

Schéma Stratégique
Décembre 2004

Département
Mathématiques et Informatique Appliquées

Sommaire

	Page
A- Les grandes orientations scientifiques	2
1- Missions, champ d'activité et orientations	2
2- Environnement et place comparative	4
3- Evolutions à engager	4
4- Les champs thématiques	5
4.1 Introduction	5
4.2 Thèmes méthodologiques	6
4.2.1 Systèmes dynamiques complexes – TM1	6
4.2.2 Méthodologie d'analyse et de gestion des risques – TM2	7
4.2.3 Représentation et exploitation d'informations dynamiques et spatialisées –	8
TM3	
4.2.4 Recherche d'une structure dans une grande quantité de données – TM4	8
4.2.5 Modèles pour l'action – TM5	9
4.2.6 Algorithmique et complexité calculatoire – TM6	9
4.3 Les champs d'intervention de l'INRA : les lignes de la matrice	10
4.3.1 Du gène à la cellule	10
4.3.2 Physiologie, écophysiologie	13
4.3.3 Processus physiques et biophysiques	14
4.3.4 Ecologie, épidémiologie : la communauté dans son milieu	14
4.3.5 Gestion des agro-éco-systèmes	15
4.3.6 Alimentation et agro-alimentaire	16
4.4 Champs thématiques	17
4.4.1 CT 1 : Bioinformatique, mathématiques et informatique pour les modèles	17
issus du gène à l'individu	
4.4.2 CT 2 : Modèles et décisions pour l'environnement, l'agriculture et	18
l'alimentation	
4.4.3 Une vision synthétique	18
5- Disciplines, compétences et métiers	19
5.1 Les chercheurs	19
5.1.1 Objectif	19
5.1.2 Moyens et mis en œuvre	20
5.2 Les ITA	20
B- Lignes directrices en matière d'organisation et de partenariat	21
1- Evolution du dispositif de recherche	21
1.1 Animation et réseaux	22
1.2 Organisation géographique	22
1.3 Risques et risques alimentaires	22
1.4 Unités pluri-départementales	23
1.5 Fonctionnement en projet ou équipes	23
2- Partenariat	23
2.1 Enseignement et formation	23
2.2 Partenariat scientifique	24
2.3 Animations collectives, plates-formes, consultation, expertise	26

A - Les grandes orientations scientifiques :

1- Missions, champ d'activité et orientations :

Comme beaucoup d'autres activités humaines, la recherche est stimulée par les évolutions toujours plus rapides de la technique. L'augmentation régulière de la puissance informatique, du nombre de données disponibles, des vitesses de calcul, des possibilités de stockage etc., permet de construire des modèles et de traiter des problèmes de plus en plus complexes, associant le temps et l'espace, travaillant à des échelles différentes et pouvant combiner des informations de nature différente. Le recours à des approches quantitatives et en particulier à la modélisation se généralise ainsi dans de nombreux domaines de l'INRA et interpelle les méthodologies mathématiques et informatiques de MIA.

Ceci s'accompagne d'une évolution des grandes disciplines de MIA. Les statistiques par exemple - "We may have to moderate our romance with mathematics" (J.H. Friedman 2001, International Statistical Review p. 5-10)-, revendiquent à juste titre de maintenir le souci de rigueur mathématique dans les développements, tout en s'ouvrant à des problématiques où la volonté opérationnelle peut être en contradiction avec la rigueur : comment utiliser des informations pertinentes mais difficiles à faire rentrer dans un modèle formel ? Les contours des disciplines se redessinent pour répondre aux enjeux en empruntant les concepts pertinents des autres disciplines. Globalement, les disciplines informatiques et mathématiques demeurent de très vastes champs disciplinaires dont le rôle s'accroît.

En face de toutes les sollicitations, MIA ne peut pas aller "partout" ni en terme de thèmes disciplinaires, qui sont trop vastes, ni en terme de domaines finalisés.

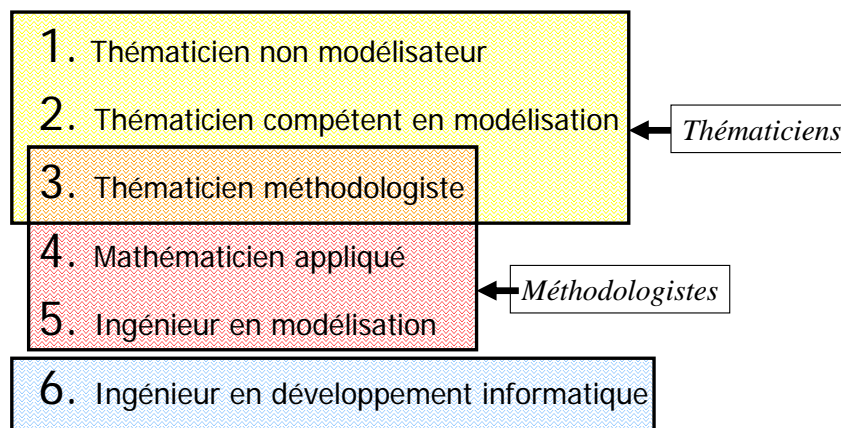
Ceci nous conduit donc à privilégier un certain nombre de thèmes finalisés (TF) prioritaires sur lesquels, en cohérence avec les recommandations de la commission d'évaluation de MIA il y a 5 ans, nous regroupons nos efforts et visons à répondre aux besoins issus des objectifs finalisés de l'INRA. Si, bien entendu, le choix de ces thèmes fait intervenir le désir des équipes MIA et d'éventuelles opportunités géographiques, ils sont d'abord choisis en fonction de leur caractère prioritaire pour l'INRA et du caractère générique et décisif d'une recherche dans nos disciplines pour les faire avancer. Il faut noter cependant que, dans certains domaines (la physique par exemple), le rôle des mathématiques est fondamental et ancien. Dans ce cas, les départements "thématiciens" ont fortement investi les domaines mathématiques correspondant et l'opportunité d'investir se pose alors en d'autres termes. Les réflexions menées avec nos partenaires des autres départements dans le cadre des missions "modélisation" et "bioinformatique" ont joué un rôle déterminant dans l'analyse du rôle de nos disciplines dans les grands champs d'intervention de l'INRA.

Sur le plan des disciplines, MIA a largement ouvert son champ depuis quelques années. Partant de sa culture de base en statistiques, MIA a investi l'analyse numérique, l'intelligence artificielle et enfin l'automatique afin de présenter une meilleure adéquation aux besoins de l'INRA. A l'intérieur de ce large spectre, nos orientations s'appuient sur l'évolution des disciplines, sur nos domaines d'excellence et sur l'identification de besoins issus de nos champs finalisés. Notre objectif est de maintenir une couverture assez large dans nos disciplines, tout en veillant à ne pas laisser de chercheurs isolés dans leur thématique. Il est par exemple envisageable de recruter dans une nouvelle thématique comme le "data mining" qui est un peu à mi-chemin entre l'analyse des données des statisticiens et l'extraction automatique dans les bases de données. C'est plus difficilement faisable par exemple dans le domaine du calcul parallèle pour lequel il faudrait faire alors appel à un appui extérieur.

Au delà de la structuration disciplinaire, la réflexion modélisation a participé à définir des grands thèmes méthodologiques (TM) prioritaires (analyse des risques, systèmes dynamiques complexes, etc.), mobilisant chacun différents domaines des mathématiques et de l'informatique et concernant plusieurs domaines finalisés. Ce sont ces grands thèmes qui orienteront et structureront nos activités méthodologiques.

Nous proposons de construire des réseaux aussi bien sur des thèmes méthodologiques que sur des champs finalisés avec nos partenaires. Ce fonctionnement en réseau permettra d'assurer une expertise collective s'appuyant sur les compétences complémentaires de ses membres et se substituera à la consultation individuelle trop gourmande en temps et ne permettant pas d'accéder à la diversité des compétences. En parallèle, MIA poursuivra l'animation du dispositif FPStat de formation à la statistique afin de permettre aux membres des départements de l'INRA de maîtriser l'utilisation des statistiques en développant préférentiellement les formations plus pointues et spécialisées (statistique des données d'expression par exemple). La construction d'un dispositif analogue pour la modélisation doit être instruit rapidement avec nos partenaires.

Il y a des mathématiciens et des informaticiens dans d'autres départements de l'INRA. Il convient donc de se poser globalement la question de la place des chercheurs de nos disciplines dans l'INRA et de l'animation de ce dispositif. La typologie proposée dans le cadre de la réflexion modélisation permet d'identifier les différents acteurs nécessaires au dispositif.



MIA est en phase avec les principes développés dans la réflexion Modélisation concernant le positionnement de ces différents types d'acteurs. On trouve dans MIA des 3., 4., 5., et 6., mais le cœur de métier de MIA se situe en 4¹.

Les missions du département énoncées dans le cadre du schéma stratégique précédent sont toujours d'actualité :

"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 dans 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

¹ Cf. Rapport Modélisation – « Les différents méthodologistes : positionnement scientifique et géographique »
Décembre 2004

l'avancement des projets et programmes évoqués précédemment. Il participe à la conception et à la mise en oeuvre des programmes de formation dans ses domaines de compétence. "

2- Environnement et place comparative :

Parmi les chercheurs en math-info au plan national, certains ont un positionnement scientifique voisin du notre et visent à développer des méthodologies pour répondre à des questions posées par d'autres disciplines, et d'autres développent une recherche plus endogène dans leur discipline. Parmi les premiers, certaines équipes (INRIA, CNRS, INSERM, Universités) travaillent sur des objets intéressant l'INRA : bioinformatique, écologie, épidémiologie ou plus généralement modélisation de systèmes biologiques ou biophysiques.

