Stochastic dynamics for epidemics based on a compartmental scheme: An application to the AH1N1 influenza

A. Reyes-Romero

Posgrado en ciencias de la tierra, Instituto de Geofísica, Universidad Nacional Autónoma de México,
Apartado Postal 22-582, 14000, Ciudad de México, México.

J. E. Fernández

Coordinación de Vigilancia Epidemiológica, Instituto Mexicano del Seguro Social,
Mier y Pesado No. 120, Col. Del Valle, Delegación Benito Juárez, 03100, Ciudad de México, México.

J. Adrian Reyes

Instituto de Física, Universidad Nacional Autónoma de México,
Apartado Postal 20-364, 01000, Ciudad de México, México.

In this paper a stochastic formulation for describing the dynamics of an epidemic is established. We state our model in the language of a compartmental scheme of three population groups susceptible, infected, and removed, for which the model parameters are extended to be statistical distributions of time. A master equation is established for the dynamics of the probability density P of finding the system in the state characterized by the values of the aleatory variables susceptible, infected, removed, and time t. Our stochastic formalism allows us to recover the associated deterministic model in terms of the expected values of susceptible, infected, and removed; whereas the second momenta of P provides us with statistical standard deviations for these three variables which delimit the region in which most of the particular realizations are to be expected. We have applied the analysis developed here, for studying the specific case of the influenza AH1N1 that took place in Mexico in 2009. The reported data by the main Mexican Health institution are in good agreement with the predictions of our model for the standard deviation of the aleatory variables.

1. Introduction

Epidemic spread in human populations is a complex phenomenon, throughout human history infectious diseases have caused debilitation and premature death to large portions of the human population, leading to serious social-economic concerns. The bubonic plague, smallpox, dengue, ebola, yellow fever, influenza, malaria, polio, avian flu, leprosy, tuberculosis, whooping cough, severe acute respiratory syndrome (SARS), among others, are examples of the most devastating ones (see [1]). Although, some viruses and bacterias causing these diseases have been completely eradicated others have only been possible to be controlled and are a health problem yet. A recent case took place in the late March and early April 2009, when reports of respiratory hospitalizations and deaths among young adults in Mexico alerted local health officials to the occurrence of atypical rates of respiratory illness at a time when influenza was not expected to reach epidemic levels. Infections with novel swine-origin influenza AH1N1 virus were confirmed in California, (United States), on April 21 (see [2]) and in Mexico on April 23 (see [3]).

The formulation of theories on the nature of infections diseases dates back to ancient times; they have evolved and currently modern theories of differential equations have been widely adopted in the analyses, and simulation of biological populations including the evolution of epidemics. The

first mathematical model to describe an epidemic was developed by D. Bernoulli which was applied to smallpox. Later, W. Heaton formulated a model to analyze the propagation of measles and R. Ross developed a work on malaria transmission. On the basis of a scheme with three elementary compartments: susceptible, infected, and removed, a mathematical model more general commonly known as SIR, was developed by [4] for describing any disease of viral kind.

Most epidemic models are based on dividing the host population into a small number of compartments, each containing individuals that are identical in terms of their status with respect to the disease in question. For instance, the standard approaches are the so-called susceptible-infected-removed (SIR) and susceptible-exposed-infected-removed (SEIR) models have contributed much in the present for understanding the nature of recurrent epidemics. Traditionally, deterministic models have been the base of mathematical epidemiology, however, stochasticity has been recognized as an important tool within epidemic modelling, but there is still much concerning its precise dynamical role and how it can be understood from a theoretical point of view. Nevertheless, stochastic models have recently been considered as a more realistic descriptions of epidemics. This is evidenced by the large amount of systematic treatments of the subject, for instance, the first stochastic simulation on epidemiological system, was implemented by Bartlett, who was inter-

ested in extinction dynamics (see [5]). In more recent studies, [6], have applied the well known Van Kampen method to the SEIR model with distributed exposed and infectious periods for studying whooping cough. Moreover, McKane *et al.* developed a method for fluctuations around cycles of forced or unforced systems, a method which was extended by [7] to a seasonally forced SEIR model in a realistic parameter region for childhood infectious diseases where different attractors exist or coexist. According to studies on avian flu [8], its is argued that a fully stochastic theory based on environmental transmission provides a simple, plausible explanation for the phenomenon of multi-year periodic outbreaks. On the other hand, a stochastic theory for the major dynamical transitions in epidemics from regular to irregular cycles, which relies on the discrete nature of disease transmission and low spatial coupling was proposed by [9], while a study on the estimation the parameters in dengue fever case studies was performed by [10]. Among the studies on epidemics in Mexico, we can mention, for instance, the study performed by [11], who used a time dependent modification of the [4] model to study the evolution of the influenza AH1N1 epidemic reported in the Mexico City area under the control measures used during April and May 2009. Meanwhile, [12] analyzed the epidemiological patterns of the pandemic during April-December 2009 in Mexico and evaluated the impact of non-medical interventions, school cycles, and demographic factors on influenza transmission.

This work is organized as follows. In order to establish the terminology and nomenclature used in our analysis, in Sec. 2 we summarize a deterministic mathematical model SIR and we give a brief description of the general master equation. An application of the stochastic formalism that we have developed on a realistic case is presented in Sec. 3. In Sec. 4, we present an analysis where the important parameters and numerical comparison are discussed. Finally in Sec. 5 we address our concluding remarks and conclusions.

2. The SIR model

In Mathematical Epidemiology, one of the most widely used models still is the simple SIR model developed in Ref. [4]. In Subsec. 2.3, we will present a stochastic extension of such model. For context, we will briefly review the deterministic SIR model. For more details on this mathematical formulation, Ref. [1] can be consulted.

2.1. Model description and assumptions

In a classic model SIR , individuals within the population N belong to one of three compartments: susceptible, infected, and removed, which are denoted by S, I, and R, respectively. These groups are defined as follows, the susceptibles: those susceptible to contract the disease, but who are not yet infected at a time t, the infected: those which are infected with the disease, and are able to spread it through contact with susceptibles. The final class are the removed: those that have

been removed from the infected and cannot spread the infection (for instance, because they are immunized).

The following hypotheses are assumed: 1) it is considered a homogeneous population without demographic dynamic, in which individuals have the same characteristics; age, sexual gender, among others, 2) it is neglected a possible seasonal period in which the dynamics of the epidemic vary, 3) it is excluded the possibility that the removed individuals lose their immunity to the disease so that they cannot belong to the susceptible group again, 4) it is ignored the virus incubation time, which implies that when a subject is infected it becomes infectious immediately, 5) it is considered that two individuals have the same probability of interacting each other, 6) it is supposed that infected persons, that are within the incubation period but do not exhibit any symptoms, are unable to contaminate to other individuals, 7) the rate of the number of infected per unit of time increases proportionally to the number of infected and susceptible, namely $\beta SI/N$ where $\beta > 0$ constant and the number of susceptible members decreases with the same rate of infected, 8) the recovery rate of infected is proportional to the number of infected, this is γI with $\gamma > 0$ constant and the number of retrieved increases at the same rate. Here, the parameter β represents the transmission rate (per capita) while γ represents the recovery rate, so the mean infectious period is $1/\gamma$.

