

Institut National de la Recherche Agronomique
Unité de Biométrie et Intelligence Artificielle
UBIA - UR875
Bilan scientifique 2005 – 2009

5 octobre 2009

Table des matières

1	Présentation et bilan général de l'unité	4
1.1	Présentation générale	4
1.1.1	Effectif et organigramme	4
1.1.2	Missions et contexte	6
1.1.3	Stratégie	7
1.2	Organisation et fonctionnement	8
1.2.1	Gouvernance	8
1.2.2	Communication interne	9
1.2.3	Animation au niveau de l'unité	9
1.2.4	Administration	9
1.3	Gestion des ressources	10
1.3.1	Ressources financières	10
1.3.2	Ressources matérielles	11
1.3.3	Gestion des compétences	11
1.4	Enseignement, formation par la recherche, ouverture internationale	12
1.5	Bref rappel de l'évaluation précédente	13
1.6	Hygiène et sécurité	14
2	Bilan par équipe	15
2.1	MAD	15
2.1.1	Modélisation et simulation de systèmes pilotés avec DIESE	16
2.1.2	RECORD - VLE	16
2.1.3	Méthodes d'exploration de systèmes complexes	17
2.1.4	Décision dans l'incertain et PDM	18
2.1.5	Analyse et décision dans les processus sur graphe	19
2.1.6	Approche PDM en écologie, épidémiologie, foresterie	19
2.2	SaAB	21
2.2.1	Localisation de QTL	21
2.2.2	Cartographie comparative et assemblage	22
2.2.3	Réseaux de contraintes pondérés et applications	23
2.2.4	Localisation de gènes d'ARN	24
2.2.5	Prédiction de gènes de protéines dans les génomes eucaryotes	25
2.3	Plate-forme bioinformatique GENOTOUL	26
2.3.1	SIGENAE	27
2.3.2	Evolution de l'infrastructure pour le traitement à grande échelle	27

2.3.3	BioMaj	28
2.3.4	Deux axes stratégiques de développement	28
2.3.5	Un catalogue de formations qui s'étoffe	29
2.3.6	Démarche qualité	30
2.3.7	Articulation avec l'équipe SaAB	30
3	Publications et autres productions de 2005 à 2009	31
3.1	Connaissances fondamentales : Publications scientifiques	31
3.1.1	Équipe MAD	31
3.1.2	Équipes SaAB & PFBI	41
3.2	Partenariats scientifiques, animation de communautés scientifiques	51
3.2.1	Organisation de congrès et conférences	51
3.2.2	Participation à des comités de programme de congrès et conférences	51
3.2.3	Participation à des comités éditoriaux de revues scientifiques	52
3.2.4	Création, animation de réseaux scientifiques et sociétés savantes	52
3.2.5	Participation à des comités scientifiques et jurys	53
3.2.6	Participation à des comités de pilotage de thèses	54
3.2.7	Directions d'ouvrages	56
3.2.8	Conception, coordination de programmes internes à un organisme ou nationaux	56
3.2.9	Rapports de conjoncture scientifique	56
3.2.10	Rapports d'évaluation, expertise scientifique	56
3.2.11	Relectures d'articles pour des revues et conférences internationales	56
3.2.12	Prix scientifiques obtenus	57
3.3	Développement de grands instruments pour la recherche	57
3.4	Connaissances opérationnelles et produits pour des acteurs socio-économiques	58
3.4.1	Bourses Cifre	58
3.5	Partenariats avec des acteurs socio-économiques	58
3.5.1	Logiciels déposés durant la période d'évaluation	58
3.5.2	Contrats de licence	58
3.6	Enseignement et Formation	58
3.6.1	HDR soutenue	58
3.6.2	Thèses soutenues	58
3.6.3	Conception et l'organisation de cursus d'enseignement supérieur	59
3.6.4	Formations assurées dans l'enseignement supérieur	59
3.6.5	Autres formations à destination de scientifiques et de professionnels	59
A	Pyramide des âges	62
B	Complément sur les ressources financières	63
C	Complément sur les enseignements dispensés	65
D	Implication des agents par activité	66
E	Rapport du comité pour l'évaluation de la période 2003-2006	67

Avant-propos

Ce rapport a été rédigé en 2009 pour l'évaluation par l'AERES (vague A) de l'unité de Biométrie et Intelligence Artificielle (UR875 de l'INRA). Cette unité comporte deux équipes de recherche (MAD et SaAB) ainsi qu'une plate-forme technologique de bioinformatique. Bien que l'évaluation des plates-formes technologiques n'entre pas dans le champ d'action de l'AERES, nous avons inclus la description de son bilan et de ses projets parce que certaines de ses activités sont imbriquées dans celles de l'équipe SaAB et parce que certains agents peuvent être à cheval sur la plate-forme et le reste de l'unité.

Chapitre 1

Présentation et bilan général de l'unité

1.1 Présentation générale

1.1.1 Effectif et organigramme

L'Unité de Biométrie et Intelligence Artificielle (UBIA) est une unité propre du département de Mathématiques et Informatique Appliquées (MIA) de l'Institut National de la Recherche Agronomique (INRA). Comme toutes les unités de MIA, l'UBIA a pour mission scientifique de développer et mettre en œuvre des méthodes mathématiques et/ou informatiques pertinentes pour résoudre des problèmes identifiés avec nos collaborateurs qui sont issus d'autres départements de l'INRA principalement. L'unité comporte actuellement (depuis janvier 2005) deux équipes de recherche : Modélisation des Agro-écosystèmes et Décision (MAD) et Statistiques et Algorithmique pour la Biologie (SaAB). Elle comprend d'autre part une équipe de service : la plate-forme de bioinformatique du GIS GENOTOUL (Génopole Toulouse Midi-Pyrénées).

Pour les aspects administratifs (personnel, budget, secrétariat, documentation) et d'informatique collective (matériel et logiciel), l'unité s'appuie sur le collectif ID&AU (Informatique, Documentation et Administration de l'UBIA) qui complète l'organigramme structurel.

A la fin de la période couverte par la présente évaluation (2005-2009), l'UBIA comprend 13 chercheurs (dont un professeur en délégation), 15 ITA, 7 doctorants (dont 4 basés dans d'autres unités où ils sont co-encadrés), 7 CDD, 1 chercheur post-doctorant. Sur cette période, 3 chercheurs sont arrivés (2 recrutements en 2008, 1 professeur en délégation) et 3 sont partis (1 mis à disposition, 1 en disponibilité, 1 en mutation à l'université). Pour les ITA, il y a eu 4 arrivées (2 mobilités, 2 recrutements) pour 3 départs (2 en disponibilité et 1 mutation vers son organisme d'origine). Trois ITA ont été recrutés en 2009 (2 IE et 1 TR) mais ne figurent pas sur l'organigramme parce que leurs dates de prise de fonction sont postérieures au 30 juin. Sont également absentes de l'organigramme les personnes rattachées administrativement à l'unité mais qui n'y sont plus physiquement ou dont les activités actuelles n'entrent pas dans le cadre couvert par la présente évaluation. Il s'agit de 3 CR (1 mis à disposition, 1 en disponibilité, 1 agent en situation d'handicap) et de 3 ITA (1 mis à disposition du GIS GENOTOUL, 2 en disponibilité). L'agent ITA mis à disposition du GIS GENOTOUL (Génopole Toulouse Midi-Pyrénées) collabore toutefois avec l'équipe administrative de l'unité du fait que cette dernière héberge le budget général de ce GIS (personnel, matériel), budget dont le Directeur de l'unité assume la responsabilité.

La fonction de directeur d'unité est remplie par Roger Martin-Clouaire depuis le 1er octobre 2006, date à laquelle il a succédé à Michel Goulard (actuellement mis à disposition d'une autre unité INRA).

La figure 1.1 donne l'organigramme de l'unité au 30 juin 2009. A noter la double appartenance d'un DR à l'équipe SaAB et à la plate-forme bioinformatique. Voir aussi la pyramide des âges en Annexe A.

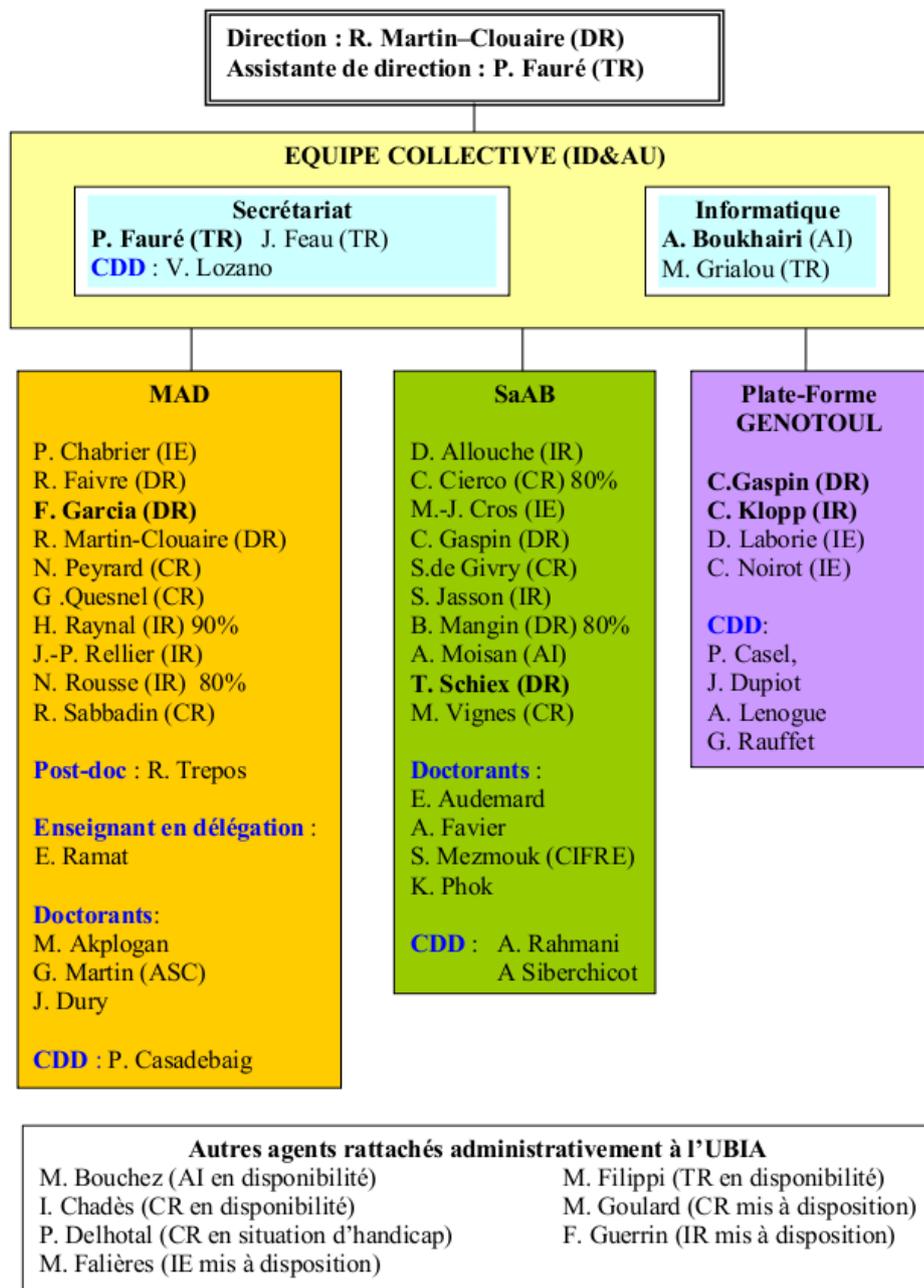


FIG. 1.1 – Organigramme de l'unité au 30 juin 2009. Les responsables d'équipes apparaissent en gras.

1.1.2 Missions et contexte

Relever le défi des sciences du vivant suppose de développer les relations entre calcul, modélisation et ces sciences (biologie, agronomie, écologie, sciences de l'environnement) afin de favoriser et d'accélérer les recherches exploratoires *in silico* sur des objets allant du gène à l'écosystème. L'informatique, la statistique et, plus généralement, les mathématiques sont indispensables pour organiser les données sur ces objets et en permettre l'accès, représenter les modes d'organisation et simuler les interactions au sein de systèmes complexes qu'ils sont ou dans lesquels ils s'inscrivent, ou concevoir des outils formels et calculatoires d'analyse de données, de prédiction et d'optimisation. Ces disciplines jouent un rôle fondamental dans la recherche de sens, dans la compréhension des fonctionnements des systèmes et dans le pilotage ou la gestion de ceux de nature anthropique. Ces besoins sont la source, pour des chercheurs en mathématiques et informatique, d'applications originales des outils de ces disciplines ainsi que de développements nouveaux. La mission de l'unité est de contribuer à apporter une réponse à ces besoins.

Chacune des équipes de recherche poursuit le même objectif général (développement d'outils mathématiques et/ou informatiques) mais pour des domaines finalisés bien distincts. Du fait des spécificités des questions et des objets d'étude, les outils méthodologiques développés ou utilisés par chacune sont variés. Chaque équipe de recherche a sa propre animation scientifique conduite par le responsable d'équipe.

Les développements méthodologiques menés par l'équipe MAD portent principalement sur les trois thèmes suivants :

- Modélisation et simulation des agro-écosystèmes, en particulier ce qui concerne les processus de décision anticipatoires et/ou réactifs en conduite de production ;
- Conception de stratégies de gestion des agro-écosystèmes. Il s'agit là de développer des méthodes d'optimisation permettant de répondre aux problèmes de configuration et de planification des agro-écosystèmes ;
- Approches spatialisées pour l'analyse et la conduite des agro-écosystèmes. Les recherches menées sur ce thème visent à développer ou adapter des modèles et méthodes statistiques ou algorithmiques pour représenter de manière compacte des processus spatio-temporels et optimiser les décisions dans leur dimension spatiale.

L'équipe MAD s'inscrit dans le second champ thématique du département MIA : Modèles et décisions pour l'environnement, l'agriculture et l'alimentation (CT2).

L'équipe SaAB est composée de statisticiens et informaticiens qui développent et mettent en œuvre des méthodes permettant d'analyser les volumes toujours croissants de données produits en biologie. Elle se focalise principalement sur des problèmes d'identification d'éléments fonctionnels dans les génomes de bactéries et de plantes (gènes d'ARN ou de protéines, régions liées à un caractère d'intérêt) en exploitant des données de séquence, d'expression et de génotypage/phénotypage dans un contexte de polymorphisme. Elle mobilise pour cela des outils statistiques (statistique inférentielle, mélanges, etc.) bien adaptés aux données bruitées de la biologie et des outils informatiques (principalement en optimisation combinatoire) bien adaptés au traitement des données discrètes de la biologie.

Les domaines finalisés abordés par l'équipe SaAB ainsi que la plate-forme bioinformatique se rattachent au champ thématique CT1 du département MIA : Bioinformatique, mathématiques et informatique pour les modèles du gène à l'individu.

La plate-forme de bioinformatique (labellisée IBISA) est une équipe de service ouverte sur l'extérieur de l'unité. Sa mission consiste à mettre à disposition de la communauté scientifique des ressources informatiques (matérielles, logicielles, banques de données) en vue de l'exploitation à grande échelle des données produites par les programmes scientifiques dans les domaines de la biologie et de la bioinformatique. Elle accompagne d'autre part les biologistes dans la mise en œuvre d'outils pour l'exploitation de leurs données. Cette équipe entretient des liens étroits avec l'équipe SaAB. Elle en était d'ailleurs une composante jusqu'en 2007 où il a été décidé d'en faire une entité distincte afin de lui donner une plus grande visibilité et de la mettre en cohérence avec la spécificité de sa mission, son environnement partenarial et, logiquement, sa stratégie et son fonctionnement.

Les socles scientifiques présents dans chaque équipe se recouvrent partiellement et permettent des échanges interdisciplinaires et inter-équipes au sein du séminaire de l'unité. Les compétences disciplinaires présentes dans l'unité couvrent un large spectre en statistique, probabilité, informatique, algorithmique, intelligence artificielle et sciences de la décision. L'unité poursuit clairement une double ambition de

production disciplinaire sur ses domaines de compétences et de production finalisée en construisant des projets et en collaborant avec des biologistes et des agronomes au sens large, ces projets se concrétisant souvent par la production d'outils logiciels. L'interaction avec nos partenaires peut être très forte, allant du processus de formulation des questions jusqu'à la validation des approches méthodologiques élaborées à partir de nos compétences disciplinaires. Les partenaires viennent de communautés différentes mais les stratégies suivies par chaque équipe de recherche sont similaires.

Par rapport aux priorités scientifiques que l'INRA a déclinées en 6 axes stratégiques (voir Orientation de l'INRA 2006-2009), l'équipe SaAB se rattache à l'axe D (Développer les recherches et produire les données génériques pour la connaissance du vivant) et l'équipe MAD intervient sur les axes E (Adapter les espèces, les pratiques et les systèmes de production agricole) et A (Gérer durablement et améliorer l'environnement, maîtriser les impacts des changements globaux et des activités productrices). Ce positionnement nous amène à avoir des projets en commun plus particulièrement avec des unités des départements EA (Environnement et Agronomie), SAD (Sciences pour l'Action et le Développement), EFPA (Ecologie des Forêts, Prairies, et Milieux Aquatiques), PHASE (Physiologie Animale et Systèmes d'Élevage), GA (Génétique Animale), GAP (Génétique et Amélioration des Plantes) et SPE (Santé des Plantes et Environnement).

L'activité de recherche du Centre INRA de Toulouse, est structurée en 3 axes. L'UBIA est fortement impliquée dans deux d'entre eux intitulés « Génome et Biotechnologie » et « Territoire et Produits ». Le premier axe étudie l'organisation et le fonctionnement de gènes d'animaux et de végétaux et, pour ce qui concerne l'UBIA, développe des outils bioinformatiques nécessaires aux approches génomiques. L'axe « Territoire et Produits » s'intéresse entre autres à l'ingénierie et à la gestion des productions agricoles, à l'utilisation raisonnée d'intrants, à la gestion/préservation des ressources naturelles et biologiques. Notre insertion se fait à travers le développement de méthodes et d'outils pour la compréhension des dynamiques des agro-écosystèmes et l'aide à leur conception. Sur ces deux axes, l'unité entretient des liens étroits plus particulièrement avec l'UMR441 Interactions Plantes-Microorganismes (dept. SPE, CNRS, Univ. Paul Sabatier), l'UR444 de Génétique Cellulaire (dept. GA), l'UR631 d'Amélioration Génétique des Animaux (dept. GA), l'UMR AGIR (dept. EA, dept. SAD, ENSAT). Le partenariat de l'unité au sein de l'INRA est aussi situé dans d'autres centres que Toulouse.

L'environnement scientifique toulousain (Université Paul Sabatier, Supaero, LAAS-CNRS, ONERA-CERT...) est bien sûr déterminant dans la vie scientifique de l'unité autour de ses socles disciplinaires et de ses domaines d'applications ciblés. L'unité partage des projets de recherche ou activités d'animation avec ces partenaires. Elle donne également des cours dans certains de ces établissements, le plus souvent au niveau Master.

1.1.3 Stratégie

Pour satisfaire leurs missions au sein de l'INRA relativement aux objectifs finalisés et pour maintenir les compétences méthodologiques et, par là, leur capacité à satisfaire durablement leurs missions, les équipes de recherche mènent ou sont impliquées dans des activités de :

- production de connaissances méthodologiques, publiées dans des conférences et journaux de statistiques ou d'informatique. L'originalité est ici un critère essentiel, plus que l'éventuelle applicabilité à des questions finalisées (jugée désirable). Pour cela, nous maintenons une activité sur un front de recherche bien identifié, où nous sommes visibles à un niveau international.
- production de méthodes bien caractérisées destinées à résoudre des problèmes bien identifiés avec nos partenaires biologistes et agronomes. Une bonne compréhension des problèmes, de leur contexte et de leur pertinence, ainsi qu'une bonne connaissance des réponses déjà apportées sont alors cruciales.
- production de connaissances biologiques ou agronomiques, en direction de journaux des communautés correspondantes se traduisant souvent par une activité de valorisation directe de données. Il peut s'agir de la simple mise en œuvre de techniques et outils bien connus ou développés par l'équipe. Dans le cas d'outils logiciels, cela peut permettre de donner une meilleure visibilité côté utilisateurs et un retour d'expérience bénéfique.

La production logicielle, souvent présente dans ces activités, s'inscrit aussi dans la stratégie de l'unité de diffusion de ses développements méthodologiques vers les biologistes et les agronomes. L'impact de

la résolution du problème, son caractère générique et la possibilité d'offrir une réponse originale sont les critères essentiels qui conditionnent le démarrage de tels développements. Ils sont parfois menés bien au-delà de l'état de simple prototype. Un outil abouti est produit et appliqué à des problèmes réels en collaboration avec des partenaires afin de vérifier sa pertinence. Ces logiciels diffusés sont souvent développés sur notre forge logicielle (collaborative) et rendus accessibles via une interface web. Leur utilisation sur le moyen terme soulève des besoins de maintenance évolutive et de formation gourmands en temps et en ressources humaines.

En ce qui concerne nos liens avec les structures locales régissant les études doctorales, l'UBIA était, avant le 1er janvier 2007, unité d'accueil de trois écoles doctorales : Mathématiques et Applications de l'Université Paul Sabatier, EDIT (Informatique et Télécommunications) de l'Université Paul Sabatier, et EDSYS (Systèmes) de l'INSA. Le cadre général des écoles doctorales a considérablement évolué ces dernières années (voir Journal Officiel du 24 août 2006) et nous a amenés à repenser notre positionnement. Nous avons fait le choix de nous associer à partir du 1er janvier 2007 à deux écoles doctorales de l'Université Paul Sabatier : MITT (Mathématiques, Informatique et Télécommunications de Toulouse, école issue de la fusion de Mathématiques Appliquées avec EDIT) qui correspond à nos socles disciplinaires et SEVAB (Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries) qui inclut les domaines scientifiques de nos thématiques finalisées. Afin de renforcer notre attractivité auprès des étudiants stagiaires et de ceux souhaitant s'engager dans des études doctorales nous intervenons dans des enseignements de masters irrigant ces écoles doctorales.

1.2 Organisation et fonctionnement

1.2.1 Gouvernance

Le directeur d'unité (DU) est responsable de la vie collective de l'unité, dans ses dimensions relationnelles, gestionnaires, organisationnelles et scientifiques. En s'appuyant sur le collège de direction et le conseil de service, il œuvre de manière à impulser, réorienter, coordonner et évaluer les actions menées par les équipes scientifiques et par l'équipe en charge des fonctions d'intérêt collectif. Pour la plate-forme bioinformatique, le DU assure essentiellement une responsabilité administrative, les missions de la plate-forme dépendant directement de l'INRA et du GIS GENOTOUL.

Par délégation du DU, chaque équipe est animée par un responsable. Son rôle est principalement de formuler et structurer les axes de recherche avec les autres membres de l'équipe et de s'assurer du bon déroulement des différents projets en cours. Au sein de son équipe, il doit aussi veiller au maintien d'une dynamique scientifique, en particulier au renouvellement ou à l'enrichissement de la problématique de recherche. Il anime les réunions d'équipe et coordonne la réalisation des documents demandés. Enfin il joue un rôle d'interface avec le département où il représente l'unité dans le champ thématique correspondant.

Le collège de direction, constitué du directeur d'unité, de son assistante administrative et des responsables d'équipe, a été mis en place en 2003 suite à une proposition de la commission d'évaluation de l'unité. Il se réunit une à deux fois par mois pour discuter de questions relatives à la politique scientifique de l'unité et des problèmes courants à traiter.

Le conseil de service a lui aussi été mis en place en 2003 après l'évaluation de l'Unité. Il est nommé pour 2 ans, se réunit en fonction des circonstances (au moins 2 fois par an) et est composé de 11 personnes : 3 membres de droit (le DU et son assistante, le responsable de l'équipe informatique), 6 membres élus (3 chercheurs, 3 ITA), un représentant des non-permanents et l'agent chargé de la prévention qui y est un invité permanent. Le conseil a un rôle consultatif et d'aide à la décision auprès du DU. Ce conseil permet d'aborder des questions susceptibles d'avoir une incidence sur les conditions d'exercice du travail, relatives à la vie de l'unité, à son organisation et à son fonctionnement tels que les moyens budgétaires et matériels, la gestion des ressources humaines (plan de formation, accueil des non permanents), la gestion des locaux (entretien) ou les questions d'hygiène et de sécurité. L'ordre du jour intègre les demandes des agents de l'unité et est communiqué par courriel à l'ensemble des agents. Les dossiers sont instruits par l'assistante du DU qui s'appuie sur l'équipe collective et l'agent chargé de la prévention. Le relevé de conclusions de chacune des réunions est diffusé sur l'intranet de l'unité.

1.2.2 Communication interne

Depuis le 1er octobre 2006, l'unité a adopté le principe d'une réunion d'information hebdomadaire (mardi à 10h15) consacrée à la communication d'informations diverses (synthèse de réunions, annonce de visites, présentation des nouveaux arrivants, rappel de règles et échéances administratives...). Ces réunions, brèves (15-45mn) car régulières, s'avèrent efficaces bien que moins formelles que les traditionnelles assemblées générales. Elles font l'objet d'un court compte rendu, diffusé sur l'intranet de l'unité. Ces réunions donnent aussi lieu à des discussions qui, si le sujet le nécessite, entraînent la programmation d'une réunion plénière spécifique.

La communication au sein de l'unité se fait aussi beaucoup par voie électronique : courriel, portail intranet (affichage de documents divers, de l'agenda des déplacements, visites et congés, réservation du matériel collectif...).

1.2.3 Animation au niveau de l'unité

Le séminaire de l'UBIA est un forum régulier, hebdomadaire, dans lequel interviennent les membres du laboratoire mais également des scientifiques extérieurs à l'unité. Ce séminaire a pour but de présenter des travaux en rapport avec les activités ou préoccupations de l'unité, dans un esprit de diversité disciplinaire, méthodologique et de domaines finalisés. Outre la présentation et la discussion des travaux de recherche en cours (thèses en particulier), le séminaire sert aussi de cadre de formation (tutoriel) sur des méthodes, techniques ou outils intéressant plusieurs membres de l'unité. Il est aussi le lieu privilégié d'échanges scientifiques entre les équipes.

En plus du séminaire, l'unité organise chaque année une assemblée générale appelée « journées d'unité ». Ce rendez-vous donne l'occasion de discussions, organisées thématiquement dans ce cas, en mettant par exemple l'accent sur la dimension stratégique des choix scientifiques de l'unité ou sur l'ingénierie informatique dans ses projets scientifiques.

Une journée « stagiaires » a lieu annuellement pour permettre d'informer l'ensemble de l'unité sur les travaux associés aux différents stages et de familiariser les stagiaires avec la pratique de présentation orale.

L'offre de séminaires scientifiques sur le centre INRA (AGIR, LIPM, LGC, SAGA...) et dans les divers établissements toulousains (LAAS, IRIT, CERT, Université de Sciences Sociales, UPS Institut de Mathématiques) est riche. Nous participons également, en qualité d'auditeurs et d'orateurs à ces séminaires.

1.2.4 Administration

Le collectif administratif de l'unité a vu s'accroître, comme partout ailleurs, l'importance de son rôle de support à l'activité de recherche. La charge de travail incombant à ce collectif est de plus en plus lourde du fait de l'augmentation des volumes à traiter (contrats et personnels accueillis en particulier), et de plus en plus complexe du fait de la multiplicité et de la sophistication des nouveaux logiciels à utiliser (S2I-finance, S2I-RH, logiciels de réservation) qui sont réputés exigeants et chronophages. Au-delà des actes de secrétariat et de gestion, ce collectif assure de manière permanente un rôle d'interface avec les SDAR (Services locaux d'Appui à la Recherche), le département MIA et l'INRA au niveau national, ainsi que d'assistance et de conseil aux agents et aux partenaires. En adaptant sans cesse leurs méthodes de travail et en mutualisant leur force de travail avec beaucoup d'abnégation, les gestionnaires de l'unité ont su et pu faire face aux changements. Il n'en reste pas moins que l'énergie dont elles ont fait preuve au cours des dernières années a désormais trouvé ses limites et un recrutement permettant de compenser la perte d'un ETP (et donc de maintenir l'effectif du collectif à 3 ETP) est nécessaire pour le maintien de son efficacité.

1.3 Gestion des ressources

1.3.1 Ressources financières

Les ressources de l'unité (voir la figure 1.2) sont composées de la subvention d'état majorée des recettes propres (colloques, expertises) et des ressources contractuelles. La subvention d'état qui est allouée à l'unité via le département MIA, est calculée en parts chercheurs. Cette subvention est stable sur la période de référence. Elle est attribuée pour les agents de catégorie A, avec une moyenne de 22,84 parts chercheurs par an représentant 32% des ressources unité ou 57% de la Subvention d'Etat.(voir la figure B.1 en annexe B).

En moyenne sur la période d'évaluation, les ressources dont dispose l'unité représentent 612 K€ ; elles sont réparties en :

- 320 K€ de subvention d'état. Les montants des années 2006 et 2009 sont induits par un financement exceptionnel « équipement lourd » pour l'investissement dans le cluster sur la plate-forme bioinformatique.
- 292 K€ de ressources contractuelles réparties entre les différents contrats de recherche de l'unité (ANR, Europe, privé). Le budget connaît une forte croissance en 2006, coïncidant avec la mise en place de l'ANR. Il est stabilisé depuis 2007.

