

Modélisation et interactions entre mathématiques et biologie : l'expérience du master professionnel didactique de l'université Paris Diderot

Michèle Artigue, Yann Dartois, Nicolas Pouyanne et Guy Rumelhard
Groupe Modélisation – IREM Paris 7

Résumé

De plus en plus, les curricula mettent l'accent sur la nécessité de connecter l'enseignement des mathématiques à celui des autres disciplines scientifiques et à des situations de la vie réelle, et par voie de conséquence sur la modélisation mathématique. Comme en témoigne l'ouvrage issu de l'Étude ICMI 14 consacrée à la modélisation et aux applications des mathématiques dans l'enseignement, très souvent, les enseignants de mathématiques, de par leur formation, sont peu préparés à une telle évolution. Ils n'ont aucune expérience d'un réel travail de modélisation et leurs interactions avec les enseignants des autres disciplines sont limitées. C'est notamment le cas en France et ceci nous a conduits à proposer un enseignement de modélisation, au sein du master professionnel Didactique de l'université Paris Diderot. Dans cette contribution, après avoir décrit l'organisation de cet enseignement, nous centrons la réflexion sur les interactions entre mathématiques et biologie qu'il permet de mettre en place. Nous le faisons en présentant et analysant un certain nombre de projets qui ont été réalisés par les étudiants depuis la création de cet enseignement.

1. Introduction

De plus en plus, les curricula mettent l'accent sur la nécessité de connecter l'enseignement des mathématiques à celui des autres disciplines scientifiques et à des situations de la vie réelle, et par voie de conséquence sur la modélisation mathématique (Conseil scientifique des IREM¹ 2004). Comme en témoigne l'ouvrage issu de l'Étude ICMI² 14 consacrée à la modélisation et aux applications des mathématiques dans l'enseignement (Blum *et al.*, 2007), très souvent, les enseignants de mathématiques, de par leur formation, sont peu préparés à une telle évolution. Ils n'ont aucune expérience d'un réel travail de modélisation et leurs interactions avec les enseignants des autres disciplines sont limitées. C'est notamment le cas en France et ceci nous a conduits à proposer un enseignement de modélisation, au sein du master professionnel Didactique de l'université Paris Diderot. Dans cette contribution, après avoir décrit l'organisation de cet enseignement, nous centrons la réflexion sur les interactions entre mathématiques et biologie qu'il permet de mettre en place. Nous le faisons en présentant et analysant un certain nombre de projets qui ont été réalisés par les étudiants depuis la création de cet enseignement.

Dans cet enseignement, la priorité a été donnée au vécu d'une expérience de modélisation sans se situer d'emblée dans une perspective de transposition didactique. Organiser et faire vivre ces expériences avec des projets renouvelés chaque année, permettre leur capitalisation progressive, a mobilisé jusqu'ici l'énergie du groupe modélisation de l'IREM. La contribution qui a été présentée au colloque n'avait d'autre prétention que de décrire une expérience sans aucun doute originale, dans le contexte français tout au moins, et permettre d'échanger à son propos. Dans la dernière partie du texte, nous revenons cependant sur les choix didactiques effectués dans cet enseignement et sur les conséquences qu'ils nous semblent avoir sur les types de travaux et d'apprentissages réalisés par les enseignants-étudiants l'ayant suivi. Ceci

¹ IREM : Institut de Recherche sur l'Enseignement des Mathématiques.

² ICMI : International Commission on Mathematical Instruction.

nous conduit à questionner cet enseignement à la lumière de travaux didactiques récents sur la modélisation et sa place possible dans la formation des enseignants, tels ceux présentés dans l'étude ICMI déjà citée ou dans les deux numéros spéciaux publiés en 2006 par la revue *Zentralblatt für Didaktik der Mathematik*. Ce questionnement n'est ici qu'ébauché, mais il montre, nous l'espérons, en quoi l'expérience menée pourrait nourrir une recherche didactique concernant la formation des enseignants et formateurs d'enseignants à la modélisation qui reste, même au niveau international, très limitée.

2. L'enseignement de modélisation du master didactique

Cet enseignement a été créé lors de l'habilitation du master professionnel didactique destiné à la formation de formateurs d'enseignants en 2004-2005³. Il s'agit d'un enseignement optionnel de 3 ECTS suivi chaque année au deuxième semestre par une vingtaine d'étudiants. Il a été créé pour les étudiants du master professionnel spécialité mathématiques qui sont des enseignants ayant au moins cinq ans d'expérience professionnelle mais est aussi ouvert aux étudiants du master recherche et aux étudiants du master professionnel spécialité sciences physiques et chimiques. Cependant vu le faible effectif de cette spécialité, pour l'instant seul un enseignant de physique a suivi cette option.

Après une introduction à la modélisation (3 séances de 3 heures) qui débute par une perspective historique et épistémologique, introduit le cycle de modélisation sous une forme inspirée de (Blomøj & Jensen, 2003) et organise la rencontre avec quelques exemples de modélisations, l'enseignement est organisé autour de l'accompagnement de projets réalisés par des groupes de 3 ou 4 étudiants, exceptionnellement 5. Il alterne sur le semestre des séances de travail de groupes sur les projets, des apports de connaissance en fonction des besoins exprimés par les étudiants, et des présentations et discussions collectives sur les projets en cours de réalisation. Après les soutenances orales, une ou deux séances sont consacrées à un travail plus spécifique sur les questions posées par la transposition didactique des projets réalisés dans l'enseignement ou en formation d'enseignants. Les présentations informatiques préparées pour les soutenances orales et les mémoires écrits des différents groupes sont collectés sur un cédérom pour chaque étudiant et ces derniers ont également accès à tous les mémoires des années précédentes. De plus, un certain nombre de mémoires sont retravaillés par les membres du groupe modélisation pour être mis en ligne sur le site web de l'IREM. Un questionnaire d'évaluation a été élaboré en 2006 et proposé aux étudiants qui avaient suivi l'enseignement les deux premières années. Il a également été soumis aux étudiants ayant suivi l'enseignement en 2007-2008.

Comme cela a été mentionné dans l'introduction, cet enseignement s'adresse à un public dont on suppose qu'il a une expérience très limitée des pratiques de modélisation et de l'interdisciplinarité même si certains des enseignants concernés ont une expérience d'encadrement des Travaux personnels encadrés (TPE) au lycée ou se sont investis dans certains des dispositifs qui, tels les Itinéraires de découverte (IDD), au collège, peuvent engager de telles pratiques. Notre point de vue est qu'une réflexion didactique sur ces pratiques et la façon dont elles peuvent être transposées dans l'enseignement secondaire ne peut faire sens sans une expérience personnelle préalable. C'est pourquoi la priorité est donnée dans l'enseignement à ce vécu d'une expérience de modélisation et à son analyse réflexive, sans lier automatiquement le travail réalisé à une exploitation en classe ou en formation. Les sujets de projets s'autorisent ainsi en particulier à sortir du cadre des seules mathématiques de l'enseignement secondaire même si l'expérience montre que les

³ Michèle Artigue, Guy Rumelhard et François Sauvageot ont mis en place cet enseignement.

mathématiques engagées restent élémentaires et ne sont pas, au moins dans leur forme, celles que l'on rencontre dans les enseignements de modélisation des masters de mathématiques.

Les premiers exemples collectivement travaillés dans la partie introductive sont des exemples historiques : modèles du système solaire de Ptolémée à Copernic et Kepler, travaux de Bernoulli sur la variole, modèles d'évolution de populations... Ils servent à mettre en évidence les questions épistémologiques posées par la modélisation telles qu'elles sont abordées par exemple dans les écrits de G. Israël (Israël, 1996) et N. Bouleau (Bouleau, 1999), des questions trop souvent occultées dans ses transpositions scolaires : pluralité des modèles que l'on peut associer à un même fragment de réalité, complémentarité/concurrence entre modèles, importance du travail de critique des modèles et diversité des façons d'y répondre, adaptabilité des modèles qui rend souvent problématique leur invalidation, problèmes posés par les changements d'échelle dans la modélisation, rapports entre modélisations déterministes et stochastiques, diversité des fonctionnalités de la modélisation et des rapports à la modélisation suivant les domaines concernés... Ces exemples historiques servent aussi à introduire de façon élémentaire un certain nombre d'outils mathématiques de base de la modélisation discrète ou continue.