Vis à vis de celles ci, MIA est globalement une force importante en statistiques, mais plus modeste, comparativement, en informatique et en automatique et systèmes dynamiques. Sur le plan de la bioinformatique, la répartition des forces entre les différents organismes est relativement équilibrée, même si l'INRIA dispose de forces plus importantes en informatique. Sur le plan de la modélisation de systèmes biologiques ou biophysiques, MIA est très présent dans les aspects statistiques mais peu, surtout par rapport à l'INRIA, dans l'analyse mathématique des modèles mécanistes.

Sur le plan international, certaines institutions ont un positionnement scientifique voisin du notre tels BioSS en Ecosse ou Biometries aux Pays-Bas, avec lesquels d'ailleurs nous construisons des relations. Notre taille plus importante que la leur nous permet de disposer d'un éventail de compétences plus large. BioSS par exemple, nous envie nos chercheurs en informatique dans le champ de la bioinformatique.

Globalement, MIA forme un groupe dont la composition est originale puisqu'il permet de mobiliser des compétences rarement associées dans une même structure, dans le cadre d'une finalité bio-agronomique. Cette force collective est un atout face aux enjeux liés au développement des approches quantitatives et en particulier de la modélisation.

3- Evolutions à engager :

Nous proposons de privilégier les thèmes méthodologiques répondant aux grands verrous identifiés dans la réflexion modélisation. Cette structuration en thèmes méthodologiques faisant appel chacun à plusieurs disciplines des mathématiques et de l'informatique, et n'étant pas une émanation directe des dynamiques endogènes des disciplines, devrait favoriser la communication entre les disciplines et l'élargissement du champ de compétence collectif sur ces thèmes. Parmi ceux-ci, le thème "Systèmes dynamiques complexes" est relativement nouveau pour MIA. MIA a déjà les compétences pour s'y investir, mais devra les compléter, en particulier aux interfaces entre statistiques, modèles dynamiques et modèles informatiques, pour répondre à tous les enjeux. Le thème "Recherche d'une structure dans une grande quantité de données" pose la question de l'investissement dans de nouveaux champs comme le data mining. Enfin, il nous faut accroître l'éventail de nos compétences dans le thème "Algorithmique et complexité calculatoire" qui reste le coeur de compétence de MIA, avec l'algorithmique des textes ou la géométrie algorithmique.

Sur le plan des thèmes finalisés, nous proposons de poursuivre les thèmes sur lesquels nous travaillons dans le domaine de la bioinformatique, en ouvrant un nouveau champ autour des réseaux de régulation. Ce thème pourra à terme, s'appuyer en partie sur les forces déjà présentes dans le champ, mais nécessitera de compléter les forces avec des compétences dans

le domaine de l'intégration des connaissances et des systèmes dynamiques complexes. D'autre part, nous identifions explicitement notre intervention dans le champ de l'écologie, sans prétendre en couvrir tous les aspects, mais en y exploitant nos compétences en statistique et plus spécifiquement autour du thème "Représentation et exploitation d'informations dynamiques et spatialisées". Nous confirmons notre investissement dans les domaines de l'épidémiologie, de la gestion des agro-éco-systèmes et des risques alimentaires. Nous affirmons notre disponibilité relativement à la problématique "Interaction Génotype X Milieu". Enfin, nous nous interrogeons, en liaison avec CEPIA, sur l'identification des problématiques du contrôle de procédés, où nos compétences sont les plus attendues.

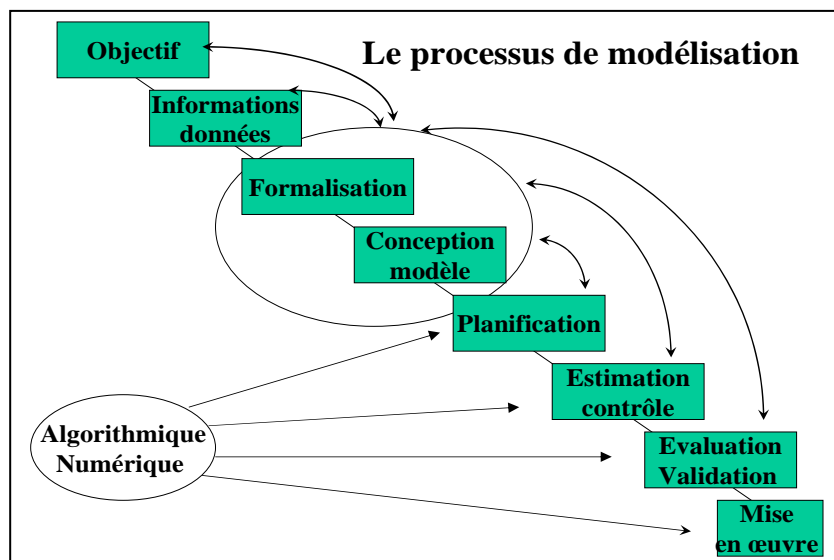
4- Les champs thématiques :

4.1 Introduction :

L'objectif premier pour MIA est de développer des méthodes mathématiques et informatiques qui permettent de mieux comprendre les phénomènes biologiques ou agronomiques et fournissent des moyens d'action, de gestion et de décision, à partir des connaissances sur ces phénomènes et des données recueillies. Cette activité peut se décomposer en différentes phases, qui participent à ce que l'on peut appeler le "processus de modélisation". Chacune de ces phases mobilise des méthodologies développées dans différentes disciplines de MIA qui répondent de manière complémentaire aux mêmes questions.

- identification des objectifs;
- conception, formalisation;
- choix et mise en oeuvre d'une classe de modèles: modélisation statistique; modélisation stochastique; modélisation mécaniste à base d'EDO ou EDP; modèles informatiques : modèles conceptuels, multi-agents, multi-acteurs, qualitatifs, de diagnostic;
- optimisation du recueil des données : planification; échantillonnage;
- gestion, représentation des données et extraction de sens: analyse d'image; analyse textuelle; bases de données réparties;
- estimation des paramètres ou des distributions, tests d'hypothèses, identification, calibration;
- validation, choix de modèles, étude de comportement, analyse de sensibilité;
- action : mise en oeuvre, contrôle, décision, optimisation.

Le schéma ci-dessous montre le fonctionnement du processus. La phase de modélisation proprement dite, est centrale, et interagit fortement avec les phases amont et aval.



La manière la plus claire de décrire les activités de MIA est une matrice où les lignes représentent les champs d'intervention de l'INRA et les colonnes, les grands thèmes méthodologiques pour nos disciplines. Nous décrivons tout d'abord les grands thèmes méthodologiques (TM) à développer puis les lignes de la matrice représentant différents niveaux d'échelle d'organisation du vivant ou de type de processus, dans lesquels sont identifiés des thèmes finalisés (TF). Nous proposons enfin une structuration en champ thématique qui représente au mieux les grandes structures de cette matrice.

4.2 Thèmes méthodologiques :

Il est essentiel d'avoir et de pérenniser par la recherche, des compétences disciplinaires générales en statistique, automatique et informatique dans les différentes phases du processus de modélisation. Mais nous voulons faire porter un effort particulier sur plusieurs grands thèmes, qui dépassent une classification disciplinaire et qui sont apparus comme déterminants pour répondre aux questions de la recherche agronomique. MIA est déjà présent activement sur les thèmes TM2, TM3 et TM5. Les thèmes TM1 et TM4 sont relativement nouveaux pour MIA et nous avons l'intention de les investir et en particulier le TM1. Enfin, le thème TM6 est transversal et au cœur des activités de MIA.

Aujourd'hui, ces thèmes méthodologiques sont plutôt des instruments de description de nos activités. L'objectif est d'en faire des objectifs de recherche avec des priorités identifiées.

4.2.1 Systèmes dynamiques complexes - TM1 :

Dans de nombreux domaines de l'INRA, les modélisateurs sont confrontés à des modèles dynamiques, construits à partir d'assemblages de sous modèles, plus ou moins spatialisés, éventuellement orientés vers l'analyse des risques, très complexes, avec un grand nombre de paramètres, 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 et qualitative 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, notamment lorsque l'on veut utiliser un tel modèle avec un objectif d'action : faut-il et comment le simplifier pour le rendre efficace (vitesse, coût, qualité, robustesse, etc.) ? Se posent aussi les problèmes d'agrégation et désagrégation d'informations pour utiliser conjointement des modèles construits à différentes échelles. Ils posent en retour des problèmes au niveau de la conception des modèles, même si cette phase est surtout prise en charge par les thématiciens.

Ces développements interpellent aussi bien les statistiques et les probabilités (agrégation, désagrégation, modèles mixtes, analyse de sensibilité, processus de branchement, etc.) que les systèmes dynamiques à base d'équations différentielles (détection de rupture dans les systèmes dynamiques, EDP pour prendre en compte l'hétérogénéité dans un système fermé, etc.) 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.