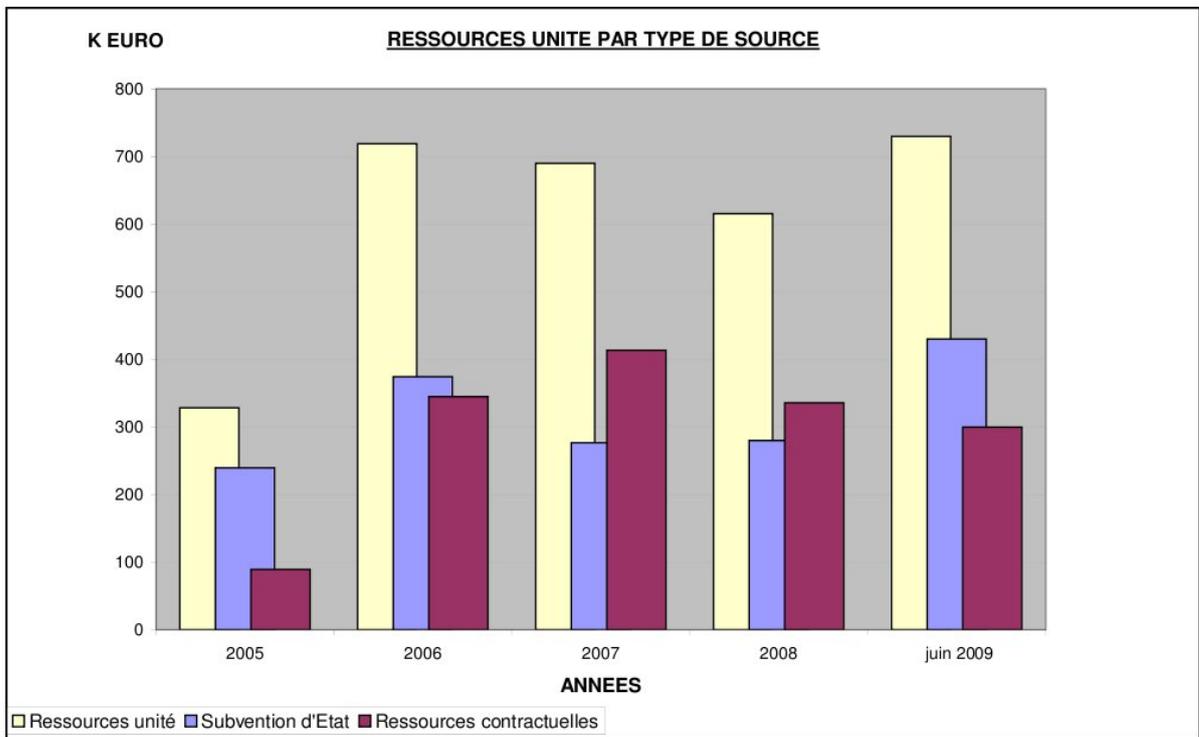


FIG. 1.2 – Ressources financières de l'unité

L'augmentation du nombre de contrats induit un accroissement substantiel du travail de gestion que doit assurer l'équipe administrative.

La gestion des ressources liées à la subvention d'état se fait à partir d'un prévisionnel défini en début d'exercice, qui est ajusté et complété grâce à un suivi mensuel et à un bilan annuel, validé par le DU.

L'ensemble des ressources est géré avec l'outil de gestion institutionnel, S2I finance, et des tableaux de bord.

Enfin, l'unité BIA a la particularité d'héberger administrativement le GIS GENOTOUL. Les fonds de cette entité sont positionnés sur la ligne budgétaire de l'Unité et ensuite reversés aux partenaires institutionnels du GIS. Les fonds restants servent exclusivement au fonctionnement de la structure adminis-

trative du GIS. Le budget global géré par l'Unité est donc composé des ressources du GIS GENOTOUL et des ressources dédiées à l'unité (voir la figure B.2 dans la Section B de l'annexe).

1.3.2 Ressources matérielles

Un AI et un TR gèrent les ressources informatiques et électroniques partagées de l'unité. L'infrastructure dédiée à l'analyse à grande échelle des données mise à disposition par la plate-forme bioinformatique est gérée indépendamment de l'équipement collectif de l'unité par un IE (voir description de la plate-forme).

La mission des 2 agents gérant l'équipement partagé de l'unité est de garantir l'accès aux ressources informatiques et électroniques dans un souci d'efficacité, de sécurité et de gestion responsable en terme d'empreinte environnementale. Les ressources se doivent de supporter la mission de recherche et de diffusion, en plus de supporter ses fonctions administratives et de services. Les ressources concernées incluent notamment 10 serveurs physiques, 43 postes fixes de travail (PC), 25 portables, 2 imprimantes laser (1 couleur et 1 NB), un équipement de visioconférence et un onduleur (30kVA). Le système informatique de l'unité est centré autour d'un serveur de fichiers qui héberge toutes les données des utilisateurs. La pérennité de ces données est assurée via un système constitué d'un robot de sauvegarde (30 Bandes de 400G). Le système d'authentification (LDAP) est lui aussi centralisé ce qui permet aux utilisateurs de l'unité d'accéder aux différents services proposés à l'aide d'un seul mot de passe. La partie acquisition de documents et reproduction est couverte par un scanner-photocopieur authentifié via le système LDAP de l'unité. Les infrastructures de communication, internes au centre INRA (réseau Ethernet) et avec l'Internet (accès au réseau RENATER et borne de connexion WIFI) sont gérées par le service informatique du centre INRA. Il en est de même pour la messagerie. Une partie de l'acquisition d'équipement informatique se raisonne au niveau du département MIA qui dispose d'une cellule d'instruction des besoins et d'allocation de moyens pour cet usage. L'évolution des ressources informatiques ces dernières années a été marquée par :

- un effort d'homogénéisation du parc (PC plutôt que terminaux) et des systèmes d'exploitation (Linux Ubuntu sur tous les postes à l'exception des quelques machines utilisées pour les outils de bureautique et pour les logiciels administratifs sous Windows) et des logiciels ;
- une réduction du nombre de serveurs physiques par partitionnement de certains serveurs physiques en plusieurs serveurs virtuels indépendants (par virtualisation 20 serveurs sont obtenus à partir des 10 serveurs physiques, contribuant ainsi à la réduction de l'empreinte écologique) ;
- l'introduction d'un collecticiel (groupware) pour améliorer la communication interne. Il comprend un Wiki Mediawiki (documentations diverses), un blog Wordpress (événements), un agenda individuel et collectif, et un système de gestion des ressources (salle, portables, etc.).

L'unité héberge la forge logicielle du département MIA afin de promouvoir le travail collaboratif sur les développements logiciels du département MIA et de partenaires externes. Pour cela, le logiciel de TCAO (Travail Collaboratif Assisté par Ordinateur) Fusion Forge a été mis en oeuvre pour remplacer la précédente Forge Express Edition et migrer vers une version Open-Source. La forge héberge actuellement 67 projets et 140 utilisateurs (dont 80% extérieurs à l'unité). Une équipe de gestion de la forge a été mise en place. Elle comprend les 2 administrateurs de l'unité et 2 autres agents du département.

Outre le matériel informatique, l'unité dispose de 3 véhicules de service et d'un fond documentaire géré au sein d'une bibliothèque partagée avec trois autres unités du centre INRA.

Enfin, l'Unité est entrée dans une politique d'achat responsable et durable : (i) en privilégiant l'achat selon les préconisations fixées par les marchés nationaux, (ii) en faisant le choix de mobilier qualifié ISO9001 et certifié PEFC (soutien à la gestion durable des forêts et respect de l'environnement), (iii) sans oublier l'approche sociale en apportant une contribution à des Centres d'Aide par le Travail (CAT) par des commandes annuelles.

1.3.3 Gestion des compétences

Comme dans toutes les entreprises publiques ou privées, tout agent de l'unité a la possibilité de se former dans différents domaines tout au long de sa carrière. Un « correspondant formation » nommé au sein de l'unité contribue à faire émerger les besoins de formation, et recueille les besoins collectifs et indivi-

duels. Un plan de formation de l'unité (PFU) est alors rédigé. Une fois élaboré, ce document est présenté en conseil de service et validé par le DU avant d'être transmis aux services de formation permanente de l'INRA. La responsabilité politique du plan de formation incombe au directeur de l'unité. Le « correspondant formation » suit la réalisation des actions de formation programmées dans le PFU. Il assure auprès du personnel de l'Unité la diffusion de l'information relative à la formation, l'accompagnement à la définition d'un besoin individuel ou collectif et est en relation avec les « responsables formation » institutionnels de centre. Il est également en relation avec des prestataires extérieurs pour spécifier les besoins et définir le programme correspondant. Toutes ces démarches sont faites en collaboration avec le service formation permanente du Centre. Le plan de formation poursuit plusieurs objectifs :

- adapter les personnels à leur poste de travail,
- contribuer aux évolutions des métiers,
- acquérir de nouvelles compétences,
- soutenir le projet professionnel des agents.

Ce plan est la rencontre entre les attentes exprimées par les agents et les besoins exprimés par l'institution (par le DU lors des entretiens biannuels avec les ITA ou par l'INRA qui propose des formations par exemple pour les nouveaux outils administratifs ou pour la gestion d'unité, d'équipes ou de projets). Le tableau en annexe 3 liste les formations suivies, leur durée en heures et le nombre d'agents concernés sur la période 2005-2009. En moyenne, 55% de l'effectif total du laboratoire a suivi une formation, soit 15% de chercheurs, 29% d'ITA et 12% de non permanents. Le volume d'heures de formation annuel moyen sur 5 ans est de 869 heures/an (soit environ 0.5 ETP par an) représentant un budget moyen de 2500€ à la charge de l'Unité. A noter qu'un certain nombre de formations sont financées sur le budget de la formation permanente (FP) :

- soit en co-financement (35% pour l'Unité et 65% pour la FP) : il s'agit essentiellement des formations prévues au PFU ;
- soit à 100% pour les agents de catégorie C, les formations aux outils institutionnels, les formations « management ».

A ce volume d'heures de formation, il faut ajouter les écoles chercheurs dans lesquelles s'investissent les chercheurs et les doctorants, les tutoriaux suivis lors des différentes conférences nationales ou internationales et la participation des doctorants à un dispositif interne à l'INRA, « Les Réflexives », auquel participent 2 doctorants et leurs encadrants, en moyenne par an.

La gestion des compétences comprend également le transfert des connaissances, des savoirs et l'accompagnement technique entre les agents de l'Unité, à l'occasion de la prise en main d'un nouvel outil, d'un recrutement de personnel temporaire ou permanent, de l'anticipation d'un départ en retraite. Cette démarche fait partie de la politique d'intégration de l'Unité.

1.4 Enseignement, formation par la recherche, ouverture internationale

Implication de l'unité dans l'enseignement Afin de répondre à notre mission d'enseignement et pour attirer les étudiants, nous dispensons en moyenne 143 heures de cours par an, du niveau IUT à Master 2, dans des universités (Université Paul Sabatier de Toulouse, Université du Littoral de Calais) et des grandes écoles (INSA de Toulouse, Supaéro de Toulouse, ESC de Rennes). Voir la section 3.6.4 et la figure C.1 de l'annexe C. Ces interventions sont assurées, en moyenne, par 5 chercheurs et 3 ingénieurs.

Cette implication nous permet également d'être visibles et reconnus par les établissements d'enseignement. Nous sommes ainsi responsables du module « Modèles probabilistes et algorithmes pour la bioinformatique » au sein du M2R Informatique et Télécommunications (Parcours Intelligence Artificielle) de l'Université Paul Sabatier. C'est sans doute en grande partie grâce à cette implication volontaire dans l'enseignement local que nous avons pu recruter en Master J. Vandel qui bénéficie d'une thèse ministérielle de l'école doctorale MITT. Nous intervenons également dans le module "Rédaction scientifique" de l'école doctorale SEVAB.

En plus de son investissement dans l'enseignement supérieur, l'unité organise des formations (voir Section 3.6.5) pour favoriser la diffusion et l'appropriation des logiciels qu'elle développe. Ces formations visent leurs usagers réels ou potentiels.

Formation par la recherche La liste des thèses et mémoires figure dans la liste des productions (voir Section 3.6). Neuf thèses encadrées ou co-encadrées par un chercheur de l'unité ont été soutenues pendant la période couverte par l'évaluation, 7 sont en cours au 30 juin 2009 et 2 débiteront à l'automne 2009.

Nous sommes d'autre part fortement sollicités (voir Section 3.2.6) pour nos compétences en mathématiques et informatique dans les comités de pilotage de thèse institués par l'INRA pour apporter un regard extérieur et un suivi du bon déroulement de la thèse. Typiquement, un comité de thèse est constitué de 3 ou 4 personnes qui se réunissent 1 fois par an pendant la durée de la thèse. Le comité examine l'avancement du travail présenté par le doctorant et conseille sur les questions pendantes.

L'unité a été fortement impliquée dans la formation doctorale via la participation de 3 de ses membres en tant que co-animateurs des séminaires Réflexives de l'INRA (<http://www.reflexives-lpr.org>). Ces séminaires permettent aux doctorants de travailler sur la construction et la communication de leur projet de thèse avec leur encadrant, en présence de co-animateurs comme personnes ressources. Enfin, c'est un chercheur de l'unité qui assure la responsabilité de l'animation des doctorants au niveau du département MIA.

L'unité accueille en moyenne 4 stagiaires de niveau M2 par an. Le recrutement de doctorants se fait de plus en plus parmi les stagiaires. La formation s'inscrit d'autre part dans notre mission de diffusion de l'information et de la culture scientifique et technique. Nous accueillons ainsi des stagiaires de M1 et L3. L'ensemble de ces stages représente en moyenne 45 mois par an d'encadrement.

Ouverture internationale de l'unité Dans la période 2005-2009, l'ouverture internationale de l'UBIA s'exprime sous diverses formes : congrès internationaux (présentations invitées, tutoriels, communications scientifiques) (voir Section 3.2.2.1), participation à des comités d'édition et de programme (voir Section 3.2.2.1), accueil de chercheurs et de post-doctorants étrangers (Canada, Australie, Tchéquie, Espagne), séjour post-doctoral à l'étranger (Australie, Ecosse), co-encadrement et co-financement d'une thèse²⁸⁶ avec une université suédoise, participation à des projets européens et à des projets bilatéraux avec l'Espagne, l'Irlande et l'Australie.

L'unité s'est également investie dans la conception d'un projet d'action COST, porté à l'origine par le département MIA, BIOS au Royaume-Uni et Biometris au Pays-Bas. Elle continue son investissement dans cette action, dénommée STATSEQ (<http://www.statseq.org>), par sa place de président-adjoint et d'adjoint chargé des séjours scientifiques entre les pays membres.

1.5 Bref rappel de l'évaluation précédente

La précédente évaluation collective de l'UBIA a été réalisée le 23 mars 2007 par la commission composée de : Sovan Lek (Président, Professeur à l'Université Paul Sabatier de Toulouse, UMR 5174 CNRS-UPS), François Charpillet (DR INRIA LORIA Nancy, projet MAIA), Antoine de Daruvar (Professeur à l'Université de Bordeaux 2, Centre de Bioinformatique de Bordeaux), Olivier François (Professeur à l'Université Joseph Fourier de Grenoble, UMR 5525 CNRS-UJF), Marianne Le Bail (Professeure UMR SADAPT INRA-AgroParisTech). Les points forts de l'unité relevés par la commission constituent la conclusion de son rapport (voir Annexe E) : « le comité est extrêmement satisfait du bilan du laboratoire, et cela aussi bien pour les aspects scientifiques que pour les aspects humains. Il approuve et soutient le projet de renouvellement, dans la composition des équipes et dans ses projets scientifiques. Finalement, le comité apprécie l'énergie de la direction à monter, soutenir et développer le laboratoire qu'elle anime, et constate qu'elle lui insuffle par là beaucoup de cohérence et de dynamisme. Le comité lui apporte toute sa confiance. » La commission attire l'attention de l'unité sur un certain nombre de points. Nous les rappelons ci-dessous et mentionnons, s'il y a lieu, les effets déjà observables des ces recommandations.

- « Augmenter le recrutement doctoral » en élargissant la recherche de financement (bourses CIFRE, BDI, Région), afin de mieux valoriser le potentiel d'encadrement de l'unité. Le nombre de doctorants a sensiblement augmenté en 2009 ; l'un d'eux a une bourse CIFRE.
- « Encourager les co-publications co-signées par plusieurs membres de l'unité ». Il ne nous a pas paru important de mettre en place des mesures particulières pour aller dans ce sens.

- « Privilégier pour SaAB une politique de développement logiciel autour de projets collectifs d'envergure ». Si aucun des projets logiciels démarrés ou poursuivis durant la période n'a monopolisé toute l'équipe, certains projets logiciels (comme DARN!, inclus dans RNASpace) ont fédéré plusieurs personnes, en croisant les compétences.
- « Favoriser la valorisation par la plate-forme bioinformatique des travaux de recherche de l'équipe SaAB ». La plate-forme bioinformatique sert de support de diffusion à différents outils, tels qu'Eu-Gène'Hom ou FrameD et son infrastructure matérielle a été mobilisée dans différents projets d'envergure (annotation de génomes eucaryotes, mise en œuvre de MCQTL). Elle a également inclus DARN!, issu d'une thèse faite dans SaAB, dans son projet RNASpace, de support logiciel à la prédiction et annotation d'ARN non codants.
- « S'attacher à la valorisation des projets logiciels (RECORD, DIESE) de l'équipe MAD ». L'équipe MAD a fortement investi dans des projets de valorisation de RECORD et DIESE à travers des formations, projets applicatifs et publications.
- « Maintenir l'équilibre entre les travaux méthodologiques et à vocation finalisée, et mettre plus en avant les aspects finalisés ». La préoccupation de maintien de l'équilibre méthodologie-application n'a jamais été délaissée même si elle conduit à une gymnastique intellectuelle difficile.
- « Utiliser les moyens de recrutement pour favoriser la cohésion interne au sein de chaque équipe et les passerelles scientifiques entre équipes ». Notre dernier recrutement de CR2 apporte vitalité au thème « génomique génétique » nouvellement créé et favorise la cohérence sur le thème « modèles graphiques stochastiques » qui intéresse des chercheurs des deux équipes de recherche.

1.6 Hygiène et sécurité

L'organisation des actions concernant l'hygiène et la sécurité dans l'unité est confiée à un agent chargé de prévention (ACP) sous la responsabilité du DU, en lien avec le délégué prévention du centre. Cette mission est remplie depuis octobre 2006 par Annick Moisan, membre de l'équipe SaAB. Afin de diffuser un esprit de prévention, les différents rôles en prévention sont répartis sur plusieurs agents : 2 sauveteurs secouristes du travail (SST) recyclés tous les ans, 1 garant ASAI (Analyse et Suivi des Accidents et des Incidents), 8 guides d'évacuation et serre-files, 1 personne habilitée électrique B0.

Les deux accidents du travail survenus durant la période considérée sont des chutes en deux roues lors de trajets domicile-travail. L'un d'eux a fait l'objet d'une semaine d'arrêt de travail.

Les agents de l'unité travaillant essentiellement dans des bureaux sur poste informatique, outre les risques routiers pour les trajets domicile-travail, les risques identifiés sont : les troubles musculo-squelettiques liés au poste de travail sur écran, les chutes de plain pied lors de la circulation dans les couloirs, les incendies et l'électricité.

Un plan de prévention annuel est proposé au conseil de service pour développer prioritairement certains aspects de la prévention. Ainsi l'accent a été mis successivement sur l'ergonomie du poste de travail (communication en séminaire d'unité du médecin prévention et étude de 3 postes), la formation (3 agents à la « Méthode générale d'évaluation des risques », 1 agent aux risques électriques, 8 personnes à la sécurité incendie, 1 agent à la formation initiale des ACP), la communication (affichage, accueil des nouveaux,...) et l'organisation de la gestion des déchets.

La montée en puissance du parc informatique pourrait entraîner l'augmentation de la température dans les salles machines, faisant croître les risques incendie. Une réflexion est en cours au niveau du centre sur la sécurisation des dispositifs informatiques.

Chapitre 2

Bilan par équipe

Ce chapitre présente les activités et les résultats principaux obtenus par les deux équipes scientifiques et l'équipe plate-forme sur la période d'évaluation. Afin d'éviter des redondances avec le reste du dossier d'évaluation (formulaire AERES pour le bilan, liste des publications et autres productions), ces points ne sont pas repris ici.

2.1 MAD

L'équipe MAD (*Modélisation des Agro-écosystèmes et Décision*) est composée actuellement de 7 chercheurs et 4 ingénieurs. Les activités de recherche de l'équipe portent sur le développement de modèles et méthodes mathématiques et informatiques pour l'analyse et la conduite des agro-écosystèmes, systèmes définis par l'interaction entre des activités agricoles et des ressources naturelles ou biologiques, aux échelles de la parcelle, de l'exploitation agricole ou du territoire, de la décision individuelle à la gestion collective, pour des horizons temporels allant de la journée à plusieurs décennies. Ces travaux méthodologiques reposent sur des collaborations finalisées avec des équipes de recherche en agronomie, épidémiologie, sylviculture, écologie, etc. dont les problématiques portent sur les organisations et modes de gestion innovants des agro-écosystèmes face au contexte actuel de changement global (démographie, climat, etc.). Nos productions prennent la forme de publications scientifiques, d'ouvrages de synthèse, de supports de cours ou encore de logiciels.

Les méthodologies manipulées et développées dans le cadre des travaux de l'équipe MAD portent de manière cohérente sur la modélisation, la simulation, l'exploration, l'optimisation de systèmes dynamiques représentant des agro-écosystèmes pilotés. Plus particulièrement, les compétences méthodologiques de l'équipe peuvent être regroupées selon trois axes forts en informatique et statistique algorithmique : i) modélisation / simulation de systèmes dynamiques, ii) optimisation de la décision dans l'incertain et iii) modèles spatiaux sur graphe. Ces compétences collectives de l'équipe, qui résultent d'une stratégie scientifique mise en place lors de sa création en 2005, nous permettent de répondre à une large gamme de problèmes de modélisation et décision pour des agro-écosystèmes variés, dans les domaines de l'agriculture et l'environnement, l'écologie, l'épidémiologie, la gestion forestière.

L'équipe a une bonne visibilité scientifique, tant en terme méthodologique, avec en particulier ses travaux sur les processus décisionnels de Markov ou sur la simulation à événements discrets, que finalisé, comme l'attestent ses nombreuses collaborations nationales ou internationales avec des équipes scientifiques de premier plan, sa participation à de nombreuses formations au niveau masters et écoles doctorales, son implication dans des projets de développement logiciel à haute visibilité. Deux HDR ont été soutenues par des membres de l'équipe, une en 2004 et une en 2009¹⁴⁶.

Nous présentons ici les résultats marquants attachés aux activités de MAD pendant la période couverte par l'évaluation. Ces résultats sont de plusieurs natures : méthodologies, applications et diffusions, animations. La liste des agents impliqués dans chacune des activités est donnée dans l'annexe D.

2.1.1 Modélisation et simulation de systèmes pilotés avec DIESE

L'objet de ce travail est de développer et mettre à disposition un cadre de représentation et de simulation de systèmes de production en portant une attention particulière à la modélisation de l'agent décideur en charge du pilotage de la production⁵². Ce cadre a été pensé pour des systèmes de production agricoles pouvant aller jusqu'à l'exploitation agricole dans sa globalité qui, compte tenu du rôle qu'y jouent les facteurs incontrôlables (climat en particulier), nécessitent des conduites flexibles (dépendantes de la situation et adaptatives). L'étude des pratiques agricoles devient en effet un courant nouveau de la recherche agronomique et génère le besoin d'outils pour les évaluer, en concevoir de nouvelles et faciliter la compréhension des interactions entre elles et avec les systèmes biophysiques sur lesquels elles opèrent.

Le résultat rapporté ici concrétise un travail entamé bien avant la période couverte par cette évaluation⁵⁶. Le cœur est constitué par un formalisme original de représentation générique basé sur une ontologie des systèmes de production qui constitue un méta-modèle à partir duquel peuvent être construits des modèles de système en réutilisant et en particulierisant des concepts de haut niveau (e.g. activité, opération, ressource). Le formalisme est implémenté dans le cadre DIESE qui comporte aussi un moteur de simulation à événements discrets, des outils d'accueil assisté des connaissances et de génération (partielle) de code compilable. Outre l'achèvement de l'implémentation, et la publication du cadre dans une revue d'informatique³⁵, le résultat principal concerne la diffusion de DIESE auprès d'utilisateurs avec qui nous avons co-développé des modèles de systèmes très divers (élevage, viticulture, grandes cultures) qui permettent de tester la généralité du cadre et la faisabilité d'une utilisation par des agronomes. Cinq projets applicatifs d'envergure ont vu le jour. Quatre d'entre eux supportent des travaux de thèses dont trois au sein de deux projets ANR (TRANS^{19;55} et SPADD^{45;104}), l'autre projet faisant l'objet d'un post-doc⁴⁶. Ces projets applicatifs jouent aussi un rôle essentiel dans la diffusion de notre méthodologie de conceptualisation de système de production auprès des chercheurs agronomes qui sont souvent désarmés dans ce type d'exercice et aussi des organismes professionnels (instituts techniques) impliqués dans le financement ou les comités de pilotage des thèses. Enfin, ils sont un vecteur essentiel pour la visibilité nationale et internationale de notre production méthodologique, d'où notre participation massive au congrès Farming Systems Design^{46;54;71;72} (organisé par l'IEMSS - International Environmental Modeling and Software Society) qui précède une série d'articles soumis dans des journaux ciblés.

2.1.2 RECORD - VLE

Le projet RECORD (« RENovation et COORDination de la modélisation de cultures pour la gestion des agroécosystèmes) est un projet à l'initiative des départements EA (Environnement et Agronomie) et MIA de l'INRA. Il a pour objet le développement et la mise en place au sein de notre institut de recherche d'une plate-forme informatique de modélisation et simulation pour l'aide à la conception et l'évaluation de systèmes de culture innovants. Cette plate-forme vise à devenir un outil modulaire, partagé par les agronomes modélisateurs de l'INRA travaillant sur les systèmes de culture, qui facilitera le développement, le partage et la réutilisabilité des modèles développés en agronomie. Il s'agit d'un projet important mené sur plusieurs années qui implique aujourd'hui une équipe projet de six chercheurs / ingénieurs, et qui vise à terme la restructuration de la démarche de travail en modélisation d'un collectif pouvant être évalué à plusieurs dizaines de scientifiques.

L'équipe projet RECORD de l'équipe MAD a en charge depuis 2006 les tâches de conception et de développement informatique de la plate-forme et les premières actions de formation et de transfert vers les équipes de modélisateurs agronomes. Cette équipe projet est épaulée dans sa tâche par un groupe de scientifiques agronomes de l'INRA. Après une première phase en 2005 d'étude d'opportunité et d'expression initiale des besoins, nous avons finalisé sur 2006-2007 l'expression des besoins en modélisation des systèmes de culture, et rédigé les cahiers des charges fonctionnel et non-fonctionnel. Après avoir étudié et comparé différents formalismes de modélisation et d'architecture pour la plate-forme, et en particulier le cadre DIESE développé au sein de l'équipe, nous avons abouti au choix de la solution informatique VLE⁴⁰ de l'Université du Littoral Côte d'Opale (ULCO).

Le projet VLE (*Virtual Laboratory Environment*) est une plate-forme de modélisation et de simulation de systèmes complexes^{27;62}. Cet environnement s'appuie sur la théorie de la modélisation et de la simulation introduite par B. P. Zeigler en 1976. Elle repose sur le formalisme de modélisation et simulation DEVS (*Discrete Event System Specification*), qui est l'objet de nombreux développements théoriques ces

dernières années dans des domaines variés de l’informatique, de l’intelligence artificielle à la parallélisation et la distribution des calculs. Les principales caractéristiques de ce formalisme à événements discrets sont : être un formalisme dit abstrait car il permet de traduire ou de représenter les autres formalismes de systèmes dynamiques qu’ils soient continus ou discrets ; de proposer un cadre formel pour le couplage de modèles avec une représentation hiérarchique, modulaire et dynamique des modèles ; d’intégrer un cadre opérationnel sous la forme d’algorithmes disponibles et publiés (les simulateurs abstraits). De par sa définition d’unification des simulateurs, DEVS permet d’encapsuler les systèmes dynamiques mais également des simulateurs existants et autorise le couplage de modèles de systèmes dynamiques reposant sur des formalismes hétérogènes (équations différentielles, équations aux différences, automates à états, événements discrets, automates cellulaires, etc.).

VLE est aujourd’hui co-développé à l’INRA au sein de l’équipe MAD, au Cirad et à l’ULCO. Au cœur de la plate-forme RECORD, VLE est aujourd’hui utilisé dans un grand nombre de projets naissants en agronomie. Au-delà de RECORD, de nombreux projets de modélisation et simulation l’ont également retenu comme, par exemple, les projets du PNEC (Programme national d’environnement côtier) et le projet Chaloupe (financé par l’ANR) sur les systèmes multiagents multi-agents^{48;65}, ou la gestion de production⁷³. Au sein de l’équipe MAD, VLE est un environnement informatique privilégié pour expérimenter des développements théoriques sur DEVS et en théorie de la modélisation et la simulation avec, par exemple, les travaux sur l’intégration, dans DEVS, de nouveaux formalismes de modélisation et de leurs couplages, ou les travaux sur modélisation décisionnelle avec les thèses d’E. Rachelson⁷⁰ et de M. Akplogan⁷⁷, ou en filtrage et simulations récursives^{42;64}.

Depuis 2008, le projet RECORD poursuit principalement des développements méthodologiques et informatiques autour de VLE, des tests de fonctionnalités de la plate-forme sur quelques projets pilotes, et met en place des premières sessions de formation⁴¹. Le déploiement de la plate-forme vers l’ensemble des unités de recherche travaillant sur les systèmes de culture est prévu pour mi-2010.

2.1.3 Méthodes d’exploration de systèmes complexes

L’équipe a une longue expérience de recherche sur la problématique de l’utilisation de modèles de simulation pour l’analyse et la conception de systèmes agro-environnementaux (e.g.^{10;41}). Sur la période 2004-2008, dans la lignée de nos travaux sur l’apprentissage par renforcement, nous avons principalement orienté notre recherche méthodologique sur le thème de l’optimisation par simulation dans l’incertain (*simulation-based optimization*). Nous avons ainsi développé une famille d’algorithmes P2 pour l’optimisation globale d’une fonction objectif aléatoire dont la distribution n’est accessible que par échantillonnage. Ces algorithmes fonctionnent selon un principe de partitionnement hiérarchique de l’espace des paramètres à optimiser, ce partitionnement adaptatif servant de guide à l’échantillonnage par simulation construit pour rechercher les optima. L’algorithme P2 considère le critère de l’espérance de la fonction objectif, cadre classique de l’optimisation par simulation¹¹. Les algorithmes P2q¹² et P2m¹³⁷, développés au cours de la thèse d’O. Crespo soutenue en 2008, apportent des solutions originales à la question de l’optimisation de critères non classiques (optimisation d’un quantile, optimisation multi-critère dans l’incertain).

Ces résultats méthodologiques ont été exploités dans le cadre d’un projet d’optimisation de stratégies de conduite de systèmes de culture sur la base du modèle MODERATO de simulation de la conduite du maïs irrigué^{91;110}. Ils sont également au cœur de notre participation à l’UMT EAU pilotée par l’UMR AGIR du centre INRA Toulouse, au sein du groupe Optimisation. Au-delà de ces applications particulières, nous avons publié un état de l’art des méthodes d’optimisation par simulation¹¹¹ au sein d’un ouvrage collectif de synthèse sur les méthodologies d’étude des modèles de culture édité chez Elsevier. Enfin, les algorithmes de la famille P2 sont actuellement en cours de réimplémentation au sein de la plate-forme VLE dans le cadre du projet RECORD.