2.2. Governing equations

Under the conditions mentioned above, the model can be mathematically expressed by the following set of ordinary differential equations,

$$\frac{dS}{dt} = -\frac{\beta}{N}SI, \quad S(0) = S_0 \geq 0,$$

$$\frac{dI}{dt} = \frac{\beta}{N}SI - \gamma I, \quad I(0) = I_0 \geq 0,$$

$$\frac{dR}{dt} = \gamma I, \quad R(0) = R_0 \geq 0, \tag{1}$$

where $N = S + I + R$ is the fixed total population. If everyone is initially susceptible ($S(0) = N$), then a newly introduced infected individual can be expected to infect other people at the rate βN during the expected infectious period $1/\gamma$. Thus, this first infected individual can be expected to infect $\mathbb{R}_0 = \beta/\gamma$ individuals. The number $\mathbb{R}_0$ is called the basic reproduction number and is unquestionably the most important quantity to consider when analyzing any epidemic model for an infectious disease. In particular $\mathbb{R}_0$ determines whether there is an epidemic or not. If $\mathbb{R}_0 < 1$ the infection dies out, while if $\mathbb{R}_0 > 1$ there is an epidemic.

2.3. General master equation

The master equation governs the stochastic dynamics of Markov process. This equation is universal and has been applied in problems in biology and population dynamics which

can be modeled as Markovian events. A process is considered Markovian if it is possible to make predictions and indeed know the full history of the process about the system, based solely on its present state, which implies that the past and future state of the system are independent. Therefore, a suitable description for Markovian systems is given by the master equation, that specifies how the probability $P(n_i, t)$ that the system is in the state n_i at time t changes with time. Its general form is given by,

$$\frac{dP(n_i, t)}{dt} = \sum_{n_i' \neq n_i} T(n_i \mid n_i') P(n_i', t)$$

$$- \sum_{n_i' \neq n_i} T(n_i' \mid n_i) P(n_i, t). \qquad (2)$$

Equation (2) manifests explicitly that the temporal changes of the probability are controlled by a balance equation that takes into account both transition probabilities per unit time $T(n_i \mid n_i')$ and $T(n_i' \mid n_i)$. The transition rate $T(n_i \mid n_i')$ corresponds to a system for which n_i individuals in the population group i migrates to the compartment i' which was originally formed by n_i' members of the whole population. Similarly, the second transition rate $T(n_i' \mid n_i)$ is measuring the probability for occur the inverse process. In this sense, the positive terms of this balance account for individuals entering into the population group i and the negative ones to those members of the population leaving the same compartment (see [13]).

2.4. One step master equation

In what follows, the description done so far on the master equation, will be limited only to the three population groups: susceptible, infected, and removed, as was discussed above are denoted by S, I, and R, respectively. We also will consider that an individual only has interaction with their nearest neighbour.

We suppose that the transition rule of the population members between the groups S, I, R is as follows: The susceptible individuals, initially healthy, may become infected with a rate $\beta SI/N$, whereas the infected individuals, have the disease and are able to transmit it to other individuals at a rate γI, the removed individuals are immune to the disease after recovery. For simplicity in this model, it is assumed that during the epidemic, the virus is transmitted only between close neighbours with the possibility that an infected individual may be removed. Figure 1 illustrates schematically the transition flows from one group to another. The elementary processes occurring into the system can be divided in two
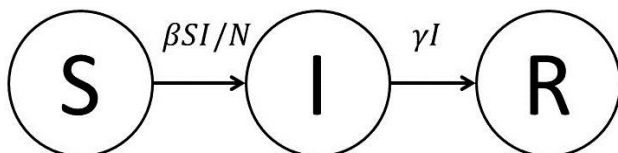
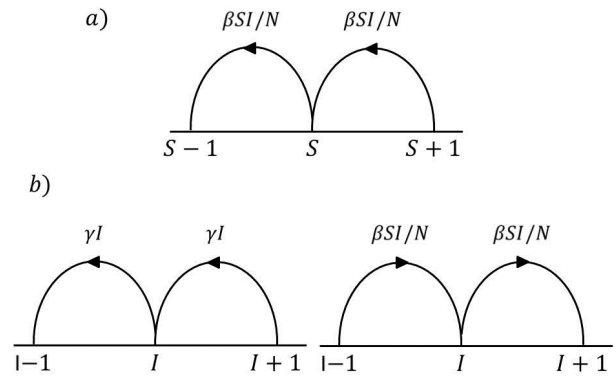

FIGURE 1. Flow chart for the SIR model.

FIGURE 2. Processes of interaction between individuals.

elementary groups; infection and recovery, which are defined as follows: An infection process implies that a susceptible individual enters in contact with an infected individual, generating therefore a newly infected individual. The recovery process signifies that an infected individual who is able to recover from the disease, cannot be susceptible again, because he has acquired immunity. In summary, the transition rates expressed in terms of the SIR model variables are defined by the following events,

$$\{S, I\} \longrightarrow \{S - 1, I + 1\}$$
$$\{S, I\} \longrightarrow \{S, I - 1\}, \qquad (3)$$

which are represented schematically in Fig. 2 together with their respective transition rates. We take just two variables $n = \{S, I\}$, because the variable R can be completely determined by the restriction $N = S + I + R$. In this way Eq.(2), can be written explicitly as

$$\frac{dP(S, I, t)}{dt} = T(S, I \mid S + 1, I - 1) P(S + 1, I - 1, t)$$

$$+ T(S, I \mid S, I + 1) P(S, I + 1, t)$$

$$- T(S - 1, I + 1 \mid S, I) P(S, I, t)$$

$$- T(S, I - 1 \mid S, I) P(S, I, t). \qquad (4)$$

Equation (4) is a master equation describing a one-step process, whose corresponding associated transition rates we shall assume by proposing the following expressions,

$$T(S - 1, I + 1 \mid S, I) = \frac{\beta SI}{N}, \qquad (5)$$

$$T(S, I - 1 \mid S, I) = \gamma I. \qquad (6)$$

Here, $T(S - 1, I + 1 \mid S, I)$ measures the probability rate for an infected individual to become susceptible I, while the number of removed R is constant. We take this quantity to be proportional to the temporal rate of change of S provided by the Kermack McKendrick model, as given by the first equation of Eq. (1), which is conceptually resembling. Similarly, $T(S, I - 1 \mid S, I)$ gives the transition probability at which

the number of infected individuals increase by one while the susceptible ones S is kept constant. Because the total number is constant, this is related with the reduction of the recovery individuals R whose temporal ratio is provided by the last equation of Eq. (1). We identify both quantities as proportional to construct our master equation. For convenience, the one-step operator E, is defined through its effect on an arbitrary function $f(m)$ as $E^{\pm 1}f(m) = f(m \pm 1)$, thus in terms of the step operator, the probability transitions given by Eqs. (5) and (6), can be rewritten as follows

$$E^{\pm 1}f(m) = T(S, I \mid S+1, I-1)P(S+1, I-1, t)$$
$$= E_S^{+1}E_I^{-1}\left(\frac{\beta SI}{N}\right)P(S, I, t) \qquad (7)$$

and

$$E^{\pm 1}f(m) = T(S, I \mid S, I+1)P(S, I+1, t)$$
$$= E_I^{+1}\gamma IP(S, I, t). \qquad (8)$$

