

Forum

De l'équation mathématique au programme informatique, dialogue entre mathématiciens et informaticiens

Jean-Pierre Treuil^a, Claude Lobry^b, Claude Millier^c, Jean-Pierre Müller^d, Dominique Hervé^e

^a Informaticien, IRD, UR079 GÉODES, 32 avenue Henri Varagnat, 93143 Bondy cedex, France

^b Mathématicien, INRIA, Université de Nice Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 Sophia-Antipolis, France

^c Bio-mathématicien, ENGREF, 19 avenue du Maine, 75732 Paris cedex 15, France

^d Informaticien, CIRAD, UPR47 GREEN, 73 rue Jean-François Breton, 34398 Montpellier cedex 5, France

^e Agronome, IRD, UR168, BP 64501, 34394 Montpellier cedex 5, France

Mots-clés :
modélisation ;
simulation ;
mathématiques ;
informatique ;
agents

Résumé – L'essor de l'informatique réactive les débats sur la nature des activités de modélisation et de simulation. Ce forum réunit des mathématiciens et des informaticiens à propos de la place respective des mathématiques et de l'informatique dans les modèles. La première partie propose deux exemples d'implication des mathématiques dans les sciences de la vie, et traite ensuite de la distinction entre deux formes de modélisation d'une même réalité, par une mise en équation ou une mise en programme, en soulignant la nécessité de théoriser les modèles informatiques. La seconde partie présente les apports complémentaires, à trois niveaux, des mathématiques et de l'informatique dans le cas des modèles à base d'agents. Le débat à plusieurs voix revient ensuite sur le choix du type de modélisation face au problème posé, les différences entre équations et programmes, les vérités issues des théorèmes ou des simulations, le rôle de l'expérimentation.

Keywords:
modeling ;
simulation ;
mathematics ;
informatics ;
agents

Abstract – From mathematical equation to computational program. Dialogues between mathematicians and computer scientists. The improvement of computer capacities has promoted a very rapid development of computer sciences and re-activated debates regarding the nature of modeling and simulation activities. This forum examines the respective places of mathematics and informatics in the models. The first part proposes two cases of implication of mathematics in life sciences, then analyses the difference between two kinds of modeling of the same reality, i.e., through mathematical equations or through program building, and concludes on the need to theorize computer models. The second part deals with the complementary contributions of mathematics and computer sciences in the case of agent-based models by examining three examples. The debate involving multiple viewpoints then discusses the kind of model that is useful to answer each question, the differences between equations and programs, the differences between truth derived from theorems or from simulations, and the role of experimentation.

Des notes d'actualités et des débats autour de la biométrie ont ponctué pendant 10 ans les colonnes de la revue *NSS* sous la plume de Jean-Marie Legay. Récemment, la revue a joué un rôle de caisse de résonance des discussions suscitées par le texte « La modélisation comme outil d'accompagnement » en permettant que s'exprime une grande variété d'opinions¹. Ces débats sur la conception et l'utilisation des modèles sont apparemment loin d'être clos et ils concernent toutes les disciplines impliquées.

Auteur correspondant : J.-P. Treuil, treuil@bondy.ird.fr

¹ Cf. la contribution du collectif ComMod et les commentaires publiés en Forum dans *NSS* (13, 2, 2005).

Aussi semblait-il opportun de rendre compte dans la revue d'une autre controverse, latente au sein des disciplines de l'information, en particulier entre informatique et mathématiques, que l'on peut voir comme l'autre face, symétrique en quelque sorte de la précédente : quelle est la légitimité et la validité des simulations informatiques qui ont fait suite, grâce aux possibilités de « mise en programme » offertes par l'ordinateur, aux certitudes affichées par les « mises en équation » mathématiques ?

La cinquième session du 6 novembre 2003 de l'Atelier Modélisation Environnement (NSS Dialogues, IRD), animé par Dominique Hervé à Montpellier, traitait de

la place respective, dans la modélisation pour la gestion des ressources renouvelables, des formalisations mathématiques et informatiques. Nous en proposons une restitution en trois volets. Sous le titre « Mathématiques et modélisation », Claude Lobry positionne la simulation informatique par rapport à la physique expérimentale et aux équations mathématiques. Jean-Pierre Treuil aborde les « approches mathématiques des modèles à base d'agents ». Le débat qui s'ensuit prend la forme d'un dialogue à plusieurs voix au cours duquel un animateur reprend la main à plusieurs reprises. Ce débat porte sur les limites des modèles à base d'agents et l'intérêt d'en faire la théorie mathématique.

Claude Lobry : « Mathématiques et modélisation »

En préambule, je citerai les propos de Jean-Marc Lévy-Leblond (1982) sur le rôle des mathématiques dans les sciences : « en physique, les mathématiques s'appliquent mais surtout s'impliquent dans la définition des concepts » (les concepts de la physique s'expriment difficilement en dehors du langage mathématique) ; « il n'y a pas d'explication à cette implication, ni à la question de pourquoi en physique plus qu'ailleurs. »

Pour en sortir, Lévy-Leblond propose de définir la physique ainsi : tout domaine des sciences expérimentales où les mathématiques s'impliquent. Il donne l'exemple de l'astronomie qui fait appel aux mathématiques impliquées, mais qui n'était pas classée dans la physique jusqu'au jour où elle est devenue l'astrophysique, lorsqu'elle s'est dotée d'une méthode de laboratoire, d'une méthode expérimentale qui permette d'explorer ce qui se passe dans le cosmos. Il ne suffit donc pas que les mathématiques s'impliquent pour que la discipline en question s'appelle de la physique. Il faut aussi respecter un certain nombre de conditions sur la façon dont on explore la réalité.

Je vais maintenant présenter deux exemples d'utilisation des mathématiques dans des domaines scientifiques sans possibilité d'expériences, qui ne se ramènent pas à la physique, et qui appartiennent plutôt aux sciences de la vie.

Exemples dans lesquels les mathématiques sont impliquées

1. Les mathématiques du chimostat (Monod, 1950 ; Novick & Szilard, 1950) : dans une cuve, on place des microorganismes et du substrat. Le bilan de masse à l'instant t est (le bilan à l'instant précédent) moins (ce qui est consommé). L'étude des équations montre que la quantité de substrat par unité de temps et par unité de biomasse est une constante qui ne dépend que de la

concentration ; ce débit à l'équilibre est choisi par l'expérimentateur.

2. Le modèle de compétition de Volterra (1931) : deux espèces X et Y sont en compétition dans deux environnements possibles. Un environnement est favorable à Y si Y élimine X dans cet environnement. Au cours du temps, les deux environnements se succèdent et ils sont tous les deux favorables à Y. Que se passe-t-il ?