En parallèle à ces développements méthodologiques sur l’optimisation par simulation, l’équipe s’est impliquée de manière importante sur la problématique de l’exploration de modèles complexes sur la base de modèles de simulation. Nous avons ainsi participé au suivi de la thèse de R. Trépos, co-financée par MIA et menée au sein de l’équipe DREAM de l’IRISA à Rennes sur le thème de l’apprentissage symbolique à partir de données issues de simulation, avec application à la gestion d’un bassin versant pour une meilleure qualité de l’eau^{47;89}. Initié dans le cadre du projet SACADEAU, ce travail se poursuit

au sein des projets APPEAU et ACCASYA, financés par l'ADD/ANR.

Enfin, l'équipe a été impliquée depuis 2006, avec deux autres unités du département MIA, dans la création et l'animation du réseau MEXICO. Ce réseau a pour objectif de fédérer les travaux autour de l'exploration et l'analyse de systèmes complexes et d'ouvrir les biologistes modélisateurs au traitement statistique de leurs simulations et à une exploration raisonnée de leurs modèles. Le réseau regroupe des chercheurs/ingénieurs de différents organismes (INRA-MIA-EA, Cemagref, Ifremer, Cirad, Université du Littoral, ...). Deux approches de l'exploration de modèles ont été abordées. La première, informatique, a consisté en l'élaboration d'une architecture informatique permettant l'enchaînement des flux de données lors de l'exploration d'un modèle (constitution de schémas XML). La seconde, plus statistique, a consisté en la maîtrise des différentes méthodes d'analyse de sensibilité de modèles. Cette démarche a été l'occasion pour l'équipe de participer au développement d'une interface informatique entre la plate-forme VLE et l'environnement statistique R, et à la conception et l'organisation d'une école chercheurs qui s'est tenue sur 4 jours et qui a regroupé 59 stagiaires issus de différents organismes. Le réseau MEXICO est par ailleurs reconnu comme sous-réseau thématique du RNSC (Réseau National des Systèmes Complexes).

2.1.4 Décision dans l'incertain et PDM

Sur le plan méthodologique, l'étude de la décision séquentielle dans l'incertain est un des thèmes de recherche importants de l'équipe MAD. Au sein de ce thème, notre recherche s'est focalisée plus particulièrement sur les sous-thèmes méthodologiques suivants : (1) les approches qualitatives de la décision dans l'incertain, (2) la planification d'actions d'acquisition de connaissances, (3) la prise en compte explicite du temps dans les problèmes décisionnels de Markov, et (4) les problèmes de décision séquentielle spatialisée.

Durant la période 2004-2008, nos travaux en décision qualitative possibiliste, domaine scientifique dans lequel nous avons une certaine expérience, se sont poursuivis. Nous avons plus particulièrement exploré les deux thèmes originaux suivants : (i) liens entre la théorie qualitative possibiliste de la décision dans l'incertain et la théorie de l'utilité espérée, en collaboration avec H. Fargier, D. Dubois et H. Prade (IRIT-CNRS-UPS) et (ii) langages de représentation concis et algorithmes pour la décision qualitative possibiliste, avec en particulier des développements sur les diagrammes d'influence possibilistes avec L. Garcia (Université d'Angers). Ces travaux ont donné lieu à deux publications dans la revue phare de la communauté de l'intelligence artificielle (AIJ^{13;16}).

La planification épistémique est une réponse au problème de décision consistant à acquérir séquentiellement de l'information sur un état de la nature, et non à contrôler directement l'état d'un processus. Ce problème recèle de nombreuses potentialités d'application, dans les centres d'intérêt classiques de l'intelligence artificielle (élicitation de préférences, diagnostic, jeux...) mais aussi dans celui de la gestion agri-environnementale (diagnostic d'invasion de pathogènes...). Nos travaux dans ce domaine ont été menés dans le cadre de la thèse de N. Ravoanjanahary, co-encadrée avec J. Lang (IRIT-CNRS-UPS) et ont conduit à une formalisation du problème général de planification d'actions d'acquisition de connaissances dans un cadre proche de celui des processus décisionnels de Markov⁷⁵. Ces travaux se poursuivent actuellement, à la fois sur le plan théorique et sur le plan des applications, par leur extension à des problèmes d'échantillonnage spatial séquentiel optimal.

Toujours autour de la thématique des PDM, nous avons proposé un cadre original pour la modélisation et la résolution de problèmes de décision dans l'incertain spatialisée : le cadre des Processus Décisionnels de Markov sur Graphe (PDMG). Ce modèle est basé sur le cadre des Processus Décisionnels de Markov factorisés et l'adapte à des problèmes de décision spatialisée en lui adjoignant une structure de graphe d'interaction entre les variables d'état et de décision du problème. Nous avons ensuite proposé différentes méthodes de résolution approchée de problèmes de décision spatialisée exprimés dans ce cadre : une approche basée sur la programmation linéaire approchée⁵¹, une approche basée sur une approximation en champ moyen⁵⁹ et une approche basée sur l'apprentissage par renforcement multi-agents⁵⁰. Ces méthodes fournissent des outils complémentaires, la première étant plus efficace en temps et la seconde en qualité de l'approximation de la valeur de la solution du problème d'optimisation, la dernière permettant de prendre en considération des modèles de simulation pour lesquels le modèle GMDP associé n'a pas à être complètement explicite. Ce cadre et ces outils algorithmiques ont démontré leur intérêt en écologie, épidémiologie et foresterie (voir dernier résultat marquant de ce bilan)

Nous avons enfin traité la question de la formalisation et de la résolution de problèmes décisionnels de Markov temporels dans le cadre de la thèse d'E. Rachelson (2009). Il s'agissait ici d'étendre la problématique classique des PDM à la prise en compte explicite du temps dans la dynamique des systèmes à piloter, pour aborder la classe des systèmes à événements discrets, au coeur des approches de modélisation de l'équipe MAD. Un premier résultat de ce travail de thèse, co-encadré avec P. Fabiani (ONERA/DCSD), a consisté à étendre le cadre classique des Temporal MDP (TMDP) à des classes d'approximation polynomiales plus larges, et à en améliorer la résolution algorithmique. Un second axe de la thèse a conduit au développement d'un algorithme d'apprentissage par renforcement de type itération de la politique (ATPI) pour le pilotage de processus temporels de grande dimension représentés dans le formalisme des *General Semi Markov Decision Processes* (GSMDP). Ces résultats ont été présentés dans de nombreuses conférences internationales, et le prix du meilleur papier nous a été décerné par l'AFIA lors de la conférence nationale d'apprentissage automatique CAP'08⁸⁷. Le lien a été fait avec la plate-forme de simulation VLE en développant une extension DEVS pour le formalisme GSMDP⁷⁰, et l'algorithme ATPI est en cours d'implémentation dans VLE.

Plus globalement, notre expertise du domaine nous a conduits à participer de manière importante (co-rédaction de cinq chapitres) à la rédaction d'un ouvrage collectif de synthèse sur les PDM récemment édité chez Lavoisier-Hermès. Une version en langue anglaise est en cours de réalisation.

2.1.5 Analyse et décision dans les processus sur graphe

Les aspects spatialisés en décision se retrouvent dans de nombreux problèmes de gestion des agro-écosystèmes, par exemple pour le contrôle d'une épidémie ou d'une espèce invasive ou pour la gestion forestière. Les entités de décision ou action étant souvent discrètes, la théorie des graphes fournit un cadre pertinent de représentation des interactions. Une partie de nos activités a ainsi commencé à s'orienter vers un apport méthodologique sur des questions à la fois d'analyse et de décision dans les processus sur graphe. En effet, mieux comprendre les facteurs qui influencent la propagation spatiale d'une épidémie permettra d'améliorer les stratégies de contrôle.

Pour ce qui est de l'analyse, un des facteurs qui reçoit de plus en plus d'intérêt de la part des épidémiologistes est la structure du réseau d'interaction de la population soumise à l'épidémie. Ce réseau est naturellement modélisé par un graphe et un modèle simple et classique de propagation est le processus de contact sur graphe. Nous avons proposé de nouvelles méthodes d'approximation de ce modèle, basées sur les méthodes variationnelles, pour l'étude de sa dynamique et son équilibre^{23;94}. Les approximations obtenues améliorent les approches précédentes en ce qu'elles permettent de prendre en compte certains éléments de la structure du réseau d'interaction et donc d'étudier leur influence sur la capacité de propagation de l'épidémie. Nous avons ainsi mis en évidence le rôle de la fréquence de certains motifs géométriques dans le graphe. Ce travail a été développé en partie dans le cadre d'une collaboration avec l'unité BioGeco de l'INRA Bordeaux et l'IIASA (Vienne, Autriche). Il a également donné lieu à la rédaction de deux chapitres de livre traitant d'épidémiologie^{121;122}.

Les développements méthodologiques au niveau de la décision spatialisée se traduisent d'une part par le cadre des PDMG et les outils algorithmiques associés décrits précédemment, d'autre part par l'initiation de travaux autour de la question de l'échantillonnage spatial optimal. Nous avons proposé une première approche exacte basée sur les champs de Markov, ainsi qu'une méthode heuristique simple, avec une première application au cas du diagnostic de présence/absence d'une espèce invasive de fourmis. Les résultats prometteurs¹⁰¹ nous encouragent à continuer à explorer cette question et une thèse débute sur le sujet à l'automne 2009 (voir section Projet).

2.1.6 Approche PDM en écologie, épidémiologie, foresterie

Faisant suite à de précédentes applications des processus décisionnels de Markov à l'Agronomie¹¹¹, nous avons eu durant la période 2006-2009 une démarche de promotion active des PDM (au sens large, incluant le cas partiellement observable ou les méthodes d'apprentissage par renforcement) auprès de la communauté des « biologistes modélisateurs ». Cette démarche a été assurée via des enseignements, réguliers dans le Master 2 « Biostatistiques et Modélisation » (Université Paul Sabatier, école doctorale SEVAB) ou exceptionnels dans le cadre d'une école chercheur de deux semaines pour étudiants scandi-

naves (en foresterie/écologie) organisée à l'université d'Umeå en 2008. Dans cette dernière école chercheur nous avons assuré la moitié de l'enseignement sur la méthodologie PDM. Nous avons également organisé un mini-workshop international (Ecology morning session at UBIA) en 2008 sur Toulouse.

Cette promotion a également été assurée dans le cadre de collaborations (principalement internationales) avec des partenaires écologues ou forestiers auprès desquels nous avons diffusé ces méthodes. Ces partenariats principaux sont les suivants. Nous collaborons de longue date avec deux laboratoires australiens : Le « spatial ecology lab » (Université de Brisbane) avec lequel nous avons appliqué le cadre des PDM complètement ou partiellement observables dans des problèmes liés à l'écologie (protection d'un groupe d'espèces menacées, détectabilité d'espèces menacées « cryptiques »^{9;81}) et le laboratoire de biologie de l'université Monash avec lequel nous avons utilisé des méthodes d'apprentissage par renforcement (afin de résoudre un problème de gestion de réserve²⁸) ou de décision en environnement partiellement observable (à travers notre participation à un projet ARC sur l'éradication d'espèces invasives). Ces collaborations se sont également matérialisées par des visites de chercheurs étrangers dans notre laboratoire (Eve Mac Donald-Madden, Univ. of Queensland, Tara Martin, CSIRO, Daniel Spring, Monash Univ.) ainsi que par les séjours postdoctoraux puis le départ de I. Chadès à l'Université de Brisbane. Nous collaborons également de longue date avec le laboratoire de gestion forestière SLU de l'Université d'Umeå en appliquant des méthodes de type PDMG ou d'apprentissage à des problèmes de gestion de risques tempête^{15;50} ou d'impact du changement climatique sur la gestion forestière¹³⁰. La thèse de N. Forsell a ainsi été réalisée en co-tutelle entre les deux laboratoires, autour de la gestion forestière permettant de minimiser le risque de dommages liés aux tempêtes. Enfin, nous avons également appliqué les PDMG à un problème de gestion de résistances variétales à la maladie du phoma chez le colza, dans le cadre du projet ADD CEDRE (2006-2009) en collaboration avec l'unité Agir de l'INRA Toulouse⁶¹.

Cette promotion de nos approches auprès de la communauté d'utilisateurs potentiels est un élément important de notre stratégie de recherche. Pour le moment cet effort de promotion a été principalement dirigé vers des partenaires étrangers, jusqu'ici plus « réceptifs » aux méthodes d'optimisation de la gestion des agro-écosystèmes. Ces premières expériences concluantes nous permettent d'attirer aujourd'hui l'attention de nouveaux partenaires potentiels, français et étrangers. Nous avons en particulier été invités à exposer nos travaux durant la première conférence « Computational Sustainability », visant à fédérer une communauté mondiale de chercheurs en math/info et biologistes travaillant autour du « développement durable » au sens large, auprès desquels nous allons continuer cet effort de promotion.

2.2 SaAB

L'équipe SaAB (*Statistics and Algorithms for Biology*) se compose de six chercheurs et de cinq ingénieurs. Son domaine d'activité principal est la bioinformatique, un domaine de recherche à l'intersection des statistiques, de l'informatique et de la biologie. Sa mission essentielle consiste à développer et à mettre en œuvre des méthodes mathématiques et informatiques pour modéliser et résoudre des questions issues de la biologie, grande productrice de données expérimentales (issues par exemple des séquençage, génotypage, expression, phénotypage, etc.) aux caractéristiques propres (grande dimension, réponse stochastique du phénomène biologique aussi bien que bruit associé au dispositif de mesure, données manquantes, interactions entre les composants du système, etc.). En sus de traditionnelles collaborations disciplinaires avec nos collègues informaticiens et statisticiens, nos travaux s'appuient généralement sur des collaborations directes avec des équipes de recherches en biologie moléculaire et en génétique du monde végétal, animal ou bactérien. Ils peuvent mener à la production de nouvelles méthodes (valorisées dans des publications méthodologiques et des logiciels dédiés) ou à la production de nouvelles connaissances biologiques, valorisées dans des revues de biologie. Un intérêt tout particulier est attaché à la mise à disposition d'outils logiciels génériques, mis en œuvre sur des données fournies par nos partenaires et immédiatement maîtrisables par d'autres biologistes, via la fourniture de documentations ou de formations adaptées.

Pour mener à bien ses travaux, l'équipe met en œuvre et développe des méthodes permettant de modéliser des phénomènes biologiques d'intérêt afin d'en extraire de la connaissance biologique et en particulier permettre l'identification d'éléments fonctionnels dans les génomes. L'équipe mobilise ainsi des compétences en statistiques (modélisation, statistique inférentielle, modèles de mélanges, statistique des processus...) et en informatique (optimisation combinatoire, modèles graphiques, réseaux de contraintes pondérés) qui lui permettent de s'attaquer à une large variété de problèmes contribuant à la compréhension du vivant en s'appuyant sur l'analyse de données issues de marqueurs moléculaires, de phénotypes ou de séquences nucléotidiques. Si les statistiques se prêtent naturellement à la modélisation de données bruitées, l'optimisation combinatoire permet souvent de prendre en compte de telles données via la minimisation d'un critère de parcimonie. Parce qu'ils définissent un critère probabiliste plus fin et permettent un rapprochement avec le point de vue statistique et l'équipe MAD, les modèles graphiques stochastiques font l'objet d'un intérêt accru.

En dehors des principales thématiques détaillées dans la suite, l'équipe a produit des résultats sur l'analyse des remaniements chromosomiques entre génomes (calcul de distance d'inversion^{159;160}, par S. Bérard qui a quitté l'équipe fin 2006) et dans l'utilisation des champs markoviens pour l'analyse de données d'expression (clustering par champ markovien, par M. Vignes recruté fin 2008 et arrivé dans l'équipe en juillet 2009 après un séjour post-doctoral au sein de l'équipe BioSS en Écosse). Elle maintient aussi une activité sur l'analyse de données de puces (génomodiagnostic).

Au delà des publications, notre production scientifique nous permet de bénéficier d'une très bonne visibilité. Notre activité méthodologique sur les réseaux de contraintes nous permet d'entretenir des collaborations avec des équipes scientifiques de premier plan au niveau international, d'intervenir de façon récurrente dans des comités de programme, d'édition de revues, d'évaluation et de sociétés savantes au niveau national et international. La visibilité de notre activité finalisée est attestée par l'adoption des méthodes et outils logiciels de l'équipe aussi bien dans le monde de la recherche publique et privée, que dans des projets d'envergure nationale ou internationale. Notre bonne connaissance du tissu de la recherche en bioinformatique nous conduit enfin fréquemment à participer à des jurys ou des comités stratégiques ou d'évaluation au niveau national.

2.2.1 Localisation de QTL

Notre spécialité scientifique dans la détection de QTL par analyse de liaison (c'est-à-dire par le suivi de la transmission allélique dans des pedigrees connus) est l'analyse de plan de croisement comportant plusieurs descendance apparentées. La modélisation de ce type de dispositif a été faite dans le cadre du modèle linéaire en se basant sur nos travaux précédents. Ceux-ci prouvaient l'équivalence asymptotique du critère des moindres carrés avec le critère de maximum de vraisemblance dans ce cas de vraisemblance non régulière. Le support de valorisation de nos travaux scientifiques sur ce sujet est le logiciel MCQTL¹⁷⁷.

Le démarrage de ce projet de développement logiciel a été rendu possible en grande partie grâce aux

sollicitations de nos partenaires que sont l'UMR de Génétique Végétale du Moulon et la société EURALIS en direction du GIS Génoplante. Depuis 2004, la société SYNGENTA, qui a acheté la première version du logiciel 20 k€, a rejoint ce groupe. Le logiciel évolue régulièrement par ajout de fonctionnalités (types de descendances : F3 jusqu'à F7, outbred ; complexification des modèles : dominance, épistasie QTLxQTL, etc.) en partie grâce aux financements que nous obtenons auprès de nos partenaires privés ou du GIS Génoplante. MCQTL a été choisi pour illustrer la contribution de Génoplante à la bioinformatique lors de l'anniversaire de ses 10 ans. Depuis l'ouverture du site de diffusion, en janvier 2006 (<http://carlit.toulouse.inra.fr/MCQTL>), environ 130 utilisateurs d'instituts académiques (50 français dont 40 de l'INRA, 80 étrangers) ont demandé une licence. L'originalité du produit, qui n'a qu'un seul concurrent au plan international, est très attractive pour le secteur privé car il permet d'analyser les plans de croisement mis en place dans les programmes d'amélioration. Il est ainsi utilisé en routine par les sociétés Biogemma, Euralis et Syngenta. La société Nestlé est venue nous rencontrer au printemps 2009 et a programmé l'achat du produit pour le premier trimestre 2010. L'évolution actuelle du logiciel, qui a pour objectif d'utiliser le déséquilibre de liaison et des informations sur le pedigree des parents du dispositif pour améliorer la cartographie des QTL, est soutenu par un projet ANR Génoplante.

Les recherches sur la cartographie fine de QTL par déséquilibre de liaison ont été amorcées dans l'équipe SaAB au début des années 2000. Elles sont inspirées par les francs succès obtenus par les généticiens humains pour cartographier de façon extrêmement précise des gènes de maladie (gènes 0-1). En 2003, nous avons constitué un groupe composé de statisticiens, généticiens des populations, des plantes et des animaux et d'épidémiologistes provenant d'organismes divers (INRA, CNRS, INSERM, Université). Ce groupe s'est réuni mensuellement jusqu'en 2008 pour partager les méthodes développées autour du déséquilibre de liaison. Un chapitre d'ouvrage ainsi que l'organisation d'un séminaire national sur les questions relatives au déséquilibre de liaison sont venus concrétiser ces collaborations. C'est au cours de la thèse de doctorat de Simon Boitard (2003-2006) que les avancées méthodologiques ont été les plus importantes. Nous avons tout d'abord amené des résultats théoriques dans le domaine de la « génétique des populations ». Nous nous sommes en effet intéressés à la distribution de probabilité des fréquences d'haplotypes pour un modèle de diffusion de Wright-Fisher à deux loci bialléliques avec recombinaison. L'expression de cette densité de probabilité est encore à l'heure actuelle un problème ouvert et on n'en connaît que certains moments calculés sous des hypothèses très particulières. Nous avons proposé une méthode numérique approchée, basée sur les différences finies, pour résoudre les équations de Kolmogorov associées à ce processus de diffusion. Nous avons ensuite étudié la précision de la solution obtenue, démontré certaines propriétés théoriques de la distribution et déterminé la complexité algorithmique de la méthode¹⁵⁷. Cette méthode a ensuite été appliquée au problème de la cartographie fine de QTL par déséquilibre de liaison. Dans une deuxième partie, nous avons généralisé une méthode de cartographie de QTL par déséquilibre de liaison déjà mise au point par l'équipe. Utilisant simultanément l'information génétique fournie par les deux marqueurs moléculaires entourant la position du QTL putatif (*interval mapping*), cette méthode s'est révélée au moins aussi précise que d'autres méthodes multi-marqueurs publiées dans la littérature, et surtout, avec des temps de calcul beaucoup plus courts¹⁵⁶. Ce travail de thèse s'est poursuivi dans un projet ANR-GenAnimal (2006-2008) « Combiner l'analyse de liaison et le déséquilibre de liaison pour cartographier finement les QTL dans les pedigrees animaux ». La méthode développée dans ce cadre a fait l'objet du dépôt d'un package R.

2.2.2 Cartographie comparative et assemblage

La cartographie génétique et d'hybrides irradiés a pour objectif une meilleure compréhension de l'organisation des génomes. S'appuyant sur des données de nature stochastique (les recombinaisons de la méiose ou les cassures par irradiation), le problème se formule idéalement comme un problème d'optimisation mixte (continu et discret) de la vraisemblance, comportant donc une facette combinatoire. Il est traité dans le logiciel CarthaGène via l'utilisation d'un algorithme EM combiné avec des méthodes d'optimisation combinatoire dédiées au traitement du problème du voyageur de commerce. Face à l'évolution des technologies de typage, la principale nouvelle difficulté réside dans la prise en compte d'un nombre croissant de marqueurs, de l'ordre de 50 par chromosome (en 2000) à plus d'un millier aujourd'hui. L'ordonnement d'un tel nombre de marqueurs nécessite d'exploiter de nouvelles informations. Pour tirer parti du nombre croissant de génomes séquencés et du phénomène de synténie, nous avons développé, en collaboration avec Thomas Faraut (LGC INRA Toulouse), une méthode de cartographie comparée

exploitant un ordre de référence des marqueurs issu d'un génome voisin déjà séquencé¹⁶⁸. Cette méthode peut également exploiter un ordre de référence issu d'un assemblage en cours d'un génome cible et offrir ainsi une aide à la construction de l'assemblage²⁴⁵. En collaboration avec une équipe canadienne (S. Moore, Univ. Alberta) disposant de données de typage haute densité (60K Illumina), nous l'avons ainsi appliquée au génome bovin^{181;185}, en cours d'assemblage. Ces méthodes sont toutes intégrées au logiciel CarthaGène, utilisé par de nombreux laboratoires publics et privés au niveau international (plus de 1000 téléchargements et plus de 80 citations depuis 2005).

Sur la période d'évaluation, le code de CarthaGène, développé sur la forge logicielle de l'unité, a également profité de la participation de développeurs extérieurs et en particulier de celle de Clare Nelson (Univ. Kansas) qui a ajouté la possibilité de traiter des pedigrees de type « backcross avancé » et a produit un tutoriel pour CarthaGène. Nous assurons également des formations CarthaGène, au niveau école toulousaine (ENSAT) ou entreprise privées (Syngenta et ASEDIS/IFR 40).

2.2.3 Réseaux de contraintes pondérés et applications

L'équipe poursuit une activité de développement d'algorithmes d'optimisation combinatoire dans le domaine des réseaux de contraintes valués et pondérés. Ce formalisme est une extension des réseaux de contraintes qui permet l'optimisation de critères de natures variées, couvrant ainsi différents modèles graphiques discrets (déterministes ou non). Ils permettent de modéliser habituellement des problèmes d'aide à la décision ou issus de la recherche opérationnelle. En bioinformatique, l'extraction d'information de nature discrète de données éventuellement bruitées se prête naturellement à ce type de modélisation, un critère de parcimonie ou de nature probabiliste permettant de capturer l'incertitude dans les données.

Nous cherchons à maintenir une activité de développement de méthodes originales, souvent motivées par nos applications en bioinformatique mais en cherchant un caractère générique. Durant la période couverte, nous nous sommes principalement intéressés à deux grandes approches : l'utilisation de techniques de filtrage par cohérence d'arc pour fournir des minorants forts et incrémentaux et l'exploitation de la structure du problème (importante en génétique) pour améliorer l'efficacité de résolution via des approches de type séparation-évaluation complètes ou partielles.

En ce qui concerne le premier point, nous continuons à faire fructifier un de nos résultats (publié en 2000) étendant la notion de filtrage, classique dans les réseaux de contraintes, aux réseaux de contraintes valués. En collaboration avec nos collègues barcelonais (Univ. Politecnica de Catalunya) et toulousains (Inst. de Recherche en Informatique, UPS), nous avons ainsi défini quatre nouvelles techniques de filtrage^{190;208;209;211}, ayant de bonnes propriétés pratiques et/ou théoriques (et permettant par exemple de traiter les problèmes sous-modulaires non booléens, non binaires – mais d'arité bornée – en temps polynomial ou encore de traiter des problèmes avec de grands domaines).

Les travaux autour de l'exploitation de la structure des problèmes ont été menés dans le cadre d'un contrat ANR blanc incluant un laboratoire marseillais (LSIS) et un laboratoire orléanais (LIFO) spécialisés dans la décomposition de graphes et de réseaux de contraintes sur la base d'un premier travail²¹⁰ réalisé en collaboration avec un collègue toulousain (ONERA-CERT). Dans le cadre du projet ANR, nous avons développé des méthodes de résolution exploitant simultanément la structure des problèmes et certaines méthodes de filtrage. Le post-doctorant étranger recruté pour 20 mois dans le cadre du projet ANR a permis de faire fortement évoluer notre plate-forme logicielle, de traiter des problèmes issus de la bioinformatique et de clore plusieurs problèmes ouverts classiques dans la communauté^{208;217}. Le projet a été invité à présenter ses résultats au colloque ANR à Paris en 2009 (9 sur 36 projets blancs en informatique).

Ces méthodes, implémentées dans notre plate-forme `toulbar2`, ont été utilisées pour modéliser et résoudre des problèmes en bioinformatique : localisation de gènes d'ARN (thèse de M. Zytnecki, élève normalien), identification de tagSNP²¹⁷, diagnostic de pedigrees animaux¹⁸⁷ menant au logiciel Mendel-Soft. Ce logiciel est intégré via une interface Web dans la plate-forme de génétique quantitative hébergée au Centre de Traitement de l'Information Génétique (CTIG) du département de Génétique Animale à l'INRA de Jouy. Il a permis au CTIG de corriger de manière optimale un pedigree ovin complexe de plus de 120000 individus. `toulbar2` a également participé à différentes compétitions MaxSAT'2006 (2nd sur 6) et MaxCSP'07-08²¹⁸ (1er sur 8 et sur 7 solveurs respectivement).

L'intérêt important accordé aux critères probabilistes dans l'équipe et dans nos problèmes nous a

amenés à nous approcher des modèles graphiques stochastiques discrets (champs markoviens, réseaux bayésiens) qui se modélisent directement sous forme de réseau de fonction de coût et à considérer une éventuelle combinaison des modèles déterministes des réseaux de contraintes avec ces modèles stochastiques. Ce travail a été mené via le co-encadrement de la thèse de C. Pralet (ingénieur Sup'Aéro, avec G. Verfaillie, LAAS-CNRS). Il a permis de généraliser le formalisme des réseaux de contraintes pondérés pour définir le cadre des réseaux PFU (Plausibilités, Faisabilités, Utilités) permettant de capturer de nombreux problèmes de décision séquentiels à horizon fini dans l'incertain^{184;216}. Les techniques de décomposition explorées dans le projet ANR blanc ont également été mobilisées dans ce cadre, fournissant des algorithmes génériques avec de bonnes propriétés théoriques^{214;215}. La thèse de C. Pralet a eu le prix national de thèse ASTI 2007 (<http://www.agence-nationale-recherche.fr/actualite/13?lnglnfold=157>). La plate-forme `toulbar2` a également participé à une évaluation sur des problèmes d'optimisation dans les modèles graphiques stochastiques discrets organisée dans le cadre de la conférence internationale UAI'2008 et a obtenu d'excellents résultats (1er sur 9 solveurs), dépassant en particulier l'efficacité de méthodes de type « *stochastic local search* », qui n'offrent aucune garantie d'optimalité.

La thèse d'A. Favier (Master à Marseille) continue dans cette direction en s'attaquant à un problème issu de la génétique (reconstruction d'haplotypes dans des pedigrees complexes) qui soulève des problèmes de dénombrement²²⁵, liés à des problèmes de calcul de vraisemblance.

Afin d'augmenter notre visibilité au niveau national et international, nous avons eu une forte activité de diffusion de nos résultats : cours de Master 2 (univ. Paul Sabatier), école d'été internationale²⁸¹, tutoriels et conférences invités^{195;196;199;204;273-275}, chapitres invités dans des ouvrages de synthèse nationaux²⁶⁸ et internationaux²⁶⁷. Nous participons activement à l'animation de la communauté internationale : organisation de plusieurs workshops (International WS on Soft constraints and Preferences, de 2005 à 2008), participation aux comités de programme de conférences nationales (JFPC, AFIA), internationales (CP, AAI, IJCAI) et aux comités d'édition des revues internationales « Constraints » et « JAIR » (Journal of Artificial Intelligence Research) ainsi qu'au comité exécutif de l'Association for Constraint Programming.

2.2.4 Localisation de gènes d'ARN

L'importance quantitative et fonctionnelle des ARN non codants (ARNnc) a été très longtemps sous estimée. Leur grande diversité, leur présence dans l'ensemble des organismes vivants, leur implication dans la régulation d'une majorité de processus cellulaires ont récemment suscité un intérêt majeur de la part des biologistes. Leur localisation et leur annotation dans les séquences génomiques constitue l'un des challenges actuels en bioinformatique. La question de la recherche de ces gènes dans le cadre de la découverte de nouvelles familles d'ARN mais aussi de manière plus classique, dans le cadre de la détection et de l'annotation de gènes appartenant à une famille d'ARN connue est traitée dans l'équipe à partir de plusieurs approches.