Une demi-séance est également consacrée à la présentation et l'analyse d'un projet réalisé une année antérieure et à ses éventuels prolongements pour aider à préciser les attentes. Les étudiants sont ensuite libres du choix de leur projet mais, chaque année, le groupe modélisation prépare un certain nombre de thèmes de projets et les présente aux étudiants. La présence au sein du groupe modélisation de l'IREM d'un biologiste, Guy Rumelhard, a conduit à proposer régulièrement des projets en relation avec les sciences de la vie et la biologie. C'est sur ces interactions modélisatrices entre mathématiques et biologie qui se sont révélées progressivement au fil des années que nous avons choisi de centrer notre contribution au thème Mathématiques et Réalité du colloque.

3. Rencontres entre mathématiques et biologie

Le mot modélisation désigne un type bien précis de rencontre entre deux disciplines. La biologie fait « feu de tous bois » pour tenter d'expliquer le vivant, son fonctionnement, son origine, son évolution. Elle va ainsi chercher ses modèles en physique, en chimie, mais aussi en linguistique (code, message), en technologie (régulation), etc. et les modèles mathématiques n'ont pas historiquement joué dans ses constructions conceptuelles le rôle fondamental qu'ils ont joué en physique. La chimie des formes emboîtées (stéréochimie) explique par exemple une grande partie des actions biochimiques dans l'organisme (hormones, récepteurs, médiateurs nerveux et immunitaires, enzymes, protéines régulatrices allostériques, etc.) et une grande partie de l'action des médicaments, en faisant appel à l'analogie opératoire suivante : « un clé entre dans une serrure de sûreté et déclenche une action d'ouverture/fermeture ». Une protéine qui a une affinité pour une autre molécule telle l'adrénaline, peut donc être trompée par une molécule artificielle (l'aténolol par exemple qui est un bloquant) qui aurait au moins en partie la même forme et peut donc entrer en compétition pour se fixer sur la protéine réceptrice. C'est le principe de nombreux médicaments. Cette analogie, inventée par Fischer au début du XX^{ème} siècle, peut se moduler. On peut ouvrir une serrure avec plusieurs clés de formes suffisamment voisines, sinon un passe partout ou des clés hiérarchisées. La spécificité est plus ou moins étroite, et présente des degrés. La modélisation mathématique de la forme des molécules qui s'emboîtent (hormone/récepteur) en se déformant réciproquement et de la force attractive des liaisons correspond à un développement plus récent des recherches mais on peut espérer qu'elle rendra le travail de recherche de molécules-médicament équivalentes plus efficace.

Historiquement en fait, une intervention majeure des mathématiques dans les sciences du vivant, « *séisme épistémologique longtemps inaperçu* » (Canguilhem 1994), s’est produite en médecine avec les travaux de Bernoulli visant à déterminer par un calcul de probabilité s’il fallait ou non vacciner les enfants contre le virus humain de la variole (et plus tard celui de la vache). Principal obstacle à la diffusion de ce travail, la médecine a longtemps pensé qu’elle soignait des individus singuliers et non pas des populations et cette approche est restée dans l’ombre. Claude Bernard (Claude Bernard, 1865) a beaucoup contribué à empêcher ce travail de calcul de probabilités et le fait que Mendel (1865) ait introduit l’aléa au cœur des explications biologiques n’a que tardivement été perçu avec sa véritable importance. Désormais les médecins soignent non plus seulement des malades mais des maladies définies objectivement par comparaison d’études faites sur des groupes randomisés définis *a priori* (ou *a posteriori* quand il n’est pas possible de faire autrement). C’est le triomphe de l’« *evidence based medecine* ».

En agronomie, en écologie, en physiologie, « utiliser » des mathématiques a longtemps été essentiellement identifié à « faire des statistiques ». Chaque étude doit être faite sur une population, mais cette population n’est qu’un échantillon d’une population plus grande. Il faut donc établir moyenne, médiane, variance, écart type, etc. avant de comparer les résultats.

Cette vision réductrice, limitée aux statistiques, des relations possibles entre mathématiques et biologie nous avait frappés lorsque notre groupe IREM s’était constitué au moment de la mise en place des TPE, la quasi-totalité des TPE math-bio recensés dans les établissements suivis par le groupe ou mentionnés par les enseignants suivant la formation que nous avons organisée étant de ce type. Dans les projets proposés en master, nous avons tenté de dépasser cette vision en montrant le rôle créateur des modèles mathématiques (Rumelhard, 2001) en biologie, qu’il s’agisse de modèles déterministes ou aléatoires. Ils permettent de créer des plans expérimentaux, d’expliquer ou de prévoir des données biologiques, de prendre des décisions concernant l’action de médicaments par exemple, en analysant des situations biologiques qui introduisent l’aléa au cœur des explications.

Nous développerons ci-dessous à titre d’exemples quelques thèmes proposés en ce sens aux étudiants du master et qui ont donné lieu à des projets effectifs :

- l’épidémiologie et la transmission des maladies contagieuses dans une population humaine,
- la dynamique des pools de gènes dans les populations animales et humaines au cours de la transmission d’une génération à l’autre,
- l’analyse de séquences d’ADN dont la fonction n’est pas connue,
- la représentation de l’ADN par la CGR : Chaos Game Representation.

Nous montrerons également comment le jeu entre continu et discret, équations différentielles et suites récurrentes, et l’utilisation des TICE (tableurs et programmation de calculatrices notamment) permettent de rendre accessible, avec les moyens limités des mathématiques de l’enseignement secondaire, ce rôle créateur des modèles mathématiques en biologie. Le travail typique de modélisation comprend bien souvent en biologie :

- la détermination des paramètres retenus et des hypothèses qui permettent d’élaborer un ou plusieurs modèles pour rendre compte de la réalité,
- l’élaboration d’un plan expérimental permettant d’obtenir des données fiables (groupes homogènes randomisés *a priori* par exemple), le recueil de données empiriques complètes ou d’échantillons représentatifs,
- l’adéquation du modèle à la réalité, sa correspondance terme à terme, ou globale,
- son caractère opératoire : explicatif et/ou prédictif permettant de prendre des décisions,
- la simulation à partir du modèle qui indique éventuellement au biologiste des observations nouvelles à rechercher,
- l’adéquation de la réalité au modèle par la recherche ou la construction expérimentale de

- situations et le recueil de données conformes au modèle,
- la détection d'écarts entre la réalité et les simulations à partir du modèle entraîne selon les cas la rectification éventuelle du modèle, ou bien la recherche de la signification biologique de cet écart.

Le caractère réducteur, simplificateur et rigide du modèle en regard de la réalité biologique continuellement variable en fonction du milieu, au cours du temps et des mutations constitue pour le biologiste un obstacle au fait d'admettre le caractère opératoire du modèle, soit positivement pour rendre compte de la réalité, soit négativement pour apprécier en quoi la réalité s'en écarte.

