



Concepts et modélisations de la diffusion géographique

Patrice Langlois, Eric Daudé

► To cite this version:

Patrice Langlois, Eric Daudé. Concepts et modélisations de la diffusion géographique. Cybergeog : Revue européenne de géographie / European journal of geography, 2007, 364, pp.24. <halshs-01062527>

HAL Id: halshs-01062527

<https://shs.hal.science/halshs-01062527v1>

Submitted on 10 Sep 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Concepts et modélisations de la diffusion géographique

Concepts and modeling of spatial diffusion

LANGLOIS Patrice, Université de Rouen, CNRS, UMR 6228 IDEES, équipe M.T.G, France

Patrice.langlois@univ-rouen.fr

DAUDE Eric, Université de Rouen, CNRS, UMR 6228 IDEES, équipe M.T.G, France

Eric.daude@univ-rouen.fr

Résumé

L'apparition d'une courbe en S dans l'exploitation de données spatio-temporelles est bien souvent associée par le géographe à une diffusion spatiale. L'assimilation presque systématique de la courbe en S au modèle logistique est trompeuse, elle conduit à un amalgame qui masque la diversité des modèles mathématiques permettant de modéliser des phénomènes de diffusion très différents.

L'idée de cet article est de montrer quelques exemples de diffusion qui s'expriment tous par une courbe en S mais qui correspondent en fait à des modèles différents.

Trois modèles sont exposés :

- le modèle logistique est le plus couramment utilisé par les géographes alors qu'il est par construction indépendant de l'espace. Nous comparons ici une version spatialisée et probabiliste de ce modèle à sa forme analytique et déterministe.
- à l'opposé du premier, le modèle de diffusion par front impose une forte contrainte spatiale, le contact étant prédominant dans la diffusion. Dans ce cas, la diffusion n'est pas soumise à une contrainte globale mais dépend de la longueur et de la forme du front à tout instant.
- le troisième modèle est composite, il procède à la fois de la croissance logistique contrainte dans un voisinage et de la propagation d'un front.

Ces trois modèles sont étudiés dans leurs formulations analytiques et comparés avec leurs dynamiques simulées. Nous montrons une continuité de passage entre ces 3 modèles entrant dans le cadre d'une classe plus large de modèles paramétrés. De plus la combinaison de leurs variantes produit un foisonnement d'autres modèles permettant de s'appliquer au mieux à une grande diversité de problématiques.

Mots clés : Diffusion spatiale, modèle logistique, modèle déterministe, modèle stochastique, simulation, automate cellulaire.

Abstract

The appearance of S-curve in the treatment of a space-time data is very often associated to a spatial diffusion. This systematic assimilation of the S-curve to the logistic model is misleading, it leads to an amalgam which masks the diversity of mathematical models to capture really different phenomena of diffusion.

The aims of this paper is to show some examples of diffusion which are all expressed by a S-curve but which correspond in fact to different models. Three models are presented:

- logistic model is the most usually used by geographers despite his independency from space. We compare here a spatialized and probabilistic version of this model with its analytical and deterministic form.

- as opposed to the first one, the front diffusion model imposes a strong space constraint, the contact being prevalent in the diffusion. In this case, diffusion is not subjected to a global constraint but depends on the length and on the form of the front at any moment.
- the third model is composite, it proceeds of the logistic growth forced by local constraints and of propagation of a front.

These three models are studied in their analytical forms and are compared with their dynamic simulated. We show a continuity between these 3 models entering within the framework of a broader class of parameterized models. Moreover their combinations give an expansion of other models making it possible to apply to a great variety of problems.

Keywords : Spatial Diffusion, logistic model, deterministic model, stochastic model, simulation, cellular automata

1 Introduction

La diffusion est une figure majeure des dynamiques qui s'exercent dans l'espace géographique, elle participe à la création, au maintien et à la transformation des structures spatiales. Les canaux et les formes de la diffusion sont multiples. Identifier les uns pour rendre compte des autres a donné lieu à de nombreuses recherches en géographie depuis les travaux pionniers de T. Hägerstrand [1953, 1965] et place ce champ de recherche parmi les plus importants en sciences sociales [Rogers 1995]. La diffusion d'une grandeur dans un milieu physique continu est aussi un phénomène qui a été très étudié en physique, en chimie, en biologie et dans beaucoup d'autres domaines. Des équations comme la diffusion de la chaleur ou celle de réaction-diffusion de Turing en sont des exemples significatifs, qui peuvent être utilisés dans bien d'autres domaines que celui d'origine. Mais la diffusion d'une grandeur en géographie s'opère dans un espace complexe qui est loin d'être continu et homogène. En effet, les phénomènes de diffusion empruntent des canaux qui ne se limitent pas à la diffusion par simple proximité spatiale, mais utilisent aussi tous les réseaux construits par l'homme, et qui constituent pour ces phénomènes des raccourcis d'espace pour sa propagation. Pensons simplement à la diffusion d'une épidémie qui emprunte les réseaux aériens, routiers, etc. pouvant se propager ainsi avec une grande rapidité sur la terre entière.

La richesse et la complexité des phénomènes de diffusion géographique nous pousse ainsi à approfondir et à préciser les concepts qui permettent de définir précisément ce qu'est la diffusion géographique et comment la modéliser. Ce travail a également pour objectif de rectifier quelques schémas de pensée par trop simplificateurs chez les géographes, comme par exemple que toute diffusion spatiale est caractérisée par une courbe en S, associée à la fonction logistique. Si la fonction logistique est bien représentée par une courbe en S, nous montrerons d'abord qu'il existe différents types de cette fonction et qu'il existe des phénomènes de diffusion spatiale qui ne suivent pas ce modèle.

Après avoir analysé les insuffisances du modèle logistique, modèle essentiellement non spatial, pour modéliser la diffusion, nous définissons un cadre théorique général de la diffusion géographique qui nous permet d'élaborer une classe de modèles assez générale de modèles de diffusion. L'équation de ce modèle permet de s'adapter à de nombreuses problématiques géographiques, et est destinée à être incorporée à des automates cellulaires ou des systèmes multi-agents appliqués à la géographie.

2 Définition du concept de diffusion géographique

Devant la multitude des phénomènes associés à la notion de diffusion, et le flou relatif qui en découle, nous proposons ici une définition formalisée de ce concept, qui permettra de mieux différencier les phénomènes de diffusion par rapport à d'autres phénomènes proches.

Pour pouvoir parler de diffusion géographique, il faut pouvoir expliciter une structure et une dynamique. La structure constitue la base d'invariance, de référence (même si elle contient des objets éventuellement mobiles) et la dynamique est décrite par un ensemble de règles définissant des processus d'évolution du contenu de la structure [Daudé & Langlois 2006].

2.1 La structure

La structure est ce qui reste invariant dans le système. Elle se compose de trois types d'objets différents :

- Considérons un espace dans lequel est localisé un ensemble d'objets, une population, dont les individus sont appelés *hôtes*, parce qu'il vont « accueillir » le phénomène de diffusion. Les hôtes possèdent un certain nombre de variables d'état. L'état global d'un hôte est défini par le vecteur des valeurs d'état des différentes variables. Une de ces variables permet de caractériser l'avancement du phénomène. Celle-ci peut être binaire : 0 : non-acquis, 1 : acquis. On utilise parfois trois valeurs d'état, ou plus, comme dans le modèle SIR : susceptible (S), contagieux (I), guéris (R). Ces états n'ont pas un statut symétrique, car le processus est irréversible. Tout hôte peut passer de l'état 0 à l'état 1 mais pas le contraire. De même, dans le modèle SIR, on a les transitions $S \rightarrow I \rightarrow R$, mais pas les transitions inverses.

- Le *vecteur* de la diffusion est composé des particules qui sont les *agents* de la diffusion. Ils constituent le médiateur de l'interaction entre les hôtes.

- Le *canal* est formalisé par les arcs d'un graphe orienté dont les sommets sont les hôtes. Ces arcs constituent le support de propagation des agents entre les hôtes. Le canal est rarement matérialisé, il constitue souvent l'ensemble des liens potentiels entre les hôtes. Ainsi, il représente les liens de voisinage lorsque les hôtes sont des lieux ou le réseau relationnel entre des personnes.

Le tableau 1 présente trois exemples d'utilisations des termes relatifs à cette structure.

	Epidémie	Incendie	Innovation
Hôtes	Personnes	Arbres	Villes
Etats	Sain / infecté (SIR)	Sain / brûlé	Nombre d'adoptants
Canal	Contacts interpersonnels	Graphe de proximité	Liens entre villes
vecteur	Agents infectieux	Chaleur et flammes	Information

Tableau 1 : exemples d'utilisation des termes

2.2 La dynamique

La dynamique opère localement selon trois *processus* qui interagissent pour produire le processus global de diffusion: celui d'*acquisition* qui formalise l'entrée des agents dans les hôtes, celui d'*émission* qui concerne la manière dont les agents sortent des hôtes, et celui de *propagation* des agents dans le canal pour changer d'hôtes.

- **Le processus d'acquisition** agit sur les hôtes. Il décide du changement d'état d'un hôte en fonction du nombre d'agents qui viennent de lui arriver, en fonction de son état global et éventuellement de son environnement. On définit aussi un **processus global d'acquisition** qui permet d'observer le phénomène global de diffusion. Il est représenté, à chaque instant, par le cumul des individus ayant acquis l'état fondamental (celui qui intéresse le modélisateur), par exemple l'état 1 en mode binaire, pour ne plus le quitter. L'effectif d'acquisition est donc une fonction non-décroissante du temps, c'est un processus **irréversible**.

- **Le processus d'émission** concerne les agents. Il est représenté par le nombre d'agents qui sortent d'un hôte. La fonction de propagation calcule, pour chaque hôte et à chaque instant, le nombre d'agents qui sortent de l'hôte pour aller dans le canal.

- **Le processus de propagation** définit le comportement des agents dans le canal. Il calcule comment se distribuent les agents sortis des hôtes vers d'autres hôtes. Ce processus caractérise l'interaction entre les hôtes.

Ainsi les trois processus interagissent pour donner la dynamique de la diffusion [fig. 1].

