

Schéma stratégique du département MIA « Mathématiques et Informatique Appliquées »

2016-2020

Version synthétique – Février 2017

I. Contexte, missions et finalités du département MIA

Le développement sans précédent des sciences et technologies du numérique a un impact considérable sur les enjeux et les pratiques de recherche de l'ensemble des disciplines scientifiques de l'Institut, depuis les sciences du vivant et de l'environnement jusqu'aux sciences humaines et sociales. De nouveaux horizons s'ouvrent grâce aux capacités de traitement et d'analyse de données massives (le « big data ») [[#OpenScience-2](#)] et à la modélisation et la simulation de systèmes complexes par du calcul intensif [[#OpenScience-3](#)]. De plus, les technologies de communication (réseaux, internet, etc.) font évoluer la recherche vers des formes plus collaboratives et ouvertes à de nouveaux acteurs ; elles transforment les productions mêmes de cette recherche en connaissances partageables sur le web et en outils d'aide au conseil et à la décision accessibles en ligne [[#OpenScience-4](#)].

Ce développement du numérique prend sa pleine mesure car il s'accompagne d'un développement continu des méthodes mathématiques et informatiques nécessaires à la représentation et à l'extraction des connaissances, à la modélisation des processus, à la conduite des systèmes, à la résolution des problèmes. Porteur de disciplines – probabilités et statistiques, systèmes dynamiques, intelligence artificielle, automatique et contrôle, recherche opérationnelle – toutes aptes à lever des verrous méthodologiques déterminants pour les enjeux et priorités de l'INRA, le département MIA est le département de référence pour les aspects méthodologiques du développement numérique des recherches de l'Institut.

Dans ce contexte, MIA a en charge une double mission:

Mission I : *mener des recherches en maths-info et à leur interface avec d'autres disciplines, en prise étroite avec de grands enjeux de la recherche en sciences du vivant et de l'environnement.*

L'ambition du département MIA est d'être à la fois un acteur reconnu de la recherche sur des fronts de science en maths-infos, et un partenaire scientifique privilégié pour les autres départements de l'INRA sur les maths-infos appliquées aux sciences agronomiques au sens large. Les recherches méthodologiques et les recherches à l'interface sont conçues comme les deux pôles d'un continuum dans lequel les chercheurs et ingénieurs de MIA doivent savoir naviguer dans leurs propres travaux et via des collaborations, que ce soit à titre individuel ou à l'échelle de leur équipe ou unité.

Mission II : *accompagner le développement des maths-infos dans le contexte de la transition numérique, à l'INRA ainsi qu'auprès de ses partenaires ; en ce sens MIA offre une plus-value à l'Institut pour l'insertion de ses travaux dans la société.*

Cette mission inclut des actions d'ingénierie, d'expertise, de formation et de suivi des partenariats en maths-infos. Elle intègre des actions nécessaires pour accompagner la transition numérique à l'échelle de l'INRA, en lien avec le Délégué à la Transition Numérique (DTN). Le développement des mathématiques et de l'informatique est un phénomène général à l'INRA, et MIA doit y jouer pleinement un rôle d'animation scientifique, de conseil, et de coordination.

II. Dispositif

En 2017, le département MIA compte 135 agents titulaires, dont 65 chercheurs, 55 ingénieurs ou assistants-ingénieurs et 15 techniciens de la recherche. Il faut y ajouter les agents permanents de nos partenaires dans les Unités Mixtes de Recherche (UMR) et un grand nombre de doctorants et post-doctorants.

Les six unités de recherche de MIA sont situées à Avignon (**BioSP**), à Montpellier (**Mistea**), à Toulouse (**MIAT**) et en Île-de-France (**MaIAGE** à Jouy-en-Josas qui dépend également du département MICA, **MIA-Paris** à Paris 5ème et le **LaMME** à Evry). Le département MIA est également pilote de l'unité mixte de service **IFB-core** à Gif-sur-Yvette sous cotutelle INRA-CNRS-Inserm-CEA-Inria, et de la toute jeune unité d'appui à la recherche **Ingenum** basée à Toulouse. Par ailleurs, six chercheurs MIA sont affectés à des unités pilotées par d'autres départements de l'INRA, à Toulouse, Grignon, Lyon et Sophia Antipolis.

Les unités MIA regroupent des équipes de recherche, mais aussi deux plateformes de bioinformatique (Genotoul Bioinfo et Migale) et une plateforme de modélisation (Record)[[#OpenScience-1](#)].

