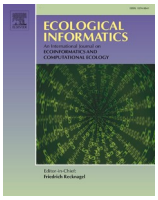
Ecological Informatics 77 (2023) 102266

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf



Application of a stochastic compartmental model to approach the spread of environmental events with climatic bias☆

Joan Boters Pitarch [a], María Teresa Signes Pont [a,*], Julian Szymański [b], Higinio Mora Mora [a]

[a] Department of Computer Science and Technology, University of Alicante, Alicante, Spain
[b] Faculty of ETI, Gdańsk University of Technology, Gdańsk, Poland

ARTICLE INFO

Keywords:
Epidemiological models
Climatic variables
SIR paradigm
Monte Carlo method

ABSTRACT

Wildfires have significant impacts on both environment and economy, so understanding their behaviour is crucial for the planning and allocation of firefighting resources. Since forest fire management is of great concern, there has been an increasing demand for computationally efficient and accurate prediction models. In order to address this challenge, this work proposes applying a parameterised stochastic model to study the propagation of environmental events, focusing on the bias introduced by climatic variables such as wind. This model's propagation occurs in a grid where cells are classified into different compartments based on their state. Furthermore, this approach generalises previous non-stochastic models, which are now considered particular cases within this broader framework. The use of the Monte Carlo method is highlighted, which allows for obtaining probabilistic estimates of the state of the cells in each time step, considering a level of confidence. In this way, the model provides a tool to obtain a quantitative estimate of the probability associated with each state in the spread of forest fires.

1. Introduction

Natural phenomena are complex systems characterised by the interaction of multiple variables with interconnected dependencies. Consequently, modelling them can be challenging due to the involvement of various factors, such as environmental conditions, physical forces, biological interactions, and stochastic processes. Moreover, these factors often interact with each other in a non-linear manner, making it difficult to establish simple cause-effect relationships.

Another significant challenge is the inherent uncertainty associated with natural phenomena. Propagation and quantification of uncertainty play a critical role when dealing with such systems since some variables are inherently uncertain and subject to random fluctuations or even the same conditions could lead to different outcomes. More, while models aim to depict the intricacy of natural occurrences, they often present a simplified version of reality. They rely on assumptions and approximations to strike a balance between accuracy and computational feasibility. In this regard, the lack of high-quality information often hinders model validation and, consequently, the testing of hypotheses.

This paper focuses on forest fires as a natural phenomenon under study. In this background, a wide range of fire spread models have been developed that consider physical aspects related to the combustion process (Asensio and Ferragut, 2002; Mandel et al., 2008; Muzy et al., 2008). These models often use mathematical tools, such as partial differential equations and ordinary differential equations, to describe the spread process. However, the parameters of these equations do not directly correspond to the aspects identified as the most influential in the fire spread such as the spatial relationships (Ganteaume et al., 2013), slope or wind (Pitts, 1991; Viegas, 2004).

For this reason, to address spatial relationships more precisely, agent-based models (ABM) emerge as a viable solution to overcome such limitations, as they take into account the dynamic interaction and unique behaviour of multiple agents immersed in a spatial environment, allowing us to discover complex patterns. When it comes to forest fires, cellular automata (CA) and graph-based models can be distinguished. Regarding CA models, it is common to consider the region where the fire occurs as a two-dimensional square grid (2D-CA models), where each element represents a land parcel and the interaction is determined by local rules enhanced by climatic variables (Karafyllidis and Thanailakis, 1997; Hernández Encinas et al., 2007; Jiang et al., 2021). However,

☆ This document is the results of the research project AICO/2021/331 funded by Generalitat Valenciana.
* Corresponding author.
E-mail address: teresa@dtic.ua.es (M.T. Signes Pont).

https://doi.org/10.1016/j.ecoinf.2023.102266
Received 30 March 2023; Received in revised form 14 August 2023; Accepted 14 August 2023
Available online 22 August 2023

depending on the context, the use of hexagonal grids (Hernández Encinas et al., 2007) could be more suitable for mitigating the limitations of square grids as representation elements. At the same time, network-based modelling (Hajian et al., 2016) is also used where the landscape is represented as a network of nodes, and the fire spread time is modelled as the shortest path problem.

Simulators and Geographic Information Systems (GIS) models play a key role in forest fire prevention and prediction. Forest fire simulators use mathematical models and geospatial data to realistically simulate how a fire would spread under certain conditions. Importantly, GIS models are not strictly ABS, as they do not imply autonomy, autonomous behaviour, and decision-making of individual agents, although they are computational systems that integrate geographic data and spatial analysis tools to represent and manipulate geographic information. Well-known examples include FARSITE (Farsite, 2004), WFDS (Noonan-Wright et al., 2011), and FBP (Andrews, 2007). Moreover, worth highlighting studies (Tuyen et al., 2021; Bui et al., 2018) include machine learning tools to understand and represent this kind of natural phenomena.

Currently, we propose a new and more realistic model (Boters Pitarch et al., 2023), which is based on previous work developed in Signes Ponnt et al. (2017), Signes-Pont et al. (2021), Signes Pont et al. (2019). We hypothesize that the phenomenon spreads in the same direction θ and the intensity ρ of the wind to reach longer distances. Therefore, we use these two variables, θ and ρ, as expansion vectors for the model, which follows the scheme based on the Susceptible-Infected-Recovered (SIR) model. The improvements of our proposal are both the innovative approach to modelling, with a flexible and probabilistic neighbourhood relationship between cells that adapts to weather conditions, and an update rule that leverages prior probabilities through the Bernoulli random variable to work with uncertainty. All this provides a highly effective stochastic model. This combination of both spatial flexibility and stochasticity makes our model a powerful tool for predicting real-world phenomena such as wildfire spreads or pest plants propagation, since we take into account the most important variables in the representation.

The outline of this research is as follows: after Introduction, Section 2 recalls the main hypotheses of the model. Section 3 realizes the sensibility analysis in order to know the impact of the different parameters on the global studied expansion. Section 4 demonstrates how our model reduces the Moore and Von Neumann dynamics to particular cases. Section 5 uses our model as the Monte Carlo method to obtain an estimate of the probability of being susceptible (infected, recovered, etc…) for every single cell, according to a given confidence level. Finally, Section 6 presents concluding remarks and proposes future topics for this research.

2. Key elements of the model

The main idea of our model is to consider both the spatial features of the system and the environmental variables. The expansion of the phenomenon occurs on a square grid with size N, where each cell can have one of three possible states: susceptible (healthy), infected or dead (fired), following the well-known SIR scheme and expansion only occurs through contact between infected and healthy cells in a 2D-CA model. To determine the cells that may become infected a neighbourhood relationship takes into account environmental variables θ and ρ and an update criterion uses this probability to determine the new state of each cell. The variables θ and ρ are considered the expansion vectors (based on polar coordinates) that represent the orientation of the expansion and the power respectively, to determine the scope or even the number of exposed cells.

Definition 2.1. An exposed cell is a susceptible cell that has a non-zero probability of infection.

It is important to point out that each exposed cell is **updated independently** by performing a single Bernoulli test, see Eq. (3). In case of success, the cell is infected, while in case of failure, it remains healthy. Fig. 1 shows the workflow and how given a direction θ we may get different outcomes.

2.1. Technical aspects

Now, we formalize the main elements above. Let us consider both the flexible neighbourhood relationship R^k that depends on θ and ρ and the update criterion that determines the state of each cell. A more detailed explanation is available in Section 3 of Boters Pitarch et al. (2023). The neighbourhood relation is depicted by function R^k, see Eq. (3)

$$\begin{array}{rcccl} R^k & : & [0,1]\times[0,2\pi[& \mapsto & \mathbb{R}^{N\times N} \\ & & (\rho,\theta) & \mapsto & \left(r_{ij}^k\right)_{i,j=0}^{N-1} \end{array} \tag{1}$$

Where r_{ij}^k stands for the probability that cell (i,j) is infected at generation k.

Remark 2.1. Our approach follows a dynamic modelling scheme, however, time is divided into discrete units called generations. For instance, the initial state is generation 0, the first step is generation 1 and so on.