S'il y a une succession rapide de ces deux environnements, une bonne approximation de ce qui se passe est donnée par la demi-somme des deux équations différentielles représentant le comportement de X et de Y. On peut trouver des valeurs telles que X gagne la compétition. En réalité, la situation n'est pas toujours favorable à Y (imprécision du langage naturel), X peut dominer dans un premier temps mais pas longtemps. Il suffit donc de commuter les environnements de façon à être toujours favorable à X.

Dans ces deux exemples, les mathématiques sont impliquées. Mais, dans le cas du chimostat, il y a un travail de laboratoire, c'est de la physique ou plutôt de la « biophysique ». Le deuxième exemple est un modèle pour penser, à partir d'un discours en langue naturelle. La logique de cette langue nous fait dire que l'issue sera favorable à Y, mais la langue naturelle n'est pas un langage formel et peut nous amener à dire des choses fausses. La modélisation mathématique montre que ce discours n'est pas correct. Lorsqu'on dit que Y finit par gagner asymptotiquement, on ne sait pas que Y est toujours en avance, que l'environnement est favorable à Y. On doit le conforter par un fait de calcul irréfutable.

Il y a 30 ans, on n'aurait jamais parlé de modèle pour l'exemple du chimostat, mais de la « théorie » du chimostat et de la « mise en équation » de cette théorie (Fig. 1).

Dans le modèle de Volterra, les mathématiques sont impliquées, mais, comme il n'y a pas de données de laboratoire, on ne peut pas parler de physique. L'écologie, l'économie, la prévision du climat sont des disciplines sans laboratoire, dans lesquelles on accepte la complexité sans la réduire au laboratoire. Ainsi, ni l'écologie ni l'économie ne sont des sciences assimilables à la physique, car elles n'ont pas recours au laboratoire, mais elles sont caractérisées par une activité intellectuelle qu'il faudrait nommer. Un écosystème complexe est différent d'une expérience contrôlée en laboratoire. Le choix délibéré de la complexité telle que la définit Legay (1997) ne la réduit pas au laboratoire.

Calcul mathématique et simulations

Depuis Newton, on résout des équations par un calcul mathématique. Depuis 1950, en plus du calcul mathématique, on dispose d'un instrument qui permet de faire des simulations sur des équations, ce qui augmente la puissance du mathématicien (Fig. 1). Mais rien d'essentiel

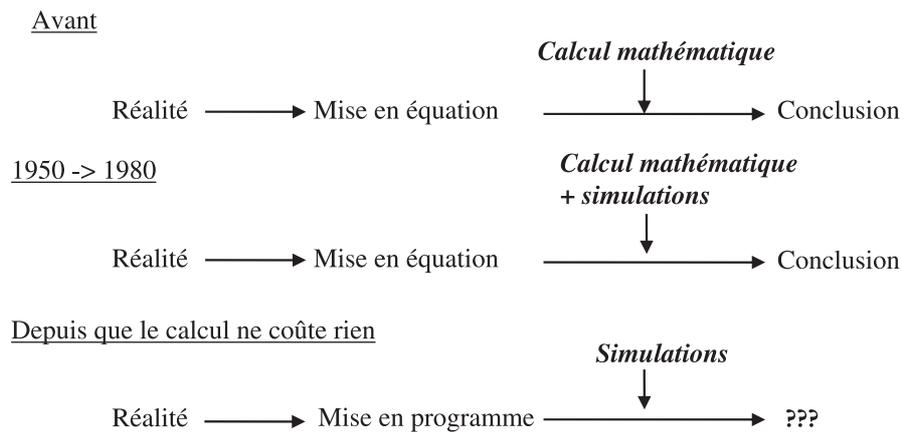


Fig. 1. De la mise en équation à la mise en programme.

n'est changé dans le processus intellectuel. On va simplement beaucoup plus vite. Ensuite, ce qui change de façon fondamentale depuis les années 1980 environ, c'est l'accès à la mise en programme (grâce à l'ordinateur). La mise en programme remplace la mise en équation. C'est très différent parce qu'avec la mise en équation mathématique, on entre dans un champ très théorisé, alors que la mise en programme suivie de simulations ne s'accompagne pas d'une théorie du modèle. L'ordinateur est incontestable car c'est une mécanique qui produit des nombres. Quand on fabrique un outil de modélisation, il faut connaître et comprendre le fonctionnement de cet outil, donc avoir une théorie du modèle, mais plus le modèle est compliqué, plus il est difficile d'en faire une théorie.

Prenons l'exemple d'un automate cellulaire, un des modèles à base d'agents les plus simples. Nous allons représenter un problème de contagion spatialisée. Un individu peut être susceptible, infecté, infectieux ou réfractaire et chaque cellule est un individu. Une cellule peut donc prendre quatre états, et la règle de changement d'état est qu'un individu susceptible devient infecté au pas de temps suivant si l'un de ses quatre voisins est infecté. La question est de savoir si cela devient endémique, c'est-à-dire si la maladie est persistante. Si vous travaillez sur une grille de dimension finie, vous faites tourner sur votre ordinateur un système dynamique sur un ensemble fini d'états possibles, qui évidemment est très grand. Donc, toute solution finira par être périodique. Il y a deux sortes de solutions périodiques : celle où tout le monde est susceptible et il n'y a plus d'infecté, ce qui conduit à une situation non endémique, et celle où il reste constamment un foyer infectieux. La question est de savoir ce que prédit ce genre de modèle. Par simulation, on ne peut pas répondre à cette question. Il est hors de question de faire des simulations exhaustives : sur une grille 100 x 100, quinze états initiaux à la puissance 10 000. Pour savoir ce qui se passe, il n'y a pas d'autre solution

que de faire une théorie. Ce type d'automate a été étudié par les mathématiques : à partir du moment où il y a plus de trois stades où l'on est infectieux, on démontre qu'avec une probabilité très proche de 1 la situation finale sera endémique. Mais on ne sait pas encore démontrer que c'est toujours le cas ; il n'y a pas de contre-exemples. On ne dispose pas encore d'une théorie mathématique complète. Le physicien peut faire cependant du modèle et de son simulateur un objet de laboratoire en étudiant ce qui se passe lorsque les paramètres varient. Il fait ainsi une « théorie du modèle ».

L'utilisation des ordinateurs pour faire des mises en programme devrait croître et ceci, pour des usages pouvant avoir de graves conséquences lorsqu'il s'agit d'aide à la décision. Il est urgent de lancer des chantiers pour théoriser les modèles informatiques.