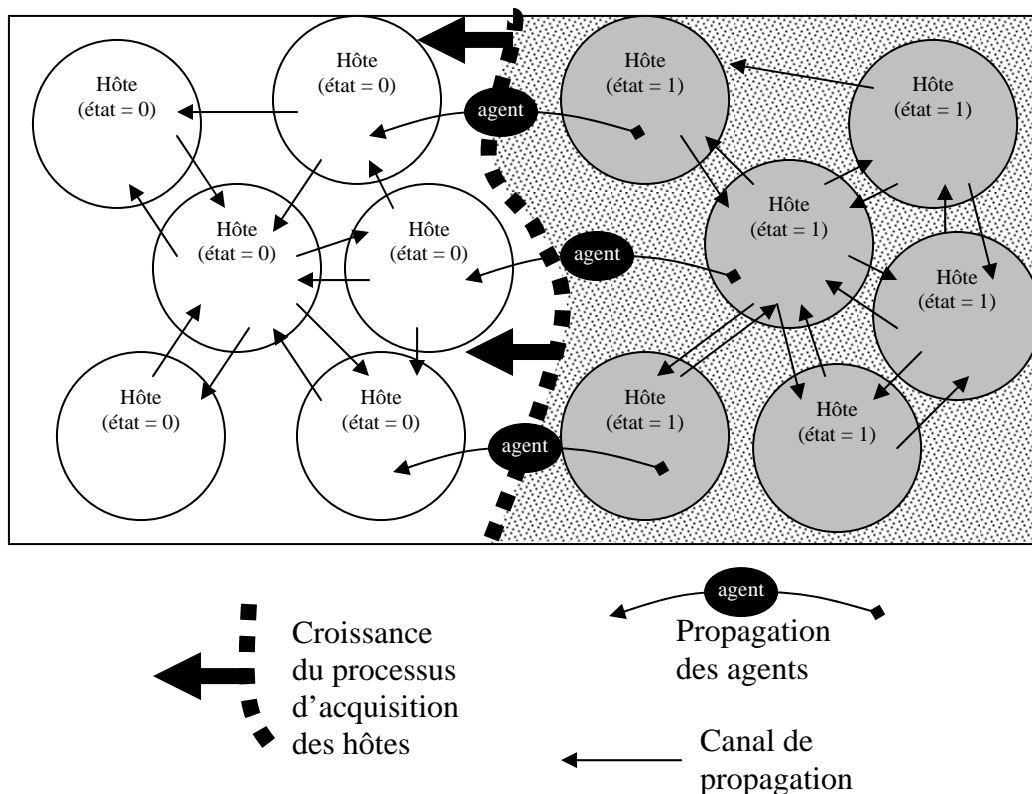


Figure 1 : les 2 processus en jeu dans la diffusion

2.3 Exemples de phénomènes de diffusion géographique

Dans certains cas c'est plutôt le processus de propagation qui est privilégié et que l'on assimile à la diffusion. C'est le cas de la **diffusion d'un nuage toxique** qui possède le plus souvent une source localisée unique, la totalité du composé toxique étant en plus produite dans un court laps de temps. Le processus de diffusion est alors conservatif, mais le volume de produit se disperse dans l'espace jusqu'à arriver à un seuil de densité au-dessous duquel il n'est plus sensible. Le processus d'acquisition concerne ici l'espace qui peut être pollué ou non par ce nuage. Il faut définir (par exemple) quelle est la densité de produit dans

l'atmosphère à partir de laquelle chaque lieu est considéré comme pollué par le produit, selon la quantité qui se dépose au sol (règle d'acquisition). De même en chaque lieu et à chaque instant, il faut définir comment le produit se dissipe dans les lieux voisins, en fonction de la densité du produit restant, de la vitesse du vent, etc. (règle d'émission).

Dans le cas d'un *incendie de forêt*, les hôtes sont les arbres, les agents sont les flammes (ou la chaleur) qui progressent d'arbre en arbre. Un arbre ne communique le feu que pendant un temps assez court, et sur une ligne de front plus ou moins complexe. Le processus d'acquisition (passage de l'état 0 « sain » à l'état 1 « brûlé ») dépend de son essence, son état de sécheresse (vecteur d'états) de la chaleur ambiante transmise par les autres arbres en flamme, etc. Le processus de propagation du vecteur « feu » est fonction de la distance entre les arbres, de la direction du vent, etc.

Mais la plupart du temps, et particulièrement dans un contexte social ou humain, le processus de propagation est mal connu et sa dynamique difficile à formaliser. C'est le cas de la *diffusion d'une épidémie*. Dans ce cas, on définit plutôt la propagation (ici la contagion) de manière probabiliste, en définissant la probabilité qu'une personne infectée communique l'agent infectieux à une personne saine, lors d'un contact personnel. Cette probabilité dépend du cycle de développement de la maladie, c'est-à-dire du cycle d'activité des agents. Contrairement au nuage toxique, la quantité d'agents infectieux n'est pas conservative, dans un premier temps, elle se reproduit activement, donc s'accroît et envahit de plus en plus l'hôte, pour décroître ensuite et disparaître ou rester endémique.

Notons que le modèle peut être beaucoup plus complexe si l'on est amené à prendre en compte plusieurs niveaux d'interaction. On est alors en présence d'un emboîtement de processus à différents niveaux d'organisation, liant hôtes et agents : en effet dans le cas du paludisme, l'agent infectieux primaire, l'hématozoaire, avant de parasiter l'homme, a pour hôte l'anophèle, qui lui-même est un agent pour infecter l'homme. Mais si l'on cherche à modéliser la diffusion du paludisme sur un territoire, l'homme, qui est l'hôte pour le repas de l'anophèle, est aussi l'agent de propagation entre les lieux de ce territoire. La complexité d'un tel système peut néanmoins se gérer par l'emboîtement de plusieurs processus de diffusion bien identifiés, à condition de bien différencier, pour chaque niveau de processus, ses agents, ses hôtes et les règles d'interaction qui s'y appliquent.

Dans le cas de la *diffusion d'une innovation*, comme le téléphone portable, le principe est assez semblable à celui d'une épidémie. L'hôte est toute personne susceptible d'acquérir le produit, le vecteur de propagation est constitué des contacts verbaux, des messages entre personnes ayant adopté et personnes adoptantes potentielles. La dynamique d'émission des messages peut suivre un cycle d'activité comparable à celui d'un agent pathogène. En effet, après l'enthousiasme initial résultant de l'acquisition, l'hôte s'habitue, commence à en déceler les inconvénients, l'effet de mode passe, et le nombre et la force des messages diminuent ou s'arrêtent au bout d'un certain temps, selon le processus d'émission. Un autre type d'influence dans la diffusion d'une innovation est constitué des médias (publicité, information, etc.). Mais ceci n'est pas un processus de propagation entre les personnes, car il agit depuis l'extérieur du système (les organes de presse, les stations radio ou télé). Les médias agissent pratiquement vers tout le monde de manière quasi-uniforme (même si chacun n'est pas également réceptif). S'il doit être pris en compte, ce type d'influence n'entre donc pas dans la définition stricte du modèle de diffusion, c'est une contrainte supplémentaire, exogène.

3 Formalisation du concept de diffusion

La définition générale donnée doit permettre de modéliser mathématiquement ces différents exemples dans un même cadre théorique, que nous présentons dans les sections suivantes.

Chaque processus se modélise dans le cadre d'un automate cellulaire (AC) ou d'un système multi-agents (SMA), par une règle (représentée par un rectangle) qui est applicable dans chaque agent (ou cellule). Cette règle code informatiquement la fonction qui transforme certaines données (les entrées de la règle, représentées par les flèches entrant sur le côté gauche du rectangle) en d'autres données (les sorties de la règle, représentées par les flèches sortant sur le côté droit du rectangle). La règle dépend aussi de facteurs environnementaux (contrôles, entrant sur le côté supérieur du rectangle).

3.1 Les objets

Nous utilisons deux catégories d'objets, l'une constituant l'ensemble des hôtes P , l'autre celui des agents V , avec :

$P = \{1, 2, \dots, n\}$ qui représente la population des hôtes. L'effectif de cette population est constant et la localisation des hôtes est fixe dans l'espace (ensemble des cellules de l'automate) ou variable (agents d'un SMA).

V représente l'ensemble des agents qui circulent dans le canal, il peut varier au cours du temps.

3.2 Le canal, les entrées, les sorties

Les agents circulent dans le canal qui est relié aux hôtes par l'intermédiaire d'entrées et de sorties.

En mode déterministe, les agents ne sont pas individualisés, ils sont représentés par un flux continu. Ainsi, les entrées (notées I) et les sorties (notées O) sont des variables quantitatives représentant ces flux.

En mode stochastique, les agents sont individualisés et I et O sont alors des ensembles (des listes) d'agents, qui circulent dans le réseau cellulaire.

On a donc pour chaque instant t :

$I(i, t)$: ensemble (ou quantité) des agents entrant dans l'hôte i à l'instant t

$O(i, t)$: ensemble (ou quantité) des agents sortant de l'hôte i à l'instant t

J : matrice des coefficients du canal (réels entre 0 et 1) qui définit la structure d'interconnexion entre les cellules. Pour toute ligne i de cette matrice, on a

$$\sum_{j=1}^n J_{ij} = 1$$

Où les termes J_{ij} de cette matrice représentent soit :

- en mode déterministe, les coefficients de redistribution du flux d'agents sortant d'une cellule i vers les entrées des autres cellules qui lui sont rattachées ;
- en mode stochastique, la probabilité pour qu'un agent sortant de la cellule i entre dans une cellule j . La matrice de transition de probabilité est J .