Construction is as follows:

1.For generation k, the new infected cell will be determined by the location of the infected ones at generation $k-1$ and their possible contact will occur due to the action of the wind. Therefore, we define:

$$I_{k-1} := \left\{\left(i,j\right) : m_{ij}^{k-1} = 1\right\} \tag{2}$$

as the set of index pairs defining the infected cells at generation $k-1$.
2.We use the θ direction of the wind, together with basic trigonometric functions (similar to polar coordinates) to determine which cells will be exposed and their infection probabilities. See Fig. 1.
3.We set a partition $\mathscr{P} = \{0, u_1, u_2, u_3, 1\}$of the interval $[0,1]$, which aims to establish the infection capacity for further cells. Thus, it allows us to understand the number of potentially affected cells, i.e., the scope. In this first approach, we assume a limited infection capacity and therefore the partitions are of this size.
4.Finally, by means of the above procedure, we will get a square matrix $P^{(i,j)}$ with size N for every $(i,j) \in I_{k-1}$. Then, $R^k(\rho,\theta)$ will be the aggregate of all previous matrices, i.e.

$$R^k\left(\rho,\theta\right) = \sum_{(i,j)\in I_{k-1}} P^{(i,j)}$$

Remark 2.2. The neighbourhood relationship construction can result in r_{ij}^k values that exceed unity. In such cases, any susceptible cell with a probability of infection higher than one will be considered infected in the next generation. This also occurs in other works, see Karafyllidis and Thanailakis (1997), Hernández Encinas et al. (2007).

Regarding the grid update criterion, we perform the following approach outlined in Eq. (3), where m_{ij}^k refers to the state (0-susceptible, 1-infected, 2-dead) of the cell (i,j) at generation k.

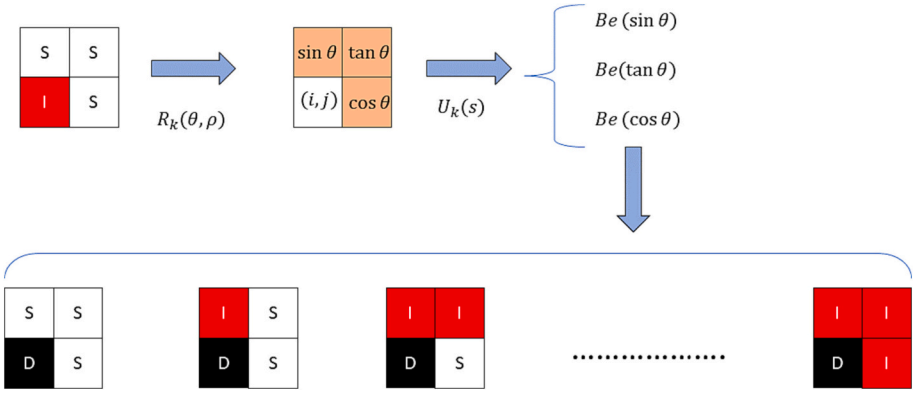


Fig. 1. Model flow overview per generation when $u_1 \leqslant \rho \leqslant u_2$.

$$
\begin{aligned}
&\forall\, m_{ij}^k = 2 \Rightarrow m_{ij}^{k+1} = 2 \\
&\forall\, 1 \leqslant m_{ij}^k < 2 \Rightarrow m_{ij}^{k+1} = m_{ij}^k + \frac{1}{\Delta_{ID}} \\
&\forall\, m_{ij}^k = 0 \Rightarrow m_{ij}^{k+1} := \begin{cases} Be\left(r_{ij}^k\right) & if \quad 0 \leqslant r_{ij}^k < 1 \\ 1 & if \quad r_{ij}^k \geqslant 1 \end{cases}
\end{aligned} \tag{3}
$$

Thus, using Eq. (3), we can conclude that an infected cell remains as infected during Δ_{ID} generations. In addition, dead elements are not recovered. Furthermore, as can be seen in Property 3.1 of Boters Pitarch et al. (2023), we can theoretically compute the number E of infections expected by the model, given its parameters. This measure can be of great help in our study to understand their impact and for a future search according to a real case.

2.2. Limitations of the model

When assessing the suitability of our model for representing different issues, it is crucial to take into account its limitations. These limitations are intricately related to both the 2D-CA models and the neighbourhood relationship R^k.

Our model is well-suited for scenarios influenced by wind and limited terrain, such as forest fires, pest plant spread, or object tracking by ocean currents. It could accurately represent the phenomenon's spread and expansion in such cases. It should be noted that, due to the 2D-CA approach, our model performs better the flatter the terrain. However, a careful selection of the parameters may overcome this drawback.

On the other hand, the simplicity of the neighbourhood relationship in our model may not adequately capture the complexity of certain problems. While this simplicity facilitates the interpretation of the model, it also limits its ability to address phenomena involving more complex interactions between individual agents or even non-local neighbourhoods. For example, problems such as population dynamics influenced by traffic networks or complex population growth may require alternative ABM modelling schemes for more realistic representations, as seen in previous work (Gómez et al., 2010; Huang et al., 2019).

3. Study of the parameters

Compared with Karafyllidis and Thanailakis (1997), Hernández Encinas et al. (2007) which use similar schemes, one of the main advantages of our model, is the existence of parameters that allow modulating and tuning the expansion rate. The aim of this section is to improve our understanding of the interdependencies between parameters and justify why measure E of expected infections (see Property 3.1 of Boters Pitarch et al. (2023)) serves as a leading indicator of the spread rate. Table 1 presents a summary of the role played by each parameter.

Table 1
Description of the role of the main parameters.

	Role
Δ_{ID}	delay between infected and dead state
$\mathscr{P}$	partition which determines the influence of the wind intensity
C	loss factor for far cells
p_0	infection probability for surrounding cells, when the wind has no effect

Regarding the interdependence of the variables, it should be noted that the delay parameter Δ_{ID} is independent of the others, as its function is solely focused on determining the number of generations between the infected and dead states. However, partitions $\mathscr{P} = \{0, u_1, u_2, u_3, 1\}$ are key parameters as they define the extent of infection for each generation. For instance, when $\rho > u_2$, the loss factor C impacts the neighbourhood relationship and consequently, the propagation rate. Similarly, if $\rho < u_1$, the parameter p_0 influences the situation; otherwise, it does not. Therefore, we can assert that there is a strong interdependence between the partitions and C and p_0. Simultaneously, it can be deduced that there is no possible dependence between C and p_0 since they come into play when intensity takes different values. Tables 2 and 3 show the interdependence between variables, showing how the marginal infection capacity E, varies at a greater or lesser rate depending on the features of the partition.

Thus, Table 2 shows that for any partition, when C increases, the value of E decreases, i.e. the probability of infection for distant cells decreases and therefore the expected number of cells decreases as well. However, in Table 3 we see that if p_0 increases, so does the value of E, since in the absence of wind, the value of p_0 indicates the probability of infection of the cells. Also, in both tables, we can see how the impact of these parameters depends on the partitions.

As an example, for $\mathscr{P}_1$ we have a very small u_1, which implies that the absence of wind is somewhat unusual, i.e. a limited impact of p_0 as shown in Table 3. In contrast, the value $(1 - u_2)$ is relatively large. This means that the infection reaches distant cells more easily, and therefore a large impact of the parameter C can be expected, as shown in Table 2, for column $\mathscr{P}_1$.

Another important aspect to address is the range of values these parameters can take. Regarding the partitions, the values should be determined based on the wind intensity which depends on the environmental phenomenon and its location. Sometimes, there is not enough knowledge of this influence, so we consider it an adjustable parameter. For C and p_0, we can use the specific information of the phenomenon. As an example, it is well-known that high temperatures and low humidity favour a quick expansion of forest fires (Ganteaume et al., 2013; Pitts, 1991; Viegas, 2004). This could lead to determine a low loss value C and a high p_0 value in such cases. Therefore, these values may vary even within the same phenomenon if the environmental conditions are significantly different. This is a potential aspect for improvement in our

Table 2
Marginal infectiousness according to partitions and loss factor.

E	$\mathscr{P}_1$	$\mathscr{P}_2$	$\mathscr{P}_3$	$\mathscr{P}_4$
$C = 1$	5.8376	5.0039	3.5165	4.1252
$C = 1.5$	4.0450	3.6188	2.8415	3.1359
$C = 2$	3.3059	3.0362	2.5355	2.7198
$C = 2.5$	2.9127	2.7219	2.3619	2.4953
$C = 3$	2.6716	2.5271	2.2504	2.3561

Table 3
Marginal infectiousness according to partitions and p_0.

E	$\mathscr{P}_1$	$\mathscr{P}_2$	$\mathscr{P}_3$	$\mathscr{P}_4$
$p_0 = 0.1$	3.2939	2.9762	2.4155	2.4198
$p_0 = 0.4$	3.3179	3.0962	2.6555	3.0198
$p_0 = 0.7$	3.3419	3.2162	2.8955	3.6198
$p_0 = 1$	3.3659	3.3362	3.1355	4.2198

model that we address in Section 6.