L'INRIA ou le département STIC du CNRS ont des compétences importantes sur ce thème, mais la spécificité des modèles issus des sciences du vivant rend nécessaire un effort spécifique à l'INRA. L'investissement dans ce thème est une priorité pour MIA.

Pour lancer la dynamique, un programme est en cours de construction :

Exploration numérique des propriétés des modèles

L'objectif est à la fois d'assurer la diffusion et le partage d'informations, de susciter des recherches (par exemple dans le domaine de l'analyse automatique et non supervisée des trajectoires, ou dans la représentation minimale), et de faciliter la mise en œuvre en proposant un support informatique cohérent. Des collaborations externes sont indispensables (IFREMER, IRD, INRIA, etc.).

Quelques autres questions à développer, sans prétendre à l'exhaustivité :

- Echantillonnage dans l'espace (très grand) des paramètres et des variables pour analyser la sensibilité du modèle, et estimer les paramètres,
- Simplification ou réduction de modèle à partir d'analyse de robustesse, en utilisant en particulier l'optimisation convexe, ou bien à partir de méthodes d'apprentissage automatique,
- Détection de rupture dans un système dynamique à partir de méthodes non paramétriques.

4.2.2 Méthodologie d'analyse et de gestion des risques - TM2 :

De nombreux modèles dans les champs d'intervention de l'INRA 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, avec des effets de seuil. De plus, ces modèles sont construits le plus souvent pour répondre à des questions de gestion et de contrôle. Il n'existe pas encore de méthodologie satisfaisante pour répondre à toutes ces questions, en particulier lorsque l'on cherche à analyser des événements exceptionnels (changement climatique par exemple).

Un cadre général adapté à ce type de problèmes est le cadre bayésien, et en particulier les réseaux bayésiens. Ces modèles permettent assez naturellement de prendre en compte les données hétérogènes et celles subjectives issues d'experts. La théorie des valeurs extrêmes est bien évidemment essentielle. Plus généralement, on retrouve les problèmes que pose la modélisation de systèmes complexes et, l'importance sociale des sujets nécessite tout particulièrement la prise en compte de multiples sources d'information.

Les équipes leaders sur les problématiques finalisées liées au risque (qu'il soit environnemental ou alimentaire) se trouvent essentiellement en Europe du Nord (Royaume Uni, Pays-Bas, Danemark) et aux USA. Nous entretenons des relations avec elles sous forme de post-docs en particulier.

Quelques questions à développer :

- Estimation de queues de distribution pour des données reconstruites,
- Développement de méthodes robustes pour petits échantillons.

4.2.3 Représentation et exploitation d'informations dynamiques et spatialisées - TM3 :

Les informations concernant un processus dynamique à modéliser sont le plus souvent spatialisées. 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'arrangements spatiaux 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 ? D'autre part, de nombreux problèmes de décision ou de contrôle sont spatialisés et demandent le développement de méthodologies adaptées.

Les méthodes statistiques (modèles bayésiens, modèles mixtes, mélanges spatiaux, statistique des processus, etc.) jouent un rôle essentiel, avec celles de l'analyse d'image (suivi et analyse de mouvement dans des scènes 3d par exemple).

Quelques équipes aux USA et au Danemark travaillent sur les statistiques spatiales et leurs applications dans des domaines biologiques ou agronomiques. En analyse d'images, un grand nombre d'équipes travaillent sur l'imagerie médicale, mais relativement peu sur des images biologiques aux échelles qui nous intéressent (parmi celles-ci, nous avons des contacts réguliers avec des équipes de l'Institut Pasteur).

Une priorité dans ce thème : associer fortement les aspects spatiaux et dynamiques. Dans ce but, un profil apportant des compétences en systèmes dynamiques en complément des compétences en modélisation spatiale, est ouvert en 2005 en Avignon.

De nombreuses questions, plus spécifiquement liées aux aspects spatiaux, restent ouvertes :

- Fusion d'images de différentes modalités et à différentes échelles,
- Utilisation des modèles bayésiens pour données incomplètes,
- Appréhension de la non stationnarité par déformation de l'espace.

4.2.4 Recherche d'une structure dans une grande quantité de données - TM4 :

Dans de nombreux domaines, génome 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". De plus en amont du modèle, se posent les questions de la construction et l'interrogation de bases de données qui doivent être inter-opérationnelles pour fonctionner en réseau, intégrer des informations de natures multiples et disposer d'interfaces graphiques performantes.

Au niveau de la construction des bases de données, les méthodologies concernent les applications client serveur, la télégestion, les graphes conceptuels flous, ou l'intégration et la représentation de données complexes. En ce qui concerne l'analyse, des méthodes existent : méthodes d'apprentissage, choix de variables à l'aide de simulations MCMC, data mining, fouille de données, représentation des connaissances, extraction de connaissances textuelles, apprentissage de règles d'extraction, etc. Certaines sont issues des statistiques, mais peuvent être moins exigeantes en matière de mathématiques, demandent surtout à être jugées sur leur efficacité et nécessitent un dialogue renforcé entre thématiciens et spécialistes des méthodes.

Notons cependant, qu'il y a de véritables verrous mathématiques dans la data mining. Enfin, 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.

Quelques questions à développer :

- Semi automatisation d'une stratégie avec des règles expertes,
- Filtres particuliers à convolution pour le filtrage non linéaire avec incertitudes.

Une instruction à mener :

Pour l'instant la problématique « Data Mining » concerne plutôt les thématiques méthodologiques à même de porter le questionnement thématique. MIA analysera s'il y a une problématique générique méthodologique derrière, mal prise en compte à l'extérieur de l'INRA et légitimant donc un investissement

4.2.5 Modèles pour l'action - TM5 :

Beaucoup de modèles sont destinés à l'action, et en particulier la décision, la gestion et le contrôle. Cependant, relativement peu sont construits en intégrant fortement cet objectif. Un enjeu est de modéliser la perception qu'ont les acteurs humains de leur environnement, les processus de décision, les processus de mise en oeuvre des décisions, et les interactions entre les différents acteurs ou groupes d'acteurs.

Les méthodes de l'intelligence artificielle (processus décisionnel de Markov, modèles multi-agents, animation pour l'aide à la décision de groupes à partir de simulations, ainsi que les concepts de la gestion de production) sont sollicitées pour modéliser les processus cognitifs et sociaux. Les difficultés apparaissent lorsque plusieurs acteurs sont impliqués, lorsque les décisions dépendent de critères multiples et évolutifs, et notamment de la façon dont les acteurs s'approprient les informations lors de négociations.

Dans le domaine du contrôle, les méthodes concernent le contrôle robuste, multivariable et le contrôle optimal stochastique. Les phénomènes de non-linéarité souvent rencontrés en biologie (inhibitions, rétro actions négatives) posent des problèmes de stabilité difficiles si on veut contrôler le processus. D'autre part, la supervision des procédés agro-alimentaires nécessite de construire des méthodes hybrides intégrant des processus continus et discrets, ainsi que des informations symboliques et quantitatives.

Quelques questions à développer :

- Construire une méthode de supervision hybride associant règles expertes et processus de commande,
- Assurer la commande des modèles de réacteurs à paramètres distribués,
- Construire une décision spatialisée à partir de processus décisionnel de Markov.

4.2.6 Algorithmique et complexité calculatoire - TM6 :

La dimension paramétrique des modèles, le nombre 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, par exemple :

- 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 très vastes,
- Recherche de motifs semblables dans des séquences nucléotidiques.

La difficulté essentielle de ces problèmes réside dans la grande dimension des espaces explorés qui conduit habituellement à des algorithmes de complexité temporelle exponentielle. La recherche d'algorithmes efficaces et rapides, et d'approximations raisonnables et calculables est donc un thème de recherche actif et permanent dans les communautés mathématiques et informatiques. Les communautés scientifiques sont fortement structurées par les formalismes mathématiques utilisés pour modéliser le problème (réseaux bayésiens et modèles graphiques, réseaux de contraintes, processus décisionnel de Markov, programmation linéaire en nombres entiers, etc.) mais aussi par la nature des problèmes traités (continus, discrets, optimisation, intégration, ...) menant à l'utilisation de techniques variées.

Pour tous ces problèmes, la théorie de la complexité montre bien que le point bloquant réside dans le caractère exponentiel (en temps et parfois en espace) des algorithmes. Cette théorie montre toutefois qu'il y a peu d'espoir de voir apparaître un jour une méthode exacte et générale permettant de résoudre toutes ces questions. Les lents et tenaces travaux de recherche permettent de grignoter lentement la courbe de la dimensionnalité et de rendre traitable des problèmes de plus en plus larges avec des modèles de plus en plus réalistes.

Il est essentiel que MIA continue à être présent dans cette dynamique pour que les problèmes spécifiques liés aux domaines d'intervention de l'INRA bénéficient de ces progrès.

Concernant les modèles probabilistes, les leaders sont statisticiens (Danemark, UK) ou informaticiens (USA). La communauté française participe au développement d'algorithmes approchés. Concernant les réseaux de contraintes, l'Europe et la France partagent le leadership avec les USA.

