

LYCÉE LAKANAL

Simulation d'épidémies à l'aide d'automates cellulaires

ARTHUR LACON - TIMOTHÉ RIOS

Table des matières

1	Première version	2
1.1	Présentation	2
1.2	Résultats	4
1.3	Limites	7
2	Deuxième version	7
2.1	Présentation	7
2.2	Résultats	8
2.3	Comparaison	9
3	Méthode différentielle	12
3.1	Équations	12
3.2	Propriétés	12
3.3	Résultats	13

Abstract

Within the theme of transport, we have looked at how an epidemic spreads. To be able to represent this phenomenon, we chose to use a cellular automaton. A cellular automaton is a matrix of cells whose states vary over time. Here, at every moment, each cell can be Healthy, Sick, Deceased or Healed, according to its state as well as that of its neighbours at the moment before. Thus, with the information of an initial state, it is possible to materialize the spread of an epidemic. Of course, in order to accurately represent reality, it is necessary to make this model more complex by taking into account many factors that play a role in how a disease spreads.

INTRODUCTION

L'étude de la propagation des maladies est une science qui remonte à la Grèce Antique, lorsque Hippocrate, au quatrième siècle avant Jésus-Christ, créa les termes d'endémie et d'épidémie afin de qualifier respectivement des maladies liées à des endroits et à des périodes donnés. Mais cette science ne se développa vraiment qu'à partir de 1854, quand John Snow, un médecin britannique, étudia la propagation de l'épidémie de choléra dans les quartiers de Londres. Ce sont là les prémices de l'épidémiologie, l'étude du transport des infections. De nos jours, les méthodes ont évolué et la méthode différentielle, que l'on présentera plus loin, est la plus fréquemment utilisée. Nous nous sommes aussi intéressés ici à la simulation par automates cellulaires.

1 Première version

1.1 Présentation

Définition formelle

Un automate cellulaire est une matrice de cellules, chacune ayant un état appartenant à un ensemble prédéfini. L'état de chaque cellule peut varier au cours du temps suivant une fonction de transfert : l'état de la matrice à l'instant $t + 1$ dépend ainsi de son état à l'instant t . L'automate doit donc posséder un état initial, c'est-à-dire la matrice des états initiaux des cellules. Même si son principe de base est simple, l'automate cellulaire se complexifie grandement lorsque la taille de la grille augmente, ce qui en fait un modèle couramment utilisé dans l'étude des systèmes complexes.

Historique

Les automates cellulaires sont plutôt récents. Ils ont été mis au point par John Von Neumann dans son étude des systèmes auto-réplicatifs dans les années 1940. Cette notion a été grandement popularisée par le "jeu de la vie" de John Conway, un automate cellulaire en deux dimensions paru dans les années 1970. À ce jour, les automates cellulaires ont de nombreuses applications dans divers domaines :

- diffusion d'un gaz en s'appuyant sur les équations de Navier-Stokes
- simulation des feux de forêts
- simulation du trafic routier

Notre automate cellulaire

Notre automate cellulaire est un automate en deux dimensions qui vise à simuler la propagation d'une épidémie. Les cellules ont donc un état parmi les quatre suivants : Sain, Malade, Guéri, Mort. L'état Sain est l'état initial que chaque cellule sauf une possède à la première étape de la simulation. Lorsqu'une cellule est saine, elle peut tomber malade avec une probabilité p_1 s'il y a des malades dans son entourage. L'état Malade est au départ donné à une seule cellule, le "patient zéro". Une cellule malade peut soit mourir avec une probabilité p_2 soit guérir avec une probabilité p_3 . Les cellules mortes ou guéries restent dans cet état indéfiniment, l'hypothèse prise étant qu'une cellule guérie est immunisée et ne peut plus attraper la maladie. Enfin, la simulation s'arrête lorsqu'il n'y a plus de cellule malade, car l'état de l'automate ne peut alors plus varier.

Calcul des probabilités

Les probabilités d'infection, de soin et de mort que nous avons utilisé ont été calculées à partir de paramètres empiriques tirés d'études réelles. On utilisera en particulier τ , le temps moyen qu'un individu reste infecté, et R_0 , le nombre moyen d'individus qu'un malade contamine durant son temps de contamination.

On cherche d'abord à déterminer la probabilité d'infection. On se place donc dans (Ω, P) un espace probabilisé fini. On étudie un individu malade entouré de n voisins sains. On considère que l'infection d'un individu sain est indépendante des autres.

On établira d'abord un résultat préliminaire.

Soit X, Y deux variables aléatoires réelles sur Ω et à valeur dans $\llbracket 0, n \rrbracket$, $n \in \mathbb{N}$, telles que $\forall k \in \llbracket 0, n \rrbracket$, X sachant que $Y = k$ suit une loi binomiale de paramètres $(n - k, p)$, $p \in]0, 1[$.

Alors $E(X) = (n - E(Y))p$.

On a donc $\forall k \in \llbracket 0, n \rrbracket$, $E_{\{Y=k\}}(X) = (n - k)p$. Par formule de l'espérance totale,

$$\begin{aligned} E(X) &= \sum_{k=0}^n P(Y = k) E_{\{Y=k\}}(X) \\ &= \sum_{k=0}^n P(Y = k)(n - k)p \\ &= np \sum_{k=0}^n P(Y = k) - p \sum_{k=0}^n kP(Y = k) \\ &= np - pE(Y) \\ &= p(n - E(Y)) \end{aligned}$$

Soit $k \in \llbracket 1, \tau \rrbracket$, X_k est la variable aléatoire donnant le nombre d'individus qui ont été infectés au jour k .

$X_1 \hookrightarrow B(n, p)$ car on infecte les individus indépendamment.

Au jour k il reste

$$n - \sum_{i=0}^{k-1} X_i$$

individus sains.