3.3 Les états

L'état de la $i^{\text{ème}}$ cellule-hôte à l'instant t est noté $s(i, t)$

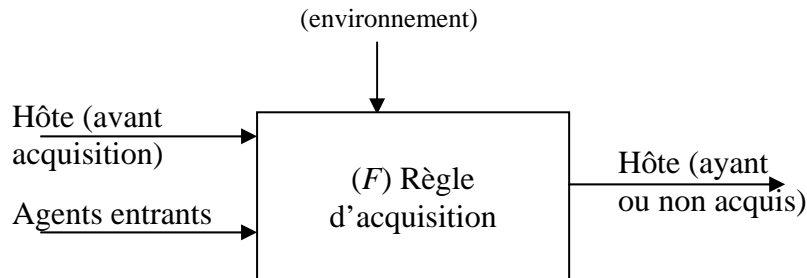
3.4 Les processus

Le modèle de diffusion implique de formaliser trois règles qui sont associées aux éléments décrits ci-dessus :

Règle locale d'acquisition F :

C'est une règle de transition qui définit le changement d'état de l'hôte i :

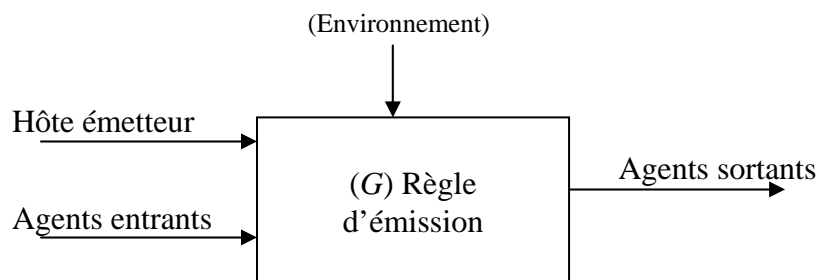
$$s(i, t+1) = F(I(i, t), s(i, t))$$



3.4.1 Règle locale d'émission G

Cette règle détermine le flux d'agents sortant de l'hôte :

$$O(i, t+1) = G(I(i, t), s(i, t))$$



Les deux règles F et G définissent donc un automate de Moore sur chaque cellule.

3.4.2 Règle de propagation H

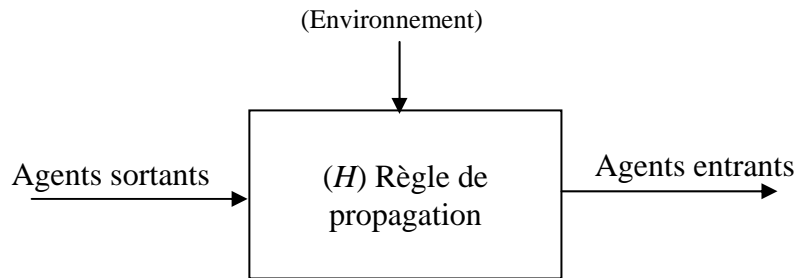
Cette règle définit le comportement des agents dans le canal, elle calcule les entrées dans les hôtes en fonction des sorties.

En mode déterministe, cette règle est une multiplication matricielle :

$$I(i, t+1) = \sum_{j=1}^n O(j, t+1) J_{ji}$$

En mode stochastique, la règle de propagation est un tirage de Monte-Carlo.

On pose $p_0^j = 0$ et $p_i^j = \sum_{m=1}^i J_{jm}$ et pour chaque agent de la liste $O(j, t+1)$, on tire un nombre p dans $]0, 1]$, si $p \in]p_{i-1}^j, p_i^j]$ on transmet alors l'agent à la liste d'entrée de la cellule i .



3.4.3 Processus global de croissance u

Pour représenter globalement le processus de diffusion, on mesure d'abord la croissance du processus d'acquisition qui est modélisé par le cumul des hôtes ayant acquis l'état fondamental. En mode binaire, l'état fondamental étant 1, ce cumul à l'instant t vaut :

$$u(t) = \sum_{(i,j) \in I \times J} s(i, j, t)$$

On peut aussi représenter le cycle d'activité des agents dans les hôtes. En situation déterministe, les agents sont déjà globalisés en une variable représentant la quantité d'agents. Leur activité peut être définie par une simple règle multiplicative, en supposant qu'elle est constante $a(i, t) = k \cdot I(i, t)$. En situation stochastique, l'activité peut être individualisée. L'entrée $I(i, t)$ est alors constituée à chaque instant par la liste des m agents entrant dans l'hôte, $I(i, t) = \{a_j(i, t)\}_{j=1, \dots, m}$ et l'activité globale des agents dans la cellule i est donnée par la somme:

$$a(i, t) = \sum_{j=1}^m a_j(i, t)$$

Dans une épidémie par exemple, l'agent pathogène possède une activité qui peut dépendre d'un cycle de vie de la maladie de l'hôte. Son activité infectieuse est alors limitée dans le temps, à partir de l'instant t_j où il a été infecté. a_j est alors une fonction à support positif $f_j(t - t_j) = a_j(i, t)$, qui dépend du temps de développement de la maladie :

$$a(i, t) = \sum_{j=1}^m f_j(t - t_j)$$

On pourrait modéliser de la même manière ce processus dans le canal, il représente alors l'activité des agents pathogènes lorsqu'ils sont en dehors des hôtes.

Après avoir présenté le concept général de diffusion et sa formalisation à travers un modèle générique, nous allons préciser maintenant les modèles globaux de croissance les plus courants, répondant à un minimum d'hypothèses réalistes. Nous analyserons plus finement le modèle logistique qui possède des propriétés intéressantes, dont le comportement peut devenir chaotique dans certaines conditions. A partir des limites constatées de ces modèles globaux, nous verrons en quoi ce modèle générique peut nous aider à formaliser des modèles locaux plus spécifiquement géographiques.

4 Les modèles de croissance

Les modèles de croissance sont parmi les modèles les plus utilisés en géographie [Saint-Julien T. 1985], [Dauphiné A. 2003]. Bien que simples dans leur formulation, leur non-linéarité nous plonge déjà dans les sciences de la complexité. Nous rappelons d'abord les définitions formelles et la signification physique d'un modèle de croissance soit dans un contexte de ressource infinie, soit dans un contexte de ressource limitée. En particulier, dans ce dernier cas, nous insistons sur le fait qu'il existe deux modèles mathématiques distincts : la *suite logistique* (temps discret) et la *fonction logistique* (temps continu).

Bien que récent par rapport à la progression géométrique qui remonte aux grecs, le modèle logistique n'est pas nouveau. Il est étudié pour la première fois par le mathématicien belge Pierre-François Verhulst vers 1840 [Verhulst P. 1838], et porte parfois le nom de modèle de Verhulst. Il a souvent servi à caractériser les modèles de diffusion en géographie, même s'il est par construction indépendant de l'espace. En effet, c'est un modèle de croissance d'une grandeur qui « s'alimente » dans un stock limité de ressource, comme la croissance d'une population, les arbres non brûlés pour un incendie de forêt, les personnes non infectées pour une épidémie, les consommateurs en état d'acheter pour la vente d'un bien, les adoptants potentiels pour une innovation, etc.

4.1 Modèle en ressource infinie

Utilisant la métaphore épidémique, appelons $u(t)$ le nombre d'individus (hôtes) infectés à l'instant t dans une population. Lorsque la *ressource est inépuisable* ou très grande devant la valeur de $u(t)$, le taux de croissance r est supposé constant et la variation de u ne dépend que du nombre d'individus déjà infectés. Dans ce cas, le nombre de nouveaux malades à chaque pas de temps est proportionnel au nombre d'infectés déjà présents. On aboutit ainsi, pour un temps discret, à une *progression géométrique* où l'accroissement $\Delta u(t) = u(t+1) - u(t)$ est proportionnel à $u(t)$, ce qui donne une relation de la forme

$$u(t+1) = u(t) + r.u(t) \quad (1)$$

La résolution de cette équation ne dépend que de l'état initial de u_0 et du taux de croissance r :

$$u(t) = u_0(1+r)^t \quad (2)$$

Dans le cas d'un temps continu, cela s'exprime par une équation différentielle :

$$\frac{du}{dt} = \alpha.u$$

dont la solution est une *fonction exponentielle* :

$$u(t) = \exp(\alpha.t + \beta) \quad (3)$$

En posant $u_0 = e^\beta$ et $r = e^\alpha - 1$, on retrouve exactement la formulation (2), la courbe en temps continu passe donc exactement par les points de la progression géométrique de croissance r , correspondant aux valeurs de t entières, donnée en (2). Le modèle discret est compatible avec le modèle continu, le premier n'étant que la restriction aux valeurs entières de t du second.

4.2 Modèle en ressource limitée

Utilisons encore la métaphore de l'épidémie pour construire le modèle de croissance lorsque la *ressource est limitée*, ce qui est en général plus réaliste. On se donne une population de taille constante u_{\max} . Dans cette population, il y a un nombre $u(t)$ d'individus infectés à l'instant t et u_0 à l'instant initial. On suppose qu'à chaque pas de temps, une

proportion p constante des $u(t)$ individus infectés sont contagieux (émetteurs) et une proportion p' constante des $u_{\max} - u(t)$ individus sains sont contaminables (récepteurs). Enfin une proportion p'' des couples émetteurs-récepteurs vont se former pour transmettre la maladie.

En somme, à chaque pas de temps, il y a une proportion $r = p.p'.p''$ parmi les $u(t).(u_{\max} - u(t))$ liens possibles qui vont provoquer une contamination. Nous appelons r le **taux d'interaction** entre les individus.

On obtient ainsi le modèle itératif en temps discret suivant, appelée ici *suite logistique*¹ et qui permet de calculer la valeur de u à un instant donné à partir de sa valeur à l'instant précédent. Cette formulation est très bien adaptée à la simulation :

$$\begin{cases} u(0) = u_0 \\ u(t+1) = u(t) + ru(t).(u_{\max} - u(t)) \end{cases} \quad (4)$$

Les valeurs de la suite dépendent de la valeur initiale u_0 qui donne le nombre de personnes infectées à l'instant initial. Les autres termes de la suite sont dépendants de cette condition initiale, ainsi que du taux d'interaction r . Remarquons que si u_0 est nulle, la suite reste identiquement nulle. Elle ne pourra « décoller » qu'avec une valeur positive de u_0 , aussi petite soit-elle.

Remarquons aussi que le nombre de liens potentiels $l = u(t).(u_{\max} - u(t))$ entre les individus peut être élevé. Il est maximum lorsque la moitié de la population est contaminée, il vaut alors

$l = \frac{u_{\max}^2}{4}$. Cette valeur dépasse largement la population totale u_{\max} . Par exemple pour $u_{\max} = 1000$ individus, le nombre de liens potentiels peut atteindre $l = 250\,000$. Rien n'empêche alors, sauf si le taux d'interaction r est suffisamment petit, un accroissement de $u(t)$ en une seule itération qui lui fasse dépasser le seuil de population u_{\max} . Ceci explique que ce modèle puisse présenter un comportement chaotique. Nous analyserons plus loin cette situation, sachant bien qu'un tel modèle utilisé pour la croissance d'une épidémie, ne doit pouvoir produire, à aucun moment, une population infectée supérieure à la population totale !

