps explained in Section 2, the penalties matrix (P) is constructed from the OPV vector, obtained by applying the AHP to each scenario. Then, the matrix P is constructed from the matrix of the following OPV, given by the AHP:

$$\begin{bmatrix} 0.495 & 0.554 & 0.596 & 0.420 \\ 0.505 & 0.446 & 0.404 & 0.580 \end{bmatrix}$$

Once normalized, the matrix of penalties is obtained, where each c_{ij} represents the penalty incurred when choosing an alternative a_i when a scenario s_j has been given. Therefore, P becomes:

$$\begin{bmatrix} 0.101 & 0.042 & 0.021 & 0.176 \\ 0.091 & 0.150 & 0.192 & 0.016 \end{bmatrix}$$

Applying Equation (12) of Section 2, a measure of the uncertainty of each alternative is determined, from which it follows that:

$$g_1{}^0 = 1.4598 \tag{16}$$

$$g_2{}^0 = 1.6920 \tag{17}$$

The value 0 is assigned to the greatest c_{ij} in each column and 1 to the smallest. The other values are assigned from the linear interpolation between 0 and 1. Since it has the smallest uncertainty, alternative 1 should be chosen. However, according to the methodology exposed in Section 2, it is necessary to calculate the weighted sum to avoid high penalties. Then, calculating W_i for each alternative:

$$W_1 = 0.0630 \tag{18}$$

$$W_2 = 0.1070 \tag{19}$$

According to the methodology used, the best alternative is the one with the highest W_i, which would be alternative 2. Because both uncertainty ($g_i{}^0$) and weighted sum (W_i) led to more than one alternative, then the alternative with the highest final sum *FS* must be selected:

$$FS_1 = W_1 + (\max g_i{}^0 - g_1{}^0) + 1/\Sigma c_{1j} = 3.2364 \tag{20}$$

$$FS_2 = W_2 + (\max g_i{}^0 - g_2{}^0) + 1/\Sigma c_{2j} = 2.3342 \tag{21}$$

Eventually, the alternative with the highest FS turns out to be 1.

Finally, uncertainty was implemented using the GMUBO method, which is easy to use and useful to help decision-making under uncertainty, as well as providing a very useful graphical tool (Figure 7) [18]. GMUBO provides two vectors that are combined subsequently. On the one hand, a measure of the uncertainty given by the degree of greyness of each alternative [24]. On the other hand, for a given alternative, the inverse of the sum of its penalties is measured. Both vectors can provide a clear and singular alternative. However, there are situations where each vector leads to a different alternative. In these cases, the final sum (*FS*) would resolve the discrepancy.

FS measures the suitability of an alternative considering all the scenarios that can arise. *FS* controls and minimize the effect of a very small uncertainty that could modify the choice of the best alternative. It is considered that for a given alternative, the inverse of the sum of its penalties should be as large as possible to avoid alternatives with both high penalty values and a very small uncertainty.

GMUBO considers uncertainty and allows to measure the robustness of the selected alternative. Furthermore, it is a helpful method for decision-makers since it has a great ease of use.
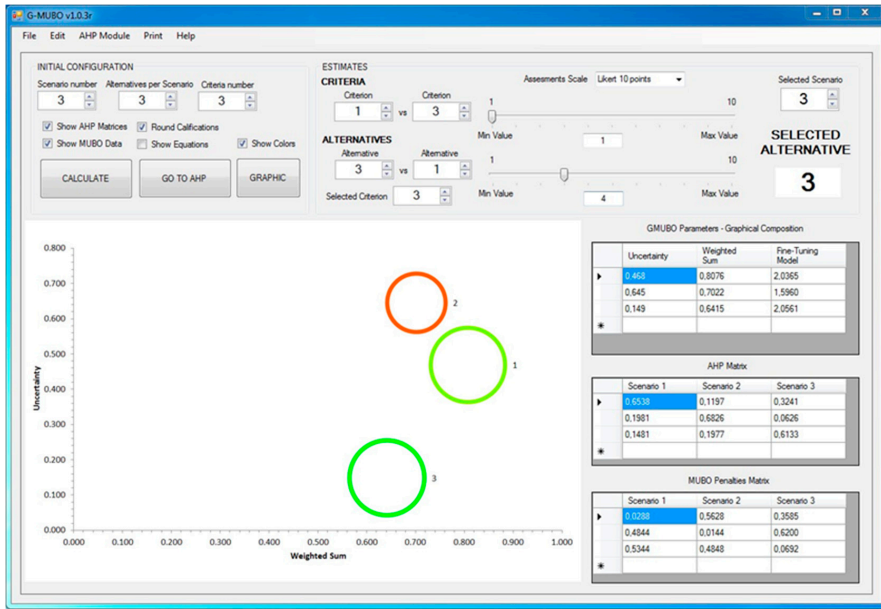
Figure 7. Graphical interface of GMUBO [17].