Montrons par récurrence que $\forall k \in \llbracket 1, \tau \rrbracket$, $E(X_k) = n(1 - p)^{k-1}p$

$\forall k \in \llbracket 1, \tau \rrbracket$, notons $P(k) = \forall i \in \llbracket 1, k \rrbracket, E(X_i) = n(1 - p)^{i-1}p$.

$P(1)$ est vraie car $X_1 \hookrightarrow B(n, p)$.

Ou bien $\tau = 1$ et c'est fini.

Ou bien $\tau \geq 2$ et on suppose que $P(k)$ est vraie pour $k \in \llbracket 1, \tau - 1 \rrbracket$. Montrons que $P(k + 1)$ est vraie.

$\forall i \in \llbracket 1, k \rrbracket$, $E(X_i) = n(1 - p)^{i-1}p$ par $P(k)$. Il reste à calculer $E(X_{k+1})$.

$$\begin{aligned} E(X_{k+1}) &= p(n - \sum_{i=1}^k E(X_i)) \\ &= p(n - \sum_{i=1}^k n(1 - p)^{i-1}p) \\ &= np(1 - p)^k \end{aligned}$$

On conclut par récurrence multiple finie.

Armés de ces résultats, nous pouvons quasiment calculer p , car on a le nombre moyen d'infectés par un malade au cours de sa période d'infection, R_0 . On a donc :

$$\begin{aligned}
R_0 &= \sum_{k=1}^{\tau} E(X_k) \\
&= \sum_{k=1}^{\tau} np(1-p)^{k-1} \\
&= np \frac{1 - (1-p)^{\tau}}{p} \\
&= n(1 - (1-p)^{\tau})
\end{aligned}$$

Ce qui nous amène par de brefs calculs à la conclusion :

$$p = 1 - \sqrt[\tau]{\frac{n - R_0}{n}}$$

On cherchera désormais à calculer la probabilité de ne plus être malade, c'est à dire ou bien de guérir ou bien de mourir, c'est à dire $p_{\text{mort}} + p_{\text{guérison}} = q$. Si l'on considère un individu qui est infecté au jour 0, et que l'on considère la variable aléatoire X donnant le jour de guérison (sachant qu'on ne peut guérir le jour de son infection), $\forall k \in \mathbb{N}^*, P(X = k) = (1 - q)^{k-1}q$.

X suit donc une loi géométrique et son espérance, c'est à dire le nombre de jour où l'individu restera malade (τ) vaut par propriété $\frac{1}{q}$.

On a donc

$$q = \frac{1}{\tau}$$

Enfin pour calculer la probabilité de mourir, il suffit de faire mq , où m est la mortalité de la maladie, et la probabilité de guérir vaut $(1 - m)q$.

Les résultats réels sont récapitulés dans le tableau suivant :

Maladie	$p_{\text{infection}}$	p_{mort}	$p_{\text{guérison}}$
Ebola	0.0187	0.0818	0.00909
Grippe	0.0406	0.0002	0.1998

1.2 Résultats

Voici quelques étapes d'une simulation de notre automate. Le foyer originel -la première cellule infectée- est placée au centre de la grille à l'étape 1. Sur ces images, représentant la grille de l'automate, les cellules saines sont représentées en blanc, celles malades sont rouges, celle guéries sont vertes et celles mortes sont noires.