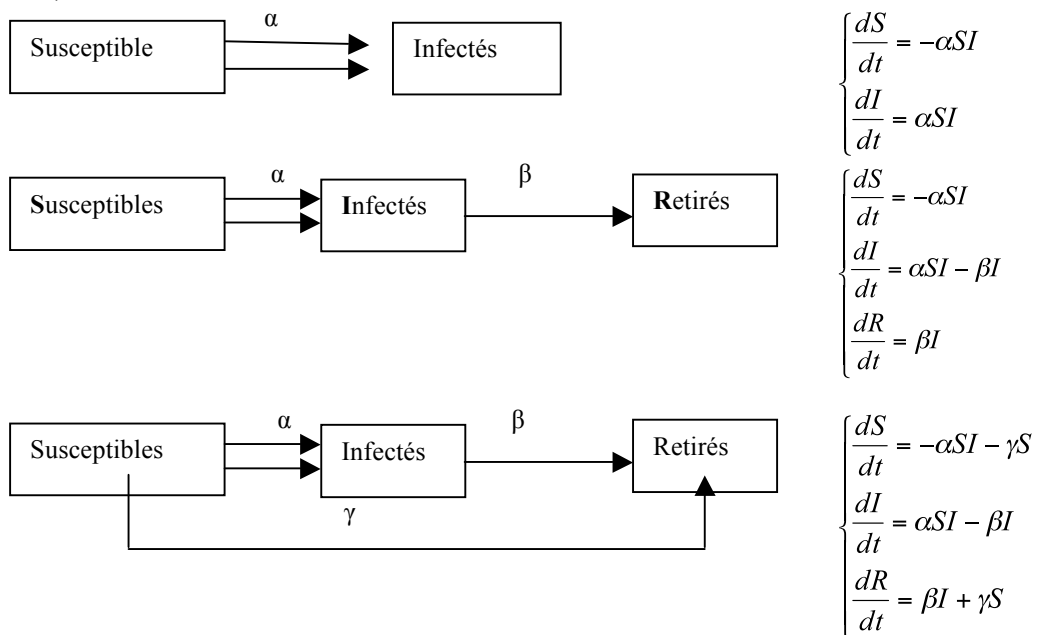
4. Quatre exemples de projets

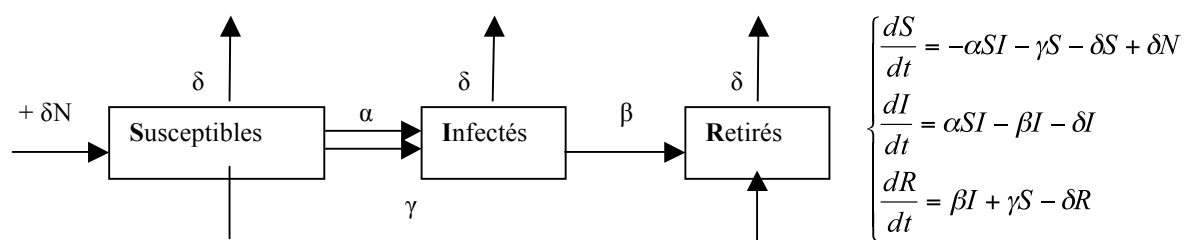
4.1. Transmission d'une maladie contagieuse dans une population humaine

Notre premier exemple rend succinctement compte de projets réalisés par plusieurs groupes d'enseignants-étudiants du master au cours des quatre dernières années, sur le thème de la modélisation de la transmission d'une maladie contagieuse dans une population. C'est un thème sur lequel diverses ressources destinées aux enseignants existent déjà et qui a fait l'objet par ailleurs de transpositions didactiques en TPE (cf. par exemple le site web www.dma.ens.fr/culturemath/html/epidemies.html/ et (Guinnebaud, 2004).

Plusieurs démarches étaient possibles : une approche probabiliste se situant davantage à l'échelle des individus, ou une approche déterministe plus globale à l'échelle des populations. C'est cette deuxième voie qui a été plus particulièrement explorée, en continuité avec les modèles « classiques » d'évolution des populations (Malthus, Verhulst) à l'aide d'équations différentielles (en continu) ou de suites numériques (en discret).

Les travaux réalisés ont, entre autres choses, permis de montrer comment, à partir du modèle compartimental le plus élémentaire utilisé en épidémiologie (modèle « S.I. » de Hamer), par raffinements successifs, on parvient à élaborer des modèles rendant mieux compte de la réalité (modèle « S.I.R. » de Kermack et MacKendrick) ou mieux adaptés à l'étude de problématiques particulières (modèle « S.I.R. » avec vaccination et/ou dynamique vitale).





Les enseignants-étudiants ont par ailleurs replacé la question de la modélisation des épidémies dans une perspective historique, depuis la deuxième moitié du XVIII^{ème} siècle (modèle de Bernoulli) jusqu'à nos jours (modèle(s) d'Anderson et May ...).

La discrétisation des modèles a permis des transpositions didactiques accessibles dès la classe de première par l'utilisation du tableur pour calculer et représenter les situations (activités autour de la propagation d'une épidémie dans une colonie de vacances testées ensuite en première L, simulations d'épidémies en TPE en premières S ...). En terminale S, l'utilisation des suites numériques, des fonctions et des équations différentielles a permis de construire des activités abordant les modèles tant sous l'angle du discret que sous celui du continu, et là aussi, tableurs et graphes ont été mis à contribution. La forme variable de ces activités a reflété la richesse du thème tant du point de vue de son exploitation possible en classe de mathématiques que de celui de l'interdisciplinarité c'est-à-dire du va et vient, de la relance des questions entre les mathématiques et la biologie. Les connaissances biologiques sont ici peu nombreuses et largement connues. Ces transpositions, évoquées ou développées, et pour certaines testées en classe, ont aussi débouché sur des questions concrètes en prise avec la réalité, concernant la propagation de maladies dans des populations humaines (S.I.D.A., peste, lèpre, variole, rougeole) ou animales (brucellose) à partir de situations tantôt simulées, tantôt réelles.

Pour comparer le modèle à la réalité et éventuellement le rectifier, ou bien prendre des décisions à partir des simulations, il faut savoir que les données empiriques sur les maladies contagieuses ne sont pas toujours bien connues. Pour une maladie comme la rougeole due à un virus et pour laquelle il n'y a pas de traitement, mais qui en France n'est pas très grave, dans la majorité des cas, il n'y avait pas jusqu'à une date récente de recueil systématique des cas de maladie ni non plus de vaccination obligatoire. Les sites de « veille sanitaire » recueillent désormais des données sur des échantillons de populations grâce à un réseau de médecins. Ces résultats sont disponibles sur Internet. C'est une condition indispensable pour tester la valeur du modèle. On a également pu effectuer des « expérimentations » c'est à dire enregistrer les effets d'une vaccination systématique d'échantillons d'enfants de différentes tranches d'âge de façon à confronter le modèle aux résultats de cette expérimentation. Ce modèle permet alors de décider quelle proportion de la population il suffit de vacciner pour éviter une propagation. Ces questions de biologie médicale sont elles aussi facilement accessibles car les connaissances biologiques sont peu nombreuses et largement diffusées (virus, bactéries, contagion directe ou indirecte, vaccination préventive, immunisation spontanée ou acquise, mortalité ou guérison spontanée des maladies). Elles ont été plus particulièrement travaillées dans les projets les plus récents sur ce thème.

Pour plus de détails, le lecteur intéressé pourra consulter le site de l'IREM Paris7 : http://iremp7.math.jussieu.fr/sections/groupe_modelisation/, où une partie du travail réalisé est accessible.

4.2. Modélisations de la dynamique du pool des gènes d'une population animale ou humaine au cours des générations : modèle de Hardy Weinberg et modèle de Wright

Les modèles concernés ici font intervenir l'aléa au cœur de la formation des gamètes et de leurs rencontres (on imagine, par delà les individus, une « urne des gamètes »). Ils reposent par ailleurs sur cinq autres hypothèses concernant :

- le grand nombre des individus concernés,
- l'absence de mutation, de migration, de sélection,
- l'absence de croisements entre générations (ce qui dépend de la taille de la population car plus elle est petite, plus les croisements entre générations augmenteront).