Les développements réalisés dans l'équipe sur cette thématique ont été en partie soutenus par un projet GENOPLANTE, deux projets ANR et un projet européen. Ils ont donné lieu à plusieurs présentations et publications proposant des résultats originaux en bioinformatique^{183;188;191;192;229} et en biologie^{155;173;242}, à 3 thèses (dont deux soutenues), à plusieurs logiciels (ApolloRNA^{252;256}, RNAsim, Darn!, RNAspace²⁴⁹, LeARN¹⁸³), à la découverte de nouvelles familles d'ARN dans 3 génomes procaryotes et à la découverte de nouveaux membres de la famille des petits ARNs nucléolaires dans des génomes d'archae et de plantes^{151;164;186}. Depuis 2005, nous participons aussi de manière active à l'animation de la communauté nationale en bioinformatique de l'ARN par l'organisation d'ateliers (respectivement 3 jours en 2005, 2008, 2009). Ces ateliers ont bénéficié du soutien du projet ARENA (ACI ImpBio) et du réseau national de bioinformatique ReNaBi.

Trois approches ont été plus spécifiquement investies. Une première réflexion a permis de mettre en place dès 2005 les bases d'une approche comparative multi-génomiques en s'appuyant sur une représentation des similarités entre génomes par un graphe. Cette approche, appliquée à plusieurs génomes de staphylocoques, a fait l'objet d'une analyse approfondie des régions conservées dans le génome de *S. aureus* N315, et a permis de proposer plusieurs dizaines de candidats parmi lesquels les biologistes en ont validé et caractérisé 10 dans les conditions testées¹⁷³. La thèse d'Ana Kozomara²⁸⁷ a permis de poursuivre ces travaux en supprimant notamment la redondance par fusion des régions chevauchantes.

Ce travail a permis de mettre en évidence des structures de type « graphes petits mondes » récurrentes qui devraient faire l'objet dans le futur de travaux méthodologiques. Le travail réalisé au cours de cette thèse a aussi permis de proposer de nouveaux ARNnc dans le génome de la bactérie *R. solanacaerum*. Dans le cadre de cette même thèse, nous avons utilisé la théorie basée sur le modèle linéaire généralisé pour valider l'intérêt d'une deuxième approche basée sur l'exploitation du biais de composition pour segmenter le génome. Ce travail reste à valoriser. Les modèles de Markov cachés ont été sélectionnés comme première approche de segmentation d'une séquence génomique. Ils ont aussi montré leur intérêt dans le cadre de l'analyse du génome de *S. aureus* N315. Enfin, La thèse de Matthias Zytnecki constitue un résultat original obtenu sur la question de la détection de gènes appartenant à une famille d'ARN connue. Le travail réalisé a permis de croiser problématique ARN et réseaux de contraintes pondérés. Cette thèse s'est concentrée sur la formalisation et le développement d'algorithmes. Elle a permis d'une part d'aboutir à un outil logiciel de bioinformatique (DARN!) qui, tout en étant aussi efficace que les outils existants, permet en sus de prendre en compte des caractéristiques jusqu'ici ignorées : interactions ARN-ARN, superposition de propriétés sur un même élément fonctionnel¹⁹¹. Cet outil est accessible via un site web dédié (<http://carlit.toulouse.inra.fr/Darn/>) et a récemment été intégré à la plate-forme de détection et d'annotation d'ARN non codant, RNAspace, développée en collaboration avec les plate-formes bioinformatique INRA Bioinfo Génotoul et SIGENAE mais aussi avec l'Université d'Orsay (IGM) et l'INRIA de Lille (LIFL). Motivée par les caractéristiques spécifiques du problème modélisé, cette thèse a aussi amené des résultats méthodologiques originaux^{190;219} pour le filtrage de réseaux de contraintes pondérés ayant de très grands domaines. Ces développements, d'abord implémentés dans DARN!, ont été ensuite intégrés dans la plate-forme `toulbar2` et ont permis d'améliorer l'efficacité de résolution sur différents benchmarks (dont certains sans lien avec la bioinformatique).

2.2.5 Prédiction de gènes de protéines dans les génomes eucaryotes

Une étape fondamentale pour le décodage des séquences génomiques consiste à identifier les régions codant pour des protéines. Ce problème a été très largement étudié dans la littérature et l'équipe a développé depuis plusieurs années une méthode de prédiction s'appuyant sur un modèle graphique inspiré des réseaux de contraintes pondérés, permettant l'intégration d'une grande variété de données et incarnée dans le logiciel EuGène, développé en collaboration avec le laboratoire du LIPM (Lab. Interactions Plante-Microorganismes, UMR INRA CNRS, Toulouse) et le VIB (Flanders Institute for Biotechnology, Gand, Belgique).

Durant la période d'évaluation, EuGène a été utilisé pour construire l'annotation officielle de différents génomes eucaryotes : celui de la légumineuse modèle *Medicago truncatula*¹⁶², celui de la tomate et celui du nématode *Meloidogyne incognita*¹⁵⁰. Ce travail important n'aurait pu être réalisé sans la participation forte de nos collègues bioinformaticiens du LIPM (INRA-CNRS) et sans le support matériel de la plate-forme bioinformatique. Effectué dans le cadre de consortiums internationaux de séquençage et d'annotation, via des projets européens (PI GLIP et EU-SOL), il permet à la fois de faire mûrir l'outil EuGène et de lui donner une visibilité internationale²⁰² tant à l'intérieur du consortium que via les publications (dans des revues à fort facteur d'impact). Les prédictions originales d'EuGène sur le génome d'*Arabidopsis thaliana* ont été validées expérimentalement¹⁵² et intégrées dans la base internationale TAIR (the Arabidopsis Information Resource).

EuGène a ainsi, entre autres, été rendu capable de prédire des variants d'épissage¹⁷² sur la base de données expérimentales de transcription (EST par exemple). L'arrivée des nouvelles technologies de séquençage devrait rapidement amener une avalanche de séquences, de qualité variable. Afin de pouvoir traiter de telles séquences, le petit frère procaryote d'EuGène (FrameD) a été utilisé pour construire un « pipeline » automatique d'annotation de contigs de transcrits (EST)¹⁷⁴. Un travail similaire est en cours sur EuGène lui-même. La thèse en cours d'E. Audemart a pour but d'enrichir les types de données dont EuGène se nourrit en exploitant en particulier les duplications intra-génomiques.

Plusieurs formations à la mise en œuvre d'EuGène ont été réalisées sur la période, attirant à chaque fois entre 6 et 10 participants (dont des sémienciers tels que Syngenta) et menant, par exemple, à l'annotation du génome de *Botrytis cinerea* à l'unité de recherche Génomique-Info de l'INRA¹⁷¹.

2.3 Plate-forme bioinformatique GENOTOUL

L'équipe « Plate-forme Bioinformatique » (<http://bioinfo.genotoul.fr>), créée en 2000, constitue, depuis mai 2007, la troisième équipe de l'unité. Cette équipe est constituée par 5 agents permanents (3 IE, 1 IR et 1 DR) dont deux y exercent leur activité sur une partie seulement de leur temps (IR et DR, 30%). La plate-forme GéoToul Bioinfo est l'une des 5 plates-formes fondatrices du projet Génopole Toulouse Midi-Pyrénées initié en 1999. Cette plate-forme, labellisée plate-forme nationale RIO en 2001, 2003 et 2006 vient d'obtenir le label IBiSA en juillet 2009. A l'occasion du changement du responsable scientifique, en 2003, les missions de la plate-forme ont été redéfinies dans le cadre de la Génopole Toulouse Midi-Pyrénées et se déclinent en plusieurs points visant à offrir à la communauté locale une infrastructure matérielle et logicielle performante et adaptée au traitement à grande échelle, à en former les utilisateurs, à fournir un appui aux programmes scientifiques prioritaires de biologie et de bioinformatique, et à développer des programmes en concertation avec la communauté locale et les autres plates-formes de bioinformatique. Les agents recrutés dans cette équipe se sont appropriés ces missions et les déclinent de manière à satisfaire au mieux l'ensemble des utilisateurs. La plate-forme exerce ces missions dans un contexte scientifique local riche et actif dans le domaine de l'analyse des séquences génomiques et de la cartographie génétique (environ 40 chercheurs et ingénieurs permanents en bioinformatique sur le centre d'Auzeville). Au-delà de ce contexte scientifique local qui positionne la thématique scientifique, les missions s'inscrivent dans le contexte plus large de la Génopole Toulouse Midi-Pyrénées et sont tournées vers l'ensemble des laboratoires et entreprises régionales présentant des besoins dans l'analyse à grande échelle des séquences génomiques. Ce contexte est aussi très diversifié dans la nature des applications et des organismes considérés : plantes, animaux (y compris l'humain), micro-organismes.

Le renforcement progressif de l'équipe et le soutien financier du réseau national des génopoles ont permis de mettre en place, dès 2005, une première infrastructure tournée vers le traitement à grande échelle des séquences génomiques, des actions en collaboration avec les autres plates-formes de bioinformatique françaises, mais aussi une activité de développement en soutien bioinformatique aux programmes de biologie régionaux (1 CDD/an en moyenne). Cette période a aussi marqué le développement du projet BioMaJ, qui sera détaillé par la suite. Les arrivées en 2007 de C. Noirot (IE, Bioinformatique) et C. Klopp (IR, nouveau responsable opérationnel de la plate-forme) ont marqué très rapidement un virage dans la stratégie de la plate-forme et ont été l'occasion de définir une stratégie de développement en bioinformatique. A cette période, les thématiques de recherche des biologistes étaient renouvelées par au moins deux faits majeurs avec 1) l'arrivée des nouvelles techniques de séquençage à haut débit qui permettent aujourd'hui d'explorer de façon différente et à moindre coût plusieurs grands domaines d'applications d'intérêt comme, par exemple, le séquençage de génomes complets (avec des capacités atteignant, pour une personne, le séquençage d'un génome de quelques mégabases avec une exactitude proche de 100%, en à peine quelques jours et en s'affranchissant totalement des étapes de clonage bactérien) mais aussi, 2) la découverte récente d'une multitude d'ARN non codant (miRNA, piRNA, petits ARN régulateurs bactériens, CRISPR,...) dont l'importance majeure dans la majorité des processus cellulaires au sein de l'ensemble des organismes ne cesse d'être démontrée et questionne sur la prédiction de ces gènes, leur fonction et leur analyse. Depuis 2007, la politique de la plate-forme vise à développer son activité autour de ces deux faits majeurs et des enjeux de traitement à grande échelle associés, avec l'ambition de permettre aux programmes scientifiques de rester compétitifs. Il nous a semblé aussi particulièrement important à cette période de mettre en place des outils d'accompagnement pour les biologistes. La mise en place d'un catalogue de formations nous a paru constituer l'un des premiers outils à produire.

Les plates-formes nationales ont vocation à être ouvertes à l'ensemble des acteurs de la recherche académique et industrielle, pour des activités de prestations et/ou de partenariat. Elles se doivent donc d'offrir à leurs utilisateurs la garantie de la qualité du travail réalisé. Conformément à la charte RIO et à la politique du réseau de Génopoles, les plates-formes régionales, avec l'aide du RNG et des établissements de recherche (CNRS, INRA, INSERM), ont souhaité s'engager dans une démarche qualité coordonnée pour mettre à profit l'expérience acquise par celles d'entre elles qui sont le plus avancées. Nous déclinerons plus bas l'état de notre investissement dans cette démarche.

2.3.1 SIGENAE

L'équipe SIGENAE (<http://www.sigena.org>) constitue la plate-forme bioinformatique des départements impliqués dans l'étude des génomes d'animaux à l'INRA (PHASE, GA, SA). Les personnels de cette plate-forme (4 IE, 1 IR) sont répartis sur 3 localisations (Toulouse, Rennes, Tours) et 4 unités d'accueil. La responsabilité de cette équipe a été confiée en 2001 à un IR, recruté dans l'Unité BIA et aussi responsable de la plate-forme bioinformatique GenoToul depuis 2007. En 2001, les biologistes de l'INRA travaillant sur l'analyse des génomes des animaux d'élevage se sont organisés au sein du programme AGENAE pour avoir accès aux outils de la génomique et de la transcriptomique à haut débit. Le comité d'orientation scientifique du programme a décidé de créer une équipe de bioinformatique de proximité, répartie sur le territoire, apte à traiter les données issues du séquençage et des puces à ADN. Le début de la période 2005-2009 a vu la stabilisation de l'équipe avec la création de trois postes d'ingénieur d'étude en plus des deux postes déjà présents : un poste de chef de projet créé en 2002 dans l'unité UBIA et un poste d'ingénieur d'étude ouvert en 2004. Ce début de période a concorde avec la migration des applications et des bases de données développées et maintenues par la plate-forme SIGENAE de Jouy-en-Josas à Toulouse. Cette migration a eu lieu suite à une réponse conjointe à un appel à financement dans le cadre de l'appel d'offre INRA sur les équipements lourds. SIGENAE a été partenaire de différents projets européens entre 2005 et 2009. Tout d'abord le réseau d'excellence EADGENE (<http://www.eadgene.info>) qui étudie les interactions hôte/pathogène dans les espèces d'intérêt agronomique et aquacole. Christophe Klopp est responsable de l'animation du WP 1.3 (integrating bioinformatic facilities) ce qui a permis à l'équipe de bénéficier de 48 mois de CDD d'ingénieur en développement financés par le réseau et de collaborer sur l'annotation des oligo-nucléotides déposés sur les puces à ADN. Les autres projets européens de la période ont été : Aquafirst (<http://aquafirst.vitamib.com/>), Aquafunc et Aquagenome (<http://genomics.aquaculture-europe.org>). Au plan national la plate-forme a participé à plusieurs projets ANR : Delisus, SheepSNPQtl, Meetac. L'ensemble des collaborations et des prestations a permis à l'équipe et à ses membres d'être auteur(s), remercié(s) ou cité(s) dans une cinquantaine d'articles parus dans des revues à comité de lecture. Avec l'arrivée des nouvelles technologies de séquençage les compétences de l'équipe sont et seront fortement mobilisées sur les différents projets en cours ou en définition. L'hébergement intégré sur la plate-forme bioinformatique, de moyens de calcul, de banques de séquences et des compétences associées, est un atout important pour pouvoir mettre rapidement à disposition des biologistes des résultats pertinents.

2.3.2 Evolution de l'infrastructure pour le traitement à grande échelle

Les dix années d'existence de la plate-forme peuvent être segmentées en trois périodes. La première, couvrant les années 2000 à 2003, correspond à la mise en place. Une seule machine servait alors à tous les besoins de stockage, de traitement et de mise à disposition des données. Cette architecture s'est rapidement révélée inadaptée aux missions de traitement à grande échelle. Au début de la seconde période, s'étendant de 2003 à 2005, l'architecture précédente a été remplacée par un ensemble de machines dédiées au calcul (cluster de 20 processeurs), aux base de données (baie de disques de 3 To) et à la mise à disposition de résultats et des applications (serveur d'application). En 2005, l'architecture précédente a été étendue, portant l'espace de stockage à 10To et le cluster de calcul à 88 processeurs. Le serveur d'application a été doublé et un serveur de test a été ajouté. L'évolution récente de l'infrastructure, réalisée avec le soutien financier (500K€) du CPER (2M€ pour la période 2007-2013) permet aujourd'hui de mettre à disposition de la communauté un cluster de calcul de plus de 400 coeurs et un espace de stockage totalisant 50 To. Ces évolutions ont permis d'une part de suivre le rythme de croissance de la production des données des biologistes et bioinformaticiens utilisant la plate-forme et d'autre part de maintenir à jour en local les banques de données biologiques internationales qui sont des outils indispensables à l'analyse à grande échelle des données générées dans le cadre des programmes de biologie. La tenue à jour de ces banques nécessite une capacité de stockage importante car outre le volume des données brutes, il est indispensable pour leur exploitation de disposer de versions d'index adaptées aux types de recherches effectuées (ex : recherche par similarité par Blast). Le renouvellement des serveurs a aussi été l'occasion d'isoler certains services au moyen de machines virtuelles, permettant une plus grande flexibilité dans la mise en oeuvre et offrant une sécurité accrue aux utilisateurs.

2.3.3 BioMaj

Les banques de données biologiques constituent l'un des outils indispensables à l'analyse des données produites dans le cadre des programmes scientifiques de biologie. Elles se caractérisent par une localisation distribuée, un format et un contenu hétérogènes et en constante évolution, des volumes très différents allant de quelques Méga à plusieurs Téra mais aussi par une fréquence de mise à jour variable. Dans le contexte actuel de production et d'analyses à grande échelle, il est devenu indispensable de disposer sur site de ces données. Il peut aussi s'avérer nécessaire de les formater en vue de leur utilisation par un outil spécifique (ex : Blast). Le suivi des mises à jour et des résultats de formatages produisent des fichiers de « traces » qu'il faut analyser afin de détecter d'éventuels problèmes liés à ces traitements. L'ensemble des tâches à réaliser rend ainsi la gestion de banques lourde et fastidieuse. C'est dans ce contexte qu'a été initié fin 2004 le projet BioMaj (Biologie, Mise à Jour) avec l'objectif d'automatiser au mieux les *workflows* liés à la gestion sur site des banques de données. Suite à une évaluation de plusieurs solutions, le logiciel « Citrina », développé par la communauté GMOD (Generic Model Organism Project), a été sélectionné pour être étendu avec l'objectif de permettre le suivi de l'état des banques biologiques pour l'ensemble des tâches considérées, la robustesse des traitements et l'ergonomie. Ce projet a fait l'objet d'une collaboration en 2005 avec Josh Goodman (auteur de Citrina, Indiana University Department of Biology), collaboration qui s'est étendue par la suite à deux plates-formes de bioinformatique du Réseau National des plate-formes bioinformatiques, la plate-forme GenOuest (INRIA Rennes) et la plate-forme MIGALE du centre INRA de Jouy. Le logiciel BioMaJ est aujourd'hui déployé sur plusieurs plates-formes (dont la plate-forme bioinformatique GenoToul), en France et à l'étranger. Il a fait l'objet d'une publication en 2008¹⁷⁰.

2.3.4 Deux axes stratégiques de développement

La constitution au plan régional du GIS GenoToul, la création au plan national du GIS IBISA et la nécessité de mettre en avant des spécificités par plate-forme ont accéléré la réflexion menée au sein de la direction de la plate-forme sur les axes stratégiques à développer. La direction de la plate-forme a pris en considération simultanément les compétences présentes et les besoins identifiés en 2007 dans la communauté des biologistes. La plateforme est accueillie dans l'UBIA qui développe depuis un peu plus de 10 ans une activité de recherche méthodologique en informatique autour de la prédiction et de l'analyse des ARN non codant. La responsable scientifique de la plate-forme bioinformatique a été impliquée dans plusieurs programmes scientifiques (2 projets ANR, 3 co-encadrements de thèse sur la période) autour de ces questions. D'un autre côté, le responsable technique de la plateforme possède une longue expérience dans l'ingénierie des données génomiques, notamment sur les traitements d'assemblage et d'annotation de séquences génomiques. Cette expérience, acquise dans le cadre du programme national SIGENAE de séquençage et d'annotation des séquences d'animaux d'élevage, dépasse le cadre de l'espèce considérée. Les thématiques de recherche sur lesquelles la plateforme se positionne sont donc en parfaite adéquation avec les orientations scientifiques de l'équipe de recherche et des autres acteurs locaux et régionaux.

Notre activité sur la prédiction et l'annotation des ARN non codant vise la mise en oeuvre et le développement d'outils. Dans le cadre de la mise en oeuvre d'outils, l'expertise acquise ces dernières années autour de la recherche de gènes appartenant à la famille des snoRNAs nous a permis d'être sollicités sur deux études portant respectivement sur les acteurs des modifications de bases chez les archaebactéries et les plantes. Dans la première étude, notre rôle a consisté à rechercher les snoRNA responsables de modifications de type 2'O méthylation et pseudouridylation identifiées chez l'archaee *Haloferax volcanii*, en collaboration avec plusieurs équipes au plan national et international. Ce travail a fait l'objet d'une publication dans BMC Genomics¹⁷⁵. Dans la deuxième étude, nous avons extrait des bases de données l'ensemble des snoRNA à boîtes C/D connus chez le riz, nous les avons repositionnés sur le génome, nous avons recherché de nouveaux snoRNA à boîtes C/D chez cette plante, puis analysé l'ensemble des régions en amont de tous ces snoRNAs. Ce travail a permis de mettre en évidence un motif conservé dans plus de 50% des régions analysées confirmant l'hypothèse d'une transcription réalisée par Pol III. Ce travail a fait l'objet d'une collaboration entre le LIPM (INRA, Toulouse) et le CIRAD (Montpellier). Un article sur ce travail est actuellement en cours d'évaluation. Le développement d'outils a porté sur la réalisation d'une plateforme d'annotation des ARN non codants, en collaboration avec plusieurs équipes locales (Plate-forme SIGENAE, équipe SaAB) et nationales (IGM/CNRS Orsay, LIFL/CNRS-INRIA,

Lille). Ce projet a été financé par le Réseau National des Génomies pour deux ans. La première version de cette plate-forme d'annotation est accessible à l'URL <http://www.rnaspace.org> et fait l'objet de la rédaction d'un article. Cet axe bénéficiera d'un soutien d'IBISA pour la période 2010-2011.

Le séquençage haut débit s'affirme depuis 2005 comme une technologie incontournable dans le cadre de nombreuses applications en biologie. En 2007, nous avons décidé d'investir le traitement des données issues du séquençage haut débit en tant qu'axe stratégique de la plate-forme. Une première expérience sur l'analyse de données produites à partir du séquenceur 454 de Roche a été acquise entre 2007 et 2009 dans le cadre d'une prestation de service (V. Neugnot-Roux, société ADISSEO, Toulouse) autour du traitement de séquences d'un champignon en vue de leur assemblage et de leur annotation pour la recherche de protéines d'intérêt industriel. Cette première expérience a permis de définir et d'initier des traitements de base, de sélectionner des outils appropriés et de les mettre en oeuvre dans un contexte d'annotation. La mise à disposition des données traitées dans un environnement d'annotation, une formation de nos clients à l'utilisation de cet environnement et un accompagnement de 12 mois sur l'utilisation de cet environnement ont conclu cette première étude. Une deuxième étude a été réalisée en 2008-2009 dans le cadre d'une collaboration avec G. Véronèse (Equipe Ingénierie Enzymatique Moléculaire, INSA Toulouse). Cette étude a porté sur la recherche de protéines d'intérêt pour la dégradation de fibres. Le travail réalisé a abouti à l'assemblage et à l'annotation de séquences de 40 fosmidites produites au Génomètre à partir d'un séquenceur 454. L'arrivée planifiée du 454 sur la plate-forme Génomique fin 2008 a fondé la mise en place d'un partenariat étroit entre les deux plates-formes. Outre le fait que la plate-forme bioinformatique, pour un coût de 500 €, s'engage sur la mise en oeuvre de traitements de base (nettoyage et assemblage) et la mise à disposition des séquences produites par la plate-forme Génomique, le partenariat engagé se concrétise aussi par des réponses communes à des appels d'offre (CPER 2006-2013), des stratégies de développements informatiques et de mise en oeuvre concertées et partagées ayant, par exemple, conduit à la réalisation de l'environnement NG6 (<http://vm-bioinfo.toulouse.inra.fr/ng6/>), le rassemblement au sein d'un même CATI des personnels informatiques des deux plates-formes, et la mise à disposition par la plate-forme génomique d'un ingénieur d'étude en CDD (24 mois) accompagnant la mise en place des premiers outils de traitement et leur mise en oeuvre sur les données produites. Depuis février 2009, la plate-forme bioinformatique a réalisé dans le cadre de ce partenariat les traitements de base pour 9 projets.

L'activité de la plate-forme bioinformatique sur cet axe s'annonce en forte croissance dans le cadre du partenariat établi avec la plate-forme Génomique mais aussi, de manière plus large, dans le cadre de collaborations régionales et nationales avec les biologistes. Cette activité mobilise actuellement 3 ingénieurs permanents de la plate-forme (environ 1,5 ETP). L'arrivée prochaine de l'IE en CDD devrait encore renforcer cet axe.

2.3.5 Un catalogue de formations qui s'étoffe

Dans le cadre de son ouverture à l'ensemble de la communauté scientifique, la plate-forme bioinformatique a mis en place, en 2007, un catalogue de formations (voir au sein de la section 3.6.5) offrant la possibilité à la communauté scientifique de s'approprier l'utilisation des outils disponibles autour de l'infrastructure proposée. L'objectif visé est d'accompagner voire d'augmenter, pour les utilisateurs, leurs capacités d'analyse des données par la formation. 103 personnes/jour ont ainsi été formées à l'utilisation de l'infrastructure disponible. Enfin, de manière ponctuelle ou récurrente, la plateforme peut prendre en charge, en collaboration avec d'autres équipes locales ou nationales, l'organisation de formations liées à l'utilisation de son infrastructure, à son domaine de compétences et/ou à des besoins exprimés par la communauté en bioinformatique. Depuis 2005, 4 formations ont répondu à ce besoin, portant respectivement sur l'analyse des données du transcriptome (126 personnes/Jour), l'analyse de séquences (36 personnes/jour), le système de gestion documentaire Typo3 (8 personnes/jour) et sur l'ordonnanceur SGE (14 personnes/jour). Les perspectives sur cette activité restent liées aux objectifs affichés. En 2009, une formation plus spécifique sur l'environnement d'annotation Apollo a été mise en place, avec le soutien de l'équipe SaAB, permettant à 10 personnes/jour de se former à l'utilisation de ce logiciel.

2.3.6 Démarche qualité

Dans le respect de la charte des plates-formes de recherche en sciences du vivant, la plateforme bioinformatique s'est engagée dans une démarche qualité en participant, comme 4 autres plates-formes régionales, au projet ANR RIOQUALITOUL, coordonné par C. Chevalet, directeur du GIS GenoToul. Dans le cadre de ce projet, la plate-forme bénéficie de l'appui d'un CDD (0,5 ETP) niveau IE, qui accompagne la démarche depuis fin 2008. Les premiers objectifs sont la mise en place d'un système de management de la qualité et l'obtention de la certification ISO-9001 :2000 dans le courant du premier trimestre 2010. Pour atteindre ces objectifs, l'ensemble de l'équipe a suivi en 2008 une journée de formation visant la sensibilisation à la norme ISO-9001 :2000. Deux membres de l'équipe ont aussi suivi en 2008 une formation plus approfondie de Responsable Management et Qualité (RMQ). A ce jour la plate-forme s'est dotée d'un logiciel, eGroupWare, permettant d'enregistrer les activités du personnel de la plate-forme et a mis en place les éléments fondateurs de la démarche : écriture de la politique qualité et du manuel qualité, définition des processus, description des procédures et modes opératoires. Un premier audit interne (avril 2009) a permis d'évaluer le système de gestion documentaire sans remettre en cause les échéances planifiées.

2.3.7 Articulation avec l'équipe SaAB

Il peut apparaître des activités ci-dessus que peu de liens aient été tissés avec l'équipe de recherche SaAB. Cela s'explique par la nature des premiers objectifs visés par la plate-forme, tournés prioritairement vers la mise en place d'une infrastructure de traitement à grande échelle mais aussi par le peu de moyens humains accompagnant cette mise en place. Sur cette période, l'équipe SaAB avait aussi peu de besoins en termes de calcul à grande échelle et les moyens de l'unité permettaient d'accompagner les développements méthodologiques. Les évolutions récentes et la masse de données à considérer et à intégrer dans les développements méthodologiques commencent néanmoins à modifier la nature des besoins de l'équipe SaAB en terme d'infrastructure de calcul. Nous considérons aujourd'hui que l'infrastructure de calcul et de stockage répond aux besoins de la communauté régionale et nous envisageons de manière sereine les prochaines évolutions, grâce au soutien du CPER. Sur la période considérée, la plate-forme bioinformatique a connu plusieurs mutations. Rattachée à l'origine à l'équipe SaAB, elle est tout d'abord devenue équipe de l'unité au regard de l'augmentation de ses effectifs et de la nature de ses activités tournées vers le service. En 2007, les arrivées de l'IR co-responsable de la plate-forme et d'un IE ont conduit à définir deux axes stratégiques pour la plate-forme, visant l'accompagnement des biologistes, et à établir un partenariat avec la plate-forme Génomique GenoToul. Les projets en cours, les demandes croissantes autour du traitement des séquences issues du haut débit ainsi que le soutien d'IBiSA sur ces deux axes montrent que les choix réalisés répondent aux besoins de la communauté. Nous pensons aussi qu'à terme, les compétences opérationnelles pointues acquises sur le traitement des données issues du haut débit pourraient bénéficier à l'équipe SaAB.

Chapitre 3

Publications et autres productions de 2005 à 2009

3.1 Connaissances fondamentales : Publications scientifiques

3.1.1 Équipe MAD

3.1.1.1 Articles dans des revues internationales ou nationales avec comité de lecture répertoriées dans les bases de données internationales ou équivalentes

- [1] J. E. BERGEZ, N. COLBACH, O. CRESPO, F. GARCIA, M. H. JEUFFROY, E. JUSTES, C. LOYCE, N. MUNIER-JOLAIN et W. SADOK : Designing crop management systems by simulation. *European Journal of Agronomy*, 31, 2009.
- [2] P. BESSE, H. CARDOT, R. FAIVRE et M. GOULARD : Statistical modelling of functional data. *Applied Stochastic Models in Business and Industry, special issue on Statistical Learning*, 21:165–173, 2005.
- [3] H. CARDOT : Conditional functional principal components analysis. *Scandinavian Journal of Statistics*, 34:317–335, 2007. URL <http://www.blackwell-synergy.com/toc/sjos/0/0>.
- [4] H. CARDOT, C. CRAMBES, A. KNEIP et P. SARDA : Smoothing spline estimators in functional linear regression with errors-in-variables. *Computational Statistics and Data Analysis*, 51:4832–4848, 2007.
- [5] H. CARDOT, C. CRAMBES et P. SARDA : Quantile regression when the covariates are functions. *Nonparametric Statistics*, 17:841–856, 2005.
- [6] H. CARDOT, P. MAISONGRANDE et R. FAIVRE : Varying-time random effects models for longitudinal data : unmixing and temporal interpolation of remote-sensing data. *Journal of Applied Statistics*, 35:827–846, 2008.
- [7] H. CARDOT et P. SARDA : Estimation in generalized linear models for functional data via penalized likelihood. *J. of Multivariate Analysis*, 92:24–41, 2005.
- [8] B. CARGNELUTTI, A. COULON, A. J. HEWISON, M. GOULARD, J. M. ANGIBAULT et N. MORELLET : Testing gps performance for wildlife monitoring using mobile collars with known reference points. *Journal of Wildlife Management*, pages 1380–1387, 2006.
- [9] I. CHADÈS, E. McDONALD-MADDEN, M. A. MCCARTHY, B. WINTLE, M. LINKIE et H. P. POSSINGHAM : When to stop managing or surveying cryptic threatened species. *Proc Natl Acad Sci U S A*, 105(37):13936–40, Sep 2008. URL <http://view.ncbi.nlm.nih.gov/pubmed/18779594>. Featured in The Australian, ABC Science and TREE by MacKenzie, D. I. 2009. Getting the biggest bang for our conservation buck. *Trends in Ecology & Evolution* 24 :175-177.