En termes de partenariat, deux unités MIA sont des UMR sous la cotutelle d'une école d'agronomie : MIA-Paris avec AgroParisTech et Mistea avec Montpellier Supagro. Le LaMME est une unité sous contrat (USC), sous tutelle de l'Université d'Evry, du CNRS et de l'ENSIIE. Les unités franciliennes font partie de l'Université Paris-Saclay.

III. Axes méthodologiques

Les priorités scientifiques de MIA se déclinent selon deux grands axes méthodologiques liés aux données et à la modélisation, complétés par un troisième qui consacre l'importance croissante de la prise en compte d'objectifs finalisés explicites (prédiction, optimisation, conception, etc.) dans le développement de nouvelles méthodes.

AM1 : extraction de connaissances à partir de données

La science des données se développe actuellement en mixant les apports des différentes communautés scientifiques du numérique. MIA s'y investit pour répondre à la croissance des besoins et à la diversité des questions posées dans les sciences du vivant, en génétique et bioinformatique notamment, mais aussi dans les sciences travaillant à de plus grandes échelles.

Les priorités méthodologiques pour MIA portent sur :

- les représentations formelles et les traitements permettant d'extraire des connaissances de données issues de sources d'information multiples et hétérogènes (résultats expérimentaux, enquêtes, images, données numériques ou textuelles recueillies via des démarches de science participative ou sur le Web) ;
- les méthodes d'apprentissage en grande dimension, supervisé ou non, soulevant des questions de réconciliation de données, de réduction de dimension et de visualisation, sur des données ayant souvent une structure spatiale, dynamique ou hiérarchique ;
- la modélisation, l'inférence statistique et l'échantillonnage de processus spatio-temporels ou de processus ayant pour support des réseaux dynamiques, dans des contextes où les données sont fréquemment dispersées et hétérogènes.

AM2: modélisation intégrative de systèmes complexes.

La modélisation permet d'intégrer des mécanismes et processus, connus à des degrés de précision divers, dans la description mathématique ou informatique d'un système. Les approches pluri-disciplinaires sont incontournables, que ce soit en biologie des systèmes, en épidémiologie ou en agroécologie, sur des échelles allant pour MIA de la molécule au paysage agricole, en passant par l'individu et la communauté de populations.

Du point de vue des maths-infos, la complexité des systèmes étudiés nous conduit à mobiliser et coupler des approches diversifiées pour construire, analyser et implémenter des modèles, avec des recherches méthodologiques ayant pour priorités :

- la description dans des modèles mathématiques cohérents, déterministes ou stochastiques, de systèmes qui suivent des processus dynamiques et multi-échelles composés d'entités multiples, hétérogènes et en fortes interactions ;
- l'estimation des paramètres et l'évaluation de la qualité prédictive des modèles, en présence d'observations partielles et hétérogènes des variables d'état ;

- le couplage informatique de modèles issus de formalismes différents, conçus à plusieurs échelles de temps et d'espace, croisant des variables discrètes et continues ;
- l'exploitation des différents leviers du calcul intensif pour optimiser le rapport coût-précision des simulations (réduction de modèles, métamodélisation, calcul multi-fidélité, calcul distribué ou haute performance).

AM3: optimisation et conception à partir de modèles.

Du fait de la complexité croissante des problèmes abordés par la science des données et la modélisation, les méthodes guidées par des objectifs finalisés explicites prennent une importance croissante, qu'il s'agisse de mieux comprendre, de prédire, de contrôler ou de piloter des systèmes.

Les priorités méthodologiques de MIA pour répondre à ces objectifs finalisés portent sur :

- la quantification globale des incertitudes sur les composantes d'un modèle (qualité des données, variabilité et structure de dépendance des entrées, incertitudes sur les paramètres) et leur réduction par des méthodes statistiques et informatiques ;
- l'échantillonnage et la métamodélisation dans des espaces de grande dimension pour comparer par simulation des scénarios, optimiser une fonction de réponse ou résoudre des problèmes inverses sous incertitude ;
- les méthodes d'optimisation uni- et multi-critères et l'identification de critères adaptés aux problèmes réels, dans des contextes de forte variabilité et de grande dimension ;
- l'assimilation de données dans des modèles complexes, via des approches bayésiennes ou d'autres méthodes de statistique computationnelle.