Lorsque $u_0 > 0$ et que le taux r est suffisamment petit², la croissance de la suite se traduit par une courbe en S. Elle augmente avec le temps jusqu'à devenir maximale lorsque la moitié de la population est infectée, puis la croissance se ralentit ensuite, $u(t)$ ne pouvant atteindre u_{\max} .

On peut utiliser une formulation équivalente de ce modèle, en considérant non plus le nombre d'individus infectés mais la proportion $w(t) = u(t)/u_{\max}$ d'individus infectés dans la population. Il suffit pour cela de diviser les deux membres de l'équation (4) par u_{\max} . Apparaît alors le produit $r.u_{\max}$ que l'on nomme coefficient d'interaction $s = r.u_{\max}$:

$$\begin{cases} w(0) = w_0 = u_0 / u_{\max} \\ s = r.u_{\max} \\ w(t+1) = w(t) + s.w(t).(1 - w(t)) \end{cases} \quad (5)$$

¹ On trouve souvent dans la littérature la dénomination de *fonction logistique* pour la fonction $f(x) = rx(1-x)$ et de *suite logistique* pour la suite $u_{n+1} = f(u_n) = ru_n(1-u_n)$. Dans notre exposé, nous utilisons la forme cumulative de ce modèle : $u_{n+1} = u_n + ru_n(1-u_n)$, qui est un peu différente mais dont le comportement s'en déduit facilement. De plus, nous appelons dans cet article *fonction logistique* (ou modèle logistique continu) la solution de l'équation différentielle (5) et non la fonction f donnée ici.

² Voir inégalité (13)

Si l'on veut construire un modèle avec un pas plus fin, défini par un réel positif Δt , on peut écrire l'équation (4) de manière un peu plus générale:

$$\begin{cases} u(0) = u_0 \\ u(t + \Delta t) = u(t) + ru(t)(u_{\max} - u(t))\Delta t \end{cases} \quad (6)$$

Remarquons que la suite obtenue avec $\Delta t = 0,2$ ne fait pas coïncider un point sur deux de la suite avec $\Delta t = 0,4$.

Dans le cas d'un temps continu, à partir de (6), on déduit la variation de la fonction $u(t)$ pour un pas de temps Δt :

$$\frac{u(t + \Delta t) - u(t)}{\Delta t} = ru(t)(u_{\max} - u(t))$$

Lorsque Δt devient infiniment petit, ce rapport est la dérivée de $u(t)$, notée $\frac{du}{dt}$, on obtient ainsi, *l'équation différentielle logistique* :

$$\frac{du}{dt} = \alpha u(u_{\max} - u) \quad (7)$$

où α est le taux de croissance instantané. Cette équation possède une solution exacte, appelée ici *fonction logistique* (voir note 1). Elle est donnée par l'une ou l'autre des expressions:

$$u(t) = \frac{u_{\max}}{1 + C \exp(-\alpha u_{\max} t)} \quad (8)$$

$$u(t) = \frac{u_{\max}}{1 + \exp(-\alpha u_{\max} t + \beta)} \quad (9)$$

où les constantes C et β dépendent de la condition initiale $u_0 = u(0)$ qui ne doit pas être nulle, et LN représente le logarithme népérien: $C = \frac{u_{\max} - u_0}{u_0}$ et $\beta = \text{LN}(C)$

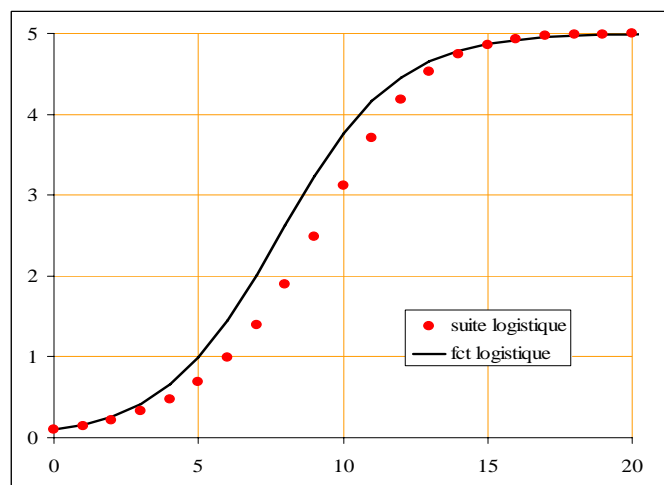


Figure 1 : fonction et suite logistique
($u_{\max} = 5$ et $u_0 = 0,5$ et $r = \alpha = 1$)

Dans le cas particulier où $u_{\max} = 1$ et $u_0 = 0,5$ cela donne un coefficient C qui est égal à 1, ce qui simplifie la formule. Cette logistique particulière, qui ne dépend plus que d'un seul

paramètre α s'appelle *logit*. Elle joue un rôle important en science car elle apparaît dans de nombreux domaines scientifiques (physique, statistique, théorie de l'information, etc.). Elle s'écrit :

$$\text{logit}_\alpha(t) = \frac{1}{1 + e^{-\alpha t}} \quad (10)$$

Remarquons que toute fonction logistique $u(t)$ se ramène à une *logit* par changement d'unités ou changement d'origine des temps.

4.2.1 Comparaison entre suite et fonction logistique

Les différences entre suite logistique (4) et fonction logistique (8), sont de trois ordres :

- La première est évidente : la suite logistique est à temps discret alors que la fonction est à temps continu.

- La deuxième porte sur le caractère explicite ou non de la fonction³ : dans la suite logistique, la fonction $u(t)$ n'est pas connue explicitement, elle n'est calculable que de manière itérative. Il faut avoir calculé tous les termes de la suite $u_0, u(1), u(2), u(3), \dots, u(t-1)$ pour pouvoir calculer $u(t)$. Ce n'est pas le cas de la fonction logistique puisqu'on dispose d'une formule directe pour calculer $u(t)$ en fonction de toute valeur réelle de t .

- La dernière différence est la plus importante : à partir de la même valeur initiale u_0 , les deux modèles ne coïncident ensuite en aucun de leur point, même avec le même taux de croissance. Néanmoins, on pourrait penser, comme c'est le cas pour le modèle exponentiel, qu'en choisissant α judicieusement par rapport à r , la fonction logistique obtenue puisse passer exactement par les points de la suite, or il n'en est rien. La fonction logistique définie en (8) n'est jamais une solution exacte de $u(t)$ vérifiant la formule (4) ou (5).

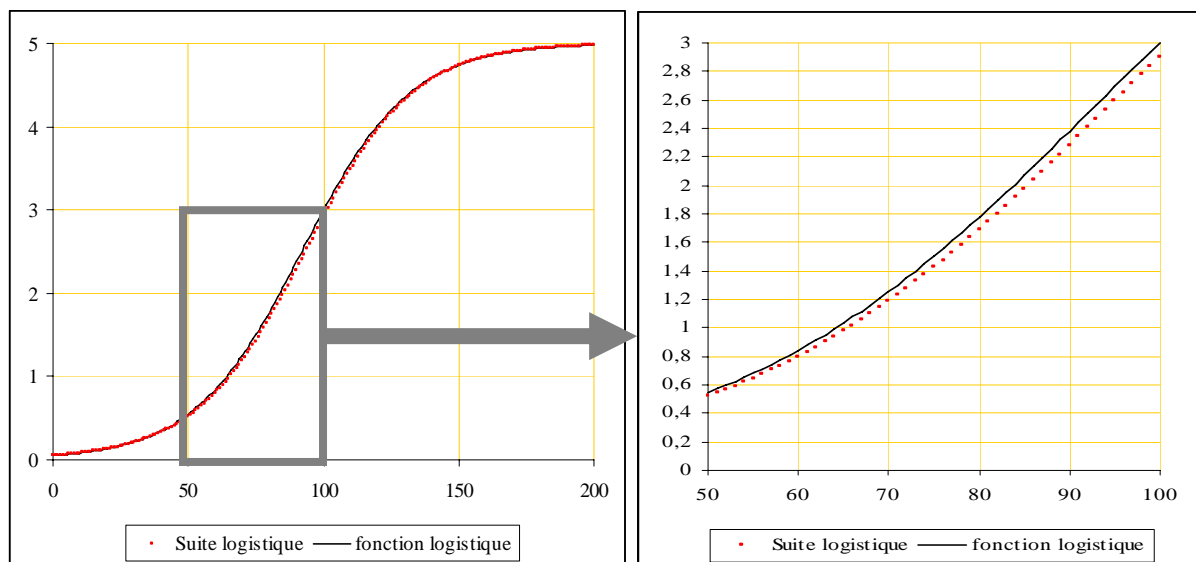


Figure 2 : Comparaison entre suite logistique et fonction logistique

($u_{\max} = 5$, $u_0 = 0,05$ et $r = 0,01$).

Nous ne connaissons la solution explicite $w(t)$ de la suite logistique sous sa forme (5) que dans un seul cas particulier, celui où $s = 1$. (ce qui correspond à $r = 1/u_{\max}$ dans le modèle (4)).

³ Une fonction explicite est de la forme $y = f(x)$, une fonction implicite s'exprime par une expression du type $f(x, y) = 0$

Dans cette situation, la formule (5) se simplifie et permet d'obtenir une expression de la fonction $w(t)$ qui ne dépend que de la condition initiale w_0 et du temps t :

$$w(t) = 1 - (1 - w_0)^{2^t} \quad (11)$$

Il faut remarquer que cette solution n'est pas une fonction de type logistique. On peut simplifier son expression si l'on appelle $v(t) = 1 - w(t)$ et $v_0 = 1 - w_0$, la fonction $v(t)$ correspond alors à la proportion d'individus sains. Avec ces notations, la formule (11) donne une expression particulièrement simple :

$$v(t) = v_0^{2^t} \quad (12)$$

Malgré ces différences importantes, les deux modèles (discret et continu), même en prenant $\alpha = r$, restent assez proches dans leur comportement, à condition que le taux d'interaction r reste suffisamment petit, comme nous le préciserons dans le paragraphe suivant. La différence dépend aussi de la valeur initiale u_0 . Par exemple, pour $u_{\max} = 5$ et $r = 0,01$, avec $u_0 = 0,5$, l'erreur relative ne dépasse pas 1.5% entre les deux courbes, alors que pour $u_0 = 0,05$ elle monte à 5,35% .