4.3 Les champs d'intervention de l'INRA : les lignes de la matrice :

Les 6 thèmes suivants suivent en partie des échelles croissantes d'organisation du vivant ou plus généralement des catégories de processus. Pour chacun de ces thèmes, nous présentons notre analyse des besoins, les points sur lesquels nous sommes déjà présents ou sur lesquels nous désirons investir en priorité (thèmes finalisés TF) en précisant dans quelles phases du processus de modélisation et au travers quelles grandes thématiques méthodologiques.

4.3.1 Du gène à la cellule :

Les mathématiques et informatique appliquées dans ce cadre se confondent maintenant avec le terme "bioinformatique". La confrontation de l'offre de recherche des bioinformaticiens de l'INRA avec les besoins des départements utilisateurs des méthodes et outils est en cours dans le cadre de la mission "Bioinformatique" confiée à MIA. Ce que nous donnons ci-dessous représente l'analyse au niveau MIA des différents thèmes du domaine. Elle sera amenée à être modifiée à l'issue de la mission bioinformatique.

Il s'agit de développer des méthodes qui permettent d'atteindre des connaissances biologiques à l'échelle du génome, à partir d'informations au niveau de la séquence, mais aussi de données de nature variées : données d'expression, protéome, phénotypes, etc. Les méthodes qui seront développées feront l'objet, dans certains cas en fonction des opportunités, d'un développement logiciel destiné aux biologistes. Les méthodologies sollicitées concernent tout particulièrement :

- TM4 Recherche d'une structure dans une grande quantité de données. C'est un enjeu fondamental ici avec la quantité formidable de données issues de la biologie moléculaire.

- TM6 Algorithmique et complexité calculatoire. Les modèles construits comportent le plus souvent de très grands nombres d'état ou de paramètres.
- TM1 Les systèmes dynamiques complexes. Les réseaux de régulation sont des systèmes dynamiques très complexes.
- TM3 Représentation et exploitation d'informations dynamiques et spatialisées. Il s'agit ici de l'analyse d'image 2D, 3D ou 4D (mouvement).

Thèmes finalisés (TF) prioritaires

4.3.1.1 Localisation de régions fonctionnelles sur les séquences - TF1:

C'est un problème relativement ancien, pour lequel il existe déjà des méthodes et logiciels, dans lequel MIA a déjà obtenu de nombreux résultats, et qui reste pleinement d'actualité pour améliorer les méthodes (meilleure modélisation des régions fonctionnelles, algorithmes plus efficaces), localiser de nouveaux types de régions (micro ARN) et intégrer d'autres informations que la séquence (informations de conservation entre organismes différents, données d'expression). Il reste un long chemin à parcourir avant d'atteindre une annotation semi-automatique sur les génomes eucaryotes.

Les développements méthodologiques utilisent la modélisation par chaînes de Markov cachées, l'étude des statistiques de test, l'apprentissage automatique (réseaux de neurones ou SVM), et enfin l'algorithmique (programmation dynamique, satisfaction de contraintes, pattern matching).

MIA a les compétences pour poursuivre dans ce domaine, et pourrait utilement bénéficier de compétences complémentaires en apprentissage automatique ou en algorithmique des textes et motifs (combinatorial pattern matching).

4.3.1.2 Annotation fonctionnelle - TF2 :

Il s'agit de prédire la fonction biochimique et biologique de la région fonctionnelle localisée. Une première source d'information est l'homologie, détectable par similarité de séquence ou de structure. Une seconde réside dans l'analyse bibliographique qui est très lente. Le développement de méthodes d'extraction automatique d'informations à partir de la littérature est donc nécessaire.

MIA compte poursuivre ses recherches dans le domaine de l'extraction automatique, dont l'intérêt dépasse largement cette thématique.

4.3.1.3 Détection et localisation de QTL - TF3 :

Il s'agit de détecter et de localiser des gènes ayant une action sur un caractère quantitatif (QTL). L'approche traditionnelle (interval mapping) a montré ses limites en terme de précision de localisation des QTL (intervalle de confiance de 20 cMorgan). Pour améliorer cette précision, l'utilisation du déséquilibre de liaison et la mise en correspondance avec des données d'expression sont des approches prometteuses.

MIA, fort de son expérience dans le domaine, a les compétences statistiques (statistique des processus, statistique asymptotique) à compléter en matière de méthodes de coalescence pour obtenir des résultats en partenariat avec les généticiens de GA et GAP dans un domaine qui est assez spécifique de la recherche agronomique.

4.3.1.4 Analyse des données d'expression : transcrits et protéines - **TF4** :

Les données d'expression, transcrits et protéines, sont obtenues à l'aide de différentes techniques : électrophorèse 2D, spectrométrie de masse, biopuces, microscopie de fluorescence, Q-PCR. Elles se présentent alors sous forme d'images ou signaux numériques qu'il s'agit de traiter afin d'extraire des données pertinentes qui permettront de faire du clustering ou d'identifier des sur ou sous expressions significatives, et enfin participeront à terme à l'élucidation des réseaux de régulation.

Les méthodologies de MIA sont sollicitées tout d'abord au niveau de la planification des expériences. Si l'analyse d'images 2D est relativement automatique, l'analyse des images obtenues en microscopie 3D, 4D (3D + temps) et 5D (4D + multispectrale) qui peut permettre la localisation spatiale ou spatio-temporelle dans la cellule, nécessite du travail de recherche. Enfin, le traitement des données obtenues, qui mobilise des techniques relativement standards, pose des problèmes liés à la taille des lots de données à traiter.

MIA est reconnue dans ces 3 domaines. Dans le cadre de l'analyse d'images, MIA a une équipe qui travaille en partenariat avec des équipes INRA et non INRA (Curie, INRIA). Dans le cadre du traitement statistique, devant le nombre très important de demandes, MIA doit jouer un rôle (par exemple au travers de FPStat) dans le transfert de méthodes (déjà connues ou en développement) auprès des biologistes.

Thème en émergence

4.3.1.5 Réseaux de régulation - **TF5** :

Un des véritables challenges de la bioinformatique est de construire un modèle prédictif des interactions et régulation entre gènes dans la cellule. Des modèles sont déjà développés à partir des données d'expression, mais il s'agit à terme de savoir intégrer des données de natures très diverses: données d'expression éventuellement localisées dans l'espace et le temps, données phénotypiques, knockout, analyse bibliographique, métabolisme, etc. Au delà de la construction, la simulation d'un réseau est aussi un enjeu où il faut pouvoir mêler régulation continue et discrète.

MIA et l'INRA sont peu présents sur ce terrain à l'heure actuelle. L'INRIA, le CNRS et le CEA développent des travaux dans le domaine. Nous pensons qu'il est absolument nécessaire que l'INRA investisse ce champ dans une perspective "biologie intégrative" en liaison avec les objectifs finalisés de l'INRA. MIA est donc prêt à ouvrir des projets dans ce domaine en collaboration étroite avec des biologistes. Dans un premier temps, il apparaît préférable de privilégier les organismes dont le fonctionnement est plus simple, comme les micro-organismes ou les levures. Dans un deuxième temps, il faudra envisager des projets concernant les organismes supérieurs, en intégrant les informations aux échelles supérieures. La présence de mathématiciens et de biologistes sur des mêmes campus est un atout pour l'INRA.

Sur le plan des méthodes, la problématique interpelle fortement les thèmes TM1, TM4 et TM6. De nombreuses approches sont envisageables et une forte coordination des organismes investissant ce champ est à construire.

Un projet en cours de construction

Sont impliqués à Jouy-en-Josas, deux unités MIA et plusieurs biologistes du centre. Un chercheur spécialiste des systèmes dynamiques est muté pour conduire ce projet. Il s'agit de construire une représentation « système » d'un réseau de régulation concernant une problématique biologique particulière (non encore précisée) en menant en parallèle l'aspect intégration des informations, qui prolonge notre investissement actuel dans le domaine, et l'aspect modélisation qui est résolument nouveau. Le fonctionnement du projet sera évalué dans un an et sera, le cas échéant, fortement soutenu par le département.

Autres thèmes

Phylogénie et cartographie comparée. En dehors de ces thèmes, MIA n'a pas a priori l'intention d'investir de manière importante dans la phylogénie et la cartographie comparée au sens descriptif, où son expérience est limitée. Cependant, MIA exploitera l'approche comparative dans les nombreux thèmes où il est d'une aide précieuse.

Cartographie génétique. Le logiciel Carthagène fera l'objet de développements nouveaux pour inclure l'analyse comparative.

Détermination de la structure 3D des protéines. Il s'agit d'un problème central pour déterminer leurs fonctions. Cette thématique est couverte par une équipe PHASE dans l'unité MIG. MIA n'envisage pas pour l'instant de s'y impliquer, mais reste attentif à son évolution.

4.3.2 Physiologie, écophysiologie :

MIA a pour l'instant une activité relativement faible à ce niveau. En effet, les biologistes de l'INRA impliqués dans ces thématiques sont relativement autonomes. Cependant plusieurs éléments peuvent modifier notre position vis à vis de ce thème.

Les modèles à l'échelle d'un individu sont souvent l'assemblage de sous-modèles associés chacun avec des connaissances physiologiques fines. La confrontation aux données pouvant se faire à l'échelle supérieure. Ces éléments et la complexité croissante de ces modèles interpellent alors la thématique TM1.