FIGURE 1 – Étape 36

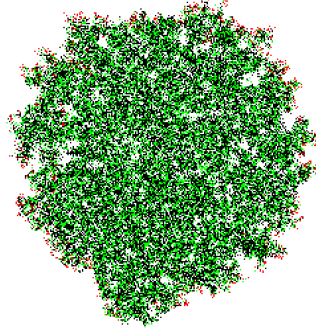


FIGURE 2 – Étape 195

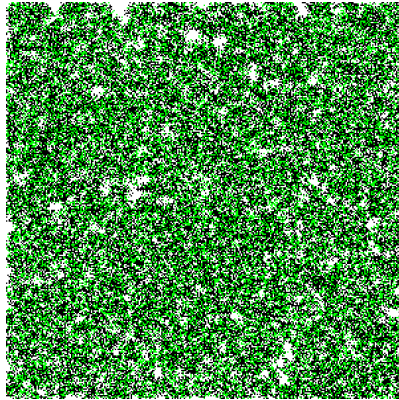
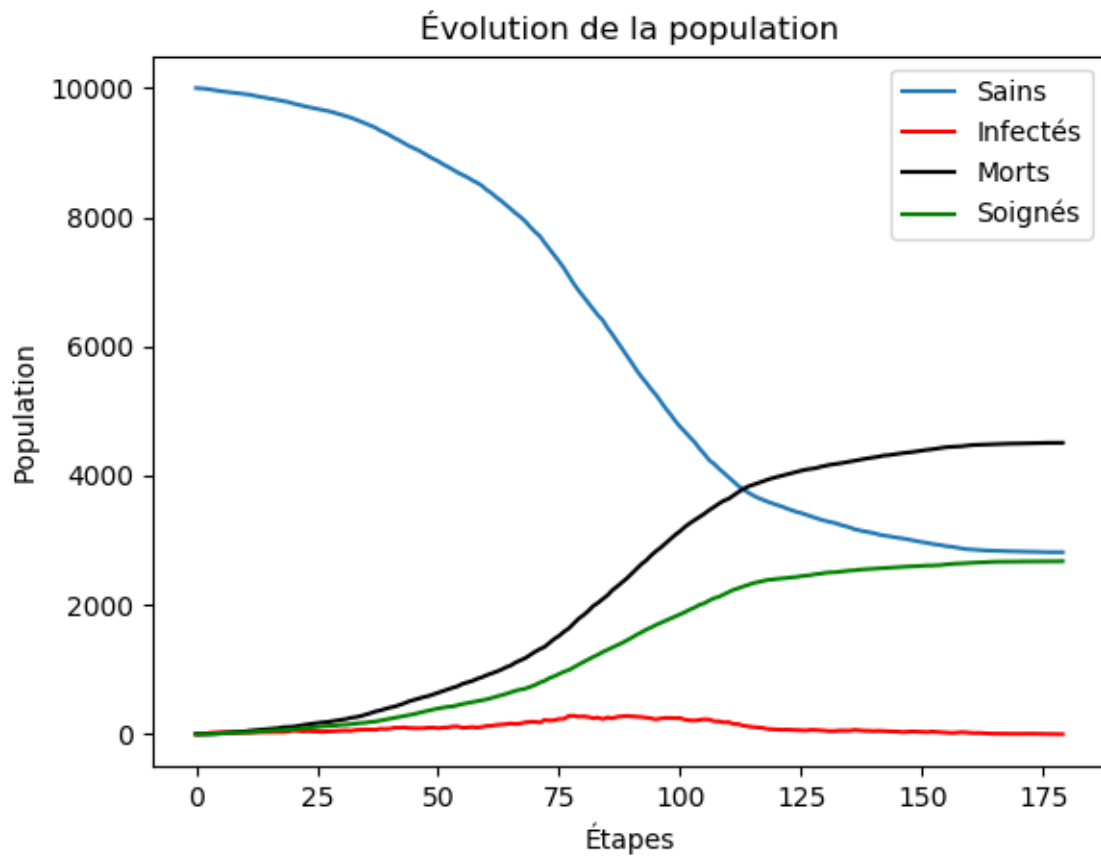
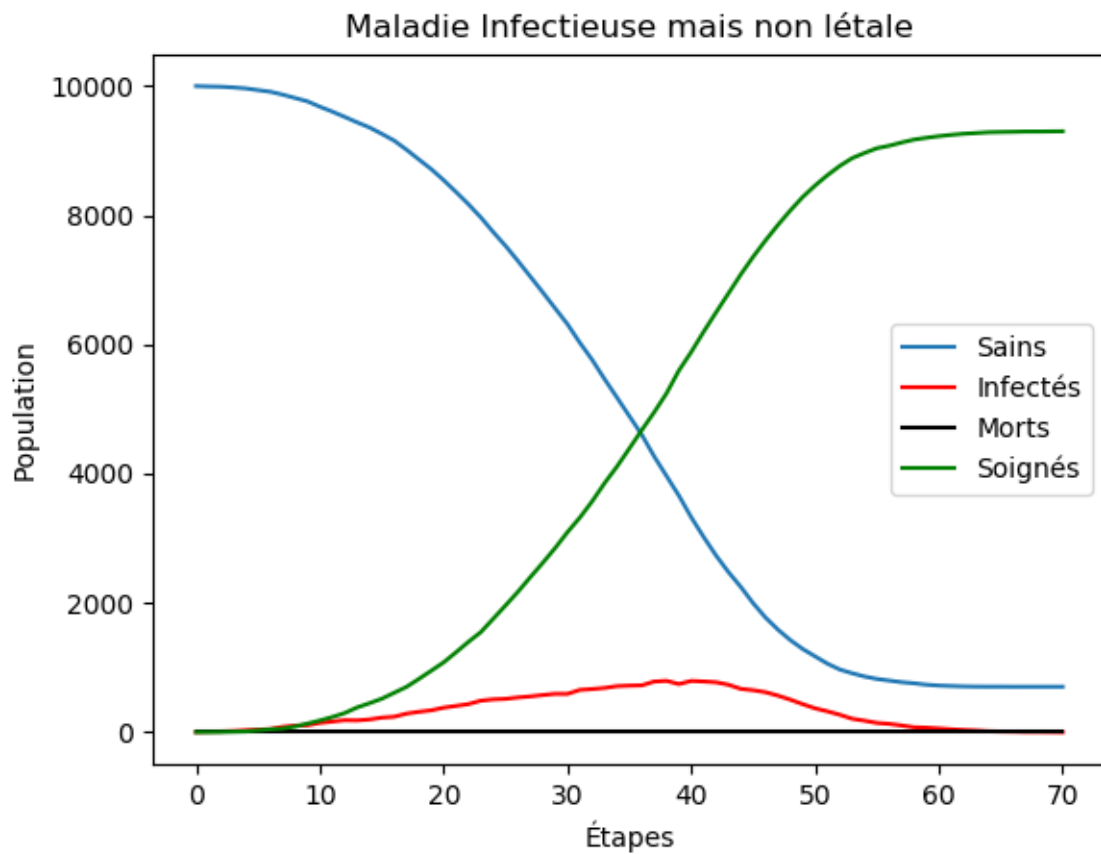