4. Managerial Implications

The design and construction of a warship is characterized by fundamental managerial implications for the EMA. One of the most important implications is the need of decision-making under uncertainty and unforeseen events. Thus, decisions about the capabilities of a warship involve not only engineering, but also a managerial level of analysis.

GMUBO manages the uncertainty combining it with the AHP and a weighted sum method. Weighted sum methods have advantages for multi-criteria decision-making: they are easier for understanding and to be handled by a decision-maker than complex calculations or simulations.

The approach proposed in the work would allow to managing uncertainty and risks to the decision-makers of the EMA. Hence, the findings of this work would be useful for enhancing the decision-making performance of the Spanish Navy.

When managing a MCDM problem, it is very important to know how sensitive or decisive each criterion is. In this way, managers can improve the decisions they must make. The sensitivity analysis show how sensitive the actual alternative is to changes on the current weights of the decision criteria. To learn more about the sensitivity analysis, the works [28–30] are available in the literature on the topic. In this work, a preliminary sensitivity analysis was done to assess the performance of the AHP. In Figures 5 and 6 of Section 3 this analysis could be used to demonstrate the convenience of improving the performance obtained with the AHP. Once the GMUBO method was applied, using a computer tool, it was possible to verify how this new approach allows the robustness of the selected alternative to be measured (some data and information have not been provided due to the confidentiality of the possible implementation of the method in the Navy).

5. Conclusions

Due to the inherent complexity of the ASW, the Spanish Navy has established that the new frigates of the F-110 class must possess a remarkable ASW capability. To get it, the F-110 frigates should be equipped with a soft kill protection based on anti-torpedo decoys that allows the distraction of the underwater threat. This work seeks to provide an approach that can help the EMA to decide on the best anti-torpedo decoy for the F-110.

The methodology initially used was based on the AHP method, which does not consider uncertainty. However, the uncertainty arises in the scenarios where a frigate develops its missions. Hence, the solution obtained with the AHP method does not provide sufficient robustness. Moreover, the problem of selecting the best anti-torpedo decoy is a complex decision and another approach is required. However, AHP is a good starting point to implement a better solution, even though it does not handle uncertainty. Since AHP does not manage uncertainty, it is interesting to look for an approach that allows it to be managed. To consider uncertainty is fundamental in decision-making processes. Some of the existing MCDM methods take uncertainty into account, but require additional knowledge to be used. In general, decision-makers usually do not have a strong theoretical foundation of mathematical knowledge. In a general way, ELECTRE, TOPSIS, MAUT, and PROMETHEE are less adapted for changing easily from one to another of various situations or fields. Moreover, a good knowledge of these methods may be difficult to achieve for an inexperienced user.

GMUBO are based in the AHP, which is capable of many applications and has a wide use. A great advantage of the AHP is that it decomposes a decision problem in parts and builds hierarchies. The above-mentioned methods do not provide this structuring capability. The integration of uncertainty in an intuitive and visual way, as well as simplicity, are some of the advantages of the proposed approach.

GMUBO provides a simple method for managing uncertainty, considering different scenarios that decision-makers do not control. GMUBO allows graphically visualizing the solution. The graphic tool that provides the method not only allows a better visualization of the problem, but also helps the decision-making process.

References

1. González-Cela, G.; Bellas, R.; Martínez, J.; Touza, R.; Carreño, R. Optimal Design of Spanish Navy F-110 Frigates Combat Information Center. *Nav. Eng. J.* **2018**, *130*, 79–90.
2. González del Tánago de la Lastra, A. La fragata F-110, una apuesta de presente y futuro. *Rev. Gen. Mar.* **2018**, *274*, 495–506.
3. Higgins, E.J.; Higgs, R.L.; Parkins, G.R.; Tionquiao, V.S.; Wells, C.K. Expeditionary warfare: Force protection. Available online: https://calhoun.nps.edu/handle/10945/6917 (accessed on 2 January 2019).
4. Riewerts, R.K. The Bleak Future of Undersea Warfare. Available online: https://apps.dtic.mil/docs/citations/AD1058272 (accessed on 2 January 2019).
5. Liang, K.H.; Wang, K.M. Using simulation and evolutionary algorithms to evaluate the design of mix strategies of decoy and jammers in anti-torpedo tactics. In Proceedings of the 2006 Winter Simulation Conference, Monterey, CA, USA, 3–6 December 2006; pp. 1299–1306.
6. Pamučar, D.; Stević, Ž.; Zavadskas, E.K. Integration of interval rough AHP and interval rough MABAC methods for evaluating university web pages. *Appl. Soft Comput.* **2018**, *67*, 141–163. [CrossRef]
7. Stević, Ž.; Pamučar, D.; Zavadskas, E.K.; Ćirović, G.; Prentkovskis, O. The selection of wagons for the internal transport of a logistics company: A novel approach based on rough BWM and rough SAW methods. *Symmetry* **2017**, *9*, 264. [CrossRef]

