

La modélisation à l'Inra

Liste des participants

Groupe de réflexion Modélisation coordonné par Bruno Goffinet

Nom / Prénom	Département
Amigues Jean-Pierre	SAE2, Toulouse
Brunet Yves	EA, Bordeaux
Clément Frédérique	INRIA, Roquencourt
Courtois Francis	ENSIA Massy, CEPIA
Della Valle Guy	CEPIA, Nantes
Di Pietro Liliana	EA, Avignon
Duru Michel	EA, Toulouse
Faivre Robert	MIA, Toulouse
Faverdin Philippe	PHASE, Rennes
Fourichon Christine	ENV Nantes, SA
Franc Alain	EFPA, Paris
Ginot Vincent	MIA, Avignon
Godon Jean-Jacques	MICA, Montpellier
Goffinet Bruno	MIA, Toulouse
Hospital Frédéric	GAP, Le Moulon
Lardon Sylvie	SAD, Clermont Ferrand
Martin-Clouaire Roger	MIA, Toulouse
Monod Hervé	MIA, Jouy en Josas
Seegers Henri	ENV Nantes, SA
Sinoquet Hervé	EA, Clermont Ferrand
Traas Jan	BV, Versailles
Trystram Gilles	ENSIA Massy, CEPIA
Vila Jean-Pierre	MIA, Montpellier
Wallach Daniel	EA, Toulouse

Sommaire

Introduction	1
---------------------------	---

Première partie analyse par thèmes

Thème 1 - Du gène à la cellule	7
Thème 2 - Physiologie, écophysiologie	10
Thème 3 - Physique et biophysique	15
Thème 4 - Ecologie des populations et des communautés, épidémiologie	19
Thème 5 - Gestion des agro-écosystèmes	24
Thème 6 - Agroalimentaire et génie des procédés	29
Thème 7 - Sciences sociales	34

Seconde partie synthèse générale

1 - Rôle de la modélisation dans la démarche de recherche	37
2 - Place de l'INRA	38
3 - Quelques grandes classes de problèmes en modélisation à l'INRA	38
4 - Une culture sur la modélisation à développer collectivement	44
5 - Comment s'organiser à l'INRA pour répondre aux défis de la modélisation ?	45
Encadré : Quelle(s) plate(s)-forme(s) informatiques pour les modèles de culture ?	48
6 - Conclusions et conséquences opérationnelles	51
Points de vue	
Par Paul Bourguine, Henri Caussinus, Jean-Paul Haton, JD Lebreton, Eric Walter.....	53

La modélisation à l'Inra

Introduction

La modélisation prend une place croissante dans tous les domaines d'intervention de l'INRA. La direction générale de l'INRA a donc confié au département Mathématiques et Informatique Appliqués (MIA, anciennement BIA) une mission d'analyse prospective sur le rôle de la modélisation et la place de l'INRA dans les différents domaines concernés, les grands problèmes rencontrés et les moyens d'y répondre (voir lettre de mission ci-après).

La liste des personnes ayant participé aux groupes de travail est donnée à la fin de ce document. Les chercheurs de MIA ont été choisis pour représenter les différentes disciplines mathématiques et informatique concernées par la modélisation (statistique, analyse de systèmes, intelligence artificielle). Les chercheurs "thématiciens" ont été choisis en concertation avec les chefs de département de l'INRA, afin de couvrir l'ensemble des champs de recherche de l'INRA, sans pour autant qu'il y ait un représentant par département.

Modélisation : terminologie et inventaire des problèmes à traiter

1. Définition

On appellera modèle, une représentation mathématique, graphique ou informatique des objets et des relations entre ceux-ci dans un domaine restreint du monde réel, objet d'un questionnement. La modélisation est un moyen d'explicitier la complexité afin de mieux comprendre le fonctionnement d'un système et de prendre les décisions le concernant. Cette définition est relativement cohérente avec ce que propose Jean Marie Legay (L'expérience et le modèle; 1996 INRA Sciences en Question) "... la décision de donner au modèle dans le cadre de la recherche scientifique, une acception instrumentale".

La position de Nicolas Bouleau (Philosophies des mathématiques et de la modélisation; 1999 L'Harmattan) est aussi en phase avec la notre "Les modèles sont des représentations partiales, au bon sens du terme, c'est à dire qu'au sein de multiples possibilités d'expression et de représentation, ils sont le choix d'un parti."

2. Les objectifs

Il est important dans un premier temps de préciser les objectifs du modèle que l'on construit. En effet, la définition de l'objectif est déterminant pour identifier les méthodes pertinentes et la démarche à suivre. Notons que les objectifs sont souvent multiples et liés entre eux.

Deux grands objectifs, à l'intérieur desquels on peut en identifier plusieurs :

- **La connaissance**
 - i. Représenter, formaliser
 - ii. Comprendre
 - iii. Echanger, communiquer (structurer une information massive)
 - iv. Enseigner
 - v. Enrichir (dont l'extraction)
 - vi. Tester (une hypothèse biologique, par exemple)
 - vii. Documentation ou support d'argumentation
- **L'action**
 - i. prédire
 - ii. diagnostiquer
 - iii. concevoir
 - iv. contrôler
 - v. décider, planifier
 - vi. classifier
 - vii.

3. Définition du phénomène

Il s'agit de faire le modèle conceptuel qui précise ce que l'on connaît ou suppose du phénomène étudié. Cette phase est plutôt du domaine des biologistes, physiciens, économistes, ou sociologues (ou plus généralement des "thématiciens" tels que le décrit Jean Boiffin). Cependant, il existe des méthodologies, comme la systémique par exemple, qui peuvent être mobilisées pour aider le thématique à construire ce modèle conceptuel.

Cette définition utilisera trois types de concepts :

- **Les objets**
 - i. ils peuvent être contrôlables ou non contrôlables
 - ii. mesurables, observables ou non
 - iii. simples ou composés (plus ou moins complexes)
 - iv. physiques ou conceptuels
 - v. il peut s'agir d'agents passifs ou actifs
- **les relations fonctionnelles**
 - i. causales
 - ii. structurelles
- **Les événements : il s'agit des événements de discontinuité du phénomène.**

De plus, il est nécessaire de préciser la nature des données dont on pourra disposer relativement aux objets décrits ci-dessus. Un des enjeux du processus de modélisation sera justement de mettre en regard les possibilités d'accès aux données et les objectifs du modèle d'une part, avec le niveau de détail d'une modélisation d'autre part. Ces données peuvent être :

- i. quantitatives : discrètes, continues, spatialisées, fréquentielles, etc.
- ii. qualitatives
- iii. symbolique ou logique (appartient ou pas à un ensemble par exemple)
- iv. déterministes ou stochastiques ou encore floues
- v. issues d'images
- vi. statiques ou dynamiques
- vii. ou enfin, on peut ne pas disposer de données relatives aux objets.

4. Le processus de modélisation

Le processus de modélisation va donc être de mobiliser les connaissances dont on dispose (relations fonctionnelles, informations) pour atteindre les objectifs fixés. Ce processus, couvre alors plusieurs phases qui ne seront pas toujours nécessaires, en fonction des types d'information et des objectifs du modèle. A priori cet ensemble de procédures décrit aussi bien le processus de modélisation mécaniste que la modélisation statistique. Le processus ne parcourt pas linéairement les phases, mais est fait de multiples allers-retours.

- i. Clarification des objectifs précis du modèle et bilan des connaissances du phénomène.
- ii. Conception, formalisation, caractérisation. C'est à ce niveau que se réfléchit le niveau de détail du modèle, en fonction de l'objectif visé et des informations potentiellement accessibles. Il s'agit de traduire sous une forme mathématique (équations différentielle par exemple) le modèle conceptuel. L'analyse systémique peut être mobilisée dans cette phase, de même que des langages informatiques graphiques (comme UML) qui peuvent aider à visualiser les relations entre les objets. C'est à ce niveau qu'interviennent les méthodes de représentation des connaissances de l'Intelligence Artificielle, tel l'analyse ontologique.

- iii. Implémentation et utilisation, y compris la vérification que le modèle conceptuel a été traduit correctement sous forme mathématique (ou logique) et que le code informatique est correct. Pour les modèles d'EDO, il existe de nombreux outils d'implémentation. L'offre est rare pour les modèles individus centrés. Pour les modèles logiques, la phase d'implémentation est très liée à celle de formalisation. Se posent à ce niveau les problèmes de vitesse d'exécution du programme, les problèmes d'algorithmique, et les problèmes de discrétisation dans les résolutions numériques.

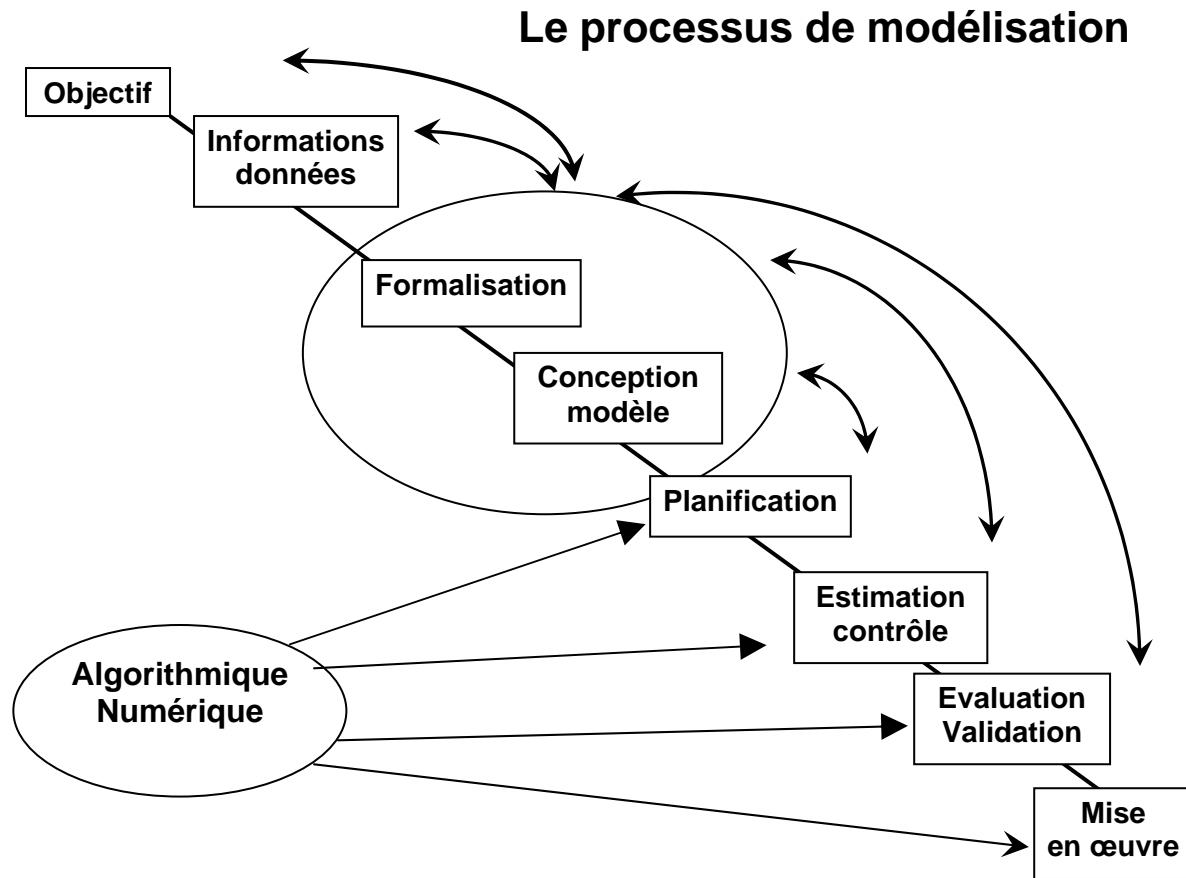
- iv. Compréhension des propriétés. (Remarque : pour ces différentes phases i à iv, les données ne sont pas nécessaires). On peut mettre ici la validation qualitative, ou l'analyse de sensibilité par exemple, ou bien encore l'étude du comportement asymptotique d'un système d'EDP. On peut utiliser des études analytiques quand une telle solution existe. Sinon, on peut faire une étude numérique ou le modèle est considéré comme un monde virtuel inconnu. Les méthodes statistiques peuvent alors s'appliquer.

- v. Planification. Il peut s'agir d'optimiser le choix des données à recueillir ou bien de définir de quelles données il faudrait disposer pour répondre à un objectif. Faut-il enquêter, expérimenter, faire de la bibliographie, faire intervenir des experts, etc. C'est un aspect décisif en forte interaction avec les autres phases. Des méthodologies existent pour le problème d'optimisation du choix des données.

- vi. Identification, estimation des paramètres du modèle. Les méthodes statistiques ne sont pas toujours bien adaptées aux modèles complexes, avec de fortes non linéarités, un grand nombre de paramètres, des données qualitatives. Les méthodes Bayésiennes sont peut-être une solution à certains problèmes.

vii. Validation, et délimitation d'un domaine de validité du modèle pour un critère que l'on définit. Il s'agit bien de validation qualitative et quantitative. Les méthodes statistiques de validation quantitative sont relativement bien développées, mais on peut se poser le problème de leur extension à des quantités vectorielles, comme des distributions ou des images.

Le processus de modélisation



Ces différentes phases achevées, se posent alors les questions de l'action : prédire, commander, diagnostiquer, décider etc. qui font l'objet de méthodologies spécifiques. Par exemple, dans le domaine de la commande, le développement de méthodes de commande robuste est un enjeu important.

Structuration de la réflexion

La modélisation est présente dans presque la totalité des champs de recherche de l'INRA. Il est alors difficile de prétendre décrire exhaustivement les activités de l'INRA concernées par la modélisation. Nous avons donc choisi de définir 7 champs qui couvrent une grande partie des activités de l'INRA, et qui représentent d'une certaine façon une typologie des types de processus rencontrés. En grande partie, mais pas exclusivement, ces champs sont définis comme des échelles d'organisation du vivant.

- 1 Du gène à la cellule
- 2 Physiologie, écophysiologie
- 3 Milieux physiques
- 4 Ecologie des populations et des communautés, épidémiologie
- 5 Gestion des agro-éco-systèmes
- 6 Agroalimentaire et génie des procédés
- 7 Sciences sociales

Pour chacun de ces thèmes, nous avons répondu aux 3 premières questions de la lettre de mission. Cette typologie apparaît a posteriori efficace puisqu'elle nous a permis d'identifier et de rassembler une grande majorité des problèmes qui se posent à l'INRA. Ces analyses sont données dans la deuxième partie du document. Elles ont été transmises aux chefs de département pour qu'ils en valident le contenu. Elles ont conduit à rédiger la synthèse qui suit. Cette première partie a fait l'objet d'échanges lors du conseil scientifique de l'INRA des 24 et 25 juin 2004, qui ont en particulier conduit le groupe à rajouter le thème « Sciences Sociales » absent dans une première version.

Dans un deuxième temps le groupe de travail a rédigé les paragraphes qui répondent aux deux dernières questions de la lettre de mission (ci-après).

Lettre de Mission

A Monsieur Bruno Goffinet
Chef du département
"Biométrie et intelligence artificielle"

Paris, le 25 juin 2003

Objet : Elaboration du périmètre scientifique : « Mathématiques et informatique appliquées, bioinformatique, modélisation »

Le Collège de direction considère que les missions formulées précédemment ¹ pour le Département BIA et récemment précisées à l'occasion de votre nomination comme chef de département, sont pleinement d'actualité et justifient l'existence d'un département autonome, à définition principalement disciplinaire. Il n'envisage pas d'en modifier substantiellement les contours, mais souhaite que son ouverture aux autres départements soit encore accentuée.

Je vous demande donc d'animer dans cet esprit les réflexions préparatoires au futur schéma stratégique de ce département en les orientant résolument vers la prise en compte des besoins et activités des autres départements, et en visant à optimiser la répartition des moyens et l'organisation des coopérations à l'échelle de l'organisme. Deux domaines transversaux, qui sont des enjeux majeurs pour l'INRA, sont visés par cet élargissement : **la bioinformatique et la modélisation.**

Sans oublier les recommandations organisationnelles et thématiques déjà formulées (notamment à propos de l'analyse des risques et de l'épidémiologie), cette réflexion élargie pourra vous amener à confirmer, étendre ou infléchir les actuels champs thématiques du département BIA, ainsi que sa politique en matière d'acquisition et gestion des compétences. Cette réflexion pourra également amener les autres départements à mieux prendre en compte, dans leur propre schéma stratégique, le potentiel d'appui que représente pour eux un département spécialisé en mathématiques et informatique appliquées. Il est donc souhaitable que le schéma stratégique de ce département soit élaboré en forte interaction avec les autres.

La bio-informatique

Le chantier en cours dans l'organisme a dépassé le stade de la réflexion pour déboucher sur certaines décisions opérationnelles. Il doit néanmoins se prolonger, notamment pour définir l'organisation générale et les moyens nécessaires à l'ensemble de l'organisme. La réflexion préparatoire au schéma stratégique que vous animez doit être l'occasion de poursuivre et d'approfondir le volet scientifique de ce chantier.

Il s'agira :

- d'identifier de façon plus exhaustive et plus précise les besoins méthodologiques correspondant aux principaux problèmes rencontrés dans l'exploitation des données de génomique et de post génomique ;
- d'analyser la nature des investissements à consentir pour satisfaire ces besoins : importation d'outils ou méthodes déjà existants, développement de recherches nouvelles, acquisition de compétences correspondant à l'une et/ou l'autre de ces options ;
- de proposer des priorités et lignes directrices pour l'INRA quant à ces investissements, compte tenu de ses atouts, faiblesses et contraintes, comparativement à ceux des autres organismes français et étrangers, ainsi que des coopérations possibles ;
- de proposer les partenariats qui permettront, en complément de cet investissement propre, de répondre à l'ensemble des besoins des biologistes de l'INRA.

La modélisation

Le développement conjoint de la biologie intégrative, de l'écologie et des sciences de l'environnement, de l'analyse des risques de toute nature, et des approches mathématiques en sciences sociales, amène à prévoir une généralisation de la démarche de modélisation au sein de l'INRA, non seulement dans les travaux de recherche

¹ Rappelées en annexe

proprement dits, mais également dans le développement d'applications. Cette généralisation peut faire apparaître des problèmes méthodologiques nouveaux, et des besoins en compétences et moyens importants.

Il vous est donc demandé d'anticiper autant que faire se peut et d'instruire dès à présent et de façon transversale les points suivants :

- Rôle de la modélisation en tant que démarche de recherche dans les domaines propres à l'INRA, et tout particulièrement ses secteurs prioritaires : est-elle effectivement une démarche incontournable qui conditionne le déblocage d'un certain nombre de verrous majeurs ?

- Positionnement de l'INRA en modélisation selon les différents secteurs disciplinaires : sommes nous producteurs, utilisateurs avertis, utilisateurs partiels, voire sous-utilisateurs de modèles ? Comment nous situons nous de ce point de vue sur les scènes nationale et internationale ?

- Identification et analyse des problèmes et blocages communs rencontrés dans les différents départements : difficultés de nature conceptuelle (mathématique et/ou informatique), liées aux moyens de calcul disponibles, ou à l'accessibilité des données... Ces difficultés impliquent-elles une recherche ou une veille méthodologique et/ou technologique spécifiques ? Les travaux récents ou en cours à l'extérieur de l'INRA dans les communautés de modélisation permettent-ils d'y répondre ?

- Analyse des compétences nécessaires, pour atteindre nos objectifs d'utilisation et de développement de la modélisation : comment les acquérir et les répartir ? Comment organiser la coopération entre un département de mathématiques et informatique appliquées, et des départements thématiques ?

- Moyens, outils et plate formes de modélisation : comment optimiser l'affectation des ressources, assurer les synergies et les interconnexions nécessaires, sans pour autant renouer avec un centralisme illusoire ?

Pour mener à bien la réflexion correspondant à ce périmètre thématique, notamment sur les deux rubriques développées ci-dessus, vous vous entourerez de groupes de travail largement ouverts aux chercheurs des autres départements concernés et procéderez à une consultation appropriée de ces départements.

S'agissant plus particulièrement du travail d'analyse et de prospective relatif à la modélisation, vous aurez recours à l'appui d'une personnalité externe qualifiée, de façon à consolider et approfondir l'examen transversal des questions ci-dessus, et à assurer que des propositions éventuellement novatrices par rapport à l'organisation actuelle auront pu être identifiées. Par ailleurs, vous veillerez à mener la réflexion en lien avec le conseil scientifique de l'organisme, notamment grâce à un dialogue régulier avec les membres que ce conseil désignera en son sein pour suivre plus particulièrement ce thème.

Il vous est recommandé de procéder en deux étapes, correspondant au calendrier de travail général de l'organisme : dans une première étape, les réflexions serviront à alimenter les schémas stratégiques de département, entre autres celui de l'actuel département BIA.

Sur ce point, conformément à ma lettre de cadrage général du 12 mars 2003, je vous demande de bien vouloir élaborer pour le 20 septembre 2003 une ébauche de schéma stratégique de votre futur département pour la période 2004-2007. Une approche de la liste des champs thématiques proposés vous est demandée pour le 12 septembre (date limite impérative) de façon à préparer la réunion du conseil scientifique des 23 et 24 septembre. Vous présenterez cette ébauche de schéma stratégique lors des Directoriales à tenir entre le 20 octobre et le 14 novembre 2003. La validation définitive des priorités fixées à chacun des départements, après avis du conseil scientifique, interviendra en février 2004 avec la nomination des chefs de départements et la rédaction de leur lettre de mission. Les schémas stratégiques seront, quant à eux, validés en mars-avril 2004.

J'attire enfin votre attention sur le fait que les centres élaborent actuellement des schémas de centre dont l'un des objectifs est d'aider à la construction de la stratégie de l'INRA au sein de la dynamique nationale des pôles de recherche et d'enseignement supérieur. Consciente des difficultés de calendrier que vous allez rencontrer, je vous invite néanmoins à intégrer dans votre réflexion les propositions contenues dans les premières ébauches de ces schémas que les présidents de centre devraient vous communiquer avant les directoriales.

La deuxième étape correspondra à l'élaboration du document d'orientation 2005-2008 pour l'organisme, qui s'appuiera sur les schémas stratégiques de département, mais également sur un certain nombre de réflexions transversales, dont celles sur la bioinformatique et la modélisation. Dans cette perspective, je vous saurais gré de remettre les rapports de synthèse sur ces sujets pour juin 2004.

Les directeurs scientifiques AAT (chef de file) APA et SED seront vos interlocuteurs au cours des prochains mois pour vous aider dans votre travail d'élaboration.

Je vous remercie grandement d'avoir accepté la charge d'animer la réflexion de votre communauté de recherche dont nous attendons, vous l'aurez compris, que, sans changement structurel notable, elle sache néanmoins remettre en cause ses objectifs et ses priorités de façon à permettre à l'INRA de remplir toujours mieux ses missions.

Marion Guillou

Copie : ensemble des responsables de périmètres thématiques

Annexe : rappel de la Mission du département BIA (schéma stratégique (1999-2003))

Le département a pour mission de mettre à la disposition de l'INRA des méthodes et des compétences à jour en mathématiques et informatique appliquées, en particulier dans le cadre de collaborations avec les autres départements sur des projets et programmes cohérents avec les axes stratégiques de l'Institut. Pour cela, il assure des activités de veille scientifique et technique et collabore avec des équipes françaises ou étrangères dans ses domaines de compétence. Il développe des méthodes originales dans ses propres disciplines, nécessaires à l'avancement des projets et programmes évoqués précédemment. Il participe à la conception et à la mise en œuvre des programmes de formation dans ses domaines de compétence.

Première partie
Analyse par thèmes

Thème 1

Du gène à la cellule

Remarque : L'analyse qui est donnée ici, utilise en partie celle effectuée dans le cadre de la mission « Bioinformatique » menée en parallèle. On trouvera des éléments plus précis dans le rapport « La recherche en bioinformatique à l'INRA ».

1. Rôle de la modélisation dans la démarche scientifique

Il s'agit ici des modèles qui décrivent le fonctionnement de processus à l'échelle de la cellule, l'expression de ces processus pouvant se faire à l'échelle du phénotype.

1.1 Rôle de la modélisation pour la connaissance

La démarche de modélisation va devenir de plus en plus incontournable pour faire face à l'afflux massif de données et à la complexité des processus que l'on cherche à connaître. C'est le modèle qui va guider l'expérimentation fine. Seul un modèle est capable d'appréhender les processus et les structures complexes que l'on ne peut identifier autrement. Bientôt il ne sera plus possible de publier un résultat d'expérimentation qui ne soit pas issu d'une démarche où un modèle est présent.

On peut structurer le rôle de la modélisation en quatre grands thèmes.

Analyser la séquence nucléique pour identifier les gènes, rechercher des motifs, détecter des signaux, etc.

C'est ce que l'on peut appeler l'annotation structurale. La modélisation est immédiatement apparue essentielle pour tirer une information biologique pertinente de la quantité considérable de données issues du séquençage. De très nombreuses formes de modèles ont été proposées pour analyser les séquences, et identifier des gènes ou des signaux. La modélisation par chaîne de Markov, qui exprime que l'élément suivant dans une chaîne de nucléotide dépend des nucléotides précédents selon une certaine loi de probabilité est très populaire et efficace dans cette analyse. Cette modélisation permet d'identifier les zones codantes et non codantes d'un génome, et les recherches se poursuivent pour améliorer et automatiser ces méthodes, et pour localiser d'autres éléments comme les sites de régulation ou les zones d'épissage.

Identifier la fonction et le mode d'expression des gènes et des protéines

On parle ici d'annotation (ou génomique) fonctionnelle. Il s'agit de caractériser la fonction biologique des régions fonctionnelles identifiées. Les modèles jouent un rôle moins important que dans l'annotation structurale. Il s'agit plutôt de combiner des méthodes diverses, basés sur des modèles empiriques et de faire intervenir l'expertise

humaine. Recherche d'homologie par similarité de séquence, extraction automatique d'information à partir de la littérature, méthodes fondées sur le contexte (colocalisation sur les chromosomes), analyse d'image à l'échelle cellulaire, méthodes d'analyse des relations structure fonction (structure 3d des protéines).

La production massive de données issues de l'analyse d'expression des gènes et des protéines, conduit à des besoins très importants d'exploitation immédiate à l'aide de modèles statistiques empiriques pour tirer de l'information de ces données très variables. Il s'agit de modèles classiques de classification, d'analyse des données, de tests de comparaisons multiples, qu'il faut adapter à des dimensions beaucoup plus importantes que celles pour lesquels ces méthodes ont été initialement développées.

Localiser les gènes, localiser les caractères, localiser les QTL

La cartographie génétique est profondément liée au modèle décrivant la recombinaison génétique lors de la méiose. De nombreux enrichissements du modèle initial combiné avec des développements statistiques sophistiqués, permettent de cartographier des marqueurs et des QTL (gènes participant à un caractère à effet quantitatif) dans les génomes. Pour aller plus loin dans la cartographie et identifier précisément ces gènes, les modèles de génétique des populations sont mobilisés pour utiliser un grand nombre de méioses historique, utilisant des méthodes de coalescence.

Modéliser et simuler le fonctionnement intégré des gènes et des protéines

L'identification des mécanismes complexes en œuvre dans la cellule permettant de réaliser une fonction biologique spécifique, est un thème de recherche fondamental pour la biologie, dans lequel la modélisation sous différentes formes est amené à jouer un rôle fondamental. Il s'agit d'abord de reconstruire (par exemple sous forme de réseaux bayésiens dynamiques) les réseaux géniques à partir de données de natures multiples : données d'expression, phénotypiques, d'interaction intermoléculaires, localisation de gènes, etc. Dans un deuxième temps, il s'agit de simuler un réseau semi quantitatif ou qualitatif. Il faut pour cela manipuler des modèles qui associent des régulations continues et discrètes, mobilisant des méthodes de l'automatique et de l'intelligence artificielle (modélisation qualitative).

1.2 Rôle de la modélisation pour l'action

L'utilisation de modèles pour différentes formes d'action ou décision est ancienne et bien rodé pour certains objectifs, en particulier ceux qui concernent les généticiens. Il s'agit de sélectionner les meilleurs génotypes en fonction des milieux, gérer la diversité, construire des résistances aux agresseurs, etc.

Les modèles de génétique quantitative et de génétique des populations ont une centaine d'années, continuent à être raffinés, et sont très efficaces vis à vis de leurs objectifs. Ils fonctionnent comme des boîtes noires, et il n'y a pas d'entrée pour les données de la génomique dans ces boîtes noires. Un enjeu fondamental est donc de construire des modèles qui intègrent la connaissance fine au niveau du génome pour des objectifs de décision. Des tentatives existent qui ne sont pas pour l'instant très satisfaisantes. Par exemple, l'utilisation des QTL en sélection n'est plus efficace que la sélection classique "boîte noire" que dans des situations particulières.

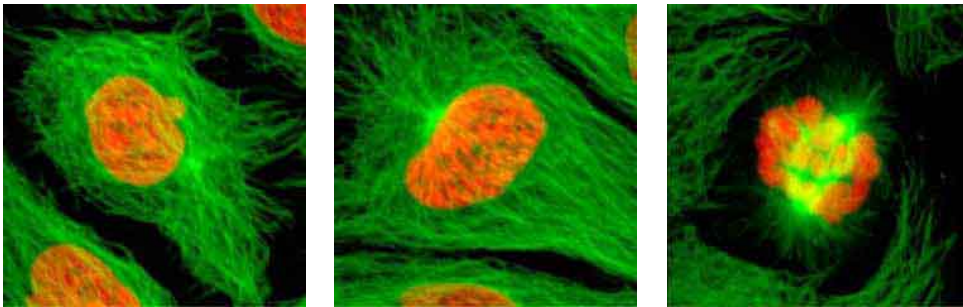
2. Positionnement de l'INRA

La construction de modèles pour la connaissance concerne de très nombreux départements de l'INRA : BV, PHASE, SA, GA, GAP, SPE, MICA, MIA etc. Les généticiens exploitent la diversité allélique pour participer à cette compréhension, tandis que d'autres départements travaillent plutôt avec un individu type. L'INRA est relativement bien impliqué dans les trois premiers thèmes décrits dans 1.1, mais il n'est pas encore véritablement présent dans le thème 4. La technologie a progressé plus vite que les concepts. Nous sommes à une période clé dans la révolution de la biologie et il n'est pas trop tard pour "prendre le train", mais il ne faut pas le rater ! Notons que l'INRIA a déjà initié des projets.

Le deuxième type d'objectif concerne un nombre plus réduit de départements. GAP et GA s'intéressent bien entendu à la sélection et à la gestion de la diversité. GAP et EA s'intéressent à la sélection dans un contexte d'interaction génotype X milieu X itinéraire technique. Historiquement, les généticiens quantitatifs de l'INRA ou d'ailleurs sont porteurs d'une culture statistique (Fischer est un père de la génétique des populations, de la génétique quantitative et de la statistique) qui est adaptée à la construction de modèles "boîte noire". Une ouverture vers d'autres méthodologies est peut être nécessaire pour répondre à l'enjeu d'intégrer les connaissances du génome dans les modèles de décision.

3. Les enjeux méthodologiques

- Concevoir des modèles. A priori de très nombreux types de modèles peuvent être pertinents et efficaces.
- Recherche d'une structure dans une grande quantité de données. C'est un enjeu fondamental avec la quantité formidable de données issues de la biologie moléculaire.
- La mise en œuvre des modèles pose des problèmes d'algorithme et de complexité calculatoire face au grand nombre de données, d'état, ou de paramètres.