Jean-Pierre Treuil : « Approches mathématiques des modèles à base d'agents »

L'ensemble des modélisations informatiques que l'on appellera ici modèles à base d'agents regroupe une grande variété d'expressions qui dépendent des disciplines concernées : « approches particulières » (physique, hydrodynamique), « modèles individus centrés » (éthologie, écologie, dynamique des populations), « simulations multi-agents » (économie, sociologie, géographie, mais l'expression est aussi utilisée dans un sens générique) ; « micro-simulations, approches micro-analytiques » (démographie, géographie).

Existe-t-il, derrière cette diversité, des éléments communs suffisamment consistants pour qu'un regroupement ait réellement un sens et qu'il soit envisageable d'en faire une théorie ? La question est ouverte, et nous devons nous garder d'y répondre trop vite. Observons cependant la structure qui se répète dans tous ces modèles : selon

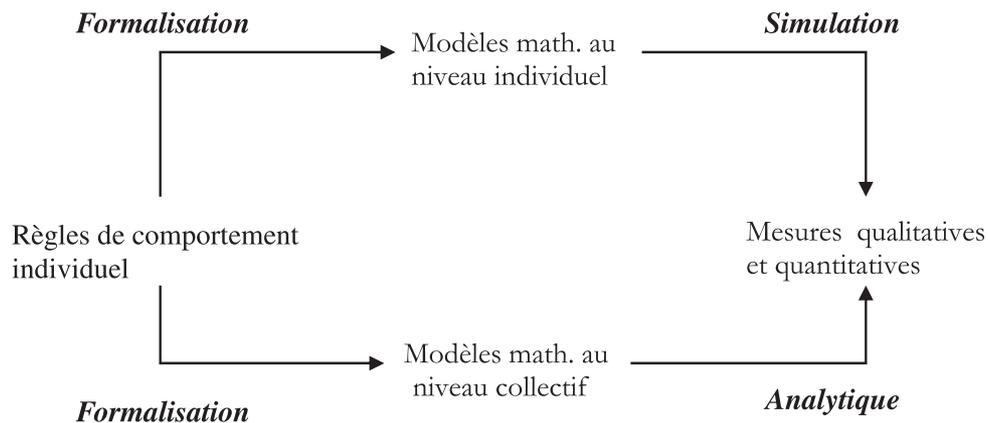


Fig. 2. Schéma situant l'intervention des mathématiques.

une formulation classique, celle de systèmes de granularité fine, constitués d'un grand nombre d'entités appelés agents, évoluant dans un environnement, et ayant entre eux et avec cet environnement des interactions déterminant leur dynamique collective. Ces agents représentent des objets concrets du monde étudié (gouttes d'eau, cellules, particules, individus du monde végétal ou animal, personnes, agglomérations urbaines). Bien sûr, le terme d'objet concret prête à ambiguïté, car, puisqu'il s'agit de modèles, ce sont toujours des concepts qui sont en jeu. Dans certains cas, par exemple, pour des raisons de performance de calcul, les agents seront des groupes d'objets rassemblés selon des critères n'ayant pas de réelle signification du point de vue scientifique concerné. L'exploitation de ces modèles est faite essentiellement par des simulations sur ordinateur, à partir d'une « mise en programme » où chaque agent et l'environnement sont représentés individuellement par des modules séparés. Du moins en est-il comme cela en théorie, car il existe toujours une distance entre le modèle conceptuel de départ et sa traduction informatique. À partir de ces simulations, on observe les dynamiques qui s'expriment au niveau de groupes d'agents ou du système tout entier. Les différents angles d'approche mathématique des modèles à base d'agents sont indiqués sur la figure 2.

Les mathématiques interviennent dans les modèles à base d'agents sur plusieurs plans :

- dans la caractérisation de la dynamique du système ;
- dans la modélisation directe d'une dynamique « agrégée » du système sous forme d'équations portant sur des variables où les agents ne sont plus explicitement représentés ;
- dans l'expression formelle des règles de comportement individuel ;
- dans la recherche de techniques permettant de passer des formulations mathématiques des règles individuelles aux équations de la dynamique agrégée.

Caractériser des systèmes d'agents

Un modèle à base d'agents est un modèle comportant un grand nombre de degrés de liberté. Il produit à chaque pas de temps simulé une information considérable. L'exploitation des simulations requiert donc des méthodes pour résumer cette information, en extraire l'essentiel du point de vue de l'observateur. Il faut pouvoir notamment bien caractériser les structures spatiotemporelles du système des agents, telles que leur répartition spatiale, l'organisation de leur réseau d'interactions, la forme de leurs trajectoires et la manière dont ces trajectoires sont corrélées entre elles. Il faut pouvoir reconnaître sur tous ces éléments des « patrons » spécifiques. Un tel travail utilise souvent de simples indicateurs – moyennes, écarts types – relevant de la statistique élémentaire. Il peut également faire appel à des notions mathématiques élaborées relevant de la géostatistique, de la géométrie stochastique, de la théorie des graphes, de la morphologie mathématique et de la théorie du signal. C'est ainsi que la caractérisation de la distribution spatiale des semis de points (« processus ponctuels ») ou d'autres objets géométriques, la caractérisation de la complexité des réseaux ont pu donner lieu à des travaux théoriques importants.

Modéliser les dynamiques agrégées des systèmes d'agents

La recherche d'une formulation mathématique directe de la dynamique agrégée d'un modèle à base d'agents peut avoir deux objectifs concrets :

- déterminer analytiquement les propriétés de cette dynamique (par exemple, l'existence d'un état stationnaire stable), de manière à avoir des certitudes quand on fait des simulations et non de simples présomptions sur le comportement du système ;

- simuler cette dynamique, quel que soit le nombre d'agents, sans passer par l'explicitation de leurs comportements au niveau individuel.

Bien sûr, une question se pose : si l'on peut construire un modèle agrégé qui donne des résultats équivalents au modèle à base d'agents, quelle est l'utilité de ce dernier ou, du moins, quelle est l'utilité d'une double modélisation ? Nous apporterons à cette question la réponse suivante : tenter de construire un modèle agrégé force à réfléchir aux hypothèses qu'il faut prendre pour assurer l'équivalence et à examiner si ces hypothèses sont vérifiées dans la réalité. Cette réflexion contribue à la connaissance en profondeur du fonctionnement du système étudié – quelle est, par exemple, l'influence réelle des hétérogénéités de l'environnement ou du comportement des agents ? Elle est par là même un garde-fou contre une tentation constamment présente dans la pratique des modélisations agents, celle de l'exhaustivité qui conduit à un trop grand réalisme et donc à une trop grande complexité.