De plus, les modélisations en physiologie tendent à impliquer de plus en plus les informations à l'échelle du gène et de la cellule dans une perspective de biologie intégrative (4.3.1.5). La prise en compte d'informations issues d'images peut être ici décisive, et nous avons déjà une activité en ce sens.

Enfin, si ces modèles sont développés avant tout pour des objectifs de connaissance à l'échelle de l'organe ou de l'individu, ils sont aussi la base de modèles pour l'action et la décision à d'autres échelles (voir 4.3.4, 4.3.5) dans lesquels nous intervenons.

Nous proposons de participer activement à l'instruction de la question « Rôle des modèles dans la biologie intégrative ». Notre investissement dans ce thème sera précisé suite à cette instruction.

4.3.3 Processus physiques et biophysiques :

MIA a une activité assez réduite dans ce thème. En effet, les physiciens sont tout à fait autonomes dans la modélisation mathématique qui fait partie intégrante de leur discipline. Cependant, le rôle et le nombre croissants des données, donnent plus d'importance à la problématique de confrontation des modèles aux données où les compétences de MIA peuvent être sollicitées. Cela est vrai en particulier pour les modèles liés à l'action : prédiction, planification, etc. et ceux qui mettent en oeuvre des changements d'échelle spatiale ou temporelle.

4.3.4 Ecologie, épidémiologie : la communauté dans son milieu :

Cette échelle de modélisation a connu récemment un regain d'intérêt à l'INRA et MIA a été et continue à être fortement sollicité pour s'y investir, en particulier en épidémiologie. Quatre grands thèmes méthodologiques pour MIA sont sollicités à cette échelle :

- TM1 Les modèles dynamiques complexes : Il y a une forte demande d'aide à l'utilisation des modèles relativement standards à base d'équations différentielles, que les "thématiciens" devraient s'approprier à terme. MIA doit assurer la maîtrise de modélisations plus complexes mathématiquement (EDP stochastique, par exemple) qui sont difficilement transférables auprès des biologistes. Au delà, de nombreuses questions de recherche se posent tels l'emboîtement d'échelles multiples dans un même modèle ou l'étude des comportements transitoires;
- TM2 Analyse et gestion des risques : De nombreuses modélisations concernent in fine l'analyse d'un risque environnemental, de santé animale ou végétale, ou de santé humaine. Cela pose les questions méthodologiques de l'analyse des risques comme la prise en compte d'événements rares ou la présence de variables cachées;
- TM3 Représentation et exploitation d'informations dynamiques et spatialisées : Les informations à cette échelle sont de plus en plus spatialisées et posent des problèmes de forte variabilité spatiale ou temporelle des données qui sont de plus lacunaires, hétérogènes, spatialement et temporellement très corrélées.
- TM5 Modèles pour l'action : De nombreux modèles sont (ou devraient être) conçus pour la gestion. Nous comptons développer et mettre en oeuvre des méthodes de contrôle et gestion des interventions pour ces modèles. L'intervention de ce thème méthodologique dans ce champ est nouvelle. Nous tenons à la promouvoir en concertation avec nos partenaires des autres départements pour en préciser les attentes.

Plus spécifiquement :

4.3.4.1 Ecologie - TF6 :

Notre objet principal d'étude est d'analyser le lien entre la biodiversité d'une part et l'environnement et les pratiques culturelles ou sylvicoles d'autre part, de mesurer son impact sur les agrosystèmes et de proposer des méthodes de gestion durable de la biodiversité. Il concerne des phénomènes de diffusion de gènes, de pollen (même si on peut considérer que la diffusion de pollen est plutôt un phénomène physique), ou d'invasion.

MIA investira, notamment via le projet MERE (voir 4.3.6.2) le champ de l'écologie microbienne qui offre des quantités de données sensiblement plus importantes et qui permettrait l'étude de modèles dynamiques plus fins.

4.3.4.2 *Epidémiologie - TF7 :*

Il s'agit d'étudier le développement des maladies animales ou végétales et de proposer des stratégies de gestion.

L'étude du comportement du modèle pour prédire le devenir d'une épidémie soumise à différentes formes de contrôle et de mode gestion, est particulièrement déterminante.

MIA est impliqué dans la transversalité "Epiemerge" et co-encadre 4 ASC en partenariat avec SPE et SA sur ce thème. Plusieurs chercheurs de MIA proposent de s'impliquer dans les réseaux en cours de construction au sein du département SPE.

4.3.4.3 *Interaction Génotype X Milieu X Mode de culture - TF8 :*

Il s'agit de développer des stratégies de sélection et d'évaluation variétale face à un environnement écologique et socio-économique changeant.

Un enjeu pour MIA est l'utilisation conjointe de modèles mécanistes et statistiques pour modéliser l'interaction. Un réseau EA, GAP, MIA existe sur le papier, mais n'est pas vraiment opérationnel.

4.3.5 Gestion des agro-éco-systèmes :

A cette échelle, trois grandes thématiques méthodologiques sont mobilisées :

- TM5 Modèles pour l'action. A cette échelle les systèmes étudiés sont anthropisés et la prise en compte des comportements des acteurs est nécessaire. D'autre part, les méthodes de contrôle peuvent jouer un rôle important dans la gestion de certains systèmes (eau, forêt).
- TM3 Représentation et exploitation d'informations dynamiques et spatialisées. Les phénomènes modélisés sont toujours spatialisés à cette échelle.
- TM1 Les modèles dynamiques complexes. Les modèles construits à cette échelle intègrent le plus souvent des processus à des échelles inférieures, conduisant à des systèmes complexes.

Le TM2, Méthodologie d'analyse et de gestion des risques, n'est pas mobilisé pour l'instant dans ce champ. Nous étudions l'opportunité de son implication.

4.3.5.1 *Gestion de systèmes de production - TF9:*

Il s'agit de développer des méthodes d'aide à la conception de conduite de systèmes de production à l'échelle de l'exploitation agricole ou à celle d'un bassin versant, ou de systèmes de gestion environnementale à l'échelle régionale. Ces méthodes intègrent des modèles écophysologiques et des modèles de comportement d'acteurs.

Un des enjeux est de construire ces modèles de comportement et d'intégrer les modèles écophysologiques. Les modèles doivent prendre en compte la différence de nature et l'antagonisme entre les différents critères qu'ils soient socio-économiques ou environnementaux, et entre les acteurs ("collecticiel" pour acteurs multiples, par exemple). De plus, l'optimisation de stratégie pose des problèmes algorithmiques importants face à la dimension des systèmes à résoudre.

Cette thématique prolonge la transversalité "Décision" et bénéficie de partenariats avec le

CIRAD, l'IRD et le CEMAGREF. Les départements EA, SAD et MIA (avec éventuellement PHASE et SAE2) ont l'intention de construire un réseau autour de ce thème en parallèle avec l'animation des ASC recrutés autour de la thématique « Gestion dans les systèmes biotechniques ». Au niveau MIA, cette thématique est en grande partie portée par les méthodologies informatiques et pourrait bénéficier d'une connexion avec les statistiques autour de la décision spatialisée. Notons qu'un nombre important d'équipes travaillant à ce niveau dans les différents départements concernés sont à Toulouse et pourraient se structurer en un pôle fort sur ce thème.

Autres thèmes

Spatial et Agriculture. Il s'agit de construire des modèles statistiques décrivant l'hétérogénéité spatiale au niveau de la parcelle (agriculture de précision), ou bien d'utiliser les données satellites pour estimer un plan d'occupation des sols ou assurer un suivi de culture. D'autres questions émergent au niveau d'un bassin versant, comme l'identification des parcelles où la mise en place d'une démarche d'agriculture de précision serait pertinente. L'investissement de MIA est relativement limité, mais très cohérent avec ses compétences et important pour les partenaires EA.

Exploitation des ressources. Il s'agit de développer des méthodes de gestion durable de ressources comme l'eau ou les forêts. Dans le domaine de l'irrigation, il s'agit de développer des méthodes pour augmenter la flexibilité des systèmes de distribution. Un des enjeux en automatique est le traitement de l'effet "retard" et la maîtrise de systèmes multivariés. Les partenaires se trouvent essentiellement au CEMAGREF.

4.3.6 Alimentation et agro-alimentaire :

Pour ce qui concerne l'alimentation les thèmes méthodologiques fortement sollicités sont :

- TM2 Méthodologie d'analyse et de gestion des risques. C'est bien entendu fondamental pour l'analyse des risques alimentaires.

- TM4 Recherche d'une structure dans une grande quantité de données. L'intégration, la représentation et la gestion répartie des données, est un enjeu décisif face à un afflux massif des données lié en particulier à l'instrumentation des procédés de fabrication ou aux exigences de traçabilité. Il s'agit de donner des recommandations pour la collecte, de représenter des données hétérogènes, qualitatives ou quantitatives, incomplètes ou imprécises, de stocker et mettre à jour les informations, de combiner et intégrer des sources d'informations dispersées, de valider les bases et enfin de construire des méthodes d'extraction et interrogation de ces bases, et assurer la traçabilité.

- TM5 Modèles pour l'action. Les méthodes du contrôle sont centrales en conduite de procédés.

4.3.6.1 Analyse des risques alimentaires -TF10 :

Il s'agit de donner une évaluation quantitative des risques dus à la consommation de denrées alimentaires contaminées biologiquement ou chimiquement. Un premier enjeu est d'intégrer des données et connaissances hétérogènes et dispersées. L'intégration de données de contamination et de données de consommation permet d'estimer l'exposition à un risque.