*Etapes de la Mitose (microscopie confocale), Philippe Denoulet et Roger Prat, Biologie et Multimédia
<http://www.snv.jussieu.fr/bmedia/Mitose/index.htm>*

Thème 2

Physiologie, écophysiologie

Le thème « Physiologie et Eco-physiologie » couvre un vaste ensemble.

La physiologie est la « science qui étudie les phénomènes dont les êtres vivants sont le siège, les mécanismes qui régulent le fonctionnement de leurs organes, les échanges qui ont lieu dans leurs tissus » (Dic Univ Francophone Hachette). Cette définition peut s'appliquer bien sûr à de nombreuses recherches menées à l'INRA sur l'animal, le végétal, ou l'être humain. Par le terme d'*éco*-physiologie, on insiste sur le fonctionnement d'un organisme en interaction avec son environnement, ou encore sur la réponse d'un individu aux facteurs du milieu.

En pratique, ce que recouvrent ces deux termes va de recherches expérimentales pointues et centrées sur le fonctionnement d'un organe ou d'une grande fonction chez un organisme donné, jusqu'aux démarches de modélisation dites souvent "éco-physiologiques" de plantes entières. Pour les animaux cet aspect global porte généralement une autre dénomination, mais est bien développé sur les aspects alimentation et nutrition.

De nombreuses disciplines scientifiques sont mobilisées (neurosciences, endocrinologie, nutrition, biologie des grandes fonctions d'intérêt zootechnique ou agronomique [telles que développement, croissance, reproduction, lactation, production d'œufs], éthologie, génomique, protéomique, immunologie, biochimie, enzymologie, chimie analytique). La dimension intégrative entre différents processus et échelles d'étude (moléculaire, cellulaire, organe, organisme) est souvent présente.

Dans ce thème, il est donc possible de distinguer plusieurs types de modélisation, correspondant à des niveaux d'organisation différents et, corrélativement, à des échelles d'espace et de temps également différents :

- Les modèles métaboliques, animaux ou végétaux, qui décrivent des processus de nature biochimique et leurs régulations dans ou entre des organes ou tissus, sur des pas de temps courts (minute-jour).

- Les modèles physiologiques, qui décrivent une fonction de l'organisme ou le développement d'un organe sur un pas de temps plus long (semaine-mois) : par exemple chez les animaux, la lactation, l'ovulation, etc. ; chez les végétaux, la croissance d'un organe, le remplissage du grain.

- Les modèles éco-physiologiques, qui décrivent l'évolution d'un individu dans son milieu ou d'un peuplement sur une saison ou plusieurs années, qui ont généralement une vocation intégratrice.

Principaux départements concernés :

Pour l'animal :

- PHASE (Physiologie animale et systèmes d'élevage) [issu des ex- Physio A, ENA, partie de HFS] dont les thèmes de recherche couvrent « la plupart des grandes fonctions : la reproduction, l'ingestion, la digestion et les métabolismes, les fonctions productives et les fonctions d'adaptation au milieu » ;

- SA (Santé Animale) : en particulier, génétique fonctionnelle des bio-agresseurs, analyse des mécanismes immunologiques, physiopathologie musculaire, etc.

- SPE (Santé des Plantes et Environnement) : physiologie de l'insecte

Pour l'homme :

- ALIM.H. (Alimentation humaine) : perception sensorielle des aliments, fonction digestive, mécanismes de l'adaptation métabolique au régime alimentaire

Pour le végétal :

- BV (Biologie Végétale) : biologie du développement, métabolisme et nutrition, stress biotique et abiotique, adaptation à l'environnement

- EA (Environnement et Agronomie) : physiologie et éco-physiologie

- EFPA (Ecologie des Forêts et Milieux Aquatiques)

- GAP (Génétique et Amélioration des Plantes)

- SPE (Santé des Plantes et Environnement).

1. Rôle de la modélisation en tant que démarche de recherche

Quel est le rôle de la modélisation en tant que démarche de recherche dans cette thématique ?

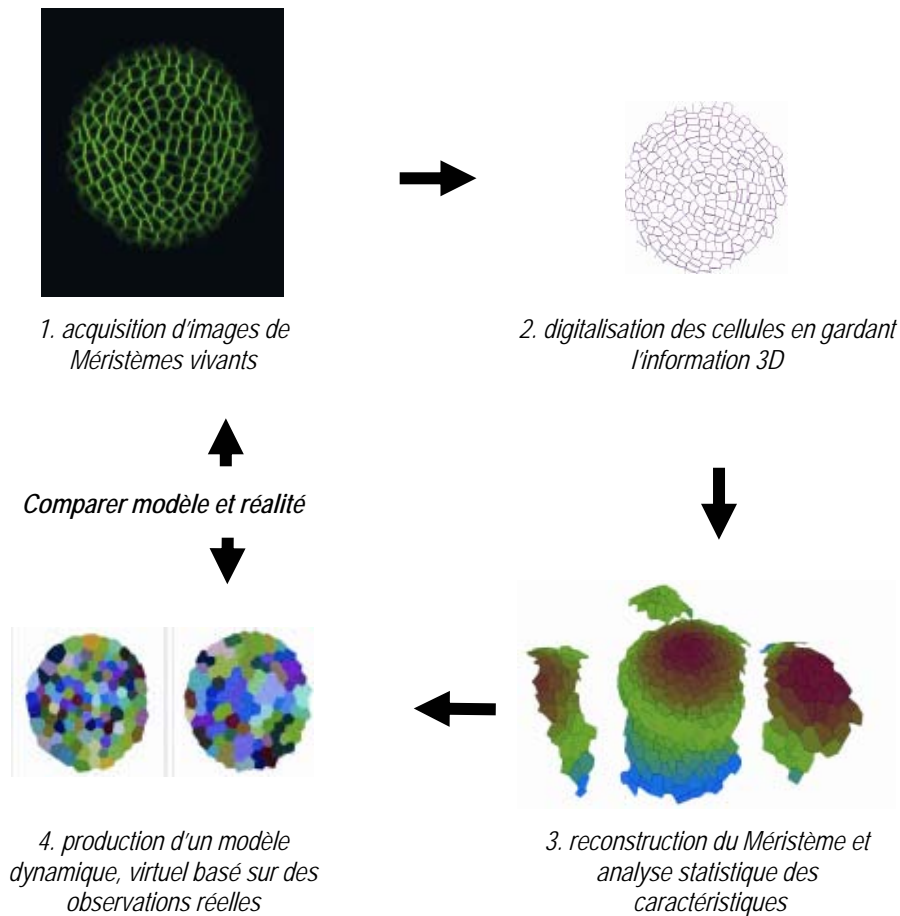
1.1 Plusieurs niveaux d'intervention de la modélisation

En physiologie, de nombreux travaux de recherche suivent une démarche expérimentale analytique « classique ». La modélisation y est soit absente, soit présente sous forme de modèles statistiques, généralement simples, dans lesquels les relations entre entrées et sorties sont basées sur des relations empiriques. La modélisation est alors un support à la mise en forme et à l'interprétation des données, tenant compte de leur variabilité et permettant de quantifier des relations qui peuvent éventuellement être intégrées dans des modèles plus globaux. Ces approches de modélisation empiriques se sont également développées comme outils d'intégration des résultats en une connaissance quantitative synthétique à travers les méta-analyses des données expérimentales.

La modélisation dans un sens plus mécaniste est d'abord un outil de formalisation incontournable pour la description et l'intégration de phénomènes biophysiques ou biochimiques impliquant des flux de matière ou de chaleur entre cellules, entre organes ou plus généralement dans ou entre des compartiments plus ou moins homogènes (ex :

transferts de chaleur dans un fruit, circulation de l'eau ou d'assimilats, photosynthèse ; métabolisme digestif,

musculaire, etc.).



Cycle d'observation et de modélisations que nous essayons de mettre en place. Le Méristème observé (vivant) est d'abord digitalisé et des reconstructions en 3 dimensions de plusieurs méristèmes sont ensuite utilisées pour faire un modèle purement virtuel. Ce modèle est comparé avec les Méristèmes vivants afin de tester sa validité.

Source J. Traas, C Godin

La modélisation se place enfin au cœur de la démarche de recherche avec une approche systémique :

- lorsque l'objet d'étude est le « système » en lui-même, impliquant des flux, des interactions entre constituants, et tout particulièrement des mécanismes de régulation (ex : fonction ovarienne, système nerveux ; croissance du méristème) ;
- dans les modèles écophysiologiques qui intègrent des processus intervenant à différents stades de développements d'un individu, à différentes échelles de temps et d'espace, et en interaction avec le milieu.

1.2 Rôle de la modélisation pour la connaissance

En (éco-)physiologie, la modélisation a d'abord comme objectif de représenter, formaliser et transmettre ou

partager des connaissances. Dans ce cadre, les modèles permettent en particulier de construire des représentations dynamiques de phénomènes, et de formuler et tester des hypothèses sur le comportement de systèmes qui sont trop complexes pour être appréhendés de façon intuitive, mais que l'on peut simuler grâce à la modélisation. Ce rôle de la modélisation correspond souvent à un changement d'échelle : la modélisation permet par exemple de passer de connaissances à l'échelle cellulaire, à des comportements (ou des propriétés) à l'échelle d'un organe ou d'une fonction globale, voire à l'échelle de l'organisme entier. Elle joue donc un rôle essentiel dans le processus de biologie intégrative.

Le lien entre la modélisation et l'expérimentation ressort comme un élément fort. En effet, la modélisation s'appuie généralement sur une acquisition intensive de données de natures variées (mesures quantitatives, images, etc.), au

cours d'expérimentations très instrumentalisées. Par ailleurs, la formalisation de résultats expérimentaux par la modélisation mathématique met souvent en évidence la nécessité d'acquies de nouvelles données, soit pour estimer des paramètres, soit pour préciser des mécanismes.

1.3 Rôle de la modélisation pour l'action

La plupart des modèles mécanistes fins sont développés, à l'heure actuelle, à des fins de connaissance fondamentale et ne sont pas adaptés à des objectifs de prédiction ou d'action. Il n'empêche que même pour ces modèles, un objectif plus applicatif est souvent présent en filigrane, comme une perspective à moyen ou long terme.

Dans le cas, par exemple, de la modélisation de la fonction ovarienne, la prédiction et le contrôle de la chronologie et du taux d'ovulation font partie des objectifs affichés. Pour les travaux sur le système nerveux, c'est le comportement que l'on envisage de prédire et de contrôler. Les travaux de modélisation de digestion et de métabolisme sont à la base de la conception des modèles de recommandations de systèmes d'alimentation et de nutrition des animaux. Dans le domaine végétal, des applications de modèles architecturaux sont envisagées, par exemple, pour définir des modes de conduite d'arbres fruitiers.

En parallèle à ces modèles plutôt orientés pour la connaissance, d'autres modèles sont développés plus spécifiquement pour l'action. Dans le domaine végétal, il s'agit de modèles de culture avec des objectifs de prédiction (e.g. zonage climatique) ou de comparaison d'itinéraires techniques. L'un des enjeux actuels est d'intégrer dans ces modèles de l'interaction génotype-environnement-itinéraire technique. Dans le domaine animal, des modèles sont développés pour prédire des lois de réponse multicritères (efficacité, qualité des produits, impact environnemental, bien-être...) des animaux aux systèmes d'élevage, aux pratiques alimentaires en particulier. De tels modèles peuvent être utilisés comme outils de communication avec des professionnels, ou dans des démarches plus quantitatives (ils peuvent constituer alors un élément des modèles du thème 5). Ils comportent une gamme assez étendue de niveaux de complexité, allant de modèles mécanistes relativement simplifiés mais restant riches en paramètres (ex : STICS), à des modèles basés sur des relations plus empiriques (on est là à la limite entre le thème 2 et les thèmes 4 ou 5), en passant par des modèles hybrides entre ces démarches.

2. Positionnement de l'INRA

Quel est le positionnement de l'INRA en modélisation dans cette thématique : producteur, utilisateur ?

2.1 Points remarquables

L'INRA est producteur et utilisateur de modèles, dans les différents niveaux identifiés précédemment. Des approches variées sont représentées, entre démarches analytiques, modélisation fine proche des mécanismes, ou modélisation plus empirique dans le cadre de modèles plus proches de l'action. L'une des caractéristiques INRA mentionnées

dans les domaines animaux et végétaux est le lien fort entre expérimentation et modélisation.

Dans le domaine végétal, plusieurs domaines d'excellence de l'INRA méritent d'être signalés :

- le développement collectif de modèles de culture génériques, dont l'exemple type est STICS, qui permet l'intégration de connaissances issues d'équipes différentes, et qui associe des compétences biologiques, modélisatrices et statistiques ;

- les approches de modélisation architecturale 2D et surtout 3D et la conception de modèles de plantes virtuelles, pour lesquelles l'INRA dispose d'une avance certaine et regroupe une communauté de chercheurs couvrant plusieurs centres (Montpellier, Clermont-Ferrand, Avignon, Grignon, etc.). A Versailles, ces approches sont appliquées au niveau cellulaire pour modéliser la croissance du méristème.

Les modèles éco-physiologiques interviennent également dans la modélisation de grands cycles biogéochimiques, celui du carbone en particulier.

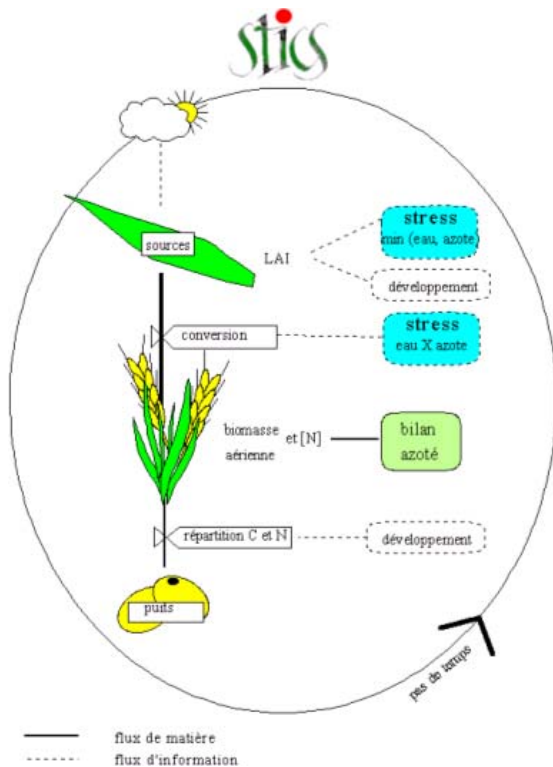
En biologie animale, il existe globalement moins de modèles qu'en agronomie, à l'INRA mais également à l'extérieur de l'INRA. Il existe cependant un passé important de modélisation dans le secteur de l'ingestion, de la digestion et de la nutrition animale. On assiste actuellement à un développement de nouveaux projets autour de la modélisation : flux de nutriments entre organes, métabolisme cellulaire, digestion et sa régulation chez les ruminants et chez les monogastriques, réception et traitements des signaux olfactifs, modélisation de l'activité ovarienne (projet en commun avec l'INRIA). L'intégration des mécanismes de régulations dans les modèles est un des enjeux importants des développements actuels, d'abord pour comprendre les mécanismes de contrôle endogène, puis à terme pour définir des stratégies de contrôle exogène.

2.2 Animation autour de la modélisation, politique de recrutements

Dans plusieurs départements concernés par la physiologie et l'éco-physiologie, la modélisation est identifiée comme un thème prioritaire et fait l'objet d'une animation active et d'actions spécifiques. Dans l'ancien département ENA, un effort spécifique et important d'animation (groupe "modélisation systémique"), mais également de recrutement et de formation (réalisation d'une Ecole-Chercheur) a eu lieu pour développer cette démarche au cours des dernières années. Il existe également un groupe de travail sur la modélisation dans le département EA, et une communauté active autour des modèles architecturaux.

Dans le domaine animal, l'acquisition de compétences en modélisation s'est fait en particulier par le recrutement d'ingénieurs, avec des profils soit plutôt biologistes, soit plutôt maths-infos. L'approche modélisation de systèmes reste néanmoins insuffisamment partagée par les scientifiques.

Dans le domaine végétal, il y a également des recrutements d'ingénieurs, mais aussi de scientifiques avec de fortes compétences en modélisation.



Stics : un modèle de culture, (Bisson, 1998, Agronomie)
<http://www.avignon.inra.fr/stics>

3. Enjeux méthodologiques

Quelles sont les démarches méthodologiques pratiquées dans cette thématique ? Identification et analyse des problèmes et blocages rencontrés et évolutions en cours ou prévisibles ?

3.1 Principales démarches méthodologiques

Globalement, les objets d'étude sont le plus souvent des systèmes complexes dont on cherche à décrire la structure, la dynamique et parfois la régulation en distinguant un sous-système opérant et un sous-système régulateur. C'est un domaine où la modélisation dynamique basée sur des modèles à compartiments, sur des équations différentielles ordinaires ou partielles, joue un rôle primordial. On s'intéresse autant aux régimes transitoires qu'aux systèmes en équilibre. Par ailleurs on tient de plus en plus compte de l'aspect spatial (2D ou 3D) et des interactions de voisinage liés à la structure spatiale.

Les modèles sont développés à différentes échelles temporelles :

- échelle métabolique, à pas de temps court ;
- échelle (éco)-physiologique, à pas de temps plus long (semaine, mois, année).

Dans le premier cas, on utilise plutôt des modèles en temps continu, et des modèles en temps discret dans le second (typiquement un pas de temps journalier).

Les méthodes d'automatique sont encore peu utilisées explicitement mais sont sans doute appelées à se développer, au moins pour les concepts, dans les modèles impliquant des mécanismes de régulation ou de contrôle. Les démarches de modélisation statistique sont utilisées soit en amont pour interpréter des résultats expérimentaux, soit en aval pour évaluer la qualité prédictive de modèles.

Enfin d'autres démarches sont mobilisées de façon plus ponctuelle. Par exemple des méthodes de modélisation symbolique et de simulation comportementale (réseaux de Petri) ont été utilisées dans le projet de modélisation de la fonction ovarienne mené en collaboration avec l'INRIA. Un autre exemple est l'application de démarches de méta-analyse sur des résultats bibliographiques.

3.2 Problèmes et blocages

En physiologie, la modélisation n'est pas une démarche suffisamment répandue parmi les chercheurs, par rapport à la quantité de travaux réalisés (et avec des nuances bien sûr selon les domaines). Il y a donc un aspect « culturel » à développer plus largement. Il y a également besoin de mieux reconnaître cette activité, souvent coûteuse en temps. Le problème n'est pas si net en éco-physiologie, où la modélisation fait partie intégrante de la démarche scientifique. Le besoin de formation, d'information ou d'échanges est néanmoins ressenti également par les modélisateurs avertis, pour pouvoir faire évoluer leurs démarches de modélisation. Des actions de mutualisation des moyens existent (ex : groupes de travail, plate-formes de modélisation) mais sont encore insuffisantes.

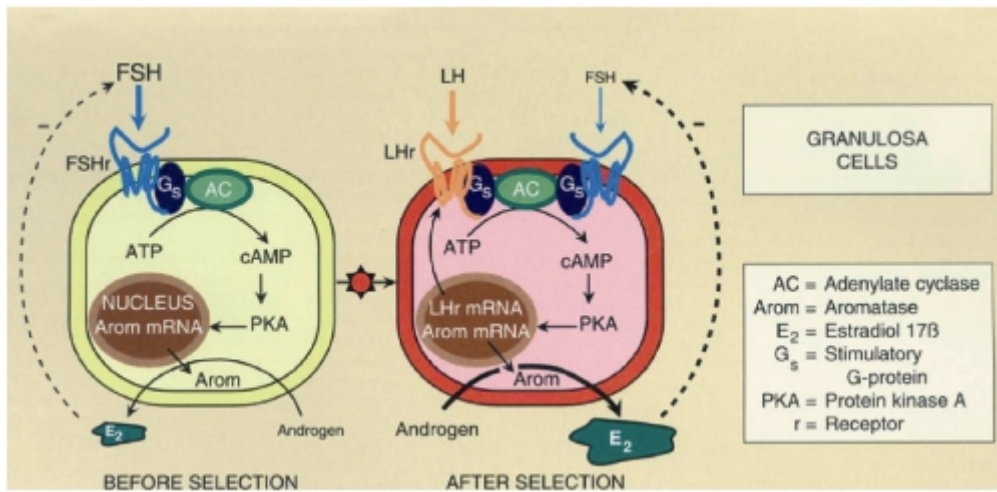
En ce qui concerne les outils de modélisation, divers logiciels et environnements de modélisation et de simulation sont utilisés (Matlab, Simulink, Vensim, etc.). Ils sont bien adaptés aux besoins dans de nombreux cas, au moins en première approche. Néanmoins, ils ne permettent pas de répondre à tous les besoins et le manque d'algorithmes ou d'environnement de simulation adaptés peut donc être un verrou dans les projets de modélisation les plus innovants. De façon générale, les modélisations les plus complexes soulèvent des problèmes méthodologiques et d'implémentation numérique que l'on rencontre dans d'autres contextes de modélisation, en calcul numérique, en automatique ou en statistique notamment.

Parmi les problèmes méthodologiques, ceux liés à l'articulation entre données et modélisation méritent une mention à part. En effet, les données sont souvent elles-mêmes de nature complexe et diversifiée. L'intégration de ces données pour le paramétrage ou pour la validation des modèles représente donc un problème général. La complexité des modèles et leur nombre élevé de paramètres rendent également nécessaire le fait de développer les méthodes d'analyse de sensibilité et d'incertitude.

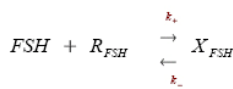
Globalement, de nombreux efforts restent à faire pour assurer une bonne cohérence aux différents niveaux d'une démarche de modélisation : adéquation de la représentation du système modélisé aux objectifs du projet (niveaux d'organisation, pas de temps, entités, relations), prise en compte des données, paramétrisation du modèle,

validation, intégration à des outils d'aide à la décision...
C'est sans doute autour d'une diffusion plus large de
"codes de bonnes pratiques" en modélisation et par des

démarches inter-disciplinaires que des progrès peuvent
être réalisés le plus efficacement.



Détection du signal

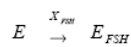


$$\frac{d}{dt} R_{FSH} = (k_- - k_+) X_{FSH} - (k_+ FSH + k_+) R_{FSH} - k_+ Xp_{FSH} + k_+ R_+$$

$$\frac{d}{dt} X_{FSH} = k_+ FSH R_{FSH} - (\rho + k_-) X_{FSH}$$

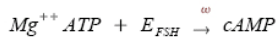
$$\frac{d}{dt} E_{FSH} = \beta [\sigma X_{FSH} - E_{FSH}] E_{FSH}$$

Relai et amplification du signal

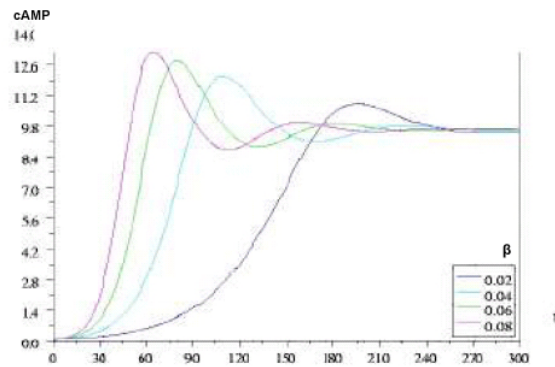


$$\frac{d}{dt} cAMP = \omega E_{FSH} - k_{pde} cAMP$$

$$\frac{d}{dt} Xp_{FSH} = \rho X_{FSH} - k_i Xp_{FSH}$$



Extinction du signal



Evolution du cAMP en fonction du temps pour différentes valeurs du paramètre

Modélisation du rôle de AMPc dans la médiation de FSH sur les cellules de granulosa.
Source F. Clément, schéma d'après Ginther et al., 1996, Biol. Reprod.

Thème 3

Physique et Biophysique

Ce thème comprend tous les aspects liés à la physique de l'environnement, notamment l'environnement des peuplements végétaux, aussi bien en ce qui concerne le sol que l'atmosphère.

Les **sous-thèmes** concernés sont :

- les sols dans leurs premières couches superficielles ; leur fonctionnement, les états physiques du milieu, les transferts de masse et d'énergie dans les sols et à leurs interfaces physiques et biologiques, l'hydrologie de surface, les cycles biogéochimiques, l'organisation spatiale et le fonctionnement des écosystèmes cultivés ;
- le climat dans la basse couche de l'atmosphère ; les mécanismes de transfert de masse et d'énergie, notamment les flux de matières et d'énergie dans l'atmosphère à leur interface avec les peuplements végétaux, les phénomènes de dispersion de particules dans l'atmosphère et par l'atmosphère.

Les **phénomènes étudiés** sont, entre autres :

- le déterminisme et la genèse des états physiques du milieu (états de surface, structure du sol), leur organisation spatiale et les rétroactions biologiques et anthropiques ;
- les transferts de masse et d'énergie entre les divers compartiments du sol ;
- les couplages transferts-réactivité (cycles de l'eau et biogéochimiques, devenir de polluants) ;
- la dispersion atmosphérique de particules (pollen, spores...) et de gaz (eau, carbone, polluants...) à partir de (ou vers des) parcelles prenant en compte, dans les conditions aux limites, toutes les hétérogénéités de surface souhaitables ;
- le comportement mécanique d'un arbre en fonction des sollicitations turbulentes, quelle que soit sa position dans le paysage.

Les **objets d'étude** sont des objets physiques, mesurables ou observables, et des variables quantitatives, déterministes ou des fonctions agronomiques et environnementales du sol, de la zone non saturée profonde et des nappes ainsi que de leurs interfaces physiques et biologiques ou au rôle de la structuration du paysage sur les flux et leurs dynamiques.

La caractéristique principale des phénomènes étudiés est leur **dynamique complexe** caractérisés par des **échelles spatiales et temporelles**, ainsi que des **temps de réponse variés**.

Les **données traitées** sont le plus souvent discrètes avec des échantillonnages spatialisés ou fréquentiels à des pas d'espace et de temps variables. Elles sont aussi des données issues d'images (télétection, méthodes géophysiques, images tomographiques de la structure). Enfin, le milieu étant évolutif, les données sont en conséquence dynamiques.

Les **approches de modélisation** mises en œuvre sont variées, mécanistes, stochastiques, marcheurs aléatoires, gaz sur réseau, réseau des neurones, empiriques, ..., et les maquettes numériques, comme modèles descriptifs du milieu, sont de plus en plus utilisées.

Principalement effectués en Environnement et Agronomie (EA), les travaux concernent également par leurs applications les départements Ecologie des Forêts, Prairies et milieux Aquatiques (EFPA), Santé de Plantes et Environnement (SPE). Les compétences mobilisées relèvent de différentes disciplines : physique, chimie, géochimie, microbiologie, mathématiques appliquées, hydrologie, pédologie, agronomie, biologie, géophysique, thermodynamique, mécanique des fluides, mécanique des solides, climatologie, météorologie.

1. Rôle de la modélisation en tant que démarche de recherche

La modélisation est depuis longtemps une démarche privilégiée et nécessaire aux recherches développées à l'INRA dans le domaine du fonctionnement du milieu physique (sol, atmosphère) et des phénomènes de transfert dont il est le siège. Elle répond au **besoin de disposer d'outils génériques** capables de traiter des interactions complexes et permettant de réaliser des généralisations.

La modélisation, dans ce domaine de recherche, joue un **rôle de structuration**. Elle est tout d'abord un outil essentiel d'intégration des connaissances. Autour des modèles, les différents secteurs disciplinaires expriment et synthétisent leurs savoirs. Elle fournit ainsi un moyen d'expérimentation virtuelle, mettant en évidence des interactions, permettant de hiérarchiser les phénomènes. Elle permet également la recherche de dispositifs expérimentaux optimaux en reproduisant le fonctionnement *a priori* d'un système que l'on veut observer car dans de nombreuses situations, on n'a pas *a priori* la connaissance suffisante d'un système naturel pour être à même de définir les protocoles d'observations (nombre et emplacement des observations, fréquence des observations). La planification de l'observation étant toujours liée à un modèle conceptuel préalable, c'est-à-dire à une hypothèse, développer le modèle mathématique de ce modèle conceptuel permet de préciser les caractéristiques de l'observation.

2. Modélisation pour la connaissance

Deux niveaux de connaissances relatives aux processus physiques et biophysiques sont recherchés par le biais de la modélisation. Une première approche a pour objectif d'améliorer les connaissances de base sur les processus en confrontant par la voie de la modélisation des

hypothèses de mécanismes à des expérimentations pour en interpréter les écarts afin d'affiner ou de modifier les hypothèses et éventuellement de mettre en évidence de nouveaux mécanismes. On peut ainsi estimer des variables non accessibles à la mesure. Par exemple, en ce qui concernent les flux dans le sol, on cale ou on vérifie le modèle sur les observations disponibles ; le modèle servant alors à l'estimation des flux non observables. Une seconde approche a pour objet de « changer d'échelle » en agrégeant des mécanismes connus et d'étudier les interactions entre ces mécanismes, et de rechercher le comportement émergent d'un système complexe soumis à une multitude de mécanismes élémentaires.

Dans ces deux approches, la formalisation passe dans la grande majorité par une mise en équations du système considéré sur la base des disciplines concernées (thermodynamique, mécanique des fluides ou du solide, hydrologie, géochimie...).

La première approche est toujours d'actualité, certains mécanismes de base n'étant pas encore complètement formalisés. On peut citer le transfert de polluants, mais aussi les voies d'écoulement dans les bassins versants ou bien l'existence de transferts préférentiels dans les sols.

Cependant, depuis quelques années, l'accent est mis de plus en plus sur l'utilisation des modèles en simulation ou prédiction (par le couplage des modèles d'évolution de l'environnement, des modèles descriptifs du milieu et des modèles phénoménologiques), plus que sur le développement lui-même des modèles à des fins de connaissances des mécanismes en tant que tels. Cette seconde approche est actuellement dominante car l'INRA a pour mission de s'intéresser à des problématiques complexes intéressant des objets complexes (sol-système de culture- bassins versants) et est donc plus en charge de développer des approches intégrées de modélisation que d'autres instituts de recherche non finalisée. Les "connaissances" alors générées sont autant liées à l'application des modèles qu'à leur construction.

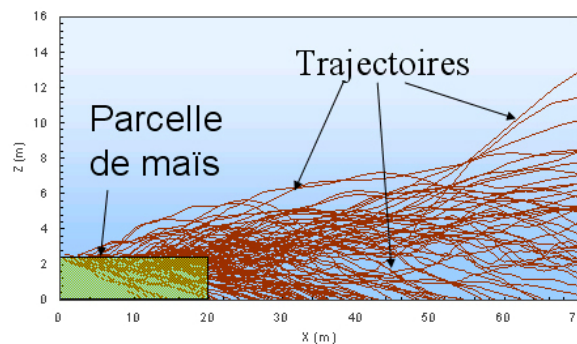
3. Modélisation pour l'action

On travaille en effet de plus en plus autour de questions de recherche ciblées, pour lesquelles on ne peut pas éviter le recours à la modélisation physique dans la mesure où elle apparaît comme une démarche incontournable.