Nous prendrons deux exemples de construction de modèles agrégés greffés sur des modèles à base d'agents.

Le premier exemple est celui d'interactions entre robots (Martinoli *et al.*, 1999). On demande à des robots de sortir des bâtons de trous où ils sont logés (Lerman *et al.*, 2001). La présence et la collaboration de deux robots sont nécessaires pour sortir un bâton d'un trou. Un robot a deux états : 1) il recherche un bâton ; 2) il est devant un bâton en attente d'un autre robot. Des équations donnent l'évolution du nombre de robots dans chaque état en faisant l'hypothèse d'une distribution aléatoire permanente. C'est un système d'équations différentielles à retard, c'est-à-dire dans lequel la valeur des variables à l'instant t est fonction de la valeur des variables à un (ou des) instant(s) antérieur(s). L'étude de ce système d'équations met en évidence des propriétés importantes, par exemple le niveau d'équilibre de la fraction nombre de robots/nombre de bâtons ou le nombre de robots nécessaire pour que les objectifs donnés à l'ensemble des robots soient remplis.

Le second exemple est celui d'interactions entre agriculture et forêt, dans un contexte de culture sur brûlis à Madagascar. Des travaux menés sur place par une équipe de l'IRD ont d'abord donné lieu à une réflexion sur l'apport de la modélisation dans ce contexte, et un modèle conceptuel à base d'agents a été élaboré et présenté dans un cadre pédagogique (Hervé & Treuil, 2005). Ce modèle met en œuvre, d'une part, une représentation spatialement explicite d'un territoire villageois, divisé en parcelles d'un hectare, d'autre part, une représentation explicite des agriculteurs avec leurs règles de décision : choix de nouvelles parcelles à défricher et mettre en culture, mise en jachère de parcelles cultivées. Les agents de ce modèle sont les agriculteurs, et l'environnement, la grille des parcelles. Les discussions ont conduit à penser que,

sous certaines hypothèses d'homogénéité du territoire et du comportement des agriculteurs, et sous certaines hypothèses sur la dynamique végétale, ce modèle conceptuel à base d'agents pouvait être réduit à un modèle mathématique agrégé contrôlant l'évolution de l'état des parcelles. Les deux variables-clés de ce second modèle sont : 1) la distribution du nombre de parcelles en jachère selon l'âge de la jachère ; 2) la distribution du nombre de parcelles cultivées selon la durée depuis laquelle elles sont cultivées et la durée de la période de jachère qui a précédé. Le système d'équations du modèle est un système d'équations aux dérivées partielles qui contrôle l'évolution de ces deux distributions. Il s'agit donc bien d'un modèle agrégé dans lequel toute représentation explicite des agriculteurs et des parcelles a disparu.

Formuler mathématiquement les règles de comportements individuels

Les mathématiques fournissent également des formalismes pour exprimer les règles de comportements individuels et les interactions entre agents. La mise en programme de ces règles peut alors s'appuyer sur cette expression mathématique préexistante et communicable en tant que telle. Nous en donnons trois exemples, sans entrer dans le détail des équations.

Dynamique de polarisation

La dynamique d'alignement d'individus entrant dans un réseau d'interactions est illustrée dans le cas d'un banc de poissons (Adiou *et al.*, 2003). Il y a compétition entre deux mécanismes locaux : la tendance des poissons à s'aligner sur les autres et l'agitation globale. La formulation mathématique est exprimée en termes de réaction à un signal ou à un stress local. Le système d'équation donnant la variation de l'angle d'orientation de chaque agent est un système d'équations différentielles stochastiques, c'est-à-dire comportant des termes aléatoires. Sous certaines hypothèses (symétrie de la relation de voisinage, homogénéité spatiale), il existe une fonction de l'état de toutes les orientations (état global du système), et tout poisson, tout en réagissant localement, se comporte en fait comme s'il cherchait à minimiser cette fonction globale.

Processus de naissance et de mort

On considère ici une dynamique très simplifiée d'une population d'arbres, basée sur deux fonctions :

- une fonction donnant pour chaque position une probabilité de naissance dépendant des positions de tous les arbres existants ;

- une fonction donnant pour chaque arbre existant une probabilité de mortalité dépendant de la position des autres arbres.

De nombreux modèles de dynamiques de formations végétales sont basés sur ce schéma. L'expression des fonctions est donnée quelle que soit la taille de la population, par exemple en faisant intervenir une somme pondérée des distances entre l'agent ou le lieu concerné par l'événement et les agents voisins. L'expression des probabilités se trouve donc ramenée à un système d'équations. Ici aussi, dans certaines conditions de spécification des probabilités, tout se passe comme si le système cherchait à minimiser une fonction globale.

Des agents dans des champs

Un troisième schéma, très fréquent, fait intervenir un environnement actif qui influence avec sa propre dynamique le comportement des agents et peut être modifié en retour par l'action de ces derniers. C'est l'exemple de systèmes de traces (phéromones) laissées par des insectes et déterminant en retour leurs trajets, d'animaux dans une prairie consommant de l'herbe qui se régénère, de l'émergence et de la structuration de centres économiques (la densité de population est représentée par un champ et les entreprises qui se délocaliseront, par des agents).

Ce schéma d'interactions a trouvé des formulations mathématiques, en particulier chez Schweitzer² (2003). Cela se traduit par des équations du mouvement des agents en fonction du gradient du champ et d'un mouvement brownien, des équations de la dynamique du champ intégrant un terme autonome de type réaction-diffusion et un terme d'interaction avec les agents, et enfin des équations contrôlant la dynamique des états internes des agents en lien avec l'état local du champ.

Techniques d'agrégation

L'intérêt porte ici sur les voies mathématiques permettant de dériver un système d'équations portant sur des variables agrégées à partir de l'expression formelle des règles de comportement individuel d'un modèle à base d'agents. On retrouve le problème général de la réduction de la complexité des modèles, décliné dans ce cas par le passage entre niveau « micro » et niveau « macro ». Ces techniques sont donc, dans leur esprit et leur objectif, à rapprocher des méthodes d'« agrégation de variables » (Auger et Bravo de la Parra, 2000). Ces dernières s'appliquent lorsque deux processus sont en œuvre, dont l'un

est beaucoup plus rapide que l'autre, par exemple une migration rapide combinée à une infection ou à un processus de reproduction lent. On fait alors l'hypothèse que cette rapidité d'un processus permet d'établir à chaque temps un état d'équilibre. L'évolution du système est vue comme une succession d'états d'équilibre. C'est cette évolution, sous l'effet du processus lent, qui est décrite par le modèle agrégé.