After discussing the interdependency and range parameters can take, our aim now is to justify why E is a reliable measure to anticipate the rate of expansion and therefore to establish a hierarchy and a direct search for the optimal parameters. When we want to conclude the impact of parameters on the **expected** rate of expansion, the study cannot be limited to specific data from a particular phenomenon, since there is a great variety of climatic conditions and potential wind-driven events. This could introduce a bias into our study so, the best way for conducting the aforementioned study is to simulate climatic data by means of

$$\theta \sim Unif(0, 2\pi) \text{ and } \rho \sim Unif(0, 1).$$

First of all, we set an arbitrary partition

$$\mathscr{P} := \{0, 0.1, 0.5, 0.9, 1\}$$

and check whether those parameter combinations with higher expected marginal infection are the ones that give rise to a higher expansion rate. In Fig. 2, you can see the surface generated by calculating E over its parameters C, p_0.

Thus, in Fig. 3, we can see how those parameter combinations with higher E in Tables 2 and 3 have a higher rate of expansion which leads us to think that there is an important connection between E and the spread rate.

However, this is not enough to complete the study. To validate this assumption, we calculate the **Pearson correlation** between E and the average spread rate d_k, given by

$$d_k = \frac{S_0 - S_k}{k}, \quad k \geqslant 1$$

where S_k refers to the number of susceptibles cells at generation k, see Fig. 4. This allows us to investigate whether a linear relationship exists between the two measures or not.

Remark 3.1. The parameters used to carry out the correlation study are

$$\mathscr{P}_1 = \{0, 0.01, 0.2, 0.5, 1\}, \mathscr{P}_2 = \{0, 0.05, 0.3, 0.65, 1\}$$

$$\mathscr{P}_3 = \{0, 0.1, 0.5, 0.9, 1\}, \ \mathscr{P}_4 = \{0, 0.25, 0.5, 0.75, 1\}$$

$C \in \{1, 2, 3\}, p_0 \in \{0.1, 0.4, 0.7, 1\}$ and $\Delta_{ID} \in \{1,2,3\}$. We consider them as a wide range of parameters which provide us with enough information to draw conclusions.

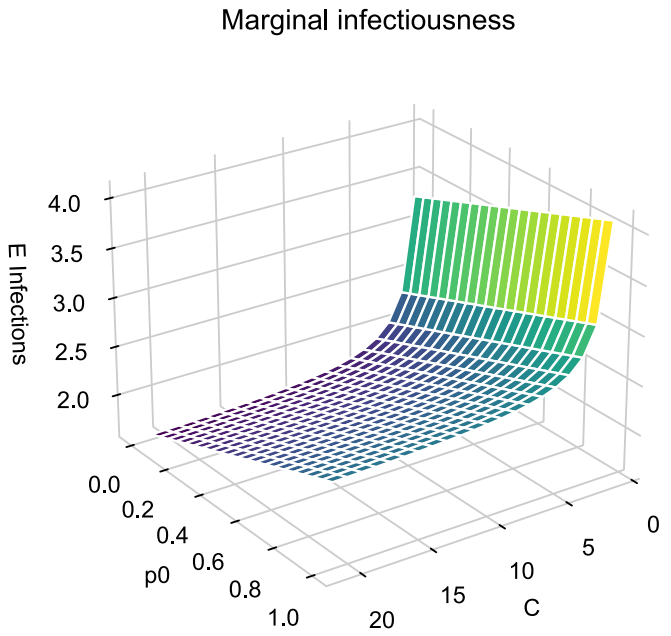


Fig. 2. The surface of marginal infectiousness E when $\mathscr{P} := \{0, 0.1, 0.5, 0.9, 1\}$.

One of the main weaknesses of the marginal infectiousness measure is that it assumes all cells around the infected cell are susceptible (can become infected). This is certainly true in the early stages of the spread but, as the spread progresses, the number of infected or dead cells grows fast. However, the Pearson correlation values are consistently high (above 0.9) for the first 10 generations for all $\Delta_{ID} \in \{1,2,3\}$. Therefore, we can state the presence of a **strong linear relationship** between the marginal infection and the spread rate that points to the marginal infection value can serve as an **effective indicator** of the expected spread pace, based on our model parameters.

4. Space of solution of our model

In this section, our aim is to study the solution space of our model. Obviously, this space is determined by the climatic conditions, since it depends on θ and ρ and on the chosen parameters.

First, we will focus on the well-known Moore and Von Neumann dynamics. These dynamics have been applied in a wide range of situations (Chen and Mynett, 2003; Chen and Horng, 2010; Roy et al., 2021). However, since our model is **parameterisable**, an interesting question is whether our model can meet the same dynamics, given particular parameters. In other words, are Moore's and Von Neumann dynamics included in the solution space of our model?.

Recall that in the performance of our model, the partitions are a critical parameter, since they determine when the wind direction θ and the intensity ρ have an influence. So when $\rho < u_1$, the wind is considered to have no influence since the intensity does not exceed a minimum threshold to guide the expansion of the phenomenon. For $u_1 = 1$, the interaction only occurs between the infected cell and the surrounding healthy cells (with a probability of infection p_0). Therefore, for this type of partition, we can affirm that the result of the model does not depend either on θ or ρ, but only on the parameter p_0, as shown in Fig. 5.

Property 4.1. Moore's Dynamics is contained within the solution space of our model.

Proof. Let us have a square grid of size N. It is enough to consider a partition $\mathscr{P}$ such that $u_1 = 1$, and $p_0 = 1$. By means of the partition, it is clear that the neighbourhood will always be that of our model shown in Fig. 5. In addition, given $p_0 = 1$, for each generation, the cells around the infected cells will have a probability of infection equal to or greater

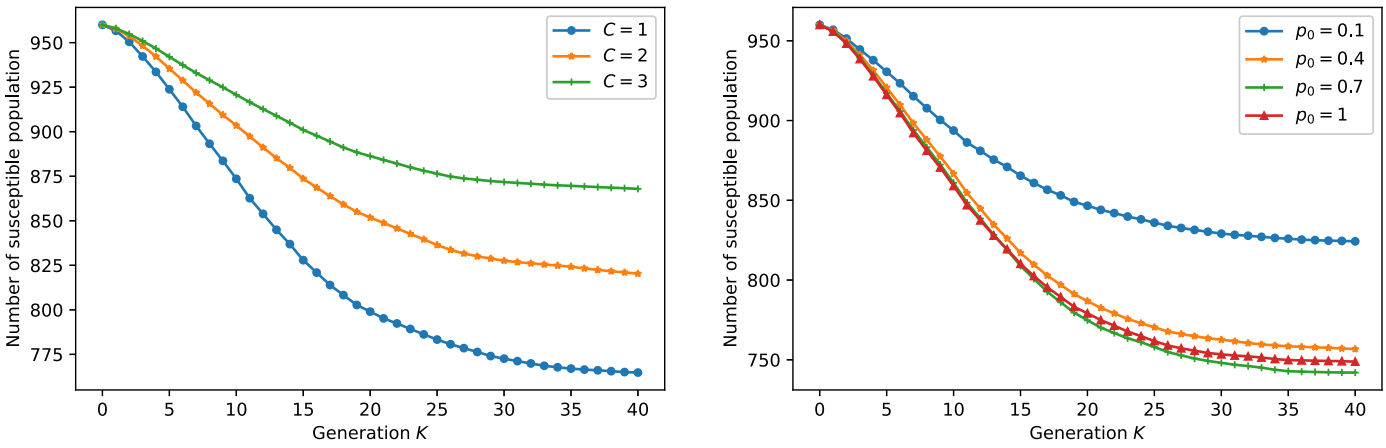


Fig. 3. Influence of C and p_0 in our model for $\mathscr{P}_3$ and $\Delta_{ID} = 1$.