FIGURE 3 – Étape 412

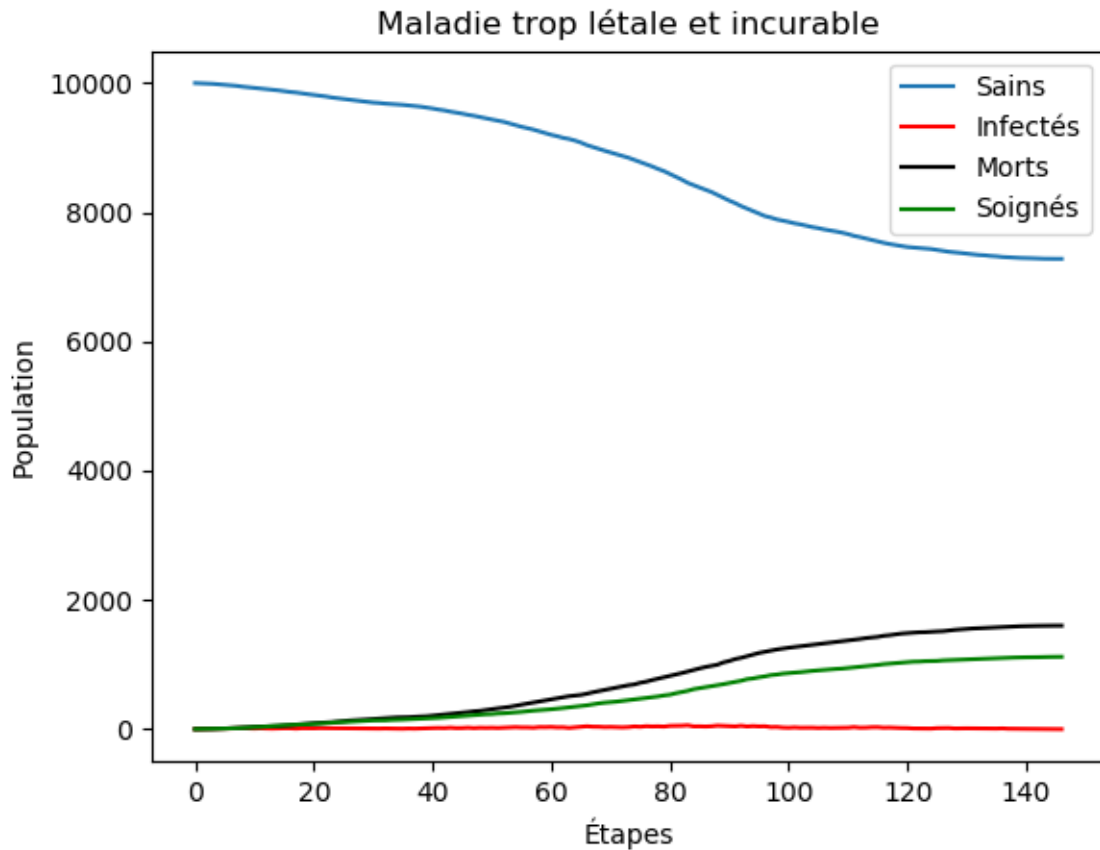
Dans la simulation présentée ci-dessus, les probabilités d'infection, de guérison et de mort ont été choisies arbitrairement afin de procurer un visuel clair du comportement de l'automate. On avait $p_{\text{infection}} = 0.2$, $p_{\text{mort}} = 0.25$, et $p_{\text{guérison}} = 0.2$. Voici encore d'autres exemples de cas possibles de comportements de maladie, dont l'évolution est cette fois-ci représentée sur un graphique donnant le nombre de personnes saines, malades, guéries et mortes au cours des étapes de la simulation.



Dans ce cas, p_1 valait 0.2, p_2 valait 0.25, p_3 valait 0.2 et la maladie pouvait se transmettre dans un rayon de deux cellules autour de l'infectée.



Nous avons utilisé ici les probabilités du virus de la grippe cités plus haut. La courbe de la population décédée est bien sûr beaucoup moins haute qu'au cas précédent $p_{\text{infection}} = 0.406$, $p_{\text{mort}} = 0.0002$, et $p_{\text{guerison}} = 0.1998$



Ici on avait $p_{\text{infection}} = 0.2$, $p_{\text{mort}} = 0.7$, et $p_{\text{guerison}} = 0$

Ce dernier exemple présente les résultats en utilisant les probabilités d'Ebola.

1.3 Limites

Cette version de l'automate cellulaire fournit déjà des résultats intéressants. Cependant, le modèle reste trop simple pour pouvoir modéliser la réalité d'une épidémie. En effet, il est ici supposé que les individus de la population étudiée restent sur place tout en étant étroitement collés les uns aux autres. Il apparaît ainsi qu'un grand nombre de paramètres a été négligé : il est ici supposé que la population est uniformément répartie, alors qu'en réalité cette densité dépend d'un nombre important de facteurs, comme le type de terrain, les constructions humaines (villes par exemple) ou les conditions météorologiques. De plus, nous supposons aussi que l'épidémie se déroule en un temps assez court pour que les décès naturels et les naissances soient négligeables, tout comme nous négligeons les mesures que pourraient prendre certains gouvernements en cas de pandémies majeures.

2 Deuxième version

2.1 Présentation

Ce deuxième automate prend en compte plus de paramètre :

- nous avons cette fois considéré les cellules comme des zones géographiques plutôt que des individus, ayant une population ainsi qu'une répartition Sains-Malades-Morts-Guérés propre.
- nous avons pris en compte les densités de population, celles-ci dépendant du 'type' de case, à savoir Ville, Campagne, Montagnes ou Route.
- nous avons pris en compte les mouvements de population, en fonction de deux paramètres : la probabilité qu'une partie de population quitte une case, et la probabilité qu'une case a d'attirer la population des cases adjacentes.

Ces ajouts de paramètres permettent une simulation plus fine et plus proche de la réalité. Cela permet d'utiliser de bien meilleure façon les automates cellulaires, qui ont pour avantage de prendre en compte les paramètres géographiques de la simulation. Cette simulation est aussi plus intéressante pour les simulations à grande échelle, étant donné que plus de population peut-être traitée sur une grille de la même taille.

Pour les paramètres nous avons choisi les mêmes que ceux du premier automate, étant donné que les fonctionnements restent proches.