From transition rates represented by expressions (7) and (8), Eq. (4) can be written in a compact form as

$$\frac{dP}{dt} = \left[\left(E_S^{+1}E_I^{-1} - 1\right), \left(\frac{\beta SI}{N}\right) + \left(E_I^{+1} - 1\right)\gamma I\right]P. \qquad (9)$$

In order to solve Eq. (9), we use the method that has been proposed by [13]. This method suppose that the probability distribution P behaves like a sharp Gaussian statistical distribution, where the random variables are rewritten as the superposition of a macroscopic deterministic variable and a mesoscopic random variable. According to this formulation, the following ansatz for the independent variables are suggested, namely,

$$S(t) = N(\phi(t) + qx) \qquad (10)$$
$$I(t) = N(\psi(t) + qy), \qquad (11)$$

where ϕ and ψ are the macroscopic variables which represent both to susceptible individuals and infected individuals, respectively, and x and y are small stochastic corrections to these variables, $q = 1/\sqrt{N}$ is a smallness parameter. The Van Kampen approximation allows us to find the deterministic evolution of $\phi(t)$ and $\psi(t)$ as well as the two first momenta of the distribution P in terms of the corrections x and y. If we apply the step operators $E_S^{\pm 1}$ and $E_I^{\pm 1}$ to expressions (10) and (11), we have to replace $x \to x + q$ and $y \to y + q$, so the resulting expressions $E_S^{\pm 1}(x+q)$ and $E_I^{\pm 1}(y+q)$ are expanded in Taylor's series,

$$E_S^{\pm 1} = 1 \pm q\frac{\partial}{\partial x} + \frac{q^2}{2}\frac{\partial^2}{\partial x^2} + O(q^3), \qquad (12)$$

$$E_I^{\pm 1} = 1 \pm q\frac{\partial}{\partial y} + \frac{q^2}{2}\frac{\partial^2}{\partial y^2} + O(q^3). \qquad (13)$$

Inserting expressions (10-13) into Eq.(9), and multiplying the hand right side by NP, it becomes

$$\left[\beta\left\{\left(1 + q\frac{\partial}{\partial x} + \frac{q^2}{2}\frac{\partial^2}{\partial x^2}\right)\left(1 - q\frac{\partial}{\partial y} + \frac{q^2}{2}\frac{\partial^2}{\partial y^2}\right) - 1\right\}\right.$$
$$\left. \times(\phi + qx)(\psi + qy) + \gamma q\left(\frac{\partial}{\partial y} + \frac{q}{2}\frac{\partial^2}{\partial y^2}\right)(\psi + qy)\right]. \qquad (14)$$

Notice that the expression between square brackets up to second order in q^2 can be approximated as

$$\left(1 + q\frac{\partial}{\partial x} + \frac{q^2}{2}\frac{\partial^2}{\partial x^2}\right)\left(1 - q\frac{\partial}{\partial y} + \frac{q^2}{2}\frac{\partial^2}{\partial y^2}\right) - 1$$
$$= q\left[\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right] + \frac{q^2}{2}\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} - 2\frac{\partial^2}{\partial x\partial y}\right] + O(q^3). \qquad (15)$$

Thus, if we expand the other terms and group the resulting ones in powers of q we can obtain

$$\frac{dP}{dt} = \frac{1}{q}\left(\beta\phi\psi\frac{\partial P}{\partial x} + (\gamma\psi - \beta\psi\phi)\frac{\partial P}{\partial y}\right) + \left[\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right]$$
$$\times(x\psi + \phi y)P + \frac{\psi\phi}{2}\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} - 2\frac{\partial^2}{\partial x\partial y}\right]P$$
$$+ \frac{\gamma}{2}\psi\frac{\partial^2 P}{\partial y^2} + \gamma\frac{\partial}{\partial y}yP, \qquad (16)$$

where the terms of order q have been neglected. It is important to remark that Eq. (16) is the Fokker-Planck's equation (see [13]). Moreover, it is useful to express the probability distribution in terms of S and I, which in turn can be written as a function of the continuous variables x and y, namely, $P(S, I, t) = \Pi(x, y, t)$, it yields

$$\frac{dP}{dt} = \frac{\partial\Pi}{\partial t} - \frac{1}{q}\frac{d\phi}{dt}\frac{\partial\Pi}{\partial x} - \frac{1}{q}\frac{d\psi}{dt}\frac{\partial\Pi}{\partial y}. \qquad (17)$$

By comparing the right hand side of the Eq. (16) and that of the Eq. (17), it can be infered by equating the lowest power in q that

$$\frac{d\phi}{dt} = -\beta\phi\psi, \qquad (18)$$

$$\frac{d\psi}{dt} = \beta\phi\psi - \gamma\psi. \qquad (19)$$

Notice that the system of Eqs. (18-19), has the same functional form that SIR model described by Eq. (1), whose solutions $\phi = E[S]$ and $\psi = E[I]$, are the expected value of the Eqs. (10-11). On the other hand, up to zero order q the following equation is fulfiled.

$$\frac{\partial\Pi}{\partial t} = \beta\left[\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right](x\psi + \phi y)\Pi + \beta\frac{\psi\phi}{2}\left[\frac{\partial^2}{\partial x^2}\right.$$
$$\left. + \frac{\partial^2}{\partial y^2} - 2\frac{\partial^2}{\partial x\partial y}\right]\Pi + \frac{\gamma}{2}\psi\frac{\partial^2\Pi}{\partial y^2} + \gamma\frac{\partial}{\partial y}y\Pi \qquad (20)$$

which can be expressed as a continuity balance equation for the probability Π given by

$$\frac{\partial \Pi}{\partial t} + \nabla \cdot \mathbf{J} = 0, \tag{21}$$

where the two dimensional probability current vector $\mathbf{J}$ is defined as

$$\mathbf{J} = \beta \begin{pmatrix} 1 \\ -1 \end{pmatrix} (x\psi + \phi y)\Pi$$
$$+ \beta \frac{\psi\phi}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \frac{\partial \Pi}{\partial x} \\ \frac{\partial \Pi}{\partial y} \end{pmatrix}$$
$$+ \begin{pmatrix} 0 \\ 1 \end{pmatrix} \left(\frac{\gamma}{2} + \frac{\partial \Pi}{\partial y} + \gamma y \Pi \right). \tag{22}$$

We remark that Eq.(21) guaratees the conservation of probability.