La possibilité du passage d'un modèle à base d'agents vers un modèle plus simple implique une certaine dynamique du système complet décrit par le modèle agents. Au-delà d'une phase initiale, les trajectoires du système complet dans l'espace de ses états possibles finissent par se restreindre à demeurer dans une certaine région de cet espace ; on dit alors que le système perd ses degrés de liberté théoriques (la totalité de l'espace possible) pour ne garder que ses degrés de liberté effectifs (la dimension de la région finale) (Dieckmann et Law, 2000). Cette perte de degré de liberté, liée aux attracteurs du système, signe une propriété du modèle agents lui-même.

Mais, en pratique, la possibilité réelle de passer d'un modèle agents à une formulation mathématique agrégée est liée à un certain nombre d'hypothèses simplificatrices rarement réunies. De plus, ces hypothèses simplificatrices (par exemple, une homogénéité spatiale du système considéré à une certaine échelle) peuvent se trouver vérifiées au cours d'une certaine période de la vie du système et devenir invalides au cours d'une autre période, le passage d'une période à l'autre manifestant un « changement de phase » du système : à l'instant t , le système peut être agrégé, alors qu'à un instant ultérieur, cela devient impossible.

Ces réserves étant faites plusieurs approches, liées d'ailleurs les unes aux autres, peuvent être envisagées.

1. Les approches partant d'un bilan local décrivant à chaque pas de temps les mouvements d'entrée et de sortie des agents dans une portion donnée de leur espace d'état individuel. De telles approches sont notamment pertinentes lorsque la nature de cet espace d'état individuel et le nombre des agents permettent de parler de distribution de densité – nombre d'agents par unité de volume – sur cet espace. Elles ont comme résultat des systèmes d'équations de différents types, à choisir selon le degré d'hétérogénéité spatiale observé :

- des équations en dérivées partielles qui donnent, à chaque temps t et en chaque point r de l'espace, la variation de la densité des agents (équations de réaction-diffusion de type Fokker Planck : voir, par exemple, Flierl *et al.*, 1999) ;
- un système d'équations qui donnent à chaque temps t la variation de la densité moyenne, d'une part, la variation des corrélations de densité à des distances données, d'autre part (Law et Dieckmann, 2000).

² Voir la page de F. Schweitzer sur le site Internet du Department of Management, Technology, and Economics (D-MTEC) du Federal Institute of Technology (ETH) de Zurich (<http://www.ais.fhg.de/~frank/index.html>).

2. Les approches partant de bilans globaux décrivant le mouvement de l'ensemble du système dans l'espace de ses états possibles. Les trajectoires considérées sont bien ici celles du système, et non celles des agents pris individuellement. On simule un grand nombre de systèmes. Les successions des configurations suivent un processus de Markov³. On essaye d'établir la loi d'évolution de la probabilité de trouver des systèmes dans une aire donnée de l'espace d'état.

L'application des mathématiques aux approches agents rencontre la difficulté de formaliser l'espace des états possibles du système, autrement dit celle de savoir ce qu'on calcule dans une simulation agents. Certaines modélisations font intervenir un nombre fixe d'agents de nature identique, comme les automates cellulaires, des modélisations dites lagrangiennes de phénomènes d'écoulements (Servat *et al.*, 1999), de formation de bancs de poissons (Adiou *et al.*, 2003). En ce cas, la formalisation de ce qu'est l'espace des états possibles du système est simple. Mais la nature informatique des modèles agents permet aisément de faire intervenir des agents de nature différente, pouvant apparaître ou disparaître du monde simulé (Picioreanu *et al.*, 2004), dotés de structures internes complexes et entretenant entre eux des « relations sociales » variables (Lamjed, 2003). Certains ont pu envisager des modèles agents où le nombre et la nature des attributs d'un même agent pouvaient évoluer au cours du temps (Alain Cardon, communication personnelle).

Cette variabilité et cette diversité jointes au caractère stochastique et souvent discret des processus posent effectivement des problèmes de représentation. Mais les mathématiques ne sont nullement démunies d'outils pour les aborder : elles proposent des formalismes restant valables que le nombre d'agents soit variable ou fixe, ou encore qu'il soit grand ou petit, et donc que l'on puisse raisonner ou non en termes de densité.

Débat

Le débat a porté sur les rapports entre modélisation mathématique, simulation informatique et approches agents. Les questions posées sur l'exposé de Jean-Pierre Treuil reviennent sur les avantages et les limites des modèles à base d'agents. Puis, à partir du parallèle entre mise en équation et mise en programme, se dégage l'intérêt qu'il y aurait à théoriser les modèles à base d'agents.

³ Dans le cas d'un temps discret, l'évolution d'un système suit un processus markovien lorsque les états possibles de ce système à l'instant $t + 1$ ne dépendent que de son état à l'instant t . L'histoire antérieure du système n'intervient pas. L'état à l'instant t résume tout le passé. Des formalismes mathématiques élaborés transposent cette notion dans le cas d'un temps continu.

Claude Millier lance la discussion en situant les approches agents parmi les autres méthodes de modélisation et en s'interrogeant sur leur domaine spécifique.

Claude Millier : Les considérations exposées par Jean-Pierre Treuil renvoient à la dualité entre les approches eulériennes et les approches lagrangiennes en mécanique. Les premières sont centrées sur la description de ce qui se passe en des points fixes de l'espace. Elles agrègent les caractéristiques des particules dans des propriétés spatiales. Dans les secondes, le monde est vu par les particules : les équations donnent les lois gouvernant le destin de chacune d'elles. Prolongeant ces approches lagrangiennes, l'individualisme méthodologique conduit à construire des modèles agents en mettant l'individu au centre. Leur avantage est de décrire de façon réaliste les interactions. En biologie, en écologie, en sciences sociales, il y a un avantage effectif à raisonner en modèle agents, à cause précisément de cette capacité à fournir un protocole de représentation d'interactions complexes. Mais dans quels contextes exactement ces avantages sont-ils déterminants ou constituent-ils un « amusement sérieux » ? Est-ce qu'ils permettent réellement de mieux traiter le complexe ? Pour aborder cette question, il faut distinguer entre complication (travail d'un ingénieur, quelque chose qui se décompose en tâches et sous-tâches) et complexité (la solution ne résulte pas d'une succession de décompositions). À quel moment cette décomposition des choses selon les différentes spécialités des sciences (selon la démarche analytique de Descartes) pose-t-elle problème ? On paye, sur certains problèmes liés à l'étude des écosystèmes, le fait que l'on raisonne en pensant que c'est compliqué et en utilisant la division normale des sciences. Les représentations complexes dans lesquelles interviennent plusieurs échelles, depuis le champ cultivé jusqu'aux écosystèmes et encore plus si l'on introduit l'homme, sont un domaine d'application des modèles agents. Par contre, ces derniers ne semblent pas pertinents pour des questions seulement populationnelles et quantitatives, où, en dehors des cas où les hétérogénéités sont trop importantes, les mathématiques apportent des solutions agrégatives.