2.2 Résultats

Les résultats graphiques de cet automate permettent d'afficher la densité de population des quatre états au sein de toutes les cellules de la grille. Bien sûr, il n'est maintenant plus possible d'afficher tous les paramètres de la grille sur une seule image. Voici par exemple la répartition des individus sains sur une grille

FIGURE 4 – Étape 10



FIGURE 5 – Étape 129

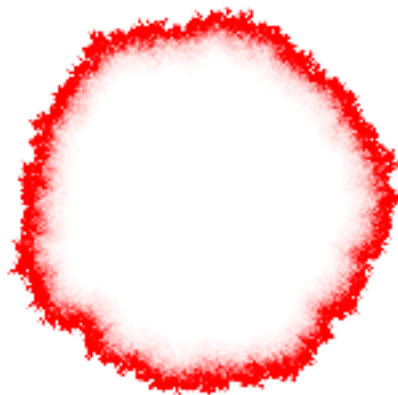


FIGURE 6 – Étape 214



Ici, plus une cellule contient de personnes malades, plus le pixel la représentant sur l'image est rouge. Les cellules sont donc toutes légèrement rougies à la première étape, non pas parce qu'elles contiennent des individus malades, morts ou guéris, mais parce que le programme étalonne automatiquement les couleurs en fonction de la taille de la grille et de sa population afin de pouvoir intensifier la couleur si la population saine d'une cellule augmente. Bien sûr, cette simulation servant d'exemple pour présenter ce nouvel automate, ses cellules possèdent des probabilités de déplacement des populations bien plus élevées que dans la réalité. On remarque dans la dernière image que l'épidémie a complètement envahi le côté droit avant les coins de gauche. Cet état est bien dû à l'incertitude causée par les probabilités.

2.3 Comparaison

Ce nouvel automate nous permet donc une modélisation plus précise de l'évolution d'une maladie. Ainsi, nous avons pu comparer nos résultats avec ceux de précédentes études. Les résultats montrés ci-dessous utilisent la nouvelle fonctionnalité de la géographie : les cartes affichent ainsi en rouge l'emplacement des cellules de type Ville, influant ainsi sur les résultats de la simulation.