Equation différentielle de la trajectoire

$$\begin{cases} dw = \left(-\frac{1}{2} \frac{C_0 \varepsilon}{\sigma_w^2} w + \frac{1}{2} \frac{\partial \sigma_w^2}{\partial z} \left(1 + \frac{w^2}{\sigma_w^2} \right) \right) dt + \sqrt{C_0 \varepsilon} d\xi_w(t) \\ dz = (w - w_s) dt \end{cases}$$

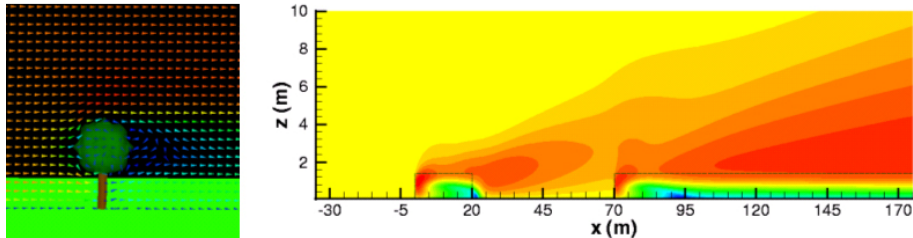
Vitesse du vent
Vitesse de sédimentation



Modèle de stochastique (lagrangien) de dispersion de pollen par simulation des trajectoires individuelles, source Y. Brunet

$$\rho_r \frac{d\bar{u}_i}{dt} = -\frac{\partial \bar{P}_i}{\partial x_i} + \frac{\partial}{\partial x_j} \left[(\mu + \mu_t) \left(\frac{\partial \bar{u}_i}{\partial x_j} + \frac{\partial \bar{u}_j}{\partial x_i} \right) \right] - \rho_i g \delta_{ij} - 2\rho_r \varepsilon_{ijk} \Omega_j \bar{u}_k$$

$$\frac{d\bar{c}}{dt} = \frac{\partial}{\partial x_j} \left[v_t \frac{\partial \bar{c}}{\partial x_j} + v_s \bar{c} \delta_{j3} \right] - \bar{S}$$



Modèle déterministe (eulérien) de dispersion par représentation continue (écoulement, concentration), source Y. Brunet

Exemples actuels :

- dans le cadre des travaux sur l'impact environnemental des OGM, on cherche à déterminer la dispersion d'un nuage de pollen et son dépôt à partir d'une parcelle source, pour une configuration de paysage quelconque, incluant tous ses éléments d'hétérogénéité (haies, discontinuités, taille et disposition des parcelles...) ; on bâtit donc des modèles de dispersion atmosphérique (mécanique des fluides) prenant en compte, dans les conditions aux limites, toutes les hétérogénéités de surface souhaitables ;
- dans le cadre de programmes de recherche liés aux tempêtes de 1999, on cherche à prédire le comportement mécanique d'un arbre en fonction des sollicitations turbulentes, quelle que soit sa position dans le paysage ; on développe pour cela simultanément des modèles type mécanique du solide (déformation d'une structure) et des modèles type mécanique des fluides (à l'échelle de la parcelle, de lisières, d'un paysage fragmenté).

Ces actions sont orientées : on cherche en particulier à quantifier le rôle de la structure du paysage et définir des modes de gestion moins sensibles à des effets indésirables (pollution génétique, casse d'arbres...). De façon générale, on va dans le sens des items mentionnés dans le document préliminaire d'introduction, c'est-à-dire prédire, diagnostiquer, concevoir, contrôler, planifier.

Pour l'action, des synthèses ou des simplifications des modèles de connaissance sont également recherchées ou des outils sont spécialement conçus pour la prédiction et la simulation de scénarios, ainsi que pour le diagnostic et pour l'aide à la décision. Ainsi, des modèles simplifiés de risque d'érosion sont établis pour fournir une cartographie régionale de ces risques. Une adaptation de l'approche de modélisation hydrologique distribuée à l'échelle du bassin versant est en cours de développement pour permettre la représentation de bassins versants à géométrie complexe ; le test et l'application du modèle sont envisagés pour évaluer le rôle des actions anthropiques et l'impact de différents scénarios d'aménagements du milieu agricole et de différents scénarios climatiques sur les flux d'eau.

4. Positionnement de l'INRA

Dans ce domaine de recherche, l'INRA est à la fois producteur et utilisateur de modèles et dans le Département EA, principalement concerné par ce thème, beaucoup de chercheurs et ingénieurs ont une formation initiale en physique avec des bases mathématiques suffisantes pour être plus que de simples utilisateurs de modèles.

La catégorie des modèles utilisés, notamment pour rendre compte des flux, repose le plus souvent sur du calcul numérique lourd (équations aux dérivées partielles à résoudre sur des domaines complexes). Après quelques essais "locaux" infructueux, les chercheurs INRA concernés se sont tournés vers des équipes spécialisées en calcul numérique (ex : Laboratoire Master à Bordeaux 1 ; laboratoire de Mécanique des Fluides de l'Ecole Centrale de Nantes ; laboratoires de recherche atmosphérique). Nous nous basons sur leurs codes (les chercheurs concernés sont utilisateurs, en général avertis) mais nous représentons nous-mêmes notre système d'étude (introduction de la végétation, des interactions fluide-structure, d'équations de conservation...) ; à ce niveau, l'INRA est donc également producteur, du moins dans le cadre de sous-modèles pour ce qui concerne la connaissance et la prise en compte de la végétation, généralement cultivée, dans les premières couches de l'atmosphère et les couches superficielles du sol.

La spécificité et le point fort de la contribution de l'INRA se situent à ce niveau ; ils reposent sur le fait que les chercheurs de l'INRA sont plus en position d'intégration et de modélisation de systèmes complexes, croisant biologie et physique, et sans doute aussi sciences sociales (comportements d'acteurs).

Par ailleurs nous continuons également de travailler, de manière plus classique, sur des modèles (type transferts végétation-atmosphère en conditions homogènes) pour lesquels nous avons les compétences et les moyens de les développer de A à Z.

Pour résumer, l'INRA est développeur, sauf pour ce qui est des moyens de calcul lourds ou dans des domaines pour lesquels nous n'avons pas les compétences nécessaires (exemple des modèles hydrologiques, ou des modèles atmosphériques à grande échelle). En tant que producteur, nous intervenons dans l'ensemble des étapes du processus de modélisation : la clarification des objectifs, la conception, la paramétrisation, l'implémentation numérique, logique ou analytique, la validation. Pour cela, l'expérimentation et la mise en place de dispositifs de mesure sont des aspects essentiels. Nous réalisons aussi la phase de valorisation du modèle (interface utilisateur, transfert).

5. Enjeux méthodologiques : obstacles, verrous et évolution

Un des défis actuels est de réussir le couplage des modèles descriptifs du milieu et des modèles phénoménologiques. Ceci implique de pouvoir résoudre dans ces géométries des systèmes d'équations relatives à plusieurs processus. Les difficultés des concepteurs de modèles résident principalement dans la modélisation de la complexité.

L'hypothèse d'un milieu homogène ou continu n'est pas applicable dans la plupart des situations d'intérêt. Les équations « classiques de transport » (pas seulement celles de la diffusion), valables pour les milieux homogènes, ne peuvent pas être appliquées en présence de discontinuités du milieu car les coefficients de transport ne sont pas définis (impossibilité de définir un volume élémentaire représentatif). Des solutions existent : approches multi-milieux continus, milieux équivalents dans certaines conditions, approches distribuées avec termes d'échange et plus récemment approches fractionnelles (les opérateurs incluent des dérivées d'ordre fractionnel). En fonction de l'approche utilisée, on aura à décrire l'hétérogénéité d'une façon fine ou pas (description explicite ou à travers des coefficients).

Les obstacles rencontrés sont souvent liés soit à des domaines de compétences particuliers (certains domaines de la physique, calcul numérique lourd), soit à des moyens de calcul disponibles (nécessité fréquente de gros calculateurs). Ces difficultés impliquent une recherche des collaborations adéquates, ce qui est fait régulièrement ; les réseaux existent et sont pour l'instant suffisants.

On se dirige de plus en plus vers une utilisation en réseaux de modèles, ou vers un développement de modèles en collaboration ou partenariat. On voit également naître des initiatives communautaires de développement de modèles « à grandes échelles » : exemple de l'initiative SEVE, regroupant une dizaine d'équipes françaises, dont quelques équipes INRA, et visant à développer un ensemble d'outils communautaires aptes à simuler le fonctionnement physique de paysages complexes ou la constitution du GIS Aquitain "Simulation Environnementale aux échelles sub-régionales", créé en Aquitaine autour de Bordeaux 1, UPPA, INRA, CEA, etc., et qui vise à rassembler les compétences existantes pour développer des projets pluri-disciplinaires, offrir une force d'expertise, faciliter le transfert, etc. Une des difficultés réside dans la possibilité d'informer les comportements locaux : description fine des hétérogénéités pour la description des flux et des transports aériens par exemple.

Dans ces cadres, le lien entre approches synthétiques et détaillées est difficile à mettre en place. La recherche d'une modélisation conjointe de processus de natures variées et se déroulant à des échelles différentes implique une diversification des variables et des échelles de contrôle. L'assimilation des différents types de données, la calibration et la validation, demeurent des problèmes cruciaux notamment dans toute approche de modélisation hydrologique distribuée. La gestion et la représentation de l'hétérogénéité des systèmes naturels nécessitent le développement de concepts de représentation des hétérogénéités. L'instrumentation et l'échantillonnage spatial demeurent dans certaines conditions des points de blocage pour procéder à la validation spatiale des modèles.

Thème 4

Ecologie des populations et des communautés, épidémiologie

La problématique de ce thème est centrée sur l'interaction, souvent tributaire de la qualité des données de terrain :

- Population : individus d'une même espèce en interaction dans un milieu ;

- Communauté : ensemble des organismes partageant un même habitat ;

- Epidémiologie : dynamique spatiales et temporelles de propagation des maladies. Par extension, étude des déterminants des états de santé des populations.

L'écologie (des populations et des communautés) et l'épidémiologie ont comme problématique commune d'être toutes deux essentiellement centrées sur les interactions entre les individus, et/ou entre les individus et leur milieu, et sur la manière dont ces interactions structurent le devenir des populations dans le temps et dans l'espace.

Les travaux de modélisation sont très anciens dans ce domaine, avec dès les années 1920 les travaux fondateurs de Lotka et Volterra en écologie et les travaux de Kermack et McKendrick en épidémiologie. Dans les deux cas, il s'agit de systèmes d'équations différentielles que l'on peut également écrire sous la forme de modèles à compartiments. Les premiers étudiaient comment les relations de prédation, et plus généralement de compétition, peuvent conditionner l'évolution des populations, et ceci plutôt dans le temps que dans l'espace. Les seconds comment l'évolution d'une maladie dépend des relations entre les trois grands états possibles d'un malade, l'état Susceptible, Infectieux, et Retiré (les célèbres modèles SIR).

Les départements les plus concernés sont Ecologie des systèmes Forestiers Prairiaux et Aquatiques (EFPA), Santé Animale (SA), et Santé des Plantes et Environnement (SPE).

Il ressort avec force des débats et discussions que cette thématique est très liée à la qualité des données acquises sur le terrain. Et que ces données sont en générale complexes, de par leurs structurations dans le temps et dans l'espace. Par conséquence, dans ce domaine, les chercheurs **utilisent tout autant les modèles statistiques** pour analyser les corpus de données, **que les modèles dynamiques et mécanistes** pour décortiquer les comportements. La modélisation sera donc vue ici au sens large, avec deux volets.

1. La statistique : une discipline plutôt bien intégrée par les chercheurs mais qui se complexifie

1.1 Rôle de la statistique en tant que démarche de recherche

Les modèles statistiques sont incontournables, ce n'est pas une nouveauté, et ils servent essentiellement à :

- caractériser les structures sous-jacentes dans les données (cycles temporels, structurations spatiales, structurations multi-factorielles) ; résumer les données ;
- identifier et hiérarchiser les facteurs explicatifs potentiels (les facteurs de risques en épidémiologie) ;
- donner des estimations de risques de maladies dans les populations, ou la probabilité de présence d'une espèce ;
- quantifier l'effet des maladies sur les fonctions zootechniques.

Face à ces questions, les chercheurs sont des utilisateurs plutôt avertis des méthodes statistiques et font un réel effort d'adaptation. Citons en particulier la généralisation de l'utilisation du modèle linéaire généralisé pour calculer des facteurs de risques ou des probabilité de présence d'une espèce.

1.2 Positionnement de l'INRA

Les départements cités sont plutôt utilisateurs. Les producteurs se trouvent dans le département Mathématiques et Informatique Appliquées (MIA), avec par exemple les développements en géostatistiques ou autour des méthodes de validation ou de choix de modèles statistiques. Il n'est cependant pas certain que le découpage actuel soit optimal, et il conviendrait de réfléchir à une meilleure association, pour la conception des modèles statistiques, entre concepts de biologie et outils mathématiques.

1.3 Points de blocage

Voici quelques points de blocage pour les outils de biostatistiques :

- beaucoup de variables explicatives et peu d'observations (données de recherche) ;
- ou au contraire beaucoup d'observations et peu de variables renseignées (données agricoles) ;
- les états mesurés sont incertains, les fréquences observées dans les populations souvent éloignées des lois de distributions usuelles ;

- les effets étudiés sont souvent non cumulatifs (interactions, seuils, hiérarchie des effets).

Et une complexité croissante, avec de plus en plus la prise en compte des contraintes spatiales et temporelles.

1.4 Evolution / propositions

Une évolution importante de ces types de modélisation est bien la prise en compte des structures spatiales, depuis les années 70 dans les laboratoires d'écologie et épidémiologie théorique, que ce soit hors ou dans l'INRA. Un premier pas est de mieux diffuser ces progrès vers les biologistes. Une deuxième évolution, plus proche de l'histoire de l'Inra, est de renforcer le dialogue entre concepts biologiques et outils mathématiques dans les travaux d'épidémiologie à l'Inra, notamment par la prise en compte des acquis récentes de la biologie évolutive (exemple : virulence / résistance en pathologie).

Par ailleurs, face à la complexification croissante des questions posées et des méthodes statistiques qui s'y rapportent, se pose la question de la limite de la formation des biologistes en mathématiques et du dialogue entre biologistes et statisticiens. Ce dialogue est jugé insuffisant et surtout difficile à entretenir.

2. Modèles mécanistes de comportement

Par modèles mécanistes, on considère ici des modèles qui intègrent un descriptif mécaniste, supposé posséder un certain degré de réalisme, des relations entre les entités modélisées.

2.1 Rôle de la modélisation en tant que démarche de recherche : aussi utiles pour la connaissance que pour l'action

Par ordre décroissant de citation dans les exposés ou les enquêtes, la modélisation est vue comme une démarche incontournable pour :

d'abord

- hiérarchiser des processus ;
- prédire, optimiser, aider à la décision, tester des scénarios : évaluer l'efficacité et/ou la rentabilité de stratégies (stratégies de lutte contre les maladies, de gestion d'une population, d'un itinéraire technique forestier...);

ensuite

- tester des hypothèses ;
- formaliser, quantifier, intégrer une certaine connaissance dans un domaine ;

un cran au dessous

- aide à la compréhension des phénomènes ;
- identification de trous ou de verrous dans la connaissance, de pistes de recherche ;
- aide à la définition d'expérimentations ;
- rôle pédagogique ;
- accès à des variables non mesurables.

On retrouve donc pratiquement à égalité la dualité cognitif / applicatif dans l'usage de la modélisation.

Pour autant, les modélisateurs (et corrélativement les travaux de modélisation) sont perçus comme insuffisamment nombreux. Peu de chiffres existent dans le domaine. Une enquête quasi-exhaustive réalisée en épidémiologie végétale (SPE, Ch. Lannou) a montré que 10% des scientifiques permanents pratiquent la modélisation, mais que 40% en ressentent le besoin. Par ailleurs, il serait extrêmement intéressant d'enquêter pour savoir quelle est la part de la modélisation dans les travaux actuellement réalisés par les doctorants. Une enquête plus ancienne menée à HYFS (V. Ginot) avait montré des chiffres similaires. Ce constat, allié à la difficulté de trouver des compétences épidémiologie/modélisation sur le "marché" des doctorants, a conduit les départements concernés à former eux même leurs doctorants, de concert avec le département MIA, selon une procédure d'agent scientifique contractuel (ASC).

On ressent donc à la fois le rôle important de la modélisation, tant dans sa dimension cognitive qu'appliquée, et la faiblesse des effectifs dans ce domaine.

2.2 Positionnement de l'INRA : plutôt producteur... de connaissances

Une réponse claire à cette question est difficile à obtenir. Les gens sont modestes et répondent difficilement à la question. Etre "producteur" de modèles suppose que ces derniers soient validés et dans une certaine mesure utilisés par d'autres. Ces autres pouvant être soit la profession ou d'autres chercheurs. Notons qu'il ne suffit pas qu'un modèle soit publié dans une bonne revue pour qu'il soit utilisé.

Dans l'ensemble, la quasi-totalité des chercheurs estime que leur modèles sont des outils d'appui à la recherche, plutôt que des modèles destinés à une utilisation par des tiers. Un exemple peut en être donné par les forestiers autour de la plate-forme logiciel Capsis : sur une vingtaine de projets, seuls un ou deux ont vraiment vocation à être "exportés" vers les utilisateurs / gestionnaires. En épidémiologie animale, secteur en développement récent, l'INRA (SA) souhaite effectivement produire des modèles. En hydrobiologie, plusieurs modèles sont développés, qui permettent de définir ou de tester des politiques de gestion, en particulier le prélèvement admissible de poissons. Mais là encore, même si à l'origine du projet cet objectif était parfois affiché, ils ne sont pas "livrés" au gestionnaire mais servent à affiner l'expertise du chercheur qui peut par ailleurs participer à diverses instances de gestion de ces milieux.

Une contrainte importante de tous ces modèles est qu'ils sont difficiles à caler. Et l'on retrouve ici la difficulté déjà évoquée de la collecte des données. Par exemple, les survies et les migrations, qui sont des composantes

essentielles de la dynamique sont fortement soumises aux conditions locales. Ces modèles, souvent développés sur des cas particuliers, sont donc difficiles à généraliser, et ont peu de lien avec des travaux de modélisation plus théoriques. En définitive, bien peu des projets de modélisation de ce domaine aurait donc vocation à sortir des laboratoires de recherche car les conditions de leur utilisation (domaine de validité) sont en général mal définies. Il y a donc là une certaine contradiction avec les objectifs affichés en 2.1 qui laissent plus de place à l'action.

Pour autant, l'INRA paraît être extrêmement bien positionnée pour proposer à la société des modèles conceptuels ou appliqués dans ce thème écologie/épidémiologie. Car bien peu d'organismes peuvent associer à la fois les compétences disciplinaires nécessaires à la modélisation et les moyens techniques d'observation/expérimentations sur le terrain, ces derniers représentant l'indispensable source d'hypothèses et le juge de paix de la validation.

2.3 Points de blocage méthodologiques ou pratiques

Le manque de compétences (formation, plates-formes) et les difficultés liées aux liens modèles – données (calage) éclipsent souvent la question de la complexité des processus. Les points les plus souvent évoqués sont :

Pour la construction du modèle

Se détachent :

- manque de formation bio-mathématique ;
- bases pour choisir une approche de modélisation (stochastique/déterministe ; continu/discret ; EDO, EDP, matriciel ; individu-centré...) ;
- manque de compétences en programmation, ou manque de moyens humains ;
- comment choisir un environnement de développement ? manque de plate forme génériques, de capitalisation des expériences ;
- intégration de données spatiales ET temporelles (à quelle échelle regarder ?).

Egalement cité :

- Méthodologie de complexification ou de simplification des modèles, passage d'un modèle théorique à un cas concret ;
- intégration de modèles d'échelles différentes, compatibilité des modèles ;
- difficulté à caractériser l'état du parasite ou de la plante (champignon/plante) ; difficulté de transposer les concepts d'épidémie humaine ou animale aux plantes ;
- présence de variables cachées.

Pour l'étude de son comportement et la validation

La question des données se détache très nettement :

- données difficiles à obtenir
- forte variabilité temporelle et/ou spatiale
- données spatialement et temporellement très corrélées

Vient ensuite :

- identification des paramètres, validation des modèles
- prise en compte de données incertaines
- prise en compte des événements rares.

Egalement cité :

- élaboration d'indicateurs synthétiques des sorties du modèle (validation et analyse de comportement)
- étude des comportements transitoires (typologie de trajectoires)
- méthodologie d'analyse des résultats
- problèmes de temps de calcul.

Le manque de formation, et dans une moindre mesure **la question des environnements de programmation** ressort donc comme un frein important et immédiat à la construction des modèles. La question de la complexité ne vient qu'ensuite comme si dans l'ensemble, le biologiste considère qu'il connaît bien son domaine et que sa transcription dans un modèle ne pose pas de problème méthodologique insurmontable. Mais ce constat est sans doute trop influencé par les travaux en cours de réalisation actuellement. Si on se focalise sur l'avenir, alors la **question de la complexité** (surtout observée via la variabilité) spatio-temporelle revient au premier plan, avec la question de la stratégie méthodologique pour naviguer dans cette complexité (choix des échelles, couplages de modèles, méthodologies de simplification). En ce qui concerne l'étude du comportement l'identification et la validation des modèles, le souci majeur concerne **les données**. Cette question, déjà évoquée, est visiblement majeure et éclipse presque toutes les autres. Dans le domaine de l'écologie des populations et de l'épidémiologie les données apparaissent souvent difficiles à obtenir, fortement bruitées et fortement autocorrélées. D'où le fort recourt aux statistiques déjà évoqué. On peut également penser que c'est un domaine où la métrologie (mesures automatiques, mesures en continue) est difficile, et a donc moins progressé que dans d'autres domaines. Curieusement, la question des SIG est peu évoquée, sauf par les forestiers. Vient ensuite la question de **l'identification des paramètres**, notamment dans le cadre de données incertaines (processus stochastiques) et d'événements rares.

2.4 Evolution prévisible

L'évolution possible oscille entre la tentation de l'hyperréalisme et la rigueur mathématique des modèles plus simples, vers une intégration des composantes spatiales et évolutives.

Vers de "gros" modèles. On observe, sinon dans les faits au moins dans les déclarations d'intention, une tendance forte à toujours plus complexifier et interconnecter les modèles. Beaucoup a déjà été écrit sur le mythe et les dangers de cette fuite vers l'hyper-réalisme. Mais le fait est qu'il se fait. Dans cette optique, le modèle n'est plus construit en fonction d'une question précise, mais tend à vouloir représenter et tester la cohérence d'un certain *Savoir* en cours dans un domaine donné. Un même modèle, en tant qu'analogie supposé d'une réalité

complexe, est alors censé pouvoir répondre à tout un tas de questions. C'est la notion en émergence de modèle dit "de connaissances". Une caractéristique de ces modèles est qu'ils ne peuvent plus être construits par un individu ou par une équipe, mais sont progressivement co-construits par plusieurs équipes. Sans trop forcer la note, on peut penser que dans quelques temps, et dans bien des domaines, il y aura des lobbies (scientifiques ou non) qui auront "leur modèle" face aux autres qui n'en auront pas. Et ces modèles seront probablement d'autant plus utilisés qu'il seront construits dans un cadre européen, avec des fonds de recherche. Et ceci, probablement relativement indépendamment de leur validation effective. **Il va donc falloir apprendre à vivre avec des modèles complexes. Et c'est en soit un enjeu pour la recherche.** Tout comme on a mis au point des méthodes d'analyse du réel il faut sans doute poursuivre la mise au point d'outils d'analyse de ces vastes mondes numériques, y compris sans doute la mise au point de méthodes de simplification. Car si un modèle "de connaissance" est par essence complexe, on peut penser qu'il est simplifiable dès qu'il est utilisé avec un objectif précis. Et l'on peut donc tout à fait imaginer que différents modèles simplifiés d'un même modèle de connaissance puissent co-exister, répondant à des questions différentes. Un modèle pour quoi, un modèle pour qui ? Avec ce type de modèle la réponse n'est plus triviale...

Naviguer dans la complexité. Il existe en écologie des communautés et en épidémiologie une tradition de modèles simples, que l'on peut comprendre mathématiquement. Ils servent également d'outils de réfutation, pour savoir en quoi l'assemblage de processus élémentaires (ex. : type mortalité / dispersion / colonisation) est suffisant ou non pour comprendre des structures et pattern à l'échelle de communautés (ex. : répartition spatiale). On peut citer les modèles de type SIR en épidémiologie, les modèles de réseaux trophiques, les modèles de dynamiques forestières à base de distributions. Pour chacun de ces modèles, il existe un ensemble de raffinements qui permettent de prendre en compte, en cas de besoin, les structures spatiales, les effets retard, une variabilité environnementale, ou une description plus fine des organismes concernés.

Sur ces créneaux, les équipes Inra sont plutôt positionnées (mais pas exclusivement) sur des modèles assez complexes, plus proches de la simulation, et fidèles par rapport aux données. Et nous l'avons vu, la tentation est forte de toujours chercher à les complexifier. Il existe donc un champ à portée, qui est au contraire la simplification, sans les dénaturer, des modèles réalistes plus complexes, à base de processus. Pour ce faire, il est utile de s'appuyer tant sur les connaissances des systèmes biologiques impliqués (exemple : les pathosystèmes en épidémiologie végétale), les concepts de biologie qui permettent la formalisation (par exemple, la notion de métapopulation), que sur des outils mathématiques parmi ceux que l'on sait maîtriser (par

exemple, la théorie des systèmes dynamiques). Ce compromis entre compréhension et réalisme peut être construit par une étroite collaboration, au sein d'équipes ou de projets, entre ces compétences. Par ailleurs certains modèles sont suffisamment simples pour être appropriés sans trop grande difficulté par des biologistes/agronomes, et cette possibilité d'une plus grande autonomie du biologiste par rapport à la modélisation doit être explorée.

Intégrer le spatial et les emboîtements d'échelles. Il est aujourd'hui acquis que l'écologie et l'épidémiologie ne peut faire l'impasse de la structuration spatiale, voire de la multiplicité des échelles spatiales et temporelles. En particulier, il existe une forte convergence méthodologique entre la notion de méta-population en écologie des communautés, et les processus de contact et les modèles SIR spatialisés en épidémiologie. Mais les progrès seront probablement assez lents car cette intégration pose des problèmes conceptuels (comment modéliser ces emboîtements) méthodologiques (comment choisir les échelles pertinentes) et pratiques (comment résoudre numériquement ces systèmes, et comment collecter suffisamment de données pour caler et valider de tels modèles). Et là encore, des hypothèses simplificatrices (champs moyens, fermeture des moments, approximations par paires...) peuvent souvent éviter d'avoir systématiquement recours à la modélisation explicite de l'espace qui tend à beaucoup alourdir les modèles. Ces champs mathématiques en émergences mériteraient donc d'être développés.

L'intrusion de la génétique, et donc des phénomènes évolutifs dans les données populationnelles. C'est une révolution dans la caractérisation de leur diversité, et c'est particulièrement vrai en écologie microbienne (mais aussi en épidémiologie, et de plus en plus en écologie) où la caractérisation des espèces est justement difficile. Mais c'est également une révolution dans la perception de la dynamique des systèmes écologiques : les traits d'histoire de vie d'une population ne sont plus figés, mais peuvent évoluer significativement en l'espace de quelques générations pour s'adapter à de nouvelles conditions environnementales. Même si ces considérations fonctionnelles sont connues depuis longtemps, et déjà intégrées dans des modèles plus théoriques comme ceux développés en biologie évolutive, elles sont encore peu prises en compte dans les travaux de modélisation actuels.

3. Conclusions

La modélisation apparaît comme un formidable outil de réfutation d'hypothèses biologiques, d'organisation de la connaissance, et de réflexion pour l'action.

L'audience de l'INRA apparaît encore faible dans ce domaine, alors que la santé (plantes et animaux) et l'environnement sont des enjeux de société majeurs et que l'INRA semble idéalement placée pour y répondre.

La mobilisation de compétences humaines (formation, recrutement) est un frein important. La modélisation est un art difficile qui demande du temps et des compétences. Il est illusoire d'espérer augmenter l'audience de l'INRA dans ce domaine sans y mettre les moyens.

L'écologie et l'épidémiologie sont des domaines où la qualité des données est souvent primordiale, surtout dans l'optique de la spatialisation des processus. L'INRA doit donc garder les moyens techniques d'acquérir ces données, et sans doute développer ses moyens d'archivage. Le traitement statistique de ces données est un préalable indispensable à tout travail de modélisation mécaniste. Or les techniques se complexifient, pour suivre la complexification des structures de données elles-mêmes. L'INRA doit donc garder un fort potentiel de développement/adaptation de ces techniques et peut-être se re-poser la question de la formation avancée des chercheurs en statistiques.

Les processus sont souvent stochastiques, les données incertaines, bruitées et fortement auto corrélées, et les variables souvent cachées. Des compétences doivent être développées dans le domaine des modèles mécanistes stochastiques, tant en ce qui concerne l'écriture des modèles que leur calage et leur validation.

En particulier, le calage de modèles probabilistes de grande dimension est encore une question méthodologique non résolue.

La modélisation des emboitements d'échelles est souvent perçue comme nécessaire. Mais tout ou presque reste à faire dans ce domaine...

En écologie/épidémiologie, comme ailleurs, on semble se diriger vers de gros modèles "de connaissances". Une méthodologie d'exploration systématique des propriétés de ces modèles est à développer. Elle est rendue d'autant plus difficile que les sorties des modèles sont également de plus en plus complexes.

Pour autant le recours à des modèles simples, i.e. mathématiquement compréhensibles, est à privilégier, y compris par simplification de modèles existants plus complexes. Ces modèles peuvent parfaitement inclure certaines composantes spatiales et génétiques/évolutives. Ils permettent en outre une plus grande autonomie du biologiste. Il est donc important que l'INRA développe des compétences mathématiques dans ce domaine.

Il paraît indispensable d'organiser une concertation étroite entre biologistes, mathématiciens, statisticiens et informaticiens dans le domaine de la modélisation.

Thème 5

Gestion des agro-écosystèmes

On s'intéresse ici aux systèmes de production agricoles considérés au sens large, couvrant différents niveaux d'organisation spatio-temporelle (itinéraire technique sur une parcelle et une saison, système de culture d'élevage ou exploitation agricole sur plusieurs années, bassin ou terroir de production sur une ou plusieurs décennies). Ces systèmes incluent une composante humaine assurant une fonction de gestion qui affecte un système naturel ou artificialisé mais soumis à des facteurs incontrôlables et plus ou moins aléatoires (climat, prix, réglementation). Ce thème fait l'objet de recherches intégrant notamment des connaissances agronomiques et zootechniques (voir Thème 2), physiques et biophysiques (voir Thème 3), écologiques (voir Thème 4), socio-économiques et de gestion technique de processus de production (voir thème 7). Ces recherches sur l'agro-écosystème visent à mieux comprendre et faire comprendre son fonctionnement et à contribuer à la mise au point et à la conduite de systèmes de culture, de systèmes d'élevage, de systèmes techniques de production, répondant à des cahiers des charges multicritères (environnementaux, techniques, économiques, juridiques, sociaux, sans oublier les exigences des agriculteurs quant à leur conditions de travail et à leur mode de vie). Dans le contexte actuel, plus contraignant et plus fluctuant que par le passé, la satisfaction de ces critères exige une capacité d'adaptation continue et convergente des modes d'organisation des activités agricoles, des technologies disponibles, mais aussi des modes d'interaction (négociation, coordination, coopération) des acteurs agricoles entre eux et avec d'autres.