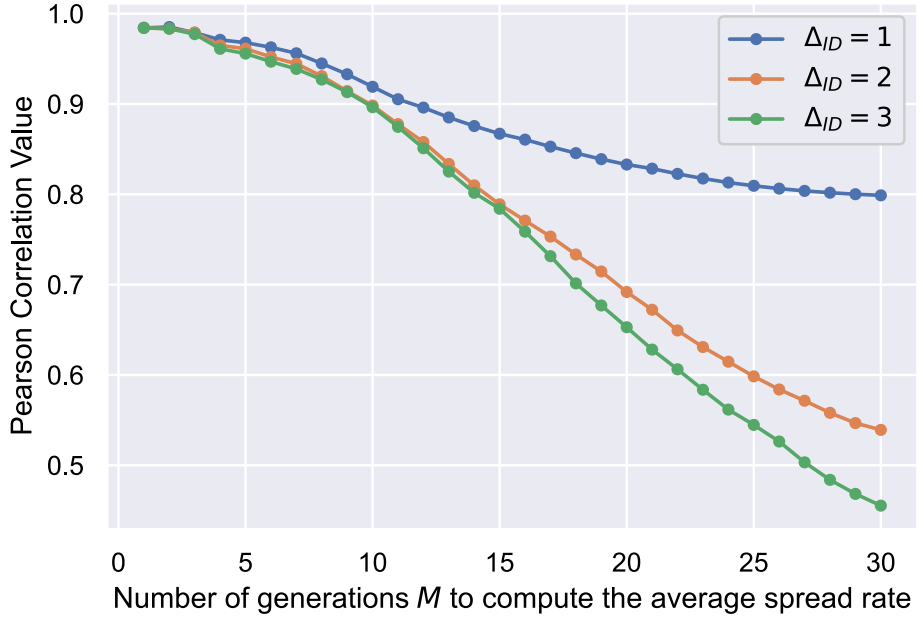


Fig. 4. Pearson correlation values between marginal E infection values and the average of the spread rate d_M.

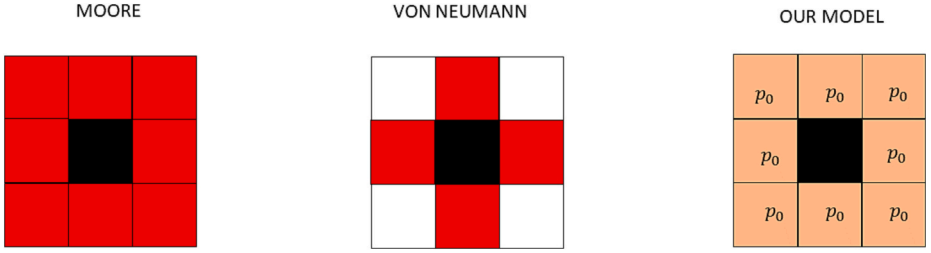


Fig. 5. Neighbourhoods in the Moore (left), Von Neumann (centre) and our model (right).

than 1. □

Remark 4.1. Note that it is not necessary to take stochasticity into account, since following the update criterion of Eq. 3, when the probabilities are greater than or equal to 1, Bernoulli tests are not performed.

Therefore, if we consider Γ in a square grid such that

$$\Gamma = (\mathscr{P}, C, p_0) = (\{0,1,1,1,1\}, c, 1),\ c \geqslant 1,\ \Delta_{ID} = 1 \tag{4}$$

then we reach the Moore model, as shown in Fig. 6.

In the case of Von Neumann dynamics, the derivation is not so straightforward. Here, stochasticity must be taken into account that is we need to identify the specific random seed that results in such dynamics. Considering Fig. 5, we must find the random seed that infects the central elements buts avoids infecting the corner elements, for every single generation. Thus, by Property 4.2 we will prove the existence of this particular random seed and by Property 4.3 we will compute the

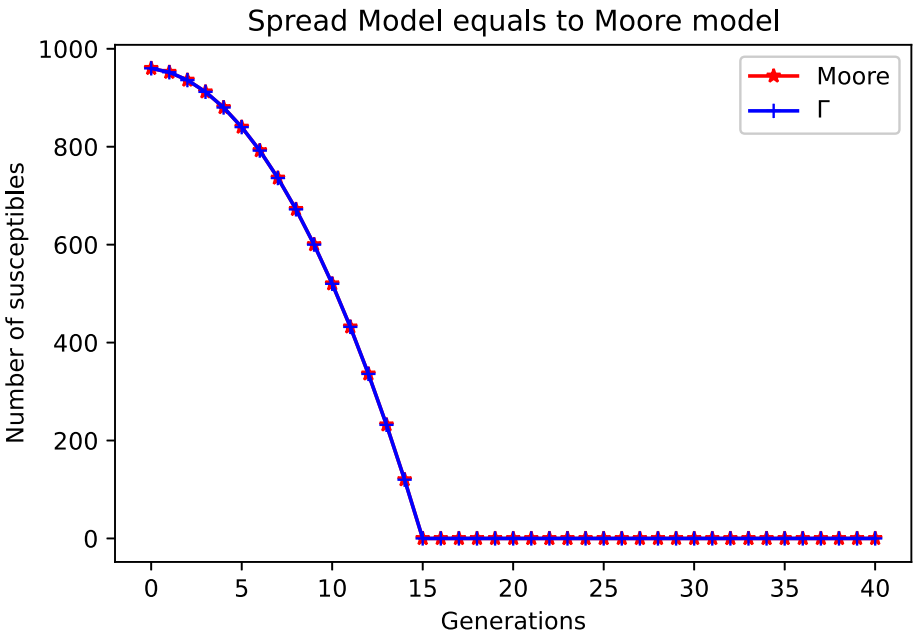


Fig. 6. Comparison between our model and Moore for parameter values in Γ.

probability of getting it randomly.

Property 4.2. Let $N, K \in \mathbb{N}$ and model parameters

$$\mathscr{P} = \{0, 1, 1, 1, 1\},\ c \geqslant 1,\ \Delta_{ID} = 1 \text{ and } p_0 = 0.5$$

then some random seed value exists for which our spread model is equal to the Von Neumann model.

Proof. It is clear that if we have one infected cell, then all its surrounding eight cells have 50% chance of being infected in the next step, due to p_0. At the same time, since our grid is finite, we will have a finite number $\widetilde{N}$ of possible outcomes. Moreover, Von Neumann's outcome is among these possible outcomes because of our model construction when $\rho \leqslant u_1$, where we know that $u_1 = 1$ due to $\mathscr{P}$. Thus, if we use different random seeds, our outcomes will be independent then we can compute the probability of getting Von Neumann according to the number n of samples. In this sense, let us consider the event

$$S_n := \{\text{Von Neumann's outcome is not in the first } n \text{ samples}\}$$

Obviously,

$$P\left(S_1\right) = \frac{\widetilde{N} - 1}{\widetilde{N}}$$

But as the n draws are independent thanks to the random seeds then,

$$P\left(S_n\right) = \left(\frac{\widetilde{N} - 1}{\widetilde{N}}\right)^n = \left(1 - \frac{1}{\widetilde{N}}\right)^n \underset{n \mapsto \infty}{\mapsto} 0$$

Thus, we can conclude that for a sufficiently large n, the probability of obtaining the Von Neumann model is sure. Therefore, some random seed exists for which our model and the Von Neumann model are equivalent. □

Property 4.3. Let N be an odd integer such that $N = 2n + 1$, and let the tuple of model parameters be

$$\Gamma = (\mathscr{P}, C, p_0) = (\{0, 1, 1, 1, 1\}, c, p_0),\ c \geqslant 1,\ \Delta_{ID} = 1 \text{ and } p_0 \in [0, 1].$$

Then, the probability p_{vn} of obtaining the Von Neumann dynamics depends on n and p_0, and its expression is given by

$$p_{vn}\left(n, p_0\right) = 16^{n^2} p_0^{4n(n+1)} (1 - p_0)^{4n^2}$$

Proof. Let's consider a square grid of size N. Notice that in the Von Neumann dynamics, we also start with a primitive matrix, meaning that in generation $k = 0$, we have a single infected element at position (n,n). However, in the next generation ($k = 1$), 4 elements get infected, two of which belong to row n and two to column n. The same dynamics occur for this row and column until generation $k = n$ (inclusive), where all the elements in the mentioned row and column have been infected. Thus, the mentioned elements get infected in the next generation after being exposed. Therefore, given that there are $4n$ elements with an infection probability of p_0, we have that

$$p_{vn} = K p_0^{4n}, \quad K \in \mathbb{R}$$

On the other hand, it is important to highlight that starting from generation $k \geqslant 2$, we obtain infected elements that do not belong to row or column n. Furthermore, the $N \times N$ grid can be divided into 4 blocks of size $n \times n$, along with row and column n. In this sense, we can reason by symmetry and focus on a single $n \times n$ block. Because of the model's construction and the Von Neumann dynamics, it is verified that every cell in the $n \times n$ block does not get infected the first time it is exposed with an infection probability of p_0. However, in the next generation, they do get infected when the infection probability is $2p_0$, as there are two infected cells around them that expose them to the infection. Since the above holds for every entry in the $4n \times n$ blocks, we have $4n^2$ entries that do not get infected with an infection probability of p_0 and get infected in the next generation with a probability of $2p_0$. Therefore, it follows that

$$\begin{aligned} p_{vn}(n, p_0) &= p_0^{4n} (1 - p_0)^{4n^2} (2p_0)^{4n^2} \\ &= 16^{n^2} p_0^{4n(n+1)} (1 - p_0)^{4n^2} \end{aligned}$$