FIGURE 7 – Ebola - 0 ville
Évolution de la population en fonction du temps

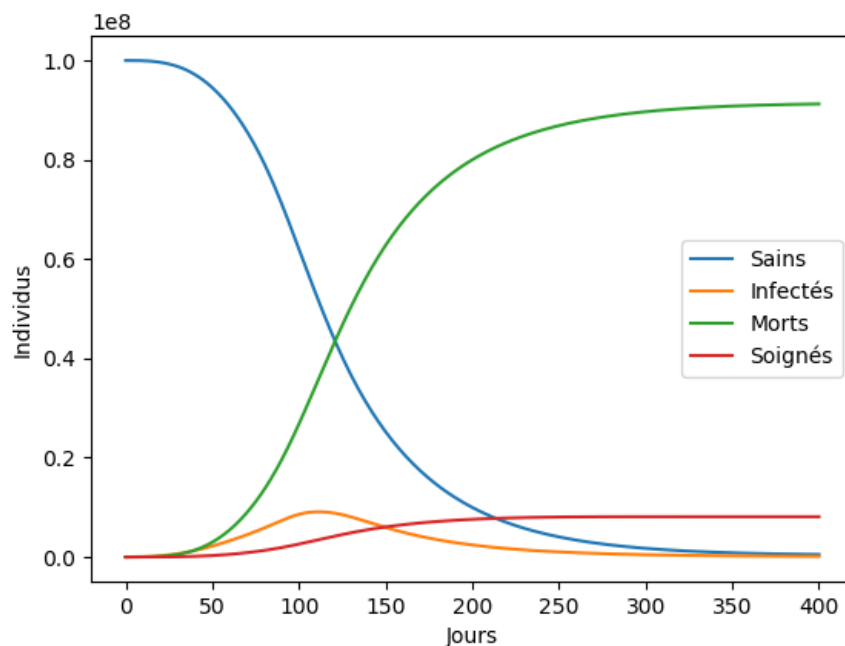


FIGURE 8 – Ebola - 1 ville
Évolution de la population en fonction du temps

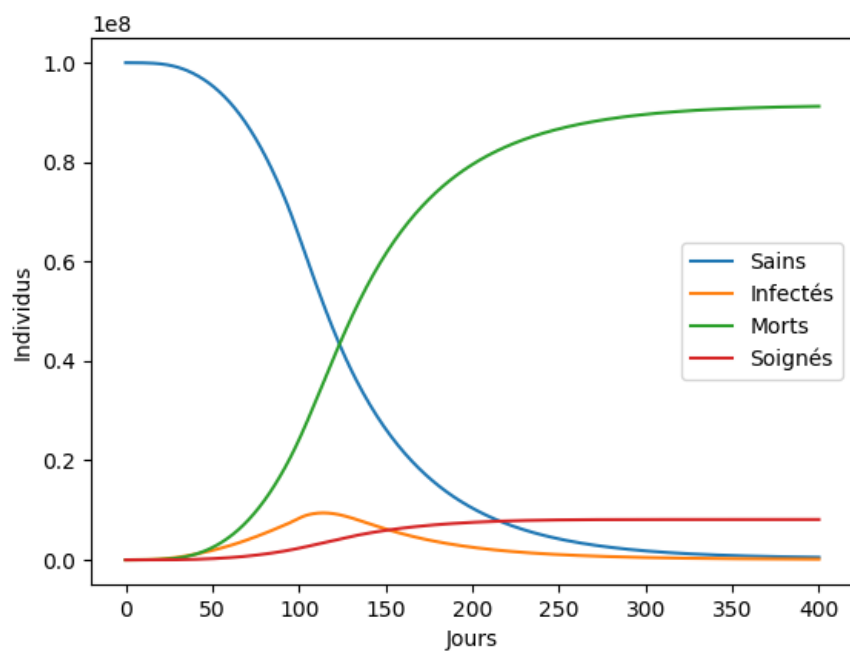


FIGURE 9 – Carte - 1 ville

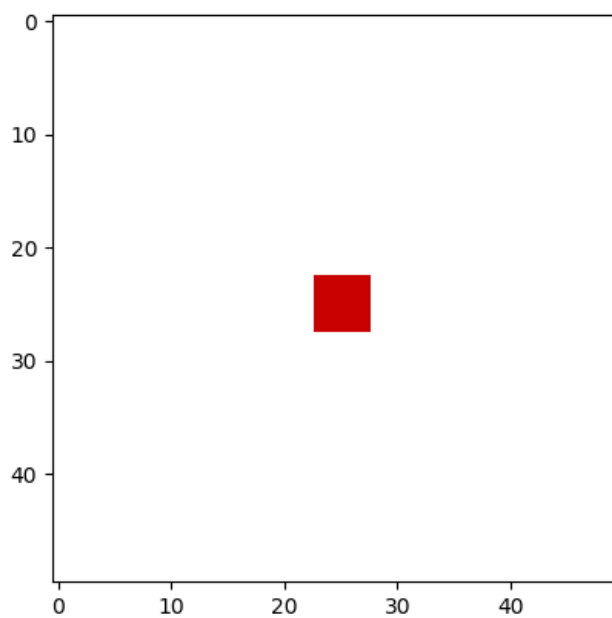


FIGURE 10 – Ebola - 2 villes

Évolution de la population en fonction du temps

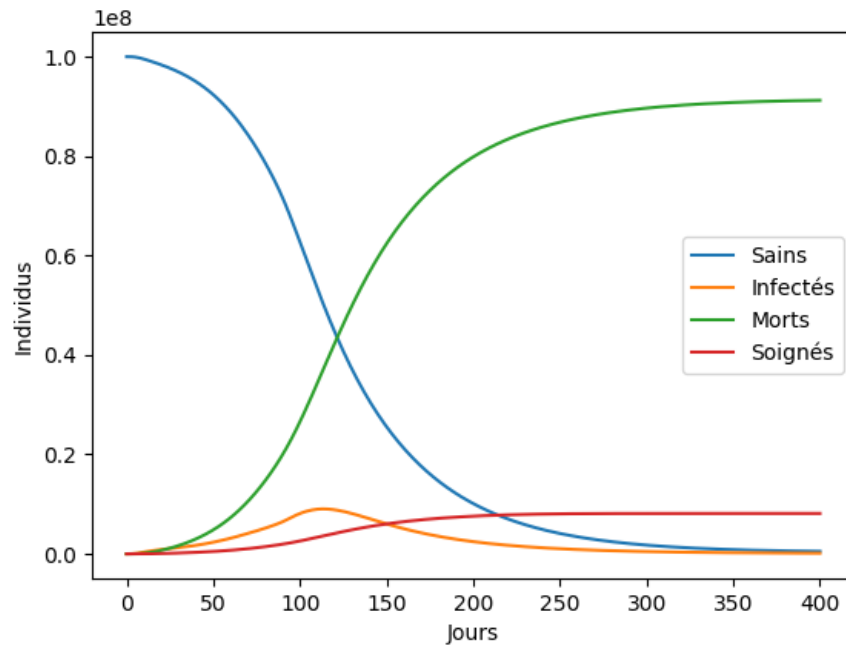
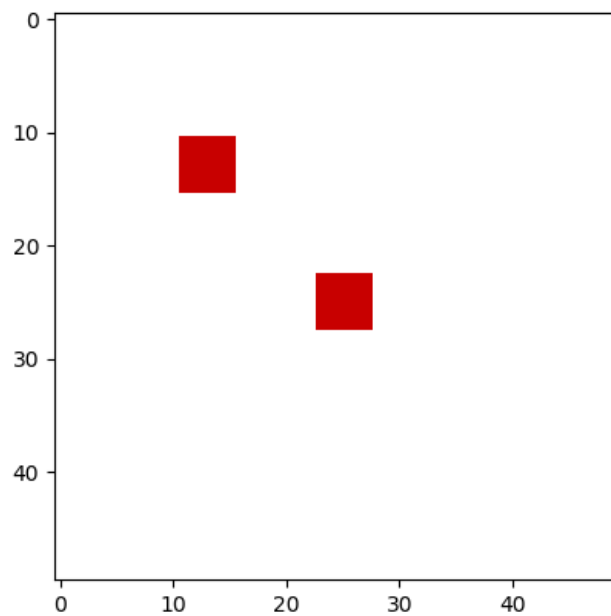
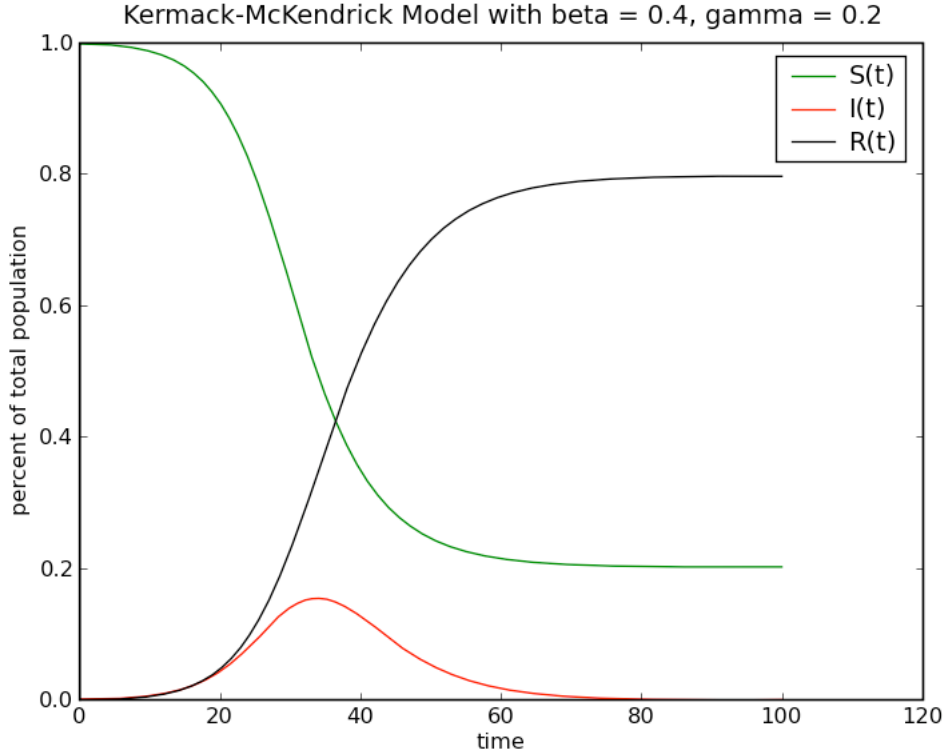


FIGURE 11 – Carte 2 villes



Ci-dessus sont présentés les résultats d'une autre étude avec un automate cellulaire, il apparaît clairement que leurs courbes sont similaires aux nôtres.