La modélisation tant mathématique qu'informatique s'avère indispensable et constitutive de la démarche de recherche. Elle vient en complément aux expérimentations en condition de production qui restent lourdes à mener et limitées. Elle permet d'élargir la gamme de scénarios qu'il est possible de considérer et d'explorer à grande échelle des phénomènes complexes (ex : effets environnementaux de combinaison de pratiques dans un bassin versant).

Les départements les plus directement concernés (selon leur schéma stratégique) sont :

- Environnement et agronomie (EA) ;
- Mathématiques et informatique appliquées (MIA) ;
- Physiologie animale et systèmes d'élevage (PHASE) ;
- Santé Animale (SA) ;
- Sciences pour l'action et le développement (SAD) ;
- Sciences sociales, agriculture et alimentation, espace et environnement (SAE2).

Les rôles et attendus de la modélisation varient bien entendu en fonction des objectifs de chaque type de

questions abordé dans ces départements. Du fait de l'étendue du domaine couvert et des compétences nécessaires pour aborder le thème de la gestion des agro-écosystèmes, les départements cités sont souvent amenés à collaborer sur des projets transdisciplinaires et à partager des bases méthodologiques de modélisation (modèles de simulation de systèmes de production par exemple).

1. Rôle de la modélisation dans la démarche de recherche

Pour les productions végétales (EA, SAD, SA2E, MIA) on s'intéresse au champ cultivé, à l'exploitation ou, plus largement, à l'écosystème cultivé dont on cherche à comprendre le fonctionnement sous l'effet des actes techniques. Les modèles concernés dans ce thème portent sur le système sol-plante dans son environnement (climat, bioagresseurs, autres plantes) et les pratiques techniques de culture (itinéraires techniques, successions culturales). La modélisation doit intégrer des modèles de culture (ils relèvent du thème 2) avec des modèles de décision. Les modèles de culture peuvent être très variés : statiques ou dynamiques, empiriques (statistiques) ou mécanistes, déterministes ou stochastiques. D'une manière générale, ils visent à estimer certaines grandeurs agronomiques ou environnementales à partir du calcul dynamique de flux de matière entre différents compartiments plus ou moins nombreux du système biophysique (différents horizons du sous-sol, différents organes de la plante). Ils peuvent aussi s'évaluer à partir d'indicateurs observables sur le terrain par les agriculteurs eux-mêmes. Les modèles de décision sont constitués soit d'un ensemble de prescriptions fixées (dates/quantités d'intrants), soit d'un ensemble structuré de règles liant des indicateurs sur l'état du système cultivé (culture et milieu) à des actions. La notion de règle de décision permet de représenter des séquences conditionnelles d'actions ou des enchaînements répétitifs d'actions. Lorsque le processus de décision porte sur un ensemble de cultures, sur la gestion des ressources (le matériel outils, la main d'œuvre par exemple), et sur un horizon pluri-annuel, il faut recourir à des structures plus élaborées que les règles de décision.

Les modèles de culture et de décision sont typiquement utilisés dans des approches de simulation du processus de production. La simulation reproduit l'interaction dynamique entre les processus biologiques et physiques et le processus de détermination de l'action en fonction de l'état courant du système. Un interpréteur de règles est nécessaire pour déterminer à différents instants, la ou les actions à réaliser. La simulation nécessite aussi en

entrée des données climatiques journalières qui sont soit issues de séries temporelles relevées sur certains sites, soit créées en utilisant des générateurs climatiques. La modélisation permet de :

- tester la réponse environnementale du système (une parcelle ou un ensemble de parcelles dans une petite région) à des scénarios d'événements climatiques et de pratiques culturales ;
- faire de la prédiction de performances agronomiques (rendement, calibre, teneur en protéine, taux de sucre, etc.) ;

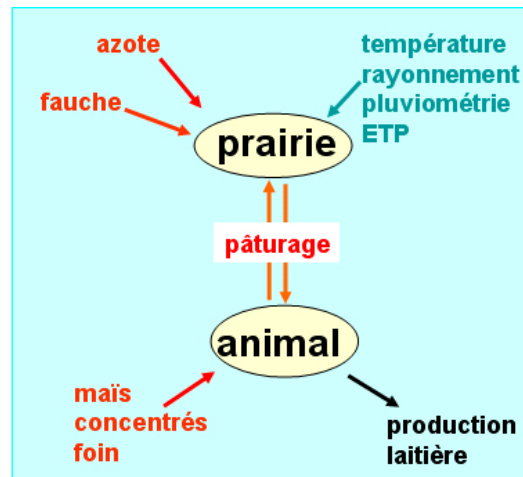
- rechercher les causes ou l'explication d'un phénomène (diagnostic agronomique d'un faible rendement) ;
- aider à la décision en conduite de production (e.g. comment organiser les actes techniques et l'utilisation des moyens, comment raisonner l'assolement) ;
- aider à la conception-sélection de variétés.

Ces capacités peuvent être mises à profit pour concevoir de nouveaux systèmes de production à l'échelle de la parcelle, de la sole ou de l'exploitation et à aider au pilotage de ces systèmes.

Eléments biophysiques :
prairies, vaches, stocks, climat

Eléments décisionnels :
Plans conditionnels concernant fertilisation, fauche, pâturage, complémentation alimentaire

- Echelles/niveaux :
- de la parcelle à la sole prairiale
 - de la journée à la saison
 - de la vache au troupeau



Modélisation de la conduite du pâturage – système de simulation SEPATOU, Cros et al., 2001

Une problématique semblable est abordée en production animale (PHASE, SA, SAD, MIA). Les modèles dynamiques jouent alors un rôle essentiel pour évaluer des techniques innovantes de maîtrise de la production ou de configuration du système de production. Pour l'aide à la décision en conduite (gestion du pâturage ou de l'alimentation par exemple) des modèles de simulation ou des modèles plus simples (se prêtant à une forme d'optimisation) sont utilisés. Dans les modèles de simulation, la composante biologique et physique concerne les ateliers animaux (troupeau bovin allaitant ou laitier, porcs, lapins, poissons, etc.) et en particulier les processus physiologiques et infectieux. L'étude de l'impact des règles de conduite des animaux (reproduction, réforme, vente) dans la gestion dynamique des effectifs des différents lots ou des différentes classes d'animaux dans un troupeau requiert ce type de simulateur. Parfois on s'intéresse aussi à l'interaction entre la production animale et la production végétale qui contribue à l'alimentation des animaux (pâturage, cultures fourragères). L'intégration des modèles zootechniques et de production végétale qui sont en forte interaction est une difficulté en soi. Les modèles de décision peuvent être complexes : le nombre de décisions et d'actions peut être grand, les actions opèrent sur des objets composés et évolutifs (lots d'animaux, portions d'espace), les décisions doivent tenir compte de

contraintes sur les ressources, les facteurs à prendre en compte dans la décisions sont nombreux et divers, différents niveaux d'organisation sont à considérer (animal, troupeau, exploitation) et à différents pas de temps (alimentation, santé animale, gestion des effluents, reproduction). Des modèles multi-agents d'animaux sont aussi considérés en relation avec des partenaires extérieurs à l'INRA pour modéliser le comportement animal (préférences alimentaires en prairies hétérogènes, répartition spatiale) et l'articuler avec les modes de conduite des troupeaux et leur impact sur l'environnement.

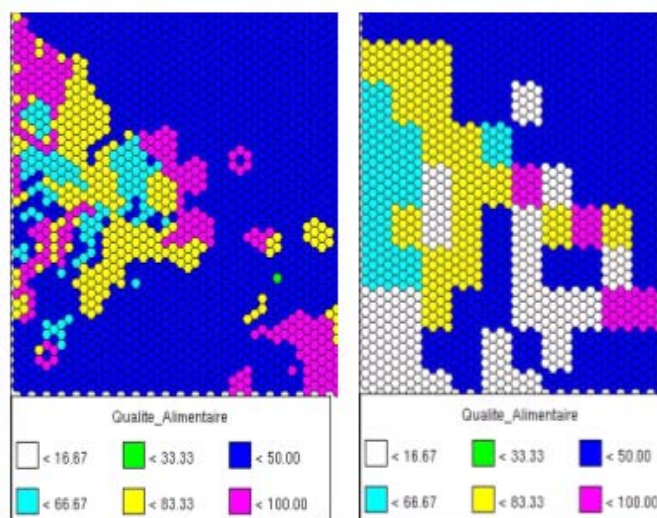
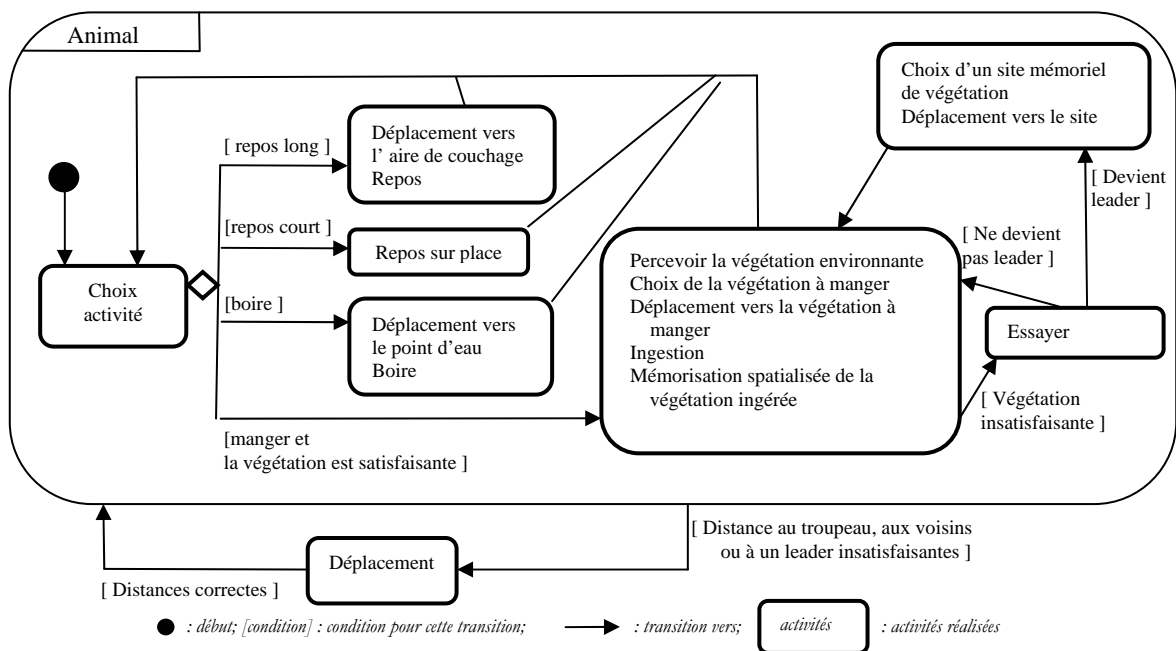
Pour l'analyse des systèmes techniques de production agricole (SAD), des approches plus intégrées sont développées. Elles rendent compte en particulier des trajectoires d'évolution des exploitations agricoles et des modes d'organisation des territoires d'exploitations, par des approches qualitatives (typologies ou catégorisation, recherche de formes archétypiques, raisonnement à partir de cas, inférences spatiales...).

Les approches de gestion des agro-écosystèmes diffèrent significativement suivant l'objectif poursuivi et, en particulier, les cibles visées par ces approches (décideurs politiques, acteurs du développement, agriculteurs, chercheurs). Ainsi, outre les modèles évoqués de simulation des systèmes de production, sont

développés des cadres analytiques, relevant de l'ingénierie, pour analyser l'organisation du territoire et la gestion de l'environnement (SAE2, SAD). Ils exploitent en particulier des modèles quantitatifs empiriques (statistiques spatiales et de processus temporels) et des modèles d'optimisation (programmation linéaire et dynamique) pour faire, en particulier, des calculs économiques et de l'analyse de risques.

Des modèles qualitatifs (modèles d'agents, représentation sociales, modélisation de structures spatiales, modèle de réseaux) et des modèles de choix multicritères relevant de l'ingénierie socio-politique sont aussi élaborés.

Ils nécessitent d'analyser les logiques d'action des différents acteurs de l'agriculture, des filières, de l'environnement et des territoires, afin de faciliter leur coordination et d'élaborer des scénarios réalistes ou fictifs d'évolution des activités pour étudier l'adéquation entre les dispositifs d'actions et les projets des différents acteurs. Lorsque des aspects d'analyse et de conception de nouveautés techniques et organisationnelles visant à améliorer l'efficacité des agro-écosystèmes, les modèles doivent s'intéresser à la gestion du travail (organisation des activités et des tâches, gestion des concurrences entre elles) et y intégrer le comportement décisionnel face à l'incertitude (organisation et utilisation de l'information pertinente).



Modélisation du système parcelles/troupeau – Système multiagents PARIS, Baumont et al., 2002
 En haut : description des activités d'un animal / En bas : représentation de la parcelle par l'animal (perçue à gauche, mémorisée à droite)

Les approches (MIA) issues de la modélisation d'entreprise et les travaux d'intelligence artificielle sur la prise de décision en univers évolutif (agents cognitifs) ont alors tout leur intérêt mais sont encore peu utilisées. Enfin les modèles peuvent être utilisés comme support pédagogique de communication et de transfert de connaissances entre des chercheurs, des acteurs du développement et des agriculteurs. Cette production de connaissances dans une démarche participative induit des besoins particuliers de modélisation qui commencent à être abordés à l'INRA, en collaboration avec d'autres partenaires (INRIA, CNRS).

Les travaux de modélisation évoqués s'inscrivent dans des projets motivés par des applications particulières. Toutefois des développements méthodologiques génériques sont engagés pour aller vers une boîte à outils de modélisation pour certaines classes de problèmes en gestion des agro-écosystèmes. Il s'agit, par exemple, de formaliser et étendre (MIA) la notion de stratégie de conduite introduite par des agronomes dans les années 80 pour développer les bases d'un (méta-) modèle générique de simulation des systèmes de production agricoles. Les divers modèles informatiques de simulation développés dans les différents départements pourraient être couplés avec des procédures d'optimisation (apprentissage par renforcement, optimisation stochastique) maintenant opérationnelles : elles permettent de calculer des stratégies optimales selon certains critères et d'explorer efficacement l'espace de décision et d'action. Des travaux plus récents visent à développer des approches mathématiques et informatiques de modélisation spatialisée et multi-acteur des pratiques agricoles en étendant les approches fondées sur les processus décisionnel de Markov de façon à pouvoir étudier les effets environnementaux de ces pratiques.

2. Positionnement de l'INRA en modélisation dans la gestion des agro-écosystèmes

Pour la conduite de culture au niveau de la parcelle ou de la sole de nombreux modèles sont développés dans le monde (Etats Unis, Australie, Pays-Bas en particulier). L'Australie semble avoir une longueur d'avance sur l'utilisation pratique de ces modèles en relation avec la profession. Pour les modèles ciblés sur la problématique de gestion des systèmes d'élevage, il existe d'autres tentatives aussi ambitieuses et intégratrices que celles engagées à l'INRA ; signalons en particulier les travaux néo-zélandais et australiens sur les systèmes bovins laitiers. Mentionnons aussi les travaux originaux des danois sur la production animale hors sol (porcs) en utilisant des modèles et des techniques d'optimisation (programmation dynamique). Les modèles traitant de l'organisation du travail sont rares.

Jusqu'à présent les travaux de modélisation des agro-systèmes à grande échelle (région ou ensemble de pays) pour évaluer des impacts économiques et

environnementaux de politiques agricoles sont plus avancés aux Pays-Bas et aux Etats-Unis qu'en France.

Sur le thème "gestion des agro-écosystèmes", l'INRA est clairement dans une position de concepteur de modèle biophysique, ce qui n'empêche pas occasionnellement de repartir de modèles étrangers pour les adapter et/ou les étendre à une problématique plus large. Pour la modélisation des systèmes de décision, l'INRA a été un précurseur (grâce en particulier aux travaux de l'école Sébillotte pour les productions végétales) et reste en pointe en abordant des systèmes de plus haut niveau (l'exploitation agricole). Pour la plupart des départements concernés, l'INRA est producteur de modèles adaptés aux questions posées et joue un rôle de leader méthodologique vis-à-vis des partenaires extérieurs (ceux représentant la demande sociale). La problématique de la gestion des agro-écosystèmes du point de vue des aspects phytosanitaires est reconnue importante à l'INRA comme ailleurs mais le pont entre la prise de décision en fonction de l'état biophysique et les conséquences sur le système biophysique est encore loin de pouvoir se faire du fait de la complexité du modèle biophysique nécessaire.

A noter la quasi absence à l'INRA de travaux sur les problèmes de gestion en sylviculture alors que les scandinaves (suédois et finlandais en particulier) et les canadiens ont déjà beaucoup investi sur le sujet.

L'IRD, le CIRAD et le CEMAGREF travaillent aussi sur ce thème, souvent en collaboration avec l'INRA et sur les mêmes bases méthodologiques. Ces organismes ont toutefois une plus grande expérience que l'INRA de l'utilisation des techniques de modélisation multi-agent.

3. Verrous, difficultés, besoins

Les discussions du groupe de réflexion sur la gestion des agro-écosystèmes ont mis en évidence un certain nombre d'obstacles que nous avons classés en trois catégories : les verrous qui demandent un investissement de recherche méthodologique sur la modélisation de certains aspects des problèmes abordés, les difficultés qui relèvent soit d'un travail assez conséquent d'ingénierie (essentiellement mais pas forcément exclusivement) soit d'empêchements technico-économiques, et enfin les besoins qui concernent le développement informatique et l'amélioration de l'exploitation des modèles construits.

Verrous

- Détermination du niveau approprié d'abstraction dans la modélisation de systèmes composites
- Modélisation des problèmes de décision sur un horizon long induisant un phénomène de dilution sur le temps de la qualité des informations, le bruit sur l'information ayant à terme plus d'impact que les décisions modélisées
- Approche des problèmes de décision liés à une structuration spatiale (e.g. agriculture de précision)
- Problèmes de décision dont les effets des actions déclenchent une propagation spatiale plus ou moins aléatoire

- Modélisation de l'anticipation dans un processus de décision (attitude réactive à base de plan ou proactive en raisonnant sur les buts et sur les prévisions, attitude optimiste ou prudente)
- Optimisation de politiques dans les problèmes de décision comportant plus de quelques (3) acteurs
- Choix multicritère dans les processus de décision, en particulier lorsque les conséquences sont obtenues à des termes différents, concernent des sujets différents, ne sont pas du tout commensurables et sont incertaines (souvent pas probabilisables)
- Caractérisation de la généralité des approches proposées.

Difficultés

- Simulation comme expérimentation virtuelle: donc besoin de générer les entrées de la simulation par une sorte de plan d'expériences
- Méthodologie de validation (acceptabilité du modèle eu égard aux objectifs poursuivis)
- Analyse des sorties selon des critères multiples, variés, non agrégeables

- Méthodologie de la modélisation par un processus de co-construction avec de non-modélisateurs
- Guides d'exploitation des modèles selon l'utilisation visée (aide à la décision, construction-analyse de scénarios, support d'apprentissage collectif ou de négociation...)
- Evaluation de la pertinence d'un modèle du point de vue de l'applicabilité pour l'aide à la décision
- Accessibilité des données nécessaires au modèle contrainte par des problèmes de coût ou de moyen physique de mesure ou de moyen matériel de stockage ou de transport de l'information (absence de norme) ou d'hétérogénéité de la fiabilité et de la précision des sources
- Restriction sur l'utilisation des données pour raisons de confidentialité.

Besoins

- Ressources et compétences professionnelles de développement informatique
- Archivage et mise à disposition des modèles et des résultats d'utilisation obtenus (retour d'expérience).

Thème 6

Agroalimentaire et génie des procédés

Plusieurs problématiques sont à considérer pour ce thème à l'INRA, du point de vue de la modélisation et de son utilisation.

Parmi d'autres nomenclatures possibles, on distinguera ici :

- mécanismes de transformation
- analyse de structures
- ingénierie reverse
- analyse de risques.

Les trois premières activités se rencontrent essentiellement au sein d'équipes travaillant dans le domaine du génie industriel alimentaire et plus ponctuellement dans d'autres équipes. Elles se distinguent par leurs finalités, les outils méthodologiques qu'elles mobilisent et leurs degrés de maîtrise à l'INRA. La quatrième activité possède une spécificité particulière bien reconnue.

1. Mécanismes de transformation

Ces mécanismes concernent des procédés de Génie Industriel Alimentaire (GIA) à base d'opérations biophysiques (opérations unitaires : séchage, cuisson, ultra-filtration, etc), et les procédés biotechnologiques rencontrés en GIA (fermentations par exemple) mais aussi en sciences de l'environnement (dépollution biologique). Ces deux types de procédés ont en commun plusieurs caractéristiques fondamentales liées à leur nature (systèmes dynamiques), aux objectifs auxquels ils correspondent, aux problèmes qu'ils soulèvent (modélisation et incertitudes, observations directes ou indirectes, supervision) et donc aux approches méthodologiques (mécanistiques, statistiques, automatiques, qualitatives) dont ils relèvent. Ils présentent aussi des particularités spécifiques qui font que les communautés scientifiques correspondantes ne sont pas toujours confondues, indépendamment des différences de contextes d'application.

Les activités de modélisation correspondantes se concentrent principalement au sein des départements CEPIA (GIA), EA (dépollution) et MIA (méthodologies de l'analyse et du contrôle de systèmes). Mais les équipes concernées sont peu nombreuses. Au niveau national, ce déficit n'est que très partiellement compensé par les activités hors institut, des domaines public ou industriel.

1.1 Rôle de la modélisation

C'est l'approche fondamentale sur laquelle repose le génie des procédés en tant que science de l'ingénieur. Les objectifs visés sont multiples :

- représentation des connaissances

- inférence
- simulation
- prévision des états et des comportements dynamiques ;

mais aussi :

- reconstruction de données non mesurables
- optimisation de fonctionnement (rendement, temps minimum)
- détermination de trajectoires de conduite (consigne) appropriée
- suivi de ces trajectoires (commande)
- surveillance et diagnostic des dysfonctionnements
- gestion de conduites coordonnées
- évaluation de risques.

1.2 Positionnement de l'INRA

Ce thème présente la particularité de correspondre à une activité de mathématiques appliquées pour laquelle l'INRA possède à la fois des équipes de concepteurs (en MIA) et de concepteurs utilisateurs en prise en directe sur la conduite de procédés réels (en CEPIA et en EA). Cette configuration permet à ce thème de bénéficier à l'INRA d'une activité de recherche significative par rapport à sa situation hors institut. D'autres équipes du département CEPIA, tournées plus prioritairement vers certains types de procédés de GIA, possèdent également des compétences opératives reconnues en modélisation et automatique des procédés.

Une partie de cette communauté s'est longtemps retrouvée au sein du Groupe Automatique de l'INRA, qui l'associait à des utilisateurs de l'automatique de l'Institut hors du domaine des procédés (pilotage de serres, notamment), confrontés au même type de problèmes. Le réseau ainsi formé est resté actif au moins en mode bilatéral. Son ouverture vers l'extérieur (écoles d'ingénieurs, CNRS, INRIA) a permis et permet des échanges de niveau méthodologique et de valorisation, qui contribuent à la représentativité de l'INRA en ce domaine.

1.3 Démarches méthodologiques

La nature du procédé (opération unitaire, transformation complexe), les incertitudes associées à sa dynamique, la nature des données et informations disponibles, conditionnent la nature de sa formalisation (analytique ou symbolique et qualitative).

1.4 Approches analytiques

Elles correspondent à des représentations qui reposent sur des considérations liées à la nature des procédés et des réactions qui les sous-tendent :

- bilans-matières (bioprocédés)
- transferts de matière et chaleur (Opérations Unitaires)

- quantité de mouvement (O.U.)
- génie des réactions (transformations physico-chimiques).

Selon la nature des informations disponibles, mais aussi des cultures en présence, ces représentations reposent sur des approches déterministes (largement majoritaires) et des approches stochastiques, différentes dans leurs exigences opératoires et partiellement, selon leurs contextes d'application.

Les premières, le plus souvent à temps continu, utilisent des représentations par EDO (sous hypothèse d'homogénéité spatiale) ou par EDP (systèmes distribués) complétées par des conditions initiales et limites (EDP). Les secondes, le plus souvent à temps discret, utilisent des représentations markoviennes sous formes de processus autorégressifs contrôlés.

Selon les objectifs de la modélisation (représentation des connaissances, inférence, simulation, prédiction, commande) les exigences dans la représentativité analytique des phénomènes ne sont pas les mêmes. C'est ainsi que la puissance des concepts de l'automatique permet de concevoir des lois de commandes robustes aux incertitudes de modèles, ceux-ci pouvant n'être que des approximations linéaires de dynamiques qui ne le sont pas. Les approches dites entrées-sorties peuvent même reposer sur des approximations de type boîte noire. De leur côté les approches stochastiques proposent une prise en charge totalement différente du problème de l'incertitude de modélisation et de non stationnarité, par l'estimation fonctionnelle des dynamiques inconnues.

Quand il s'agit d'objectifs de conduite (stabilisation, suivi de consigne, optimisation), la théorie du contrôle permet le calcul de loi commande adaptées aux caractéristiques du système (peu ou fortement non linéaire, mono ou multi variable en entrée et sortie), à la nature des incertitudes (paramétriques, fonctionnelles) et des perturbations.

Quand le système n'est observé qu'indirectement mais que la connaissance des variables d'état est souhaitée ou nécessaire (dans un but de supervision par exemple, cf. § suivant) on doit faire appel à des méthodes de reconstruction d'état (approche déterministe) ou de filtrage (approche stochastique) qui, en langage d'ingénierie des procédés, aboutissent à des <<capteurs logiciels>>.

Pour compléter la panoplie modélisation/filtrage/commande appliquée à un procédé donné, il est nécessaire d'y adjoindre des procédures de surveillance aptes à détecter en ligne l'occurrence d'une anomalie et opérer son diagnostic (panne d'actionneur, de capteur, contamination bactérienne, etc.). L'offre méthodologique à ce niveau est encore insuffisante et assez disparate. Elle emprunte à l'automatique (observateurs et filtres), à la statistique (tests de rupture de modèle, tests d'hypothèse multiple), à l'analyse des données (classification) et à l'intelligence artificielle (inférence

floue). La panoplie ainsi complétée constitue ce qui est appelé la supervision de procédé.

Signalons enfin l'utilisation, à l'INRA comme dans d'autres structures, d'approches neuronales (réseaux à couches essentiellement) en modélisation prédictive, contrôle et plus généralement supervision de procédés. Leur légitimité théorique reste quelquefois à consolider, malgré des réussites pratiques indéniables.

1.5 Approches qualitatives et symboliques

Elles s'adressent aux nombreux mécanismes de transformations agro-alimentaires complexes qui ne bénéficient pas de connaissances suffisamment formalisables analytiquement au moins pour certaines de leurs composantes, qui par contre peuvent donner lieu à l'établissement de règles de fonctionnement (procédé de fabrication fromagère, procédé de cuisson industrielle, etc.). Ces règles peuvent traduire une connaissance d'expert ou résulter d'une induction automatique par apprentissage à partir d'une base d'exemples. Elles peuvent être optimisées par leurs paramètres. Des architectures de rétroaction peuvent alors être construites au travers d'algorithmes reposant sur des équations logiques et la théorie des ensembles flous (contrôleurs flous), pour permettre une conduite du procédé satisfaisant des contraintes imposées (qualité finale). Ces approches ont donné et donnent lieu à plusieurs réalisations à l'INRA (CEPIA ENSIA, MIA) et au CEMAGREF notamment.

1.6 Points de blocage

Les thèmes précédents (modélisation, contrôle, filtrage, supervision, raisonnement flou) sont actuellement le support de travaux de recherches et de collaboration des équipes INRA concernées.

Les points de blocage rencontrés par les utilisateurs coïncident souvent avec ceux rencontrés dans d'autres contextes de modélisation :

- discernabilité, choix et validation de modèles
- intégration d'échelles multiples
- résolution analytique des équations
- analyse de sensibilité par rapport aux constituants des modèles (paramètres, composantes fonctionnelles, conditions initiales) aspects numériques : simulation, schéma de discrétisation d'EDO et d'EDP, identifiabilité et estimation de paramètres propagation des incertitudes paramétriques planification expérimentale pour la caractérisation et l'identification de ces modèles
- intégration des modèles (aspects multivariés, etc.)
- approche hybride : continu/discret, quantitatif/symbolique
- réduction de modèle (approche top-down).

D'autres sont plus spécifiques à des objectifs de conduite automatique :

- effet retard dans la commande
- planification expérimentale pour la commande.

Il faut signaler enfin deux autres difficultés propres au domaine du génie des procédés agro-alimentaires :

- l'urgence de la constitution de bibliothèques de modèles, facilitant leur standardisation, leur sélection, leur réutilisation et leur adaptation ;
- l'importance, dans un contexte social pressant, de la formalisation de modèles de qualité des produits, intégrables dans la conduite automatisée d'un procédé de transformation.

2. Analyse de structures

Elle constitue une importante activité du département CEPIA, que l'on retrouve dans d'autres contextes à l'INRA (sciences des sols, par exemple).

Sous le seul angle de la modélisation qui nous occupe ici, il s'agit fondamentalement de caractériser les structures et déterminants structuraux de produits alimentaires bruts ou élaborés, à différentes échelles et d'étudier les relations entre ces structures et les propriétés qu'elles présentent : mobilité, déformabilité, réactivité à différents agents. Cet objectif peut être lié ou non à la caractérisation d'une notion de qualité (tendreté de la viande, texture des fruits, par exemple). Mais la formalisation de cette notion est elle-même délicate.

2.1 Rôle de la modélisation

La modélisation, ici à base mécaniste essentiellement, doit permettre de comprendre et prédire la mise en place des caractéristiques structurales de ces produits, leurs

évolutions provoquées ou non et les conséquences de ces dernières. Elle doit intégrer des lois de changement d'échelle pour relier ces caractéristiques à des comportements mécaniques, de transport et de réactivité vis à vis de divers agents mécaniques, thermiques, chimiques, enzymatiques, etc.

2.2 Positionnement de l'INRA

Le département CEPIA possède des compétences en ce domaine. Ses besoins sont en partie satisfaits en interne.

2.3 Démarches méthodologiques et blocages

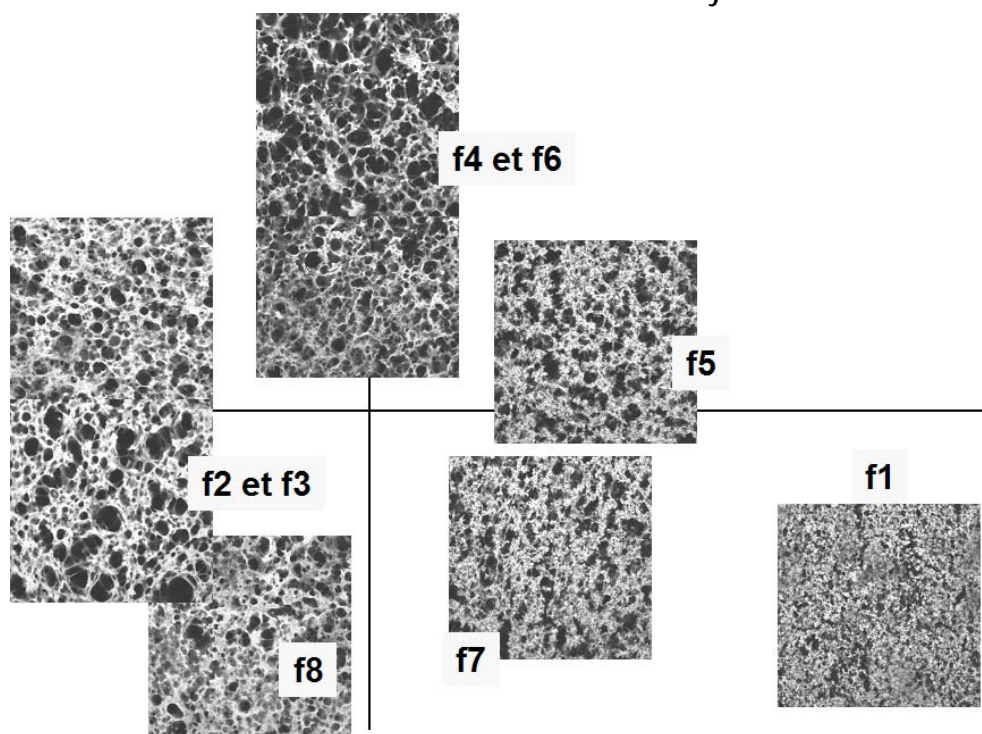
Elles associent des techniques d'exploration des mésostructures (analyse d'image, diagramme d'état), de modélisation sous forme d'EDO et de simulation.

