

Modélisation : complexifier ou simplifier ?

Hubert Charles

▶ **To cite this version:**

Hubert Charles. Modélisation : complexifier ou simplifier ?. Modélisation : succès et limites, publication des actes du colloques de l'académie des technologies du 6/12/2016, CNRS, pp.81-88, 2016. hal-03506007

HAL Id: hal-03506007
https://hal.inrae.fr/hal-03506007

Submitted on 1 Jan 2022

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.

Actes du colloque

Organisé par le CNRS et l'Académie des technologies

MODÉLISATION

succès et limites

6 DÉCEMBRE 2016

CNRS - CAMPUS GÉRARD MÉGIE
3 RUE MICHEL ANGE - PARIS 16E



cnrs
www.cnrs.fr

ACADÉMIE DES TECHNOLOGIES
POUR UN PROGRÈS RAISONNÉ, CHOISI ET PARTAGÉ

Modélisation : complexifier ou simplifier ?

Réflexions sur la deuxième session du colloque « Modélisation : succès et limites » organisé par le CNRS & l'Académie des Technologies (Paris, 6 décembre 2016)

Hubert Charles

Univ Lyon, INSA-Lyon, INRA, BF2I, UMR0203,
Bât. Pasteur - 11, av. Capelle
F-69621, Villeurbanne, France
hubert.charles@insa-lyon.fr

RÉSUMÉ : Le choix entre petits et grands modèles est un problème récurrent, guidé par des raisons méthodologiques et par des considérations pratiques. Soit on part d'un petit modèle qu'on élargit et complexifie autant qu'il en est besoin, soit on élabore un modèle le plus exhaustif possible, qu'on simplifie si nécessaire. Le choix de la démarche dépend d'une part de la confrontation avec l'expérience, d'autre part des objectifs de la modélisation, mais aussi de la prise en compte de la capacité heuristique des modèles, c'est-à-dire du rôle qu'ils jouent dans la conception de procédés nouveaux.

MOTS CLES : modélisation, simulation numérique, modèles KISS et KIDS

1. Une diversité de modèles pour une diversité d'objectifs

La diversité actuelle des modèles est très importante. Comme le dit Frank Varenne [VAR 2016] [VAR 2014], cette diversité a augmenté au fil des années en suivant les différentes révolutions scientifiques, notamment mathématiques, statistiques puis numériques. Mais la typologie des modèles n'est cependant pas infinie et peut être étudiée. Une classification très basique permet de distinguer les modèles de structures de données, les modèles mathématiques et les modèles numériques (pour une analyse plus fine, voir [PAV 2012]).

Ces modèles empruntent des formalismes particuliers comme les mathématiques, la logique, la simulation numérique, les bases de données relationnelles, la théorie des graphes pour les modèles de réseaux, ou l'informatique, par exemple, pour les modèles individus centrés.

Les objets modélisés sont également de nature très différente : biologique, physique, sociale ou économique. Tous ces objets ont en commun de pouvoir être analysés soit de façon ponctuelle, soit selon un suivi temporel ou encore selon une certaine distribution géographique entrainant alors une structure de corrélation interne des données et des paramètres qu'il est nécessaire d'intégrer dans le formalisme de la modélisation.

On peut souhaiter modéliser ces objets à différentes échelles allant du niveau atomique au moléculaire (*p. ex*., pour les modèles de repliement tridimensionnel des protéines), de l'échelle microscopique à l'échelle macroscopique (*p. ex*., de la cellule à la physiologie humaine). Certains modèles cherchent à atteindre des échelles beaucoup plus grandes, notamment en écologie ou en météorologie où l'écosystème planétaire est modélisé globalement. Même si la physique des particules pour la compréhension de l'univers n'a pas été évoquée ici (de l'infiniment petit à l'infiniment grand), cette intégration des différentes échelles reste encore aujourd'hui un problème très ardu et nous verrons qu'il peut être abordé soit par une complexification, soit par une simplification extrême des modèles selon les objectifs recherchés.