□

Fig. 7 represents in Log scale the probability of getting the Von Neumann dynamics according to n, with $p_0 = 0.5$. We observe that the probability decreases when n increases, so even though such a seed exists, it is computationally expensive to find it. Moreover, thanks to Property 4.3 we can deduce that the optimal value for getting this dynamic also depends on the grid size and is

$$p_0 = \frac{n+1}{2n+1} \xrightarrow[n\to\infty]{} \frac{1}{2}$$

Remark 4.2. Let's see what the probability is, i.e. how many (expected) trials should be performed to find a seed for which we obtain the Von Neumann dynamics, on a small grid. Let us suppose $N = 5$ (i.e. $n = 2$) and we want to get Von Neumann's dynamic, then we know that the optimal value for p_0 is

$$p_0 = \frac{3}{5}$$

Moreover, the probability of getting this dynamic is

$$p_{vn}\left(2, 0.6\right) = 16^4\,(0.6)^{24}\,(1-0.6)^{16} \approx 1.3337 \times 10^{-7}$$

Therefore, we should carry out 10^7 tests approximately to find at least one random seed which provides us with the same dynamic.

Example 1. As the proof of Property 4.3 may be quite complicated, we will carry out an example using the same reasoning, which boils down to forcing the elements to be infected for each generation. Let's assume that $N = 5$ (or $n = 2$), then the initial state is

$$M^0 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Now, we compute the sequence

$$M^0, R^0, M^1, R^1, \ldots, M^4$$

where R^k is the neighbourhood relation determined by the parameter tuple Γ of Property 4.3 and M^k is the Von Neumman dynamics for generation k. Thus, one has

$$R_0 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & p_0 & p_0 & p_0 & 0 \\ 0 & p_0 & 0 & p_0 & 0 \\ 0 & p_0 & p_0 & p_0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad M^1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad q_1 = p_0^4\,(1-p_0)^4$$

$$R_1 = \begin{bmatrix} 0 & p_0 & p_0 & p_0 & 0 \\ p_0 & 2p_0 & 0 & 2p_0 & p_0 \\ p_0 & 0 & 0 & 0 & p_0 \\ p_0 & 2p_0 & 0 & 2p_0 & p_0 \\ 0 & p_0 & p_0 & p_0 & 0 \end{bmatrix}, \quad M^2 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 1 & 2 & 2 & 2 & 1 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \quad q_2 = p_0^4\,(2p_0)^4\,(1-p_0)^8$$

$$R_2 = \begin{bmatrix} p_0 & 2p_0 & 0 & 2p_0 & p_0 \\ 2p_0 & 0 & 0 & 0 & 2p_0 \\ 0 & 0 & 0 & 0 & 0 \\ 2p_0 & 0 & 0 & 0 & 2p_0 \\ p_0 & 2p_0 & 0 & 2p_0 & p_0 \end{bmatrix}, \quad M^3 = \begin{bmatrix} 0 & 1 & 2 & 1 & 0 \\ 1 & 2 & 2 & 2 & 1 \\ 2 & 2 & 2 & 2 & 2 \\ 1 & 2 & 2 & 2 & 1 \\ 0 & 1 & 2 & 1 & 0 \end{bmatrix}, \quad q_3 = (2p_0)^8\,(1-p_0)^4$$

$$R_3 = \begin{bmatrix} 2p_0 & 0 & 0 & 0 & 2p_0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 2p_0 & 0 & 0 & 0 & 2p_0 \end{bmatrix}, \quad M^4 = \begin{bmatrix} 1 & 2 & 2 & 2 & 1 \\ 2 & 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 & 2 \\ 1 & 2 & 2 & 2 & 1 \end{bmatrix}, \quad q_4 = (2p_0)^4 \tag{5}$$

Then, the Property 4.3 is satisfied since

$$p_{vn}\left(2, p_0\right) = \prod_{i=1}^{4} q_i = 16^4\,p_0^{24}\,(1-p_0)^{16}$$

Note that this numerical example proves that the reasoning carried out in the proof of Property 4.3 holds for $N = 5$. Moreover, it helps in the understanding of the calculation of probabilities for cells belonging to blocks of size $n \times n$.

Corollary 4.1. The space of possible outcomes for our spread model contains Von Neumann and Moore outcomes.

Proof. It is clear by Property 4.1 and 4.3 □

5. The current model as a Monte Carlo Method

One of the key strengths of this model is its stochastic nature. When updating the grid state, we often need to run Bernoulli trials, which involve random sampling and therefore rely on pseudo-random numbers (i.e., random seed values). Even if we were to set the climatic conditions, the model would still exhibit stochastic behaviour due to the way the grid is updated. Consequently, we can see in Fig. 8 our model may be applied as Monte Carlo method since setting the climate conditions and parameters, we can generate different outcomes by changing the random seed values.

Remark 5.1. For ease of notation, let us consider

$$X_i = \left\{S_k^i\right\}_{k=0}^{K} \qquad \forall 1 \leqslant i \leqslant n$$

$$\text{Rel.Freq.of}\{X \in A\} \approx P(X \in A)$$

By applying the *law of large numbers* to our model, we can assert that

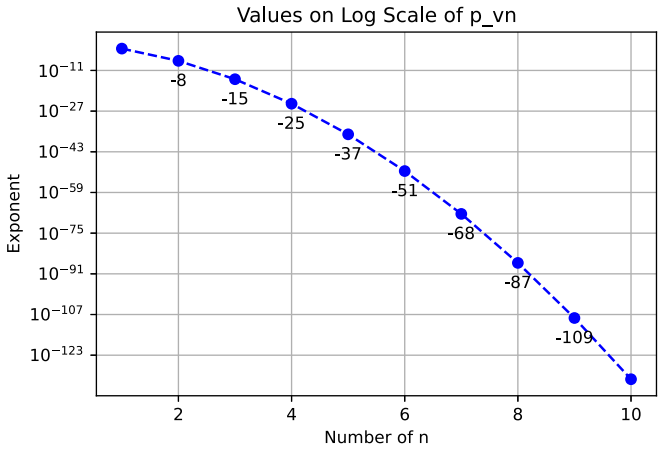


Fig. 7. Probabilities of getting the Von Neumann's dynamics according to n and being $p_0 = 0.5$.

the probability of cell (i,j) being susceptible (resp. infected, dead) in generation k converges almost surely to the relative frequency of being susceptible (resp. infected, dead). Thus, for a sufficiently large n, by computing the relative frequencies of sequence $X_1, X_2, \ldots, X_n$ for each cell and generation, we can calculate the probability of being healthy (or infected, dead). We define $X \in [0,1]^{N\times N\times K}$ as the variable that stores the estimated probabilities of the cells being healthy (or infected, dead). This outcome will be very important since it will help to understand the phenomenon scope and which areas are more likely to be affected according to our approach, climate conditions and parameters.

Remark 5.2. Moving forward, we will focus on studying the probability of the cells being healthy. However, reasoning can be symmetrically extended to study the probability of cells being infected or dead. In this regard, the sequence $X_1, X_2, \ldots, X_n$ i.i.d. refers to whether the cell is healthy or not, that is, $X_i \in \{0,1\}^{N\times N\times K}$.