Le modèle (Serre, 1997) se présente sous forme d'une formule à établir à l'aide d'un graphe des cas possibles de descendants, tenant compte de la probabilité de la formation des gamètes et de leurs rencontres. Les notions de biologie à connaître sont les suivantes : un caractère visible (groupe sanguin, maladie génétique, etc.) dépend chez l'adulte de deux gènes allèles. Si les deux gènes allèles sont identiques (AA ou aa) les individus sont dits homozygotes, s'ils sont différents les individus sont dits hétérozygotes (Aa). La lettre majuscule désigne un caractère « dominant » c'est-à-dire qui s'exprime toujours lorsqu'il est présent, et la lettre minuscule désigne un caractère « récessif » qui ne s'exprime que s'il est seul (aa) et ne s'exprime pas s'il est en présence du caractère dominant (Aa). Un individu peut donc être porteur de ce gène allèle récessif sans le savoir. La formation d'une génération nouvelle implique la rencontre de deux gamètes (cellules sexuelles) qui ne contiennent toujours qu'un seul des gènes allèles. En effet au moment de la formation des gamètes chez l'adulte, les deux gènes allèles se séparent dans des cellules sexuelles différentes. On modélise cette séparation comme un événement *aléatoire*. On peut établir le fait que les probabilités des AA, Aa, et aa sont respectivement p^2 , $2pq$, q^2 telles que $p^2 + 2pq + q^2 = 1$ et que ces proportions sont statistiquement constantes au cours des générations successives. Le mathématicien peut calculer la fréquence des porteurs du gène récessif (Aa) qui ne le savent pas, connaissant empiriquement q^2 (aa) dans une population donnée. Ce calcul prend de l'importance si (aa) est une maladie grave, et il ouvre la question de l'élimination éventuelle de ce gène allèle dans une population donnée, ce que l'on nomme l'eugénisme (Jacquard 1970, 1977, 1978).

Il se présente un obstacle pour le biologiste. Ce modèle décrit un *équilibre* alors que le biologiste recherche, dans le cadre de la théorie darwinienne, une *évolution* de la fréquence des gènes. Ce modèle étant accepté, le biologiste recherche d'abord des situations empiriques qui y correspondent effectivement. Le cas des groupes sanguins est intéressant car on ne choisit pas son conjoint en fonction de son groupe sanguin et il n'est donc pas soumis à sélection, mais les répartitions géographiques des groupes sanguins ne sont pas toutes homogènes cependant. Le cas du groupe sanguin nommé M, N y correspond statistiquement (Notons qu'en dehors des groupes A, B, O et Rhésus il y a onze autres groupes sanguins dont on ne tient pas compte lors des transfusions sanguines courantes). Le biologiste souhaite ensuite que l'on remette en cause chacune des cinq hypothèses qui vont déterminer des évolutions dans la répartition des gènes allèles dans une population au cours des générations successives. La première étant celle du grand nombre, elle conduit le mathématicien à proposer au biologiste une fluctuation et une disparition éventuelle d'allèles dans une petite population qui n'est pas due à un effet de sélection (modèle de Wright élaboré en 1917) mais qui est due à une variation aléatoire. On parle de dérive génétique et non pas de sélection, ce qui ne contredit pas la théorie de l'évolution mais limite seulement le champ, l'extension de ce concept darwinien de sélection. Les populations humaines primitives, il y a 100 000 ans et plus, étaient certainement peu nombreuses et réparties en petits groupes. La seconde rectification du modèle consiste à considérer le cas de maladies génétiques récessives qui ne

permettent pas la reproduction des (aa) par exemple. La probabilité de l'allèle (a) doit donc diminuer progressivement. Ceci se calcule de deux manières différentes : continue et discrète. Le biologiste doit alors chercher à expliquer les situations empiriques qui s'accordent à ce modèle et celles qui s'en écartent. Cette fois, l'écart signifie qu'il y a un effet de sélection par élimination. Pour rechercher un accord entre la simulation à partir du modèle et la réalité, le biologiste a réalisé des expériences sur de petits animaux (mouches dont le nombre de générations dans des conditions optimales est de 25 par an) dans des cages. Le cas de l'homme ne peut donner lieu à expériences et implique par ailleurs des observations sur des milliers d'années pour disposer de nombreuses générations. C'est rarement envisageable. Les expériences en laboratoire ont été critiquées car elles permettent rarement d'observer des mutations, des migrations et des sélections comme dans la nature. Elles sont donc délicates à mettre en regard des simulations à partir du modèle. Il faut alors rechercher des observations empiriques réalisées dans la nature et non pas au laboratoire dans un espace clos et limité. On voit là à l'œuvre le va et vient entre le biologiste et le mathématicien, entre les observations spontanées, les observations provoquées au laboratoire, les observations empiriques réalisées dans la nature et les divers modèles proposés et raffinés ou rectifiés. Le rôle positif de cette modélisation consiste à « *réduire le caractère métaphorique* » (Conry, 1981) du concept darwinien de sélection qui prend source dans une mauvaise analogie avec la sélection par l'homme des espèces domestiquées. Cette sélection humaine est nécessairement finalisée et l'explication darwinienne élimine au contraire toute finalité dans l'évolution. Le biologiste ne doit pas oublier non plus que l'ensemble des gènes d'une population ne se comporte pas nécessairement comme les billes dans une urne, et peut présenter des processus d'interaction et de régulation. Cette réserve n'invalide pas le modèle proposé. Elle souligne seulement la nécessité éventuelle de le rectifier, ou simplement de constater des écarts entre la simulation à partir du modèle et la réalité observée dans la nature et non pas en laboratoire. La fonction de la modélisation est donc ici différente du cas de l'épidémiologie. Mais, encore une fois, les connaissances biologiques ne sont pas ici très nombreuses ce qui permet d'aborder aisément ce travail.

Les enseignants-étudiants du master ont construit les arbres de transmission des gènes allèles au cours des générations. Ils ont calculé sous forme de suites récurrentes l'évolution de la fréquence du gène qui détermine une maladie génétique : la mucoviscidose, qui diminue l'espérance de vie et ne permet pas, dans les conditions actuelles d'avoir une descendance. Le cas de la mucoviscidose (Serre, 1997) est intéressant car la fréquence de cette maladie en Europe est particulièrement élevée. Ils ont recherché en combien de générations, conformément au modèle, la fréquence du gène qui détermine cette maladie devrait diminuer de moitié. Les calculs conduisent à une estimation de 51 générations donc environ 1000 ans. Ils ont exploré également la piste d'un facteur privilégiant la reproduction des (Aa) pour expliquer la résistance d'une maladie qui ne permet pas la reproduction dans les conditions actuelles. Le travail réalisé a été ensuite transposé pour des élèves de TS sous forme d'une suite organisée d'activités.

4.3. Analyse statistique de l'ADN et modélisation par les chaînes de Markov

Si l'on accepte de réduire le fonctionnement de l'ADN, molécule formée de deux très longues chaînes, au fonctionnement d'une seule chaîne, en oubliant également temporairement le rôle des protéines qui l'entourent et le rôle du deuxième brin, c'est une première étape qui permet la modélisation. Parmi les molécules qui constituent cette chaîne, seules les quatre bases adénine, cytosine, guanine et thymine étant différentes l'une de l'autre, on peut à nouveau réduire l'analyse à celle de la succession des quatre bases notées par une seule lettre A, C, G, T. Par analogie avec l'étude du langage, l'ADN peut être considéré comme un « texte » écrit

avec quatre lettres, ayant un début, une fin, un sens unique de lecture mais aucune séparation entre les « mots ». Il faut encore, et c'est un obstacle important, admettre que si le rôle du chimiste est de rechercher comment cet ADN permet la synthèse d'une protéine, autrement dit sa fonction, le rôle de la modélisation est ici de rechercher une structure formelle, une organisation qui permettra ensuite de proposer au biologiste la recherche de la fonction d'un fragment détecté au milieu d'une longue chaîne dont la signification échappe encore actuellement (Rumelhard, 2006). Chez les Mammifères, plus de 90% de l'ADN n'a pas de rôle actuellement connu. À l'opposé de l'analyse littéraire d'un texte écrit en langue française et pour lequel l'essentiel est de rechercher le sens ou d'apprécier l'esthétique, on va proposer une analyse statistique des lettres isolées, puis des successions de toutes les 16 combinaisons de deux lettres (AA, AC, AG, AT, CA, CC, CG, CT, etc.), de toutes les 64 combinaisons de trois lettres (AAA, AAC, AAG, AAT, etc.) ou plus. On s'appuie sur les chaînes de Markov et on admet que l'existence de la lettre A (par exemple) à un endroit donné de la chaîne dépend de la (ou des 2, 3, 4,) lettre(s) qui précède(nt) immédiatement.