- [10] M. H. CHATELIN, C. AUBRY et F. GARCIA : A novel method for the sustainable management of wheat crops : exploration by simulation. *Agronomy for Sustainable Development*, 27:337–345, 2007.
- [11] O. CRESPO, J. E. BERGEZ et F. GARCIA : P2 hierarchical decomposition procedure : application to irrigation strategies design. *Operational Research : An International Journal*, 2009.
- [12] O. CRESPO, J. E. BERGEZ et F. GARCIA : P2q hierarchical decomposition algorithm for quantile optimization : application to irrigation strategies design. *Annals of Operations Research*, 2009.
- [13] H. FARGIER et R. SABBADIN : Qualitative decision under uncertainty : back to expected utility. *Artificial Intelligence*, 165:245–280, 2005.
- [14] F. FORBES, N. PEYRARD, C. FRALEY, D. GEORGIAN-SMITH, D. GOLDHABER et A. RAFTERY : Model-based region-of-interest selection in dynamic breast mri. *Journal of Computer Assisted Tomography*, 30(4):576–687, 2006.
- [15] N. FORSELL, P. WIKSTRÖM, F. GARCIA, R. SABBADIN, K. BLENNOW et L. ERIKSSON : Management of the risk of wind damage in forestry : a graph-based markov decision process approach. *Annals of Operations Research*, 2009.
- [16] L. GARCIA et R. SABBADIN : Complexity results and algorithms for possibilistic influence diagrams. *Artificial Intelligence*, 172(8-9):1018–1044, 2008.
- [17] C. GASCUEL-ODOUX, P. AUROUSSEAU, M. O. CORDIER, P. DURAND, F. GARCIA, J. SALMON-MONVIOLA, F. TORTRAT et R. TREPOS : A decision-oriented model to evaluate the effect of land use and agricultural management on herbicide contamination in stream water. *Environmental Modelling & Software*, 2009.
- [18] G. MARTIN, P. CRUZ, J. P. THEAU, C. JOUANY, P. FLEURY, S. GRANGER, R. FAIVRE, G. BALENT, S. LAVOREL et M. DURU : A multi-site study to classify semi-natural grassland types. *Agriculture, Ecosystems and Environment*, 129:508–515, 2009.
- [19] G. MARTIN, L. HOSSARD, J. P. THEAU, O. THÉRON, E. JOSIEN, P. CRUZ, J. P. RELIER, R. MARTIN-CLOUAIRE et M. DURU : Characterizing potential flexibility in grassland use. application to the french aubrac area. *Agronomy for Sustainable Development*, 29(2):381–389, 2009.
- [20] T. G. MARTIN, I. CHADÈS, P. ARCESE, P. P. MARRA, H. P. POSSINGHAM et D. R. NORRIS : Optimal conservation of migratory species. *PLoS One*, 2(1):e751, 2007. URL <http://view.ncbi.nlm.nih.gov/pubmed/17710150>.
- [21] L. MATON, D. LEENHARDT, M. GOULARD et J. E. BERGEZ : Linking irrigation strategy and farm systems : an attempt in south western france. *Agricultural Systems*, 86:293–311, 2005.
- [22] N. PEYRARD et P. BOUTHEMY : Motion-based selection of relevant video segments for video summarization. *Multimedia Tools and Applications Journal, Special Issue : Video Segmentation for Semantic Annotation and Transcoding*, 26:255–274, 2005.
- [23] N. PEYRARD, U. DIECKMANN et A. FRANC : Long-range correlations improve understanding the influence of network structure on per contact dynamics. *Theoretical Population Biology*, 73/3:383–394, 2008.
- [24] N. PEYRARD et A. FRANC : Cluster variation approximations for a contact process living on a graph. *Physica A*, 358:575–592, 2005.
- [25] N. PEYRARD, F. PELLEGRIN, J. CHADŒUF et D. NANDRIS : Statistical analysis of the spatio-temporal dynamics of rubber bark necrosis : no evidence of pathogen transmission. *Forest Pathology*, 36:360–371, 2006.
- [26] D. PÉPIN, P. C. RENAUD, Y. BOSCARDIN, M. GOULARD, C. MALLET, F. ANGLARD et P. BALLON : Relative impact of browsing by red deer on mixed coniferous and broad-leaved seedlings-an enclosure-based experiment. *Forest Ecology and Management*, 222:302–313, 2006.

- [27] G. QUESNEL, R. DUBOZ et E. RAMAT : The virtual laboratory environment – an operational framework for multi-modelling, simulation and analysis of complex dynamical systems. *Simulation Modelling Practice and Theory*, 17:641–653, April 2009.
- [28] C.-E. RABIER, R. SABBADIN et D. SPRING : Dynamic reserve site selection under contagion risk of deforestation. *Ecological Modelling*, 201:75–81, 2007.
- [29] D. SPRING, O. CACHO, R. M. NALLY et R. SABBADIN : Preemptive conservation versus fire-fighting : a decision-theoretic approach. *Biological Conservation*, 136:531–540, 2007.
- [30] M. TCHAMITCHIAN, R. MARTIN-CLOUAIRE, B. JEANNEQUIN, J. LAGIER et S. MERCIER : Ser-riste : a daily set point determination software for glasshouse tomato production. *Computers and Electronics in Agriculture*, 50:25–47, 2006.
- [31] G. THÉBAUD, N. PEYRARD, S. DALLOT, A. CALONNEC et G. LABONNE : Investigating disease spread between two assessment dates with permutation tests on a lattice. *Phytopathology*, 95:1453–1461, 2005.
- [32] S. VENNER, I. CHADÈS, M. J. BEL-VENNER, A. PASQUET, F. CHARPILLET et R. LEBORGNE : Dynamic optimization over infinite-time horizon : Web-building strategy in an orb-weaving spider as a case study. *Journal of Theoretical Biology*, 241(4):725–733, 21 2006.

3.1.1.2 Articles dans des revues avec comité de lecture non répertoriées dans des bases de données internationales

- [33] I. CHADÈS : Algorithmes de co-évolution simultanée pour la résolution approchée de pdm multi-agent. *Revue d'Intelligence Artificielle*, 20:345–382, 2006.
- [34] L. GARCIA et R. SABBADIN : Diagrammes d'influence possibilistes : complexité et algorithmes. *Revue d'Intelligence Artificielle*, 21(4):521–554, 2007.
- [35] R. MARTIN-CLOUAIRE et J. P. RELIER : Modelling and simulating work practices in agriculture. *International Journal of Metadata, Semantics and Ontologies*, 4(1-2):42–53, 2009. URL <http://carlit.toulouse.inra.fr/diese/docs/IJMSO09.pdf>.
- [36] N. PEYRARD, A. CALONNEC, F. BONNOT et J. CHADŒUF : Explorer un jeu de données sur grille par tests de permutation. *Revue de Statistique Appliquée*, 1, 2005.
- [37] L. PÉRET et F. GARCIA : Recherche en ligne pour les processus décisionnels de markov. *Revue d'intelligence artificielle*, 20(2-3):181–202, 2006.
- [38] G. THÉBAUD, N. PEYRARD, S. DALLOT, A. CALONNEC, G. LABONNE et J. CHADŒUF : Testing the spatial association of disease patterns between two dates in orchards. *Acta Horticulturae*, 2008. sous presse.

3.1.1.3 Articles dans des revues sans comité de lecture

3.1.1.4 Conférences données à l'invitation du comité d'organisation dans un congrès national ou international

3.1.1.5 Communications avec actes dans un congrès international

- [39] S. BACKÉUS, L. O. ERIKSSON et F. GARCIA : Impact of climate change uncertainty on optimal forest management policies at stand level. In *Proceedings of the International Congress on Modelling and Simulation, Advances and application for management and decision-making*, Melbourne, Australia, dec 2005.

- [40] J. E. BERGEZ, P. CHABRIER, F. GARCIA, C. GARY, M. H. JEUFFROY, R. MARTIN-CLOUAIRE, H. RAYNAL et D. WALLACH : Studying cropping system by simulation : the record platform project. *In Proceedings of Farming Systems Design : Methodologies for Integrated Analysis of Farm Production Systems*, Catane, Italie, 2007.
- [41] J. E. BERGEZ, P. CHABRIER, F. GARCIA, C. GARY, D. MAKOWSKI, G. QUESNEL, E. RAMAT, H. RAYNAL, N. ROUSSE et D. WALLACH : Record : a new software platform to model and simulate cropping systems. *In Proceedings of Farming Systems Design*, 2009.
- [42] B. BONTÉ, R. DUBOZ, G. QUESNEL et J. MÜLLER : Recursive simulation and experimental frame for multiscale simulation. *In SCSC'09 : Summer Computer Simulation Conference*, Istanbul, Turkey, July 2009.
- [43] P. CHABRIER, F. GARCIA, M. MARTIN-CLOUAIRE, G. QUESNEL et H. RAYNAL : Toward a simulation modeling platform for studying cropping systems management : the record project. *In D. OXLEY, Les ; Kulasiri, éditeur : Proc. of MODSIM07, International Congress on Modelling and Simulation*, pages 839–845, Christchurch, NZ, 2007. Modelling and Simulation Society of Australia and New Zealand. ISBN 978-0-9758400-4-7.
- [44] I. CHADÈS, T. G. MARTIN, J. M. CURTIS et C. BARRETO : Managing interacting species : a reinforcement learning decision theoretic approach. *In D. OXLEY, Les ; Kulasiri, éditeur : Proc. of MODSIM07, International Congress on Modelling and Simulation*, pages 2209–2215. Modelling and Simulation Society of Australia and New Zealand, 2007. ISBN 978-0-9758400-4-7.
- [45] X. CHARDON, C. RIGOLOU, C. BARATTE, A. LE GALL, S. ESPAGNOL, R. MARTIN-CLOUAIRE, J. P. RELIER, C. RAISON, J. C. POUPA et P. FAVERDIN : Melodie : A whole-farm model to study the dynamics of nutrients in integrated dairy and pig farms. *In D. OXLEY, Les ; Kulasiri, éditeur : Proc. of MODSIM07, International Congress on Modelling and Simulation*, pages 1638–1645. Modelling and Simulation Society of Australia and New Zealand, 2007. ISBN 978-0-9758400-4-7. URL http://www.mssanz.org.au/MODSIM07/papers/25_s25/MELODIE_s25_Chardon_.pdf.
- [46] N. CIALDELLA, J. P. RELIER, R. MARTIN-CLOUAIRE, M. H. JEUFFROY et J. M. MEYNARD : Silasol : A model-based assessment of pea (*pisum sativum* l.) cultivars accounting for crop management practices and farmers' resources. *In Proceedings of Farming Systems Design 2009*, Monterey, CA, August 2009.
- [47] M. O. CORDIER, F. GARCIA, C. GASCUEL-ODOUX, V. MASSON, J. SALMON-MONVIOLA, F. TORTRAT et R. TREPOS : A machine learning approach for evaluating the impact of land use and management practices on streamwater pollution by pesticides. *In Proceedings of the International Congress on Modelling and Simulation, Advances and application for management and decision-making*, Melbourne, Australia, dec 2005.
- [48] R. DUBOZ, D. VERSMISSE, G. QUESNEL, A. MUZZY et E. RAMAT : Specification of dynamic structure discrete event multiagent systems. *In Proceedings of Agent Directed Simulation, Spring Simulation Multiconference*, Huntsville, Alabama, USA, April 2006.
- [49] N. FORSELL, L. O. ERIKSSON, F. GARCIA, R. SABBADIN et P. WIKSTROM : A graph-based markov decision process approach for managing forests under windthrow risk. *In Proc. of MODSIM07, International Congress on Modelling and Simulation*, Christchurch, New-Zealand, 2007.
- [50] N. FORSELL, F. GARCIA et R. SABBADIN : Reinforcement learning for spatial processes. *In 18th Inter. Congress on Modelling and Simulation (MODSIM'09)*, Cairns, Australia, 2009.
- [51] N. FORSELL et R. SABBADIN : Approximate linear-programming algorithms for graph-based markov decision processes. *In Proc. of European Conference on Artificial Intelligence (ECAI'06)*, pages 590–594, Riva del Garda, Italy, Août 2006.
- [52] F. GARCIA, F. GUERRIN, R. MARTIN-CLOUAIRE et J. P. RELIER : The human side of agricultural production management - the missing focus in simulation approaches. *In R. M. ZERGER,*

- Andre; Argent, éditeur : *Proc. of MODSIM05, International Congress on Modelling and Simulation*, pages 203–209. Modelling and Simulation Society of Australia and New Zealand, 2005. ISBN 0-9758400-2-9. URL <http://www.mssanz.org.au/modsim05/papers/garcia.pdf>.
- [53] L. GARCIA et R. SABBADIN : Possibilistic influence diagrams. In *Proc. of European Conference on Artificial Intelligence (ECAI'06)*, pages 372–376, Riva del Garda, Italy, Août 2006.
- [54] G. MARTIN, M. DURU, R. MARTIN-CLOUAIRE, J. P. RELIER et J. P. THEAU : Taking advantage of grassland and animal diversity in managing livestock systems : a simulation study. In *Proceedings of Farming Systems Design 2009*, August 2009.
- [55] G. MARTIN, M. DURU, R. MARTIN-CLOUAIRE, J. P. RELIER, J. P. THEAU, O. THÉRON et L. HOSSARD : Towards a simulation-based study of grassland and animal diversity management in livestock farming systems. In J. C. J. R. A. G. G. SÀNCHEZ-MARRÈ, M; Béjar, éditeur : *Proceedings of the iEMSs Fourth Biennial Meeting : International Congress on Environmental Modelling and Software (iEMSs 2008). Barcelona, Catalonia, July 2008.*, pages 783–791. International Environmental Modelling and Software Society, 2008. ISBN 978-84-7653-074-0. URL <http://www.iemss.org/iemss2008/index.php?n=Main.S8>.
- [56] R. MARTIN-CLOUAIRE et J. P. RELIER : Representing and interpreting flexible production management plans. In A. F. C. G. N. BARROS, Fernando; Bruzzone, éditeur : *Proceedings of Conceptual Modeling and Simulation Conference (CMS 2005)*, pages 69–76, LSIS, Domaine Scientifique de Saint Jérôme, Marseille, 2005. ISBN 2-9520712-2-5.
- [57] E. McDONALD-MADDEN, I. CHADÈS, M. A. MCCARTHY, B. WINTLE et H. POSSINGHAM : Should i spread my risk or concentrate my efforts : Is triage of a subpopulation ever the best decision? In D. OXLEY, Les; Kulasiri, éditeur : *Proc. of MODSIM07, International Congress on Modelling and Simulation*, pages 2244–2249. Modelling and Simulation Society of Australia and New Zealand, 2007. ISBN 978-0-9758400-4-7. URL http://www.mssanz.org.au/MODSIM07/papers/41_s34/ShouldISpread_s34_McDonald_.pdf.
- [58] A. OLIOSO, V. RIVALLAND, R. FAIVRE, M. WEISS, J. DEMARTY, T. WASSENAAR, F. BARET, H. CARDOT, P. ROSSELLO, F. JACOB, C. B. HASAGER et Y. INOUE : Monitoring evapotranspiration over the alpillles test site by introducing remote sensing data at various spatial resolutions into a dynamic svat model. In M. A. M. J. D'URSO, G; Osann Jochum, éditeur : *Proc. of Conf. Earth observation for vegetation monitoring and water management*, volume 852, pages 234–241, Napoli, Italie, 2006. AIP Conference proceedings, American Institute of Physics. ISBN 0-7354-0346-5. URL <http://proceedings.aip.org/>.
- [59] N. PEYRARD et R. SABBADIN : Mean field approximation of the policy iteration algorithm for graph-based markov decision processes. In *Proc. of European Conference on Artificial Intelligence (ECAI'06)*, pages 595–599, Riva del Garda, Italy, Août 2006.
- [60] N. PEYRARD, R. SABBADIN et J. N. AUBERTOT : A graph-based markov decision process framework for optimising integrated management of diseases in agriculture. In *Proc. of MODSIM07, International Congress on Modelling and Simulation*, Christchurch, New-Zealand, 2007.
- [61] N. PEYRARD, R. SABBADIN et J.-N. AUBERTOT : A markov decision process for integrated pest management. In *International Congress on Modelling and Simulation*, 2007.
- [62] G. QUESNEL, R. DUBOZ, E. RAMAT et M. TRAORÉ : Vle : A multimodeling and simulation environment. In *Proceedings of the Summer Simulation Multiconference*, pages 367–374., San Diego, California, USA, July 2007.
- [63] G. QUESNEL, R. DUBOZ, D. VERSMISSE et E. RAMAT : Devs coupling of spatial and ordinary differential equations : Vle framework. In *Proceedings of the Open International Conference on Modeling and Simulation*, pages 281–294, Clermont-Ferrand, France, June 2005.
- [64] G. QUESNEL, H. RAYNAL, E. RAMAT et D. MAKOWSKI : Software environments for implementing the particle filter with dynamic models. In *Proceedings of Farming Systems Design*, 2009.

- [65] G. QUESNEL, D. VERSMISSE et E. RAMAT : Coupling of physical models and social models : multi-modeling and simulation with vle. *In Joint Conference on Multi-Agent Modeling for Environmental Management*, Bourg Saint Maurice, France, March 2005.
- [66] C. E. RABIER, R. SABBADIN et D. SPRING : Dynamic reserve site selection under contagion risk of deforestation. *In Proc. of MODSIM05, International Congress on Modelling and Simulation*, pages 2102–2108, Melbourne, Australia, Décembre 2005.
- [67] E. RACHELSON, P. FABIANI et F. GARCIA : Adapting an mdp planner to time-dependency : case study on a uav coordination. *In 4th Workshop ICAPS on Planning and Plan Execution for Real-World Systems : Principles and Practices for Planning in Execution*, Thessaloniki, Greece, September 19-23rd 2009.
- [68] E. RACHELSON, P. FABIANI et F. GARCIA : Timdppoly : an improved method for solving time-dependent mdps. *In 21st International Conference on Tools with Artificial Intelligence*, Newark, New Jersey, USA, November 2-5 2009.
- [69] E. RACHELSON, F. GARCIA et P. FABIANI : Extending the bellman equation for mdps to continuous actions and continuous time in the discounted case. *In 10th International Symposium on Artificial Intelligence and Mathematics*, Fort Lauderdale, Florida, USA, 2008.
- [70] E. RACHELSON, G. QUESNEL, F. GARCIA et P. FABIANI : A simulation-based approach for solving generalized semi-markov decision processes. *In European Conference on Artificial Intelligence*, Patras, Grèce, July 2008.
- [71] C. RIGOLOT, X. CHARDON, J. P. RELIER, R. MARTIN-CLOUAIRE, J. Y. DOURMAD, A. LE GALL, S. ESPAGNOL, C. BARATTE et P. FAVERDIN : A generic framework for the modelling of livestock production systems : Melodie. *In In Proceedings of Farming Systems Design*, Monterey, CA, August 2009.
- [72] A. RIPOCHE, J. P. RELIER, R. MARTIN-CLOUAIRE, A. BIARNÈS, N. PARÉ et C. GARY : Modeling dynamically the management of intercropped vineyards to control the grapevine water status. *In Proceedings of Farming Systems Design 2009*, August 2009.
- [73] O. ROUX, D. DUVIVIER, G. QUESNEL et E. RAMAT : Optimization of preventive maintenance through a combined maintenance-production simulation model. *In International Conference on Industrial Engineering and Systems Management (IESM 2009)*, Montreal, Canada, 13-15 May 2009.
- [74] R. SABBADIN, E. BERGONNIER et D. SPRING : A reinforcement-learning application to biodiversity conservation in costa-rican forest. *In Proc. of MODSIM07, International Congress on Modelling and Simulation*, Christchurch, New-Zealand, 2007.
- [75] R. SABBADIN, J. LANG et N. RAVOANJANAHRY : Purely epistemic markov decision processes. *In 22nd AAAI Conference on Artificial Intelligence (AAAI'07)*, pages 1057–1062, Vancouver, Canada, 2007.
- [76] V. THOMAS et M. AKPLOGAN : Using social actions and rl algorithms to build policies in decpomdp. *In Proceedings of International Conference on Intelligent Systems and Agents (IADIS ISA)*, Lagoa, Portugal, 2009.

3.1.1.6 Communications avec actes dans un congrès national

- [77] M. AKPLOGAN, F. GARCIA, A. JOANNON, R. MARTIN-CLOUAIRE et G. QUESNEL : Un modèle devs d'agent intelligent : application à la conduite des systèmes de culture. *In M. SIGAUD, O ; Vié, éditeur : Journées Francophones Planification Décision Apprentissage (JFPDA)*, juin 2009.
- [78] J. E. BERGEZ, D. LEENHARDT, L. MATON, F. GARCIA, D. SALLES et J. P. AMIGUES : Comment modéliser les pratiques des agriculteurs pour estimer la demande régionale en eau d'irrigation ? *In Actes du Colloque national du programme INRA-PSDR*, Lyon, France, mar 2005.

- [79] N. FORSELL et R. SABBADIN : Algorithme de résolution approchée basé sur la programmation linéaire pour les processus décisionnels de markov sur graphe. *In Premières Journées Françaises de Planification, Décision et Apprentissage (JFPDA 2006)*, pages 115–122, Toulouse, France, Mai 2006.
- [80] M. GOULARD, P. CHABRIER, H. JACTEL, D. GUYON, J. C. SAMALENS et I. HALDER (VAN) : Cartographie de probabilités d’attaques dans un peuplement forestier à partir d’une information exhaustive d’attaques et d’informations partielles sur le peuplement. *In Actes des XXXVIIèmes journées de Statistique de la SfdS*, page 6 p., Pau, 2005.
- [81] G. MARTIN, M. DURU et R. MARTIN-CLOUAIRE : Simuler la diversité des végétations au sein de l’exploitation d’élevage : des solutions virtuelles pour faire face à l’aléa climatique. *In Actes des Journées de printemps de l’AFPF*, pages 202–203, 2007.
- [82] R. MARTIN-CLOUAIRE et J. P. RELIER : Représentation et interprétation de plans de production flexibles. *In G. GARCIA, F; Verfaillie, éditeur : Journées Francophones sur la Planification, la Décision et l’Apprentissage pour la Conduite de Systèmes (JFPDA06)*, pages 149–156, INRA, Toulouse, 2006. URL <http://carlit.toulouse.inra.fr/JFPDA/procedings/martin-clouaire.pdf>.
- [83] N. PEYRARD, A. DENIS et F. FORBES : Comparaison de deux modélisations pour la classification de données géostatistiques. *In XXXVIIème Journées de Statistique*, Pau, France, 6 - 10 juin 2005.
- [84] N. PEYRARD et R. SABBADIN : Itération de la politique approche pour processus décisionnels de markov sur graphe. *In 15ème congrès Reconnaissance des Formes et Intelligence Artificielle (RFIA 2006)*, Tours, janvier 2006.
- [85] E. RACHELSON, P. FABIANI, J. L. FARGES, F. TEICHEL et F. GARCIA : Une approche du traitement du temps dans le cadre mdp : trois méthodes de découpage de la droite temporelle. *In G. GARCIA, F; Verfaillie, éditeur : Actes des Journées Francophones de Planification, Décision, Apprentissage (JFPDA)*, Toulouse, France, 2006. INRA.
- [86] E. RACHELSON, P. FABIANI et F. GARCIA : Un algorithme amélioré d’itération de la politique approchée pour les processus décisionnels semi-markoviens généralisés. *In Journées Françaises Planification, Décision, Apprentissage*, Mertz, France, 2008.
- [87] E. RACHELSON, P. FABIANI, F. GARCIA et G. QUESNEL : Une approche basée sur la simulation pour l’optimisation des processus décisionnels semi-markoviens généralisés. *In Conférence Francophone sur l’Apprentissage Automatique*, Porquerolles, France, Mai 2008.
- [88] N. RAVOANJANAHARY et R. SABBADIN : Processus de décision séquentielle purement épistémique en environnement incertain et partiellement observable. *In 5èmes journées nationales autour des PDM en IA (PDMIA’05)*, Lille, France, Juin 2005.
- [89] R. TRÉPOS, V. MASSON, M. O. CORDIER et C. GASCUEL-ODOUX : Induction de motifs spatiaux décrivant les chemins de ruissellement. *In Actes Représentation et Raisonnement sur le Temps et l’Espace (RTE’08)*, Montpellier, France, juin 2008.
- [90] D. VERSMISSE, J. SOULIÉ et G. QUESNEL : Une étude de cas dans le domaine des pêcheries pour la simulation de systèmes complexes. *In Actes de la 6ème Conférence Francophone de Modélisation et Simulation Modélisation, Optimisation et Simulation des Systèmes : Défis et Opportunités*, Rabat, Avril 2006.

3.1.1.7 Communications orales sans actes dans un congrès international ou national

- [91] J. E. BERGEZ, O. CRESPO et F. GARCIA : A multiobjective simulation-based method for optimizing irrigation strategies. *In Proceedings of the 21st European Conference on Operation Research (abstract)*, Reykjavik, Iceland, 2006.

- [92] X. CHARDON, C. RIGOLOT, C. BARATTE, R. MARTIN-CLOUAIRE, J. P. RELIER, C. RAISON, A. LE GALL, J. Y. DOURMAD, J. C. POUPA, L. DELABY, T. MORVAN, P. LETERME, J. M. PAILLAT, S. ESPAGNOL et P. FAVERDIN : A whole farm model to simulate the environmental impacts of animal farming systems : Melodie. *In MODNUT 2009 : 7th Int. Workshop on Modelling Nutrient Digestion and Utilization in Farm Animals*, Paris, sept 2009.
- [93] L. O. ERIKSSON, N. FORSELL, F. GARCIA, R. SABBADIN et P. WIKSTRÖM : A graph markov decision process model for harverst planning under risk of windthrow. *In Proceedings of the 21st European Conference on Operation Research (abstract)*, Reykjavik, Iceland, jul 2006. The Association of European OR Societies.
- [94] A. FRANC et N. PEYRARD : Epidemiology, metapopulations and invasive species : a common modelling framework? *In 9th International Workshop on Plant Disease Epidemiology - Plant Disease Epidemiology : Facing 21st century challenges*, Lacanau, France, avril 2005.
- [95] R. MARTIN-CLOUAIRE et J. P. RELIER : Dynamic resource allocation in agricultural production systems. *In Proceedings of the 21st European Conference on Operation Research (abstract)*, Reykjavik, Iceland, 2006. The Association of European OR Societies.
- [96] R. MARTIN-CLOUAIRE et J. P. RELIER : A generic framework for simulating agricultural production systems. *In MODNUT 2009 : 7th Int. Workshop on Modelling Nutrient Digestion and Utilization in Farm Animals*, Paris, sept 2009.
- [97] N. PEYRARD et A. FRANC : Pair and bethe approximation for the contact process on a graph. *In 9th International Workshop on Plant Disease Epidemiology - Plant Disease Epidemiology : Facing 21st century challenges*, Lacanau, France, avril 2005.
- [98] N. PEYRARD et A. FRANC : On a bethe approximation for a contact process on a graph. *In Mathematics and its applications*, Turin - Italie, 3-7 juillet 2006. Session speciale "Mathematical Population Dynamics, Mathematical Epidemiology and Other Biomedical Applications".
- [99] N. PEYRARD, A. FRANC et U. DIECKMANN : Improving the understanding of the influence of the network structure on epidemic dynamics using long range correlation pair approximations. *In IMA Conference on Mathematics in the Science of Complex Systems*, Warwick, RU, 18-21 septembre 2006.
- [100] N. PEYRARD et R. SABBADIN : Mean field approximation of the policy iteration algorithm for graph based markov decision processes. *In IMA Conference on Mathematics in the Science of Complex Systems*, 2006.
- [101] N. PEYRARD et R. SABBADIN : Model-based adaptive spatial sampling for occurrence map reconstruction. *In First Conference on Computational Sustainability, CompSust09*, june 2009.
- [102] E. RACHELSON, G. QUESNEL, F. GARCIA et P. FABIANI : Approximate policy iteration for generalized semi-markov decision processes : an improved algorithm. *In European Workshop on Reinforcement Learning*, Villeneuve d'ascq, France, June 2008.
- [103] T. J. REGAN, I. CHADÈS et H. P. POSSINGHAM : Making optimal management decisions for weed eradication when observation is incomplete. *In Conference of the Ecological Society of Australia*, 2005.
- [104] C. RIGOLOT, J. P. RELIER, R. MARTIN-CLOUAIRE, P. FAVERDIN, C. BARATTE et J. Y. DOURMAD : A model relating management practices and nutrient flows in indoor farming. *In MODNUT 2009 : 7th Int. Workshop on Modelling Nutrient Digestion and Utilization in Farm Animals*, Paris, sept 2009.

3.1.1.8 Communications par affiche dans un congrès international ou national

- [105] M. CHOISY, N. PEYRARD et R. SABBADIN : A probabilistic decision framework to optimize the dynamics of a network evolution : application to the control of childhood diseases. *In European Conference on Complex Systems*, 2007.
- [106] M.-O. CORDIER, V. MASSON, A. SALLEB, R. TREPOS, C. GASCUEL et F. GARCIA : Sacadeau project : recommending actions from simulation results. *In Proceedings of the 5th ECAI Workshop on Binding Environmental Sciences and Artificial Intelligence (poster)*, 17th European Conference on Artificial Intelligence, Riva del Garda, Trento, Italy, aug 2006.
- [107] J. DURY, N. SCHALLER, M. AKPLOGAN, C. AUBRY, J. E. BERGEZ, F. GARCIA, A. JOANNON, B. LACROIX, P. MARTIN, A. REYNAUD et O. THÉROND : Modelling crop allocation decision-making processes to simulate dynamics of agricultural land uses at farm and landscape level. *In proceedings of Farming Systems Design*, Monterey, CA, USA, August 23-26 2009.
- [108] N. PEYRARD, A. DENIS et F. FORBES : How to classify point-referenced data using the EM algorithm? *In International workshop on Spatio-Temporal Modelling (METMAT3)*, Pamplune - Espagne, 27-29 septembre 2006.
- [109] N. PEYRARD, R. SABBADIN, E. LO-PELZER et J.-N. AUBERTOT : A graph-based markov decision process applied to the optimization of strategies for integrated management of diseases. *In American Phytopathological Society and Society of Nematologist joint meeting*, 2007.