8. Stević, Ž.; Pamučar, D.; Vasiljević, M.; Stojić, G.; Korica, S. Novel integrated multi-criteria model for supplier selection: Case study construction company. *Symmetry* **2017**, *9*, 279. [CrossRef]

9. Sremac, S.; Stević, Ž.; Pamučar, D.; Arsić, M.; Matić, B. Evaluation of a Third-Party Logistics (3PL) Provider Using a Rough SWARA–WASPAS Model Based on a New Rough Dombi Agregator. *Symmetry* **2018**, *10*, 305. [CrossRef]

10. Stević, Ž.; Pamučar, D.; Subotić, M.; Antuchevičiene, J.; Zavadskas, E.K. The location selection for roundabout construction using Rough BWM-Rough WASPAS approach based on a new Rough Hamy aggregator. *Sustainability* **2018**, *10*, 2817. [CrossRef]

11. Pamučar, D.; Stević, Ž.; Sremac, S. A New Model for Determining Weight Coefficients of Criteria in MCDM Models: Full Consistency Method (FUCOM). *Symmetry* **2018**, *10*, 393. [CrossRef]

12. Radovic, D.; Stevic, Z.; Pamučar, D.; Zavadskas, E.K.; Badi, I.; Antuchevičiene, J.; Turskis, Z. Measuring of performance in transportation companies in developing countries: A novel rough ARAS model. *Symmetry* **2018**, *10*, 434. [CrossRef]

13. Pamučar, D.; Badi, I.; Korica, S.; Obradović, R. A novel approach for the selection of power generation technology using a linguistic neutrosophic combinative distance-based assessment (CODAS) method: A case study in Libya. *Energies* **2018**, *11*, 2489. [CrossRef]

14. Pamučar, D.; Pejčić-Tarle, S.; Parezanović, T. New hybrid DEMATEL-MAIRCA model: Sustainable selection of a location for the development of multimodal logistics center. *Econ. Res. Ekon. Istraz.* **2018**, *31*, 1641–1665.

15. Kumar, A.; Sah, B.; Singh, A.R.; Deng, Y.; He, X.; Kumar, P.; Bansal, R.C. A review of multi criteria decision making (MCDM) towards sustainable renewable energy development. *Renew. Sustain. Energy Rev.* **2017**, *69*, 596–609. [CrossRef]

16. Brege, E.D. Design and Construction of a Low Cost, Modular Autonomous Underwater Vehicle. Master's Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2011.

17. Saaty, T.L. How to make a decision: the analytic hierarchy process. *Eur. J. Oper. Res.* **1990**, *48*, 9–26. [CrossRef]

18. Carreño, R.M.; Comesaña, A.; Bouza, J.B. GMUBO: A new graphical method that integrates uncertainty into the process for assessing and selecting design alternatives. *Res. Eng. Des.* (under review).

19. Cao, L.; Wang, H.Y.; Lu, F.X. Operational Capability Evaluation of Surface Warship Formation Based on Set Pair Analysis. In Proceedings of the 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Hangzhou, China, 27–28 August 2016; pp. 492–495.

20. Saaty, T.L. Decision making with the analytic hierarchy process. *Int. J. Serv. Sci.* **2008**, *1*, 83–98. [CrossRef]

21. Brunelli, M. Priority vector and consistency. In *Introduction to the Analytic Hierarchy Process*; Springer: Cham, Switzerland, 2015; pp. 21–34.

22. Marzouk, M.; Moselhi, O. A decision support tool for construction bidding. *Constr. Innov.* **2003**, *3*, 111–124. [CrossRef]

23. Liu, S.; Lin, Y. *Grey Information: Theory and Practical Applications*; Springer: London, UK, 2006; pp. 3–43.

24. Liu, S.; Fang, Z.; Yang, Y.; Forrest, J. General grey numbers and their operations. *Grey Syst. Theory Appl.* **2012**, *2*, 341–349. [CrossRef]

25. Forman, E.H.; Gass, S.I. The analytic hierarchy process—An exposition. *Operat. Res.* **2001**, *49*, 469–486. [CrossRef]

26. Saaty, T.L. Axiomatic foundation of the analytic hierarchy process. *Manag. Sci.* **1986**, *32*, 841–855. [CrossRef]

27. Partovi, F.Y. An analytic model to quantify strategic service vision. *Int. J. Serv. Ind. Manag.* **2001**, *12*, 476–499. [CrossRef]

28. Triantaphyllou, E.; Sánchez, A. A sensitivity analysis approach for some deterministic multi-criteria decision-making methods. *Decis. Sci.* **1997**, *28*, 151–194. [CrossRef]

29. Pamučar, D.; Božanić, D.; Ranđelović, A. Multi-criteria decision making: An example of sensitivity analysis. *Serb. J. Manag.* **2017**, *12*, 1–27. [CrossRef]

30. Mukhametzyanov, I.; Pamucar, D. A sensitivity analysis in MCDM problems: A statistical approach. *Decis. Mak. Appl. Manag. Eng.* **2018**, *1*, 1–27. [CrossRef]