Jean-Pierre Treuil : La réponse à cette question – dans quel contexte les modèles agents sont un amusement sérieux – ne peut être donnée que par la démonstration de la capacité de ces modèles à s'intégrer dans des problématiques scientifiques et à fournir des solutions aux problèmes posés. Le danger, c'est la facilité avec laquelle les modèles agents permettent de créer un « monde », une belle interface, sans se demander quelle est la question scientifique à laquelle on veut répondre. Par exemple, pour répondre à une question délimitée : « Pourquoi les anchois naissent à un endroit précis en Afrique du Sud, à l'est du Cap, dans l'océan Indien ? », C. Mullon et P. Curry construisent un modèle agents inverse, et montrent qu'à

partir du moment où certaines contraintes existent pour la reproduction, le seul endroit où les anchois peuvent se reproduire est bien cet endroit-là (Mullon *et al.*, 2002). Un modèle agents d'un écosystème est d'abord un modèle écologique avant d'être un objet informatique. Il faut l'évaluer du point de vue de la discipline pour laquelle il a été formulé, l'écologie en l'occurrence ; et, pour ce faire, cette discipline « commanditaire » doit expliciter l'objectif proposé. Par ailleurs, il faut se demander, dans chaque cas, quel degré de granularité est nécessaire par rapport à cet objectif. Il s'agit de trouver la granularité – la décomposition en éléments interagissant – nécessaire pour exhiber le comportement collectif observé tout en conservant un souci de parcimonie. Cela devrait permettre d'éviter le risque d'« explosion naturaliste » tendancielle de ces modèles. C'est avec ce point de vue également qu'il faut aborder la question de la validation. Le premier niveau de validation d'un modèle agents, c'est qu'il reproduise le monde observé sur un certain nombre de motifs, de *patterns*, caractérisés qualitativement, voire quantitativement.

Jean-Pierre Müller repose alors la question des relations entre mathématiques et informatique, en s'adressant à Claude Lobry.

Jean-Pierre Müller : Écrire des programmes informatiques, c'est un moyen de créer des formalismes. Accepteriez-vous le terme de « formalisme » (dans le sens de jeu de formes auquel on associe des opérations internes, mécaniques, de transformation de ces formes) pour qualifier les mathématiques ? Les mathématiques ne sont-elles pas un cas particulier de création de formalismes à partir de la langue naturelle et de transformation de formes ? Les automates cellulaires, les automates d'une manière générale, la logique d'inférence sont autant d'outils pour des modèles à penser.

Claude Lobry : Il y a des langages formels, ceux qui permettent de faire fonctionner les ordinateurs, ou qui permettent avec des multiplications ou des additions d'arriver à un fait de calcul, indiscutable. Les mathématiques sont une discipline qui produit des faits de calcul et cherche à donner du sens à des faits de calcul. Les faits de calcul sont acceptés par tout le monde. L'ordinateur peut calculer, mais la plupart de ces faits de calcul ne sont pas intelligibles, ils sont trop compliqués pour l'homme. Dans une simulation sur ordinateur, on se pose la question : « Est-ce que j'ai bien programmé ce que je voulais ? »

Les physiciens expérimentent et font des théories en laboratoire, ils ne sont pas mathématiciens, mais ils utilisent les mathématiques, ce qui entraîne d'ailleurs un renouvellement du champ de réflexion des mathématiques. Pour les modèles informatiques, les gens qui les utilisent vont être obligés de demander à ce que soit construite une théorie de ces modèles. À l'heure actuelle, on n'est pas raisonnable en travaillant sans théorie. Y aura-t-il

une discipline scientifique qui s'appellera « la modélisation » ? Je pense qu'il y aura plutôt des théoriciens de l'écologie, de l'économie, et des mathématiciens s'empareront de ces modèles, comme celui du jeu de la vie, pour travailler dessus.

Jean-Pierre Treuil : Un fait de calcul mathématique aboutit à une certitude, alors qu'un fait de calcul informatique ne permet d'avoir aucune certitude quant à sa généralisation. Il me semble cependant qu'il n'y a pas de différence de nature entre mise en équation et mise en programme. Les deux sont des manières de spécifier des calculs. Par exemple, les formules que j'ai rappelées au sujet des taux de naissances et de disparitions dans les processus naissance-mort, avec une intégrale sur la densité, ne sont que des notations condensées pour décrire un calcul à faire. Le passage de la formule au calcul peut être plus ou moins facile selon le degré de « condensation » de ces notations, leur caractère plus ou moins abstrait.

La différence – mais il me semble qu'on n'a pas assez insisté sur ce point –, c'est que les mathématiques permettent de faire « des calculs sur des spécifications de calcul », par exemple de montrer l'équivalence de deux spécifications, comme dans le cas que nous a exposé Claude Lobry, des calculs du second ordre en quelque sorte. Elles permettent de spécifier de tels calculs sur des spécifications de calculs.

Bien que des formalismes, comme le lambda-calcul (formalisme parmi bien d'autres qui modélise ce qu'est un calcul), aient établi depuis longtemps la continuité entre données et programmes, l'informatique, me semble-t-il, ne s'est pas encore montrée capable de l'exploiter. On ne fait pas de programmes pour transformer des programmes, pour en déduire des propriétés, sauf dans des cas très marginaux (preuves ou synthèses de programmes).

Claude Lobry : Une équation ne spécifie pas un calcul mais un « être mathématique ». Par exemple, π est le « nombre qui, multiplié par le diamètre d'un cercle, donne la longueur de sa circonférence ». Il existe de nombreuses manières de calculer des approximations de π . Une équation aux dérivées partielles exprime des contraintes exercées sur une fonction ; des théorèmes disent que les contraintes ne sont pas contradictoires (théorème d'existence) et spécifient effectivement (unicité de la solution) différentes procédures de calcul. Ici, « calculer » doit être pris dans le sens de décrire des propriétés de l'objet spécifié, par exemple : « il est périodique », et pas uniquement dans le sens de calculer une « valeur numérique ».