L'enjeu est ensuite de produire une estimation globale des risques dus à la présence de contaminants dans un produit alimentaire. Cette estimation doit intégrer de nombreux aspects : prospective sur les risques émergents, effets chez l'homme, existence de risques conjoints et en interaction, variabilité à tous les niveaux (hommes, souches, etc.), etc. Les méthodes d'analyse des risques sont évidemment au cœur de la problématique. Les méthodes d'analyse

et contrôle de systèmes, pourront aussi être mobilisées autour de l'écologie microbienne.

L'unité pluridépartementale MIA-ALIMH « Mét@risk », située à l'INA-PG, doit jouer un rôle moteur dans la thématique, en coordination étroite avec l'équipe MIA-Jouy et les autres équipes INRA travaillant sur le thème. Assurer à cette unité les moyens en terme de recherche et d'ingénierie est une priorité pour MIA.

4.3.6.2 Conception et conduite de procédés - TF11 :

La thématique concerne aussi bien les procédés dans le cadre agroalimentaire que ceux de la dépollution. Un enjeu est d'intégrer les différents types de modélisation et les différents types de contrôle des procédés intervenant successivement dans la fabrication d'un aliment, que ces procédés fassent appel à des connaissances expertes ou à des modèles formalisés.

Un autre enjeu est de développer des modèles et des commandes qui tiennent compte de la répartition non homogène des bactéries au sein des bioréacteurs. Enfin, il faut développer des méthodes de surveillance des procédés, par exemple avec des tests statistiques de rupture.

Concernant la dépollution, cette thématique fait l'objet d'une collaboration formalisée avec le LBE (EA) et l'INRIA sous forme d'un projet (au sens INRIA) commun (MERE).

En ce qui concerne les procédés agro-alimentaires, nous nous interrogeons avec CEPIA sur les attentes et les priorités en matière de commande et de conduite. A priori, cette instruction devrait conduire à une diminution de notre investissement dans ce champ.

4.4 Champs thématiques :

Cette analyse conduit à proposer deux champs thématiques :

-CT1 Bioinformatique : Mathématiques et informatique pour les modèles du gène à l'individu. Ce champ réunit les thèmes "Du gène à la cellule" et "Physiologie, Ecophysiologie".

-CT2 Modèles et décision pour l'environnement, l'agriculture, et l'alimentation.

Ce champ réunit les thèmes "Milieux physiques et biophysiques", "Ecologie, épidémiologie : la communauté dans son milieu", "Gestion des agro-éco-systèmes" et "Alimentation et agroalimentaire".

4.4.1 CT1 – Bioinformatique, mathématiques et informatique pour les modèles du gène à l'individu :

Ce champ thématique demande un fort investissement en biologie pour répondre aux grands enjeux. La bioinformatique n'est plus seulement un accompagnement de la biologie, mais elle devient prospective et innovante, et est en position d'émettre des hypothèses que les biologistes expérimentateurs devront vérifier. L'investissement dans ce champ doit donc être relativement pérenne.

Un premier enjeu pour ce CT1 est d'être à l'origine d'un résultat biologique important. Un autre enjeu fondamental est de construire et démarrer un projet ambitieux concernant les réseaux de régulation. Enfin, il faut produire des méthodes et logiciels qui soient largement utilisés.

Il est pertinent d'organiser une animation de la bioinformatique à l'INRA, dépassant les frontières du département MIA et de confronter régulièrement les bioinformaticiens avec les biologistes utilisateurs des méthodes. Cependant, cette animation doit être coordonnée avec l'animation nationale et ne pas se substituer à elle au niveau de chaque sous-thème où elle existe (données d'expression par exemple).

4.4.2 CT2 - Modèles et décisions pour l'environnement, l'agriculture, et l'alimentation :

Dans ce champ thématique, on trouve des objets assez différents, mais il est plus facile de passer de l'un à l'autre que de changer de champ thématique. Les thèmes méthodologiques "Analyse et gestion des risques" et "Modèles pour l'action" sont essentiellement mobilisées dans ce champ, ce qui lui donne de plus une cohérence méthodologique.

Même si le thème concerne l'ensemble de MIA, c'est un enjeu pour ce champ d'investir le thème "TM1 Systèmes dynamiques complexes" et de mettre en œuvre ces méthodologies dans les différents thèmes finalisés du champ. Sur le plan finalisé, un objectif est d'être impliqué en profondeur dans plusieurs projets importants à vocation opérationnelle : gestion d'une épidémie, d'un risque alimentaire, d'un risque environnemental, d'une production etc.

Contrairement au CT1, ce CT2 n'a pas vocation à construire une animation unique. Il sera amené, en revanche, à jouer un rôle déterminant dans la construction, la coordination et l'animation des différentes structures : réseaux, formations, etc., qui seront mises en place suite à la mission modélisation.

C'est dans ce champ qu'interviennent les trois grands programmes transversaux de l'INRA : Agriculture et développement durable, Ecologie pour la gestion des écosystèmes et de leur ressource, Alimentation. Les thèmes particuliers, sur lesquels les équipes MIA interviendront dépendront donc en partie des arbitrages pris dans chacun de ces programmes. Au-delà de la participation à des projets construits sous la responsabilité de nos partenaires, nous avons l'intention de proposer des projets articulés autour de problèmes méthodologiques communs à plusieurs sous projets « thématiques ». Cependant, fédérer des partenaires autour d'un projet méthodologique reste une vraie difficulté. Tout l'enjeu pour MIA est d'arriver à la lever.

4.4.3 Une vision synthétique

Le tableau ci-après décrit les forces en présence dans la matrice TM X Champs, et les propositions d'évolution reprenant les différents points présentés dans l'analyse des lignes et colonnes de la matrice.

	TM5	TM2	TM3	TM4	TM1	TM6	
Du gène à la cellule							CT1
Physiologie, écophysiologie							
Processus physiques et biophysiques							CT2
Ecologie, épidémiologie							
Gestion des agro-écosystèmes							
Alimentation et agro-alimentaire							

5- Disciplines, compétences et métiers :

Globalement, l'investissement de MIA dans l'animation de réseaux et dans la formation tel qu'il est apparu nécessaire dans le cadre de la réflexion modélisation va poser des problèmes de force de travail disponible. Il posera aussi des problèmes de valorisation de ces activités dans les carrières.

5.1 Les chercheurs :

5.1.1 Objectif :

Le coeur de compétence des chercheurs MIA est, et doit rester, les mathématiques et l'informatique. Il est nécessaire de maintenir et développer une compétence pointue dans les domaines qui répondent aux grands thèmes méthodologiques décrits dans le paragraphe 4.2. Globalement, les compétences actuelles de MIA sont pertinentes pour investir ces thèmes, mais certains thèmes méthodologiques, en particulier les TM1 et TM4 vont demander d'élargir le champ de recherche de MIA. Concernant le TM1 que nous tenons à investir en priorité, il nécessitera le recrutement de nouvelles compétences et le renforcement de nos liens avec nos partenaires mathématiciens et informaticiens (INRIA, CNRS, Universités).

Cependant, au-delà de la maîtrise des méthodes mathématiques ou informatiques, les chercheurs de MIA doivent être capable de formaliser un problème de biologie ou d'agronomie. Il ne s'agit pas de se substituer aux "thématiciens", mais de savoir dialoguer avec eux et de savoir identifier le type de modélisation approprié face au problème posé et analyser les verrous et difficultés.

Dans ce but, il faut accroître la palette des méthodes maîtrisées collectivement dans une équipe à l'intérieur d'un thème méthodologique sans que forcément ces méthodes soient toutes des thèmes de recherche de l'équipe.

5.1.2 Moyens mis en oeuvre :

La participation conjointe de plusieurs équipes, aux compétences complémentaires, dans des projets finalisés, est un moyen de confronter des approches variées face à un problème. Il faut les encourager.

Le recrutement de nouvelles compétences sera faite avec précaution afin d'éviter l'isolement d'un nouveau recruté. Il est nécessaire que l'équipe ou l'unité d'accueil possède des compétences assez proches de celles à recruter et qu'elle ait établi des relations avec des équipes porteuses de la compétence, extérieures à l'INRA.

Si cela est possible, nous tenons à privilégier les recrutements en CR2. En effet, l'expérience montre que le nombre de candidats de qualité est sensiblement plus important en CR2. Cependant, il pourra être opportun de recruter en CR1 dans le cas de l'ouverture d'un nouveau champ méthodologique.

Si l'on veut un investissement fort dans l'animation et la formation, il faut être vigilant dans la prise en compte de ces activités dans les carrières des chercheurs.

5.2 Les ITA :

Il est nécessaire de maintenir et développer une compétence forte en développement informatique qui constitue un mode de diffusion fondamental de nos travaux. Pour autant, les scientifiques ne peuvent pas déléguer dans la totalité la responsabilité de ces développements.

Une implication de plus en plus étroite dans l'analyse et le traitement des données, dans les projets en partenariat avec des équipes d'autres départements, est nécessaire si l'on veut s'impliquer en profondeur dans ces projets. Ceci demande d'accroître notre potentiel en particulier en matière de gestion des données.