To analyze the dynamics of the expected values of the ramdom variables x and y we multiply by x or by y Eq.(20) and integrate over the whole domain. It yields

$$\frac{d<x>}{dt} = -\beta <x\psi + \phi y>, \tag{23}$$

$$\frac{d<y>}{dt} = \beta <x\psi + \phi y - \gamma y>, \tag{24}$$

where as usual the expected value is given by

$$<u> = \int\limits_{-\infty}^{\infty} \int\limits_{\infty}^{\infty} u\Pi(x,y) dx dy. \tag{25}$$

Here u is a variable which can be a function of x and y. By adding Eqs. (23) and (24) we get the expression

$$\frac{d}{dt} (<x> + <y> - \gamma y) = 0, \tag{26}$$

which allows us to find a time invariant of the system. To find the dynamics of the second momenta, we proceed similarly and integrate by parts the resulting expressions by assuming that the density of probability Π and its partial derivatives with respect to x and y vanish at $x = \pm\infty$. It leads to

$$\frac{d<x^2>}{dt} = -2\beta <x(x\psi + \phi y)> + \beta\psi\phi, \tag{27}$$

$$\frac{d<y^2>}{dt} = 2\beta <y(x\psi + \phi y)>$$
$$+ \beta\psi\phi + \gamma\psi - 2\gamma <y^2>, \tag{28}$$

$$\frac{d<xy>}{dt} = \beta <(y-x)(x\psi + \phi y)>$$
$$- \gamma <xy> - \beta\psi\phi. \tag{29}$$

After adding the first and second expressions and substracting the third one we obtain following relation

$$\frac{d}{dt} <(x-y)^2> = \gamma\psi - 2\gamma <y(x+y)>, \tag{30}$$

It is convenient to stress that the model described by the set of Eqs.(23), (24), (27), (28), and (29) is general and therefore, it can be applied to any epidemic of viral transmission fulfilling the restriction of having a narrow statistical distribution; it means that he solution $P(S, I, t)$ of Eq. (9) obeys a Gaussian distribution with averages $\phi(t)$ and $\psi(t)$, whose fluctuations bands $\phi(t) \pm \sigma_S$ and $\psi(t) \pm \sigma_I$, are given by the solutions of the Eqs.(23), (24), (27), (28), and (29). Here the respective standard deviations of each variable are defined as usual: $\sigma_S^2 = E[x^2] - E[x]^2$ and $\sigma_I^2 = E[y^2] - E[y]^2$.

3. Application of this model to study a realistic case

Before applying our stochastic formulation, it is good to mention that although the behavior of the groups S and R can be predicted by using a deterministic model, in contrast with infected group, its quantification for the whole population is very difficult, and, in consequence, there is not enough field data. The infected group is considered as the variable with more impact and as such, it is therefore the most commonly used in epidemiological predictions. For these reasons, we focus our discussion mainly in infected cases I.

In what follows, we apply the mathematical model developed in Sec. 2.1 and summarized by Eqs. (23)-(29) to analize the influenza AH1N1 spread in Mexico in 2009. We use real data reported by Mexican Institute for Social Security (IMSS) in that year. In Table I it is presented the total number of infected people, this information includes 35 geographical regions stipulated by the [14]. We use the following methodology: even though it is well known that one epidemiological year encompasses a whole period of 52 weeks, of which we only consider those weeks in which there was a significant number of infected cases N_0. In regard to the proportion of the susceptible members $S(t)$, in the relation $N(t) = S(t) + I(t) + R(t)$, it is taken those individuals who finally became infected, which implies the initial condition $I(t) = S(0)$. In addition, to estimate the order of magnitude of the parameters β and γ we apply the Levenberg-Marquardt algorithm, whereas the value $\mathbb{R}_0$ it is computed directly from the ratio β/γ. In order to complete the problem formulation, we have also imposed the initial condition $S(0) = 1$, which implies that final epidemic size for the deterministic SIR epidemic model can be computed from the implicit relation

$$S(\infty) = \frac{N\gamma}{\beta} \ln\left(\frac{S(\infty)}{N-1}\right) + N. \tag{31}$$

Hence, the total number infected during the outbreak is $I = N - S(\infty)$ and the above expression can be expressed as

$$\frac{N\gamma}{\beta} \ln\left(\frac{N-I}{N-1}\right) + I, \tag{32}$$

this implicit relation is used to calculate the number of susceptible people, by using the Generalized Reduced Gradient

TABLE I. Total number of cases and parameters estimated by geographical region [14].

Geographical region	N	Infected cases	β	γ	$\mathbb{R}_0$
Aguascalientes	20,910	825	6.69	6.04	1.11
Baja California	43,918	1,050	9.98	9.64	1.03
Baja California Sur	11,289	565	5.82	5.13	1.13
Campeche	4,703	151	7.14	6.62	1.08
Coahuila	26,988	516	7.78	7.32	1.06
Colima	20,346	924	5.19	4.77	1.09
Chiapas	22,694	1,311	8.29	7.77	1.07
Chihuahua	21,309	458	9.56	8.85	1.08
Durango	17,154	752	5.00	4.43	1.13
Guanajuato	30,909	557	8.43	7.98	1.06
Guerrero	15,536	484	11.16	10.25	1.09
Hidalgo	15,193	758	6.58	5.71	1.15
Jalisco	91,003	1,576	8.52	8.12	1.05
Mexico Oriente	66,957	1,677	10.67	10.17	1.05
Mexico Poniente	40,153	1,245	7.21	6.55	1.10
Michoacan	22,708	607	6.76	6.25	1.08
Morelos	6,482	187	6.59	6.04	1.09
Nayarit	15,605	569	4.74	4.38	1.08
Nuevo Leon	69,512	1,571	10.49	10.01	1.05
Oaxaca	30,087	1,484	6.84	6.52	1.05
Puebla	24,470	685	8.54	7.79	1.10
Queretaro	30,237	834	8.91	8.59	1.04
Quintana Roo	6,365	203	17.25	15.43	1.12
San Luis Potosi	43,847	1,989	5.87	5.31	1.11
Sinaloa	17,395	441	7.82	7.14	1.09
Sonora	21,681	777	9.76	8.55	1.14
Tabasco	13,616	1,020	10.75	8.74	1.23
Tamaulipas	43,708	1,672	11.12	9.94	1.12
Tlaxcala	9,151	491	5.97	5.17	1.16
Veracruz Norte	29,404	749	10.54	9.82	1.07
Veracruz Sur	32,401	870	11.39	10.62	1.07
Yucatan	26,875	1,187	8.12	7.49	1.09
Zacatecas	9,915	385	8.85	7.94	1.11
D.F. Norte	60,936	2,063	11.18	10.68	1.05
D.F. Sur	52,399	1,253	8.09	7.47	1.08
Whole country	1,441,331	31,887	13.09	12.77	1.02

(GRG) nonlinear algorithm. Also, for the other expected variables we assume the following initial conditions: $E[x = 0] = E[y = 0] = E[x^2] = E[y^2] = E[xy] = 0$. Under these conditions, we have solved numerically the set of equations given by Eqs.(23), (24), (27), (28), and (29) by using a standard Runge-Kutta algorithm (see [15]).