Therefore, it follows that

$$X = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \text{with} \quad E\left[X\right] = \mu, \quad Var\left[X\right] = \frac{\sigma^2}{n} \tag{6}$$

We can define the following metrics for the estimator X of μ:

$$F\left(X\right) := \sum_{k}\sum_{i,j} X_{i,j}^{k}; \quad \text{and} \quad F\left(X,k\right) := \sum_{i,j} X_{i,j}^{k}$$

And for estimator $\hat{s}$ of σ^2, it would be enough with:

$$\hat{s} := \max\left\{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - X\right)\right\}$$

Example 2. We will now consider a square grid of size $N = 21$and a number of generations $K = 15$. We will apply our spread model to study the probability of being healthy for each cell, under simulated weather conditions $\mathscr{D}$ and parameters Γ. Specifically, we set

$$\Gamma = (\mathscr{P}, C, p_0) = (\{0, 0.1, 0.5, 0.9, 1\}, 2, 0.25),\ \Delta_{ID} = 1$$

We will use the resulting probability estimator X for each cell. The numerical results are presented in Table 4 and Fig. 9. Our estimations achieve a high level of accuracy, as $F(X,k)$ converges for all $k \in \{5, 10, 15\}$. Assuming that the error is uniformly distributed across every cell of the grid, we obtain an error of approximately 1.5×10^{-3} for each cell. Additionally, we observe that the estimates of the grid-wide mean quickly converge to fairly good confidence levels. Lastly, regarding the

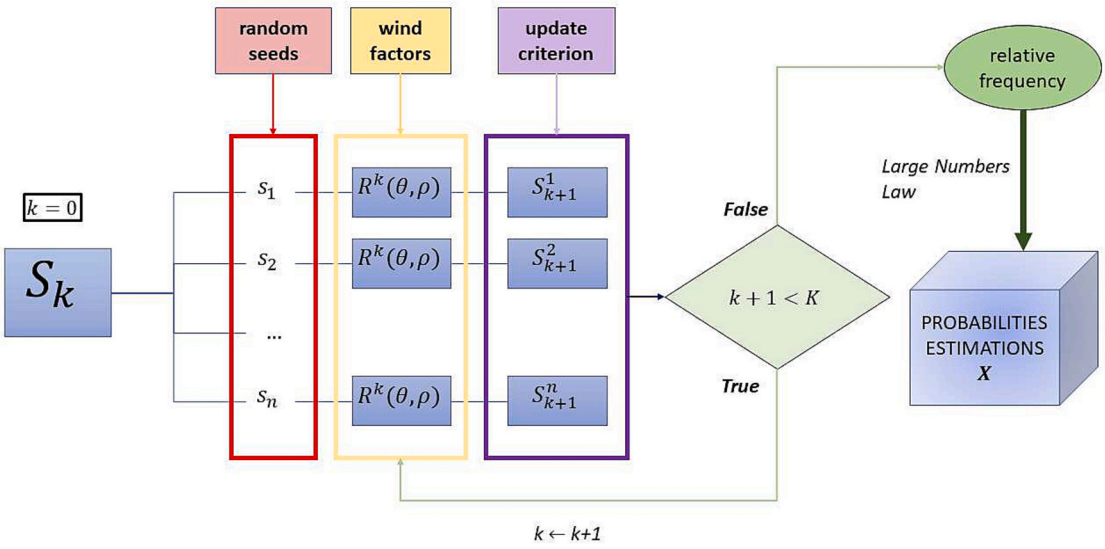


Fig. 8. Overview of the operational principles of the proposed model as Monte Carlo Method.

Table 4
Estimations of the population mean and variance using Monte Carlo simulations with $\Delta_{ID} = 1$.

n	$F(X,5)$	$F(X,10)$	$F(X,15)$	$F(X)$	$F(\lvert X^{i+1} - X^{i}\rvert)$	$\hat{s}$
10	417.3000	377.4000	364.3000	6368.5000	–	–
10^2	404.5400	350.0500	336.8400	6079.7300	288.7700	0.4955
10^3	401.7320	343.4400	330.7950	6012.9610	66.7690	0.2720
10^4	401.0213	342.5184	330.4957	6002.9682	9.9928	0.2513
10^5	401.4771	343.1444	331.0231	6009.9809	7.0127	0.2501

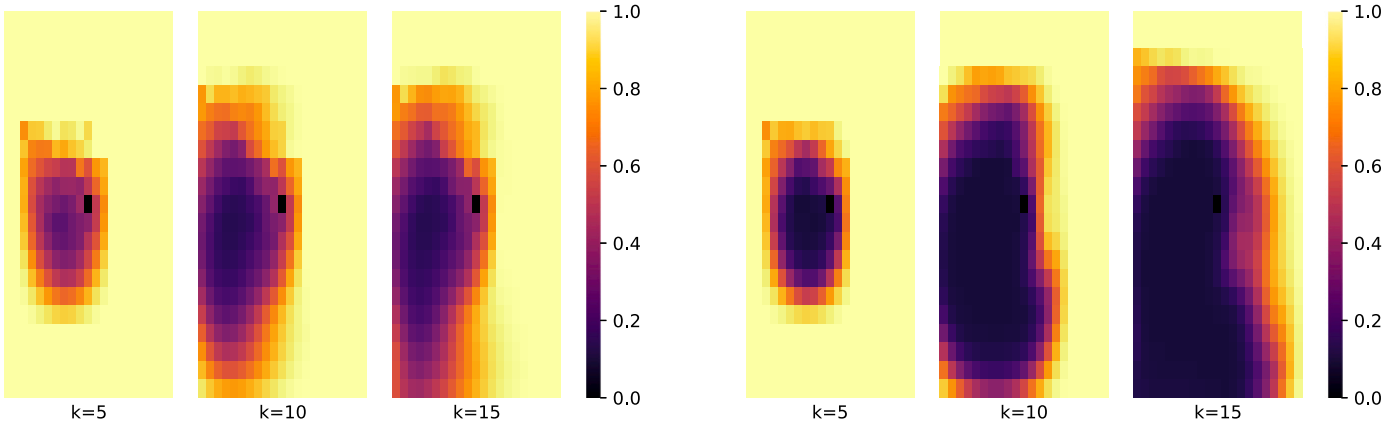


Fig. 9. Comparison of heatmaps when $\Delta_{ID} = 1$ (left) and $\Delta_{ID} = 2$ (right).

Table 5
Estimations of the population mean and variance using Monte Carlo simulations with $\Delta_{ID} = 2$.

n	$F(X,5)$	$F(X,10)$	$F(X,15)$	$F(X)$	$F(\lvert X^{i+1} - X^i \rvert)$	$\widehat{s}$
10	399.90000	306.60000	265.10000	5623.20000	–	–
10^2	389.63000	280.41000	227.06000	5322.48000	337.48000	0.397753
10^3	386.15500	269.55400	211.82900	5203.31500	135.01500	0.262959
10^4	385.50440	267.95510	210.18240	5186.82160	25.00460	0.251203
10^5	385.87992	268.70916	211.10374	5195.11573	9.54519	0.250145

estimations $\widehat{s}$ of the population variance, we may assume that

$$\left(\sigma^2\right)_{i,j}^{k} \lessapprox 0.250115 \qquad \forall 0 \leqslant i,j \leqslant N-1, \quad \forall 0 \leqslant k \leqslant K$$

Remark 5.3. It is considered X^{i+1} as the estimation of X using $n = 10^i$.

Example 3. In order to study the effect of the delay Δ_{ID}, we will repeat Example 2, but with $\Delta_{ID} = 2$. The results are shown in Table 5 and Fig. 9.

On the other hand, it is well-known that confidence interval CI for μ with a level of confidence α can be calculated by Eq. (7), where $z_{\frac{\alpha}{2}}$ refers to the critical value from the standard normal distribution that corresponds to the level of significance $\frac{\alpha}{2}$ in a two-tailed test

$$CI_\alpha\left(\mu\right) := \left[\widehat{\mu} - z_{\frac{\alpha}{2}}\frac{\widehat{\sigma}}{\sqrt{n}}, \widehat{\mu} + z_{\frac{\alpha}{2}}\frac{\widehat{\sigma}}{\sqrt{n}}\right] \tag{7}$$

In our case, we assume that $\widehat{\mu} := X$, and $\widehat{\sigma} := \sqrt{s}$. Then, if we consider as standard error SE_α as:

$$SE_\alpha := z_{\frac{\alpha}{2}}\frac{\widehat{\sigma}}{\sqrt{N}}$$

Let us suppose that we can achieve an accuracy around 10^{-2}, then it is enough with

$$SE_\alpha \leqslant 10^{-2} \Rightarrow N \geqslant z_{\frac{\alpha}{2}}^2 10^4 \widehat{\sigma}^2$$

Thus, by applying the aforementioned equation for the specific case where $\alpha = 0.95$, it can be concluded that $n = 9604$ guarantees an accuracy of 10^{-2} for each cell. Fig. 9 presents a heatmap of $\widehat{\mu}$, which displays the outcomes of Examples 2 and 3 and allows identification of the affected cells.

5.1. Limitations of the model as a classification model

Our classification model is built upon the Monte Carlo method. While Monte Carlo codes have achieved significant sophistication, simulations suffer from the drawback of demanding extensive computational resources to attain an adequate level of result confidence. Hence, it is crucial to acknowledge the associated limitations.

Specifically, a substantial number of tests may be required. Insufficient testing could lead to a heightened sampling error, resulting in overly broad confidence intervals. Ensuring an ample number of trials is crucial to obtain more precise and reliable estimates.