Le réseau informatique est notre outil de travail. Il faut en maintenir la qualité en maintenant le potentiel humain qui en assure le fonctionnement.

Certains ingénieurs ont une fonction liée à une plate-forme relativement extérieure aux préoccupations des unités de recherche. Il semble cependant important qu'ils restent rattachés à une unité de recherche, ce qui leur permet de participer à la dynamique de l'unité, de faire évoluer leurs compétences, et donne de la souplesse dans leur évolution professionnelle.

La formation permanente est essentielle dans les métiers de l'informatique qui évoluent très rapidement. Le réseau des informaticiens de MIA permet des rencontres régulières et la confrontation des expériences dans les différentes unités.

B- Lignes directrices en matière d'organisation et de partenariat :

1- Evolution du dispositif de recherche :

La structuration en champs thématiques est cohérente avec les équipes (qui se confondent avec l'unité dans certains cas) en place. Dans le CT2, certaines équipes ont une identité plutôt liée à un thème méthodologique (**TM3** à Avignon par exemple) avec une implication dans différents thèmes finalisés du champ, alors que d'autres ont une identité centrée autour d'un thème finalisé (Mét@risk par exemple). Dans le CT1, l'identité des équipes est plus globalement la bioinformatique, avec quelques **TF** privilégiés. Nous donnons ci-dessous les thèmes qu'ils soient finalisés ou méthodologiques, qui font l'identité de ces équipes (entre parenthèses quelques identifiants secondaires).

Unités	Equipes	Nbre* Cat. A INRA	Nbre EC	C.T.	Thèmes identitaires	Identifiants secondaires
Jouy MIG		8		CT1	TF1 / TF2 TF4	TM4 TM6
Jouy BIA	Mathématiques pour la biologie cellulaire	5		CT1	TF4	TM3 / TM4 TM6
	Mathématiques pour les risques	13		CT2	TM2	TM1 / TM3 TF6 / TF7 TF8 / TF10
Evry UMR		3	5	CT1	TF1	TM4 / TM6
INA-PG UMR	Génome	1	2	CT1	TF4	TM3 / TM4 TM6
	Modèles mathématiques pour la biologie et l'environnement	1	1	CT2	TM3 TM2	TF6
	Méthodes informatiques pour la maîtrise des risques alimentaires		2	CT2	TM4	TF10 TM2
INA-PG Mét@risk		4		CT2	TF10	TM2 / TM4
Toulouse	Statistiques et informatique appliquées à la génétique et à la biologie moléculaire	11		CT1	TF1 TF3	TM4 TM6
	Modélisation des grands systèmes	4		CT2	TM3 TM1	TF6 TF8
	Méthodes mathématiques et informatiques pour la décision	5		CT2	TF9 TM5	TM6
Montpellier UMR	Génome		2	CT1	TF4	TM4
	Analyse des systèmes	8	1	CT2	TF11 TM5	TM1 TM4
Avignon		9		CT2	TM3	TM6 / TF6 TF7

* Y compris les recrutements 2004 mais sans compter les AI.

1.1 Animation et réseaux :

Les conclusions du rapport Modélisation mettent en évidence la nécessité de construire deux types de réseaux. D'une part, les réseaux sur des thématiques liées aux grands axes prioritaires de l'INRA, qui doivent être portés par des « thématiciens » ou par des associations « thématiciens X méthodologistes » (Gestion de systèmes biotechniques, Epidémiologie, Risques alimentaires, etc.) et auxquels MIA est prêt à participer activement. D'autre part, des réseaux sur des thèmes méthodologiques (cohérents avec les thèmes méthodologiques identifiés dans le rapport Modélisation) que MIA se propose de construire et d'animer. Par exemple, un réseau autour des problèmes de changement d'échelles est en cours de construction, animé par l'équipe « Modélisation MIA Toulouse ».

Au-delà de leurs rôles pour le développement et la diffusion de méthodes dans l'INRA, ces réseaux constitueront l'ossature de l'animation scientifique de MIA. Les thèmes méthodologiques (TM), ou finalisés (TF), serviront de support à ces réseaux mais n'auront pas de rôle fonctionnel. Comme indiqué dans le rapport Modélisation, ces réseaux feront l'objet de lettre de mission pour en préciser l'objectif et exprimer explicitement l'intérêt de la hiérarchie.

1.2 Organisation géographique

La construction des collaborations autour de projets s'appuie souvent, mais pas uniquement, sur des interactions de proximité. Le positionnement géographique des unités MIA joue donc un rôle important dans la construction de collaborations aussi bien avec nos partenaires « thématiciens » qu'avec nos partenaires académiques. Quelques illustrations :

-Modélisation des réseaux de régulation : La contrainte d'insertion dans une équipe de bioinformaticiens pour mobiliser les méthodes et outils d'intégration, et la nécessaire proximité avec un thématicien méthodologiste fortement concerné par cette problématique, nous ont conduit à construire la thématique à Jouy, à partir d'un chercheur de Montpellier, compétent en analyse de systèmes.

-Biologie intégrative : comme indiqué dans 4.3.2, nous proposons de participer à l'instruction de ce thème. La présence dans MIA à Montpellier d'une compétence forte en analyse de système, et l'existence d'un pôle important en biologie végétale à Montpellier, sont des éléments importants de cette réflexion.

1.3 Risques et risques alimentaires :

L'unité pluri-départementale MIA-ALIMH « Mét@risk », située à l'INA-PG, doit jouer un rôle central dans les activités liées aux risques alimentaires. Il est cependant pertinent de conserver une activité en la matière à Jouy dans le cadre d'une équipe « Mathématiques pour les Risques » visant à développer des méthodologies génériques autour de la problématique « Risque », que ceux-ci soient sanitaires ou environnementaux. La coordination de l'activité concernant les risques alimentaires doit cependant être sous la responsabilité de Mét@risk quant au choix des problématiques finalisées. De même, l'activité de recherche autour de la construction et l'exploitation des bases de données pour l'analyse des risques, se poursuit dans l'UMR INA-PG en étroite coordination avec l'unité voisine Mét@risk.

1.4 -Unités pluri-départementales :

Il y a actuellement 2 unités de ce type dans MIA : MIG et Mét@risk. Concernant MIG, l'évaluation collective de l'unité après les 4 premières années de fonctionnement montre déjà l'efficacité du système. Le mode de fonctionnement des unités pluri-départementales est en phase avec les propositions de Bernard Sauveur. Cependant, ce ne doit pas être le mode universel de fonctionnement des unités MIA. En ce qui concerne l'écologie à Avignon, le système mis en place avec EFPA (un CR EFPA dans une unité MIA) semble un moyen pertinent pour construire des activités de recherche en commun et nous pourrions faire le point dans 4 ans sur ce fonctionnement. Plus généralement, faire de nos unités un lieu de passage plus ou moins long pour des "biologistes" des autres départements, et inversement que les unités de biologistes puissent accueillir des agents MIA, pourrait être un moyen intéressant de construction de thématiques transversales.

Nous n'avons pas a priori de projets de création de nouvelles unités pluri-départementales dans les 3 années à venir. Cependant, à Toulouse, avec le renforcement des liens d'une part, avec l'UMR Arche et d'autre part, avec nos partenaires en bioinformatique, la question d'une restructuration de l'ensemble pourra se poser.

1.5 Fonctionnement en projet ou équipes :

Il n'y a pas lieu de promouvoir un mode unique de fonctionnement. Les équipes apportent le sentiment d'appartenir à un groupe, un collectif. Les deux modes peuvent coexister. En général, mais c'est loin d'être une règle absolue, c'est l'équipe qui produit la vision prospective et les projets qui produisent l'activité scientifique.

2- Partenariat :

2.1 Enseignement et formation :

MIA compte poursuivre son investissement dans la formation aux statistiques dans l'INRA à travers FPStat. Cependant, le système mis en place il y a 15 ans est peut être à revisiter. D'une part, il est de plus en plus difficile de trouver des formateurs dans les différentes implantations de l'INRA. C'était plus facile au moment de la construction du dispositif où les formateurs étaient formés par des spécialistes dans les phases de mise au point des modules, et ils en tiraient alors un bénéfice en terme de formation approfondie. La formation autour d'un module existant est moins motivante. De plus, l'activité de formation n'est pas toujours bien valorisée dans la carrière des formateurs. D'autre part, les demandes de formation qui arrivent via le dispositif FPStat sont de plus en plus spécifiques et pas toujours en correspondance directe avec les modules de formation existants. Une solution pourrait être de repositionner les formations dans le cadre des grands projets transversaux finalisés, comme la dynamique des populations ou l'analyse des données d'expression par exemple, en liaison avec des écoles chercheurs. Les statistiques, et peut être plus généralement les méthodes mathématiques et informatiques pour la modélisation, seraient décrites en liaison approfondie avec le champ finalisé visé, et non pas comme à l'heure actuelle autour des méthodes génériques des statisticiens. Ceci présuppose la participation d'agents MIA aux groupes de coordination et animation de ces projets transversaux pour aider à mieux définir les besoins en formation. Certaines parties telles qu'elles existent à l'heure actuelle pourraient être sous-traitées (ce sera de toute façon le cas pour le niveau I) assez facilement à l'extérieur de l'INRA et ne feraient donc plus l'objet d'une intervention systématique du réseau de formateurs actuels.