Comme montré dans (Rumelhard, 2006), de façon à se familiariser avec cette modélisation, on peut commencer à l'utiliser sur un texte en langue française, et comparer les résultats avec ceux d'un texte en langue anglaise, etc. On peut également réduire la comparaison à la seule succession de voyelles et de consonnes.

Des fragments de 5 000 à 10 000 lettres de séquences réelles d'ADN sont en fait accessibles gratuitement dans des banques de données sur divers sites internet et peuvent être téléchargés dans Excel. En utilisant les fonctions de ce logiciel, les calculs de proportions se font aisément. On va rechercher en particulier les fragments dont les fréquences réelles s'écartent des fréquences calculées dans le modèle M0 (Prum 2000, 2004). Dans ce modèle M0, on suppose que la lettre qui succède à une lettre donnée ne dépend que d'un « tirage au sort » dans une urne de Bernoulli dont la composition correspond aux proportions observées empiriquement sur une séquence donnée d'ADN. Ce modèle M0 étant peu convaincant, on travaille avec des modèles M1, M2, etc. On recherche alors en quoi la séquence réelle diffère ou non de ces modèles. Ce sont donc ces fragments qui s'écartent du modèle qui sont désignés au biologiste qui se tourne alors vers le chimiste d'une part, et vers la bibliothèque des séquences déjà analysées d'autre part.

Une autre direction de travail consiste à détecter la présence, la position et le nombre de « petits mots » (Schbath, 2003) de deux lettres tel CG dont on connaît le rôle dans l'inactivation d'un chromosome, de trois lettres qui représentent les « codons » c'est-à-dire les lieux de fixation des éléments constituant les protéines et que l'on nomme acide-aminés, ou plus spécialement de TAA, TAG, TGA qui sont des codons start (début de séquence) et stop (fin de séquence), de quatre lettres ou plus (de 8 à 13) dont la signification est connue comme site de fixation de protéines régulatrices ou d'enzymes sur l'ADN. Inversement on peut rechercher le nombre, la répartition et la position de « petits mots » qui n'ont *a priori* aucune signification connue telle la succession ACGT.

L'une des difficultés réside ici dans la quantité de connaissances biologiques nécessaires au mathématicien pour envisager ce travail, ainsi que les connaissances mathématiques pour que le biologiste comprenne ce que le modèle de Markov propose.

Pour se familiariser avec ce type de modélisation et d'approche, les enseignants-étudiants du master ont commencé par analyser un texte écrit en langue française en omettant les accents, les espaces entre les mots, etc. à l'aide du modèle M0, retrouvant approximativement les régularités connues sur les fréquences des différentes lettres. Ils ont ensuite « mélangé » le texte en lui faisant subir une succession de permutations de lettres et, se limitant à une distinction entre voyelles (v) et consonnes (c), ont exploré l'effet du mélange effectué sur les paramètres du modèle M1, en considérant les fréquences respectives des successions vv, vc, cc, cv. Des exercices ont ensuite été composés à partir de ce travail, à destination des élèves

de 1STG, TES, TS, TSTG. Les enseignants-étudiants ont ensuite téléchargé un fragment de 8000 « lettres » d'une séquence réelle d'ADN dans Excel. À l'aide des fonctions de ce logiciel, ils ont calculé les fréquences des quatre lettres isolément (modèle M0), puis des 16 groupes de deux lettres AA, AC, AG, AT, etc. (modèle M1), puis des quatre groupes de trois lettres dont la fonction est connue comme début ou fin de séquences (start : TAA, TAG, TGA, et stop ATG), et comparé les résultats obtenus avec les modèles M0 et M1. Ils ont ensuite conçu des activités exploitant ces données réelles à destination des élèves, en s'appuyant sur l'utilisation d'Excel.

4.4. L'objet et les modèles : de l'ADN à la CGR

Le dernier exemple que nous avons choisi de présenter concerne également l'ADN mais le travail qui y est mené se situe à un autre niveau : celui des représentations associées à la modélisation. Comme rappelé ci-dessus, la modélisation des séquences d'ADN comme des suites finies de lettres de l'alphabet $\{A, C, G, T\}$ constitue un premier niveau de modélisation qui est un affaiblissement drastique de la réalité mettant en avant la question vague de la recherche de régularités structurelles ou statistiques de telles suites, relatives à un individu, une espèce ou une fonction biologique. Idéalement, une telle approche scientifique suppose un aller-retour permanent entre le biologiste, l'informaticien (à cause de la taille des séquences) et le mathématicien : établissement du modèle, étude du modèle en soi, test de la pertinence des résultats qui infirme ou valide le modèle dans les conditions du phénomène observé, ajustement éventuel du modèle, etc.

La modélisation mathématique par des chaînes de Markov constitue un second niveau de modélisation, interne aux mathématiques. On y considère une séquence d'ADN comme une suite de variables aléatoires $(X_n)_n$, finie ou infinie, à valeurs dans l'alphabet $\{A, C, G, T\}$. C'est une manière de prendre en compte toutes les façons dont les lettres peuvent se succéder dans la séquence. En termes plus précis, la question posée est alors celle de la loi du tirage de la suite des lettres : la $n^{\text{ième}}$ lettre est-elle tirée indépendamment des précédentes ? Peut-on déduire la loi du tirage de la $n^{\text{ième}}$ lettre de la connaissance des $n-1$ précédentes ? Certains motifs sont-ils répétés, avec quelle fréquence, selon quelle distribution ? Toute hypothèse sur les lois (jointes) de probabilité des X_n consiste en un nouvel affaiblissement du phénomène réel (nouvelle modélisation). À l'intérieur des mathématiques, compte tenu des séquences « réelles » fournies par l'observation biologique *via* l'ordinateur, joue la dualité entre probabilités et statistiques comme celle entre un modèle et sa mise à l'épreuve de l'expérience (tests).

Le projet dont il est ici question sur la CGR (Chaos Game Representation) se situe à un troisième niveau. Il ne s'agit pas d'une nouvelle modélisation mais plutôt d'une représentation de l'objet modélisé par les modèles ci-dessus (les séquences X_n), introduit pour la première fois par Jeffrey (Jeffrey, 1990).

Décrivons ce système de représentation (Roy *et al.*, 1998) par un exemple : on dessine dans un carré un nuage de points, le $n^{\text{ième}}$ point dessiné correspondant à la $n^{\text{ième}}$ lettre de la séquence. À chaque sommet du carré correspond une lettre de l'alphabet, comme sur le dessin ci-dessous. Prenons une séquence commençant par GAGCACAGTGGGAAGGG... On part du centre du carré, noté O. Le premier point du nuage est placé au milieu du segment formé par le centre et le sommet G, qui est la première lettre de la séquence. Le deuxième point du nuage est placé au milieu du segment formé par ce premier point et le sommet A qui est la deuxième lettre de la séquence. Le troisième point est placé au milieu du segment formé par ce deuxième point et le sommet G qui est la troisième lettre de la séquence, et ainsi de suite par construction récursive. La figure 1 ci-dessous relie les sept premiers points de la CGR de l'exemple.

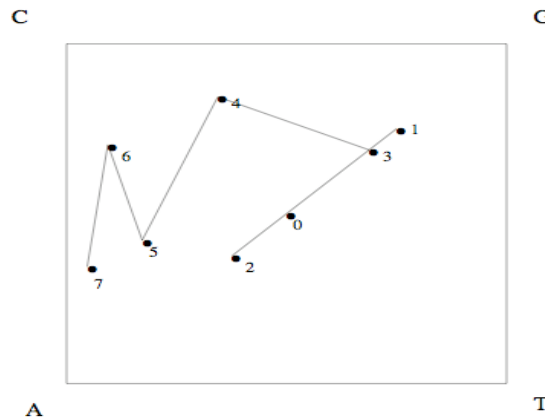


Figure 1 - Représentation CGR du début de la séquence *GAGCACAGTGGGAAGGG*

La figure 2 est le dessin de la CGR de séquences d'ADN des quatre espèces différentes (quelques milliers de points). Elle est extraite de (Deschavanne, Giron, Vilain, Fagot, Fertil, 1999).