4.2.2 Le comportement chaotique de la suite logistique

Nous avons vu que la suite logistique telle que décrite en (5) modélise l'évolution en temps discret d'une grandeur $w(t)$ associée à une population limitée. Dans le contexte d'une épidémie, $w(t)$ désigne la proportion d'individus infectés à l'instant t , le taux d'interaction r indique la proportion des contacts entre individus infectés et individus sains qui se traduisent par des contaminations effectives. Nous montrons dans ce qui suit que la suite logistique ne devient chaotique que dans des circonstances qui sont hors des conditions normales de modélisation. Une absence de précaution dans le choix de la discrétisation peut malgré tout déboucher sur un comportement chaotique du modèle. Par exemple, en faisant un raisonnement grossier, s'il y a en moyenne une infection nouvelle par jour dans une population et si on prend un pas de temps de un an pour discrétiser le phénomène, cela représente en moyenne 365 infections par an. Or s'il n'y a que 100 individus, ce mode de discrétisation entraînera dès la première année plus d'infectés que d'individus. Donc, lorsque le pas de temps choisi est disproportionné par rapport à la dynamique du phénomène, le taux d'infection est tellement élevé que la suite devient chaotique. Notre objectif est donc de fixer une limite claire au modélisateur, (dans la formule 13) de manière à ce qu'il puisse se rendre compte à travers le choix du taux d'infection r , s'il reste dans des conditions normales d'utilisation de son modèle et en particulier si le choix de son pas de temps est correct.

La première condition de réalisme est que le nombre de malades ne puisse jamais dépasser la population totale... Ainsi, quel que soit t , $w(t)$ ne doit jamais dépasser 1. Cette condition se traduit, à partir de (5) par l'inégalité

$$(s + 1)w(t) - s.w(t)^2 \leq 1$$

En étudiant le signe du trinôme : $sw^2 - (s+1).w + 1$, on montre alors qu'il faut que $s \leq 1$, ce qui impose la **condition de stabilité** suivante sur r :

$$\boxed{w(t) \leq 1 \Rightarrow r \leq 1/u_{\max}} \quad (13)$$

Ainsi, r doit rester inférieur à $1/u_{\max}$ pour que le nombre d'infectés ne dépasse jamais la population totale. Cette contrainte est assez restrictive, mais convient à une utilisation

normale de modélisation. Le comportement chaotique peut apparaître d'autant plus facilement que u_{\max} est grand. Par exemple, pour une population maximale de $u_{\max} = 3000$ individus, une valeur de $r = 0,001$ donne déjà un comportement totalement chaotique. Il faut prendre r inférieur à $1/3000$, par exemple $r = 0,0001$ pour garantir un comportement stable.

Lorsque la valeur de r dépasse $1/u_{\max}$, le régime devient pseudo-périodique, puis chaotique pour des valeurs de r encore supérieures. Cette curiosité a été étudiée en 1975 par le physicien Feigenbaum.

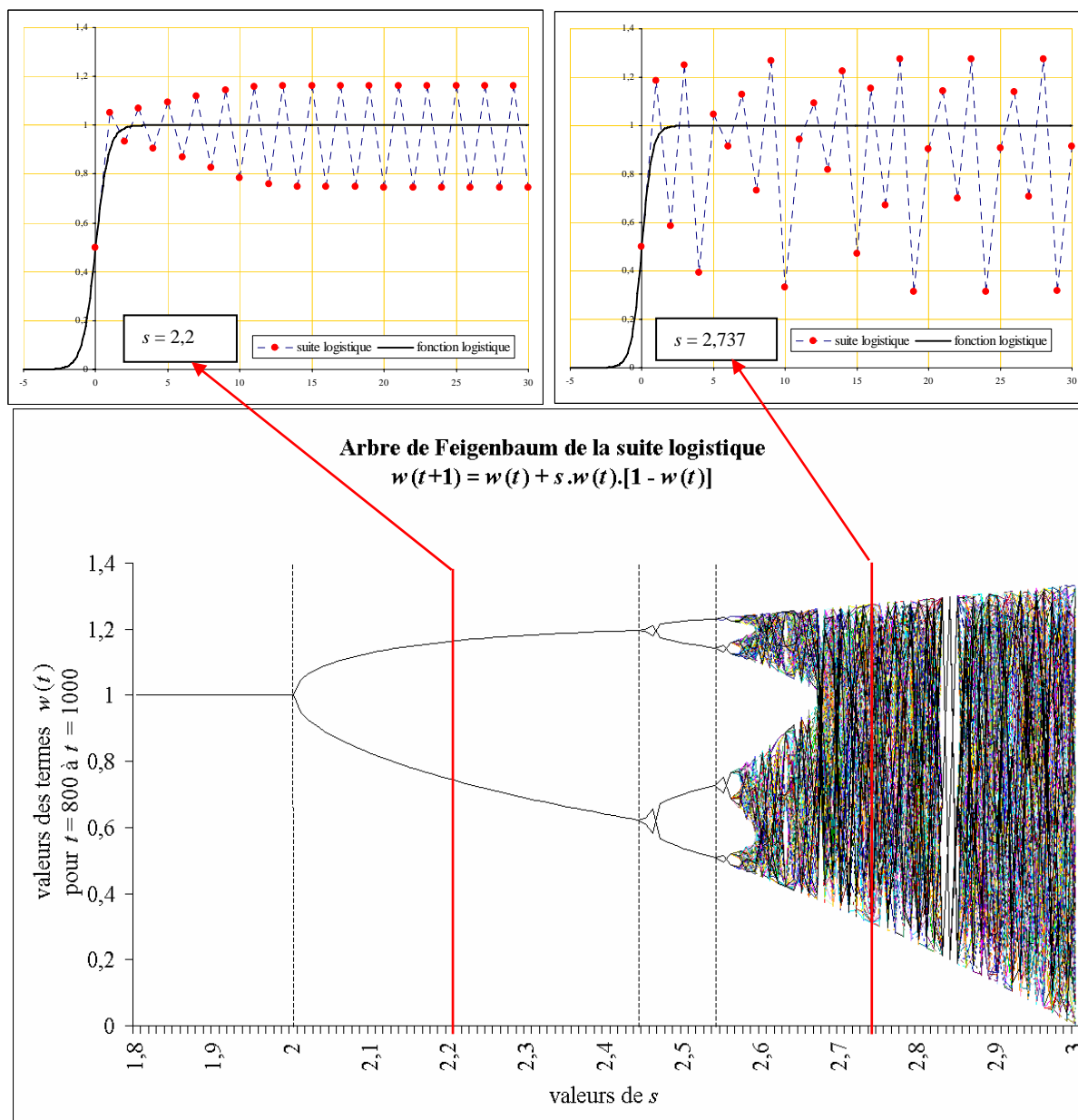


Figure 3 : Comportement chaotique de la suite logistique selon les valeurs de s

Selon les différentes valeurs du coefficient d'interaction $s = r.u_{\max}$, nous résumons les différents comportements possibles de la suite logistique définie par :

$$w(t+1) = w(t) + s \cdot w(t) \cdot (1 - w(t))$$

pour $s = 0$:	un attracteur de type point fixe : 0
pour $0 < s \leq 1$	comportement stable un seul attracteur inférieur à 1
pour $1 < s < 2$:	la suite dépasse 1 mais se stabilise à un attracteur : 1
pour $2 < s < \sqrt{6} = 2,4495$:	un cycle d'ordre 2
pour $2,45 < s < 2,59$:	un cycle d'ordre 4
pour $2,6 < s \leq 3$:	chaos
pour $s > 3$	divergence vers $-\infty$

L'arbre de Feigenbaum (appelé aussi diagramme de bifurcation) présenté dans la figure 3, résume ces comportements dans un graphique où l'axe des abscisses donne les différentes valeurs du taux de croissance s et l'axe des ordonnées donne les valeurs $w(t)$ de la suite logistique au delà d'un certain rang t (de $t = 800$ à $t = 1000$).

Pour conclure sur cette section, nous pouvons dire que les deux modèles (discret et continu) sont incompatibles car la fonction logistique n'est jamais solution de l'équation itérative qui génère la suite logistique, contrairement à la fonction exponentielle qui permet de calculer exactement les termes d'une progression géométrique (qui modélise une croissance dans un environnement sans limite de ressource). Cette différence entre les deux modèles va très loin puisque la suite logistique peut montrer, pour certaines valeurs de son taux de croissance, un comportement anarchique qui est connu comme un des exemples les plus simples du chaos déterministe, alors que la fonction logistique reste très régulière quelle que soit la valeur de son paramètre.

4.2.3 Modèle logistique déterministe ou probabiliste

Le phénomène étudié (épidémie, etc.) peut se calculer au niveau macroscopique de manière déterministe, le taux d'interaction r de la suite logistique (4) représente alors la proportion des liens potentiels qui se traduisent par une infection. Mais le modèle peut se concevoir au niveau des individus, microscopique, le coefficient r représente alors la probabilité pour qu'un contact potentiel entre individu sain et individu infecté se traduise par une contamination de l'individu sain.

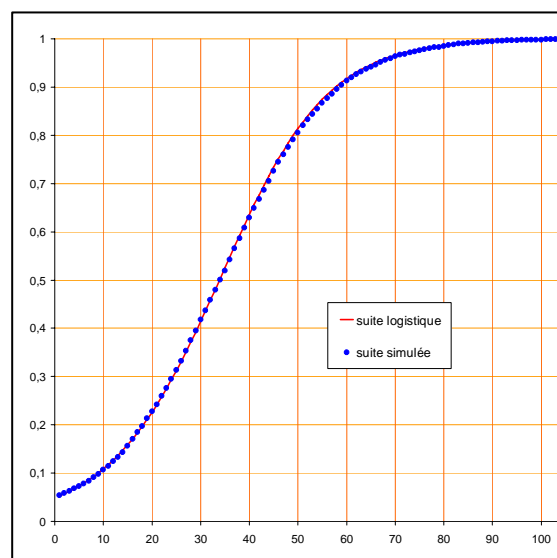


Figure 4 : Comparaison entre la suite logistique déterministe et probabiliste ($u_{max} = 10201$, $u_0 = 497$ et $r = 0,9$).