4. Results and discussion

In Table I, we present the corresponding values of the typical parameters β, γ, and $\mathbb{R}_0$ that we have estimated. The main results for parameter $\mathbb{R}_0$ are as follow: we have estimated the parameter magnitude $\mathbb{R}_0 = 1.05$ and $\mathbb{R}_0 = 1.08$ for the
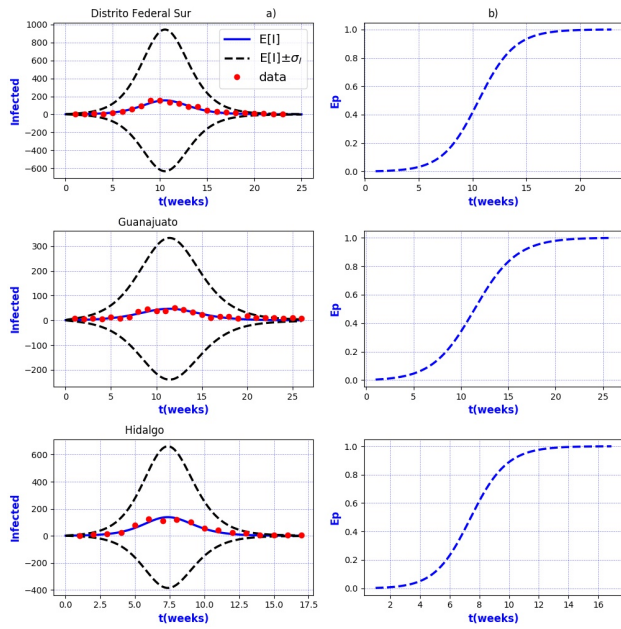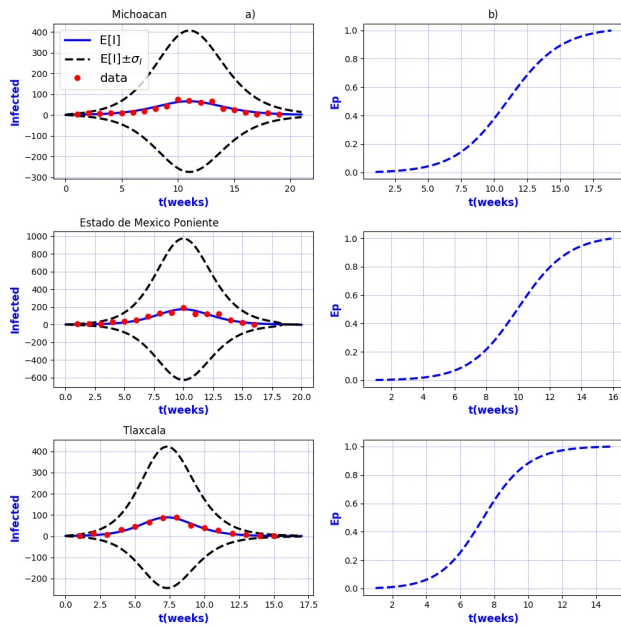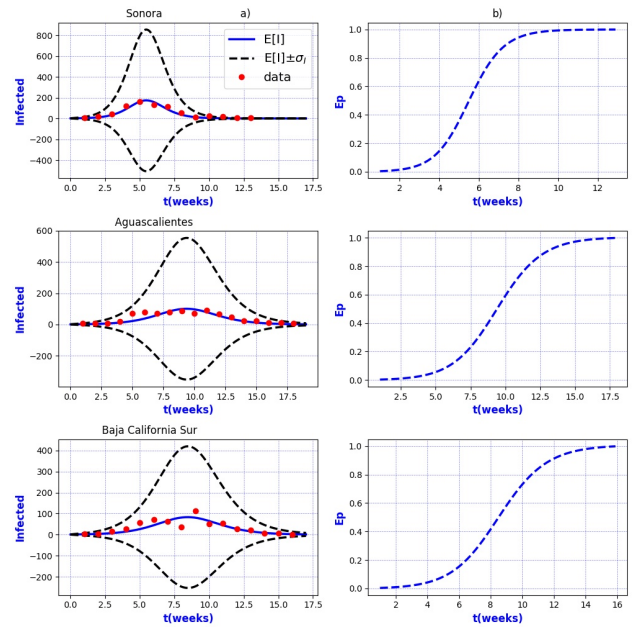
FIGURE 3. a) Probability distribution theoretical and observed. b) Ep=Extinction probability. Here $\psi = E[I]$ is the expected value of the theoretical distribution and σ_I is the standard deviation around the expected value.

FIGURE 5. a) Probability distribution theoretical and observed. b) Ep=Extinction probability. Here $\psi = E[I]$ is the expected value of the theoretical distribution and σ_I is the standard deviation around the expected value.
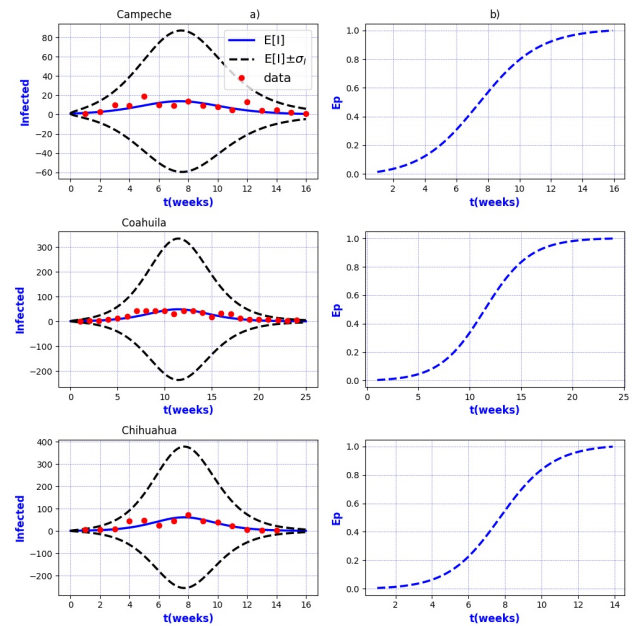

FIGURE 4. a) Probability distribution theoretical and observed. b) Ep=Extinction probability. Here $\psi = E[I]$ is the expected value of the theoretical distribution and σ_I is the standard deviation around the expected value.

FIGURE 6. a) Probability distribution theoretical and observed. b) Ep=Extinction probability. Here $\psi = E[I]$ is the expected value of the theoretical distribution and σ_I is the standard deviation around the expected value.

geographical region Distrito Federal Norte and Distrito Federal Sur respectively, whereas a value $\mathbb{R}_0 = 1.72$ has been estimated by [11] for the same geographical region so called Distrito Federal (include both regions Distrito Federal Norte and Distrito Federal Sur). On the other hand, for the whole country, we have estimated the value $\mathbb{R}_0 = 1.02$, while Fraser *et al.* estimated this value within the interval $(1.4, 1.6)$; thus,