En effet, les modèles sont développés pour une certaine finalité. On peut distinguer ainsi les modèles descriptifs utilisés pour la compréhension d'un phénomène, le diagnostic ou les études rétrospectives. Ces approches privilégient des modèles mécanistes déterministes incorporant parfois des composantes stochastiques qui permettent d'intégrer une part de variabilité issue des objets ou des processus modélisés.

D'autres modèles sont développés dans une optique de prédiction ou d'anticipation. Ces modèles parfois mécanistes ou bien purement statistiques privilégient la composante stochastique. Ils nécessitent une étape de validation sur des données observées qui sont un prérequis important de leur capacité prédictive.

Des modèles permettent encore d'explorer des hypothèses. Dans une démarche où l'expérience est guidée par le modèle, l'expérimentateur va alors réaliser un cercle vertueux de raffinements successifs de son modèle et de ses hypothèses.

Enfin, les modèles peuvent servir pour communiquer, voire pour négocier. Cette dernière finalité de la modélisation a fait l'objet de la troisième session du colloque « La modélisation : vecteur de dialogue entre recherche académique et industrie. »

2. Complexifier ou simplifier, mais à quel prix ?

L'évolution des capacités de calcul des ordinateurs a formidablement augmenté les possibilités de formalisation des modèles. On peut citer comme exemple phare la

résolution numérique des systèmes d'équations différentielles qui en physique a permis notamment l'essor des simulations par la méthode des éléments finis et en biologie le développement des modèles à compartiments associés à l'écologie, à l'épidémiologie ou à la pharmacologie.

Ainsi, les modélisateurs sont maintenant confrontés à un compromis entre simplicité et complexité de leurs modèles (axe horizontal de la Figure 1). Il s'agit en effet de choisir entre l'élaboration de modèles complexes intégrant de nombreux paramètres en interaction pour modéliser les phénomènes au plus près du réalisme physique des objets avec le risque de perdre la maitrise du comportement du modèle, ou au contraire, d'élaborer des modèles simples pour contrôler le fonctionnement global du système et en étudier ses propriétés émergentes au risque de perdre le réalisme physique des objets modélisés. Ce compromis entre les modèles KIDS (keep it descriptive, stupid, 4) et KISS (keep it simple, stupid, 5) a été relevé par de nombreux orateurs du colloque.

Léna Sanders [SAN 2017] propose un autre axe de différentiation supplémentaire des modèles en fonction de leurs propriétés génériques ou au contraire spécifiques qui augmente encore la difficulté du compromis que doit affronter le modélisateur (axe verticale de la Figure 1).

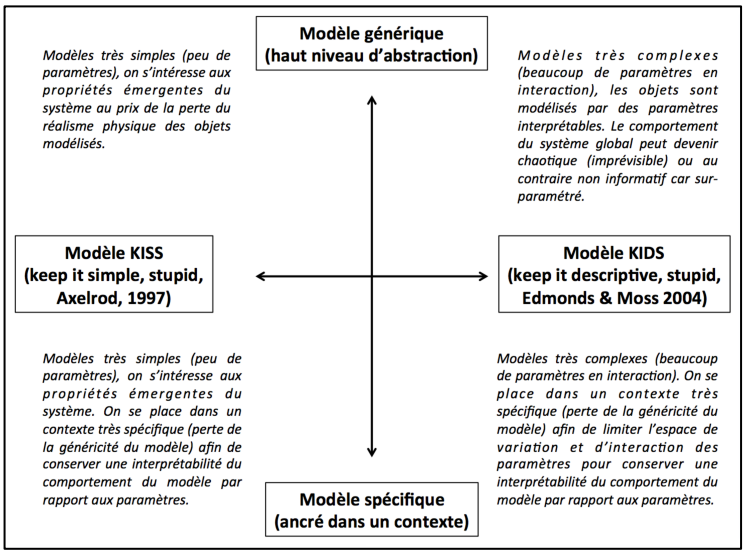


Figure 1. *Carte de compromis pour le choix d'un modèle sur les axes simplicité / complexité et généricité / spécificité, figure adaptée de l'exposé de Lena Sanders [SAN 2017].*