4.2.4 Simulation spatialisée du modèle logistique discret.

Imaginons que l'espace soit partitionné en N cellules carrées, chaque individu occupant une cellule, on a donc $N = u_{\max}$. Chaque individu i peut être dans deux états $s_i = 0$ (sain) ou $s_i = 1$ (infecté), (figure 5). On définit u_0 cellules infectées à l'instant initial, les autres ($u_{\max} - u_0$) étant saines.

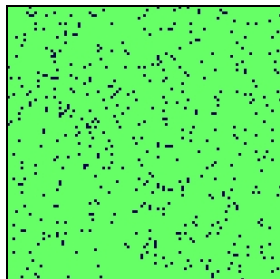


Figure 5 : Localisation aléatoire des hôtes ($s_i = 1$) à l'état initial ($u_0 = 497$), domaine de 101×101 ($u_{\max} = 10201$ hôtes).

Le principe de croissance logistique se traduit alors de la manière suivante. La probabilité pour qu'une cellule i soit saine est :

$$P(s_i = 0) = \frac{u_{\max} - u(t)}{u_{\max}}$$

Pour chaque cellule saine i , parmi les $N-1$ contacts possibles qu'elle peut avoir avec les autres individus, seuls $u(t)$ contacts concernent des personnes infectées et une proportion k de ces contacts transmettent l'infection. La probabilité pour qu'un individu soit infecté à l'instant t sachant qu'il est sain, vaut donc :

$$P(s_i = 1 / s_i = 0) = \frac{ku(t)}{u_{\max} - 1}$$

et la probabilité qu'une cellule i soit infectée à l'instant t est donc donnée par une loi logistique :

$$P(s_i = 1) = \frac{ku(t)}{u_{\max} - 1} \cdot \frac{u_{\max} - u(t)}{u_{\max}} = r \cdot u(t) (u_{\max} - u(t)) \quad (14)$$

avec $r = \frac{k}{u_{\max} (u_{\max} - 1)}$

On peut vérifier que la valeur de r ainsi définie respecte la condition de stabilité (13).

L'algorithme s'exprime alors comme suit. Pour toute cellule i entre 1 et N , si i est sain ($s_i = 0$), on fait un tirage aléatoire uniforme d'un nombre entier n entre 1 et $N-1$. Si $n \leq ru(t)$ alors la cellule i devient infectée ($s_i = 1$) sinon elle reste saine. Remarquons qu'on aurait pu aussi tirer au hasard les individus sains, pour que le processus soit entièrement aléatoire. La figure 6 montre une séquence de diffusion pour une simulation.

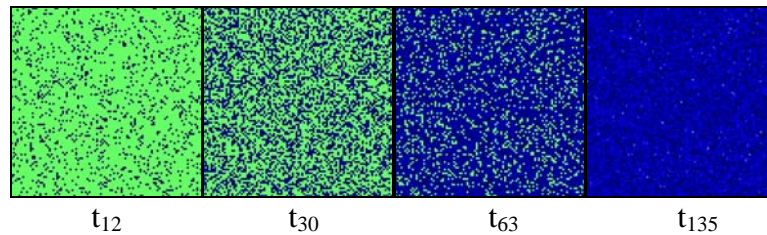


Figure 6 : Simulation spatialisée de la fonction logistiqu probabiliste.

Comme on peut le constater avec la figure 4, il y a une bonne correspondance entre modèle probabiliste et déterministe lors d'une simulation portant sur un domaine de $101 \times 101 = 10201$ cellules, dans lequel on a 497 cellules infectées au départ (soit 4,87%).

On peut dire que ce modèle n'est pas un modèle de diffusion spatiale dans la mesure où il ne tient pas compte de l'espace. Ici tout individu peut entrer en contact avec tout autre, sans considération de localisation. L'objectif des sections qui suivent est de spatialiser les modèles de croissance que nous venons d'explorer. Trois modèles sont ainsi présentés : un modèle de diffusion par front, un modèle de croissance-diffusion et un modèle générique, apte à rendre compte d'une grande variété de phénomène de propagation.

5 Modèle de diffusion par front

A l'inverse du modèle précédent, on examine maintenant un modèle dont la croissance dépend des interactions entre individus voisins. C'est le modèle de diffusion par front.

5.1 Modèle théorique du front circulaire

On se place dans un espace continu à deux dimensions et nous utilisons ici la métaphore de l'incendie de forêt. Si le foyer démarre en un point C , (ou formant un petit cercle de rayon R_0 autour de C) dans un espace homogène et sans vent, la propagation s'effectue alors au voisinage immédiat des flammes dans les parties non brûlées. Le front se propage à une vitesse constante v autour du centre C , la partie brûlée (points i où $s_i = 1$) a donc la forme d'un cercle de rayon $R(t)$ qui s'agrandit avec une vitesse v constante, on a donc $R(t) = vt + R_0$ et la surface brûlée $S(t)$ à l'instant t est donc la surface du cercle de rayon $R(t)$:

$$S(t) = \pi R(t)^2 = \pi v^2 t^2 + 2\pi v R_0 t + \pi R_0^2$$

qui est donc de la forme :

$$S(t) = at^2 + bt + c$$

Ce modèle de croissance n'est donc pas du tout de type logistiqu. De manière plus générale **la croissance est ici proportionnelle à la longueur $L(t)$ du front**, elle vérifie donc une équation du type:

$$\frac{\partial u}{\partial t} = kL(t)$$

Comme cette longueur est, pour un cercle (ou même un carré), proportionnelle au rayon $R(t)$, la solution $u(t)$ de cette équation donne toujours un modèle de croissance quadratique en fonction du temps. Evidemment lorsque l'incendie atteint les limites du domaine, cette croissance se ralentit pour s'arrêter lorsque tout est brûlé, donnant ainsi une courbe de

croissance en forme de S , mais qui n'a plus rien à voir, mathématiquement, avec la fonction logistique. A l'extrême, si le domaine est idéalement circulaire autour du centre C , le front d'incendie atteint les limites du domaine en tout point au même instant et elle s'arrête instantanément, si le domaine est carré de côté d , le front circulaire atteint le domaine progressivement :

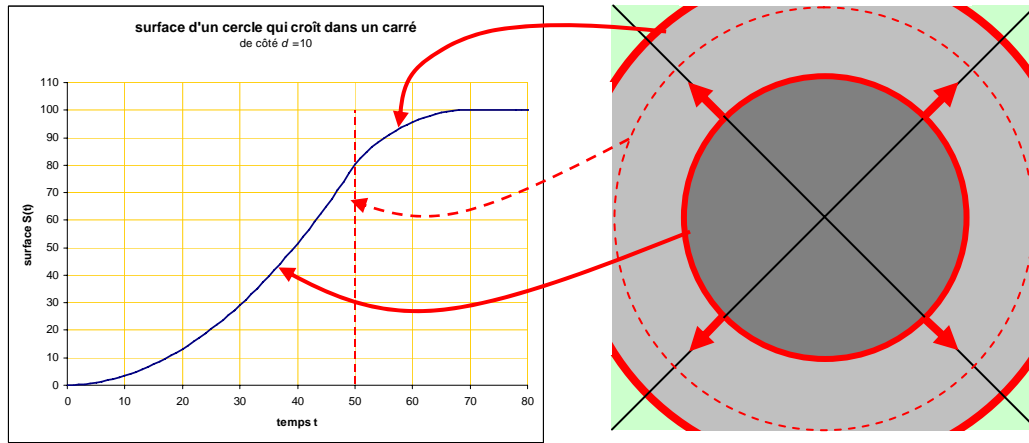


Figure 7 : Croissance quadratique d'une diffusion par front.
La forme en S vient de l'inflexion de la courbe lorsqu'elle atteint les limites du domaine

La simulation déterministe de la diffusion par front décrit ainsi la propagation d'agents par contact direct entre hôtes, à l'instar d'un incendie. La règle d'acquisition est déterministe : tout hôte i passe de l'état 0 à l'état 1 si au moins un de ses 4 voisins est à l'état 1. La figure 9 montre une séquence de la simulation à partir d'une configuration initiale où les hôtes situés au centre du domaine sont à l'état 1.

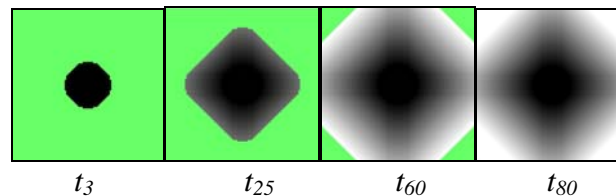


Figure 9 : Une forme déterministe de diffusion par front ($u_0 = 497$).

La population infectée augmente à chaque itération en proportion de la longueur du front, ce qui se traduit par une courbe parabolique pendant les premières phases de la diffusion (fig. 8).

La réalité est évidemment beaucoup plus complexe. Dans le modèle sans vent, l'hétérogénéité locale a pour effet de donner une forme complexe à la ligne de front, qui s'écarte du centre avec le temps. La croissance de la surface brûlée est alors plus forte dans les parties convexes du front et plus lente dans les parties concaves. Néanmoins, le front faisant un tour complet autour du centre de départ de l'incendie, la somme des convexités l'emporte sur les concavités de la valeur d'un cercle, si bien que la surface brûlée évolue selon un modèle très proche du front circulaire théorique.

6 Modèle de croissance-diffusion

Nous abordons maintenant l'étude du modèle mixte, où nous montrons que les deux premiers types de modèles présentés, croissance logistique et croissance par front, sont en fait les deux limites extrêmes d'une classe plus générale de modèles de diffusion, où la fixation des paramètres permet d'instancier⁴ chaque modèle de cette classe.

Les deux modèles précédents sont tous deux des modèles de croissance, aussi la notion de diffusion reste sous-jacente. Or la diffusion pure ne présuppose pas une croissance de l'ensemble de départ. L'exemple du produit volatile qui s'échappe d'une cuve (stock limité) pour se répandre dans l'atmosphère est un exemple typique de diffusion sans croissance. Le volume initial de produit se conserve dans le temps mais occupe un domaine de plus en plus vaste en se diluant, jusqu'à ce que la dilution soit telle qu'il devienne indétectable, ou, si le domaine est petit (comme une pièce fermée), le produit tend à s'homogénéiser dans l'espace pour arriver à un état d'équilibre, d'équidensité.