in a similar analysis done for members belonging IMSS, [12] estimated season values to $\mathbb{R}_0$ between $(1.8 - 2.1)$ during Spring, of $(1.6 - 1.9)$ for Summer and $(1.2 - 1.3)$ for Fall. We must notice that the order of magnitude of the parameter $\mathbb{R}_0$ for all the geographical regions is very close to the value that has been estimated by those authors, who have used different techniques than those we have discussed here.
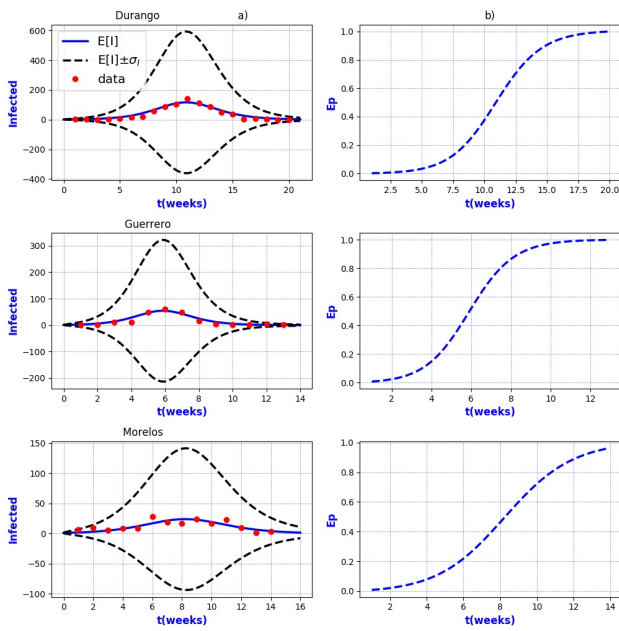
FIGURE 7. a) Probability distribution theoretical and observed. b) Ep=Extinction probability. Here $\psi = E[I]$ is the expected value of the theoretical distribution and σ_I is the standard deviation around the expected value.
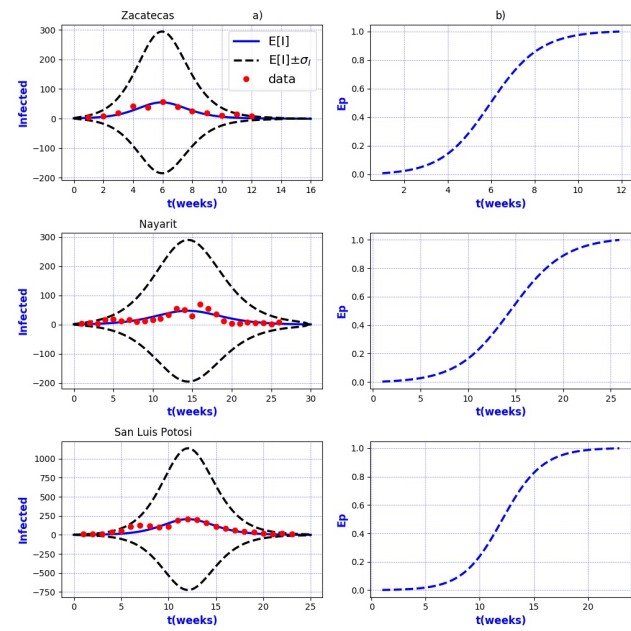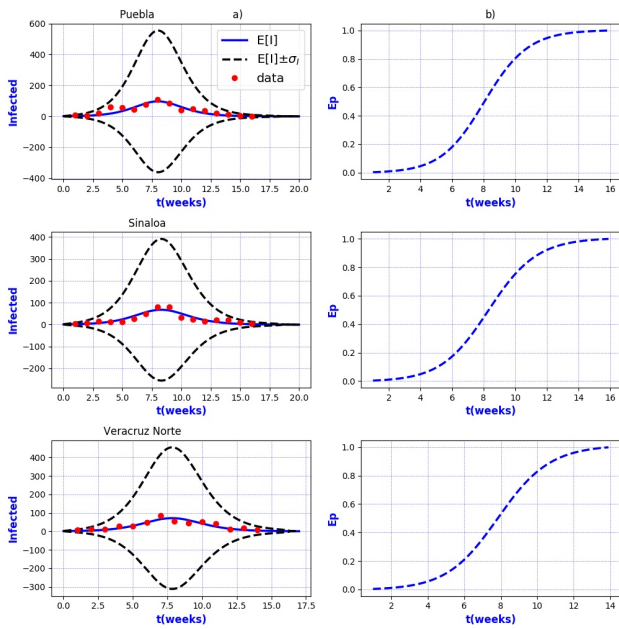
FIGURE 9. a) Probability distribution theoretical and observed. b) Ep=Extinction probability. Here $\psi = E[I]$ is the expected value of the theoretical distribution and σ_I is the standard deviation around the expected value.

FIGURE 8. a) Probability distribution theoretical and observed. b) Ep=Extinction probability. Here $\psi = E[I]$ is the expected value of the theoretical distribution and σ_I is the standard deviation around the expected value.
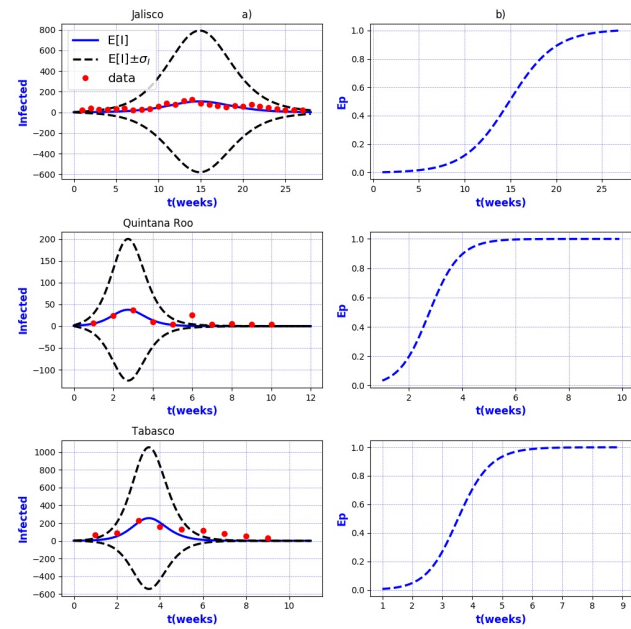
FIGURE 10. a) Probability distribution theoretical and observed. b) Ep=Extinction probability. Here $\psi = E[I]$ is the expected value of the theoretical distribution and σ_I is the standard deviation around the expected value.

Figures 3-8 exhibit all the events that we have analyzed. The corresponding comparison between real infected cases I and its respective theoretical average $E[I]$ has been presented for each geographical region. In the same Figs. 3-8, also is illustrated the corresponding associated curves with the extinction probability of the epidemic as a function of the time.