Mais alors, pourquoi simplifier ou complexifier un modèle ? Frank Varenne nous rappelle les raisons de la simplification [VAR 2016]. D'abord ontologique, la nature est économe (principe de sélection naturelle) et il n'est pas nécessaire de multiplier les paramètres sans raison, car il doit exister des mécanismes majoritaires : c'est le principe de parcimonie. D'autres raisons sont de nature épistémique, car on doit pouvoir continuer à comprendre ce que fait le modèle (éviter le piège de la boite noire). De plus, un modèle trop complexe et devenu inintelligible ne permet plus de servir de preuve ou de réfutation d'une hypothèse. Enfin, il existe encore des raisons techniques à la simplification. En effet, des modèles trop paramétrés vont mécaniquement s'ajuster aux données (problème du surajustement) et les valeurs corrélées des paramètres n'auront alors plus aucun réalisme. De même, il est souhaitable de conserver la maîtrise du comportement d'un modèle dans l'espace de fluctuation de ses paramètres (contrôle des émergences et des bifurcations), et ceci peut devenir impossible avec un modèle trop complexe.

A l'inverse, il existe des raisons qui poussent à complexifier les modèles. Du point de vue épistémique, il s'agit d'augmenter le réalisme physique des objets modélisés (hétérogénéité du terrain) au risque de perdre la reproductibilité mathématique du modèle. Les modèles de fonctionnement des écosystèmes pour le suivi et la prédiction de l'évolution de la biodiversité présentés par Michel Loreau [LOR 2017] sont tout à fait dans ce paradigme et montrent qu'il est parfois nécessaire d'intégrer une grande diversité de mécanismes pour comprendre et prédire l'évolution d'un seul processus. Du point de vue technique, la simulation numérique peut apporter certaines solutions comme de simuler un environnement avec un maximum d'hétérogénéité, puis d'utiliser ces simulations dans des modèles rematématisés dans une approche par briques (Légo) qui ne permet plus d'analyser les phénomènes d'émergence ou d'interactions non linéaires entre les briques, mais qui permet de conserver une maitrise du modèle à l'intérieur de chaque brique.

3. Quelques exemples

La table ronde animée par Philippe Davy [DAV 2017] a permis de montrer d'une part l'existence de deux communautés de modélisateurs tournées l'une vers la complexité, l'autre vers la simplification avec parfois peu de communication entre les deux.

Par ailleurs, il semble que c'est globalement la finalité de la modélisation qui dicte l'approche simplifiée ou complexe de la modélisation et non les objets physiques eux-mêmes qui peuvent toujours être appréhendés comme un tout ou comme une somme de parties en interaction.

Enfin, les participants ont tous insisté sur la nécessité de procéder à des aller-retour (ou d'avancer par raffinements successifs) entre les modèles et les données avec des étapes de validation des modèles.

Michel Loreau [LOR 2017] dans son exposé sur les potentialités et les écueils de la modélisation prédictive en écologie nous présente la nécessité de construire des modèles très complexes capable de traduire une hétérogénéité du terrain à l'échelle planétaire. Ces modèles qui nécessitent d'intégrer notamment le changement climatique global associé à l'évolution des activités humaines avec le fonctionnement des écosystèmes pour prédire l'évolution de la biodiversité représentent un formidable challenge pluridisciplinaire [LOR 2010]. Mais il montre aussi les limitations de ces modèles et leur très faible pouvoir prédictif sur de grandes échelles de temps. Il distingue également les prédictions explicatives (dépendantes d'un corpus d'hypothèses) qui peuvent servir comme outil de compréhension d'un système, des prédictions anticipatives (associées à des scénarios plus ou moins probables) qui peuvent servir d'outil de décision. Il montre alors sur un exemple célèbre de préservation de l'habitat de la chouette tachetée en Amérique du Nord qu'une approche purement statistique et extrêmement simple peut être beaucoup plus efficace qu'une approche mécaniste complexe lorsqu'il s'agit de déterminer un seul paramètre (ici le pourcentage d'habitat à conserver) dans le but précis de la préservation de l'espèce. Aurélie Méjean (10) spécialiste en modélisation économique confirme que l'évolution des systèmes économiques associée au comportement humain est imprédictible et que seule une approche par scénarios est envisageable [MEJ 2017].