Les propriétés des êtres mathématiques sont présentées en général sous la forme de théorèmes. Par exemple : « Toutes les solutions de l'équation logistique convergent vers l'équilibre non nul. » En principe, dans la version

moderne des mathématiques, un théorème (chaîne de formules, déduites les unes des autres de façon mécanique, qui relie les prémisses à la conclusion) est un « calcul », mais ce n'est pas un calcul au sens fort où $6 \times 7 = 42$. En effet, le calcul n'est que potentiel. Toute démonstration pourrait, en principe, être formalisée et ainsi être soumise à une vérification purement mécanique qui lui donnerait toute sa force. Mais le « en principe » n'est pas une clause de style. Dès qu'un résultat est non-trivial, sa formalisation complète est impossible car trop longue à réaliser matériellement. Les résultats mathématiques sont produits par des humains, de façon non-mécanique (intuition, effets de style, etc.), sous le contrôle d'une méthode (la méthode formaliste) qui permet d'aboutir à des consensus sur les points jugés délicats. D'autre part, une démonstration est par elle-même porteuse de sens sur le résultat (on dit qu'une démonstration éclaire) et il existe des philosophes (Wittgenstein, 1956) pour soutenir que le sens d'un résultat mathématique, c'est l'ensemble de ses démonstrations.

Un programme est très différent : c'est un algorithme effectif. Le calcul est effectivement réalisé de façon matérielle par l'ordinateur. Sauf fantaisies rarissimes des électrons, que nous négligeons, le résultat est toujours exact, indiscutable. Il n'a pas de sens, mais on peut tenter de lui en donner un. Dans un système expert pour l'aide à la décision dans la lutte contre les feux de forêt, l'ordinateur fournit la chaîne des décisions logiques qu'il a suivies. Un exemple, que je trouve original, d'usage de l'ordinateur est le suivant : établir les équations de la mécanique pour un système articulé à n degrés de liberté. Pour $n = 1, 2, 3$, ça se fait à la main ; à partir de 4, il vaut mieux demander à un logiciel de calcul formel. Pour $n = 6$, l'ordinateur fournit un système d'équations différentielles qui occupe des dizaines de pages et qui est donc totalement incompréhensible pour un mathématicien. Un programme transforme automatiquement ces équations en un programme qui simule les solutions de ces équations. C'est donc une « mise en programme » qui établit une « mise en équation », traitée à son tour de façon automatique.

L'exemple du chémostat cité par Claude Lobry est en dehors de cette controverse, car il relève de la démarche expérimentale propre à la physique. Claude Millier s'appuie sur cet exemple pour relancer le débat sur les relations entre expérimentation et modèles agents, en posant la question : « Qu'est-ce qui, au niveau des modèles agents, peut être matière à expérimentation ? »

Claude Lobry : Sur un automate cellulaire⁴, il est possible de tester un ensemble d'états initiaux (et non pas

⁴ Un automate cellulaire ou, plus exactement, un réseau d'automates cellulaires est un réseau spatialement régulier d'éléments de nature identique ; l'évolution au cours du temps de chaque élément est gouverné par des règles simples, prenant en compte l'état des éléments du voisinage immédiat. Il existe

tous), judicieusement choisis. Cela revient à faire des expériences comme un physicien sur un objet. Mais cet objet est l'objet informatique résultant de la mise en programme du modèle agents. On travaillera, par exemple, sur un laboratoire virtuel, en interne, sans aucun retour sur la réalité, en se demandant simplement comment cela marche. Cela permet de savoir ce que fait réellement l'outil, mais personne ne le fait, car il faut prendre du temps pour arriver à une conviction raisonnable que le modèle répond correctement à des questions. Ma propre expérience de programmation d'équations différentielles m'a montré qu'il y a toujours des erreurs dans le codage de ces équations, erreurs que j'ai détectées parce que je connais à l'avance les résultats à obtenir grâce au cadre théorique. Sans cadre théorique, comment faire ? Une autre réponse est de ne pas programmer n'importe comment, d'utiliser des langages de haut niveau, de respecter les procédures. Mais comment répondre à la question : « Est-ce que j'ai programmé ce que je voulais ? » sans disposer d'une théorie mathématique ?

Jean-Pierre Treuil : Quel est l'intérêt d'étudier un programme comme un objet expérimental et d'en faire une théorie, à travers un modèle satisfaisant, c'est-à-dire rendant compte de certains aspects d'intérêt de son comportement sous certaines conditions ? Il me semble que la première raison est de savoir, sous une forme communicable, ce que fait ce programme, quels sont les mécanismes de base mis en œuvre, afin qu'il fasse sens pour tous ceux qui veulent l'étudier. Par exemple, dans l'automate cellulaire exposé par Claude Lobry, ce sens était clair, les mécanismes de base entièrement spécifiés et reproductibles, suffisamment simples pour que l'on puisse s'assurer de l'équivalence de deux implantations informatiques différentes de ces mêmes mécanismes. La théorie à rechercher porte alors sur le comportement.

Mais que dire lorsque la base de l'objet à étudier, le programme, et sa spécification ne sont pas clairement délimitées, sont entachées d'un flou qui fait qu'on ne sait pas très bien ce qu'on étudie réellement, qu'on a des difficultés à l'explicitier ?

Jean-Pierre Müller : Trois usages des systèmes informatiques apparaissent dans ces exposés : 1) le système informatique comme modèle ; cela donne des faits de calcul non interprétables et dont on ne peut pas garantir les résultats. L'augmentation du niveau d'abstraction dans la programmation permet de s'assurer de ce que fait le programme, mais il reste effectivement beaucoup de travail dans ce sens ; 2) le programme est un nouvel objet de laboratoire, qui conduit à faire une théorie de cet objet.

une vaste littérature sur les réseaux d'automates cellulaires qui a pour objet l'étude des relations entre les règles locales et la dynamique globale observée au niveau du réseau ; cette dynamique peut être, en effet, d'une très grande complexité, difficile, voire impossible à prévoir sans simulation effective du système.

Mais la théorie d'un objet qui fonctionne demande énormément de travail ; 3) le système informatique pour aider à donner sens. En mathématiques, il y a les faits de calcul et on peut aussi utiliser les graphiques pour mieux comprendre. Est-ce que le système informatique et son écran graphique n'ont pas cette fonction-là aussi ?

Claude Millier complète cette dernière question. Le message envoyé par un écran ou un graphique est-il perçu comme le voulait le concepteur ? Que peut-on contrôler ? Quel est le meilleur support en terme de communication pour les modèles agents : une représentation UML⁵ ou un modèle mathématique ?