La réflexion modélisation a identifié la nécessité de construire un dispositif de formation à la modélisation. MIA participera à la construction de ce dispositif en coordination avec les autres départements (par exemple dans le cadre d'un bureau permanent), qui pourrait s'appuyer d'une part sur des écoles chercheurs, et, d'autre part sur un système construit sur la logique FpStat.

MIA a une activité raisonnable en matière d'enseignement dans l'enseignement supérieur (1150 heures en 2002 données par 24 agents de catégorie A sur 75). MIA compte poursuivre cette activité à ce rythme. Il est important de s'impliquer dans les écoles doctorales à travers les enseignements de façon à jouer un rôle dans la formation de nos futurs recrutés, de mieux connaître les stagiaires pour mieux les choisir, et le cas échéant d'obtenir des financements de thèse. Nous pouvons jouer un rôle déterminant dans l'enseignement des mathématiques et de l'informatique appliquées aux sciences de la vie, et pour cela être présent dans les comités qui définissent les enseignements. Mais il ne faut pas pour autant avoir la même quantité d'activité d'enseignement que les enseignants chercheurs qui n'ont pas les mêmes missions. Un maximum de l'ordre de 50 heures par personne semble un chiffre raisonnable.

Un effort particulier pourra être effectué en direction des écoles d'ingénieurs, agronomiques ou autres, qui constituent un réservoir de candidats potentiels au recrutement et qu'il faut sensibiliser assez tôt à la recherche. Des écoles d'ingénieur généralistes comme SUPELEC s'interrogent sur l'ouverture de leur formation vers la biologie pour exploiter la culture « analyse de systèmes » de l'école. Nous allons accroître nos liens avec ce type d'école.

2.2 Partenariat scientifique :

Ecoles agronomiques et CIRAD: Les laboratoires de mathématiques et informatique appliquées des écoles agronomiques ont des positionnements et des objectifs scientifiques assez voisins des nôtres et la construction d'UMR avec eux est donc naturelle. Cependant, les écoles ont besoin d'une couverture relativement large des domaines disciplinaires et finalisés, ce qui peut être difficile à concilier avec la nécessité de spécialiser les unités pour éviter la dispersion. Le nombre d'enseignants-chercheurs à l'INA-PG permet de concilier les besoins, mais il faudra veiller à maintenir l'implication de l'INRA avec le départ des chercheurs travaillant autour des risques alimentaires à Mét@risk. Le problème est plus délicat à l'ENSAM où les enseignants chercheurs ne se retrouvent pas dans la dynamique de l'unité orientée vers l'automatique et les systèmes dynamiques. La pertinence de la poursuite du fonctionnement en UMR est liée à une diversification des orientations scientifiques qui soit assumée collectivement et pleinement par l'unité. La question de la création de nouvelles UMR avec l'enseignement agronomique n'est pas posée aujourd'hui, mais il faudrait au moins que les relations de l'unité de Toulouse avec l'ENSAT soient analysées et, le cas échéant, renforcées.

Même si nous sommes plutôt plus théoriciens, nos objectifs scientifiques sont très voisins de ceux des chercheurs MIA du CIRAD. Il faut donc poursuivre la recherche de coordinations avec eux, sur nos champs finalisés et sur certains de nos thèmes méthodologiques (par exemple, dans le cadre du projet « Analyse numérique des modèles »).

Université et CNRS: L'association sous forme d'UMR de l'unité Statistique et Génome d'Evry permet de mieux coordonner les activités de recherche avec celles très proches de MIG et de l'INA-PG autour de « Statistique et Génome ». Par ailleurs, pour l'unité BIA Jouy, des connexions sont envisagées avec l'unité MAP5 (Mathématiques appliquées, Paris 5) et l'ENSAE. Certains partenaires, à orientation plus « académique » que MIA, peuvent attirer les plus académiques de nos chercheurs. Il faut cependant éviter que cela ne modifie le positionnement scientifique "médian" du département.

INRIA : L'INRIA exprime l'intérêt d'investir les sciences du vivant et a un positionnement scientifique voisin du nôtre, même si notre position au milieu d'autres laboratoires de l'INRA nous rend peut être plus sensible à la nécessité du « retour » de nos recherches sur les questions finalisées qui les ont provoquées. Pour l'instant nos liens avec les équipes de l'INRIA se font entre des équipes aux compétences voisines : d'une part, le projet (au sens de l'INRIA, c'est-à-dire une équipe) MERE à Montpellier associe 3 chercheurs INRA et deux chercheurs INRIA autour du traitement de l'eau et au-delà de l'écologie microbienne ; d'autre part, un chercheur MIA est mis à disposition du projet VISTA à Rennes et permet de créer des liens entre cette équipe et l'équipe Image de l'unité de Jouy. Au delà, l'INRIA a des compétences plus importantes que les nôtres dans certaines thématiques méthodologiques comme les systèmes dynamiques complexes, et a déjà des collaborations avec plusieurs départements de l'INRA (EFPA, PHASE, EA, GA). Nous devons donc réfléchir en interaction avec eux sur les thèmes que nous devons investir, le cas échéant en interaction forte avec eux, et ceux pour lesquels ils sont mieux à même de contribuer. La biologie intégrative fait partie des thèmes où la question se pose.

International : Les équipes entretiennent des relations avec des équipes au positionnement similaire, en Europe et dans le monde, sur un nombre important de thèmes de recherche qui se caractérisent par des invitations croisées, et quelques travaux en collaboration (exemples : Chili, Danemark, etc.). Nous essayons de construire des relations plus institutionnelles avec BioSS (notre homologue Ecossais) et Biometries (Pays-Bas) avec l'organisation d'un workshop sur le risque, qui fait suite à un premier autour de la bioinformatique. Dans les projets Européens dans lesquels nous sommes présents, nous le sommes la plupart du temps "en deuxième ligne" derrière un partenaire INRA d'un autre département. C'est relativement normal vu que les thèmes des appels d'offres européens ne sont pas directement associés à nos compétences.

Socio-économique : L'outil fondamental de valorisation de nos travaux est le logiciel. Dans de nombreux cas, il s'agit essentiellement d'un « programme » à destination de chercheurs, avec une ergonomie limitée, qui met en œuvre une nouvelle méthode. La question est de savoir dans quels cas, et avec quels moyens il est opportun de développer des logiciels plus ambitieux, à plus large diffusion, où les critères de qualité logiciel s'appliquent : ergonomie, autonomie, évolutivité, etc. La possibilité d'investir dans de tels développements est liée à la possibilité de dégager des forces de développement spécifiques pour un tel projet. Les forces du département peuvent intervenir en « direction de projet », et un peu en développement. Dans la plupart des cas, un appui supplémentaire est nécessaire. Il peut se faire par des CDD affectés au projet : contrats privés, Génoplante, Européen, CDD INRA, etc. Le département peut jouer un rôle incitateur fort, s'il dispose d'un volant de CDD affectables sur ce type de projet (il ne s'agit pas ici de dire que le développement informatique doit être fait essentiellement par des CDD, mais, dans le cadre éventuel d'une augmentation significative des CDD à l'INRA, de proposer des métiers où l'on pense que des CDD sont les mieux (ou les moins mal) appropriés). Bien sûr, il convient de confier la responsabilité du projet à des ingénieurs titulaires afin d'en assurer la pérennité et le suivi. Le département pourra alors se doter d'indicateurs de qualité des logiciels (un peu comme à l'INRIA) incluant le niveau de diffusion réel du produit et son utilité pour répondre à une question finalisée, et fera en sorte qu'ils soient pris en compte dans les critères d'évaluation à tous les niveaux (individus, unités). De manière générale, sauf contraintes apportées par un partenaire investissant fortement dans un projet, ces produits seront diffusés sous la forme de logiciels libres et il faut poursuivre, avec les autres départements concernés, la réflexion et l'information sur les aspects « protection intellectuelle » liés aux logiciels libres.

2.3 Animations collectives, plates-formes, consultation, expertise :

MIA dispose de forces relativement importantes en matière d'ingénierie informatique et est impliqué dans des domaines très variés de l'INRA qui le conduisent à prendre des responsabilités dans des plates-formes informatique ou bioinformatique. Cette implication permet aux unités MIA d'être plus proches de la gestion des données et donc de pouvoir jouer un rôle plus important dans les grands enjeux finalisés. De plus, elle permet aux ingénieurs informaticiens de se confronter à des problèmes très motivants. Cependant un investissement trop important pose le problème de l'équilibre entre les activités de recherche et d'ingénierie dans MIA, surtout lorsque les plates-formes sont intégrées aux unités de recherche.

Nous proposons d'instruire ces questions, et au-delà celle du rôle que pourrait jouer MIA dans la structuration et l'animation du développement logiciel à l'INRA, conjointement avec la DISI-SI et de discuter les conclusions de cette réflexion lors d'un séminaire mi 2005 avec les décideurs de l'INRA (DS, CD, DAR, etc.) (lettre de mission à venir).

Les formes d'animation scientifique à mettre en place autour des différents thèmes de la modélisation seront instruites dans le cadre de la mission modélisation. MIA s'y investira de manière importante. C'est dans ce cadre que pourra se construire une expertise collective, amenée à se substituer à la consultation individuelle.