En géographie humaine, la diffusion pure est assez rare, elle s'accompagne en général d'un processus de croissance. Ainsi la migration de groupes humains à partir d'un foyer initial ou les fronts pionniers contiennent en même temps un processus de diffusion et un processus de croissance. Les premiers migrants se fixent et croissent sur place, puis une partie de leurs descendants continuent de progresser.

6.1 Le modèle continu de Croissance-Diffusion

Supposons donc un espace à deux dimensions (ou plus) représentant un milieu continu pouvant contenir un certain « produit » (cela peut être un effectif de chômeurs, la proportion de personnes infectées, le taux de NO_x dans l'atmosphère, etc.) dont la mesure dans le milieu est définie par une fonction $z = u(x, y, t)$, qui dépend du lieu (x, y) et de l'instant t . On suppose la fonction u continue et on considère la surface représentant les valeurs de la fonction u sur l'axe des z .

Le modèle de diffusion dans ce milieu continu est formalisé le plus souvent par un modèle de croissance-diffusion. Son équation traduit le fait que la croissance en un lieu donné est d'autant plus forte qu'on se trouve dans un « creux » plus prononcé, et que la décroissance est d'autant plus forte qu'on se trouve sur une « bosse » accentuée. Ainsi, le processus de diffusion pure comble les « creux », et réduit les « bosses » de la surface de densité au cours du temps.

Si la dynamique est suffisamment amortie, le système se stabilise vers l'équilibre, sinon il oscille en permanence, les creux se comblant trop vite, produisent des bosses et les bosses évoluent vers des creux. Il y a donc deux membres dans l'équation, le membre de gauche exprime la variation dans le temps de la fonction u , c'est la dérivée partielle par rapport au temps. Le deuxième membre mesure la concavité locale de la surface des $z = u(x, y, t)$, c'est l'opérateur de Laplace (ou le laplacien) de u qu'on note :

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}$$

Il donne une valeur positive si la surface en (x, y) est concave (les dérivées par rapport à x et y croissent) et négatif si la surface est convexe, est nulle si, au point (x, y) , la surface est

⁴ En programmation orientée objet, une classe est une description abstraite d'un ensemble d'objets caractérisés par les mêmes attributs, propriétés et méthodes. Instancier consiste à construire un objet concret de cette classe, en allouant un espace mémoire à cet objet et en affectant des valeurs particulières à ses attributs.

localement plane (en pente ou non). Ainsi l'équation de diffusion obtenue, appelée aussi équation de la chaleur, s'exprime par :

$$\frac{\partial u}{\partial t} = k \Delta u \quad (17)$$

Si, de plus, il y a croissance de u , il faut ajouter dans le membre de droite un terme de croissance, par exemple logistique comme en (5), ce qui donne l'équation finale de croissance-diffusion, appelée aussi réaction-diffusion car elle a été initialement étudiée en chimie par Turing [TUR 52], où une réaction provoque l'augmentation locale de densité d'un produit dans le milieu, qui diffuse ensuite dans son voisinage. L'équation s'écrit donc :

$$\frac{\partial u}{\partial t} = k \Delta u + \alpha u(1 - u) \quad (18)$$

Pour plus de généralités dans ces modèles, en particulier pour les modèles qui dépendent d'un environnement proche (modèles non-locaux), voir [COV 03]. Ce modèle purement local et continu peut se formaliser aussi en termes discrets plus propice à la simulation informatique.

6.2 Modèle discret de Croissance-Diffusion

Dans un espace discret comme un réseau cellulaire à n_V voisins ($n_V = 4$ pour le voisinage dit de von Neumann, $n_V = 8$ pour le voisinage de Moore, ou plus si le rayon du voisinage est supérieur à 1), le laplacien peut se définir comme la différence entre la valeur centrale de u et la moyenne de ses voisins :

$$\Delta u(i, j, t) = \langle u(i, j, t) \rangle_V - u(i, j, t)$$

où, $\langle u(i, j, t) \rangle_V = \frac{1}{n_V} \sum_{(\varepsilon, \eta) \in V} u(i + \varepsilon, j + \eta, t)$ est la moyenne de u , pour les voisins de (i, j)

Le premier membre de (18) exprime l'accroissement de $u(i, j, t)$ dans un pas de temps, d'où l'expression discrète :

$$u(i, j, t+1) = (1 - k)u(i, j, t) + k \cdot \langle u(i, j, t) \rangle_V + r \cdot u(i, j, t)(1 - u(i, j, t)) \quad (19)$$

Si $k = 0$, la diffusion est nulle, l'équation est seulement une croissance logistique.

Si $k = 1$, la diffusion est maximale, elle ne dépend plus de la valeur de u sur la cellule (i, j) , qui est remplacée par la moyenne des valeurs du voisinage.

Si $k = \frac{n_V}{n_V + 1} = k_e$, le modèle est équilibré car la cellule centrale a le même poids que chaque cellule du voisinage.

La croissance continue purement locale n'est pas très réaliste. On peut élargir à un voisinage le champ de connaissance de la cellule (i, j) pour sa croissance. Ainsi, le modèle de croissance continu du front, qui était purement local, peut s'exprimer ici de manière plus large. De même, la croissance logistique homogène dans toute la population n'est pas plus réaliste, il est plus satisfaisant de supposer que les contacts ne sont possibles que dans un voisinage de chaque individu. Ainsi, on considère la population des individus de la cellule (i, j) et dans n_V cellules voisines. S'il y a n individus par cellule, cela représente $n(1+n_V)$ individus au total dans tout le voisinage, la cellule centrale (i, j) compris. Dans la cellule (i, j) , il y a une proportion $1-u(i, j, t)$ d'individus n'ayant pas adopté, et une proportion de $\langle u(i, j, t) \rangle_V$ d'ayant adopté dans son voisinage V' , il y donc un nombre de contacts proportionnel à $\langle u(i, j, t) \rangle_V \cdot (1-u(i, j, t))$

Le modèle de croissance non local se traduit alors par :

$$u(i, j, t+1) = u(i, j, t) + r \cdot \langle u(i, j, t) \rangle_v \cdot (1 - u(i, j, t))$$

qui peut s'écrire aussi :

$$u(i, j, t+1) = r \cdot \langle u(i, j, t) \rangle_v + (1 - r \cdot \langle u(i, j, t) \rangle_v) u(i, j, t) \quad (20)$$

La forme déterministe du modèle binaire non local de croissance, avec $u(i, j, t) = 0$ ou $u(i, j, t) = 1$ exprime l'adoption (ou non) d'une innovation, avec un individu par cellule. Dans ce cas, si $u(i, j, t) = 1$ alors (20) donne tout de suite $u(i, j, t+1) = 1$, mais lorsque $u(i, j, t) = 0$, on a $u(i, j, t+1) = r \cdot \langle u(i, j, t) \rangle_v$. Pour que $u(i, j, t+1)$ reste binaire, il faut arrondir à l'entier 0 ou 1 le plus proche, ce qui revient à prendre une fonction Y telle que si $x > 0$ alors $Y(x) = 1$ et sinon $Y(x) = 0$. On utilise alors un seuil θ (par exemple $\theta = \frac{1}{2r}$), pour calculer $u(i, j, t+1)$:

$$\begin{aligned} &\text{si } u(i, j, t) = 0 \text{ alors} \\ &\quad u(i, j, t+1) = Y(\langle u(i, j, t) \rangle_v - \theta) \\ &\text{sinon} \\ &\quad u(i, j, t+1) = 1 \end{aligned} \quad (21)$$

Sous sa forme probabiliste, on a pour toute cellule (i,j) saine :

$$P\left(u(i, j, t) = 1 \middle| u(i, j, t) = 0\right) = r \cdot \langle u(i, j, t) \rangle_v$$

qui exprime que la probabilité de transition de l'état 0 vers l'état 1 est proportionnelle à la moyenne des voisins infectés.

L'objectif est maintenant de proposer un modèle générique capable de couvrir tous les cas intermédiaires, entre le modèle logistique (voisinage global) et le modèle par front (voisinage local), basé sur un voisinage de taille donnée et une certaine profondeur temporelle. A l'intérieur du voisinage, on a un fonctionnement de type logistique et à l'extérieur une diffusion sur un front non-local limité par la taille de ce voisinage (Coville J. 2003).

7 Vers un modèle Spatio-temporel de diffusion

On est souvent amené à restreindre les contacts possibles d'un adoptant potentiel, non seulement dans l'espace mais aussi dans le temps. Il arrive en effet fréquemment que les ayant adoptés ne restent qu'un certain temps enthousiastes à communiquer leur adoption, de même, dans une épidémie, un individu infecté ne reste contagieux qu'un certain temps, etc.

Appelons $v(i, j, t)$ la variation de la proportion d'adoptants entre l'instant $t-1$ et t . Le cumul des $v(i, j, h)$ pour $h = 0, 1, \dots, t$ est donc égale à $u(i, j, t)$. L'introduction de cette variable v indique que la cellule possède la mémoire de son passé, tout au moins ici, d'une partie de son passé, car elle ne mémorise que ses variations cumulées depuis T unités de temps dans le passé (de $t-T+1$ à t).