The results show that the best fit correspond to geographical regions Distrito Federal Sur, Guanajuato, Hidalgo, Michoacan, Estado de Mexico Poniente, Tlaxcala and Sonora, because real data are accumulated around the expectation value $E[I]$. Moreover, the geographical regions Aguascalientes, Baja California Sur, Campeche, Coahuila, Chihua-
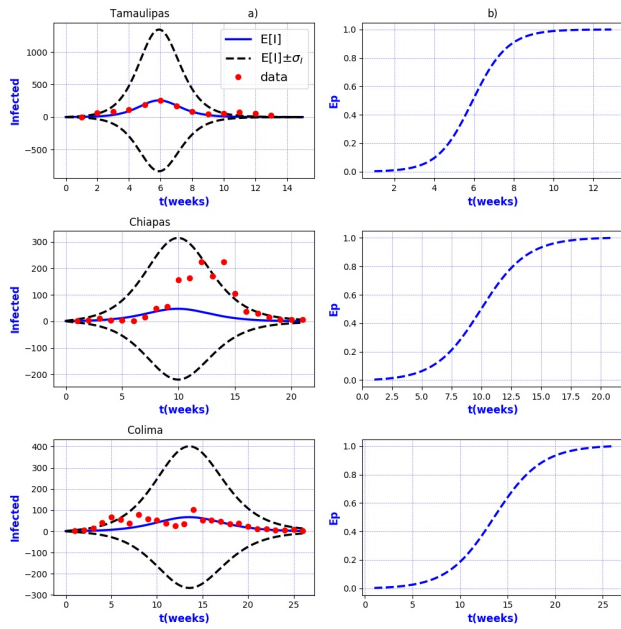
FIGURE 11. a) Probability distribution theoretical and observed. b) Ep=Extinction probability. Here $\psi = E[I]$ is the expected value of the theoretical distribution and σ_I is the standard deviation around the expected value.
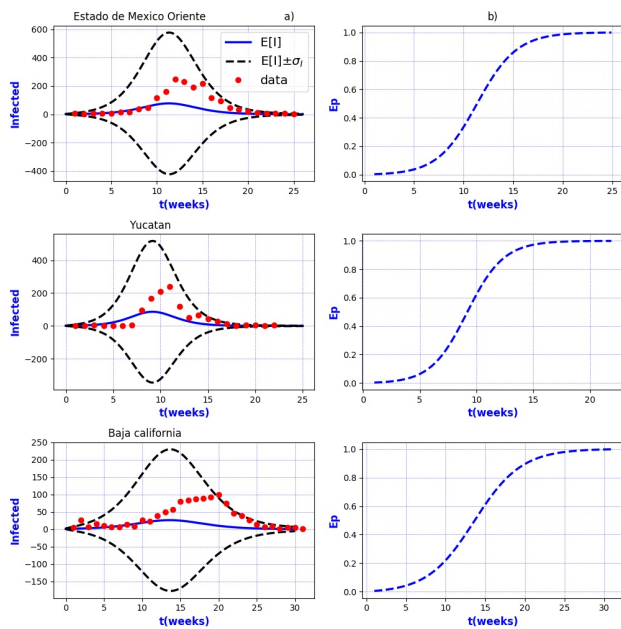

FIGURE 13. a) Probability distribution theoretical and observed. b) Ep=Extinction probability. Here $\psi = E[I]$ is the expected value of the theoretical distribution and σ_I is the standard deviation around the expected value.

FIGURE 12. a) Probability distribution theoretical and observed. b) Ep=Extinction probability. Here $\psi = E[I]$ is the expected value of the theoretical distribution and σ_I is the standard deviation around the expected value.

FIGURE 14. a) Probability distribution theoretical and observed. b) Ep=Extinction probability. Here $\psi = E[I]$ is the expected value of the theoretical distribution and σ_I is the standard deviation around the expected value

hua, Durango, Guerrero, Morelos, Puebla, Sinaloa, Veracruz Norte, Zacatecas, Nayarit, and San Luis Potosi, also exhibit a good fit with real data more scattered around its theoretical average. However all of them are still within the calculated probabilistic bands given by $E[I] \pm \sigma_I$. Similar behavior around the expected value $E[I]$ is observed in the regions
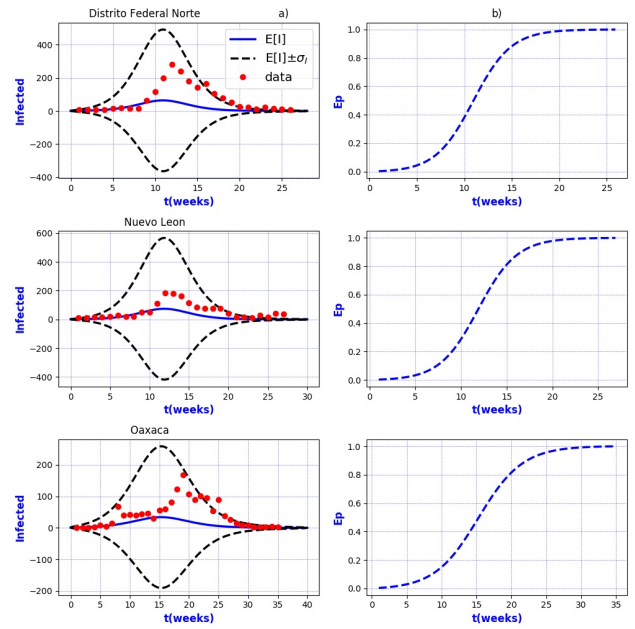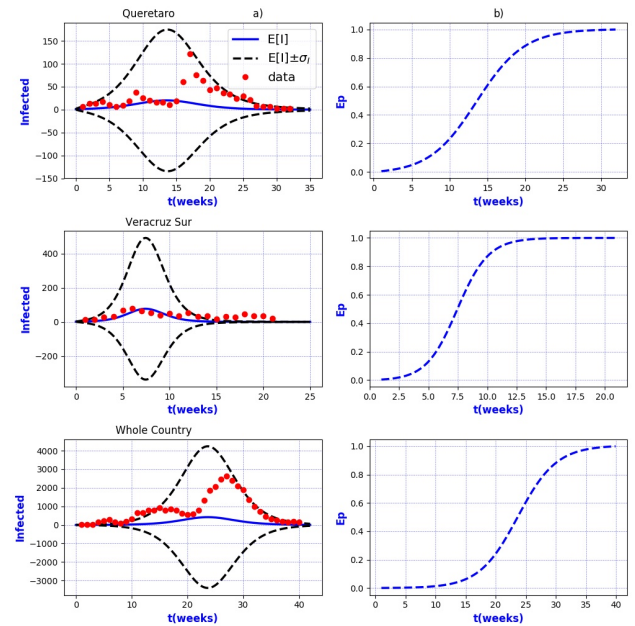
Jalisco, Quintana Roo, Tabasco, and Tamaulipas, but some real cases are slightly out of the bands $E[I] \pm \sigma_I$.

In contrast, the results show that the largest spread of real cases around the theoretical average $E[I]$ correspond to geographical regions Chiapas, Colima, Estado de Mexico Oriente, Yucatan, Baja California, Distrito Federal Norte, Nuevo Leon, Oaxaca, Queretaro, and Veracruz Sur as well as for the

whole country. It can seen that in some of these regions and in the whole country, a considerable number of real cases are outside of the theoretical standard deviation bands $E[I] \pm \sigma_I$.

On the other hand, it is also important to point out that in the geographical regions Puebla, Sinaloa, Colima, Jalisco, Estado de Mexico Oriente, Nayarit, San Luis Potosi, Yucatan, Distrito Federal Norte, Nuevo Leon, Oaxaca, Queretaro, and Veracruz Sur an oscillatory trend is observed particularly notorious at the level of the entire country. The origin of these discrepancies probably, is owing to two main reasons: 1) the simplicity of the model that we have utilized and / or 2) inaccuracies in the acquisition of field data. A possible solution in order to have a better fit, could be the implementation of a more realistic model that takes into account infected death, vaccination, births, population migration, incubation period of the virus, among other factors which we have neglected. Moreover, the oscillatory behavior observed, characteristic in large populations, suggests the necessity of introducing seasonal forcing into the model. In parallel, suitably trained staff would represent a direct improvement on the quality and reliability of the data.