3.1.1.9 Ouvrages scientifiques (ou chapitres de ces ouvrages)

- [110] J. E. BERGEZ, F. GARCIA, D. LEENHARDT et L. MATON : *Optimising Irrigation Management at the Plot Scale to Participate at the Regional Scale Water Resource Management*, chapitre 7, pages 141–160. Elsevier, 2006.
- [111] J. E. BERGEZ, F. GARCIA et D. WALLACH : *Modeling and optimising decisions within crop models*, chapitre 6, pages 173–208. Elsevier, 2006. ISBN 978-0-444-52135-4.
- [112] M. BOUSSARD, M. BOUZID, A. I. MOUADDIB, R. SABBADIN et P. WENG : *Critères non classiques*, volume 1, chapitre 5, pages 179–218. Lavoisier / Hermes Science Publications, 2008.
- [113] H. CARDOT, C. CRAMBES et P. SARDA : Ozone pollution forecasting using conditional mean and conditional quantiles with functional covariates. *In P. V. P. HARDLE, W ; Mori, éditeur : Statistical Methods for Biostatistics and Related Fields*, pages 221–244. Springer Verlag, 2005.
- [114] H. CARDOT et P. SARDA : Linear regression models for functional data. *In W. HARDLE, éditeur : The art of semiparametrics*. Physica Verlag, Contributions to Statistics, 2006.
- [115] M. CERF, P. FAVERDIN et F. GARCIA : *Formalisation de l'interaction entre processus décisionnels et processus biophysiques*. Quae, 2009.
- [116] I. CHADÈS : *Conservation de la biodiversité*, chapitre 7. Lavoisier / Hermes Science Publications, Paris, France, 2008.
- [117] M. J. CROS, F. GARCIA, R. MARTIN-CLOUAIRE et J. P. RELIER : *Modeling and simulation*, volume 6, chapitre 3, pages 109–124. CIGR The International Commission of Agricultural Engineering, ASABE, St. Joseph, Michigan, USA., 2006. ISBN 1-892769-54-9.
- [118] D. DUBOIS, H. FARGIER, H. PRADE et R. SABBADIN : *Concepts et méthodes pour l'aide la décision*, volume 2 de *Traité IC2, série Informatique et Systèmes d'Information*, chapitre Critères qualitatifs pour la décision dans l'incertain, pages 99–137. Hermès Sciences, 2006. Eds. D. Bouyssou, D. Dubois, M. Pirlot et H. Prade.
- [119] D. DUBOIS, H. FARGIER, H. PRADE et R. SABBADIN : *Critères qualitatifs pour la décision dans l'incertain*, volume 2 de *Traité IC2, série Informatique et Systèmes d'Information*, pages 99–137. Hermès Sciences, 2006.

- [120] D. DUBOIS, H. FARGIER, H. PRADE et R. SABBADIN : *Decision-making Process - Concepts and Methods*, chapitre A Survey of Qualitative Decision Rules under Uncertainty, pages 435–473. ISTE-Wiley, 2009.
- [121] A. FRANC et N. PEYRARD : Rôle de la géométrie du réseau d'interaction dans l'émergence d'une maladie. In I. BANOUIN, J. Sache, éditeur : *Les maladies émergentes chez le végétal, l'animal et l'homme. Enjeux et stratégies d'analyse épidémiologique*. QUAE, quae édition, 2009.
- [122] A. FRANC, N. PEYRARD et B. ROCHE : *Introduction à l'épidémiologie quantitative des maladies infectieuses*, chapitre Propagation d'agents pathogènes dans les réseaux, pages 81–116. De Boeck Université, 2009.
- [123] F. GARCIA : *Processus décisionnels de Markov*, chapitre 1. Lavoisier / Hermes Science Publications, 2008.
- [124] C. LOYCE, J. P. RELIER et J. M. MEYNARD : *Managing wheat for ethanol production. A multiple criteria approach*, chapitre 20, pages 419–427. Elsevier, 2006. ISBN 978-0-444-52135-4.
- [125] L. PÉRET et F. GARCIA : *Approches de résolution en ligne*, chapitre 5. Lavoisier / Hermes Science Publications, 2008.
- [126] R. SABBADIN : *Concepts et méthodes pour l'aide à la décision*, volume 2 de *Traité IC2, série Informatique et Systèmes d'Information*, chapitre Planification dans l'incertain et processus décisionnels de Markov, pages 203–252. Hermès Sciences, 2006. Eds. D. Bouyssou, D. Dubois, M. Pirlot et H. Prade.
- [127] R. SABBADIN : *Planification dans l'incertain et processus décisionnels de Markov*, volume 2 de *Traité IC2, série Informatique et Systèmes d'Information*, pages 203–252. Hermès Sciences, 2006.
- [128] R. SABBADIN : *Decision-making Process - Concepts and Methods*, chapitre Planning under Uncertainty with Markov Decision Processes, pages 541–577. ISTE-Wiley, 2009.
- [129] O. SIGAUD et F. GARCIA : *Apprentissage par renforcement*, chapitre 2. Lavoisier / Hermes Science Publications, 2008.

3.1.1.10 Ouvrages de vulgarisation (ou chapitres de ces ouvrages)

3.1.1.11 Directions d'ouvrages ou de revues

3.1.1.12 Autres productions : bases de données, logiciels enregistrés, traductions, comptes rendus d'ouvrages, rapports de fouilles, guides techniques, catalogues d'exposition, rapports intermédiaires de grands projets internationaux, etc.

- [130] S. BACKÉUS, L. O. ERIKSSON et F. GARCIA : Bringing climate change uncertainty into the forest management planning problem. Rapport technique, 2009.
- [131] J. E. BERGEZ, H. CARDOT et F. GARCIA : Quantile regression trees for yield prediction using a simulation model. Rapport technique, 2005.
- [132] J. E. BERGEZ, P. CHABRIER, F. GARCIA, C. GARY, M. H. JEUFFROY, H. RAYNAL et D. WALLACH : Projet record : expression des besoins fonctionnels. Rapport technique, INRA, juin 2007.
- [133] J. E. BERGEZ et F. GARCIA : Integrating anticipatory behaviour in decision making model. the case of irrigation management in south-western france. Rapport technique, 2009.
- [134] N. BRISSON, P. CHABRIER, B. GABRIELLE, F. GARCIA, M. H. JEUFFROY, B. MARY, N. MOITRIER, H. MONOD, D. WALLACH et J. WERY : Projet record : Document de vision. Rapport technique, INRA, 2006.

- [135] P. CHABRIER et H. RAYNAL : Projet record : expression des besoins non fonctionnels. Rapport technique, INRA, juin 2007.
- [136] I. CHADÈS, T. G. MARTIN et J. M. R. CURTIS : Are recovery targets for endangered species realistic the case of sea otter and abalone. Rapport technique, 2009.
- [137] O. CRESPO, J. E. BERGEZ et F. GARCIA : Multiobjective optimization subject to uncertainty : application to agricultural resources management. Rapport technique, 2009.
- [138] M. DURU, J. E. BERGEZ et R. MARTIN-CLOUAIRE : Cognitive tools to support learning about grazing management. Rapport technique, 2009.
- [139] F. GARCIA et H. RAYNAL : Rapport d'activité du projet record. Rapport technique, INRA, décembre 2008.
- [140] B. GOFFINET, R. FAIVRE et R. MARTIN-CLOUAIRE : La modélisation à l'inra. Rapport technique, INRA, 2006.
- [141] G. MARTIN, R. MARTIN-CLOUAIRE, J. P. RELIER et M. DURU : A conceptual model of grassland-based livestock systems. Rapport technique, 2009.
- [142] G. MARTIN, R. MARTIN-CLOUAIRE, J. P. RELIER et M. DURU : Sediver : a simulation model to design grassland-based livestock systems and their management strategies. Rapport technique, 2009.
- [143] R. MARTIN-CLOUAIRE et J. P. RELIER : Fondements ontologiques des systèmes pilotés. Rapport technique, UBIA-INRA, 2006. URL http://carlit.toulouse.inra.fr/diese/docs/ri_ontologie.pdf.
- [144] E. McDONALD-MADDEN, I. CHADÈS, M. A. MCCARTHY, M. LINKIE et H. P. POSSINGHAM : Allocating conservation resources between areas where persistence of a species is uncertain. Rapport technique, 2009.
- [145] T. J. REGAN, I. CHADÈS et H. P. POSSINGHAM : Optimal strategies for managing invasive plants in partially observable systems. Rapport technique, 2009.
- [146] R. SABBADIN : Mémoire d'hdr : Modèles et algorithmes pour la décision dans l'incertain. Rapport technique, Université Paul Sabatier, Toulouse, Février 2009.
- [147] J. SALMON-MONVIOLA, F. GARCIA, C. GASCUEL-ODOUX, F. TORTRAT, M. O. CORDIER, V. MASSON et R. TRÉPOS : Effect of spatial and temporal distribution of herbicide applications on stream water pollution by using a decision model. Rapport technique, 2009.
- [148] R. TRÉPOS : *Apprentissage symbolique à partir de données de simulation pour l'aide à la décision. Gestion d'un bassin versant pour une meilleure qualité de l'eau.* Thèse de doctorat, Université de Rennes, janvier 2008.
- [149] R. TRÉPOS, A. SALLEB, M. O. CORDIER, V. MASSON et C. GASCUEL-ODOUX : Building actions from classification rules. Rapport technique, INRA, 2009.

3.1.2 Équipes SaAB & PFBI

3.1.2.1 Articles dans des revues internationales ou nationales avec comité de lecture répertoriées dans les bases de données internationales ou équivalentes

- [150] P. ABAD, J. GOUZY, J.-M. AURY, P. CASTAGNONE-SERENO, E. G. J. DANCHIN, E. DELEURY, L. PERFUS-BARBEOCH, V. ANTHOUARD, F. ARTIGUENAVE, V. C. BLOK, M.-C. CAILLAUD, P. M. COUTINHO, C. DASILVA, F. DE LUCA, F. DEAU, M. ESQUIBET, T. FLUTRE, J. V. GOLDSTONE, N. HAMAMOUCHE, T. HEWEZI, O. JAILLON, C. JUBIN, P. LEONETTI, M. MAGLIANO, T. R. MAIER, G. V. MARKOV, P. MCVEIGH, G. PESOLE, J. POULAIN, M. ROBINSON-RECHAVI, E. SALET, B. SÉGURENS, D. STEINBACH, T. TYTGAT, E. UGARTE, C. van GHELDER, P. VERONICO, T. J.

- BAUM, M. BLAXTER, T. BLEVE-ZACHEO, E. L. DAVIS, J. J. EWBANK, B. FAVERY, E. GRENIER, B. HENRISSAT, J. T. JONES, V. LAUDET, A. G. MAULE, H. QUESNEVILLE, M.-N. ROSSO, T. SCHIEX, G. SMANT, J. WEISSENBACH et P. WINCKER : Genome sequence of the metazoan plant-parasitic nematode *meloidogyne incognita*. *Nat Biotechnol*, 26(8):909–15, Aug 2008. URL <http://view.ncbi.nlm.nih.gov/pubmed/18660804>.
- [151] S. AUBOURG, V. BRUNAUD, C. BRUYERE, M. COCK, R. COOKE, A. COTTET, A. COULOUX, P. DEHAIS, G. DELEAGE, A. DUCLERT, M. ECHEVERRIA, A. ESCHBACH, D. FALCONET, G. FILIPPI, C. GASPIN, C. GEOURJON, J. GRIENENBERGER, G. HOULNE, E. JAMET, F. LECHAUVE, O. LELEU, P. LEROY, R. MACHE, C. MEYER, H. NEDJARI, I. NEGRUTIU, V. ORSINI, E. PEYRETAILLADE, C. POMMIER, J. RAES, J. RISLER, S. RIVIERE, S. ROMBAUTS, P. ROUZE, M. SCHNEIDER, P. SCHWOB, I. SMALL, G. SOUMAYET-KAMPETENGA, D. STANKOVSKI, C. TOFFANO, M. TOGNOLLI, M. CABOCHE et A. LECHARNY : GeneFarm, structural and functional annotation of Arabidopsis gene and protein families by a network of experts. *Nucleic Acids Res*, 33(Database issue):D641–6, 2005. URL <http://view.ncbi.nlm.nih.gov/pubmed/15608279>.
- [152] S. AUBOURG, M.-L. MARTIN-MAGNIETTE, V. BRUNAUD, L. TACONNAT, F. BITTON, S. BALZERGUE, P. E. JULLIEN, M. INGOUFF, V. THAREAU, T. SCHIEX, A. LECHARNY et J.-P. RENOU : Analysis of CATMA transcriptome data identifies hundreds of novel functional genes and improves gene models in the Arabidopsis genome. *BMC Genomics*, 8:401, 2007. URL <http://view.ncbi.nlm.nih.gov/pubmed/17980019>.
- [153] G. BLANC, A. CHARCOSSET, B. MANGIN, A. GALLAIS et L. MOREAU : Connected populations for detecting quantitative trait loci and testing for epistasis : an application in maize. *Theoretical and Applied Genetics*, 113(2):2006–224, 2006.
- [154] J. BLANCHET et M. VIGNES : A model-based approach to gene clustering with missing observation reconstruction in a markov random field framework. *Journal of Computational Biology*, 16(3):475–486, mar 2009. URL <http://view.ncbi.nlm.nih.gov/pubmed/19254185>.
- [155] S. BOISSET, T. GEISSMANN, E. HUNTZINGER, P. FECHTER, N. BENDRIDDI, M. POSSEDKO, C. CHEVALIER, A. C. HELFER, Y. BENITO, A. JACQUIER, C. GASPIN, F. VANDENESCH et P. ROMBY : Staphylococcus aureus RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. *Genes Dev*, 21(11):1353–66, Jun 2007. URL <http://view.ncbi.nlm.nih.gov/pubmed/17545468>.
- [156] S. BOITARD, J. ABDALLAH, H. de ROCHAMBEAU, C. CIERCO-AYROLLES et B. MANGIN : Linkage disequilibrium interval mapping of quantitative trait loci. *BMC Genomics*, 7:54, 2006. URL <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?cmd=prlinks&dbfrom=pubmed&retmode=ref&id=16542433>.
- [157] S. BOITARD et P. LOISEL : Probability distribution of haplotype frequencies under the two-locus Wright–Fisher model by diffusion approximation. *Theoretical Population Biology*, 71(3):380–391, 2007. URL <http://view.ncbi.nlm.nih.gov/pubmed/17316725>.
- [158] J. BURSTIN, P. MARGET, M. HUART, A. MOESSNER, B. MANGIN, C. DUCHENE, B. DESPREZ, N. MUNIER-JOLAIN et G. DUC : Developmental genes have pleiotropic effects on plant morphology and source capacity, eventually impacting on seed protein content and productivity in pea. *Plant Physiology*, 144(2):768–781, jun 2007. URL <http://view.ncbi.nlm.nih.gov/pubmed/17449650>.
- [159] S. BÉRARD, A. BERGERON, C. CHAUVE et C. PAUL : Perfect sorting by reversal is not always difficult. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 4(1):4–16, 2007. URL <http://www.cecm.sfu.ca/~cchauve/Publications/TCBB-WAB105.pdf>.
- [160] S. BÉRARD, C. CHAUVE et C. PAUL : A more efficient algorithm for perfect sorting by reversals. *Information Processing Letters*, 106(3):90–95, 4 2008. URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.132.3556&rep=rep1&type=pdf>.

- [161] S. BÉRARD, F. NICOLAS, J. BUARD, O. GASCUEL et E. RIVALS : Fast and specific alignment method for minisatellite maps applied to human y chromosome minisatellite msy1. *Evolutionary Bioinformatics Online*, 2:327–344, 2006.
- [162] S. B. CANNON, L. STERCK, S. ROMBAUTS, S. SATO, F. CHEUNG, J. GOUZY, X. WANG, J. MUDGE, J. VASDEWANI, T. SCHIEX, M. SPANNAGL, E. MONAGHAN, C. NICHOLSON, S. J. HUMPHRAY, H. SCHOOF, K. F. X. MAYER, J. ROGERS, F. QUÉTIÉ, G. E. OLDROYD, F. DEBELLÉ, D. R. COOK, E. F. RETZEL, B. A. ROE, C. D. TOWN, S. TABATA, Y. Van de PEER et N. D. YOUNG : Legume genome evolution viewed through the medicago truncatula and lotus japonicus genomes. *Proc Natl Acad Sci U S A*, 103(40):14959–64, Oct 2006. URL <http://view.ncbi.nlm.nih.gov/pubmed/17003129>.
- [163] R. CATHELIN, F. LOPEZ et C. KLOPP : Agscan : a pluggable microarray image quantification software based on the imagej library. *Bioinformatics*, 23(2):247–8, Jan 2007. URL <http://view.ncbi.nlm.nih.gov/pubmed/17092988>.
- [164] B. CLOUET-D’ORVAL, C. GASPIN et A. MOUGIN : Two different mechanisms for tRNA ribose methylation in Archaea : a short survey. *Biochimie*, 87(9-10):889–95, Sep-Oct 2005. URL <http://view.ncbi.nlm.nih.gov/pubmed/16164996>.
- [165] S. de GIVRY et L. JEANNIN : A unified framework for partial and hybrid search methods in constraint programming. *Computer & Operations Research*, 33(10):2805–2833, 2006. URL <http://www.inra.fr/internet/Departements/MIA/T/degivry/GivryJeannin04.pdf>.
- [166] S. de GIVRY, M. BOUCHEZ, P. CHABRIER, D. MILAN et T. SCHIEX : Carthagene : multipopulation integrated genetic and radiation hybrid mapping. *Bioinformatics*, 21(8):1703–4, Apr 2005. URL <http://view.ncbi.nlm.nih.gov/pubmed/15598829>.
- [167] C. DÉSSERT, M. J. DUCLOS, P. BLAVY, F. LECERF, F. MOREEWS, C. KLOPP, M. AUBRY, F. HERAULT, P. LE ROY, C. BERRI, M. DOUAIRE, C. DIOT et S. LAGARRIGUE : Transcriptome profiling of the feeding-to-fasting transition in chicken liver. *BMC Genomics*, 9:611, 2008. URL <http://view.ncbi.nlm.nih.gov/pubmed/19091074>.
- [168] T. FARAUT, S. de GIVRY, P. CHABRIER, T. DERRIEN, F. GALIBERT, C. HITTE et T. SCHIEX : A comparative genome approach to marker ordering. *Bioinformatics*, 23(2):e50–6, Jan 2007. URL <http://view.ncbi.nlm.nih.gov/pubmed/17237105>.
- [169] T. FARAUT, S. de GIVRY, C. HITTE, Y. LAHLIB-MANSAIS, M. MORISSON, D. MILAN, T. SCHIEX, B. SERVIN, A. VIGNAL, F. GALIBERT et M. YERLE : Contributions of radiation hybrids to genome mapping in domestic animals. *Cytogenetic and Genome Research*, A paraître, 2009.
- [170] O. FILANGI, Y. BEAUSSE, A. ASSI, L. LEGRAND, J.-M. LARRÉ, V. MARTIN, O. COLLIN, C. CARON, H. LEROY et D. ALLOUCHE : Biomaj : a flexible framework for databanks synchronization and processing. *Bioinformatics*, 24(16):1823–5, Aug 2008. URL <http://view.ncbi.nlm.nih.gov/pubmed/18593718>.
- [171] S. FOISSAC, J. GOUZY, S. ROMBAUTS, C. MATHÉ, J. AMSELEM, L. STERK, Y. van de PEER, P. ROUZÉ et T. SCHIEX : Genome annotation in plants and fungi : Eugène as a model platform. *Current Bioinformatics*, 3(2):87–97, 5 2008. URL <http://www.bentham.org/cbio/openacsarticles/cbio3-2/0003CBIO.pdf>.
- [172] S. FOISSAC et T. SCHIEX : Integrating alternative splicing detection into gene prediction. *BMC Bioinformatics*, 6:25, 2005. URL <http://www.biomedcentral.com/1471-2105/6/25>.
- [173] T. GEISSMANN, C. CHEVALIER, M. CROS, S. BOISSET, P. FECHTER, C. NOIROT, J. SCHRENZEL, P. FRANÇOIS, F. VANDENESCH, C. GASPIN et P. ROMBY : A search for small noncoding RNAs in staphylococcus aureus reveals a conserved sequence motif for regulation. *Nucleic Acids Research*, 2009.
- [174] J. GOUZY, S. CARRERE et T. SCHIEX : Framedp : sensitive peptide detection on noisy matured sequences. *Bioinformatics*, Jan 2009. URL <http://view.ncbi.nlm.nih.gov/pubmed/19153134>.

- [175] H. GROSJEAN, C. GASPIN, C. MARCK, W. A. DECATUR et V. de CRÉCY-LAGARD : Rnomics and modomics in the halophilic archaea *haloferax volcanii* : identification of rna modification genes. *BMC Genomics*, 9:470, 2008. URL <http://view.ncbi.nlm.nih.gov/pubmed/18844986>.
- [176] F. HERAS, J. LARROSA, S. de GIVRY et T. SCHIEX : 2006 and 2007 max-sat evaluations : Contributed instances. *Journal of Satisfiability, Boolean Modeling and Computation*, 4:239–250, 2008. URL http://www.st.ewi.tudelft.nl/jsat/content/volume4/JSAT4_12_Heras.pdf.
- [177] M. JOURJON, S. JASSON, J. MARCEL, B. NGOM et B. MANGIN : MCQTL : multi-allelic qtl mapping in multi-cross design. *Bioinformatics*, 21(1):128–30, 2005. URL <http://view.ncbi.nlm.nih.gov/pubmed/15319261>.
- [178] J. LARROSA, F. HERAS et S. de GIVRY : A logical approach to efficient max-sat solving. *Artificial Intelligence*, 172(2-3), 2008.
- [179] A. LUCAS et S. JASSON : Using amap and ctc packages for huge clustering. *R News*, 5(6):58–60, 2006.
- [180] B. MANGIN, P. GARNIER-GERE et C. CIERCO-AYROLLES : The estimator of the optimal measure of allelic association : Mean, variance and probability distribution when the sample size tends to infinity. *Statistical Applications in Genetics and Molecular Biology*, 7(1), 2008. URL <http://view.ncbi.nlm.nih.gov/pubmed/18597666>.
- [181] E. MARQUES, S. de GIVRY, P. STOTHARD, B. MURDOCH, Z. WANG, J. WOMACK et S. S. MOORE : A high resolution radiation hybrid map of bovine chromosome 14 identifies scaffold rearrangement in the latest bovine assembly. *BMC Genomics*, 8:254, 2007. URL <http://view.ncbi.nlm.nih.gov/pubmed/17655763>.
- [182] M. MORISSON, M. DENIS, D. MILAN, C. KLOPP, S. LEROUX, S. BARDES, F. PITEL, F. VIGNOLES, M. GÉRUS, V. FILLON, M. DOUAUD et A. VIGNAL : The chicken RH map : current state of progress and microchromosome mapping. *Cytogenet Genome Res*, 117(1-4):14–21, 2007. URL <http://view.ncbi.nlm.nih.gov/pubmed/17675840>.
- [183] C. NOIROT, C. GASPIN, T. SCHIEX et J. GOUZY : LeARN : a platform for detecting, clustering and annotating non-coding RNAs. *BMC Bioinformatics*, 9:21, 2008. URL <http://view.ncbi.nlm.nih.gov/pubmed/18194551>.
- [184] C. PRALET, G. VERFAILLIE et T. SCHIEX : An algebraic graphical model for decision with uncertainties, feasibilities, and utilities. *Journal of Artificial Intelligence Research*, 29:421–489, 2007. URL <http://www.jair.org/media/2151/live-2151-3404-jair.pdf>.
- [185] A. PRASAD, T. SCHIEX, S. MCKAY, B. MURDOCH, Z. WANG, J. E. WOMACK, P. STOTHARD et S. S. MOORE : High resolution radiation hybrid maps of bovine chromosomes 19 and 29 : comparison with the bovine genome sequence assembly. *BMC Genomics*, 8:310, 2007. URL <http://view.ncbi.nlm.nih.gov/pubmed/17784962>.
- [186] M. RENALIER, N. JOSEPH, C. GASPIN, P. THEBAULT et A. MOUGIN : The Cm56 tRNA modification in archaea is catalyzed either by a specific 2'-O-methylase, or a C/D sRNP. *RNA*, 11(7):1051–63, 2005. URL <http://view.ncbi.nlm.nih.gov/pubmed/15987815>.
- [187] M. SANCHEZ, S. de GIVRY et T. SCHIEX : Mendelian error detection in complex pedigrees using weighted constraint satisfaction techniques. *Constraints*, 13(1):130–154, 2008. URL <http://dx.doi.org/10.1007/s10601-007-9029-5>.
- [188] P. THÉBAULT, S. de GIVRY, T. SCHIEX et C. GASPIN : Searching rna motifs and their intermolecular contacts with constraint networks. *Bioinformatics*, 22(17):2074–80, Sep 2006. URL <http://view.ncbi.nlm.nih.gov/pubmed/16820426>.
- [189] M. VIGNES et F. FORBES : Gene clustering via integrated markov models combining individual and pairwise features. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 6(2):260–270, apr 2009. URL <http://view.ncbi.nlm.nih.gov/pubmed/19407350>.

- [190] M. ZYTNICKI, C. GASPIN, S. de GIVRY et T. SCHIEX : Bound arc consistency for weighted csps. *Journal of Artificial Intelligence Research*, 35, 2009.
- [191] M. ZYTNICKI, C. GASPIN et T. SCHIEX : DARN ! a soft constraint solver for RNA motif localization. *Constraints*, 13(1):91–109, 2008. URL <http://dx.doi.org/10.1007/s10601-007-9033-9>.

3.1.2.2 Articles dans des revues avec comité de lecture non répertoriées dans des bases de données internationales

- [192] C. GASPIN et P. THÉBAULT : Identification d'ARN non codants. *Biofutur*, 251:33–36, 2005.
- [193] C. PRALET, T. SCHIEX et G. VERFAILLIE : Algorithmes et complexités génériques pour différents cadres de décision séquentielle dans l'incertain. *Revue d'Intelligence Artificielle*, 21:459–488, 2007. URL <http://www.inra.fr/mia/T/schiex/Export/ria2007.pdf>.
- [194] C. PRALET, G. VERFAILLIE et T. SCHIEX : Un cadre graphique et algébrique pour les problèmes de décision incluant incertitudes, faisabilités et utilités. *Revue d'Intelligence Artificielle*, 21(3), 2007. URL <http://www.inra.fr/mia/T/schiex/Export/ria2006.pdf>.

3.1.2.3 Articles dans des revues sans comité de lecture

3.1.2.4 Conférences données à l'invitation du Comité d'organisation dans un congrès national ou international

- [195] R. DECHTER, S. de GIVRY, R. MARINESCU et T. SCHIEX : Combinatorial optimization for graphical models. In *Selected tutorial at IJCAI'2009.*, Pasadena (CA), USA., 2009.
- [196] J. LARROSA et T. SCHIEX : Soft constraint solving. Invited tutorial at CP'2006, 2006.
- [197] B. MANGIN : Combiner l'analyse de liaison et le déséquilibre de liaison pour cartographier finement les QTL dans le pedigree animal. In *Bilan des projets GENANIMAL 2003-2006*, Dourdan, 10 2007.
- [198] B. MANGIN : Une contribution de génoplatte en bioinformatique. In *Les 10 ans de GÉNOPLANTE*, Paris, apr 2009.
- [199] P. MESEGUER et T. SCHIEX : Constraint processing. Invited tutorial at IJCAI'05, 2005.
- [200] C. PRALET et T. SCHIEX : Des réseaux de contraintes valués aux problèmes de décision séquentielle. In *Journées Nationales de l'Intelligence Artificielle Fondamentale*, 2007. URL <http://gdri3iaf.info.univ-angers.fr/spip.php?rubrique22>.
- [201] C. PRALET, T. SCHIEX et G. VERFAILLIE : Quantified cost functions for reasoning under uncertainty. In *International Workshop on Mathematics of Constraint Satisfaction : Algebra, Logic and Graph Theory*, St Anne's College, University of Oxford, 2006. URL <http://www.cs.rhul.ac.uk/home/green/mathscsp/>.
- [202] T. SCHIEX : From gene prediction to integrated annotation pipeline (Invited talk). In *Plant'GEMS*, Amsterdam, The Netherlands, 2005.
- [203] T. SCHIEX : Cost function optimization and local consistency. In *International Workshop on Mathematics of Constraint Satisfaction : Algebra, Logic and Graph Theory*, St Anne's College, University of Oxford, 2006. URL <http://www.cs.rhul.ac.uk/home/green/mathscsp/>.
- [204] T. SCHIEX : Combinatorial optimization for pedigree analysis using graphical models. Invited talk at the 2nd channel conference of the International Biometry Society, apr 2009.