La variation d'adoptants entre t et $t+1$ est donc proportionnelle au nombre d'adoptants potentiels dans la cellule ($1 - u(i, j, t)$), ainsi qu'au nombre d'adoptants récents (moins de T

pas de temps) dans la cellule et son voisinage V , représenté par la première parenthèse de (22) :

$$\begin{cases} v(i, j, t+1) = r \cdot \left(\sum_{\tau=0}^{T-1} \frac{1}{n_V} \sum_{(\varepsilon, \eta) \in V} v(i + \varepsilon, j + \eta, t - \tau) \right) \cdot (1 - u(i, j, t)) \\ u(i, j, t+1) = u(i, j, t) + v(i, j, t+1) \end{cases} \quad (22)$$

la première parenthèse n'est autre que la proportion moyenne des adoptants (récents) dans le voisinage spatio-temporel $V \times T$, posons alors :

$$\langle u(i, j, t) \rangle_{V \times T} = \frac{1}{n_V} \sum_{\tau=0}^{T-1} \sum_{(\varepsilon, \eta) \in V} v(i + \varepsilon, j + \eta, t - \tau)$$

l'équation (22) devient alors :

$$u(i, j, t+1) = u(i, j, t) + r \cdot \langle u(i, j, t) \rangle_{V \times T} \cdot (1 - u(i, j, t)) \quad (23)$$

Dans le cas d'une variable u binaire, (23) se traduit par une même formulation qu'en (21), mais avec le voisinage spatio-temporel $V \times T$

Tout se passe alors comme si chaque agent, une fois transmis à un hôte, avait une durée de vie limitée à T itérations. Passé à l'état 1, l'hôte ne reste donc actif que pendant les T itérations suivantes. Ceci nécessite que chaque hôte possède non seulement une connaissance de son voisinage, mais aussi de son passé à travers une fonction $\tau_i(t)$ qui donne l'âge de sa transition depuis l'instant t_i de passage à l'état 1. Tant que l'hôte est dans l'état 0, la valeur de τ_i n'est pas définie (par exemple on la fixe à $\tau_i = -1$), lorsque son état passe à 1, τ_i prend la valeur 0, et sa valeur s'incrémente ensuite de 1 à chaque pas de temps, c'est-à-dire $\tau_i = t - t_i$. Dès que τ_i atteint la valeur seuil T , l'hôte devient inactif. L'activité de l'hôte i est représentée par la fonction d'émission (ou de sortie) $v_i(t)$ qui est définie par :

$$v_i(t) = \begin{cases} 1 & \text{si } \tau_i(t) \in [0, T[\\ 0 & \text{sinon} \end{cases}$$

La transition à l'état 1 d'un hôte i dépend de la proportion d'hôtes actifs dans son voisinage spatio-temporel $V \times T$, qui est la moyenne de v dans ce voisinage, notée :

$$\langle v_i(t) \rangle_{V \times T} = \frac{1}{n_R} \sum_{j \in V(i)} v_j(t)$$

où n_R représente le nombre de voisins dans un rayon R . Dans une approche stochastique, la probabilité de transition à l'état 1 de l'hôte i est alors donnée par :

$$P\left(\frac{s_i = 1}{s_i = 0}\right) = r \cdot \langle v_i(t) \rangle_{V \times T}$$

La règle d'émission de l'agent est donc importante, elle tend à diminuer la force de l'agent avec le temps et peut être plus ou moins complexe selon le phénomène étudié. La figure 10 montre des dynamiques différenciées de diffusion selon la durée de vie T des agents.

On constate que la règle d'émission a ici une forte incidence sur l'évolution de la diffusion. Une faible durée de vie de l'agent (fig. 10, courbe bleue) entraîne un arrêt rapide de la diffusion, bien avant la saturation du domaine (fig. 11, simulations 1 et 2), malgré une règle d'acquisition peu contrainte ($r = 1$). Au contraire, lorsque la durée de vie de l'agent est importante (fig. 10, courbe rouge), le modèle privilégie la règle d'acquisition et a un

comportement proche d'un modèle stochastique de diffusion par front, le vecteur saturant presque le domaine (fig. 11, simulations 5 et 6). C'est dans des niveaux paramétriques intermédiaires (fig. 10, courbe verte) que le modèle produit ses formes les plus intéressantes, à l'instar des simulations 3 et 4 (fig. 11). Dans ce cas les règles d'acquisition et d'émission sont pleinement en interactions, la forme finale dépend en partie de la direction initiale prise par le front.

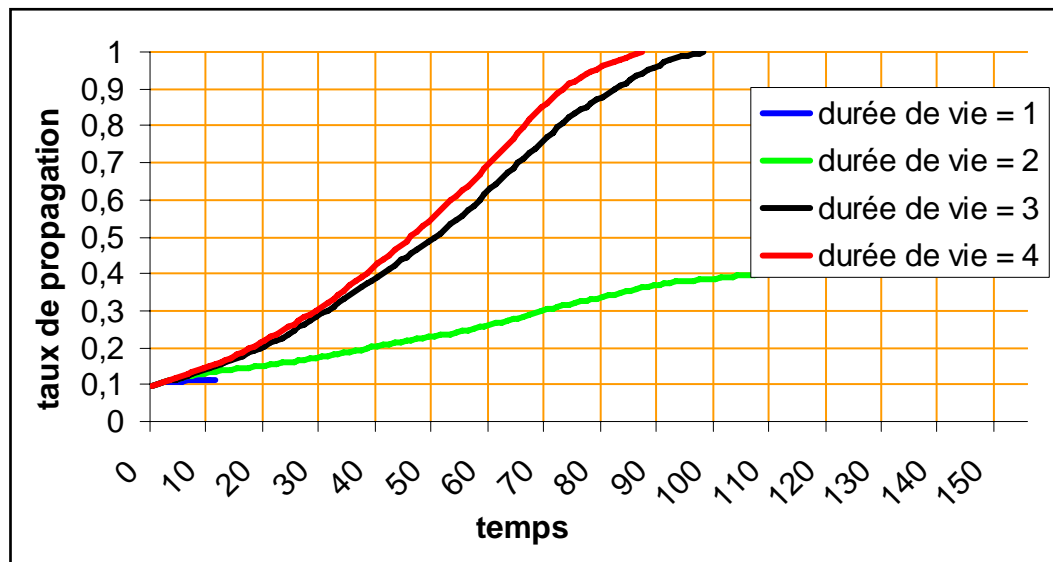


Figure 10 : Dynamique de diffusion selon une règle d'émission évolutive.

Par construction et à la différence des modèles précédents, ces deux règles rendent possibles la non saturation du domaine par le processus de diffusion. Ceci n'est pas lié à des règles d'acquisition différenciées selon les hôtes, ce qui est envisageable, mais au fait qu'un hôte peut se retrouver rapidement entouré d'hôtes infectés mais non contagieux, inactifs.

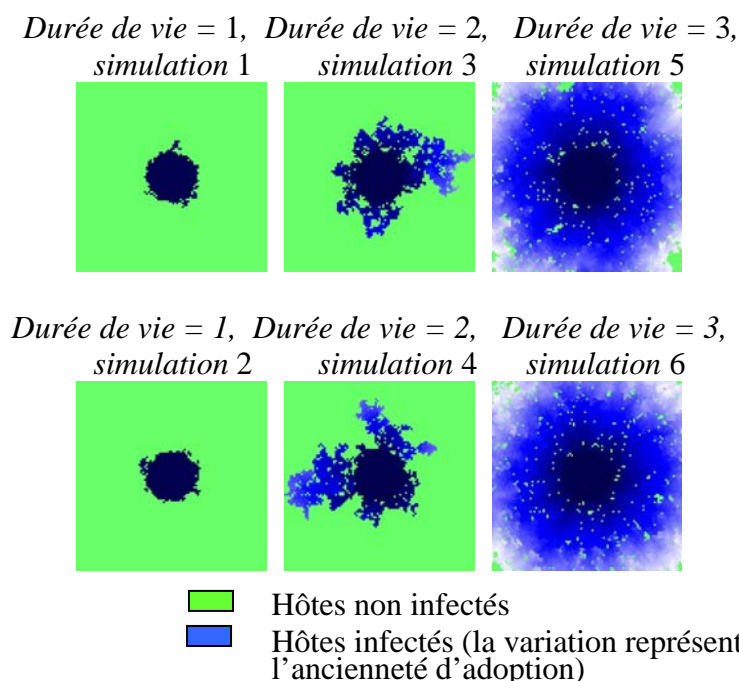


Figure 11 : Différentes formes spatiales de diffusion selon la virulence de l'agent, après arrêt du processus ($u_0 = 497$, localisation au centre du domaine, $R = 1$)

8 Conclusion

Nous avons tenté de poser les bases d'une théorie géographique de la diffusion. Après avoir mis en place la définition et les concepts de la diffusion, nous avons montré comment ces concepts pouvaient se traduire en équations que ce soit sous forme déterministe ou stochastique. Nous avons privilégié une approche discrète, apte à être directement utilisée en simulation. Mais l'approche continue a été aussi abordée et permet d'introduire l'approche discrète.

Nous avons formalisé le concept de diffusion comme étant la combinaison de trois processus : l'acquisition, l'émission et la propagation. Ces processus agissent sur deux niveaux d'entités, les hôtes et les agents. Cette formalisation est parfaitement adaptée à l'utilisation des outils actuels que sont les automates cellulaires (Dubos-Paillard E. & al. 2003) et les systèmes multi-agents (Daudé E. 2004). Elle permet de simuler correctement les phénomènes classiques de la diffusion en géographie, tels que l'innovation ou les épidémies et devrait pouvoir s'appliquer avec succès aux autres processus spatiaux aussi bien naturels que sociaux.

9 bibliographie

- Coville J. 2003 - Equations de réaction-diffusion non-locale, thèse de mathématiques, université Paris 6
- Daudé E. 2004 - Apports de la simulation multi-agents à l'étude des processus de diffusion. *Cybergeo : Revue européenne de géographie*, n°. 255, 15 p.
- Daudé E. Langlois P. 2006 - "Les formes de la diffusion", La forme en géographie, Actes du Colloque GéoPoint'04, Avignon, pp 171-175.
- Dauphiné A. 2003 - Les théories de la complexité chez les géographes, ed. Anthropos-Economica, Paris, 248 pages.
- Dubos-Paillard E., Guermont Y., Langlois P., 2003 - Analyse de l'évolution urbaine par automate cellulaire, le modèle SpaCelle", *L'espace Géographique*, pp 257-378, vol 4, 2003
- Hägerstrand T., 1953 - *Innovations förloppet ur korologisk synpunkt*, C.W.K Gleerup, Lund, Sweden. (Traduit et réédité par Pred A., *Innovation Diffusion as a Spatial Process*, Chicago, University of Chicago Press. 1967)
- Hägerstrand T., 1965 - A Monte Carlo Approach to Diffusion, *European Journal of Sociology*, vol. 63, pp 43-67.
- Rogers E., 1995 - *Diffusion of innovations*, New York: Free Press, Macmillan Publishing Co, 4th edition (1er ed. 1962)
- Saint-Julien T., 1985 - *La diffusion spatiale des innovations*. Montpellier, RECLUS, coll. RECLUS modes d'emploi.
- Turing A., 1952 - The Chemical Basis of Morphogenesis. *Phil. Trans. Roy. Soc. B* 237: pp 37-72.
- Verhulst P., 1838 - Notice sur la loi que la population suit dans son accroissement. *Correspondance Mathématique et Physique*, n° 10, pp 113-121.