Moreover, the computational cost grows exponentially with the increase in the number of variables in the model and the grid size for analysis. Despite the potential for parallelization to improve efficiency, obtaining accurate results under such circumstances can be challenging due to computational demands.

A critical consideration involves the estimator used in the Monte Carlo method, which relies on the average of test results Eq. 6. However, it is noteworthy that this estimator may not always minimize variance. This observation presents an opportunity for future research to explore techniques for variance reduction, see Section 6.

Therefore, while our classification model harnesses the power of the Monte Carlo method, its practical application requires careful consideration of the computational challenges and the need for a sufficient number of trials to ensure reliable results. Exploring innovative variance reduction techniques (Kleijnen et al., 2010; Botev and Ridder, 2017) would help to improve the model's accuracy and broaden its applicability in future research and real-world applications.

5.2. Application to a real wildfire

Since our model can be used as a Monte Carlo method, we apply it to a real case. We consider the wildfire in Vall de Gallinera (Valencian Community, Spain) in August 2020, see Fig. 10. This is part of a database of wildfires in our community, which we are setting up, see **Supplementary Data**.

Then, the estimate that our model would make is given by Fig. 12, assuming Fig. 11 as our climate conditions for θ and ρ.

Thus, the outcomes obtained in this study provide a **credible and realistic starting point** and lay a solid foundation that can be improved and refined in future work, see Section 6.

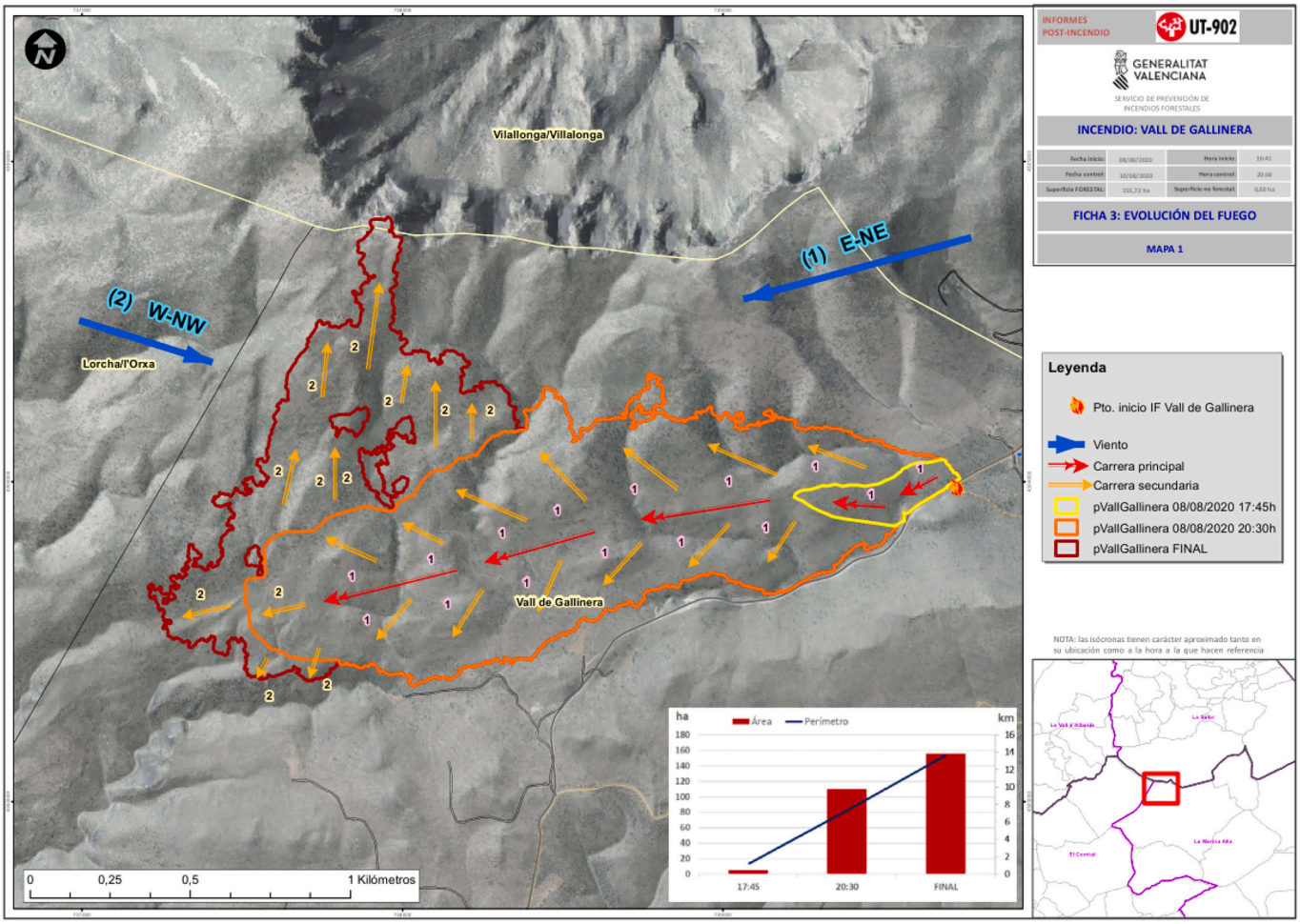


Fig. 10. Wildfire representation in Vall de Gallinera.

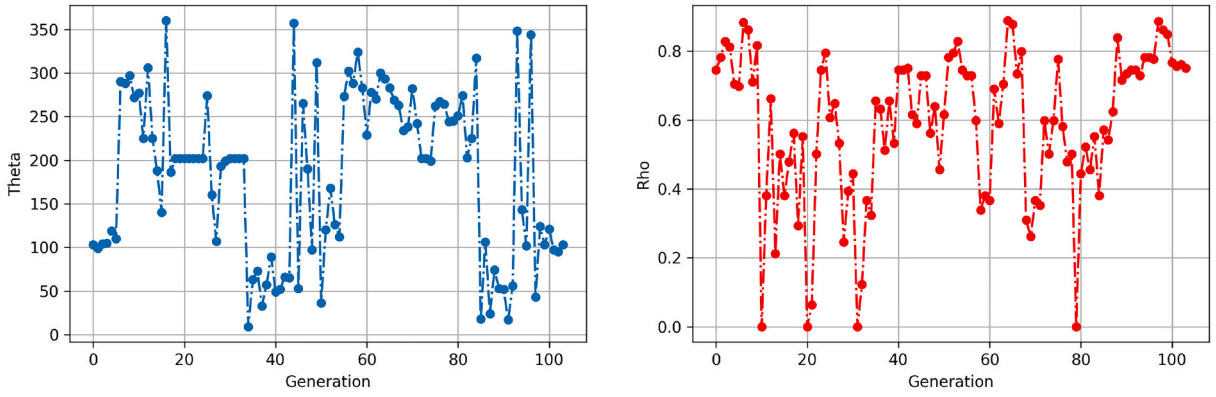


Fig. 11. Climate conditions θ, ρ in the wildfire.

6. Conclusions and future work

In this paper, we have explored the potential of our 2D-CA model driven by wind direction θ and intensity ρ. Our approach has demonstrated the ability to modulate the expansion of the phenomenon by various parameters. Furthermore, it has been observed that by carefully selecting these parameters, it is possible to consider the Moore and Von Neumann dynamics as particular cases, taking advantage of the stochasticity of the model.

Another important result is the identification of the measure of expected infections E as an early indicator of the rate of expansion since we have observed a strong linear relationship between them. This fact could help us to establish a hierarchy between tuples of parameters, and therefore to optimise the parameters by means of the spread rate. Furthermore, stochasticity has allowed us to use the model as a Monte Carlo method, which means that we can estimate the probability of

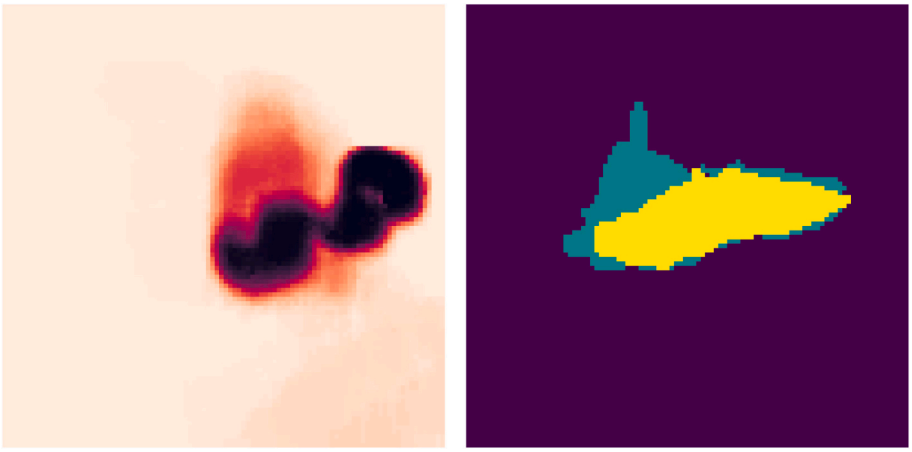


Fig. 12. Estimation parts fired heatmap (left). Representation of the real parts affected (right).

belonging to each state for each generation, given a confidence level.