3.1.2.5 Communications avec actes dans un congrès international

- [205] S. BOITARD : Computation of the haplotype frequency distribution under the two-locus wright-fisher model. a diffusion approximation method. *In Proc. of European mathematical genetics meeting in annals of human genetics 69(6) ,p. 765*, pages 764–774, Le Kremlin-Bicêtre, France, 2005.
- [206] S. BOUVERET, S. de GIVRY, F. HERAS, J. LARROSA, E. ROLLON, M. SANCHEZ, T. SCHIEX, G. VERFAILLIE et M. ZYTNICKI : Max-csp competition 2007 : toolbar/toulbar2 solver brief description. *In C. R. O. van DONGEN, MRC; Lecoutre, éditeur : Proceedings of the Second International CSP Solver Competition*, pages 19–21, 2008.
- [207] S. BÉRARD, A. BERGERON, C. CHAUVE et C. PAUL : Perfect sorting by reversal is not always difficult. *In Proc. of the 5th Workshop on Algorithms in Bioinformatics (WABI'05)*, numéro 3692 in LNCS, pages 228–238, 2005.
- [208] M. COOPER, S. de GIVRY, M. SANCHEZ, T. SCHIEX et M. ZYTNICKI : Virtual arc consistency for weighted csp. *In Proc. of AAAI'2008*, Chicago, USA, 2008. URL <http://www.aaai.org/Papers/AAAI/2008/AAAI08-040.pdf>.
- [209] M. COOPER, S. de GIVRY et T. SCHIEX : Optimal soft arc consistency. *In Proc. of IJCAI-07*, pages 68–73, Hyderabad, India, 2007. URL <http://www.ijcai.org/papers07/Papers/IJCAI07-009.pdf>.
- [210] S. de GIVRY, T. SCHIEX et T. VERFAILLIE : Exploiting tree decomposition and soft local consistency in weighted CSP. *In Proc. of AAAI'2006*, Boston (MA), USA, 2006. URL <http://www.aaai.org/Papers/AAAI/2006/AAAI06-004.pdf>.
- [211] S. de GIVRY, M. ZYTNICKI, F. HERAS et J. LARROSA : Existential arc consistency : getting closer to full arc consistency in weighted CSPs. *In Proc. of IJCAI-05*, pages 84–89, Edinburgh, Scotland, 2005. URL <http://www.ijcai.org/papers/0827.pdf>.
- [212] T. FARAUT, S. de GIVRY, P. CHABRIER, T. DERRIEN, F. GALIBERT, C. HITTE et T. SCHIEX : A comparative genome approach to marker ordering. *In Proc. of ECCB'2006*, 2006. URL <http://www.inra.fr/mia/T/schiex/Export/ComparativeMapping.pdf>.
- [213] A. FAVIER, S. de GIVRY et P. JÉGOU : Exploiting problem structure for solution counting. *In Proc. of CP'2009*, LNCS, Lisbon, Portugal, 2009.
- [214] C. PRALET, T. SCHIEX et G. VERFAILLIE : Decomposition of multi-operators queries on semiring-based graphical models. *In Proc. of CP'2006*, Nantes, France, 2006. URL <http://www.inra.fr/mia/T/schiex/Export/cp2006.pdf>.
- [215] C. PRALET, T. SCHIEX et G. VERFAILLIE : From influence diagrams to multioperator cluster dags. *In Proc. of UAI'2006*, Boston (MA), USA, 2006. URL <http://www.inra.fr/mia/T/schiex/Export/uai2006.pdf>.
- [216] C. PRALET, G. VERFAILLIE et T. SCHIEX : Decision with uncertainties, feasibilities and utilities : towards a unified algebraic framework. *In Proc. of European Conference on Artificial Intelligence (ECAI'06)*, Trento, Italy, 2006. URL <http://www.inra.fr/mia/T/schiex/Export/ecai2006.pdf>.
- [217] M. SANCHEZ, D. ALLOUCHE, S. de GIVRY et T. SCHIEX : Russian doll search with tree decomposition. *In Proc. of IJCAI'09*, Pasadena (CA), USA, 2009. URL <http://www.inra.fr/mia/T/schiex/Export/IJCAI2009.pdf>.
- [218] M. SANCHEZ, S. BOUVERET, S. de GIVRY, F. HERAS, P. JÉGOU, J. LARROSA, S. NDIAYE, E. ROLLON, T. SCHIEX, C. TERRIOUX, G. VERFAILLIE et M. ZYTNICKI : Max-csp competition 2008 : toulbar2 solver description. *In Proceedings of the Third International CSP Solver Competition*, 2008.
- [219] M. ZYTNICKI, C. GASPIN et T. SCHIEX : A new local consistency for weighted CSP dedicated to long domains. *In Proceedings of the 2006 ACM Symposium on Applied Computing*, pages 394–398, Dijon, France, 2006. URL <http://www.inra.fr/mia/T/zytnicki/Files/230406.pdf>.

3.1.2.6 Communications avec actes dans un congrès national

- [220] S. BÉRARD, L. TICHIT et C. HERRMANN : Simct, a tool to visualize ontology based relationships. *In Actes des Journées Ouvertes Biologie Informatique Mathématiques (JOBIM'05)*, pages 447–457, Lyon, 2005.
- [221] M. COOPER, S. de GIVRY, M. SANCHEZ, T. SCHIEX et M. ZYTNICKI : Cohérence d'arc virtuelle pour csp pondérés. *In Actes des Quatrièmes Journées Francophones de Programmation par Contraintes*, Nantes, 2008. URL http://hal.archives-ouvertes.fr/index.php?halsid=k2g54dhpeid43mrq5rbl4fvtk2&view_this_doc=inria-00292640&version=1.
- [222] S. de GIVRY, T. SCHIEX et G. VERFAILLIE : Décomposition arborescente et cohérence locale souple dans les csp pondérés. *In Actes des Deuxièmes Journées Francophones de Programmation par Contraintes*, Nîmes, 2006. URL http://hal.archives-ouvertes.fr/index.php?halsid=k2g54dhpeid43mrq5rbl4fvtk2&view_this_doc=inria-00085804&version=1.
- [223] L. DUPUY, M. VIGNES, B. MCKENZIE et P. WHITE : Meristematic waves, a new approach to model root architecture dynamics. *In Proceedings of the 10th JOBIM*, Nantes, June 2009.
- [224] T. FARAUT, S. de GIVRY, P. CHABRIER, T. DERRIEN, F. GALIBERT, C. HITTE et T. SCHIEX : A comparative genome approach to marker ordering. *In Actes de JOBIM'06*, page 11, Bordeaux, France, 2006.
- [225] A. FAVIER, S. de GIVRY et P. JÉGOU : Comptage de solutions en exploitant la structure du graphe de contraintes. *In Y. DEVILLE, éditeur : Actes des Cinquièmes Journées Francophones de Programmation par Contraintes*, Orléans, 2009.
- [226] C. PRALET, T. SCHIEX et G. VERFAILLIE : Une nouvelle architecture de calcul pour résoudre des diagrammes d'influence. *In Actes des JFPDA'2006*, Toulouse, France, 2006.
- [227] C. PRALET, G. VERFAILLIE et T. SCHIEX : Décision avec incertitudes, faisabilités et utilités : vers un cadre algébrique unifié. *In Actes des JFPDA'2006*, Toulouse, France, 2006.
- [228] M. SANCHEZ, D. ALLOUCHE, S. de GIVRY et T. SCHIEX : Algorithme des poupées russes exploitant une décomposition arborescente. *In Y. DEVILLE, éditeur : Actes des Cinquièmes Journées Francophones de Programmation par Contraintes*, Orléans, 2009. URL http://hal.archives-ouvertes.fr/index.php?halsid=k2g54dhpeid43mrq5rbl4fvtk2&view_this_doc=hal-00387845&version=1.
- [229] M. ZYTNICKI, C. GASPIN et T. SCHIEX : DARN ! a soft constraint solver for RNA motif localization. *In Proc. of JOBIM'2007*, Marseille, France, 2007.
- [230] M. ZYTNICKI, F. HERAS, S. de GIVRY et L. LARROSA : Cohérence d'arc existentielle : un pas de plus vers la cohérence d'arc complète. *In Actes des Premières Journées Francophones de Programmation par Contraintes*, pages 139–148, Lens, France, 2005. URL http://hal.archives-ouvertes.fr/index.php?halsid=k2g54dhpeid43mrq5rbl4fvtk2&view_this_doc=inria-00000061&version=1.

3.1.2.7 Communications orales sans actes dans un congrès international ou national

- [231] D. ALLOUCHE, S. de GIVRY, M. SANCHEZ et T. SCHIEX : Tagsnp selection using weighted csp and russian doll search with tree decomposition. *In CP-09 workshop on Constraint Based Methods for Bioinformatics*, 2009.
- [232] S. BOITARD : Approximation diffusion du modèle de wright-fisher à deux loci et distributions des fréquences génotypiques. *In Actes des XXXVII èmes Journées de la Société Françaises de Statistique*, Pau, France, 2005.
- [233] S. BOITARD : Asymptotic properties of several association tests in a structured population. *In Proc. of Statistic Mathematic and Applications (communication orale)*, Marseille, France, 2006.
- [234] S. BOITARD : Cartographie fine de traits à caractère quantitatif par déséquilibre de liaison. *In Actes du colloque Jeunes Probabilistes et Statisticiens (communication orale)*, Aussois, France, 2006.

- [235] S. BOITARD, J. ABDALLAH, H. de ROCHAMBEAU, C. CIERCO-AYROLLES et B. MANGIN : Linkage disequilibrium interval mapping of quantitative trait loci. *In Proc. of 8th World Congress on Genetics Applied to Livestock Production*, Belo Horizonte, Brazil, 2006.
- [236] M. COOPER, S. de GIVRY et T. SCHIEX : The quest for the best arc consistent closure in weighted CSP. *In Proc. of the Soft'06 workshop*, Nantes, France, 2006.
- [237] S. de GIVRY, I. PALHIÈRE, Z. VITEZICA et T. SCHIEX : Mendelian error detection in complex pedigree using weighted constraint satisfaction techniques. *In ICLP-05 workshop on Constraint Based Methods for Bioinformatics*, Sitges, Spain, 2005.
- [238] S. de GIVRY, T. SCHIEX et G. VERFAILLIE : Combining tree decomposition and local consistency in max-csps. *In Proc. 7th International CP-05 Workshop on Preferences and Soft Constraints*, Sitges, Spain, 2005.
- [239] D. DELANNOY, R. CATHELIN, M. F. JOURJON, S. JASSON, J. MARCEL, B. NGOM et B. MANGIN : MCQTL : multi-allelic QTL mapping in multi-cross design. *In Proc. of PAG XIV (poster et démonstration)*, San Diego, USA, 2006.
- [240] C. GASPIN, S. de GIVRY, T. SCHIEX, P. THÉBAULT et M. ZYTNIKI : A new local consistency for weighted csp applied to ncRNA detection. *In ICLP-05 workshop on Constraint Based Methods for Bioinformatics*, Sitges, Spain, 2005.
- [241] C. GASPIN, T. SCHIEX et M. ZYTNIKI : Bound arc consistency for weighted CSPs. *In Proc. 7th International CP-05 Workshop on Preferences and Soft Constraints*, Sitges, Spain, 2005.
- [242] T. GEISSAMN, E. HUNTZINGER, S. BOISSET, P. FECHTER, M. POSSEDKO, N. BENDRID, Y. BENITO, A. JACQUIER, C. GASPIN, F. VANDENESCH et P. ROMBY : Regulatory RNA in *Staphylococcus aureus* : a common antisense regulatory mechanism for the expression of several virulence factors. *In Conférences Jacques-Monod*, Roscoff, France, 2006.
- [243] C. PRALET, G. VERFAILLIE et T. SCHIEX : Composite graphical models for reasoning about uncertainties, feasibilities, and utilities. *In Proc. 7th International CP-05 Workshop on Preferences and Soft Constraints*, Sitges, Spain, 2005.
- [244] M. SANCHEZ, D. ALOUCHE, S. de GIVRY et T. SCHIEX : Russian doll search with tree-decomposition. *In Soft'08 Workshop, attached to CP'08 conference*, Sydney, Australia, 2008.
- [245] B. SERVIN, S. de GIVRY et T. FARAUT : Modélisation de l'incertitude des cartes comparées et application à l'étude des assemblages génomiques. *In Séminaire Génétique Animale*, Lacanau, France, 2008.
- [246] P. THÉBAULT, S. de GIVRY, T. SCHIEX et C. GASPIN : Combining constraint processing and pattern matching to describe and locate structured motifs in genomic sequences. *In 5th IJCAI workshop on modelling and solving problems with constraints*, 2005.
- [247] J. VANDEL et S. de GIVRY : Inférence de réseaux de régulation de gènes à partir de données d'expression et de polymorphisme. *In JOBIM-09 workshop sur Modèles graphiques probabilistes pour l'intégration de données hétérogènes et la découverte de modèles causaux en biologie*, Nantes, France, 2009.

3.1.2.8 Communications par affiche dans un congrès international ou national

- [248] S. AUBOURG, M.-L. MARTIN-MAGNIETTE, V. BRUNAUD, L. TACONNAT, F. BITTON, V. THAREAU, T. SCHIEX, M. CABOCHE, A. LECHARNY et J.-P. RENOU : Identification of novel functional genes in the Arabidopsis genome supported by transcriptome data. *In Poster dans la conférence Plant GEMS*, Venice, Italy, 2006.
- [249] P. BARDOU, M. CROS, C. GASPIN, D. GAUTHERET, J. LARRÉ, B. GRENIER-BOLEY, J. MARIETTE, A. de MONTE et H. TOUZET : RNAspace non-coding RNA annotation web platform. *In JOBIM*, Nantes, 2009.

- [250] C. CIERCO-AYROLLES, M. CROS, C. GASPIN, A. KOZOMARA, A. MOISAN et E. SALLET : Prediction ab initio d'ARN non codant. *In GenoToul*, Toulouse, France, 2006.
- [251] C. CIERCO-AYROLLES, M. CROS, C. GASPIN, A. MOISAN et E. SALLET : ARN non codant : extension de l'environnement Apollo. *In GenoToul*, Toulouse, France, 2006.
- [252] C. CIERCO-AYROLLES, M. CROS, C. GASPIN, A. MOISAN et E. SALLET : Visualizing genome information for ncRNA search. *In RIBOREG-ESF Workshop on Mechanisms Involving Non-Protein-Coding RNAs*, Carry, France, 2007.
- [253] M. CROS : Classification d'ARN non-codant-protéine par SVM. *In Ecole de Printemps Théorique en Informatique sur Apprentissage automatique*, Porquerolles, 2008.
- [254] M. CROS : Développement d'une plateforme web avec cheerypy. *In PYCON*, Paris, 2009.
- [255] M. CROS, E. SALLET, A. MOISAN, C. CIERCO-AYROLLES et C. GASPIN : ApolloRNA : Visual support to ncRNA identification and analysis. *In The new RNA frontier Symposium*, Colmar, 2007.
- [256] M. CROS, E. SALLET, A. MOISAN, C. CIERCO-AYROLLES et C. GASPIN : Non-protein-coding RNA identification and analysis : choice and extension of apollo. *In JOBIM*, Marseille, 2007.
- [257] M. CROS, E. SALLET, A. MOISAN, C. CIERCO-AYROLLES et C. GASPIN : ApolloRNA : Visual genome environment to support ncRNA identification. *In SIFRARN*, Montpellier, 2008.
- [258] S. de GIVRY, Z. VITEZICA, I. PALHIÈRE et T. SCHIEX : Mendelsoft : Mendelian error detection in complex pedigree using weighted constraint satisfaction techniques. *In Poster at the 8th World congress on Genetics Applied to Livestock Production*, Belo Horizonte, Brazil., 2006.
- [259] S. JASSON et B. MANGIN : MCQTL. *In QTLMAS meeting*, 2007.
- [260] B. MANGIN, S. BOITARD, P. GARNIER-GÉRÉ, H. de ROCHAMBEAU et C. CIERCO-AYROLLES : Revisited statistical properties of the optimal measure of allelic association estimator. *In European Mathematical Genetics Meeting*, Cambridge, apr 2006.
- [261] B. MANGIN, H. de ROCHAMBEAU et H. GILBERT : LDLmapQTL. *In Bilan des projets GENANIMAL 2003-2006*, Dourdan, 10 2007.
- [262] B. MANGIN, L. MOREAU, M. SAWKINS et N. VUILLAUME : Use of linkage disequilibrium in mcqtl. *In The Génoplane generic and targeted ANR project for 2005, 2006, 2007*, apr 2008.
- [263] C. NOIROT, C. GASPIN, T. SCHIEX et J. GOUZY : LeARN : the Legume ncRNA database; a platform for detecting, clustering and annotating non coding RNA. *In Poster à JOBIM'2006*, Bordeaux, France, 2006.
- [264] K. PHOK, A. MOISAN, M. CROS, C. GASPIN et B. CLOUET D'ORVAL : Caractérisation de nouveaux ARN non-codant dans les génomes de Thermococcales. *In SIFRARN*, Montpellier, 2008.
- [265] K. PHOK, A. MOISAN, M. CROS, C. GASPIN et B. CLOUET D'ORVAL : Identification and characterisation of new non-coding RNA in *Pyrococcus abyssi*. *In Regulatory RNA in Prokaryotes*, Berlin, 2009.

3.1.2.9 Ouvrages scientifiques (ou chapitres de ces ouvrages)

- [266] S. GENIN, M. SALANOUBAT, J. GOUZY, A. MOISAN, T. SCHIEX, S. CUNNAC, M. LAVIE, C. ZISCHEK, P. BARBERIS et C. BOUCHER : The *Ralstonia solanacearum* complete genome sequence : outputs and prospects. *In Bacterial wilt disease and the Ralstonia solanacearum species complex*, pages 335–342. American Phytopathological Society (APS Press), St. Paul, USA, 2005.
- [267] P. MESEGUER, F. ROSSI et T. SCHIEX : *Soft Constraints*, chapitre 9, pages 281–328. Elsevier, 2006. ISBN 978-0-444-52726-4. URL http://www.elsevier.com/wps/find/bookdescription.cws_home/708863/description.

- [268] G. VERFAILLIE et T. SCHIEX : *Problèmes de satisfaction de contraintes*, volume 1 de *Traité IC2, série Informatique et Systèmes d'Information*, chapitre 6, pages 287–344. Hermès Sciences, 2006. ISBN 2-7462-1289-7. URL <http://www.lavoisier.fr/fr/livres/index.asp?texte=2746212890&select=isbn&from=Hermes>.
- [269] G. VERFAILLIE et T. SCHIEX : *Constraint Satisfaction Problems*, chapitre 6, pages 265–320. Numéro ISBN 978-1-84821-116-2. ISTE-Wiley, 2009. URL <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-1848211163.html>.

3.1.2.10 Ouvrages de vulgarisation (ou chapitres de ces ouvrages)

- [270] C. GASPIN : Bioinformatique. In J.-G. GANASCIA, éditeur : *Communication et Connaissance, Supports et médiations à l'âge de l'information*, Sciences et Techniques de l'Ingénieur, pages 63–66. CNRS, 2006.

3.1.2.11 Directions d'ouvrages ou de revues

- [271] P. MESEGUER, F. ROSSI et T. SCHIEX : Constraint-based approaches to preference modelling and reasoning. *Constraints, Special issue.*, 2010.

3.1.2.12 Autres productions : bases de données, logiciels enregistrés, traductions, comptes rendus d'ouvrages, rapports de fouilles, guides techniques, catalogues d'exposition, rapports intermédiaires de grands projets internationaux, etc.

- [272] M. CROS, E. SALLET, A. MOISAN, C. CIERCO-AYROLLES et C. GASPIN : Visualizing and exploring genomic information for non-protein-coding RNA identification using ApolloRNA. *Nature Protocols*, DOI : 10.1038/nprot.2007.285, 2007. URL http://www.natureprotocols.com/2007/07/06/visualizing_and_exploring_genos.php.
- [273] S. de GIVRY, M. COOPER, M. SANCHEZ, T. SCHIEX et M. ZYTNICKI : Virtual arc consistency for weighted csp and applications of wcsp techniques (toulbar2). In *Présentation invitée à Cork Constraint Computation Centre*, Cork, Ireland, 2008.
- [274] S. de GIVRY, T. SCHIEX et G. VERFAILLIE : Exploiting tree decomposition and soft local consistency in weighted csp. In *Présentation invitée à Cork Constraint Computation Centre*, Cork, Ireland, 2006.
- [275] S. de GIVRY, T. SCHIEX, G. VERFAILLIE et S. BOUVERET : Méthodes hybrides dans les réseaux de contraintes pondérées. In *Présentation invitée à l'Université de Caen*, Caen, France, 2006.
- [276] S. DEJEAN, N. OUMOUHOU, D. ESTIVALS et B. MANGIN : HAPim. Web, 2008. URL <http://cran.r-project.org/web/packages/HAPim/index.html>.
- [277] C. GASPIN : *Traduction du chapitre 20*, chapitre Analyse des génomes et biologie des systèmes, pages 545–567. Pearson Education France, 6 édition, 2009. ISBN 978-2-7440-7348-9.
- [278] S. JASSON, D. DELANNOY, R. CATHELIN, B. ESCALIERE, M. F. JOURJON, S. LAMBERT, J. MARCEL, B. NGOM et B. MANGIN : MCQTL version 4. Web, 2008. URL <http://carlit.toulouse.inra.fr/MCQTL/>. n° APP IDDN.FR.001.360016.003.S.P.2003.000.103000.
- [279] B. MANGIN, R. CATHELIN, D. DELANNOY, S. LAMBERT, J. MARCEL, B. NGOM, M. F. JOURJON et S. JASSON : MCQTL reference manual. Rapport UBIA Toulouse, 2007.
- [280] B. MANGIN, M. F. JOURJON, D. DELANNOY et S. JASSON : Supplement to MCQTL reference manual : outbred families. Rapport UBIA Toulouse, 2005.
- [281] T. SCHIEX : Soft Constraint Processing. First international Summer School on Constraint Programming, 2005.

[282] T. SCHIEX, T. FARAUT, S. de GIVRY, P. CHABRIER et M. BOUCHEZ : *CarthaGène rev 1.0 : User Documentation*. INRA, 2006.

3.2 Partenariats scientifiques, animation de communautés scientifiques

3.2.1 Organisation de congrès et conférences

3.2.1.1 Congrès, conférences et rencontres internationales

- Session « Natural Resources and Biodiversity Management », MODSIM07 (Christchurch, New Zealand)
- Ateliers internationaux sur les préférences et contraintes souples (Nantes, 2006; Sydney, 2008)
- Atelier international sur les méthodes de comptage dans les CSP, SAT et problèmes proches (Sydney, 2008)
- co-organisation avec P. Rouzé (VIB - Ghent, Belgique) de la session « Bioinformatique » à la conférence « Federation of European Societies of Plant Biology » 2006.
- Organisation du « Unique Informal Workshop on Cost Function Processing » (14/15 Feb 2008), Aleyria, France. 26 participants (France, Espagne, UK, Portugal). <http://carlit.toulouse.inra.fr/cgi-bin/awki.cgi/CostFunctionWorkshop>.
- SSIAB 08 : 7ième workshop Spatial Statistics and Image Analysis in Biology, Toulouse (31) du 13 au 16 mai 2008 - 40 participants - M. Goulard et N. Peyrard, membres du comité d'organisation.

3.2.1.2 Congrès, conférences et rencontres nationales

- BioARN 2005 : Mondonville (31) du 7 au 9 décembre 2005 - 50 participants - C. Gaspin INRA
- JFPDA'06 : Journées Francophones Planification, Décision, Apprentissage pour la conduite de système - Toulouse (31) du 10 au 12 mai 2006 -40 participants - F. Garcia INRA / G. Verfaillie ONERA
- GenoToul 2006, «Bioinformatique : offre, développements, attentes», 200 participants, Toulouse, 23 novembre 2006 - C. Gaspin INRA
- RECORD 2007 : Modélisation des systèmes de culture - Toulouse (31) 18 et 19 janvier 2007 - 40 participants et Toulouse 4 septembre 2007 - 40 participants - H. Raynal / F. Garcia INRA
- BioARN 2008 : Labourgade (82) du 26 au 28 février 2008 - 40 participants - C. Gaspin INRA
- BioARN 2009 : Boussens (31) du 14 au 16 avril 2009 - 40 participants - C. Gaspin INRA

3.2.2 Participation à des comités de programme de congrès et conférences

3.2.2.1 Comités de programme internationaux

- ECAI 2006, 2008 : European Conference on Artificial Intelligence.
- EWRL 2005, 2008 : European Workshop on Reinforcement Learning.
- LFA 2008 et 2009 : Congrès Francophone de Logique Floue et Applications.
- MICAI 2005 : Mexican International Conference on Artificial Intelligence.
- WCB 2006, 2007, 2008 : International workshop on Constraint Based Methods for Bioinformatics.
- CP 2006, 2007, 2008, 2009 : International Conference on Principles and Practice of Constraint Programming.
- AAAI 2006 : conférence de l'American Association for Artificial Intelligence.
- IJCAI 2007 : International Joint Conference on Artificial Intelligence.
- ICML 2009 : International Conference on Machine Learning.
- IJCAI 2009 : International Joint Conference on Artificial Intelligence (Senior PC).
- ESI 2009 : EURO Summer Institute "OR in Agriculture and Forest Management", Lleida.

3.2.2.2 Comités de programme nationaux

- Membre du comité d'organisation des 37èmes journées de Statistique organisées à Pau (64) par la Société Française de Statistique, 2005.
- JFPDA 2006, 2007 : Journées françaises de Planification, Décision, Apprentissage.
- JOBIM 2009 : Journées Ouvertes en Biologie, Informatique et Mathématiques, Nantes.
- IAF 2008 : Journées « Intelligence artificielle fondamentale », Paris, Dauphine.
- CAp 2009 : Conférence francophone sur l'Apprentissage artificiel.
- ROADEF 2010 : conférence de la société française de Recherche Opérationnelle et d'Aide à la Décision, Toulouse.

3.2.3 Participation à des comités éditoriaux de revues scientifiques

3.2.3.1 Revues internationales

- Revue «International Journal of Agricultural and Biological Engineering (IJABE)» (<http://www.ijabe.org>)
- Guest editor du Special Volume of the Annals of Operations Research (2009)
- Revue «Constraints» (<http://www.springer.com/computer/artificial/journal/10601>), fondateur du comité d'édition de la revue (de 1996 à novembre 2008).
- Revue «Journal of Artificial Intelligence Research» (<http://www.jair.org>), depuis 2009.

3.2.4 Création, animation de réseaux scientifiques et sociétés savantes

3.2.4.1 Réseau et sociétés internationaux

- Réseau européen COST TD0801 (<http://www.statseq.eu>), membres du comité exécutif (Vice-chair, co-auteur de la proposition d'action et Short term scientific mission vice-coordinator).
- Réseau européen COST FP0804 « Forest Management Decision Support Systems » (FORSYS).
- Membre du board du EURO working group ORAFM (OR in Agriculture and Forest Management).
- Correspondant français du *Biometric Bulletin*, bulletin de l'*International Biometric Society*.
- International Commission of Agricultural and Biological Engineering, Membre de la section Information Systems.
- Association for Constraint programming (<http://www.a4cp.org>), Membre du comité exécutif et trésorier.

3.2.4.2 Réseau et sociétés nationaux

- Présidence de la Société Française de Biométrie 2005-2008.
- Animateur de l'Atelier M., cellule de coordination de la modélisation de l'Inra depuis 2006 (avec Th. Simonneau (EA), Ph. Faverdin (PHASE)).
- Animateur depuis 2006 du réseau Mexico, réseau méthodologique du département MIA, avec H. Monod et H. Richard (MIA Jouy et Avignon). co-fondateur avec H. Monod et V. Ginot (MIA).
- Animateur depuis 2006 du réseau GIAE, Gestion de l'Information Agri-Environnementale avec R. Delécolle (DISI/PSI) et D. Maurice (EFPA), ayant trait notamment aux ORE INRA. Ce réseau a pour vocation de laisser la place au dispositif éco-informatique en train de mettre en place.
- Co-animateur dans le cadre des formations Réflexives® doctorants-encadrants.
- Participation aux entretiens de Cargèse fin 2006 pour définir la feuille de route du réseau national des systèmes complexes (RNSC).
- Responsable (de 2004 à 2007) de la plate-forme opérationnel INRA-ACTA-ICTA régie par une convention signée entre l'INRA, l'ACTA et 9 Instituts et Centres Techniques Agricoles.
- Animation de la Cellule Bioinformatique de l'INRA.
- Membre fondateur de l'association française pour la programmation par contrainte (<http://www.afpc-asso.org>). Membre du CA de sa fondation à 2005.
- Création et animation depuis 2006 du réseau national MSTGA (<http://carlit.toulouse.inra.fr/MSTGA/>). (financements RNCS et département MIA).

- Co-animateur du réseau INRA « méthodes d'évaluation des impacts environnementaux », 2009.

3.2.5 Participation à des comités scientifiques et jurys

3.2.5.1 Rapporteurs et membres de jury de thèses ou HDR

- Jury de thèse de C. Bru, 2005, bioinformatique, Univ. Paul Sabatier.
- Rapporteur de la thèse de B. Venable, 2005, informatique, Univ. de Padoue, Italie.
- Rapporteur de la thèse de M. Sanchez, 2006, informatique, Univ. Pol. de Catalunya, Espagne.
- Rapporteur de la thèse de S. Ndiaye, 2007, informatique, Univ. Paul Cézanne.
- Rapporteur de la thèse de C. Piette, 2007, informatique, Univ. Artois.
- Rapporteur de la thèse d'E. Rollon, 2008, informatique, , Univ. Pol. de Catalunya, Espagne.
- Rapporteur de la thèse de N. Levasseur, 2008, informatique, Univ. Caen.
- Rapporteur de la thèse de A. Fontaine, 2009, bioinformatique, Univ. Lille.
- Rapporteur du mémoire d'HdR C. Lecoutre, 2007, informatique, Univ. Artois.
- Jury de thèse de F. Picard, 2005, Statistique, Agro-Paris-Tech, Paris .
- Rapporteur de la thèse de V. Segura, 2007, Génétique, Université SupAgro, Montpellier,
- Rapporteur et président du jury de la thèse de de J. Besson, 2005, bioinformatique, INSA de Lyon.
- Jury de thèse de Romain Rivière, 2005, (LRI Orsay).
- Rapporteur de la thèse d'Aurélie Lescoute, 2006, (IBMC Strasbourg), 2006.
- Rapporteur de la thèse de Rym Kachouri (IBMC Strasbourg), 2008.
- Rapporteur de la thèse de C. Lauvernet (Univ. J. Fourier, Grenoble), 2005.
- Rapporteur du mémoire d'HdR de D. Makowski (Univ. Paris Sud Orsay), 2007.
- Rapporteur de la thèse de R. Trépos (ASC INRA-INRIA, Univ. Rennes 1), 2008.
- Rapporteur de la thèse de H. Drouineau (IFREMER, Univ. Rennes 2), 2008.
- Rapporteur de la thèse de A. Pamponet Machado (Informatique, Univ. Pierre et Marie Curie, Paris), 2009.
- Rapporteur de l'HdR de J.-C. Soulié (Informatique, Univ. du Littoral Côte d'Opale, Calais), 2009.
- Rapporteur de la thèse de D. Versmisse (Informatique, Univ. du Littoral Côte d'Opale, Calais), 2008.
- Rapporteur de la thèse de T. Degris (Informatique, Univ. Pierre et Marie Curie, Paris), 2007.
- Jury de thèse de X. Chardon (Aghronomie, Agro Paris Tech, Rennes), 2008.
- Jury d'HdR de P. Régnier (Informatique, Univ. Paul Sabatier, Toulouse), 2005.
- Jury de thèse de S. Damiani (Informatique, Supaero, Toulouse), 2005.