Les trois axes AM1-AM2-AM3 ne doivent pas être dissociés. Les enjeux à venir se situeront en effet, très largement, dans la capacité à gérer les liens et les flux entre données, connaissances et modèles et dans la capacité à en exploiter efficacement les résultats par l'analyse et le calcul. L'ambition de MIA est de favoriser l'intégration et le croisement des démarches reposant sur les sciences du numérique au sens large, selon un processus dynamique apte à intégrer les innovations technologiques et conceptuelles.

IV. Champs thématiques et priorités transdisciplinaires

Les recherches menées au sein de MIA le conduisent à interagir avec l'ensemble des autres départements de l'INRA, contribuant ainsi à un grand nombre d'enjeux décrits dans les cinq #défis thématiques et les trois orientations de politique générale du document d'orientation #Inra2025. Elles sont structurées en trois champs thématiques (CT).

CT1- MIA-bio : bioinformatique et modélisation pour la biologie des systèmes et de synthèse

La bioinformatique, la génomique et la biologie des systèmes sont au cœur de ce CT. Initialement focalisées sur le génome et la compréhension du vivant à des échelles allant de la cellule à l'individu, les recherches menées à MIA se développent aujourd'hui de l'échelle moléculaire à celle de la population dans des écosystèmes microbiens, ou encore à celle de l'individu eucaryote en interaction avec son environnement. Les enjeux sont d'améliorer la compréhension des mécanismes en jeu et la capacité de les simuler, offrant ainsi la possibilité de maîtriser leur impact sur des fonctions d'intérêt. Pour cela MIA développe :

- des méthodes de représentation et d'analyse des données issues de la génomique et de la métagénomique, pour détecter et annoter des éléments fonctionnels [[#OpenScience-2](#)] [[#BioRes-2](#)];
- des méthodes d'extraction de connaissances, d'apprentissage et d'inférence sur les relations entre gènes, sur les réseaux de régulation et les réseaux métaboliques, à partir de sources d'information diverses et hétérogènes [[#OpenScience-3](#)];
- des approches de modélisation basées sur la biologie des systèmes et s'appuyant sur des données de génotypage et de phénotypage haut-débit, permettant de simuler des phénotypes microbiens, végétaux et animaux dans des environnements variés [[#3Perf-2](#)];
- des techniques d'inversion de modèles et d'optimisation pour des objectifs liés à la sélection génétique ou à la biologie de synthèse [[#BioRes-1](#)].

Adossée au CT1 et en lien avec le CT2 décrit ci-dessous, la priorité thématique PT1 porte sur l'exploitation de données génomiques et métagénomiques pour l'étude des écosystèmes microbiens. Il s'agit d'un domaine scientifique extrêmement dynamique, porteur d'enjeux importants et diversifiés (microbiote intestinal [#Food-2], du sol, pathobiomes, biofilms, bioprocédés, bioréacteurs, etc. [#3Perf-2]) et source de problèmes méthodologiques qui sont des fronts de science. MIA y est fortement impliqué depuis plusieurs années, en particulier dans le cadre du métaprogramme MEM. Les maths-infos apportent ici une valeur ajoutée incontestable, sur des champs d'application qui concernent la plupart des départements (MICA, ALIMH, CEPIA, EA, EFPA, PHASE, SA, SPE).

CT2 – MIA-pop : mathématiques et informatique pour la biologie des populations, l'écologie et l'épidémiologie

Le CT2 porte sur l'analyse et la modélisation intégrative de processus dans lesquels la biologie, la génétique et la dynamique des populations jouent un rôle majeur, sous l'influence de facteurs environnementaux et d'interventions humaines. Les objets d'étude incluent l'individu hôte d'un écosystème microbien (homme, animal, plante), à l'interface du CT1. Mais ils recouvrent aussi des échelles nettement plus larges telles que l'exploitation agricole, le parcellaire forestier, le réseau d'élevages ou la zone semi-naturelle. Les enjeux sont de comprendre, prédire et aider à la décision sur des systèmes complexes, pour contribuer in fine à une meilleure santé, à la maîtrise d'épidémies animales et végétales [#3Perf-2], au développement de l'agroécologie [#3Perf-1] ou au maintien de la biodiversité. Les recherches portent en particulier sur :