De même, les résultats d'une autre étude utilisant la méthode différentielle concordent également à ceux que nous obtenons.

3 Méthode différentielle

3.1 Équations

On peut aussi simuler les épidémies par la méthode différentielle. Nous nous sommes basés sur le modèle SIR, c'est-à-dire le modèle sains, infectés, rétablis. Les équations sont les suivantes :

$$\frac{dS(t)}{dt} = -\beta I(t)S(t) \quad (1)$$

$$\frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t) - \delta I(t) \quad (2)$$

$$\frac{dG(t)}{dt} = \gamma I(t) \quad (3)$$

$$\frac{dM(t)}{dt} = \delta I(t) \quad (4)$$

Ces équations différentielles, une fois résolues, donnent des courbes d'évolution des populations que nous présenterons plus tard.

Bien sûr, en sommant les quatre équations vues plus haut, on obtient 0, conséquence de la constance de la population au sens large du terme (incluant les morts).

$$\frac{d}{dt}(S + I + G + M)(t) = 0$$

3.2 Propriétés

Nous réintroduisons ici les paramètres de la maladie que nous avons défini plus haut, à savoir R_0 le taux de reproduction de la maladie, τ le temps moyen d'infection et S_0 la population initiale saine. Nous posons aussi $\rho = \frac{\gamma + \delta}{\beta}$. On a alors $R_0 = \frac{S_0}{\rho}$

L'utilité de R_0 est alors mise en évidence :

$$\begin{aligned}
 \frac{dI(t)}{dt} &= \beta S(t)I(t) - \gamma I(t) - \delta I(t) \\
 &= I(t)(\beta S(t) - (\gamma + \delta)) \\
 &= I(t)(\gamma + \delta)\left(\frac{\beta}{\gamma + \delta}S(t) - 1\right) \\
 &= I(t)(\gamma + \delta)\left(\frac{S(t)}{\rho} - 1\right) \\
 &= I(t)(\gamma + \delta)\left(R_0 \frac{S(t)}{S_0} - 1\right)
 \end{aligned}$$

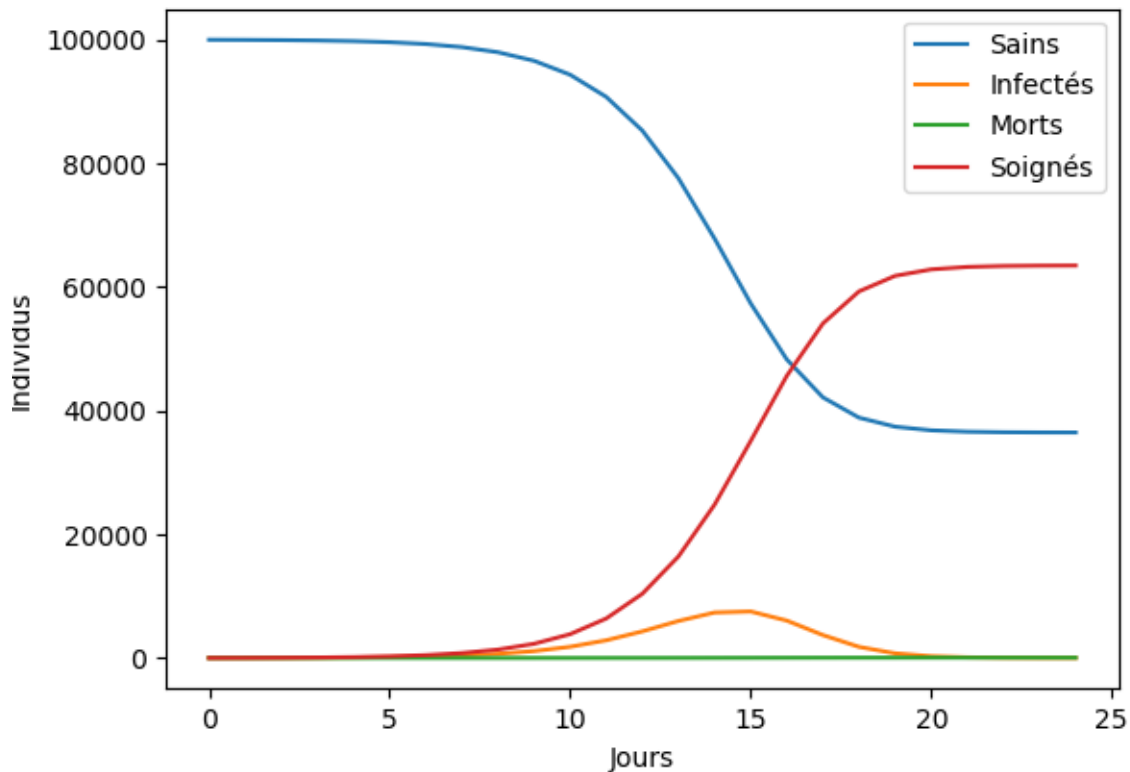
Ainsi, en 0, $\frac{dI(t)}{dt} = I(t)(\gamma + \delta)(R_0 - 1)$, ce qui montre bien que si R_0 est inférieur à 1, le nombre d'infectés va diminuer dès le début et la maladie s'éteindra alors très vite.