Léna Sanders présente des modèles de croissance de villes intégrant des informations spatiales, économiques, sociales et historiques [SAN 2017]. Ces modèles se veulent à la fois réalistes (on cherche à comprendre pourquoi une ville se développe plus qu'une autre) et très complexes (de très nombreux paramètres hétérogènes influencent la croissance d'une ville) : le compromis de modélisation sur la carte KISS-KIDS / générique-spécifique (Figure 1) est donc exacerbé ici. L'auteur propose alors une démarche (une trajectoire) consistant à partir d'un modèle spécifique complexe pour aller vers une généralisation du modèle avant la simplification permettant de comprendre les mécanismes principaux.

François Képès présente les potentialités offertes actuellement aux biologistes par la biologie de synthèse qui œuvre dans une spirale d'amélioration entre modélisation mathématique, analyse numérique, construction génétique et tests phénotypiques permettant une véritable ingénierie des systèmes vivants [KEP 2017]. Dans ce type d'approche, il est possible de simplifier le modèle en compressant ou en ignorant certains niveaux d'organisation : le cas du répressilateur d'*Escherichia coli* [ELO 2000] est présenté. Mais on pourra limiter la variabilité des oscillations en complexifiant le modèle et en intégrant plus de boucles de rétroaction. De même, la description d'un processus biologique, comme le repliement tridimensionnel d'une protéine, peut se faire en première approximation par des approches combinatoire à partir de sa séquence linéaire d'acides aminés. Face aux milliards de combinaisons possibles, des heuristiques doivent être développées pour n'analyser que quelques milliers de combinaisons. Mais là encore, Daniel Borgis [BOR 2017] évoque les progrès de la simulation moléculaire en chimie physique qui, partant de l'équation

de Schrödinger, permettait à ces débuts de modéliser le comportement d'un corps pur et s'est ensuite attachée à décrire les transitions de phase, puis les liquides moléculaires et parvient maintenant, avec des ressources de calcul très importantes, à modéliser des protéines [BOR 2017].

Gilles Bernod, spécialiste des réseaux biologiques, évoque au contraire une possibilité de modélisation en biologie qu'il qualifie de jetable (« kleenex ») [BER 2017]. Une modélisation très spécifique du problème donné et complètement dirigée par les hypothèses du chercheur dans un formalisme logique : l'expérimentateur exprime une hypothèse qui est directement transcrite informatiquement. Une expérience de validation est réalisée et les résultats permettent de reformuler l'hypothèse, puis le modèle afin de refaire une nouvelle expérience de validation. L'idée est de faire converger le modèle vers une preuve (ou une réfutation) de l'hypothèse reformulée. Cette approche se situe dans le cadrant bas et gauche de la figure 1. La qualité de l'expérimentation et de la récolte des données est ici cruciale.

4. De la nécessité de développer des formations pluridisciplinaires en modélisation

La modélisation est par essence pluridisciplinaire puisqu'elle interroge des objets (physiques, biologiques, sociaux, économiques) en utilisant un formalisme mathématique et/ou informatique. Elle est de plus omniprésente et nécessaire pour la compréhension des enjeux majeurs de notre société, comme les problématiques du réchauffement climatique, de la préservation des ressources énergétiques et de la biodiversité. Des formations alliant notamment les sciences du vivant avec l'informatique et les mathématiques sont donc absolument nécessaires pour former les cadres en entreprise et les chercheurs de demain. C'est ainsi que depuis les années 2000, de nombreuses formations se sont développées à ces interfaces disciplinaires, sous les termes de « bioinformatique » pour les sciences du vivant ou « d'analyse des systèmes complexes » pour les sciences physiques. Mais il faut en convenir, avec assez peu de succès par rapport aux enjeux et à la demande d'emplois du secteur : le pluridisciplinaire fait peur aux étudiants !

Aussi voit-on maintenant fleurir de multiples formations disciplinaires qui partant d'une spécialisation en mathématique, en informatique, en physique ou en biologie s'ouvrent à une autre discipline sous la forme de masters ou de spécialisation en dernière année d'école d'ingénieur. Cette organisation rassure les étudiants, qui se sentent spécialistes d'un domaine, et sans doute aussi, quelques employeurs. Elle est vraisemblablement responsable du cloisonnement encore très marqué entre les différentes communautés de modélisateurs mentionnées durant ce colloque.