Les difficultés rencontrées sont issues du caractère non linéaire des relations étudiées, de la forte dimension paramétrique, de l'existence et de la nécessaire intégration de données qualitatives.

Le besoin de construction de plans d'expérience adaptés à l'identification de ces modèles est réel.

Des questions fondamentales, comme la détermination des contributions de différents niveaux d'échelle aux phénomènes et propriétés structurales étudiées, ou comme la compréhension des évolutions dynamiques des structures, restent ouvertes.

Carte 1-2 de l'analyse factorielle discriminante



3. Ingénierie reverse

Cette approche, qui tire parti des méthodologies de modélisation précédentes, se distingue avant tout par sa finalité : appréhender de manière globale en les optimisant, les itinéraires technologiques conduisant à des caractéristiques déterminées pour un produit fini. Elle vise à intégrer les diverses données et informations biotechniques (choix des matières premières, modalités de transformation, de distribution, etc.) permettant une conduite de production optimale qui place le produit au départ de la réflexion et de l'action. Cette démarche est en cours d'initiation au département CEPIA.

3.1 Rôle de la modélisation

Il s'agit d'une problématique très intégratrice, qui fait appel à des référents très divers : génétiques, technologiques, économiques, logistiques, organoleptiques, etc. Chacun d'eux peut faire appel à des approches modélisatrices appropriées, parfois déjà disponibles. L'originalité porte avant tout sur l'assemblage de ces informations et leurs exploitations coordonnées pour satisfaire un cahier des charges. La modélisation à considérer à ce niveau porte sur l'élaboration de modèles de représentation de connaissances et d'indicateurs de suivi et leur utilisation en aide à la décision.

3.2 Positionnement de l'INRA

Le département CEPIA est détenteur et producteur des connaissances exploitables. Des projets intégrateurs de ce type sont en cours de démarrage (par exemple *grain virtuel* à Montpellier, *Panification* à Nantes, *Arômes* à Dijon) qui vont solliciter des compétences que ce département s'apprête à développer, ainsi peut-être que celles déjà existantes du département MIA.

3.3 Démarches méthodologiques

Elles reposent sur divers outils d'intégration, de fusion et d'exploitation de données de natures érogènes (méthodes statistiques, neuronales, IA). Elles doivent mobiliser de façon privilégiée des méthodes de représentation de connaissances expertes ou induites et des méthodes de modélisations qualitatives pour l'inférence et la décision qui s'appuient sur les premières. Les méthodes de raisonnement qualitatif à la base de ces modélisations doivent tirer parti de l'expertise disponible mais aussi de la connaissance contenue de manière implicite dans les données expérimentales. En effet, s'agissant de systèmes aussi complexes, les très nombreuses interactions entre variables ne peuvent être restituées par des experts humains. Des techniques d'induction automatique par apprentissage doivent alors être mises en oeuvre pour extraire des règles de fonctionnement aptes à modéliser au mieux les processus. Des systèmes d'inférence reposant par exemple sur la logique floue et utilisables dans l'aide à la décision peuvent alors être construits. MIA a développé plusieurs systèmes de ce type pour des applications aux transformations agro-alimentaires. Les travaux actuels portent sur l'enrichissement et la compatibilité de

différents systèmes de règles (règles conjonctives, règles implicatives, etc.).

4. Analyse de risques alimentaires

Sous l'effet d'une forte pression sociale, cette activité, s'est vu reconnaître une importance croissante à l'INRA ces dernières années. Elle porte sur les analyses quantitatives et prévisionnelles de la présence et de l'influence d'agents pathogènes divers dans les produits alimentaires : bactéries, mycotoxines, xénobiotiques, spores, etc.

Dans sa composante de microbiologie prévisionnelle il s'agit d'un thème assez ancien (Buchanan 1918), par ailleurs apparenté à l'écologie microbienne recouverte par le Thème 2.

4.1 Rôle de la modélisation

La modélisation doit permettre l'élaboration et l'évaluation probabiliste d'indicateurs spécifiques et notamment de prédicteurs d'intensité de contamination, d'indice de risque, etc., fonction du temps et de certains facteurs d'environnement : évaluation de l'exposition répétée et cumulée à un contaminant, effectifs bactériens, etc.

4.2 Positionnement de l'INRA

Plusieurs équipes notamment à Jouy (Matrisq-MIA), à Paris (Met@risk-MIA, UMR RISQUAL) à Avignon (SOPOV-CEPIA), à Theix (SRV-CEPIA), etc, en partenariat avec diverses structures (AFSSA, EV, DGAL), développent des recherches actives autour des points précédents. Ces équipes sont quelquefois multidisciplinaires (modélisateurs, microbiologistes, informaticiens). La coordination de ces différents acteurs est encore perfectible : le projet SYM'PREVIUS qui associait différents partenaires (dont des industriels) et dont l'INRA abritait la cellule opérationnelle (LGPTA-Lille) a été gelé avant sa phase III (décroissance et survie).

4.3 Approches méthodologiques

Les méthodes d'évaluation des risques microbiens reposent classiquement sur la construction de modèles de prédiction, la recherche de lois de distribution a priori (méta-analyse) pour les paramètres d'entrée et enfin l'estimation des distributions marginales a posteriori des variables d'intérêt (indicateurs), le plus souvent par des approches de types MCMC. Pour ces modèles de prédiction en microbiologie prévisionnelle, on distingue traditionnellement *les modèles primaires*, portant soit sur la croissance soit sur la décroissance temporelle de populations bactériennes, et *les modèles secondaires* portant sur les paramètres des modèles précédents. Ces « paramètres » sont en fait des fonctions complexes de facteurs écologiques, technologiques et physiologiques.

Les modèles primaires de croissance bactériennes les plus utilisés actuellement sont des développements de quelques modèles de base : logistique modifié, logistique avec rupture, Gompertz modifié, Baranyi. Les modèles primaires de décroissance sont encore moins nombreux

parce que plus récents : ex. le modèle d'Albert et Maffart, pour modéliser les effets d'un traitement thermique sur la décroissance.

Les modèles dits secondaires cherchent essentiellement à traduire les effets de certains facteurs environnementaux sur les paramètres les plus importants des modèles précédents : vitesse maximale, temps de latence, etc. Ces facteurs relèvent de quelques classes principales : caractéristiques physico-chimiques des aliments (pH, potentiel redox, etc), technologiques (traitements antimicrobiens), liés à la conservation (froid), biologiques (interactions inter-populations), et interactions entre tous ces facteurs.

Par ailleurs, la théorie probabiliste des valeurs extrêmes et des statistiques qui en découlent est au coeur des approches d'analyse des risques liées aux expositions à des contaminants récurrents (ochratoxine-A, métaux lourds, etc.) : détermination du seuil correspondant à un niveau de risque donné, probabilité de dépasser un seuil donné d'exposition, probabilité de dépasser une limite légale par une contamination maximale ou inversement détermination d'une norme sanitaire, etc.

Ce thème possède également une composante informatique très importante qui mobilise d'autres efforts concernant la modélisation des connaissances (bases de données microbiologiques) et leur exploitation optimisée (data mining, réseaux bayésiens, graphes conceptuels), à l'interface de l'informatique, de l'intelligence artificielle et de la statistique. Cette activité relève davantage d'une problématique de traitement de données complexes et de leur interprétation dans le cadre particulier des modèles de microbiologie prévisionnelle.

4.4 Points de blocages

La complexité des modèles utilisés, notamment en microbiologie prévisionnelles, soulèvent des problèmes classiques en modélisation non linéaire et en modélisation de dénombrements : échantillonnage-observabilité, identifiabilité, planification expérimentale, aspects numériques.

Par ailleurs, quelques pratiques seraient à améliorer :

- l'estimation directe des paramètres de ces modèles est souvent très délicate, quelquefois même en pratique impossible. Les approches utilisées reposent alors sur des procédures d'estimation séquentielle : phases alternées de calage et d'estimation proprement dite sur des sous-ensembles de paramètres ;
- l'acquisition des données se fait le plus souvent par des méthodes classiques de dénombrement (série de

dilutions et comptage). Mais le caractère aléatoire du comptage ainsi opéré n'est pas pris en compte généralement. La variabilité intrinsèque est donc ignorée. Pas d'intervalle de confiance associé aux réponses des modèles ;

- faible utilisation de la planification expérimentale ;
- enfin, et ce qui peut paraître étonnant, les apports possibles de l'analyse des systèmes dynamiques à l'étude des dynamiques de populations bactériennes concernées (caractérisation d'équilibre, comportements limites, etc.) semblent en grande partie ignorés.

4.5 Quelques enjeux actuels

- Apport d'une modélisation spécifique à la compréhension de l'état physiologique des bactéries lors des différentes phases de croissance ou non croissance, notamment lors de la phase de latence en début de croissance.
- Détermination de modèles secondaires plus performants en termes de prédiction de latence et de vitesse de croissance maximale (apport des analyses de sensibilité et de la planification optimale).
- Modélisation de la survie.
- Introduction d'information a priori pertinentes (experts) dans l'utilisation prédictive des modèles en analyse de risques (détermination de lois prédictives d'effectifs a posteriori par MCMC).
- Evaluation d'effet d'accumulation d'exposition à des contaminants par modèle dynamique.

5. Conclusion

Les quatre activités précédentes à finalités agro-alimentaires pour l'essentiel, ont des recours différents à la modélisation en fonction de leurs objectifs spécifiques. Toutefois plusieurs des approches méthodologiques qu'elles sollicitent peuvent se recouper, notamment celles concernant le cadre de la modélisation dynamique, qui leur est en grande partie commun, même si l'utilisation qui en est faite (simulation, prédiction, contrôle, aide à la décision) diffère d'une activité à l'autre.

La différenciation ainsi opérée entre ces usages spécifiques a permis leurs caractérisations méthodologiques respectives et va permettre, dans les étapes suivantes de cette enquête, une analyse plus appropriée des besoins et des solutions qu'ils appellent à l'INRA.

Thème 7

Sciences sociales

Un certain nombre de « spécificités » des disciplines relevant des sciences humaines et sociales, ainsi que de leur démarche de modélisation, spécificités souvent mal perçues par les chercheurs en sciences de la nature, suggèrent un avertissement préliminaire.

Les sciences humaines et sociales sont éminemment des « sciences du contexte », au sens où elles ne se réduisent pas à une science de l'action dans un monde d'agents intelligents ; ni à une analyse des relations diversement contradictoires entre un « monde des idées », lieu d'élaboration et de remaniements continuels de représentations et de symboles, et un « monde physique », lieu d'une réalité objective supposée ; ni à l'étude d'une intersubjectivité productrice de « sens » dont on pourrait décliner fonctionnellement les objectifs, les ressources et les contraintes. De cette posture intellectuelle prudente, héritée de siècles d'erreurs et de réflexions diversement fécondes, découlent plusieurs propositions problématiques.

Parmi les plus importantes citons : le statut de *l'induction par les « faits »* en sciences sociales, le fait social pouvant difficilement être assimilé aux faits « physiques » objectifs, au sens d'accessibles à l'expérimentation contrôlée, que traitent la physique ou la biologie ; le statut ambigu du « *mécanisme social* » dans ces disciplines tant du fait de la prééminence de « l'intercausalité » dans les relations humaines, que du caractère éminemment conventionnel, et donc critiquable, de toute tentative de description « mécaniste » d'une société ou du comportement humain ; le statut de *l'historicité*, en tant que moyen d'accès supposé à un ensemble réduit de règles générales objectives (des conditions « nécessaires » si l'on veut) censées rationaliser les dynamiques historiques des sociétés et leurs possibilités de développement dans les domaines économiques, politiques ou culturels.

Face à ces enjeux problématiques, les sciences humaines et sociales ont répondu par une grande variété de postures méthodologiques, voire épistémologiques, source de nombreux débats qu'il serait erroné de réduire à de simples « débats d'école ». Pour ce qui concerne la modélisation en sciences sociales, on peut citer le débat entre les approches dites « holistes », se réclamant d'une acception étendue de la notion de « système », et les approches relevant de « l'individualisme méthodologique », qui font du comportement des acteurs individuels le fondement des dynamiques sociales, ou encore le débat entre les tenants d'une rationalité substantielle des acteurs, où ces derniers sont supposés détenir des capacités cognitives illimitées de résolution de problèmes, et les partisans d'une rationalité limitée, où ces capacités sont contraintes *a priori*.

Il n'existe pas aujourd'hui de « mathématiques du contexte » et partant, la mobilisation des mathématiques et de l'informatique dans une démarche logique de modélisation qualitative ou quantitative en sciences humaines et sociales suppose un effort parallèle d'identification du contexte dans laquelle elle s'insère. La manière d'opérer cette identification est très variable selon les disciplines. L'économie mathématique ou la géographie quantitative, par exemple, vont s'appuyer sur une démarche hypothético-déductive de nature conventionnelle, où la spécification des hypothèses du modèle va permettre de borner *a priori* son domaine de pertinence ou de validité, selon une méthode inspirée des sciences physiques. La géomatique va mobiliser les outils de l'informatique et de l'intelligence artificielle pour modéliser des systèmes complexes, se rapprochant des démarches de l'écologie et de l'agronomie au sens large. La psychométrie ou la sociométrie vont emprunter les protocoles statistiques des sciences expérimentales en tentant de mimer les conditions de validité d'une expérimentation en milieu contrôlé.

1. Rôle de la modélisation en tant que démarche de recherche

Le terme « modèle » est excessivement polysémique en sciences humaines et sociales, allant du « modèle de l'inconscient » de Freud aux modèles d'options financières de Black et Scholes, une déclinaison des équations de Navier-Stokes dans le domaine de la variabilité des cours sur les marchés financiers. Il en résulte que préciser le rôle de la modélisation, ou plutôt des modélisations, dans la démarche de recherche en SHS est un exercice de catégorisation très difficile.

L'effort de modélisation résulte d'un « parti pris » sur des éléments jugés fondamentaux de la compréhension d'un phénomène social (plutôt que d'un « mécanisme » pour les raisons citées plus haut). En d'autres termes, un modèle est une représentation de la réalité ; on choisit un point de vue pour considérer cette réalité.

Il serait très hasardeux de tenter une typologie des « modèles » développés en sciences sociales (comme ailleurs sans doute). Il importe néanmoins d'établir quelques distinctions importantes. La première distingue les modèles « idéal-type », pas forcément « formalisés », modèles qui relèvent de la catégorie des « expériences de pensée », et qui visent à fournir des éléments d'interprétation et d'explication de phénomènes sociaux au moyen d'une représentation volontairement simplifiée de ces phénomènes, et des modèles au sens « objet », qui sont des outils de représentation formalisée, la plupart du temps sur support informatique, dédiés à la

représentation et à l'analyse de dynamiques sociales ou d'action. La seconde distingue des modèles « de connaissance » et des modèles « d'action ». Ce qui les distingue, ce sont les destinataires des modèles et par là, les qualités que l'on attend d'eux (comprendre un processus ou le simuler pour envisager des moyens de le contrôler, expliquer un fonctionnement ou comparer différentes alternatives...).

Certaines disciplines, comme l'économie, pratiquent des modélisations « mécanistes » des dynamiques économiques. D'autres, comme la sociologie, l'anthropologie, la géographie ou l'histoire ont des approches plus qualitatives, à forte composante heuristique, mais ne sacrifiant pas pour autant aux exigences de rigueur scientifique que l'on trouve dans la modélisation quantitative.

Toutes les sciences sociales mobilisent à des degrés variables les outils de la statistique à partir de canaux d'acquisition de données multiples, allant de l'expérimentation contrôlée à l'analyse sémiologique de discours en passant par les enquêtes de terrains ou le travail sur des archives historiques et archéologiques. La logique, le raisonnement, l'analyse de graphes ou d'images, plus spécifiques des approches en intelligence artificielle, sont également utilisés.

2. Place de l'INRA

Les SHS à l'INRA sont essentiellement présentes dans le département SAE2 et dans une moindre mesure au SAD. Dans les deux cas, l'activité de modélisation est quasi-systématique dans les questions étudiées. Le département SAE2 développe en propre des modèles dits d'équilibre général calculable à l'échelle nationale et européenne à des fins de conseil pour le pilotage de la politique agricole intra et inter sectorielle (élevage, grandes cultures et matières premières alimentaires), l'impact des réformes de la PAC et l'anticipation de ses conséquences sur la balance extérieure agricole de la France et de l'Union Européenne. Un modèle mondial est en cours de construction. Le département développe et maintient des modèles sectoriels d'offre de produits agricoles de base et transformés par les IAA (lait, filière viti-vinicole, grandes cultures, oléo-protéagineux, filière viande, filière bois) et possède de fortes compétences en termes de modélisation de la consommation alimentaire et des marchés de la grande distribution en France. Il développe également des modèles de commerce international à l'échelle de l'Europe et du monde sur de grands produits, base de conseil politique pour les négociations de l'OMC. Sont également modélisés l'analyse coût-bénéfices des politiques environnementales (outils de mesure des bénéfices non marchands en particulier), la gestion des ressources naturelles (eau, forêt, biodiversité), l'analyse des stratégies de R&D des firmes de l'agro-fourriture et de biotechnologies, les dynamiques foncières et les stratégies patrimoniales des actifs agricoles, le marché du travail agricole, ainsi que les logiques d'occupation de

l'espace en zone rurales et péri-urbaines. Le SAD développe des modélisations en propre des déterminants socio-économiques de la gestion des espaces agricoles, de l'analyse du travail agricole, des logiques de filière, des systèmes de gestion publique des mesures agri-environnementales. Il s'intéresse également à la gestion des ressources naturelles et de l'environnement dans ses dimensions économiques, anthropologiques et sociétales et à la coordination des activités dans une perspective de développement territorial.

Les SHS à l'INRA sont surtout utilisatrices de méthodologies de modélisation développées ailleurs de leurs champs de recherche. Mais ces méthodologies sont ensuite adaptées au contexte agricole pour déboucher sur des outils de modélisation (au sens objets) originaux, outils dont les départements assurent en général la totalité du développement, du recueil de données au transfert vers les utilisateurs, en passant par les phases de conceptualisation, développement informatique, validation et documentation. Ils en assurent également la maintenance pour leurs besoins de recherche propres ou pour maintenir une capacité de réponse et d'expertise face à des partenaires extérieurs, le plus souvent des acteurs publics (régions, ministères, commission européenne) ou issus du monde agricole et industriel.

A l'exception des modèles d'équilibre général et de certains modèles de consommation (analyse simultanée des déterminants de la consommation de plusieurs milliers de produits alimentaires à partir de données portant sur plusieurs millions de ménages), les modèles développés en SHS sont plutôt de « petite » taille, très orientés sur une ou quelques finalités précises, et partant ne sont donc pas toujours faciles à interfacer ou à intégrer dans des plate-formes de modélisation générique. Les outils généraux de développement de modèles sont souvent ignorés. Les outils développés sont des modèles de « chercheurs », limités aux connaissances propres de leurs auteurs, connaissances très hétérogènes. Il existe très peu de développements informatiques originaux dans ces travaux, la plupart des outils étant développés à partir de plates-formes de simulation informatique (systèmes multi-agents...) ou d'analyse statistique standard (SAS, GAUSS, GAMS, Mathematica...), voire d'outils bureautiques (tableurs, base de données, SIG). Dans quelques cas, l'usage des méthodes conceptuelles en informatique a permis de décliner des modèles spécifiques sur des outils génériques. Les modèles sont alors plus souvent des maquettes qui évoluent au fil de la conceptualisation du problème de recherche et sont utiles dans leur construction. Par ailleurs, ils peuvent être adaptés à un usage avec des acteurs, ce qui modifie leurs modalités d'élaboration. Cela demande d'articuler différentes méthodes, de combiner concepts et outils, et d'intégrer plusieurs domaines de connaissances.

La gestion des bases de données se fait le plus souvent à partir d'achats auprès de fournisseurs extérieurs (INSEE, SCEES), puis dégroupée au niveau des unités, voire d'individus ou de personnes ressources. Les

chercheurs des deux départements ont également une activité conséquente de recueil de données de terrain, à partir d'enquêtes ou d'entretiens. Ces données sont de nature quantitative et qualitative. Les systèmes génériques de gestion d'information, type clients serveurs, sont très peu développés, mais les départements concernés font actuellement de gros efforts d'organisation en la matière.

En termes de modélisation socio-économique de l'agriculture, l'INRA est en position de leader national et très souvent de monopole de fait, étant le seul institut de recherche capable d'apporter une masse critique de recherche et d'ingénierie suffisante pour développer des modèles et les maintenir sur la durée. L'Institut est aussi un acteur important et reconnu à l'échelle européenne sur ces questions, très impliqué dans les réseaux de modélisation du secteur agricole européen, et depuis plus récemment à l'échelle mondiale (nombreux développements en cours en partenariat national et international).

3. Identification et analyse des problèmes et blocages

Les principaux problèmes rencontrés par la modélisation en sciences sociales tiennent en premier lieu aux sources de données. Ces données sont fréquemment acquises auprès de prestataires extérieurs souvent au prix de négociations compliquées. Par ailleurs ces données peuvent être confidentielles (secret industriel) ou d'une fiabilité difficile à apprécier (déclarations de revenu des agriculteurs). Par ailleurs ces données peuvent être d'une utilisation problématique : quelle robustesse accorder à des sondages d'opinion ou de popularité par exemple ? D'autres sources sont les données d'enquêtes, qui posent des problèmes de formalisation, d'échantillonnage, de comparaison et de validation.

Le deuxième niveau concerne le degré d'intégration recherché des phénomènes en jeu. Un exemple sera utile pour fixer les idées. Supposons que le chercheur s'intéresse au déterminants du choix d'une innovation technique par un agriculteur. Ce choix va être influencé par l'observation d'autres agriculteurs, les avis d'organismes de développement, du négoce, de la presse et des médias spécialisés, mais aussi de paramètres comme les prix des produits et des facteurs ou le taux d'intérêt de l'emprunt finançant l'investissement. Un tel choix individuel ne peut donc pas être relié simplement à une échelle spatiale (la localisation de l'exploitation), temporelle (l'état de l'information au moment de la décision), pas plus qu'aux caractéristiques propres de l'exploitant (son âge, son niveau de technicité). Le choix des variables « exogènes », c'est à dire supposées indépendantes du choix, et « endogènes », c'est à dire influencées par le choix (supposons que l'agriculteur doive se former à la nouvelle technique ou envisager une restructuration

foncière pour pouvoir l'exploiter efficacement par exemple) est donc par nature un exercice délicat s'agissant de systèmes sociaux par construction « ouverts » au sens fort. Il en résulte que les sciences sociales sont moins préoccupées par des questions d'emboîtement d'échelles spatiales ou temporelles que les sciences naturelles, ces échelles étant généralement enchevêtrées dans tout modèle explicatif des comportements sociaux. Ce sont plutôt les interactions entre niveaux d'organisation qui sont importantes. Par contre, elles seront très attentives dans un dialogue interdisciplinaire à la définition au préalable de ce qui sera traité comme des « paramètres » et comme des « variables », ces « variables » ne se résumant pas au choix des acteurs. Par exemple, le poids économique de la production agricole française en Europe pour certains produits comme le blé, le lait ou les œufs est tel qu'une modification des décisions de production des agriculteurs sur ces produits va entraîner des modifications significatives des prix, prix qu'il ne faut donc pas traiter comme des « données ». On retrouve ici le problème évoqué en introduction de l'importance d'une bonne définition du « contexte », préalable à toute modélisation pertinente et de la nécessité d'explicitier les points de vue retenus.

Le dernier facteur de blocage concerne le niveau d'abstraction requis pour une modélisation efficace. Les sciences sociales ne peuvent pas s'appuyer sur un corpus de lois générales, lois qui se déclinent sous des formes particulières selon la nature des conditions aux limites ou d'un ensemble réduit de conditions initiales. Ces « lois » prennent davantage la forme de principes généraux dont la déclinaison sur un champ d'étude précis est à chaque fois, ou presque, à construire conceptuellement. Des modèles très abstraits mobiliseront plus efficacement ces principes généraux mais peineront à rendre compte des phénomènes observés. A l'inverse, des niveaux d'abstraction faibles « colleront » mieux aux faits mais leur capacité explicative ou interprétative sera également faible, faute de pouvoir se raccorder à un nombre réduit de principes généraux. Ce point est également problématique dans un dialogue interdisciplinaire ou en partenariat, dans la mesure où il sera nécessaire, bien que difficile, de mettre en œuvre des heuristiques de modélisation progressive, partant d'un « premier regard » un peu général sur les phénomènes à expliquer pour ensuite affiner progressivement la compréhension de ces phénomènes. Dans un tel contexte de travail, les chercheurs en sciences sociales seront rapidement contraints à des remaniements importants de leurs modèles (hypothèses, processus à intégrer) et de leur dispositif d'étude (type de données nécessaires, acteurs à prendre en compte, choix du terrain d'étude) à chaque étape.

Seconde partie

Synthèse générale

1. Rôle de la modélisation dans la démarche de recherche

Le rôle de la modélisation est croissant dans tous les domaines de l'INRA et son utilisation tend à devenir incontournable dans un grand nombre de thèmes. C'est un moyen effectivement très puissant pour intégrer et formaliser les quantités croissantes de connaissances et de données sur les processus biologiques ou socio-économiques. C'est aussi un outil fondamental pour le transfert de la recherche vers le développement.

1.1 Différentes formes d'intervention

On peut distinguer deux grandes formes d'intervention de la modélisation dans la démarche de recherche.

Le modèle statistique, basé sur des relations empiriques entre données d'entrée et observations en sortie d'un processus, qui permet de quantifier des relations et de tester des hypothèses les concernant (effet d'un gène sur un caractère phénotypique), de quantifier l'effet de différents facteurs dans un processus (effet de la température sur la croissance d'un végétal), d'estimer des risques (risque de maladie dans une population), d'identifier et de caractériser des structures dans les données (identification d'une structure spatiale dans une épidémie), d'intégrer des résultats multiples en une information synthétique (méta analyse de données de risques alimentaires), etc.

La modélisation statistique est utilisée depuis longtemps pour tester les hypothèses biologiques simples et elle est au cœur de certaines disciplines, comme la génétique quantitative. De plus, c'est pour tous les thèmes une étape incontournable lorsque l'on est confronté à une quantité importante de données.

Dans la modélisation mécaniste on intègre les connaissances (biologiques, physiques, sociales, etc.) que l'on a sur le processus sous différentes formes : analytiques dans les modèles à base d'équations différentielles ou aux dérivées partielles (l'outil fondamental de la modélisation mécaniste), qualitatives ou symboliques (intégration de connaissances d'expert), ou multi-agent lorsque l'on veut modéliser les interactions entre agents. La différence entre un modèle mécaniste et un modèle empirique est particulièrement nette lorsque le processus à modéliser est la combinaison de plusieurs processus. Le modèle empirique les représente indistinctement sous la forme d'une équation synthétique regroupant les différents processus, tandis que le modèle mécaniste conservera la formalisation de chacun des processus.

Ce type de formalisation permet de décrire des processus, de les intégrer dans un même système (intégration des différents processus impliquant des flux de matières en écophysologie), de hiérarchiser des processus, de faire de l'expérimentation virtuelle, d'évaluer l'efficacité d'une stratégie (lutte contre une maladie ; comparaison d'itinéraires techniques), d'identifier des trous dans la connaissance, etc.

Notons que les méthodes statistiques ont aussi un rôle important à jouer dans le cadre des modèles mécanistes, quand il s'agit de les confronter aux données : estimation de paramètres, validation, etc.

1.2 Modèles pour la connaissance

Globalement, on retrouve les différents objectifs des modèles pour la connaissance dans les différents thèmes : connaître les processus (connaître les fonctions des gènes, les voies d'écoulement dans les bassins versants, quantifier l'effet d'une maladie), tester des hypothèses ou des combinaisons d'hypothèses (en simulant le comportement d'un système complexe, en physiologie par exemple), intégrer des processus, ou représenter des processus (identification de structures spatiales ou de cycles temporels).

Un élément important est le lien entre la modélisation et l'acquisition de données par l'expérimentation. Les modèles s'appuient la plupart du temps sur une acquisition intensive de données qui sont nécessaires pour quantifier les éléments des processus, et réciproquement le modèle doit pouvoir être utilisé pour guider l'expérimentation, voire s'y substituer, et la rendre plus efficace pour répondre aux interrogations soulevées par le modèle (identification de « trous » dans la connaissance ; difficulté d'estimation de certains paramètres). D'autre part, les modèles permettent d'avoir accès à des données non mesurables (flux dans le sol ; variables chimiques ou biologiques dans un fermenteur) en les reconstruisant. Modélisation et expérimentation ne doivent donc pas être considérés comme antagonistes, mais comme complémentaires dans la construction de la connaissance.

La modélisation permet d'intégrer des processus à des échelles variées (spatiales, temporelles, d'organisation du vivant) et de simuler le modèle global obtenu pour tester des hypothèses sur le fonctionnement à cette échelle globale. Cette intégration doit donc jouer un rôle déterminant pour la connaissance dans le cadre de la biologie intégrative, mais aussi pour les modèles d'action qui sont souvent confrontés à des systèmes complexes, associant les aspects biologiques, physiques, techniques et socio-économiques.

Enfin, la modélisation peut aussi se trouver au cœur de la démarche de recherche dans la thématique. C'est le cas par exemple des modèles physiques (du moins pour les processus simples) où de nombreux concepts sont définis par des modèles mathématiques, ou en écologie et en économie où le modèle sert de support à l'expression de certains concepts. C'est aussi le cas dans d'autres domaines (écophysologie, agro-éco-systèmes par exemple) lorsque l'objet d'étude est le système lui-même. On peut penser que certains concepts qui sont à créer en biologie intégrative utiliseront le formalisme des modèles.

1.3 Modèles pour l'action

Les modèles pour l'action sont utilisés pour aider à la décision de différents types d'acteurs. D'une part, les « gestionnaires de processus » : agriculteurs et techniciens

agricoles (recommandations en matière d'itinéraires techniques, irrigation, stratégie de lutte contre des maladies, alimentation animale, etc.), semencier (aide à la conception de variétés), gestionnaire de forêts (itinéraire technique forestier), responsable de conduite de procédés (gestion de conduite, surveillance et diagnostic), etc. D'autre part, la décision publique : gestion d'une épidémie, aménagement du territoire (structure du paysage et risques de dissémination d'OGM ou de sensibilité aux tempêtes), pilotage de la politique agricole (impact des réformes de la PAC), etc.