Il est aussi possible de chercher le maximum des courbes, en cherchant les points où les dérivées s'annulent, ce qui n'est vraiment utile que pour les infectés :

$$\begin{aligned}
 \frac{dI(t)}{dt} = 0 &\Leftrightarrow \beta S(t)I(t) - \gamma I(t) - \delta I(t) = 0 \\
 &\Leftrightarrow \beta S(t)I(t) = (\gamma + \delta)I(t) \\
 &\Leftrightarrow S(t) = \frac{(\gamma + \delta)}{\beta} \\
 &\Leftrightarrow S(t) = \rho
 \end{aligned}$$

3.3 Résultats

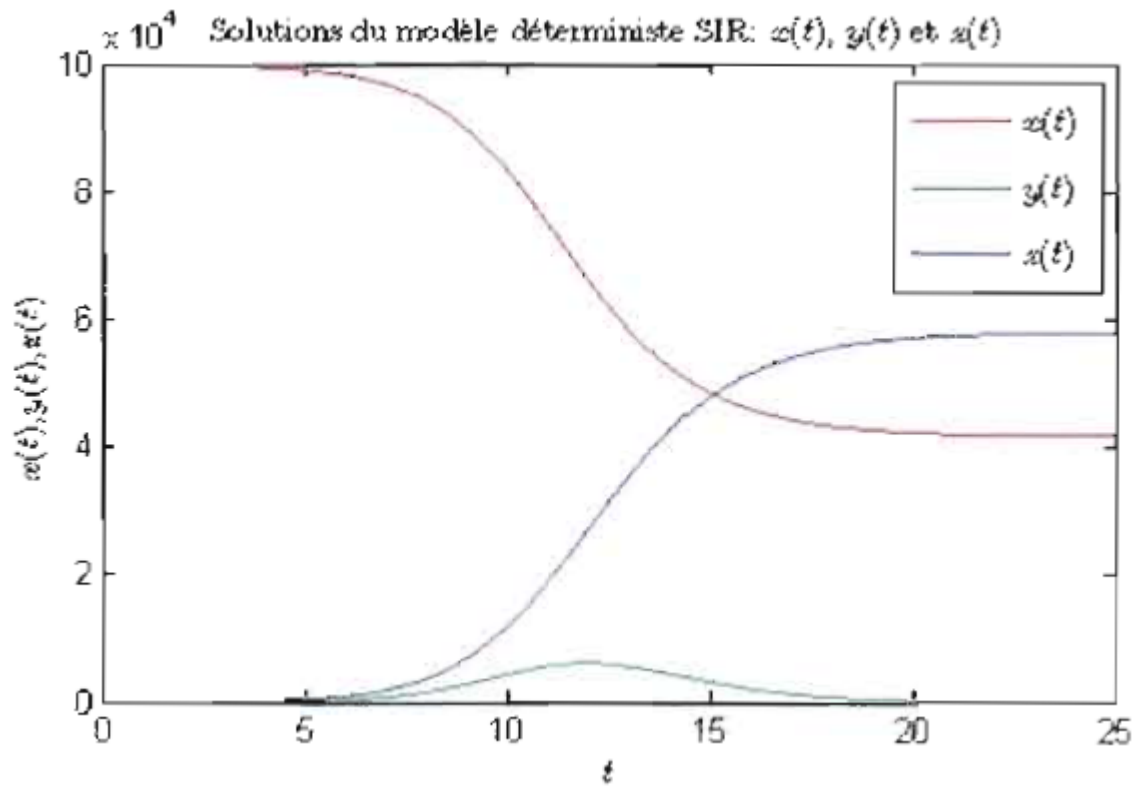
Evolution de la population en fonction du temps



Voici

un résultat de la méthode différentielle pour les probabilités de la grippe, soit $p_{\text{infection}} = 0.406$, $p_{\text{mort}} = 0.0002$, et $p_{\text{guérison}} = 0.1998$.

Après comparaison avec ce graphe effectué par une étude précédente,



nous avons observé une différence moyenne de :

- cinq pour cent pour la courbes des Sains
- un pour cent pour la courbe des Infectés
- dix pour cent pour la courbe des Morts/Guérés (somme)

Commentaire bibliographique

[4] : Cet article nous a permis de bien comprendre la théorie des automates cellulaires, et de mettre l'accent sur les paramètres intéressants qu'une simulation par automate cellulaires permettait de prendre en compte. Il présente aussi les principaux avantages des automates cellulaires par rapport à la méthode différentielle.

[2] : Cet article, en plus de présenter un autre modèle d'automates cellulaires, fournit une liste relativement exhaustive de paramètres à prendre en compte dans la simulation (même si les auteurs en ont négligé une partie qui ne présentait qu'un intérêt minime).

[1] : Cet article présente plus en profondeur les différences des autres méthodes, et fournit une approche intéressante en matière de résultats et de leur présentation.

[5] : Cet article présente la méthode différentielle, et est très fourni en graphiques bien expliqués.

[3] : Ce livre nous a fourni des éléments sur l'histoire de l'épidémiologie.

Références

- [1] Sharon Chang. Cellular automata model for epidemics.
- [2] Shih Ching Fu and George Milne. Epidemic modelling using cellular automata -introceedings of the 1st australian conference on artificial life(acal'03).
- [3] Alfredo MORABIA. A history of epidemiologic methods and concepts.
- [4] G. Rodriguez Sánchez S. Hoya White, A. Martin del Rey. Using cellular automata to simulate epidemic diseases. *Applied Mathematical Sciences, Vol. 3*, 2007.
- [5] Victor CAMEO PONZ Laura LANCE Alban MERCIER Jérémy RISSO BOURGÈS Cécile VASSEUR Aurore VIMONT. Modélisation de la propagation d'un virus.