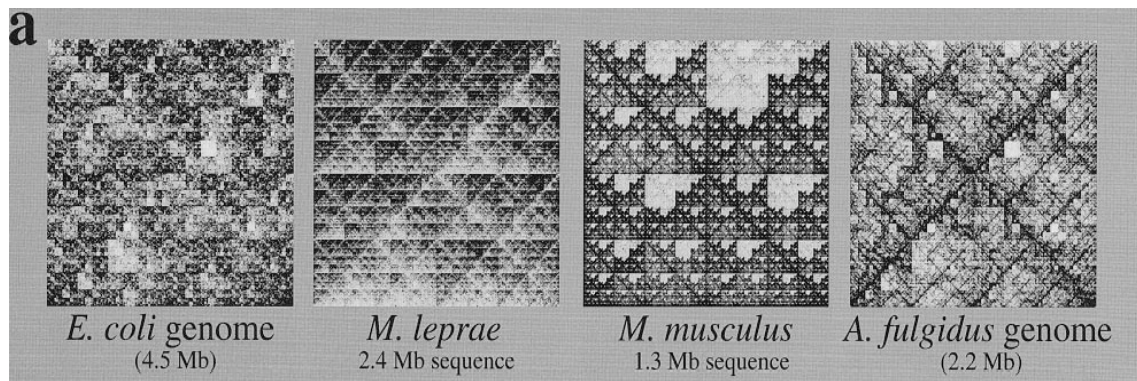


Figure 2 - Représentation CGR de l'ADN de quatre espèces

De visu, on montre ainsi que les lettres d'une séquence d'ADN ne sont pas tirées indépendamment et de manière identiquement distribuée (iid), en comparant les CGR de « vraies » séquences aux CGR de tirages iid. On pouvait se douter que la permanence des espèces n'est pas en ce sens le « fruit du hasard » ; on en a là un élément de démonstration. Ces dessins permettent une discrimination grossière des espèces par la CGR de leur ADN. Comme expliqué dans (Roy, Raychaudhury, Nandy, 1998) : « les vertébrés présentent des régions clairsemées bien distinctes, les invertébrés présentent moins de motifs que les vertébrés, certaines moisissures et les plantes présentent des striations parallèles, les bactéries révèlent des carrés remplis de points uniformément répartis »⁴.

Chaque fois qu'une lettre A apparaît dans la séquence, le point de la CGR qui lui correspond est dans le quart de carré en bas à gauche. Chaque fois que le motif AA apparaît, le second point A correspondant à ce motif est dessiné dans le seizième de carré en bas à gauche. En généralisant ce résultat, on montre par exemple que les zones noircies d'une CGR correspondent à des motifs fréquemment répétés dans la séquence et que l'aspect d'une CGR n'est guère influencé par le début ni la fin de la séquence : seules comptent les fréquences

⁴ Traduction des auteurs

d'apparition des sous-mots. En ce sens, le dessin rend compte de la statistique du « milieu » de la séquence.

Un groupe de quatre enseignants-étudiants a fait de ce thème l'objet de leur mémoire. C'est en des termes à peine plus précis que leur ont été présentés le modèle de l'ADN et la représentation graphique de la CGR. Deux références bibliographiques avancées leur ont été fournies au départ : une thèse de mathématiques en français (Cénac, 2006), un article de biologie en anglais. Dans le cours du semestre, cette bibliographie s'est enrichie. À cause du caractère récent de l'invention de la CGR, aucune source « vulgarisée » n'était alors disponible. Leur premier travail a été de se saisir du modèle et de sa représentation graphique. À la fois pour entrer dans le détail de l'objet mathématique et pour en faire une transposition didactique, ils ont dû faire le choix de notations pour décrire la suite de points de la CGR. Le domaine de travail spontanément adopté est celui de la géométrie affine euclidienne du plan (points et vecteurs), enrichi par la définition récursive d'une suite de points du plan.

Le mémoire produit contient une transposition pour l'enseignement au lycée structurée autour des quatre thèmes suivants :

- Récurrence, homothéties : combinatoire élémentaire, suite de points du plan définis par récurrence barycentrique, intervention des homothéties, exploitation combinatoire de la récurrence vectorielle, utilisation d'un tableur pour représenter et conjecturer, interprétation et comparaison de séquences « vraies » et de séquences produites par un automate probabiliste ;
- Calcul barycentrique : calcul des poids des points de la CGR comme barycentres des sommets du carré, exploitation combinatoire en relation avec l'aspect des dessins ;
- Un peu d'arithmétique : écriture des poids en base 2 ;
- Programmation d'un tableur pour dessiner des CGR, application aux séquences d'ADN des patients atteints de la chorée de Huntington, cas d'école de la biologie. Liens entre propriétés statistiques simples d'une séquence et aspect de la CGR.

Au-delà de l'appropriation du modèle, les activités proposées se résument ainsi : étude mathématique de l'objet (niveau lycée), interprétation graphique de propriétés statistiques des séquences et, réciproquement, interprétation statistique de propriétés graphiques, illustration par des exemples « vrais » issus de séquences réelles, utilisation d'un tableur pour dessiner. Elles n'épuisent pas le sujet de la CGR qui peut vraisemblablement faire l'objet d'autres développements. En particulier, le lien entre les statistiques et les probabilités, qui s'apparente au lien entre un objet et son modèle, pourrait être exploité dans ce cadre.

La variété des sujets proposés à la transposition témoigne de la richesse de l'objet mathématique. Celle-ci est peut-être une condition nécessaire à l'intérêt d'une démarche de modélisation pour l'enseignement des mathématiques (ou la formation). La partie modélisation proprement dite de l'objet biologique (ADN) est restée un peu à l'écart. Cela tient probablement à la complexité de l'ADN : le modèle (suite de lettres) est une simplification extrême de la réalité qui rend difficile l'interprétation sur le réel des résultats sur le modèle (Deschavanne *et al.*, 1999). Une interaction véritable sur le sujet entre biologie et mathématiques, si elle est envisageable dans l'environnement du secondaire, requiert vraisemblablement une collaboration étroite entre biologistes et mathématiciens. Elle est malgré tout subordonnée à l'accessibilité de la biologie du génome dans l'enseignement secondaire.

4.5. Conclusion

Ces quelques exemples, brièvement décrits et non exhaustifs, se veulent illustratifs du travail mené au sein de cet enseignement du master didactique sur les interactions entre mathématiques et biologie. Ils essaient de montrer la richesse de ces interactions quand on se

située dans une perspective de modélisation, une richesse à laquelle nous sommes devenus progressivement sensibles au fil des années. C'est une richesse qui, pour pouvoir être exploitée, nécessite une interaction effective entre mathématiciens et biologistes dans la durée. Les interactions entre mathématiques et biologie sont en effet moins « naturelles » que les interactions entre mathématiques et sciences physiques ; nos cultures respectives, nos langages sont plus distants. Au sein du groupe IREM, cette interaction s'est construite très progressivement à partir d'un travail initialement centré sur l'accompagnement des TPE. Les premiers sujets travaillés empruntaient les voies les plus usuelles, celles aussi les moins coûteuses en connaissances biologiques. Ils touchaient peu à l'aléatoire avec lequel les enseignants se sentent souvent peu à l'aise. Ce n'est plus tout à fait le cas aujourd'hui. En revanche, par rapport au contexte des formations pluridisciplinaires IREM menées sur les TPE, dans l'enseignement de master, l'interaction est limitée à des interactions au sein de l'équipe enseignante ou aux contacts extérieurs que les enseignants-étudiants peuvent être amenés à prendre pour les besoins de la réalisation de leurs projets. Vu les filières de ce master, nous n'avons en effet pas la possibilité d'avoir des groupes mixtes où travailleraient ensemble enseignants de mathématiques et de SVT.