Finally, our model has been applied to a real case and we have seen that the results provide a credible and adequate basis for future studies.

Regarding future work, we identify the following lines of research:

a.**Increasing the number of climate variables**: To establish a more accurate neighbourhood relationship, we propose to incorporate a larger number of climate variables. For example, we could consider as climate functions the parameters C and p_0, i.e.

$$C(t,h) \text{ and } p_0(t,h),$$

where t and h represent temperature and humidity resp.

b.**Modification of the update criterion and application of machine learning**: We will seek to improve our model by modifying the update criterion, maintaining stochasticity but making it differentiable. This will allow us to take advantage of machine-learning tools and turn our model into an intelligent cellular automaton.

c.**Optimising the scope**: We will make changes in the partitions used to increase the number of cells exposed in each iteration. This will become another adjustable parameter to improve the accuracy and efficiency of the model.

d.**Strategies of variance-reduction**: As previously mentioned, the estimator given by Eq. 6 may not have the minimum variance. Therefore, it opens the possibility to investigate and develop new and more sophisticated estimators by using variance reduction strategies such as the importance sampling, stratified sampling, Quasi-Monte Carlo methods (Kleijnen et al., 2010; Botev and Ridder, 2017) or even Markov chain Monte Carlo methods (Brooks, 1998)

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Links to the code and data are available at Supplementary data.

Acknowledgement

This research is funded by Generalitat Valenciana, project AICO/2021/331. This work does not have any conflicts of interest.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ecoinf.2023.102266.

References

Andrews, Patricia L., 2007. Behaveplus fire modeling system: past, present, and future. In: Proceedings of 7th Symposium on Fire and Forest Meteorology, pp. 23–25. Boston: MA American Meteorological Society.

Asensio, M.I., Ferragut, L., 2002. On a wildland fire model with radiation. Int. J. Numer. Meth. Eng. 54 (1), 137–157.

Botev, Zdravko, Ridder, Ad, 2017. Variance reduction. Wiley statsRef: Statistics reference online, pp. 1–6.

Boters Pitarch, Joan, María Teresa, Signes Pont, Cortés-Plana, José Juan, Mora, Higinio, 2023. A new stochastic approach to the spread of environmental events enhanced by the wind. Math. Methods Appl. Sci.

Brooks, Stephen, 1998. Markov chain monte carlo method and its application. J. Royal Stat. Soc.: Ser. D (Stat.) 47 (1), 69–100.

Bui, Dieu Tien, Van Le, Hung, Hoang, Nhat-Duc, 2018. Gis-based spatial prediction of tropical forest fire danger using a new hybrid machine learning method. Ecol. Inform. 48, 104–116.

Chen, Rong-Jian, Horng, Shi-Jinn, 2010. Novel scan-ca-based image security system using scan and 2-d von neumann cellular automata. Signal Process.: Image Commun. 25 (6), 413–426.

Chen, Qiuwen, Mynett, Arthur E., 2003. Effects of cell size and configuration in cellular automata based prey–predator modelling. Simul. Model. Pract. Theory 11 (7–8), 609–625.

Farsite, Finney M.A., 2004. Fire area simulator-model development and evaluation. US Department of Agriculture, Forest Service, Rocky Mountain Research Station, 2004.

Ganteaume, Anne, Camia, Andrea, Jappiot, Marielle, San-Miguel-Ayanz, Jesus, Long-Fournel, Marlène, Lampin, Corinne, 2013. A review of the main driving factors of forest fire ignition over europe. Environ. Manage. 51, 651–662.

Gómez, Sergio, Arenas, Alexandre, Borge-Holthoefer, Javier, Meloni, Sandro, Moreno, Yamir, 2010. Discrete-time markov chain approach to contact-based disease spreading in complex networks. Europhys. Lett. 89 (3), 38009.

Hajian, Mohammad, Melachrinoudis, Emanuel, Kubat, Peter, 2016. Modeling wildfire propagation with the stochastic shortest path: a fast simulation approach. Environ. Model. Software 82, 73–88.

Hernández Encinas, L., Hoya White, S., Martín Del Rey, A., Rodríguez Sánchez, G., 2007. Modelling forest fire spread using hexagonal cellular automata. Appl. Math. Model. 31 (6), 1213–1227.

Hernández Encinas, A., Hernández Encinas, L., Hoya White, S., Martín del Rey, A., Rodríguez Sánchez, G., 2007. Simulation of forest fire fronts using cellular automata. Adv. Eng. Softw. 38 (6), 372–378.

Huang, Chung-Yuan, Chin, Wei-Chien-Benny, Wen, Tzai-Hung, Yu-Hsiang, Fu., Tsai, Yu-Shiuan, 2019. Epirank: Modeling bidirectional disease spread in asymmetric commuting networks. Scient. Rep. 9 (1), 5415.

Jiang, Wenyu, Wang, Fei, Fang, Linghang, Zheng, Xiaocui, Qiao, Xiaohui, Li, Zhanghua, Meng, Qingxiang, 2021. Modelling of wildland-urban interface fire spread with the heterogeneous cellular automata model. Environ. Model. Software 135, 104895.

Karafyllidis, Ioannis, Thanailakis, Adonios, 1997. A model for predicting forest fire spreading using cellular automata. Ecol. Model. 99 (1), 87–97.
Kleijnen, Jack P.C., Ridder, Ad, Rubinstein, Reuven, 2010. Variance reduction techniques in monte carlo methods.
Mandel, Jan, Bennethum, Lynn S., Beezley, Jonathan D., Coen, Janice L., Douglas, Craig C., Kim, Minjeong, Vodacek, Anthony, 2008. A wildland fire model with data assimilation. Math. Comput. Simul. 79 (3), 584–606.
Muzy, Alexandre, Nutaro, James J., Zeigler, Bernard P., Coquillard, Patrick, 2008. Modeling and simulation of fire spreading through the activity tracking paradigm. Ecol. Model. 219 (1–2), 212–225.
Noonan-Wright, Erin K., Opperman, Tonja S., Finney, Mark A., Thomas Zimmerman, G., Seli, Robert C., Elenz, Lisa M., Calkin, David E., Fiedler, John R., 2011. Developing the us wildland fire decision support system. J. Combust. 2011.
Pitts, William M., 1991. Wind effects on fires. Prog. Energy Combust. Sci. 17 (2), 83–134.
Roy, Satyabrata, Shrivastava, Manu, Pandey, Chirag Vinodkumar, Nayak, Sanjeet Kumar, Rawat, Umashankar, 2021. Ievca: An efficient image encryption technique for iot applications using 2-d von-neumann cellular automata. Multimedia Tools Appl. 80, 31529–31567.
Signes Ponnt, M.T., Mora, H., Cortés, Castillo Antonio, 2017. The Susceptible-Infectious-Recovered (SIR) model of disease expansion: a new approach. In: 17th Edition of the Mathematical Modelling in Engineering and Human Behavior Conference, 2017.
Signes Pont, María Teresa, Ramirez-Martinez, Deivis Eduard, García-Chamizo, Juan Manuel, Mora, Higinio, et al., 2019. A multigrid approximation to the expansion of xylella fastidiosa in almond trees.
Signes-Pont, María Teresa, Cortés-Plana, José Juan, Mora, Higinio, Mollá-Sirvent, Rafael, 2021. An epidemic model to address the spread of plant pests. the case of xylella fastidiosa in almond trees. Kybernetes 50 (10), 2943–2955.
Tuyen, Tran Thi, Jaafari, Abolfazl, Yen, Hoang Phan Hai, Nguyen-Thoi, Trung, Van Phong, Tran, Nguyen, Huu Duy, Van Le, Hiep, Phuong, Tran Thi Mai, Nguyen, Son Hoang, Prakash, Indra, et al., 2021. Mapping forest fire susceptibility using spatially explicit ensemble models based on the locally weighted learning algorithm. Ecol. Inform. 63, 101292.
Viegas, Domingos X., 2004. Slope and wind effects on fire propagation. Int. J. Wildland Fire 13 (2), 143–156.