In regard to extinction probability as a function of time, we have identified the week for which there is a 90% of extinction probability of the epidemic -a threshold that has been chosen arbitrarily and could easily be changed-. Under this criteria, the average extinction time for each geographical region as well as for whole country was around 13 and 30 weeks respectively. Moreover, geographical regions in which it is observed a longer period of extinction corresponding to 90% are, Colima, Nayarit, and Queretaro with 19 weeks, Jalisco and Baja California with 20 weeks and Oaxaca with 21 weeks. Complementary, shorter periods were found for geographical regions Quintana Roo and Tabasco with 4 weeks.

As already mentioned, these estimates values could be improved with the implementation of a more realistic model and a more reliable registration of infected cases. However, a lot of the neglected processes and unrevealed variables can be incorporated to some extent, within the standard deviation band that we have computed. This is indeed the power of our stochastic approach because even though there exist various additional variables, not considered here, which modify our epidemic evolution, we can embody all of them within the statistical distribution of the parameters of a quite simple dynamic model which describes the dominant behavior of the phenomenon.

5. Concluding remarks

We have established a stochastic model based on a simple compartmental SIR formalism. Using this approach we have studied the propagation of the AH1N1 influenza that took place in Mexico in 2009. We have restricted this study to

fected cases reported by Mexican Institute for Social Security (IMSS) in that year.

Our main conclusions are the following: the order of magnitude estimated for the basic reproductive number of infection $\mathbb{R}_0$ is within the interval $(1.03 - 1.23)$ which is shorther than those previously has been reported in the literature. The model that we have developed is quite general and therefore is not only applicable to influenza viruses, but it can also be applied to any epidemic in which the following conditions are fulfilled: the infectious agent is a virus, there are not demographic dynamics, the population is homogeneous, the seasonal period is negligible, removed individuals may not be susceptible again, the incubation time of the virus is not taken into account, any two individuals have the same probability of contacting each other, and the duration of the epidemic is short compared with the life expectancy of the individuals. This type of study, applied to real cases, the first of its type in Mexico, as far as we know. In spite of the simplicity of the model that we have implemented, the results show clearly that the probability distribution predicted by the stochastic method for the infected individuals is in good agreement with the distribution obtained from the dataset provided by IMSS. This is very important because provides a solid confidence in our formalism for working in subsequent developments. Thus, the results of this research may have immediate practical application because they constitute a tool for taking decisions and for the optimization of resources against possible epidemic scenarios. The small discrepancies between the forecast and observed data in some geographical regions, might have their origin in: 1) the simplicity of the model implemented; in this case, a more complete formulation is needed in order to provide a better description with respect to the real data. For instance, the implementation of a model with seasonal forcing would be suitable for those regionswhich do not evidently exhibit a Gaussian distribution around the average and 2) data reliability. It is advisable to implement strategies that contribute to train staff effectively; hence, achieving a better quality on the information gathered.

As future work, it will be interesting to formally extend the present study by including the main health agencies that have a presence throughout the country, such as the Mexican Secretariat of Health (SSA), Institute for Social Security and Services for State Workers (ISSSTE) and Navy Secretariat (SEMAR). The data set coming from these institutions would complement the information provided by the Mexican Institute for Social Security (IMSS) for a nationwide study. We can even consider other kind of epidemics, as well as an extension to our model by means of incorporating important variables such as: vaccination, demographic dynamics (emigration, immigration, births, and deaths), information on passive immunity, age and sex of the population, quarantine period, incubation period of the infectious agent, loss of immunity, forcing seasonal, among other factors. It would also be interesting to consider an interaction between individuals beyond their immediate neighbors, that is, second and third

neighbors. Nevertheless, we should be aware of including only a few additional important variables in order to keep a very simple and easy-to-use description.

---

1. F. Brauer, P. van den Driessche, and J. Wu, *Mathematical Epidemiology* (Springer- Verlag, Berlin, 2008), `https://doi.org/10.1007/978-3-540-78911-6`.

2. MMWR Morb Mortal Wkly Rep 58. *Swine influenza A (H1N1) infection in two children-Southern California*, March-April, (2009) 400-402.

3. Centers for Disease Control and Prevention, *Swine influenza A(H1N1) infection in two children-Southern California, March-April 2009*, MMWR Morb Mortal Wkly Rep. No. 58(12) 400-2, 2009.

4. W. O. Kermack, and A. G. McKendrick, A. contributions to the mathematical theory of epidemics, *Proc. R. Soc. Lond.* **115** (1927) 700. `http://doi.org/10.1098/rspa.1927.0118`.

5. I. Nasell, On the time to extinction in recurrent epidemics, *J. Roy. Stat. Soc.* **61** (1999) 309. `https://doi.org/10.1111/1467-9868.00178`.

6. A. J. Black, and A. J. McKane, Stochasticity in staged models of epidemics: quantifying the dynamics of whooping cough, *J. R. Soc. Interface* **7** (2010) 1219. `http://doi.org/10.1098/rsif.2009.0514`.

7. G. Rozhnova, and A. Nunes, Stochastic effects in a seasonally forced epidemic model, *Phys. Rev. E,* **82** (2010) 041906. `https://doi.org/10.1103/PhysRevE.82.041906`.

8. R. H. Wang, Z. Jin, Q. X. Liu, J. van de Koppel, and D. Alonso, A Simple Stochastic Model with Environmental Transmission Explains Multi-Year Periodicity in Outbreaks of Avian Flu, *PLOS ONE* **7** (2011) e28873, `https://doi.org/10.1371/journal.pone.0028873`.

9. D. Alonso, A. J. McKane, and M. Pascual, Stochastic amplification in epidemics, *J. R. Soc. Interface,* **4** (2007) 575. `http://doi.org/10.1098/rsif.2006.0192`.

10. M. A. Capistran, J. Andrés, and J. X. Velasco-Hernandez, Towards uncertainty quantification and inference in the stochastic SIR epidemic model, *Math. Biosci.* **240** (2012) 250. `https://doi.org/10.1016/j.mbs.2012.08.005`.

11. G. Cruz-Pacheco *et al.,* Modelling of the influenza A(H1N1)v outbreak in Mexico City, April-May 2009, with control sanitary measures, *Eurosurveillance* **14** (2009) 19254.

12. G. Chowell, S. Echevarría-Zuno, C. Viboud, L. Simonsen, and J. Tamerius, Characterizing the Epidemiology of the 2009 Influenza A/H1N1 Pandemic in Mexico, *PLOS Med.* **8** (2011) e1000436. `https://doi.org/10.1371/journal.pmed.1000436`.

13. N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry.* (Elsevier, Netherlands, 2007),. `https://doi.org/10.1016/B978-0-444-52965-7.X5000-4`.

14. Sistema de Información en Línea para la Vigilancia Epidemiológica de la Influenza (SINOLAVE), IMSS Data base, January-December 2009.

15. J. Mathews and F. Kurtis, *Numerical Methods Using MATLAB*, 3rd ed. (Prentice Hall, New Jersey, 2000).