- l'analyse, la modélisation et la prédiction de systèmes dynamiques intégrant, possiblement, des interactions biotiques multi-échelles et multi-espèces, des processus de propagation et de diffusion dans l'espace ou dans des réseaux, des mécanismes évolutifs et adaptatifs [#OpenScience-3] [#Food-2];
- l'élaboration de méthodes quantitatives, basées sur la modélisation ou sur l'analyse de grands jeux de données, pour mesurer et gérer la biodiversité et les services agroécologiques qui lui sont associés [#Climat-3];
- le développement de méthodes pour l'épidémiologie-surveillance et le contrôle des populations de bioagresseurs dans les cultures, pour la préservation de la biodiversité dans les populations et les communautés, et pour l'adaptation de l'agriculture et des milieux semi-naturels au changement climatique [#3Perf-2].

Dans le cadre du CT2, en lien avec le CT3, la priorité thématique PT2 porte sur la modélisation de processus agroécologiques et son application à la conception et l'optimisation de services écosystémiques par l'agriculture [#3Perf-1]. Elle nécessite de modéliser, analyser et optimiser des systèmes complexes, à l'interface des sciences physiques, de la biologie, de l'écologie et parfois des sciences humaines et sociales. Le département MIA aborde ces questions en collaboration avec plusieurs autres départements (EA, PHASE, SPE, SA, EFPA, BAP, SAD, SAE2) et métaprogrammes (SMaCH, GISA, EcoServ).

CT3 – MIA-num : agriculture, environnement et alimentation numériques

L'application des sciences du numérique aux systèmes agricoles et alimentaires est un domaine en pleine croissance. Dans le cadre du CT3, MIA a l'ambition de développer des activités de recherche en informatique sur la gestion de données, l'extraction des connaissances et l'apprentissage à partir de sources d'information multiples et hétérogènes, en exploitant les ressources du stockage et du calcul distribué. Il s'agit également d'exploiter l'ouverture des données (climatiques, phénotypiques, environnementales) pour renouveler les approches de modélisation et d'aide à la décision à l'échelle de l'exploitation ou du territoire. Les activités du CT3 incluent des recherches sur :

- la récupération, l'organisation et l'apprentissage sur des données numériques et textuelles issues de sources d'information multiples et dynamiques (bases de données, objets connectés, sites Web), offrant de plus la possibilité d'individualiser les prédictions et les préconisations [#OpenScience-2];
- la modélisation de données spatio-temporelles, en particulier climatiques, permettant de quantifier et simuler la variabilité et la probabilité d'occurrence de valeurs extrêmes [#Climat-1] ;
- l'exploration et la comparaison de scénarios par expérimentation numérique, pour répondre à des enjeux de performance et de durabilité de systèmes agricoles et alimentaires [#3Perf-4][#Climat-2] ;
- l'évaluation quantitative de risques agro-environnementaux par des approches multi-critères, pour l'aide à la décision publique ou par des groupes d'acteurs [#3Perf-3].

Dans le cadre du CT3, la priorité thématique PT3 se focalise sur l'agriculture numérique. Face à des évolutions telles que le génotypage et le phénotypage hauts débits, l'internet des objets connectés ou les algorithmes d'apprentissage profond, l'implication de chercheurs en informatique et statistique est indispensable pour recueillir et traiter des informations massives et en extraire des connaissances fiables et exploitables [#3Perf-2]. Ce thème est mené en collaboration avec plusieurs départements (EA, PHASE, BAP, GA, SAD, SAE2, SPE) et se trouve au cœur de collaborations avec de nombreux partenaires, en particulier via l'institut de convergence #DigitAg.

De façon transversale aux trois CT, la priorité thématique PT4 porte sur les sciences participatives. Le "crowdsourcing", qui permet de démultiplier les moyens pour aborder une question de recherche, introduit par son principe même une grande hétérogénéité et une augmentation importante du volume et des flux de données. Les sciences participatives apportent des enjeux méthodologiques originaux en analyse et modélisation statistique. De plus, la nécessité de transférer au citoyen les moyens d'explorer lui-même les données ou les connaissances qu'il recueille, entraîne des enjeux informatiques en termes de plateformes ouvertes accessibles largement [#OpenInra-4].

Pour en savoir plus

Internet département MIA : <http://www.mia.inra.fr/>

Intranet département MIA : <https://intranet.inra.fr/mia>

Portail des maths-infos de l'Inra : <http://www.mathinfo.inra.fr/fr>