Claude Lobry : Les informaticiens travaillant sur la théorisation des programmes pourraient peut-être nous aider à répondre. Par exemple, le réseau Internet est un objet dont on explique le fonctionnement. De nombreuses thèses en écologie sont très codifiées jusqu'au modèle. Ce qui pose un problème de déontologie, c'est que personne ne peut vérifier les nombreuses équations. La façon dont sont faites les simulations n'est pas expliquée et il arrive qu'à la fin, on n'en tire rien ! Dans 10 ou 20 ans, la partie modélisation sera traitée suivant un formatage scientifique. Il y a eu la mode des AFC⁶ (Benzécri, 1976), jusqu'à la diagonalisation des matrices (Lebart et al., 2000), et le mouvement est maintenant stabilisé. Avec l'outil ordinateur s'est développé un langage nouveau. De quoi ont besoin les informaticiens ? D'une « informatique théorique ».

Qui doit faire la théorie du modèle et comment doit-elle se faire ? Dans le cas des modèles faits par les physiciens, il n'y a pas de problème. Ils le font dans le cadre de leurs propres théories. La communauté scientifique des sciences de la vie et particulièrement de l'écologie n'a pas, pour le moment, les outils intellectuels qui permettent de faire ce travail. Souvent, le travail de théorisation d'un modèle (mathématique ou informatique) n'est pas considéré comme un travail de recherche, car le modèle est simple et sa théorie doit rester simple pour être comprise et utilisée par le biologiste. Ma position est que ce travail devrait être considéré comme un service « public » rendu par les disciplines mathématiques et informatiques et qu'il en soit tenu compte dans les évaluations. J'ai proposé en vain à l'Inra de créer un « laboratoire de modélisation » où auraient travaillé des chercheurs à mi-temps : 50 % dans une discipline fondamentale des mathématiques, de l'informatique, et 50 % sur la théorie de modèles fournis pas des chercheurs de l'institut. J'ai l'impression qu'aux États-Unis, ils sont plus avancés dans la voie d'une solution...

⁵ L'UML (*Unified Modeling Language*) est un formalisme graphique, orienté objets.

⁶ AFC : analyse factorielle des correspondances.

On peut aussi se dire que cela n'a pas d'importance et laisser faire en attendant que les choses se stabilisent d'elles-mêmes. Mais il y a le risque de connaître un jour un « accident de modélisation », c'est-à-dire une décision catastrophique prise sur la base des prédictions d'un mauvais modèle.

Références

- Adiou, M., Treuil, J.-P., Arino, O., 2003. Alignment in a fish school: a mixed Lagrangian-Eulerian approach, *Ecological Modelling*, 167, 1-2, 19-32.
- Auger, P., Bravo de la Parra, R., 2000. Methods of aggregation of variables in population dynamics, *Comptes rendus de l'Académie des sciences, Sciences de la vie*, 323, 8, 665-674.
- Benzécri, J.-P., 1976. *L'Analyse des données, 2. L'Analyse des correspondances*, Paris, Dunod.
- Dieckmann, U., Law, R., 2000. Relaxation projections and the method of moments, in Dieckmann, U., Law, R., Metz, J.A.J. (Eds), *The Geometry of Ecological Interactions: Simplifying Spatial Complexity*, Cambridge, Cambridge University Press, 412-455 (<http://www.iiasa.ac.at/~dieckman/reprints/DieckmannLaw2000.pdf>).
- Flierl, G., Grunbaum, D., Levin, S., Olson, D., 1999. From individuals to aggregations: the interplay between behavior and physics, *Journal of Theoretical Biology*, 196, 4, 397-454.
- Hervé, D., Treuil, J.-P., 2005. Modélisation Environnement Madagascar (MEM), CD-ROM, Montpellier IRD.
- Lamjed, B.S., 2003. *Simulation multi-agent des comportements des consommateurs dans un contexte concurrentiel*. Thèse d'informatique, télécommunications et électronique, Laboratoire informatique de Paris 6, Université Pierre et Marie Curie, Paris.
- Law, R., Dieckmann, U., 2000. Moment approximations of individual-based models, in Dieckmann, U., Law, R., Metz, J.A.J. (Eds), *The Geometry of Ecological Interactions: Simplifying Spatial Complexity*, Cambridge, Cambridge University Press, 252-270100 (<http://www.iiasa.ac.at/~dieckman/reprints/LawDieckmann2000a.pdf>).
- Lebart, L., Moreneau, A., Piron, M., 2000. *Statistique exploratoire multidimensionnelle*, Paris, Dunod.
- Legay, J.-M., 1997. *L'Expérience et le modèle*, Paris, INRA Éditions.
- Lerman, K., Galstyan, A., Martinoli, A., Jan Ijspeert, A., 2001. A macroscopic analytical model of collaboration in distributed robotic systems, *Artificial Life* 7, 4, 375-393.
- Lévy-Leblond, J.-M., 1982. *Physique et mathématiques*, Paris, Le Seuil.
- Martinoli, A., Ijspeert, A.J., Montada, F., 1999. Understanding collective aggregation mechanisms: from probabilistic modelling to experiments with real robots, *Robotics and Autonomous Systems*, 29, 51-63.
- Monod, J., 1950. La technique de culture continue. Théorie et applications, *Annales de l'Institut Pasteur*, 79, 390-410.
- Mullon, C., Curry, P., Penven, A.P., 2002. Evolutionary individual-based model for the recruitment of anchovy (*Engraulis Capensis*) in the Southern Benguela, *Canadian Journal of Fishery and Aquatic Sciences*, 59, 5, 910-922.
- Novick, A., Szilard, L., 1950. Experiments with the chemostat on spontaneous mutations of bacteria, *Proceedings of the National Academy of Sciences*, 36, 12, 708-719.

- Picioreanu, C., Kreft, J.U., Van Loosdrecht, M.C.M., 2004. Particle-based multidimensional multispecies biofilm model, *Applied and Environmental Microbiology*, 70, 5, 3024-3040.
- Schweitzer, F., 2003. *Brownian Agents and Active Particles: Collective Dynamics in the Natural and Social Sciences*, Berlin, Springer Verlag.
- Servat, D., Leonard, J., Perrier, E., Treuil, J.-P., 1999. The Rivage Project: a new approach for simulating runoff dynamic, in Feyen, J., Wiyono, K. (Eds), *Modelling of Transport Processes in Soils*, Wageningen, Wageningen Press, 592-601.
- Volterra, V., 1931. *Leçons sur la théorie mathématique de la lutte pour la vie*, Paris, Gauthier-Villars.
- Wittgenstein, L., 1956. *Bemerkungen über die Grundlagen der Mathematik*, Oxford, Blackwell. Trad. fr. : *Remarques sur les fondements des mathématiques*, Paris, Gallimard, 1983.