Les modèles pour l'action prennent une place importante quand il s'agit de processus aux plus grandes échelles qu'elles soient spatiales ou d'organisation du vivant, alors qu'ils sont peu présents aux autres échelles. Cependant, il y a plutôt un accroissement de l'utilisation des modèles pour l'action dans des thèmes qui étaient plutôt historiquement investis par des concepteurs de modèles de connaissance (thèmes 2 et 3 par exemple).

Les modèles pour l'action ne sont pas toujours développés directement avec cet objectif. Ce sont souvent des modèles de connaissance qu'il faut alors adapter et souvent simplifier en diminuant le nombre de paramètres et rendre connectables aux outils de conception, de simulation et d'aide à la décision.

2. Forces et faiblesses de l'INRA

L'INRA produit des modèles dans ses différents domaines d'intervention. Le plus souvent il s'agit de modèles relativement petits d'appui à la recherche, sans vocation à une utilisation et une diffusion très large. Il est utilisateur de modèles développés dans des domaines en amont de son intervention, comme par exemple en physique et biophysique (modèles atmosphériques à grande échelle). L'INRA est utilisateur de méthodologies développées par ailleurs, et les adapte au contexte particulier de ses domaines d'intervention (adaptation de modèles économiques au contexte agricole par exemple) mais il participe aussi dans de nombreux domaines au développement des méthodes de modélisation.

Une grande force de l'INRA, que l'on retrouve dans tous les domaines, est sa capacité à associer l'expérimentation ou plus précisément la production de données contrôlées, et la modélisation. Cette capacité d'expérimentation contrôlée peut devenir une force dans la construction des modèles et d'autre part, la modélisation renouvelle les questions de planification expérimentale.

Une autre force, qui complète la précédente, est la présence de compétences en matière de modélisation stochastique et en statistiques, qui permet d'aller au-delà des modèles déterministes et de construire les méthodes statistiques de confrontation des modèles aux données. De la même façon, les compétences en mathématiques et informatique de la décision, permettent de donner aux modèles pour l'action le support méthodologique nécessaire.

Enfin, une troisième grande force de l'INRA est la capacité à modéliser à l'interface entre les disciplines (modélisation biophysique du sol et physiologique de la racine, par exemple). Ces problématiques à l'interface entre les processus de natures différentes, sont des objectifs scientifiques en eux-mêmes pour l'INRA, et sont sources de problèmes méthodologiques spécifiques. Cependant, ce potentiel ne s'exprime pas toujours dans la réalité des projets. Par exemple, l'INRA n'est pas suffisamment leader dans les projets européens construits autour de « gros » modèles d'intégration.

L'INRA n'est pas suffisamment présent dans les modèles pour l'action, y compris dans ses domaines de compétence forte. Les modèles développés en épidémiologie ou en écologie sont rarement destinés aux gestionnaires. Les modèles à l'échelle des agro-éco-systèmes destinés à la décision publique ou aux professionnels, sont souvent de bonne qualité, mais leur pénétration dans les milieux socioprofessionnels est moins forte que dans certains pays Anglo-Saxons (Pays-Bas, Australie en particulier).

3. Quelques grandes classes de problèmes en modélisation à l'INRA

Notre analyse des problèmes soulevés dans les 7 thèmes nous amène à structurer les questions qui se posent en modélisation en plusieurs grandes classes qui soulèvent chacune des difficultés de différentes natures à différents niveaux dans le processus de modélisation, de la formalisation à la résolution mathématique ou informatique.

L'identification de ces classes de problèmes permet de s'apercevoir qu'elles sont chacune communes à plusieurs thèmes, et qu'il y a donc un espace de partage entre les thématiques sur ces problèmes génériques. Cette formulation des problèmes permet de plus de mieux mesurer les priorités et de dépasser les frontières disciplinaires entre les mathématiques et l'informatique.

3.1 Recherche d'une structure dans une grande quantité de données

Dans de nombreux domaines, génome, agro-éco-systèmes et alimentation en particulier, on est confronté à un nombre généralement très important de données, associé à une connaissance faible des processus en jeu dans le système à modéliser. On parle alors de modèle "boîte noire".

Cette recherche de structure renvoie à deux grandes questions :

- Comment reconnaître un motif structurel dans les données, ce qui pose des problèmes de définition du motif et de reconnaissance des formes ? (Fig. 1)
- Comment extraire des régularités de l'analyse d'un grand nombre de données ?

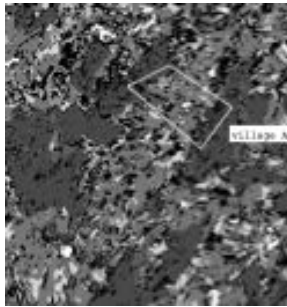
Des méthodes existent, méthodes d'apprentissage, data mining, fouille de données, représentation des connaissances, etc. Certaines sont issues des statistiques, mais sont moins exigeantes en matière de mathématiques

et demandent surtout à être jugées sur leur efficacité et nécessitent un dialogue renforcé entre thématiciens et spécialistes des méthodes. Cependant, les données mesurées ne sont pas toujours les plus pertinentes pour la modélisation ; les méthodes de filtrage ou de reconstruction de variables non mesurables prennent ici tout leur sens. On peut aussi citer l'utilisation de méthodes de reconnaissance de structures dans les langages telles

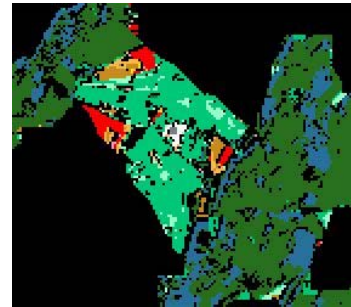
celles qui ont été utilisées en bio-informatique.

Initialement, la communauté scientifique qui a investi ce thème était plutôt concernée par des domaines éloignés de ceux de l'INRA (marketing). Mais elle commence à investir le domaine du vivant et l'INRA est très peu présent.

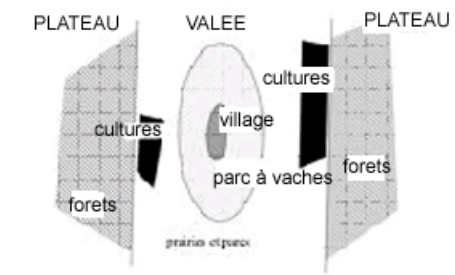
Image satellite prétraitée



Territoire affecté à un village



Modèle d'organisation spatiale «village vallée»



Reconnaissance du modèle «village vallée»

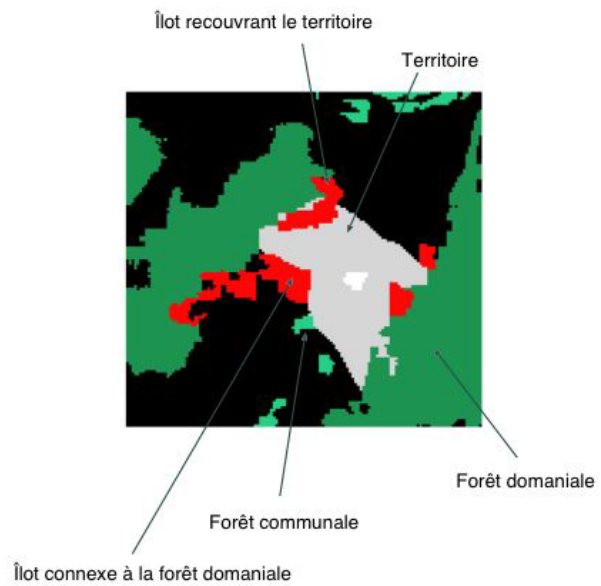


Fig. 1 - Recherche de structures spatiales dans des images

*Organisation spatiale de finages agricoles lorrains
Bachaccou et al. , 2004, Ecospace*

3.2 Représentation et exploitation d'informations dynamiques et spatialisées

Les informations concernant un processus dynamique à modéliser sont fréquemment spatialisées (Fig. 2). Elles sont alors le plus souvent hétérogènes, lacunaires et fortement variables. Il faut de plus pouvoir tenir compte des interactions de voisinage.

Des méthodes de caractérisation d'une structure spatiale existent, mais la caractérisation et l'analyse de systèmes dynamiques spatialisés posent des problèmes difficiles. La difficulté vient d'un nombre d'arrangement spatial possibles, quasiment infini. Quels sont les phénomènes génériques et ceux spécifiques à une configuration donnée? Est-il possible de trouver des propriétés génériques qui s'appliquent en dehors des répartitions aléatoires? Des pistes se développent, en particulier en physique, autour des structures qui seraient l'expression de la propagation de contraintes dans un réseau en état critique.

D'autre part, de nombreux problèmes de décision ou de contrôle sont spatialisés et demandent le développement de méthodologies adaptées.

L'INRA a une compétence reconnue dans le domaine de la modélisation spatiale, en particulier statistique. L'enjeu est maintenant de modéliser les structures spatio-temporelles.

3.3 Analyse, prévision et gestion des risques

De nombreux modèles - voir thèmes 3, 4, 5 et 6 - sont construits avec un objectif d'analyse et de gestion d'un risque. Cet objectif pose des problèmes particuliers puisqu'il s'agit souvent d'événements rares, associés à un

faible nombre de données, avec des états mesurés incertains, et avec des effets de seuil. De plus, ces modèles sont construits le plus souvent pour répondre à des questions de gestion et de contrôle.

Le problème est tout spécialement délicat lorsque l'on cherche à prédire une situation exceptionnelle, éventuellement très rarement ou jamais observée (changement climatique par exemple). Le modèle est alors extrapolé bien au-delà des données qui permettent de le construire.

De méthodologies existent comme la théorie des valeurs extrêmes. Cependant, ces méthodes sont plutôt développées pour des modèles relativement simples et posent des problèmes difficiles si l'on veut les mettre en œuvre dans des "gros modèles" ou pour le contrôle. On peut également proposer des méthodes d'apprentissage, où l'utilisateur apprend à réagir à des événements extérieurs dont l'impact est simulé par le modèle. Notons que l'analyse des risques pose des problèmes méthodologiques autres que ceux découlant du caractère rare des événements, abordés ailleurs dans ce chapitre.

L'INRA est présent dans ce champ. Cependant, il gagnerait à mieux associer les modélisations biologiques ou chimiques avec celles des sciences sociales (par exemple, le risque d'intoxication alimentaire dépend à la fois du comportement du consommateur et de la dynamique des populations de bactérie pathogènes). De plus, il faut intensifier les liens avec les chercheurs travaillant sur des thèmes différents de ceux de l'INRA (risque nucléaire, par exemple).

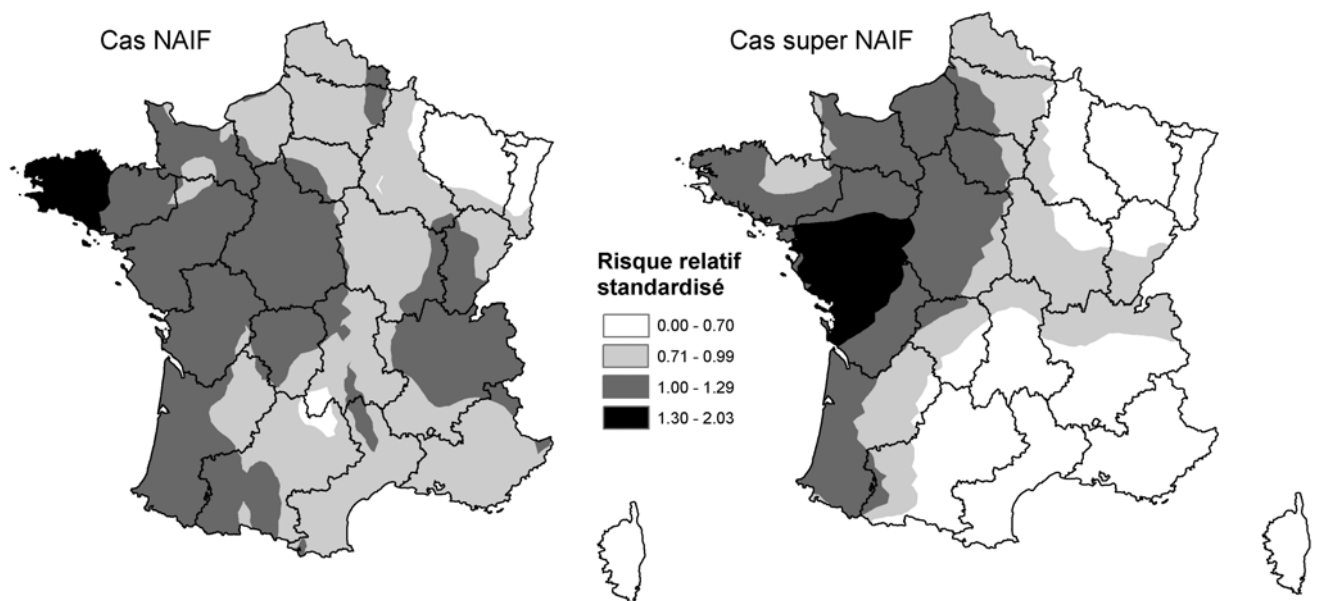


Fig. 2 - Cartographie du risque ESB en France, pour les cas NAIF et super NAIF ; les grisés indiquent l'intensité de ce risque exprimé par le risque relatif de contamination de chaque zone par rapport au risque national moyen.

Abrial D., Calavas D., Jarrige N., Ducrot C., Preventive Veterinary Medicine, 2005

3.4 Assemblage, intégration, réutilisation

Devant le coût important du développement d'un modèle spécifique à l'objectif de départ, et afin d'utiliser les connaissances contenues dans des modèles existants, le modélisateur est amené à assembler des modèles ou sous-modèles à l'intérieur d'un "grand modèle". Ce type de construction pose des problèmes à de nombreux niveaux. (Fig. 3)

Tout d'abord, les objectifs de ces sous-modèles sont différents, même si les objets sont communs, et certains processus peuvent être ignorés. Cela pose, de plus, des problèmes d'échelle spatiale ou temporelle, et peut être même des problèmes d'échelles d'organisation du vivant. A quelles conditions peut-on utiliser un modèle construit à une échelle inférieure dans un sur-modèle à une échelle supérieure ? Comment agréger le résultat à une échelle supérieure ? Il existe des méthodes d'agrégation, mais leur utilisation demande une technicité particulière et elles ne répondent pas à toutes les situations, en particulier pour les modèles complexes. Pour la désagrégation, la difficulté est d'avoir accès aux informations nécessaires pour renseigner le modèle à l'échelle inférieure.

Les modèles ainsi construits sont très complexes avec un grand nombre de paramètres, et des informations plus ou moins lacunaires et de type et de qualité hétérogènes. Les

problèmes méthodologiques liés à la manipulation des systèmes dynamiques complexes, font l'objet d'un développement particulier au chapitre 3.5.

Enfin, de conception informatique. Il s'agit notamment de faire en sorte que les différents modèles communiquent, c'est à dire que les sorties de l'un soient compatibles avec les entrées de l'autre. Les modèles doivent aussi être couplés avec des systèmes d'informations, entre autres géographiques. Les modèles peuvent aussi être articulés, soit en combinant des indicateurs locaux et globaux, soit en utilisant les résultats d'un modèle pour construire les indicateurs d'un autre modèle. Ce ne sont pas seulement les valeurs des données qui sont transmises, mais également les modalités de traitement.

Tout ceci renvoie aussi à la nécessité pour ceux qui construisent les modèles destinés à être réutilisés, à les concevoir avec l'idée de capitalisation et de communication future potentielle. Les plates-formes de modélisation tels *Capsis* sont une réponse possible.

Sur le plan des méthodologies informatiques, l'INRA est plutôt utilisateur de méthodes développées par ailleurs. Il faut assurer le relais auprès des développeurs informatiques. Sur les autres plans, peu d'équipes sont présentes dans la communauté nationale et internationale.

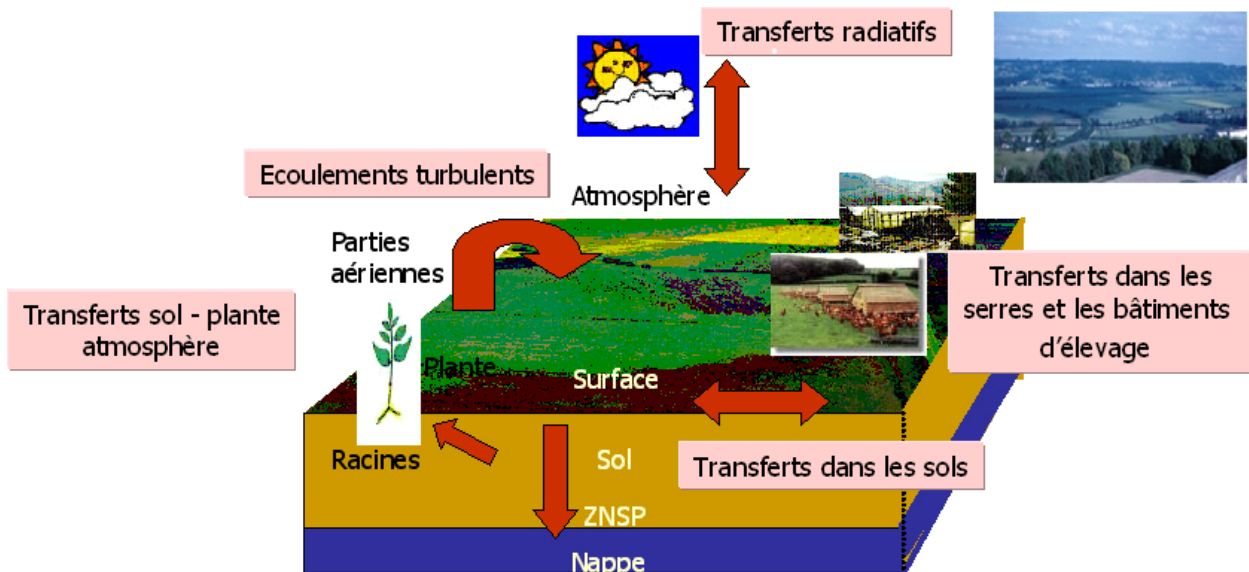


Fig. 3 - Du cm²... à la parcelle... au paysage, Y. Brunet

3.5 Systèmes dynamiques complexes ou compliqués

De nombreux modèles dynamiques, construits à partir d'assemblages de sous-modèles, plus ou moins spatialisés, éventuellement orientés vers l'analyse des risques, etc., sont très complexes, avec un grand nombre de paramètres, des interactions multiples et des informations plus ou moins lacunaires et de type et de qualité hétérogènes.

L'analyse et le traitement de ce type de modèles renouvellent les problématiques "classiques" de l'analyse mathématique, informatique et statistique et plus généralement de la confrontation aux données : planification et échantillonnage, estimation des paramètres, analyse de sensibilité, validation, analyse quantitative du modèle, calibration, filtrage, contrôle, assimilation de données, etc. Ils posent le problème de la représentation d'un système complexe par un système "manipulable", en quelque sorte une représentation minimale. C'est notamment utile lorsque l'on a un objectif d'action : faut-il et comment le réduire pour le rendre efficace (vitesse, coût) sans perdre trop d'informations (qualité, robustesse, etc.) ?

Ces développements interpellent aussi bien les statistiques que les systèmes dynamiques à base d'équations différentielles qui sont au cœur de ces modélisations. Ils peuvent aussi concerner les modèles informatiques : modèles qualitatifs, symboliques, multi-agents etc. Dans ce cadre, on a affaire à l'assemblage ou au couplage d'informations de natures différentes (données discrètes, continues, symboliques, logiques...). Dans certaines situations, les solutions ne pourront être qu'empiriques.

Dans ce champ, les compétences sont fortes à l'extérieur de l'INRA, et un objectif prioritaire est de les mobiliser sur des objets intéressants l'INRA. Cependant, l'existence de problèmes spécifiques liés aux objets de recherche de l'INRA, et le caractère central de la problématique, rendent nécessaire le développement de compétences à l'intérieur de l'INRA.

3.6 Modélisation du comportement individuel et/ou collectif

Lorsque des individus (hommes ou animaux) sont objets du modèle, un élément décisif est la modélisation de leurs comportements dans le contexte où ils évoluent. Il s'agit de modéliser la perception qu'ils ont de leur environnement, les processus de décision, les processus de mise en œuvre des décisions, et les interactions entre les différents acteurs ou groupe d'acteurs.

Pour les hommes, les sciences sociales et l'Intelligence artificielle sont sollicitées pour modéliser les processus cognitifs et sociaux. Les difficultés surviennent lorsque plusieurs acteurs sont impliqués, lorsque les décisions dépendent de critères multiples et évolutifs, ou lorsque les acteurs doivent s'approprier les informations lors de négociations. Dans ce cas, il est possible d'utiliser le modèle, non pas pour rendre compte du comportement des acteurs, mais comme outil à manipuler par les acteurs pour étayer leur raisonnement (notion de jeu de rôle, dont le concept a été particulièrement développé au CIRAD avec

des techniques multi-agents).

Ceci concerne également l'écologie, l'épidémiologie et l'éthologie. En écologie des communautés par exemple, il est courant de formuler les théories avec le langage mathématique des équations différentielles, des EDP ou des processus stochastiques (réseaux trophiques, métapopulations). Une question centrale est de comprendre en quoi des variables globales (biomasse totale, diversité, etc.) peuvent être comprises, via le modèle, comme le fruit des interactions entre individus, éventuellement localisés dans l'espace.

Dans les méthodologies qui peuvent répondre à ces questions on peut citer les modèles multi-agents, la théorie des jeux, l'évaluation contingente, les chaînes de Markov, les agents intelligents, etc.

3.7 Complexité calculatoire

La dimension paramétrique des modèles, les nombres de données, les contraintes à respecter sont des sources de problèmes algorithmiques qui peuvent être déterminants. Modéliser devient un compromis entre la pertinence et la faisabilité algorithmique. Le problème se retrouve dans de nombreuses situations :

- calcul du maximum de vraisemblance dans des "gros" modèles, spatialisés, bayésiens, associant variables continues et discrètes, etc. ;
- optimisation de décision dans des espaces de décision et d'état très vastes ;
- recherche de motifs semblables dans des séquences nucléotidiques ;
- nombre d'agents dans un système multi-agents.

La recherche d'algorithmes, efficaces et rapides, et d'approximations raisonnables et calculables est un thème de recherche actif dans les communautés mathématiques et informatiques. La méta-modélisation (au sens modèle du modèle) est une voie qui pourrait aussi offrir des solutions dans certaines situations.

L'INRA est assez bien placé dans la recherche en algorithmique qui est souvent stratégique dans la mise en œuvre des modèles. Ne perdons pas cette position.

La résolution d'EDO, EDP, EDP stochastiques, dans des systèmes où le nombre de variables est très grand et où elles peuvent être fortement non linéaires, pose des problèmes de résolution numérique et d'identification très difficiles qui réclament une expertise. Celle-ci se trouve le plus souvent en dehors de l'INRA : INRIA, CNRS, Universités. Il ne semble pas nécessaire pour l'instant d'acquérir des ressources propres au niveau de l'INRA.

3.8 Quelques verrous communs

Dans ces différents thèmes méthodologiques, on retrouve des verrous ou difficultés communs :

- Peur de commencer le processus de modélisation devant l'ampleur du projet et l'absence de certaines informations. Une organisation plus collective est un moyen de répondre à cette interrogation.
- Validation des modèles. La complexité des modèles rend

très difficile l'analyse du comportement de ces modèles. D'autre part, les données sont souvent chères, difficiles à acquérir, hétérogènes, etc.

- Développer des méthodes d'utilisation de connaissances a priori. Il existe des cadres conceptuels intéressants, comme le Bayésien, mais qui posent des problèmes algorithmiques difficiles et qui concernent surtout les données quantitatives. Mais il n'y a pas encore de cadre conceptuel satisfaisant pour utiliser conjointement des informations qualitatives et quantitatives.

- Utilisation et transposition de méthodes existantes, variées et synergiques. Il est très difficile de maîtriser l'ensemble des méthodes pertinentes pour modéliser un processus.

4. Une culture sur la modélisation à développer collectivement

4.1 Poser et formaliser le problème : un dialogue transdisciplinaire

La première difficulté en modélisation, et pour certains la plus grande, est de bien poser le problème. Cette phase est surtout le fait des "thématiciens" (biologistes, physiciens, économistes, sociologues, etc.). La question peut être disciplinaire, mais dans de nombreuses situations, le problème posé nécessite de faire travailler ensemble plusieurs de ces disciplines (biologie et économie par exemple) et le dialogue entre elles devient déterminant.

La phase suivante est celle de la formalisation, où se posent les problèmes du niveau approprié d'abstraction dans un modèle et d'échelle (spatiale, temporelle, niveau d'organisation) de représentation. Dans cette phase, les disciplines mathématiques et informatiques doivent intervenir si l'on veut mettre en face et analyser les objectifs, les moyens et les méthodes à mettre en œuvre. Le dialogue à ce niveau est encore insuffisant et difficile à entretenir, mais pourtant crucial.

Il existe des méthodologies d'aide à la formalisation, qui ne sont cependant pas encore pleinement satisfaisantes. Les méthodes conceptuelles en informatique facilitent cependant le dialogue entre disciplines et préparent l'implémentation en machine (ex. UML).

Cela demande du temps, des formations, et la reconnaissance institutionnelle. Il faudrait favoriser ces dialogues en les reconnaissant comme partie intégrante du processus de modélisation.

4.2 Acquérir une culture en modélisation

Le besoin de formation s'exprime dans tous les thèmes : en statistique, en bio-mathématique, mais aussi en développement informatique. C'est même un problème de culture à développer dans certains domaines disciplinaires.

Cette formation apparaît nécessaire pour acquérir les bases permettant de choisir la meilleure approche de modélisation face à un problème.

Elle est nécessaire aussi pour assurer la qualité des

développements informatiques. Cependant, sur ce point, la solution pourrait plutôt venir du recrutement d'informaticiens de développement professionnels.

4.3 Les forces en présence

La modélisation demande beaucoup de temps et de compétences. Les forces en présence dans certains domaines (épidémiologie par exemple) semblent insuffisantes. Si on souhaite développer la modélisation, il faudra s'en donner les moyens et faire en sorte que cette activité soit valorisée dans les carrières.

Les besoins sont aussi bien des besoins de chercheurs que des besoins d'ingénieurs, modélisateurs et/ou informaticiens. Il faut pouvoir préciser dans chacun des domaines où des besoins sont identifiés, si l'activité de modélisation est bien une activité de recherche validable dans la discipline en question, ou bien si il s'agit plutôt d'une activité d'ingénierie.

4.4 Mise en réseau des compétences, capitalisation

Il est déterminant de pouvoir capitaliser le travail, en général considérable, de construction d'un modèle afin de pouvoir en réutiliser des éléments. Le développement et l'utilisation de plate-forme ou d'environnement générique de modélisation ou la construction de bibliothèque de modèles sont des réponses possibles. Cette demande est déjà explicite dans certains domaines comme l'agro-alimentaire ou les agro-éco-systèmes par exemple.

On s'oriente de plus en plus vers des "gros" modèles issus de l'interconnexion de sous-modèles. Ceci demande de construire des réseaux pour rassembler des compétences complémentaires et, de disposer d'environnement de modélisation permettant l'interconnexion de ces modèles.

Des réseaux peuvent aussi être construits autour de problématiques génériques liées à des questionnements analogues dans différents domaines de modélisation. Les échanges porteraient alors sur l'architecture des modèles et plus seulement sur le contenu des modèles.

4.5 Les données

L'accessibilité des données est un problème fondamental dans tous les domaines. Elle peut se poser en terme de coût, de difficulté de mesure, de qualité ou de confidentialité. C'est aussi un problème de normes.

Dans certains domaines, les méthodes de filtrage ou de reconstruction de données, peuvent permettre de se dispenser de dispositifs d'acquisition très onéreux en les substituant par des mesures beaucoup plus abordables.

4.6. Calcul

Les difficultés liées au coût et au temps de calcul ne se posent pas aujourd'hui pour tous les thèmes. Cependant, la complexification des modèles rend de plus en plus nécessaire de pouvoir disposer de moyens de calcul importants en interne et en externe. L'augmentation de la vitesse des processeurs est toujours rattrapée par la quantité des calculs que l'on demande ! La distribution du

calcul sur plusieurs machines pourrait être une solution intéressante.

5. Comment s'organiser à l'INRA pour répondre aux défis de la modélisation ?

L'analyse développée dans les chapitres précédents montre bien que les grandes questions qui se posent dans la modélisation sont partagées par les différentes thématiques dans lesquelles elle est mise en œuvre et donc par les différents départements de l'INRA. D'autre part, les compétences à mobiliser sont réparties dans les différents départements. En conséquence, une réflexion collective sur l'organisation de cette activité à l'échelle de l'INRA est nécessaire si l'on veut capitaliser les avancées à cette échelle, développer harmonieusement des projets transversaux impliquant de la modélisation et réfléchir à l'articulation optimale des compétences entre départements.

5.1 Les objectifs

La question est alors de savoir de quels acteurs et sur quelles compétences en particulier, nous avons besoin pour répondre aux problématiques de l'INRA, comment les recruter, les former et en maintenir la compétence, comment valoriser les compétences existantes ou bien comment la mobiliser à l'extérieur de l'INRA et enfin comment en organiser l'activité collective afin de :

- Construire des projets de modélisation qui puissent réunir l'ensemble des compétences nécessaires pour un objectif particulier. Il peut s'agir aussi bien des différents types d'acteurs identifiés ci-dessous, mais aussi de plusieurs domaines thématiques ou disciplines concourant à la connaissance d'un processus : différentes échelles biologiques, aspects socio-économiques interagissant avec les aspects biologiques, etc.
- Favoriser et faciliter le développement des très nombreux « petits projets » de modélisation qui réclament du soutien sur le plan méthodologique.
- Développer les méthodologies de modélisation nécessaires dans chacun des domaines thématiques, ou pour les projets plurithématiques, en maintenant l'équilibre entre les méthodes génériques et spécifiques.
- Capitaliser ces développements entre les projets et entre les différents domaines thématiques. En effet, notre analyse des verrous méthodologiques de la modélisation montre que de nombreux verrous sont communs aux différents domaines.

Bien entendu, il existe déjà de nombreuses compétences à l'INRA et dans certaines situations, des coordinations relativement efficaces. L'analyse proposée ici place cet existant dans une perspective plus vaste et plus collective.

5.2 Les acteurs

Différents types d'acteurs peuvent ou doivent être mobilisés dans des projets de modélisation. Ils assurent dans leur ensemble un continuum entre des cultures différentes, qui doit permettre le dialogue et la collaboration. La typologie

proposée est forcément simplificatrice. Elle est apparue néanmoins utile par exemple pour préciser les attentes en matière de formation, ou bien analyser les problèmes d'environnement scientifique de ces acteurs. Elle permet ainsi aux départements exprimant un besoin en matière de modélisation, de mieux définir de quels acteurs ils ont besoin, et quelles sont les contraintes (environnement scientifique, formation, recrutement) associées. Dans cette typologie, le « thématicien » représente un chercheur spécialiste d'une discipline agronomique, biologique ou économique au sens large, dans laquelle la modélisation est un outil d'investigation plus ou moins développé. D'autre part, les termes « méthodologies » ou « méthodologistes » ne sont employés qu'en référence à l'activité de modélisation.