Pour répondre à ce besoin de pluridisciplinarité, une formation en bioinformatique et modélisation (BIM) a été créée en 2000 à l'INSA de Lyon sous

l'impulsion des professeurs Jean-Michel Fayard, Christian Gautier et Alain Pavé dans la lignée de l'école de modélisation lyonnaise portée par Jean-Marie Legay. Cette formation d'ingénieur du département Biosciences est volontairement non spécialisée en donnant une part équivalente à l'enseignement de la biologie, des mathématiques et de l'informatique [BIO 2017]. Les 26 ingénieurs formés chaque année se placent dans les secteurs de la santé et du médicament, de l'environnement ou de l'agroalimentaire. Même s'ils s'insèrent sans aucune difficulté sur le marché du travail y compris à l'international, ils se heurtent encore à quelques difficultés lors des entretiens d'embauches : « êtes-vous biologistes, informaticiens ou mathématiciens, demandent les recruteurs ? » Le chemin vers l'interdisciplinarité est encore long.

7. Références

[AXE 2014] AXELROD R., "Handbook of Research on Nature Inspired Computing for Economy and Management", Eds. Jean-Philippe Rennard, Hersey, 2005.

[BER 2017] BERNOT G., "*Participant à la table ronde Modélisation : complexifier ou simplifier* " , Colloque Modélisation : succès et limites, CNRS & Académie des Technologies, Paris, 6 décembre 2016.

[BIO 2017] http://biosciences.insa-lyon.fr/

[BOR 2017] BORGIS D., "*Participant à la table ronde Modélisation : complexifier ou simplifier* ", Colloque Modélisation : succès et limites, CNRS & Académie des Technologies, Paris, 6 décembre 2016.

[DAV 2017] DAVY P., "*Animateur de la table ronde Modélisation : complexifier ou simplifier* ", Colloque Modélisation : succès et limites, CNRS & Académie des Technologies, Paris, 6 décembre 2016.

[EDM 2004] Edmonds B., Moss S. "*From KISS to KIDS – an anti-simplistic modelling approach*", in Multi-Agent and Multi-Agent-Based Simulation, Eds. Paul Davidsson, Brian Logan, Keiki Takadama (dir.), Springer, 2004.

[ELO 2000] ELOWITZ M., LEIBLER S. "*A synthetic oscillatory network of transcriptional regulators*", Nature, 403:335-338, 2000.

[KEP 2017] KÉPÈS F., "*Modélisation pour la biologie de synthèse",* Colloque Modélisation : succès et limites, CNRS & Académie des Technologies, Paris, 6 décembre 2016.

[LOR 2010] LOREAU M. "*The challenges of biodiversity science.* Excell. Ecol., 17:1–151, 2010.

[LOR 2017] LOREAU M., "*Potentialités et écueils de la modélisation prédictive en écologie*", Colloque Modélisation : succès et limites, CNRS & Académie des Technologies, Paris, 6 décembre 2016.

[MEJ 2017] MEJEAN A., "*Participant à la table ronde Modélisation : complexifier ou simplifier* " , Colloque Modélisation : succès et limites, CNRS & Académie des Technologies, Paris, 6 décembre 2016

[PAV 2012] PAVÉ, A., "Modélisation des systèmes vivants", Eds. Lavoisier (Hermes Science Publications), 633p., 2012.

[SAN 2017] SANDERS L., "*Représentation stylisée versus réaliste de l'espace géographique dans la modélisation des phénomènes socio-spatiaux",* Colloque Modélisation : succès et limites, CNRS & Académie des Technologies, Paris, 6 décembre 2016.

[VAR 2014] VARENNE F., SILBERSTEIN M., DUTREUIL S., HUNEMAN P., "Modéliser & simuler, Epistémologies et pratiques de la modélisation et de la simulation", Eds. Paris, Matériologiques, Tome 1, 980p., 2013, Tome 2, 770p., 2014.

[VAR 2016] VARENNE F., "*Histoire de la modélisation : quelques jalons*", Colloque Modélisation : succès et limites, CNRS & Académie des Technologies, Paris, 6 décembre 2016.