5. Retour sur les choix didactiques effectués

Nous souhaitons revenir dans cette partie, comme annoncé, sur les choix didactiques à la base de cet enseignement et sur leurs effets, en situant cette réflexion par rapport aux travaux de recherche existants dans ce domaine de la modélisation et plus particulièrement de la formation des enseignants et formateurs d'enseignants à la modélisation. Notre formation vise la formation de formateurs d'enseignants, susceptibles d'avoir à conseiller des enseignants ayant à encadrer des TPE au lycée, participer à des itinéraires de découverte ou à un travail sur des thèmes de convergence au collège, ou préparer des formations dans ce domaine. Nous avons fait le choix de concevoir cet enseignement autour du vécu d'une expérience de modélisation et de son analyse réflexive, dans une pédagogie de projets réalisés en petits groupes. Cette importance donnée au vécu d'une expérience de modélisation se retrouve dans les travaux portant sur la modélisation en formation d'enseignants auxquels nous avons eu accès⁵. C'est ce qui ressort par exemple de l'article de Lingefjård (2007) qui synthétise les travaux menés dans l'étude ICMI sur la modélisation dans la formation des enseignants quand il écrit :

« There was no doubt about the fact that use of modeling and technology in instruction with the purpose of enriching students' mathematical learning is valued by future teachers if they are convinced of their impact on their own learning of the content. Therefore in order for modeling to become a part of a teacher's functioning and practice, experiences provided to them in the course of their own mathematics learning should assist them in constructing an image of the teaching and learning that is enhanced by modeling »

C'est aussi le cas dans l'article de Blomhøj et Hoff Kjeldsen (2006) analysant une formation continue développée au Danemark pour aider les enseignants de lycée, peu familiers avec la modélisation, à mettre en place les enseignements de modélisation prescrits par la réforme curriculaire de 2005 :

⁵ Précisons que nous n'avons pas trouvé dans la littérature didactique de travaux portant directement sur la formation de formateurs d'enseignants dans ce domaine. De plus, les travaux concernant la formation des enseignants concernent essentiellement leur formation initiale et c'est d'abord l'initiation à la modélisation des futurs enseignants eux-mêmes qui y est visée.

« Our position is that mathematical modeling – building, analyzing, criticizing – is learned by doing it. In this context problem-oriented project work in modeling presents itself as a suitable didactical choice, and it takes care of some of the central challenges in the new reform. »

Néanmoins, dans cette expérience de formation continue d'enseignants, le lien entre vécu et transposition didactique de ce vécu est bien plus fort que dans notre enseignement. Après un séminaire de trois jours d'introduction à la modélisation, les enseignants doivent, par groupes de 2 ou 3, développer un projet d'enseignement d'une dizaine d'heures, l'implémenter et en analyser la réalisation, des phases d'échanges et de bilans collectifs étant régulièrement organisées. Dans l'article, deux projets aux résultats contrastés sont présentés. Leur comparaison sert notamment à mettre en évidence l'attention à accorder à l'organisation didactique du travail des élèves qui va, dans le projet réussi, se trouver structurée en 4 phases de 3, 2, 6 et 10 séances d'enseignement, les trois premières phases servant à préparer le terrain pour rendre possible et productif le travail autonome des élèves des 10 séances sur le projet lui-même. Ce projet concerne la régulation de la prescription d'un médicament pour un asthmatique.

Cet article montre bien, nous semble-t-il, les conditions à satisfaire pour qu'une formation à la modélisation des enseignants puisse s'effectuer, de façon productive, en interaction étroite avec la pratique. Dans le contexte français, cela semble difficile dans le cadre du dispositif d'enseignement principal alors que c'est sans doute possible dans le cadre des dispositifs secondaires que sont les TPE et l'option Sciences au lycée ou les « Thèmes de convergence » aujourd'hui au collège. Conscients de cette tension entre l'objectif de faire vivre aux enseignants-étudiants du master une expérience de modélisation épistémologiquement satisfaisante et celui de les préparer à l'action didactique, nous avons choisi de donner la priorité à la dimension épistémologique, considérée comme un préalable à la réflexion didactique. Ce choix a bien évidemment des conséquences à la fois positives et négatives. Les projets qui ont été produits au fil des années montrent un travail dans le domaine de la modélisation très varié mais où, chaque année, les différentes étapes du cycle de modélisation sont présentes, si ce n'est dans chaque projet, au moins au niveau collectif. La complexité des questions de modélisation y est bien prise en compte, tout comme la dimension d'analyse critique des modèles et des processus de modélisation. Les enseignants-étudiants ne produisent pas nécessairement eux-mêmes des modèles mais font sens de ceux qui ont été développés pour répondre à telle ou telle question grâce à une enquête approfondie combinant lecture critique de textes et travail sur les modèles eux-mêmes. Ils nous semblent développer également une vision non réductrice des relations entre disciplines comme nous avons essayé de le montrer en prenant l'exemple de la biologie. Les réponses obtenues au questionnaire d'évaluation montrent que cette expérience est effectivement tout à fait nouvelle pour la quasi-totalité d'entre eux. Même ceux qui ont encadré des TPE soulignent en quoi ce travail réalisé à la première personne a été pour eux différent de l'expérience d'encadrement qu'ils avaient connue et leur a permis rétrospectivement de comprendre ce qu'était le travail demandé aux élèves et sa complexité. C'est un enseignement qu'ils perçoivent comme riche, très intéressant mais aussi très déstabilisant. Ils soulignent en particulier la difficulté pour un enseignant de faire face, au début, à l'évidence de la limite de ses connaissances, notamment mathématiques, même sur des sujets relativement élémentaires, mais aussi le plaisir voire l'excitation ensuite éprouvés à apprendre ou ré-apprendre, avec l'aide des collègues et de l'équipe enseignante. Beaucoup aussi soulignent que c'est à l'occasion de la réalisation de ce projet que la notion de travail en groupes est devenue pour eux autre chose qu'un objet de discours pédagogique, car il est devenu clair pour eux que, sans un travail de groupe comportant soutien mutuel et partage des tâches, leur projet n'aurait pu aboutir. Certains aussi découvrent à cette occasion le rôle essentiel de la technologie, tant pour accéder à des

informations et ressources, que pour travailler dans les modèles, effectuer des calculs et des simulations, et enfin présenter leur travail et le rendre plus largement accessible.