1. Tout d'abord le « thématicien non modélisateur ». Il est porteur d'une problématique pour laquelle l'utilisation d'un modèle apporterait un élément décisif, mais il ne sait pas utiliser les méthodes de la modélisation et/ou il n'est pas sensible à la modélisation.
2. Le « thématicien compétent en modélisation », par exemple un biologiste ou un agronome, spécialiste du processus (ou d'un aspect du processus) à modéliser et qui connaît et sait utiliser des méthodes de modélisation sans prétendre les maîtriser totalement.
3. Le « thématicien méthodologiste » développe des méthodes génériques de modélisation dans un domaine thématique particulier. Il peut s'agir par exemple d'un épidémiologiste, d'un écologue quantitatif ou encore d'un généticien quantitatif. Les méthodes qu'il développe sont souvent dans le champ des mathématiques appliquées ou de l'informatique. On parle souvent de double compétence.
4. Le « mathématicien appliqué » développe des méthodes génériques en mathématiques ou en informatique (on pourrait parler d'un « mathformaticien ») pour la modélisation, même si les projets qui sont à l'origine de ces développements, s'inscrivent dans un domaine thématique particulier (prototype : la plupart des chercheurs de MIA). Il est le plus souvent spécialisé dans un domaine de recherche particulier des mathématiques ou de l'informatique.
5. L'« ingénieur en méthodologie de la modélisation » maîtrise ces méthodologies dans un domaine thématique particulier ou en général.
6. L'« ingénieur en développement informatique » maîtrise les outils de développement informatique. On peut distinguer ceux qui exercent cette activité au niveau IE et IR.

Globalement, le terme « thématicien » désignera les catégories 1, 2 et 3, et le terme « méthodologiste » les catégories 3, 4, et 5.

5.3 Les projets

Il s'agit de répondre à une question de biologie, agronomie, économie, etc., pour laquelle la construction d'un modèle est un élément du projet, ou bien est au cœur du projet, mais cette construction n'est pas une fin en soi. Certains de ces projets sont de « gros projets » qui font appel à des disciplines variées associant les aspects biologiques et socio-économiques, et dans lesquels sont fréquemment associés des méthodologistes. Beaucoup de projets sont en

fait des « petits projets » portés par un thématique plus ou moins autonome en modélisation, même si ces petits projets peuvent être fédérés. Enfin, toutes les dimensions intermédiaires existent.

Les « petits projets »

La première distinction porte sur l'autonomie du thématique porteur de projet.

- Dans les thématiques où la modélisation est depuis assez longtemps incontournable dans l'activité de recherche (beaucoup de thématiques de EA par exemple), la majorité des équipes a acquis une culture de la modélisation. Les thématiques porteurs de projet sont relativement autonomes et insérés dans un environnement favorable aux échanges que ce soit dans leur équipe, ou dans des réseaux thématiques (réseau modélisation du département EA par exemple). Ces thématiques, individuellement, ou au travers ces réseaux, sont à même d'identifier des questions méthodologiques non résolues et ce sont de bons partenaires pour construire des collaborations efficaces avec les méthodologues sur ces questions.

- Dans les thématiques où l'importance de la modélisation est plus récente ou réactualisée (l'échelle du génome par exemple ; c'était peut être aussi le cas de l'épidémiologie il y a quelques années et des réponses sous forme de réseaux et de co-encadrements de thèses ont été apportées), certaines équipes ne sont pas autonomes en matière de modélisation. La solution est à trouver, soit au travers de la formation (voir le paragraphe « *La formation* ») qui doit être envisagée collectivement sur une thématique de manière à éviter l'isolement des thématiques formés, soit à travers l'apport de compétences en matière de modélisation.

Pour ces projets, l'apport de compétences peut se faire à travers un ingénieur en méthodologie de la modélisation, aidé le cas échéant d'un expert méthodologue (type 3 ou 4). En effet, il s'agit souvent d'articuler différentes méthodes et de les adapter. Cependant, cette solution peut difficilement s'envisager à long terme, car il faut éviter que les thématiques de type 1 se reposent sur ces ingénieurs et ne fassent pas l'effort d'acquiescer une culture en modélisation leur permettant de passer en type 2. La modélisation doit faire partie de la panoplie de nombreux thématiques de même que la biologie moléculaire pour certains d'entre eux. D'autre part, des problèmes de pérennité des compétences peuvent se poser si un tel ingénieur est isolé dans une unité de thématiques non modélisateur (voir « *Positionnement des méthodologues* »).

Les « gros projets »

Ce type de projet comporte par nature le rassemblement de compétences complémentaires. Il s'agit dans de nombreux cas de réponses collectives à des appels d'offre. On peut donc y impliquer des méthodologues de la modélisation émanant de différentes disciplines. C'est ainsi un excellent moyen de confronter et de faire communiquer des méthodologies différentes.

Les aspects ingénierie et en particulier informatiques (connexion entre langages, SIG, etc.) constituent souvent un des problèmes essentiels dans ces projets. Au-delà de l'ingénierie, ces projets posent aussi des problèmes

methodologiques particuliers liés à leur structure (identification de paramètres, par exemple).

Toute taille de projets

Certains projets, choisis pour leur importance stratégique pour le domaine thématique, ou pour le caractère générique des développements méthodologiques nécessaires, feront l'objet de partenariats entre thématiques et méthodologues. C'est le moyen d'ancrer les recherches méthodologiques sur les grands enjeux de la modélisation. Si plusieurs projets nécessitent des développements méthodologiques voisins, ils gagneront à être coordonnés sur ce plan. Un moyen de construire ces partenariats et cette coordination méthodologique est de l'identifier explicitement dans les appels d'offre thématiques.

Certains projets nécessitent des développements informatiques importants surtout quand ils sont destinés à être diffusés et conduire à des produits informatiques utilisables par des non chercheurs. C'est, par exemple, souvent l'aspect prépondérant dans les « gros projets » de modélisation à l'échelle européenne, mais c'est vrai aussi de nombreux « petits projets ». Toutes les équipes développant des modèles ne peuvent avoir une équipe d'informaticiens mobilisables sur ce type de développements professionnels. Une solution qui a fait ses preuves est de créer des plateformes de développement logiciel du type de celle réunissant l'INRA et les instituts techniques (voir « *Informatique et plate forme* ».)

5.4 Les différents méthodologues : positionnement scientifique et géographique

Depuis quelques années, on observe que de nombreux départements recrutent des thématiques méthodologues. Il est important d'en mesurer les conséquences en particulier en matière d'environnement scientifique nécessaire à la pérennisation des compétences.

Positionnement scientifique des méthodologues

Les trois types de méthodologues sont nécessaires dans le dispositif d'ensemble. Les thématiques méthodologues maîtrisent le domaine thématique et sont au cœur des développements méthodologiques intéressants leur domaine thématique. Les mathématiciens appliqués permettent d'assurer la capitalisation des développements méthodologiques et la diffusion entre les différentes thématiques. Leur positionnement un peu en amont leur permet de mobiliser des méthodes mathématiques ou informatiques plus pointues. Ils assurent de la souplesse au dispositif en pouvant investir relativement rapidement un domaine thématique dont les aspects méthodologiques ne sont pas encore stabilisés (les pistes méthodologiques sont nombreuses, et l'on ne sait pas encore celles qui seront efficaces et déterminantes pour répondre aux enjeux du domaine). Ce regard un peu extérieur leur permet de transposer plus facilement une méthodologie développée dans une thématique, dans une autre thématique. Par contre, la pertinence thématique de leur activité est plus difficile à garantir. Quant aux ingénieurs en méthodologie de la modélisation, ils se positionnent sur l'adaptation et l'usage combiné de différentes méthodologies.

Typiquement, même si ce n'est pas une règle absolue, les mathématiciens appliqués sont aujourd'hui dans le département MIA et les thématiciens méthodologistes dans les départements thématiques. Pour certains d'entre eux, c'est d'ailleurs plus leur « appartenance » à un département qui définit leur positionnement scientifique que leur formation et leur choix scientifique. Cependant, il n'existe pas dans tous les cas de communauté scientifique de méthodologistes dans le domaine thématique concerné. Par ailleurs, il y a à l'INRA très peu d'ingénieurs en méthodologie de la modélisation.

Les thématiciens méthodologistes sont encore relativement rares (par rapport au nombre de thématiciens) actuellement à l'INRA et il paraît difficile de parier sur une croissance très importante de leur nombre. En effet, peu d'étudiants ont réellement la double compétence nécessaire. Cependant, des écoles doctorales « double compétence » se développent. De plus, des ASC sont actuellement en cours de formation avec l'objectif d'acquérir une double compétence en bioinformatique, épidémiologie et en conduite de systèmes biotechniques. Ce type de construction de double compétence pourrait être envisagé pour d'autres domaines.

Positionnement géographique des méthodologistes

Les méthodologistes doivent être d'une part proches des thématiciens pour ne pas s'écarter des problématiques des domaines thématiques, mais aussi proches les uns des autres pour ne pas perdre leur compétence méthodologique. Cette double exigence, parfois contradictoire, est au cœur du problème de positionnement des méthodologistes. Plusieurs pistes sont proposées ici pour répondre à cette contradiction, qui visent surtout à éviter l'isolement des méthodologistes en construisant les liens entre méthodologistes et thématiciens autour de projets. La diversité des solutions présentées permet de donner de la souplesse au dispositif afin de s'adapter à la variété des situations rencontrées.

Lorsqu'il est possible de construire une équipe de taille assez importante de thématiciens méthodologistes, au cœur d'une unité de thématiciens, le système est efficace (c'est par exemple, l'unité GA de Jouy-en-Josas). Notons que ce type d'équipe a souvent vocation à développer des méthodologies pour l'ensemble des thématiciens de son champ sans forcément privilégier les projets développés dans la même unité.

Lorsque ce n'est pas possible, un moyen intéressant est d'associer dans une même unité des thématiciens méthodologistes et des mathématiciens appliqués, à proximité d'unités de thématiciens. C'est par exemple ce qui se fait en Avignon, où un écologue méthodologiste du département EFPA est positionné dans une unité MIA, à proximité d'une unité de thématiciens EFPA. L'« appartenance » de cet écologue méthodologiste au département EFPA assure une garantie sur la pertinence thématique et son positionnement dans une unité MIA apporte un environnement fécond et indispensable pour ses recherches méthodologiques. Ce « modèle » pourrait être développé quand les contraintes géographiques le

permettent. Il offre de plus l'intérêt d'associer les mathématiciens appliqués à des thématiques. Il faut rester cependant vigilant car il demande des qualités d'ouverture au thématicien méthodologiste qui ne sont pas si fréquentes et il ne faudrait pas déstabiliser le dispositif en diminuant la proportion de mathématiciens appliqués.

Une variante intéressante, mais difficile à maintenir dans la durée, est celle où le thématicien méthodologiste effectue des séjours longs et alternés entre l'unité de mathématiciens appliqués et celle de thématiciens (ce qui se fait de plus en plus avec les doctorants en modélisation).

Un autre dispositif intéressant est l'unité formée à part égale de thématiciens méthodologistes et mathématiciens appliqués, autour d'une même thématique. C'est le cas de l'unité MIG qui réunit des bio-informaticiens de MIA, de PHASE et de MICA. Pour pérenniser ce type de structure il faut que les liens avec les thématiciens des autres unités soient forts et entretenus, par exemple en participant aux séminaires organisés par les thématiciens.

Ces différents dispositifs ne mettent pas en cause l'existence d'unités composées essentiellement de mathématiciens appliqués, garantes du développement des recherches méthodologiques sur les grands thèmes identifiés au chapitre précédent.

Une alternative est de positionner des thématiciens méthodologistes dans des équipes externes à l'INRA pour en assurer l'environnement méthodologique. Le problème est alors d'assurer le lien avec les thématiciens. Cela peut être cependant intéressant si l'on veut acquérir une compétence méthodologique particulière. Un tel dispositif doit sans doute être limité dans le temps et être évalué périodiquement.

Bien entendu, les équipes de modélisateurs thématiciens aimeraient avoir des méthodologistes dans leurs équipes. C'est un moyen de développer l'activité de modélisation dans une équipe, et d'autre part cela permet de jouer un rôle de relais pour amplifier les travaux des méthodologistes. Il faut essayer de l'éviter avec un jeune chercheur méthodologiste. C'est envisageable avec un senior ou dans un dispositif en alternance sur un temps limité. C'est plus facilement envisageable avec des ingénieurs en méthodologie de la modélisation, mais là aussi l'entretien des compétences nécessite d'éviter l'isolement géographique. La présence de plusieurs ingénieurs sur un même centre est alors nécessaire, ainsi qu'un dispositif d'animation intra centre coordonné au niveau national.

Le département MIA

Les objectifs scientifiques du département MIA sont de répondre aux grands verrous de nature mathématique ou informatique, identifiés dans les chapitres précédents. Pour la plupart de ces verrous, un partenariat étroit avec des thématiciens est nécessaire pour assurer la pertinence des développements et il est plus facile à construire avec des thématiciens méthodologistes. Le niveau d'investissement des chercheurs de MIA dans chacune des thématiques dépend de la thématique, et en particulier du niveau de stabilisation des méthodologies. Par exemple, le type de méthodologie à développer à l'échelle du génome n'est pas du tout stabilisé, et MIA a choisi d'investir fortement en bio-

informatique pour y répondre. D'autre part, les partenariats avec des mathématiciens et informaticiens académiques sont la garantie de la qualité des développements

méthodologiques. Le schéma directeur MIA décrit précisément ces choix. Quelques éléments sont donnés dans l'encadré suivant.

Nous proposons ici quelques éléments d'illustration des moyens à mettre en œuvre pour répondre aux besoins identifiés précédemment dans le rapport, structurés à partir des grands thèmes méthodologiques. Il s'agit donc plutôt d'illustrations « MIA-centrés ».

- Recherche d'une structure dans une grande quantité de données. Cette problématique est fondamentale pour différents thèmes et en particulier l'analyse des données à l'échelle du génome. Le rôle des ingénieurs informaticiens y est fondamental et il faut construire des liens très étroits entre ces informaticiens et les chercheurs thématiques et méthodologistes. Sur le plan de la recherche méthodologique, il faut s'inscrire dans le data mining (ou fouille de données), qui est une problématique à la frontière des statistiques et de l'intelligence artificielle qui vise à générer des hypothèses à partir de la masse de données, qui sont ensuite soumises à un processus plus conventionnel d'analyse. Pour l'instant la problématique concerne plutôt les thématiques méthodologistes, à même de porter le questionnement thématique. Peu de chercheurs INRA sont engagés dans ce thème et il faut créer une structure d'échanges entre eux. Pour MIA, il s'agit d'analyser s'il y a une problématique générique méthodologique derrière, mal prise en compte à l'extérieur de l'INRA, légitimant un investissement.
- Représentation et exploitation d'informations dynamiques et spatialisées. Les aspects informatiques sont ici aussi déterminants, et un enjeu est de mieux coordonner les informaticiens et méthodologistes. Sur le plan méthodologique, MIA est compétent sur les aspects spatiaux et cherche à compléter cette compétence sur les aspects dynamiques.
- Analyse, prévision et gestion des risques. Concernant l'analyse des risques alimentaires, une unité spécifique pluridisciplinaire a été créée (MIA et AlimH) il y a un an qui vise donc à rapprocher thématiques et méthodologistes dans une même unité. Globalement, il faut intensifier les liens avec les méthodologistes travaillant dans des domaines thématiques différents de ceux de l'INRA et ayant déjà une expérience importante du domaine (risques nucléaires par exemple).
- Assemblage, intégration, réutilisation. De nombreux développements informatiques sont faits de manière indépendante. Il y a un enjeu fort en matière de conception informatique qui nécessiterait la mobilisation de compétence en la matière. Sur le plan méthodologique, il existe une communauté de chercheurs en informatique à l'extérieur de l'INRA (Lille 1, réseau CNRS VERSIM ou OpenMI au niveau européen) dont il faudrait se rapprocher.
- Systèmes dynamiques complexes ou compliqués. C'est un enjeu fondamental pour de nombreuses thématiques et il y a relativement peu de spécialistes à l'INRA des systèmes dynamiques. Il faut mettre en place un ou plusieurs programmes ambitieux (par exemple sur l'exploration numérique des propriétés des modèles) pour développer des recherches méthodologiques, diffuser les résultats des travaux, et faciliter la mise en œuvre des méthodes. Un tel programme peut s'appuyer sur les compétences en statistique à l'INRA (et en particulier MIA), mais nécessite le renforcement des compétences en analyse de systèmes et la mobilisation de partenaires extérieurs à l'INRA sur ces aspects.
- Modélisation du comportement individuel et collectif. Ces modèles reposent le plus souvent sur le concept des systèmes multi agents. Ils posent des problèmes de logiciels d'aide à la construction de ces modèles, et d'analyse mathématique de ces modèles. Il faut à la fois renforcer les compétences en ingénierie informatique et s'impliquer dans les réseaux d'utilisateurs de ces modèles existant essentiellement en dehors de l'INRA. Sur le plan de la recherche méthodologique, l'INRA peut difficilement prendre aujourd'hui le leadership de cette thématique mais possède des compétences pour y participer utilement.
- Complexité calculatoire. Les modèles impliquent un nombre toujours croissant de paramètres ou variables, conduisant à des véritables verrous algorithmiques que la vitesse croissante des ordinateurs ne peut ouvrir. C'est un thème fort pour le département MIA qu'il tient à poursuivre. Sur le plan du calcul numérique, la mobilisation d'experts, le plus souvent extérieurs à l'INRA, semble une réponse suffisante.

5.5 Réseaux et capitalisation

La construction de réseaux répond à différents objectifs. Ils permettent d'éviter l'isolement, favorisent la construction de projets entre acteurs différents (en particulier méthodologistes et thématiciens) autour d'un objectif commun ainsi que la capitalisation méthodologique. Ils permettent aussi d'augmenter collectivement la culture d'un ou plusieurs départements.

Concernant la modélisation, on peut envisager des réseaux de différents types : le réseau entre thématiciens pour partager les expériences en matière de modélisation, le réseau entre méthodologistes et thématiciens dans le cadre d'une thématique « finalisée » particulière (épidémiologie par exemple), le réseau autour d'une problématique méthodologique générique (comme celles identifiées au chapitre précédent : changement d'échelle par exemple). La capitalisation sur les thèmes méthodologiques génériques pourra se faire aussi par l'organisation de journées méthodologiques au niveau national.

Au-delà de la capitalisation méthodologique, les réseaux méthodologiques sont un lieu d'accueil naturel pour les thématiciens. Ils peuvent accueillir de temps en temps lors de réunions des thématiciens afin d'être confrontés à des problèmes de modélisation, d'offrir une expertise collective à ces thématiciens plutôt que des consultations individuelles, et d'infléchir leurs propres recherches méthodologiques autour des nouveaux enjeux identifiés.

La multiplication des réseaux ne résoudra pas automatiquement tous les problèmes. Il faut être attentif aux structures existantes et éviter de créer des doublons, que ce soit à l'intérieur ou à l'extérieur de l'INRA. Il existe par exemple des réseaux nationaux autour de certaines thématiques de la bioinformatique qui ne nécessitent pas d'être reproduits à l'intérieur de l'INRA.

D'autre part, il faut y mettre quelques moyens financiers pour en favoriser le fonctionnement et exprimer clairement la volonté de la part de la hiérarchie. Une lettre de mission et l'identification claire d'un responsable sont nécessaires. Un site Web peut être un excellent moyen pour mettre à disposition les informations échangées dans ces réseaux.

Le département MIA est légitime pour susciter et structurer, en concertation avec les autres départements, des réseaux sur des thèmes méthodologiques génériques concernant les mathématiques et l'informatique. Sur les autres thèmes (comme l'épidémiologie par exemple), c'est plutôt de la responsabilité des départements thématiques en concertation avec MIA. Un bureau multi-département pourrait assurer une coordination du dispositif d'ensemble afin d'en assurer la cohérence.

De nombreuses avancées méthodologiques sont obtenues dans le cadre de projets liés à des thématiques particulières. Les mathématiciens appliqués sont alors en charge, quand cela est pertinent, de donner un caractère générique à ces résultats pour en assurer la capitalisation au-delà de la thématique.

L'effort important de construction d'un modèle doit être mieux exploité. Les modèles doivent être construits avec un

souci de réutilisabilité et de communication avec les autres modèles, surtout dans la perspective d'agrégation de modèles pour construire des modèles multi-échelles ou multi-disciplinaires. Les plates-formes de modélisation telles que CAPSIS fournissent un cadre pertinent dans ce but, et, comme c'est le cas pour CAPSIS, les objectifs et les spécificités de ces plates-formes doivent être clairement explicités (voir 5.8 Informatique et Plate-forme).

5.6 Formation et information

Un nombre relativement important de thématiciens doit être formé pour devenir thématicien modélisateur. Même si des ingénieurs « modélisateurs » dans les unités peuvent aider les thématiciens à maîtriser les méthodes, il n'est pas raisonnable de penser qu'ils peuvent éviter aux thématiciens d'acquérir une culture en modélisation. La modélisation doit faire partie de la panoplie de nombreux thématiciens.

Cette formation doit être construite par les thématiciens méthodologistes du domaine, par les mathématiciens appliqués et par les ingénieurs. Le système « école chercheur », quand il ne se réduit pas à un mini colloque, est adapté à ce type de formation, puisqu'il permet de construire des formations associant l'ensemble des aspects en créant des liens entre les différents acteurs au-delà de la formation et amorçant la construction d'un réseau.

Ce système « école chercheur » pourrait être complété par des formations plus conséquentes adaptées au thématicien « débutant », prenant éventuellement exemple sur le dispositif FpStat développé pour les statistiques et pouvant s'appuyer sur des formations externes (en particulier pour les aspects informatiques). Ces différents dispositifs devraient être intégrés dans une structure de coordination plus vaste d'identification des besoins, des possibilités d'autoformation, des offres externes de formation.

Plus généralement, l'information sur les compétences en matière de modélisation existantes à l'INRA et en dehors de l'INRA doit être assurée et régulièrement mise à jour.

Il est important de plus, de se préoccuper de la formation initiale des thématiciens, en faisant remonter auprès des responsables d'école doctorale les besoins de l'INRA et en participant aux formations. Symétriquement, le problème se pose pour certaines formations de mathématiciens appliqués, peu préoccupées par les problèmes finalisés. Il faut donc globalement s'impliquer plus dans les écoles doctorales si l'on veut peser sur leurs orientations.

5.7 Evaluation des acteurs

Le processus de développement d'un modèle est long avant d'aboutir à une production scientifique valorisable. De plus il n'est pas toujours facile de publier un travail relatif à un modèle. Si le modèle a permis de faire avancer une question scientifique, ou bien s'il nécessitait un développement méthodologique original, le travail est publiable respectivement dans une revue « thématique » ou dans une revue de mathématiques ou informatique appliquées. Mais un modèle pour « l'action » est plus difficile à valoriser. Il faut donc prendre en compte ces difficultés dans l'évaluation des acteurs de la modélisation.

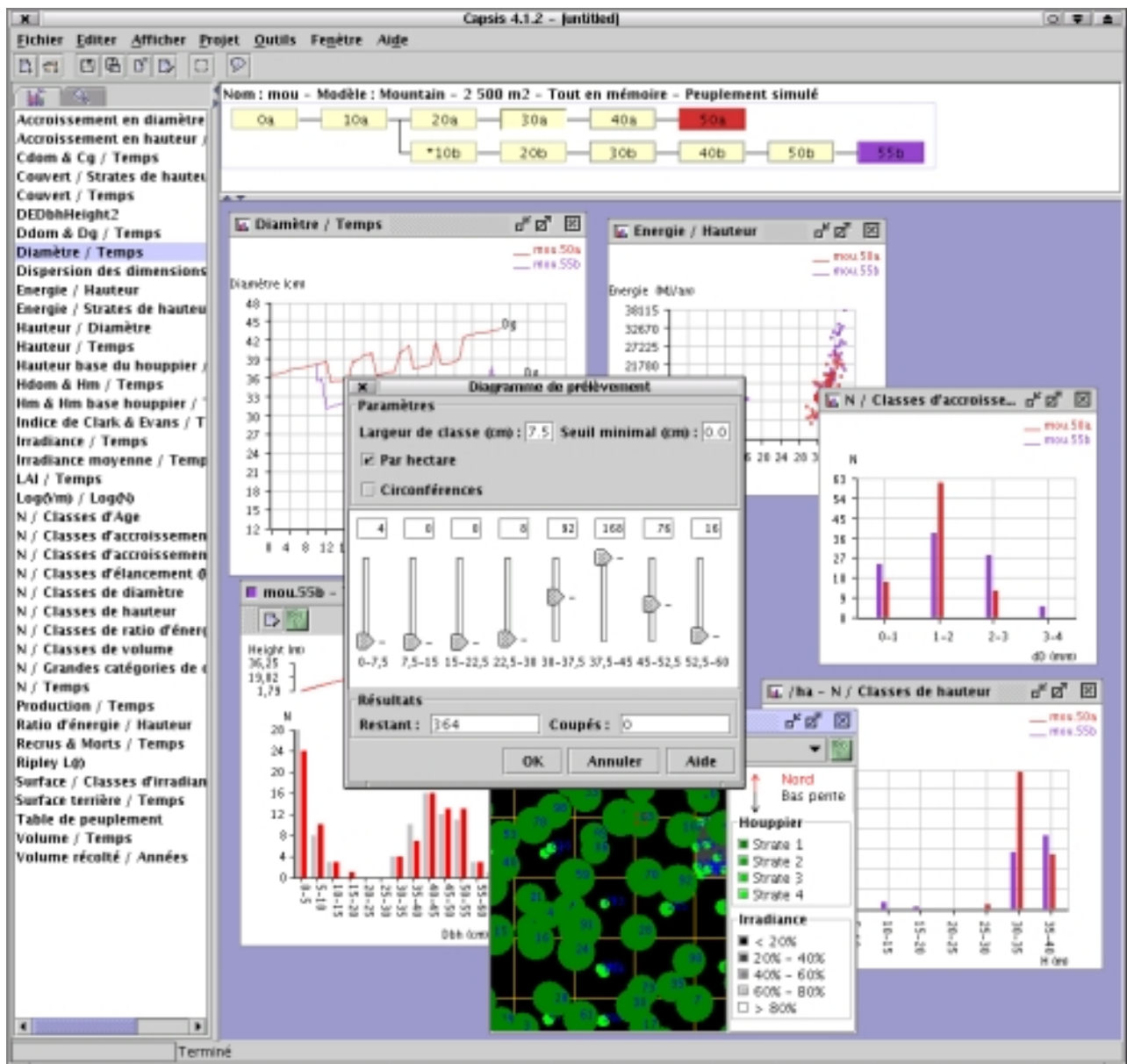


Fig. 4 - Copie d'écran du logiciel CAPSIS, F. de Coligny, <http://coligny.free.fr/>

L'effort de formation proposé ci-dessus demandera un investissement de la part de nombreux acteurs déjà en place. Il faut que cette mission soit explicite et valorisable.

5.8 Informatique et plate-forme

Le besoin en ingénierie informatique est important si l'on veut développer les logiciels qui mettent en œuvre les modèles que ce soit pour un modèle de connaissances ou plus encore pour les modèles destinés à servir à la décision et à l'action et donc à être diffusés à l'extérieur de l'INRA.

Pour cela il est nécessaire de pouvoir mobiliser des IE de développement et des IR pour assurer la compétence pointue en matière de génie logiciel. Il est fréquent et naturel de positionner les IE dans les équipes de thématiques afin d'assurer un lien fort avec les thématiques responsables

des projets. Cependant, ce positionnement peut poser des problèmes d'isolement pour ces informaticiens. Différents types de réseaux d'informaticiens doivent exister (et il en existe déjà) pour maintenir la compétence de ces informaticiens : au niveau des centres s'il y a suffisamment d'informaticiens, au niveau des départements, et au niveau de l'INRA, avec un rôle d'animation à préciser pour MIA et la DISI.

Sous sa forme « conceptuelle » le terme « plate-forme » désigne un cadre de développement informatique qui favorise la capitalisation, la réutilisabilité des modules, et surtout leur connectivité (compatibilité des entrées et sorties des modèles). Il s'agit ici de CAPSIS (Fig. 4) ou STICS. D'autre part, le terme désigne aussi un lieu physique où sont rassemblées des forces plus ou moins importantes de

développement (plate-forme INRA-ICTA à Toulouse) et le cas échéant des modélisateurs. La plate-forme APSIM en Australie conjugue à la fois les deux aspects. Globalement, il faut encourager toutes les formes de coordination entre les développements qui visent à améliorer l'efficacité du système (meilleure utilisation des forces de développement,

réutilisabilité des modules, connectivité, qualité des développements, etc.). A un niveau plus ambitieux, on peut parler de plate-forme pour désigner un projet fortement structurant réunissant thématiciens, méthodologistes et informaticiens. Ce type de structure est analysé en détail pour un objectif agronomique dans l'encadré ci-dessous.

Quelle(s) plate(s)-forme(s) informatique(s) pour les modèles de culture ?

Contexte

Un modèle de culture est un modèle mathématique qui décrit le fonctionnement dynamique du système sol-plante en interaction avec l'atmosphère et la gestion de la parcelle.

Ces modèles sont devenus des outils incontournables pour plusieurs types d'étude. A l'échelle de la parcelle, les modèles de culture sont utiles pour aider à l'analyse de résultats expérimentaux, pour prédire le comportement du système ou pour simuler différentes options de gestion. Ils sont utilisés à l'échelle d'une région ou d'un pays pour fournir des prédictions de rendement et à l'échelle d'un bassin versant pour étudier l'impact de l'agriculture sur la quantité et qualité de l'eau. Les modèles de culture sont également des outils de recherche au niveau des processus individuels, permettant de replacer ces processus dans le contexte du système global.

Les modèles de culture sont donc importants dans plusieurs domaines, mais développer un modèle de culture est un travail très lourd. Il est alors important de réfléchir sur les possibilités de mieux profiter d'une mise en commun des efforts dans ce domaine. La proposition ici est de créer une nouvelle plate-forme informatique pour modèles de culture, dont le développement et l'utilisation seraient très largement partagés.

Plusieurs plates-formes pour modèles de culture existent déjà. Il est important de profiter de ces expériences. A l'INRA, STICS est une plate-forme pour modèles de culture, et la proposition ici pourrait être vue comme une nouvelle génération de cette plate-forme. APSIM, la plate-forme d'APSRU (Agricultural Production Systems Research Unit) en Australie, est sans doute la plate-forme la plus aboutie pour modèles de culture en termes d'organisation des compétences et en termes de structure informatique. DSSAT a été développée par un consortium de groupes de recherche américains. L'objectif est de fournir une gamme de simulateurs de différentes cultures qui utilisent les mêmes données d'entrée et qui partagent certains éléments de programmation. SEAMLESS est un projet européen en cours qui doit fournir une plate-forme pour modèles de culture dans 4 ans. La plate-forme CAPSIS, développée à l'INRA, pourrait également servir de modèle, même si cette plate-forme concerne des peuplements d'arbres.

Elements d'un cahier des charges pour une plate-forme pour modèles de culture

La première activité d'un projet pour créer une nouvelle plate-forme pour modèles de culture serait d'identifier les utilisations et utilisateurs, et c'est en fonction de cela que l'on pourrait dresser un cahier des charges détaillé. On peut néanmoins prévoir des éléments de ce cahier des charges.

Flexibilité d'utilisation. La diversité des questions à étudier est grande. Différentes cultures sont concernées, et dans chaque étude il y a des conditions pédoclimatiques spécifiques, une interrogation unique et une disponibilité de données particulière. La plate-forme doit être suffisamment flexible pour faire face à cette diversité.

Identification des éléments communs. Il faudrait identifier les éléments communs entre différentes utilisations de la plate-forme, qui deviendront des modules partagés.

Qualité des modèles. L'utilité de la plate-forme dépendra fortement de la qualité des prédictions et préconisations basées sur les modèles. Cette qualité dépendra en grande partie de la qualité des modèles pour les processus sous-jacents. Il est alors essentiel de s'assurer que la plate-forme bénéficie en continu des résultats de la recherche sur les processus, ce qui demande une collaboration forte et durable entre modélisateurs de processus et la plate-forme.

Evaluation. Il est essentiel de pouvoir évaluer les modèles de culture. Cela implique que la plate-forme devrait permettre de calculer facilement des critères quantitatifs de distance entre valeurs prédites et valeurs observées. Il faudrait par ailleurs un système de documentation pour que chacun puisse consulter les résultats d'évaluations précédentes. Pour que ce soit utile, il faudrait une certaine stabilité des modèles, aussi bien par rapport aux équations que par rapport aux valeurs des paramètres.

Analyse statistique. La plate-forme devrait comprendre des outils statistiques, notamment pour l'évaluation des modèles de culture, pour l'analyse de sensibilité et pour l'estimation des paramètres d'un modèle de culture. Des outils pour l'assimilation de données pourraient également être très utiles.

Qualité informatique. L'utilité de la plate-forme dépendra pour une grande partie de sa qualité informatique. Il faudrait que la structure informatique rende la plate-forme facile à créer, facile à utiliser et facile à faire évoluer.

Disponibilité de données. L'utilité de la plate-forme dépendra en partie de la disponibilité de données (météo, sol, essais agronomiques). Il faudrait donc faire un effort particulier pour construire les bases de données nécessaires et pour les rendre disponibles aux utilisateurs de la plate-forme.

Mécanismes d'évolution. Une plate-forme doit évoluer et pour cela il faut prévoir des mécanismes d'évolution. APSIM fournit un exemple. Chaque processus ou culture a un responsable. C'est lui qui décide des changements, après discussion avec utilisateurs et informaticiens.

Conclusions

Le cahier des charges, bien que préliminaire, montre clairement que la création d'une plate-forme pour modèles de culture demande une gamme importante d'expertises, allant de l'agronomie aux statistiques en passant par la physiologie végétale, la science du sol, la bioclimatologie et l'informatique. En fait, la qualité de la plate-forme dépendra en grande partie du succès dans la mise en place des collaborations entre toutes ces disciplines. Cela plaide en faveur d'une seule plate-forme pour modèles de culture pour éviter une dispersion des efforts. Avoir une seule plate-forme serait par ailleurs efficace. Cela permettrait de profiter au maximum du fait que différentes problématiques pourraient partager des éléments d'analyse, de modélisation ou de programmation.

Au-delà de son rôle d'outil, le fait d'avoir une plate-forme unique pourrait avoir un effet important sur l'organisation de la recherche, en augmentant et enrichissant les interactions entre chercheurs ou entre chercheurs et ingénieurs du développement.

Néanmoins, il ne faudrait sans doute pas exiger que toute nouvelle modélisation soit intégrable dans une même plate-forme. Il faudrait laisser la place à des innovations qui ne pourraient pas toutes nécessairement se faire dans ce cadre.

Il y a quelques années, un projet très ambitieux de création d'une plate-forme de modèles de culture très polyvalente aurait été un pari du point de vue informatique. Depuis, les méthodes et langages informatiques ont évolué dans le sens de permettre des projets plus complexes. La preuve en est les plates-formes très sophistiquées qui existent déjà. Un projet de plate-forme pour modèles de culture nouvelle génération est réaliste, à condition d'y affecter les ressources nécessaires.

Un projet pour créer une nouvelle plate-forme pour modèles de culture a été lancé en 2005 sous l'égide des départements EA et MIA. Le projet s'appelle RECORD (REnovation et COoRdination de la modélisation des cultures).

6. Conclusions et conséquences opérationnelles

Cette réflexion sur la « Modélisation à l'INRA » a permis de mesurer l'importance croissante de la modélisation dans de nombreux champs d'activité de l'INRA. Dans certains domaines (comme ce qui concerne l'environnement physique ou l'écologie), le rôle de la modélisation est relativement ancien et certains départements ont construit des structures pour permettre aux chercheurs de s'approprier les méthodes et concepts. Un enjeu est de capitaliser les développements au-delà du domaine et de faciliter l'accès aux méthodologies mathématiques et informatiques les plus performantes. Dans d'autres domaines, et en particulier en ce qui concerne le génome et la biologie intégrative, il semble que l'on soit face à un tournant et que la modélisation peut être amenée à jouer un rôle décisif dans le futur, et participer activement à la création de concepts nouveaux. Il faut alors créer les conditions permettant de répondre aux enjeux. L'analyse ci-dessus conduit à plusieurs propositions opérationnelles dans ce but.

6.1 S'appuyer sur les « grandes classes de problèmes » en modélisation identifiées dans le paragraphe 3

Les verrous méthodologiques, communs à plusieurs thématiques, concernent tous les départements de l'INRA et en particulier le département MIA. Le schéma directeur du département MIA propose de s'impliquer prioritairement dans plusieurs des grands thèmes méthodologiques identifiés dans le paragraphe 3. Cette implication se fait dans le cadre d'un domaine thématique pour lequel ces développements méthodologiques sont les plus susceptibles

d'être décisif. Un effort particulier concerne le thème 3.5 « Systèmes dynamiques complexes ou compliqués » qui est fondamental pour la modélisation et qui n'est pas suffisamment présent aujourd'hui dans les activités du

département MIA.

Cette implication se fera en bonne partie par une adaptation des compétences des chercheurs en place. Par exemple, la maîtrise des aspects statistiques qui est un des cœurs de compétence de MIA et qui pouvait apparaître a priori comme décalée vis-à-vis des priorités, s'affirme comme une composante essentielle dans de nombreux thèmes méthodologiques. Cependant, elle doit être complétée par d'autres compétences qu'il faudra acquérir par la formation interne, par le recrutement ou bien en impliquant des partenaires extérieurs (INRIA, CNRS, Universités) dans des projets communs.

Au-delà des verrous méthodologiques, les « grandes classes de problèmes » constituent des problématiques de modélisation communes à de nombreux départements de l'INRA et devraient susciter la création de réseaux et de projets communs.

6.2 Projets

Nous proposons de construire des projets dont la cohérence est assurée par une problématique de modélisation commune, concernant différents domaines d'application. Il peut s'agir par exemple des modèles et méthodes de gestion d'écosystèmes, appliqués à différents écosystèmes. Cependant, la construction de tels projets est difficile car elle inverse un peu le leadership sur les projets. Un moyen de mobiliser les thématiciens sur ce type de projet est de l'associer avec le développement et la mise à disposition de modules informatiques d'intérêt partagé, comme dans le projet « Exploration numérique des propriétés des modèles » en cours d'élaboration.

Les projets plus classiques, dans lesquels il s'agit de développer un modèle pour une problématique particulière, sont aussi importants dans le dispositif. Il faut cependant qu'il y ait échange entre les différents projets concernant des classes de problèmes voisines, afin de capitaliser les développements. La cellule peut jouer un rôle à ce niveau.

Le thème particulier de la modélisation des réseaux de régulation, fera l'objet d'une fédération de projets construits

autour de binômes mathématicien - biologiste. Il s'agit pour l'instant d'identifier ces binômes.

6.3 Plate-forme

Il faut démarrer le projet de plate-forme « Modèles de culture ». L'objectif est d'avoir un cahier des charges précis à la fin 2005 qu'il faudra faire valider par la cellule multi départementale, et de commencer les développements informatiques en 2006. Il est fondamental que la réflexion tienne compte de l'existant.

Au-delà de l'application particulière, la réflexion sur cette plate-forme doit servir de modèle pour la construction de futures plates-formes.

6.4 Construire un ensemble cohérent de réseaux

Les initiatives peuvent venir des CD, d'une cellule de coordination (voir ci-dessous) ou des chercheurs eux-mêmes. Cependant, une manifestation d'intérêt explicite de la part des CD (participation au financement par exemple) apparaît nécessaire. La cellule sera sollicitée pour apporter un label, et le cas échéant participer au financement. Elle analysera la cohérence avec les autres réseaux INRA et extérieurs.

L'objectif commun aux réseaux est le partage d'expérience autour d'une problématique. Cependant, il est important d'y ajouter un objectif de production qui peut revêtir des formes variées : construire une formation, organiser un colloque, rédiger un livre, définir des questions de recherche, construire un projet de recherche, etc.

Plusieurs réseaux méthodologiques sont en cours de construction : changement d'échelle ; modèles et décisions autour des systèmes biotechniques, par exemple. On peut imaginer que chacune des « grandes classes de problèmes » puisse être à l'origine d'un réseau, à échelle régionale, nationale voire européenne en fonction des situations et des opportunités. En parallèle, les réseaux plutôt thématiques (autour de l'épidémiologie, par exemple) sont aussi encouragés pour permettre le partage d'expérience entre thématiciens.

6.5 Construire un dispositif de formation

Il est nécessaire de construire un dispositif de formation à la modélisation impliquant l'ensemble des départements. Il pourra concerner différents niveaux. On peut envisager un mode de fonctionnement du type de FP Stat, c'est-à-dire des formations de formateurs qui permettent de multiplier les formateurs potentiels. On peut aussi réfléchir en concertation avec d'autres organismes (CNRS, INSERM, etc.) qui ont des besoins voisins.

Pour conclure, le « réseau » interdépartemental qui a construit ce rapport a montré dans son fonctionnement qu'il est possible de construire une coordination sur la modélisation à l'échelle de l'INRA.

6.6 Construire une cellule de coordination multidépartementale

Il apparaît nécessaire de construire une cellule réunissant des scientifiques compétents en modélisation, nommés par les chefs de département. Cette cellule serait essentiellement un lieu d'échanges, élaborant des avis consultatifs. On peut classer les missions possibles à confier à cette cellule en deux catégories. Une qui concerne l'animation, ou plus généralement la construction d'une culture de la modélisation, et une autre liée aux aspects d'organisation ou de gestion des ressources humaines. Ces missions, en particulier celles de la deuxième catégorie, ne pourraient être confiées à une telle cellule qu'avec un plein accord des départements.

Missions relatives à l'animation

- Poursuivre la veille prospective sur la modélisation à l'INRA ; identifier les compétences existantes.
- Assurer l'animation globale du dispositif, par exemple en organisant un colloque ouvert à l'extérieur, sur la modélisation ; assurer l'échange d'informations entre les départements sur leurs priorités et actions dans le domaine.
- Aider à la construction de plates-formes de modélisation ; par exemple définir un cahier des charges.
- Produire un programme de formation à la modélisation. C'est fondamental si l'on veut que les thématiciens puissent prendre en charge le développement des modèles.
- Assurer la cohérence des réseaux dans lesquels la modélisation joue un rôle important, et en particulier ceux relatifs aux aspects méthodologiques. Eventuellement, la cellule pourrait disposer d'un budget incitatif pour soutenir des réseaux.
- Susciter la rédaction d'articles ou d'ouvrages de synthèse.

Missions relatives à l'organisation et à la GRH

- Coordonner les demandes de recrutement en ingénieur de type 5 ou 6, de façon à doter l'INRA d'une capacité collective à s'impliquer dans les projets européens très gourmands en ingénierie.
- Analyser les demandes de profil de recrutement des thématiciens méthodologistes et mathématiciens appliqués. Cette analyse qui ne se substitue pas à celle des départements et du collège, permettrait l'échange inter département sur les priorités scientifiques et produirait un avis sur la cohérence globale de la demande sur les plans thématique, méthodologique et géographique.
- Favoriser les constructions de thèses co-encadrées par des thématiciens et des méthodologistes. Proposer et évaluer des sujets.
- Assurer le suivi des jeunes chercheurs ou ingénieurs, par exemple en proposant des tutorats.

Points de vue

Suite à la présentation de ce rapport devant différentes instances de l'INRA (Collège de Direction, Conseil Scientifique), des avis d'experts couvrant différentes disciplines concernées par la modélisation ont été sollicités afin d'apporter un regard extérieur sur une analyse effectuée essentiellement en interne.

Paul Bourguine

École Polytechnique, Unité associée au CNRS (UMR 7656)
Centre de recherche en épistémologie appliquée (CREA), Palaiseau

Le rapport sur 'la modélisation à l'INRA' est bien structuré et fournit une synthèse approfondie et fort intéressante de l'ensemble des activités de l'INRA, relatives à ce thème.

Il y a toutefois un point discutable sur le plan épistémologique dans l'introduction du rapport. Il porte sur la définition 'partagée' à l'INRA de la modélisation. Un modèle serait :

'une représentation « partisane » d'un processus, pour un objectif donné'.

Un modèle d'un processus vise à 'reconstruire' ce processus à partir des données et des connaissances relatives à la classe des processus considérés (paradigmes, lois de la physiques, principes théoriques, biologiques...). Cette reconstruction se fait à l'intérieur d'un formalisme continu (comme les équations différentielles, aux dérivées partielles, stochastiques ou non) ou d'un formalisme discret (un programme d'ordinateur, quel qu'il soit, est un système dynamique discret). Dans les deux cas (continu et discret), un modèle peut être simulé et la reconstruction peut être posée comme un problème inverse : il s'agit de trouver le modèle, dans une classe de modèles définies par les connaissances a priori, dont la simulation fournit bien l'objectif attendu (lui aussi déclaré a priori comme dans la définition ci-dessus).

La définition suivante semble correspondre assez bien à la suite du paragraphe ainsi qu'aux différents types de modélisations pratiquées à l'INRA : « un modèle est une 'reconstruction' d'un processus à partir des données et des connaissances, pour un objectif donné »

Henri Caussinus

professeur émérite, Laboratoire de statistique et probabilités
Université Paul-Sabatier, Toulouse

J'ai lu « La modélisation à l'INRA » avec le plus grand intérêt. Ce document met beaucoup d'ordre dans un domaine difficile aux multiples facettes, il identifie bien les grandes classes de problèmes qui se posent et, sans doute parce qu'il a été conçu par un groupe aux compétences variées et complémentaires, il est équilibré dans ses considérations sur les rôles de chacun dans un domaine éminemment multidisciplinaire. C'est donc un document important, pour tous les chercheurs par sa partie la plus générale, plus particulièrement pour les sciences de la vie par les exemples qu'il donne en ce domaine et, là encore, son effort de classification (et donc de clarification), pour l'INRA enfin plus précisément (qu'on me pardonne cette appréciation concernant un institut auquel je n'appartiens pas) par l'accent mis sur l'importance de la modélisation dans le dialogue pluridisciplinaire et la complémentarité des sciences fondamentales et appliquées. Devant un travail de cette ampleur, il serait incongru de critiquer tel ou tel point de détail, comme il serait difficile d'ajouter quelques remarques pertinentes en mettant le doigt sur telle ou telle insuffisance. Je me permettrai plutôt de divaguer en donnant quelques réflexions « parallèles » qui me sont venues à l'idée durant sa lecture. Cela permettra aussi de revenir sur les appréciations qui précèdent et de les justifier sur quelques points.

Dans son article de 1962, *The future of data analysis* (Annals of Mathematical Statistics, 33, 1, pp 1-67), qui fut le texte de référence d'une génération (au moins) de statisticiens, J. Tukey cite M. Wilk : « *the hallmark of good science is that it uses models but never believes them* », ce qui met l'accent sur l'acception instrumentale du modèle (voir la citation de J.-M. Legay dans l'introduction) mais surtout indique, en creux, qu'il paraissait encore utile à cette époque de bien préciser que le modèle ne devait pas être pensé comme une vérité. Quelques années plus tard, J.-P. Benzecri (L'Analyse des Données, tome 2, 1973), énonçant ses principes sur « *l'analyse des données fondée sur l'usage des ordinateurs* » (page 3), écrit (page 6, 2^e principe) « *le modèle doit suivre les données, non l'inverse* » et stigmatise « *l'abondance de modèles, forgés a priori puis confrontés aux données par ce qu'on appelle des tests* ». Indépendamment des points de vue de

l'auteur sur la place du modèle, c'est la façon même dont le problème est posé qui importe : on en retire l'impression que, pour lui comme pour ceux qu'il critique, le modèle a une fin en soi... Ces exemples semblent témoigner de deux choses : même dans une période assez récente (et déjà dans l'ère de l'informatique), les modèles sont souvent évoqués sans discussion détaillée de leur statut ; quand ce dernier est pris en considération, c'est de façon assez dogmatique, le dogme pouvant évidemment différer des uns aux autres. Cette relation au modèle pourrait d'ailleurs expliquer que l'on ait longtemps parlé de modèle, objet idéalisé, plus que de modélisation, acte complexe mettant en jeu une foule d'outils et d'acteurs.

Mais la réflexion a beaucoup avancé ces dernières années ce qui nous ramène au document « La modélisation à l'INRA » pour observer tout d'abord qu'il vient parfaitement en son temps apporter des clarifications de première utilité, non seulement pour l'INRA, mais pour toutes les sciences de la vie et, au-delà même, pour la plupart des disciplines scientifiques et techniques. Il était important en premier lieu de souligner que le modèle dépend des objectifs et distinguer ainsi deux grandes classes selon que ceux-ci sont de l'ordre de la connaissance ou de l'action. Ce ne fut pas toujours une évidence. Par exemple, les statisticiens connaissent bien la longue et virulente querelle entre Fisher et Neyman au sujet des tests d'hypothèses, vers le milieu du XXe siècle ; pour schématiser, on peut dire que le premier se référait à l'élaboration des connaissances, le second se référait à l'action, et on mit quelque temps à se persuader que là résidait la différence majeure et que, sans doute, chacun avait raison dans l'environnement convenable. Et, même sur cette question simple qui semble aujourd'hui maîtrisée, bien préciser les buts d'une étude éviterait des approches hybrides perverses (voir par exemple l'article de D. Denis, suivi d'une discussion : *The modern hypothesis testing hybrid – R.A. Fisher fading influence* (2004), *Journal de la Société Française de Statistique*, 145, 4, pp 5-68), ce qui nous ramène encore à notre propos et à l'importance majeure d'une classification claire des objectifs. A ce sujet, j'insisterai sur l'importance du modèle en matière de communication puisque celle-ci est essentielle pour établir la connaissance et que, d'une part une connaissance non communicable est un non-sens, d'autre part l'avancée scientifique passe par le rapprochement de plusieurs connaissances partielles. Il faut noter cependant que, justement parce que le modèle sert la communication, il peut servir de passerelle entre connaissance fondamentale et action : quel type de modèle est-il alors en jeu ? Cette question est justement mise en évidence dans le rapport (fin du paragraphe 1.3, p. 42) et renvoie à un problème épistémologique et pratique important. Un autre problème délicat à souligner est celui de l'assemblage des modèles : bien entendu, comme dans une composition orchestrale, on ne fait pas jouer ensemble et n'importe comment, n'importe quels modèles (voir paragraphe 3.4, p. 45). Mais on peut aussi arriver à une difficulté spécifique : si le modèle sert de simplification nécessaire à une réalité peu accessible, dans quelle mesure un assemblage ne sera-t-il pas trop complexe pour la communication ou, tout simplement, pour être convenablement contrôlé (et ce n'est pas seulement un problème informatique) ? Voilà encore un verrou à analyser (qu'on trouve en filigrane dans le paragraphe 3.8, p. 49).

Jean-Paul Haton

Professeur, LORIA/INRIA-UMR CNRS 7503
Université Henri-Poincaré, Nancy

Ce rapport, bien rédigé et complet, résulte d'une réflexion approfondie menée au sein de l'INRA. Il montre bien les compétences nombreuses et très diverses en modélisation disponibles dans cet Institut, ainsi que le caractère pluridisciplinaire de ce thème. Le département MIA est exemplaire en ce qui concerne la coopération sur la modélisation entre des chercheurs d'horizons très différents.

Les chercheurs de l'INRA ont su tisser des liens avec les communautés de modélisation, notamment celle d'intelligence artificielle (laboratoires universitaires, INRIA, CNRS et AFIA).

Un lien de plus en plus fort est celui qui lie modélisation et simulation, notamment pour les systèmes dynamiques complexes, d'une grande importance et nécessitant un effort spécifique, comme le rapport le signale très justement.

Dans le paragraphe 3.1, p. 42, du document principal, un domaine qui me paraît important est celui de la reconnaissance de formes. En ce qui concerne les acteurs de la modélisation à l'INRA, le rapport cerne bien les diverses compétences nécessaires. On peut ajouter que le « mathématicien » pourrait être avantageusement complété par le spécialiste d'intelligence artificielle maîtrisant des concepts spécifiques à ce domaine, tels que représentation de connaissances, systèmes multi-agents, définition et utilisation d'ontologies, etc. En particulier, le développement collectif de modèles, la mise en réseau de compétences et

la capitalisation des résultats et des expériences nécessite pour les chercheurs de l'INRA de construire des ontologies dans les différents domaines considérés.

Les modèles statistiques peuvent également intégrer des connaissances sur les phénomènes étudiés, notamment sous forme de probabilités conditionnelles (cf. les modèles graphiques : HMM, réseaux bayésiens, etc.). Ces modèles ont pris une importance croissante au cours des dernières années et sont bien représentés à l'INRA.

Un point important relatif à la taille des projets est le passage à l'échelle réelle qui pose souvent des problèmes imprévus.

Au total, le présent rapport identifie clairement les problèmes liés à la modélisation sous ses différents aspects, ainsi que leur incidence à l'INRA et les moyens et méthodes nécessaires pour les traiter.

JD Lebreton

Directeur de recherche CNRS

Centre d'écologie fonctionnelle et évolutive (CEFE), Montpellier

Le texte pose bien l'omniprésence actuelle de la modélisation (dès l'introduction, puis en paragraphe 4.6, p. 50), qui justifie clairement une réflexion commune. L'importance des objectifs assignés à la modélisation est aussi clairement mise en avant, ce qui témoigne bien sûr de la maturité de la réflexion et des pratiques sur le sujet à l'INRA. Le texte fournit un panorama et des réflexions très intéressantes, et les quelques commentaires ci-dessous ne sont donnés « que pour ce qu'ils valent », et sont plutôt des notes de lecture.

D'une façon générale, les contributions de ce texte pourraient être poussées plus avant, diverses idées étant avancées avec une grande prudence. Par exemple, pages 41 et 42, la réflexion sur le rôle et la nature des modèles pourrait être poussée plus loin, notamment sur les modèles comme générateurs de concepts. De même page 42, le diagnostic sur les petits modèles, associés à l'expérimentation, la modélisation aux interfaces, et l'absence relative dans les modèles pour l'action est intéressant mais pourrait être approfondi par plus d'arguments et des références. La classification par classes de problèmes reprise page 58, est très intéressante, et mériterait d'être affinée et validée par une comparaison avec l'extérieur.

Le paragraphe 3 (page 42) et la typologie qu'il propose soulèvent immédiatement de nombreuses pistes de réflexions :

- Il me semble que la nature des questions que l'on pose au modèle est sous-jacente à cette typologie : dans 3.3 le mot prévision (ou « projection » au sens de prévision si les « paramètres » restent en l'état) est important. L'excellente analyse dans ce paragraphe dépasse les modèles de risques, et s'applique à une large part des modèles en agro-écologie. Une des pistes pourrait être d'explicitement accepter de mettre en concurrence des modèles assez variés techniquement du même phénomène et visant la même question, comme dans la confrontation des modèles de changement climatique.

- Dans 3.4, je lis implicitement « modèles pour comprendre », forcément très importants.

- Je ne mettrai pas en avant dès 3.1 la « recherche d'une structure » : Est-ce vraiment recommandable ? le « data mining » n'est-il pas un péché ? Le rôle de la discipline source doit rester majeur.

- Le paragraphe sur le spatial pourrait être complété d'un inventaire des questions et des savoir-faire, car il me semble que l'INRA a globalement une masse critique de compétences marquée.

La réflexion sur le rôle des plates-formes de modélisation dans « l'intégration » est très importante. Dans 3.4, 3.5 et 3.6 : on peut ainsi poser la question du contrôle de qualité (débugage code, validation), et du structuré par opposition au ad hoc. De façon intéressante l'appel d'offre Bio-informatique d'il y a quelques années, mettait l'accent sur les questions de validation, les méthodes de validation étant elles-mêmes un enjeu de recherche, dans la mesure où elles font partie quelques part du génie logiciel au sens large (c'est implicite mais pas traité dans 3.7 complexité calculatoire).

La partie 3.8, « verrous », mériterait d'être détaillée car elle peut permettre de définir des stratégies à assez large échelle donc intéressantes pour la direction d'un organisme comme l'INRA. Je rapprocherai de ces verrous, la difficulté à utiliser les compétences en matière de modélisation stochastique, évoquées page 42,

alors qu'il s'agit d'un des atouts de MIA et de l'INRA. Ces compétences devraient permettre de gommer la frontière modèle statistique/modèle mécaniste évoquée page 41. La mobilisation de ces compétences au service de la communauté INRA est une question à mon avis centrale car c'est un lieu de savoirs qu'il est difficile de rendre pluridisciplinaire. Les deux écueils sont bien sûr de ne pas utiliser un outil stochastique pour une question qui le réclamerait (par exemple pour un processus fortement non linéaire affecté d'une forte variation), ou au contraire, d'utiliser un outil d'une complexité démesurée par rapport à l'objet et à la question, comme un processus markovien d'étude mathématique difficile pour un problème qui peut être simplifié voire traité en déterministe. La modélisation stochastique est donc difficile à mobiliser à bon escient, et qui touche à la place et au mode de fonctionnement du département MIA à l'INRA.

Le paragraphe 4, « une culture sur la modélisation... », p. 49, est lui aussi très riche. Un des points délicats est certainement l'équilibre entre prise en compte des pratiques dominantes et renouvellement. Des thèmes forts pourraient être précisés par exemple la nécessité dans un grand nombre de cas de modèles spatialisés.

J'ai été vivement intéressé par le paragraphe 5.2 « les acteurs », car c'est bien sûr un (si non « le ») point crucial pour faire vivre et se développer une activité pluridisciplinaire. Voici mes remarques :

- p. 51 : les ingénieurs en développement informatique, vue la nécessité de travailler de façon structurée, générique, avec des outils de génie logiciel (voir plus haut), vu aussi la sous-estimation des problèmes numériques (par exemple ceux liés à la discrétisation ou aux simulations stochastiques) par beaucoup de « thématiciens », doivent aussi être pour une partie des ingénieurs en calcul scientifique (besoin de nouveaux métiers).

- p. 50 : les objectifs me paraissent relativement peu argumentés d'où découle par exemple le besoin de favoriser de très nombreux petits projets, ce qui est peut-être une fausse piste (mais voir ma remarque sur la mise en concurrence de modèles sur une même question).

- p. 50 : acteurs : on ne saurait trop insister sur l'importance du thématicien méthodologiste et de la qualité de son lien avec les mathématiciens appliqués et informaticien : c'est probablement une priorité de pousser de tels profils.

- p. 51 : je ne suis pas sûr qu'il faille distinguer – avec les outils de haut niveau actuel – entre ingénieur en méthode de modélisation et ingénieur en développement informatique. Des ingénieurs numériques et calcul scientifique acceptant de – comme on dit - "pisser du code" sont souvent le profil (hybride) le plus précieux (cf plus haut).

L'apport de compétences peut se faire avec des thésards mis sur l'interface thématique/méthode. L'absence d'une masse critique d'interdisciplinarité rendra le rôle de l'ingénieur méthode difficile. L'ingénieur méthode sera plus à l'aise dans de gros projets car il y aura une masse critique d'interdisciplinarité pour fonctionner.

Quid de l'appartenance à un laboratoire et à un département ? Un mathématicien appliqué peut-il appartenir à MIA et être dans un Labo appartenant à un département thématique ? (cf les remarques en page 52). La "direction" d'un ingénieur calcul scientifique appartient alors à un thématicien méthodologiste s'il a une "surface" suffisante en mathématiques appliquées. C'est peut-être un risque à prendre. Le "bien entendu" de la page 52 laisse entendre de forts risques. Il doit aussi laisser entendre de bonnes raisons. Sinon je suis globalement d'accord avec l'ensemble des remarques page 52.

- p. 54 : réseaux : on parle de l'intérieur de l'INRA seulement ou surtout ? Cela est-il raisonnable ?

- On sent en filigrane MIA un peu sur la défensive par rapport à un risque "d'atomisation". C'est une difficulté dont il faudrait sortir par le haut.

- Tout ce qui est dit sur la formation page 54 me paraît très pertinent. Ateliers de travail sur un projet de modélisation ???

- Apprendre à vivre avec des gros modèles : cf démarche validation/qualité bien connue pour changement global.

Eric Walter,
Directeur de recherche CNRS, Laboratoire des Signaux et Systèmes
Supélec, Gif-sur-Yvette

La réflexion sur la modélisation conduite à l'INRA à l'initiative de sa direction générale est particulièrement bienvenue. La fascinante capacité d'un relativement petit nombre de structures mathématiques à décrire le comportement d'objets d'apparence totalement différente a en effet pour conséquence la possibilité d'une mise en commun d'une réflexion méthodologique dont tous les départements scientifiques de l'INRA (et pas seulement de l'INRA) devraient bénéficier. Ceci est d'autant plus vrai que les problèmes rencontrés sont souvent trop complexes pour qu'un seul individu puisse prétendre les maîtriser et que, comme indiqué au paragraphe 5.1, p. 50, de nombreux verrous sont communs.

Le constat d'une situation éclatée où des études similaires sont conduites en parallèle par des groupes trop petits sans capitalisation et mutualisation des connaissances est courageux, et l'analyse des points forts et faibles de l'organisme honnête. L'étude demandée au département MIA a été conduite dans l'esprit d'ouverture indispensable, à la fois en interne vers les autres départements et en externe (je peux ainsi témoigner du fait que Bruno Goffinet et Hervé Monod sont venus longuement discuter avec moi dans mon laboratoire, et que j'ai été invité à de multiples occasions où le sujet était discuté). Les outils proposés pour améliorer la situation vont dans la bonne direction.

Le degré de succès de l'initiative dépendra très largement de la capacité qu'aura l'INRA de convaincre son personnel que cette collaboration est de l'intérêt de chacun et de l'organisme dans son ensemble. Ma modeste expérience des collaborations entre ce que le rapport appelle les méthodologistes et les thématiciens est que tout repose sur la confiance personnelle que les participants au projet ont pu construire les uns vis-à-vis des autres. Cette confiance est longue à construire et facile à détruire. Elle ne se décrète pas mais l'organisme peut prendre des mesures pour la favoriser, en particulier en valorisant les agents qui se lancent dans l'aventure.

Un point central me semble également être d'agir pour attirer vers l'INRA de jeunes chercheurs compétents sur les divers aspects de la méthodologie de la modélisation. Pour ne prendre qu'un seul exemple, qui est celui que je connais le mieux, il n'est pas clair dans l'esprit de nombreux jeunes automaticiens que l'INRA est l'un des endroits où ils peuvent développer une recherche en modélisation et commande des systèmes dynamiques, et ceci avec des possibilités de recrutement non négligeables.

En conclusion, ce document fort intéressant témoigne de la capacité qu'à l'INRA à conduire une réflexion à dimension collective sur des sujets qui dépassent les dimensions de chacun de ses départements scientifiques. Je forme le vœu que les autres grands organismes de recherche (CNRS, INRIA et INSERM notamment) s'associent à l'opération car les problèmes de modélisation pour la connaissance et pour l'action dont il est question ici dépassent largement l'INRA.