En revanche, comme l'on pouvait s'y attendre, sur la question de la préparation à l'action didactique, avec des élèves ou en formation, les résultats de cet enseignement sont bien moins convaincants. Concernant les possibilités de réinvestissements possibles en classe ou en formation d'enseignants de cet enseignement, les réponses au questionnaire d'évaluation de la formation sont partagées. Les enseignants de lycée s'estiment bien préparés à l'encadrement de TPE et plusieurs mémoires ont d'ailleurs incorporé des analyses de TPE, mais ceux de collège perçoivent plus difficilement les réinvestissements possibles. D'une part, l'intersection est souvent faible entre les mathématiques engagées dans les projets et les mathématiques du collège, d'autre part le travail sur les thèmes de convergence qui, dans le contexte actuel, pourrait fournir un cadre favorable à des activités de modélisation peine à s'installer. C'est pourquoi nous avons depuis 2007-2008 suscité des projets autour de thématiques de société telles que crédits, énergie et environnement, mettant au premier plan un travail sur les formules et les ordres de grandeur. Comme cela a été montré dans les descriptions faites, presque tous les projets réalisés comportent cependant une ébauche pour le moins de transposition didactique, à destination d'élèves. Pour des contraintes de temps évidentes, ces transpositions ont rarement la possibilité d'être testées pendant le semestre de cet enseignement. Elles restent donc le plus souvent à l'état de propositions, non confrontées à la contingence, discutées dans la dernière phase de la formation et éventuellement retravaillées et exploitées ultérieurement. Il semble clair que, dans le contexte de cet enseignement, les enseignants-étudiants prennent conscience des limites évidentes des pratiques de modélisation proposées usuellement dans l'enseignement et qu'ils perçoivent mieux en quoi pourraient consister des pratiques épistémologiquement plus satisfaisantes sur un certain nombre de thèmes. Les activités qu'ils fabriquent pour les élèves dans le cadre des projets nous semblent cependant très distantes de cet idéal. Compte-tenu des contraintes institutionnelles⁶, étant faites à quelques exceptions près pour pouvoir être gérées en une ou deux séances avec des élèves non initiés à la modélisation, ou proposées en devoir à la maison, elles sont extrêmement guidées, et l'élève y est généralement invité à suivre pas à pas et à réagir à un travail de modélisation dont il n'a pas l'initiative. On voit bien que le travail de modélisation s'y inscrit dans une pratique de classe marginale et est fortement contraint par cette marginalité. *A posteriori*, on peut penser que la vision épistémologique de la modélisation qui est développée dans cet enseignement et est plutôt, si l'on se réfère aux catégories introduites par Kaiser et Sriraman (2006), de nature pragmatique et socio-critique que conceptuelle et théorique, ne facilite pas la transposition didactique. La modélisation n'y est en effet pas mise en priorité au service du développement de concepts et théories mathématiques à enseigner mais en priorité à celui de la compréhension du monde extérieur aux mathématiques, de la résolution des problèmes qu'il pose et de la réflexion sur le rôle des mathématiques dans les autres disciplines et dans la société. Mais c'est aussi cette vision de la modélisation qui nous semble le plus manquer aux futurs enseignants et formateurs. Elle est sans aucun doute plus exigeante, en particulier parce qu'elle ne peut se développer de façon satisfaisante que dans l'interaction d'un certain nombre d'expertises, non réduite à la seule expertise mathématique. En ce qui concerne les projets à l'interface des mathématiques et de la biologie décrits ici, on voit bien à quel point le peu de contacts existant actuellement entre les disciplines et les enseignants de ces disciplines rend un travail de qualité, respectueux de l'épistémologie des deux disciplines et producteur à la fois en mathématiques et en SVT, est difficile à mettre en place.

⁶ Les transpositions produites ne sont en particulier pas de type TPE car, dans le contexte des TPE, ce sont les élèves eux-mêmes qui choisissent leur sujet. Les études de TPE produites dans le cadre des projets sont des analyses *a posteriori*.

Dans cette unité d'enseignement, enfin, les enseignants-étudiants ne se sont pas véritablement posés en formateurs. Les transpositions proposées sont des transpositions de premier niveau destinées à leurs élèves. Il y a à cela des contraintes temporelles évidentes : prendre connaissance du phénomène de modélisation en général, s'approprier un modèle sans source bibliographique très bien adaptée, en traiter la partie mathématique et enfin travailler à une transposition didactique simultanée des mathématiques en question et de la démarche de modélisation suffit à remplir le semestre, quand on cumule avec un enseignement à temps complet. Penser à une exploitation de la modélisation en formation est peut-être, à ce stade, prématuré mais nous voudrions souligner que plusieurs des projets réalisés ont été exploités dans les formations continues à la modélisation que le groupe IREM a mis en place dans le cadre des plans académiques de formation pour des enseignants de mathématiques, sciences physiques et SVT, un groupe IREM auquel les anciens enseignants-étudiants du master sont invités à s'intégrer.

Comme nous l'avons souligné dès le début de cette contribution, nous relatons ici une expérience d'enseignement qui ne s'est pas inscrite dans une recherche en didactique. Il n'en demeure pas moins que la masse des données recueillies au fil des quatre années écoulées constitue sans aucun doute un matériau qui mériterait d'être plus systématiquement analysé et que, pour ce faire, les catégories produites par les travaux de recherche en didactique dans ce domaine, à la fois sur le plan théorique et méthodologique, devraient être particulièrement utiles.

Michèle Artigue

Université Paris Diderot – Paris 7, Laboratoire de didactique André Revuz

artigue@math.jussieu.fr

Yann Dartois

IREM, Université Paris Diderot – Paris 7

dartois.yann@orange.fr

Nicolas Pouyanne

Université de Versailles Saint Quentin

pouyanne@math.uvsq.fr

Guy Rumelhard

Professeur de SVT, chercheur à l'INRP, IREM, Université Paris Diderot – Paris 7

guy.rumelhard@wanadoo.fr

Références

Blum W., Galbraith P. L., Henn H. W. & Niss M. (Eds). (2007). *Modelling and Applications in Mathematics Education*. The 14th ICMI Study. Springer.

Bernard C. (1865). *Introduction à l'étude de la médecine expérimentale*. Paris : Garnier-Flammarion Editeurs, édition de 1966.

Blomhøj M. & Jensen T.H. (2003). Developing mathematical modeling competence: Conceptual clarification and educational planning. *Teaching Mathematics and its Application*S, 22 (3), 123-139.

Blomhøj M. & Hoff Kjeldsen T. (2006). Teaching mathematical modeling through project work. *Zentralblatt für Didaktik der Mathematik*, 38 (2), 163-177.

Bouleau N. (1999). *Philosophies des mathématiques et de la modélisation*. Paris : L'Harmattan.

- Canguilhem G. (1968). *Études d'histoire et de philosophie des sciences concernant les vivants et la vie* Paris : Vrin (7^{ème} ed 1994).
- Cénac P. (2006). *Étude statistique de séquences biologiques et convergence de martingales*, thèse de l'université Paul Sabatier Toulouse III ; disponible à partir du lien <http://math.u-bourgogne.fr/IMB/cenac/thesis.fr.html>
- Conry Y. (1981). *Organisme et organisation : de Darwin à la génétique des populations*. Réédité In *Darwin en perspective*. Paris : Vrin 1987.
- Comité Scientifique des IREM (2004). *La modélisation*. IREM Paris 7.
- Deschavanne P.J., Giron A., Vilain J., Fagot G. & Fertil B. (1999). *Genomic signature: characterization and classification of species assessed by chaos game representation of sequences*, Mol. Biol. Evol. 16(10),1391-1399.
- Guinnebaud B. (2004). Modèles mathématiques pour une propagation des maladies contagieuses. *Bulletin de l'APMEP* n° 441, 467-479.
- Israel G. (1996). *La Mathématisation du réel*. Paris : Éditions du Seuil.
- Jacquart A. (1970). *Structure génétique des populations humaines*. Paris : Masson.
- Jacquart A. (1977). *Concepts en génétique humaine*. Paris : Masson.
- Jacquart A. (1978). *Éloge de la différence* Paris : Seuil.
- Jeffrey H.J. (1990). *Chaos game representation of gene structures*, Nucleic Acid Res. 18, 2163-2170.
- Kaiser G, Sriraman B. (2006). A global survey of international perspectives on modelling in mathematics education, *Zentralblatt für Didaktik der Mathematik*, 38 (3), 302-310.
- Lingefjård T. (2007). Modelling in teacher education. In, W. Blum, P. L. Galbraith, H. W. Henn, M. Niss (Eds). (2007). *Modelling and Applications in Mathematics Education*. The 14th ICMI Study. pp. 475-482. Springer.
- Prum Bernard (2000). Les chaînes de Markov dans l'analyse des génomes, *Matapli* 62, 24.
- Prum B. (2004). Mathématiques et biologie *APMEP* n° 440, 337-348.
- Robin S., Rodolphe F. & Schbath S. (2003). *ADN, mots et modèles*. Paris : Belin, collection Echelles.
- Roy A., Raychaudhury C. & Nandy A. (1998). Novel techniques of graphical representations and analysis of DNA sequences – a review, *J. Biosc.*, 23(1), 55-71.
- Rumelhard G. (2001). Le rôle créateur des mathématiques en sciences de la vie. Propositions pour les TPE. *Biologie-Géologie (APBG)* n°4.
- Rumelhard G. (2006). Analyse statistique de l'ADN, modélisation à l'aide des chaînes de Markov, simulation et détection de biais. *Biologie-Géologie (APBG)* n°6.
- Schbath S. (2003). À la recherche de mots de fréquence exceptionnelle dans les génomes Images des mathématiques CNRS vol 3.
- Serre J.L. (1997). *Génétique des populations*. Paris : Nathan université. Réédition (2008) Paris : Doin.