3.2.5.2 Membres de jury de recrutement

Jurys externes à l'INRA

- Président du jury de recrutement MCF (section 27) de l'ENSAR en 2008.
- Commission spécialiste 67 de l'université Paul Sabatier de Toulouse, membre suppléant 2003-2007.
- Commission des spécialistes 65ème de l'UPS, Toulouse (2005-2008).
- Concours MdC, 2006, (CS 67) .
- Concours MdC, 65ème, UPS Toulouse, 2008.
- Comité de sélection MdC (CNU 26), INSA, 2009.
- Comité de sélection professeur (CNU 27), Univ. Aix-Marseille, 2009.
- Recrutement CR2/CR1 INRIA, Grenoble, 2007.
- Recrutement CR2/CR1 INRIA, Lille, 2008.
- Recrutement CR2/CR1 INRIA, Lille, 2009.
- Concours CR1-INRIA Paris, 2006.
- Concours DR2-INRIA, Paris, 2006.
- Concours CR2-INRIA, Bordeaux, 2007.
- Concours CR2-INRIA, Lille, 2007.
- Concours CR1-INRIA, Paris, 2007.
- Concours CR2-INRIA, Bordeaux, 2008.
- Concours CNRS-IR, Toulouse, 2007.

- Concours CNRS-IE, Toulouse, 2007.
- Concours CNRS-IE, Toulouse, 2009.
- Concours interne CNRS-IE 808, Paris, 2008.
- Concours DR IRD, poste CGRA1, 2008.
- Concours CR-IRD CSS 1/ CGRA 1, 2009.

Jurys INRA

- Concours CR INRA, 2009 «Modélisation et BioInformatique».
- Concours DR de la direction scientifique ECONAT, INRA, 2006.
- Admissibilité CR2, Dép. MIA-EFPA-SPE, INRA, avril 2008.
- Concours CR2 INRA, 2008 « statistique et modélisation »
- Concours CR2 INRA, 2008 « génétique quantitative et génomique ».
- Concours CR2 INRA, 2009 « génétique quantitative et des populations »
- Concours IR-IE01 (MIA+SPE), 2004.
- Président du jury de concours IRE02 (EA-EFPA), 2005.
- Concours CRMIA (MIA-PHASE-EFPA-GA-CEPIA), 2006.
- Concours CR (MIA-PHASE-EFPA-GA-CEPIA), 2006.
- Président du jury de concours IRE01 (MIA+CEPIA+EFPA), 2007.
- Concours ASC Epidémio, 2007.
- Concours ASC Epidémio 2008.
- Concours IR-IE01 (MIA+CEPIA+EFPA), 2007.
- Président concours IR INRA, (EA+CEPIA), 2008.
- Président concours AI INRA, (SAD Corte), 2006.
- Concours CR2 INRA (MIA-+CEPIA+PHASE), 2006.
- Concours CR2 INRA (SAD+EA+PHASE), 2007
- Concours CR2 INRA (MIA+EFPA+SPE), 2008.

3.2.5.3 Membres de conseils

Conseil externes à l'INRA

- Conseil scientifique ANR Blanc, 2005-2006.
- Conseil scientifique Département STIC (CNRS), 2002-2005.
- Commission d'Evaluation de l'INRIA, 2005-2008.
- Bureau ReNaBi (Réseau National des plateformes Bioinformatiques), 2006-.
- Conseil scientifique de la Plateforme de Génomique Fonctionnelle de Bordeaux, 2008- .
- Délégation permanente de la commission CGRA1 Ingénierie et Expertise de l'IRD, 2008-
- Commission d'évaluation de l'UMR SENAH (PHASE+AlimH), Rennes, 2007.
- Commission d'évaluation de l'UMR-E Santé Végétale (SPE-ENITA), Bordeaux, 2008.
- Conseil scientifique de l'ED SEVAB, Toulouse, 2007-2010.

Conseil INRA

- Conseil scientifique du département INRA Génétique et Amélioration des plantes, membre nommé, 2002-.
- Conseil scientifique de centre INRA, Toulouse, membre élu 2008-.
- Conseil scientifique de département MIA, membre élu, 2006-.

3.2.6 Participation à des comités de pilotage de thèses

Des comités de pilotage de thèse sont associés à chaque thèse INRA depuis plusieurs années. Il s'agit d'un groupe de personnes réuni une fois par an typiquement durant une thèse et servant de conseil au doctorant et aux encadrants.

- A. Feugier, 2004-2006. Une méthode alternative de reproduction chez la lapine : un modèle pour une approche systémique du fonctionnement des élevages cunicoles. (INRA PHASE).
- I. Dieng, 2003-2007. Prédiction de l'interaction génotype x environnement par linéarisation et régression PLS-mixte. (CIRAD).

- A. Kozomara, 2005-2009. Prédiction de gènes d'ARN dans les séquences génomiques. (INRA MIA-SPE, 2009).
- L. Clavel, 2007-2010. Construire et évaluer des scénarios pour la gestion quantitative de l'eau. (INRA EA-SAD).
- X. Chardon, 2005-2008. Evaluation environnementale des exploitations laitières par modélisation dynamique de leur fonctionnement et des flux de matière : développement et application du simulateur MELODIE. (INRA PHASE).
- J. Vayssières, 2005-2008. Modélisation et Agro-zootéchnie des systèmes. (CIRAD).
- B. Leger, 2005-2008. Recueil et formalisation de procédés experts pour conduire une protection intégrée du vignoble. (CEMAGREF).
- G. Martel, 2005-2008. Pratiques d'élevage, productivité des troupeaux de truie et rythmes de travail des éleveurs en production porcine : une approche par modélisation. (INRA SAD-PHASE).
- A. Lurette, 2004-2007. Modélisation pour l'évaluation de mesures de maîtrise du portage de salmonelles chez le porc charcutier. (INRA MIA, ENV Nantes).
- C. Rigolot, 2006-2009. Modélisation de l'impact environnemental des pratiques en élevage porcin à l'échelle sectorielle (animal et effluent), de l'exploitation, et du cycle de vie (ACV). Incidence du niveau d'approche sur la perception des pratiques. (INRA PHASE).
- A. Ripoché, 2007-2010. Conception de systèmes viticoles enherbés : construction d'un outil d'aide à la réflexion du pilotage de l'enherbement et évaluation multicritère. (INRA EA).
- T. Derrien, 2003-2006. Développement d'outils et de méthodologies bioinformatiques pour l'analyse du contenu en gènes et de la diversité génétique chez le chien. (Université de Rennes 1).
- V. Lacroix, 2003-2006. Recherche de motifs dans les réseaux métaboliques. (Univ. Lyon 1).
- S. Bouveret, 2003-2006. Partage équitable de ressources sous contraintes. (ONERA).
- A. Niveau, 2006-2009. Compilation de connaissance pour la décision en ligne. (ONERA).
- F. Ytournel, 2005-2008. Déséquilibre de liaison et cartographie de QTL en population sélectionnée. (INRA, GA)
- B. Cochard, 2005-2008. Etude de la diversité génétique et du déséquilibre de liaison au sein de populations améliorées de palmier à huile. (CIRAD)
- E. Mandrou, 2007-2010. Variabilité fonctionnelle de gènes candidats impliqués dans la lignification chez l'eucalyptus. (INRA, EFPA)
- C. Lepoittevin, 2006-2009. Etude de la diversité nucléotidique de gènes candidats impliqués dans la formation du bois chez le pin maritime et association avec les composantes de la qualité du bois. (INRA, EFPA)
- B. Guitton, 2005-2008. Development of a genetic map and QTL detection of architectural traits in apple. (INRA, GAP)
- S. Allais, 2008-2011. Identification et validation de marqueurs génétiques impliqués dans les qualités de la viande. (INRA, GA)
- B. Bonté, 2008-. Evaluation et optimisation de l'observation d'un système dynamique de grande taille : Application à l'épidémiologie. (CIRAD).
- N. Schaller, 2008-. Modélisation conceptuelle de l'organisation spatio-temporelle de l'occupation du sol dans des exploitations de céréaliculture et de polyculture élevage. (INRA-SAD)
- Z. Assaghir, 2007-. Estimation et contrôle d'incertitudes des indicateurs agri-environnementaux. (INPL, Nancy)
- Y. Monnet, 2008-. Modélisation des interactions dynamiques entre le développement de la plante hôte, l'architecture du couvert végétal et le développement d'épidémies de maladies fongiques aériennes (INRA-SPE).
- R. Trépos, 2004-2008. Apprentissage symbolique à partir de données de simulation pour l'aide à la décision. Gestion d'un bassin versant pour une meilleure qualité de l'eau. (INRA-EA, IRISA).
- L. Sorel, 2005-2008. Paysages virtuels et analyse de scénarios pour évaluer les impacts environnementaux des systèmes de production agricole (INRA-EA).
- G. Beaumet, 2005-2008. Planification continue pour la conduite d'un satellite agile autonome. (ONERA).
- L. Maton, 2003-2006. Représentation et simulation des pratiques culturales à l'échelle régionale pour estimer la demande en eau d'irrigation. Application à un bassin versant maïsicole du sud-ouest de la France. (INRA-EA).

- S. Damiani, 2003-2005. Gestion d’une constellation de satellites de surveillance de la Terre : autonomie et coordination. (ONERA).
- G. Infantes, 2003-2006. Apprentissage de modèles de comportements pour le contrôle d’exécution et la planification robotique. (LAAS).
- F. Tortrat, 2003-2005. Modélisation orientée décision des processus de transfert par ruissellement et subsurface des herbicides dans les bassins versants agricoles. (INRA-EA)

3.2.7 Directions d’ouvrages

- Co-éditeur avec P. Meseguer (Espagne) et F. Rossi (Italie) d’un numéro spécial de la revue « Constraints » (T. Schiex), 2009.
- Guest editor du Special Volume of the Annals of Operations Research, « Agriculture and Forest Management » (2009)
- Co-éditeur avec F. Charpillat, P. Perny et O. Sigaud d’un numéro spécial « Décision et Planification dans l’incertain » de la Revue d’Intelligence Artificielle, 20(2-3), 2006.

3.2.8 Conception, coordination de programmes internes à un organisme ou nationaux

Se reporter à la feuille de calcul « Partenariats » du formulaire « Bilan ».

3.2.9 Rapports de conjoncture scientifique

- Goffinet B, Faivre R, Martin-Clouaire R et al. (2006) Rapport prospectif sur la modélisation à l’INRA. Mission confiée au département MIA, coordonnée par B. Goffinet. INRA.
- N. Aussenac, . . ., C. Gaspin et al. (2006), Communication et connaissance : Supports et médiations à l’âge de l’information, Sciences et techniques de l’ingénieur, dirigé par JG Ganascia.

3.2.10 Rapports d’évaluation, expertise scientifique

- Évaluation d’une unité bioinformatique pour le ministère de l’enseignement et de la recherche (2005).
- Membre du comité d’évaluation de la station de Génétique et d’Amélioration des Plantes de l’INRA de Versailles (2006).
- Expert auprès de l’AERES (évaluation d’un gros laboratoire en informatique) (2008).
- Expertise de projet pour le NWO Physical sciences free competition (Pays-Bas).
- Région des Pays de la Loire (2006) : expertise scientifique sur un projet de bioinformatique
- Evaluation à mi-parcours du quadriennal de l’IGM
- RTRA Digiteo : 1 projet
- Expert auprès de l’ANR sur le programme BLANC (2008)
- expert-évaluateur sur le programme ANR GENOMIQUE (2009)
- expert-évaluateur pour le programme de l’Agropolis-Fondation (2008).
- expert-évaluateur pour le programme EuroMéditerranée (2008)
- Expertise de projet auprès de « Natural Sciences and Engineering Research Council of Canada (NSERC) ». (2007)
- Commission d’évaluation et sélection des projets INRA-INRIA Euroméditerranée, 2008.
- Expert en modélisation des systèmes, Réflexion prospective INRA-CIRAD « Alimentation durable », 2009.

3.2.11 Relectures d’articles pour des revues et conférences internationales

Journaux

Agronomie : Computers and Electronics in Agriculture, Agricultural Systems,

Biologie : BMC Genomics, Theoretical and Applied Genetics, Heredity, BMC Genetics, Genetics Selection Evolution, Nucleic Acids Research (NAR)

Bioinformatique : Bioinformatics, BMC Bioinformatics

Informatique : Artificial Intelligence, IEEE Transactions on Pattern Analysis and Machine Intelligence, Theoretical Computer Science, Journal of Artificial Intelligence Research, Constraints, Journal of Heuristics, IEEE Transactions on Signal Processing, ACM Transactions on Computational Logic, Journal of Satisfiability, Annals of mathematics and Artificial Intelligence, Signal Processing

RO : Operations Research, European Journal of Operational Research, Annals of Operation Research, Operational Research : An International Journal

Simulation : Simulation Modelling Practice and Theory

Statistiques : Statistics and Computing, Biometrics

Conférences

Bioinformatique : RECOMB'2007

Informatique : ICAP'2007, SAT'2005, SAT'2009, CP'2005, CP'2006, CP'2007, CP'2008, CP'2009, ECAI'2006, AAAI'2006, AAAI'2007, IJCAI'2005, IJCAI'2007, IJCAI'2009, ICML'09, EWRL'05, EWRL'08

Automatique IFAC'2008

3.2.12 Prix scientifiques obtenus

La thèse de C. Pralet²⁸⁸ a eu le prix national de thèse ASTI 2007 <http://www.agence-nationale-recherche.fr/actualite/13?lnglnfold=157>.

Les articles de Eve Mac Donald-Madden (co-écrit avec I. Chadès) et de Nicklas Forsell (co-écrit avec N. Peyrard et R. Sabbadin) ont reçu respectivement les premier et deuxième prix de la conférence MODSIM 2007, catégorie "doctorants".

L'article co-écrit avec E. Rachelson et P. Fabiani⁸⁷ a reçu le prix du meilleur papier décerné par l'AFIA lors de la conférence nationale d'apprentissage automatique CAp'08.

3.3 Développement de grands instruments pour la recherche

Bases de données, logiciels scientifiques, modélisations

La tableau suivant récapitule l'ensemble des logiciels scientifiques développés dans l'unité ayant évolué durant la période d'évaluation et qui sont distribués ou mis à disposition sur le Web. La majorité de ces logiciels sont aussi accessibles via notre forge logicielle (<https://mulcyber.toulouse.inra.fr/>).

Logiciel	Finalité	Site Web
HAPim	Détection de QTL	http://cran.r-project.org/web/packages/HAPim
MCQTL	Détection de QTL	http://carlit.toulouse.inra.fr/MCQTL/
toulbar2	Contraintes pondérées	http://carlit.toulouse.inra.fr/cgi-bin/awki.cgi/ToolBarIntro
MendelSoft	Analyse de pedigree	http://www.inra.fr/mia/T/MendelSoft/
CarthaGène	Cartographie génétique & RH	http://www.inra.fr/mia/T/CarthaGene
DARN!	Localisation d'ARNs	http://carlit.toulouse.inra.fr/Darn/
EuGène	Prédiction de gènes	http://www.inra.fr/mia/T/EuGene
FrameDP	Analyse cluster EST	http://iant.toulouse.inra.fr/FrameDP/
LeARN	Annotation d'ARN	http://bioinfo.genopole-toulouse.prd.fr/LeARN
ApolloRNA	Aide identification ARNnc	http://carlit.toulouse.inra.fr/ApolloRNA
RNASpace	Identification, annotation ARNnc	http://www.rnaspaces.org
MDPToolBox	Résolution de PDM	http://www.inra.fr/mia/T/MDPtoolbox
DIESE	Environnement de modélisation/simulation	http://carlit.toulouse.inra.fr/diese/frontpage.html
VLE	Environnement de modélisation/simulation	http://vle.toulouse.inra.fr

Conception et mise à disposition de Plates-formes technologiques

Voir la section 2.3.

3.4 Connaissances opérationnelles et produits pour des acteurs socio-économiques

3.4.1 Bourses Cifre

S. Mezouk, Analyse des bases génétiques et moléculaires de la tolérance à la sécheresse chez le maïs par génétique d'association, Bourse Cifre avec Biogema, 2009.

3.5 Partenariats avec des acteurs socio-économiques

3.5.1 Logiciels déposés durant la période d'évaluation

- Logiciel MCQTL, 2008, n° APP IDDN.FR.360016.003.S.P.2003.000.10300
- Logiciel HAPim, en cours de dépôt à l'APP
- Logiciel VLE, en cours de dépôt à l'APP

3.5.2 Contrats de licence

MCQTL (2006) Licence déposée par Génoplante Valor

3.6 Enseignement et Formation

3.6.1 HDR soutenue

Sabbadin R. (2009) Modèles et Algorithmes pour la Décision dans l'Incertain. HDR de l'Université Paul Sabatier.

3.6.2 Thèses soutenues

Neuf thèses ont été soutenues dans la période couverte par l'évaluation.

- [283] S. BOITARD : *Cartographie de gènes à caractères quantitatifs par déséquilibre de liaison*. Thèse de doctorat, Université Paul Sabatier, Toulouse III, 2005.
- [284] C. CRAMBES : *Modèles de régression linéaire pour variables explicatives fonctionnelles*. Thèse de doctorat, Université Paul Sabatier, Toulouse, 2006.
- [285] O. CRESPO : *Conception par simulation pour la conduite de culture*. Thèse de doctorat, Université Paul Sabatier, Toulouse, Février 2008.
- [286] N. FORSELL : *Planning under Risk and Uncertainty : Optimizing Spatial Forest Management Strategies*. Thèse de doctorat, Faculty of Forest Science - SLU, May 2009.
- [287] A. KOZOMARA : *Prédiction de gènes d'ARN dans les séquences génomiques*. Thèse de doctorat, Université Paul Sabatier, Toulouse, Toulouse, France, Juin 2009.
- [288] C. PRALET : *Un cadre générique pour représenter et résoudre des problèmes de décision séquentiels avec incertitudes, faisabilités et utilités*. Thèse de doctorat, Ecole Nationale Supérieure de l'Aéronautique et de l'Espace, Toulouse, Toulouse, France, 2006.

- [289] E. RACHELSON : *Optimisation en ligne pour la décision distribuée dans l'incertain*. Thèse de doctorat, ISAE-SUPAERO, Toulouse, Toulouse, France, Mars 2009.
- [290] N. RAVOANJANAHARY : *Planification d'actions purement épistémiques*. Thèse de doctorat, Université Paul Sabatier - Toulouse, Décembre 2008. ED Informatique.
- [291] M. ZYTNICKI : *Localisation d'ARN non-codants par réseaux de contraintes pondérées*. Thèse de doctorat, Université Paul Sabatier, Toulouse, 2008.

3.6.3 Conception et l'organisation de cursus d'enseignement supérieur

- Responsables de la conception et organisation du cours de Master « Modèles probabilistes et Algorithmes pour la Biologie » (MAB) du Master informatique de l'université Paul Sabatier (20 heures). T. Schiex, S. de Givry en collaboration avec T. Faraut (LGC, INRA, Toulouse).
- Responsables de la conception et responsables tournant de l'organisation du Master « Méthodes de preuves, Optimisation et Satisfaction de Contraintes » (MPOSC) du Master informatique de l'université Paul Sabatier (20 heures). T. Schiex, S. de Givry en collaboration avec H. Fargier (IRIT, CNRS) et M. Cooper (Univ. Paul Sabatier).
- Responsable de la conception du cours « Décision séquentielle dans l'incertain » de l'ISAE-SUPAERO, avec P. Fabiani, ONERA (2x15h en 2ème et 3ème année). F. Garcia.

3.6.4 Formations assurées dans l'enseignement supérieur

La section 1.4 donne une analyse globale. On donne ici une liste exhaustive des enseignements assurés.

Public	Nature	Volume	Remarques
ISAE-SUPAERO 2ème et 3ème années	Cours+TD Processus Décisionnels de Markov	15h par an depuis 2002	
ESC Rennes 1ère et 2ème années	Cours+TD Recherche Opérationnelle	50h par an 2003-2007	en Anglais
Master BSM M2 (UPS)	Cours+TD Processus Décisionnels de Markov	24h en 2006-07 30h par an depuis 2008	
Master Agrofood Chain M1 (UPS)	Cours+TD RO, Simulation, Optimisation	12h par an depuis 2006	en Anglais
NOVA PhD course SLU, Umea, Suède	Cours+TD+TP Markov Decision Processes	60h en 2008	Ecole chercheur en Anglais
Université ULCO, L3, M1 et M2 info.	Projets d'études, développement informatique	20h en 2009 36h en 2008	
Master MAB	Bioinformatique	20h	démarrage en 2008
Master MPOSC	Réseaux de contraintes	20h	
Master informatique	Réseaux de contraintes pondérés	30h	Univ. Padoue, Italie
Master informatique	Complexité, Contraintes	20h	Univ. Fianarantsoa, Madagascar
Master 2 Pro Biotech. Utilisation des logiciels	Cours+TD par an Carthagene et MCQTL	16h par an	

3.6.5 Autres formations à destination de scientifiques et de professionnels

Pour favoriser la diffusion et l'appropriation des logiciels qu'elle développe l'unité organise des formations pour leurs usagers réels ou potentiels.

Public	Nature	Volume	Remarques
Chercheurs, modélisateurs	Tutoriel sur DIESE avec TD	40h en 2007	manuel disponible
Chercheurs, ingénieurs, doctorants, étudiants	Environnement UNIX	1 jour	5 formations en 2007, 2008 et 2009
Chercheurs, ingénieurs, doctorants, étudiants	Utilisation des ressources de la plateforme	1 jour	5 formations en 2007, 2008 et 2009
Chercheurs, ingénieurs, doctorants, étudiants	Formation à l'envir. d'annotation Apollo	demi-journée avec SaAB	En collaboration
Chercheurs, ingénieurs, doctorants, étudiants	Formation SGE	demi-journée seulement	Organisateur
Usagers de Typo3 sur les sites de la PF	Formation à l'envir. Typo3	demi-journée	
Ecole-chercheurs, génomique des caractères complexes	atelier logiciel	2h en 2007	site web
Ecole-chercheurs, génomique des caractères complexes	Carthagène	1h en 2007	
Ecole-chercheurs, génomique des caractères complexes	MCQTL	1h en 2007	
Chercheurs, INRA	MCQTL	2h en 2008	
Chercheurs (Pasteur)	Modèles probabilistes	3h en 2005,2006	
Chercheurs, ingénieurs, doctorants	Formations RECORD, VLE et DEVS	2×40h 2007 et 2008	Supports disponibles
Ingénieurs, chercheurs bioinfo.	EuGène	8 × 7h	Supports, TP
Assoc. Interprofession.	MCQTL et Carthagène	1h en 2009	
Chercheurs, doctorants	Réseaux de contraintes	4h	2005, Italie
Chercheurs	Prédiction de gènes	1h en 2005	Pour le LGC (INRA)
Développement des Semences du Sud-Ouest, Sélectionneurs de Biogemma, IFR40	MCQTL	4h en 2009	
SYNGENTA	Bioinformatique	16h en 2008	organisée par plate-forme bioinfo.
Syngenta	CarthaGène	1h	2006
Chercheurs, ingénieurs, doctorants, étudiants INRA-Cemagref-Ifremer Cirad-Ecoles	Analyse de sensibilité de modèles, Méthodes et TP	4 jours en 2009	Ecole-chercheurs du réseau MEXICO. 59 participants

Annexes

Annexe A

Pyramide des âges

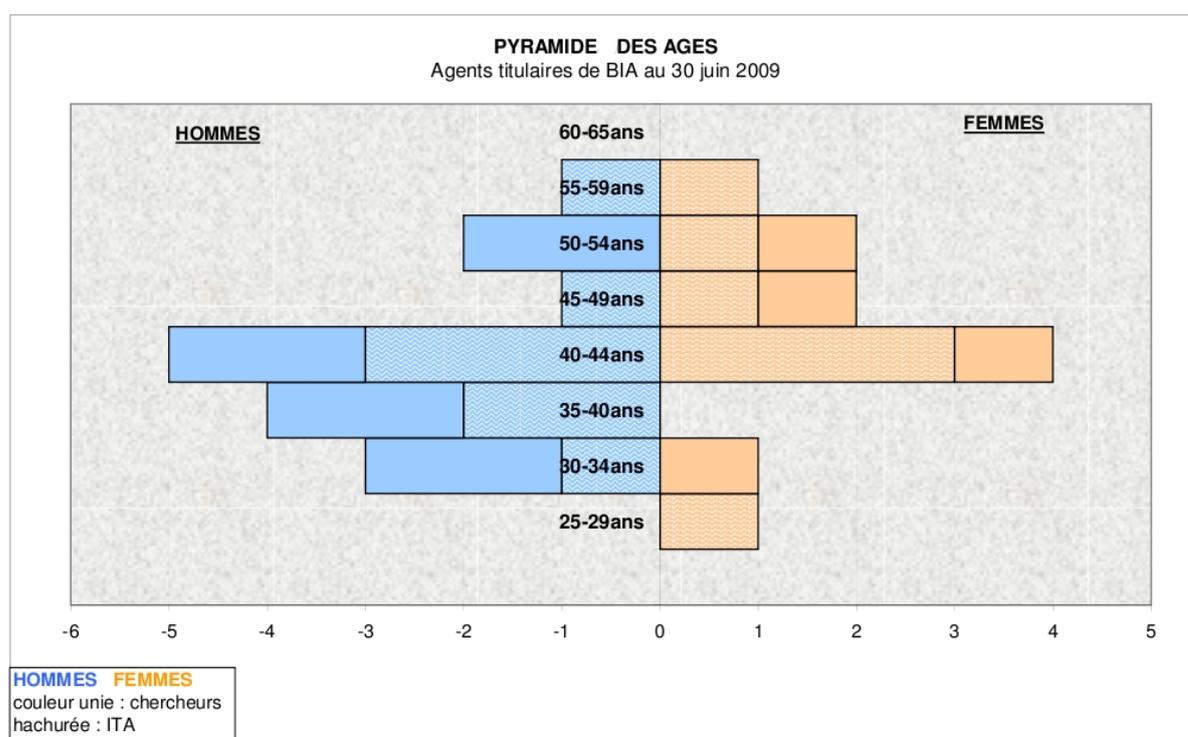


FIG. A.1 – Pyramide des âges

Cette pyramide montre une parité sensiblement équilibrée hormis pour les tranches d'âges 35-40 ans et 25-29 ans. Toutefois, les derniers recrutements qui ne sont pas pris en compte (car effectifs seulement à partir de l'automne 2009), modifieront cet affichage avec 1 ITA « femme » en 35-40ans et 2 ITA « homme » en 25-29 ans. Concernant la tranche d'âge 55-59 ans : 1 agent a fait valoir ses droits en retraite et le renouvellement de son poste a déjà été arbitrée. Il faut rester attentif à la tranche 50-54 ans dont les agents pourraient faire valoir leur droit au départ dès la fin du prochain quadriennal. Une réflexion sur les conséquences qu'aurait la perte potentielle de compétences et de moyens humains reste à mener.

Annexe B

Complément sur les ressources financières

La figure B.1 montre la proportion de la subvention d'état attribuée par chercheur dans le budget de l'unité. La non-proportionnalité du montant de la subvention d'état au nombre de parts chercheur provient de la décision du département de diminuer progressivement le montant de la part chercheur.

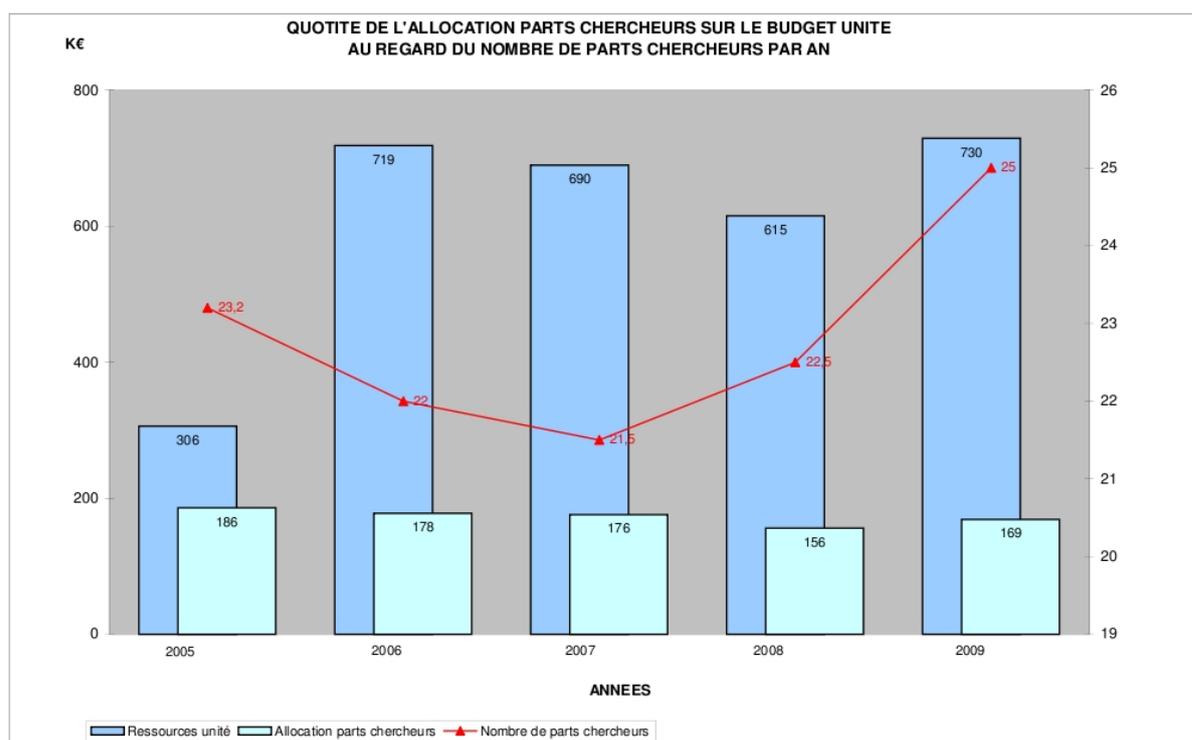


FIG. B.1 – Parts chercheur sur la subvention d'état

Le tableau suivant indique les dépenses, en K€, concernant les plus gros pôles financiers de l'Unité, toutes ressources confondues.

Années	Informatique	Missions à l'étranger	Missions en France	Fluides	Abonnements et doc.	Gratifications de stages
2005	275	59	74	16	21	20
2006	106	26	45	21	29	25
2007	54	21	59	32.5	11	12
2008	106	41	34	13	18	23
2009	469	23	48	17	2	27

La figure B.2 donne une vue globale des masses financières gérées (ressources de l'unité et ressources du GIS GENOTOUL).

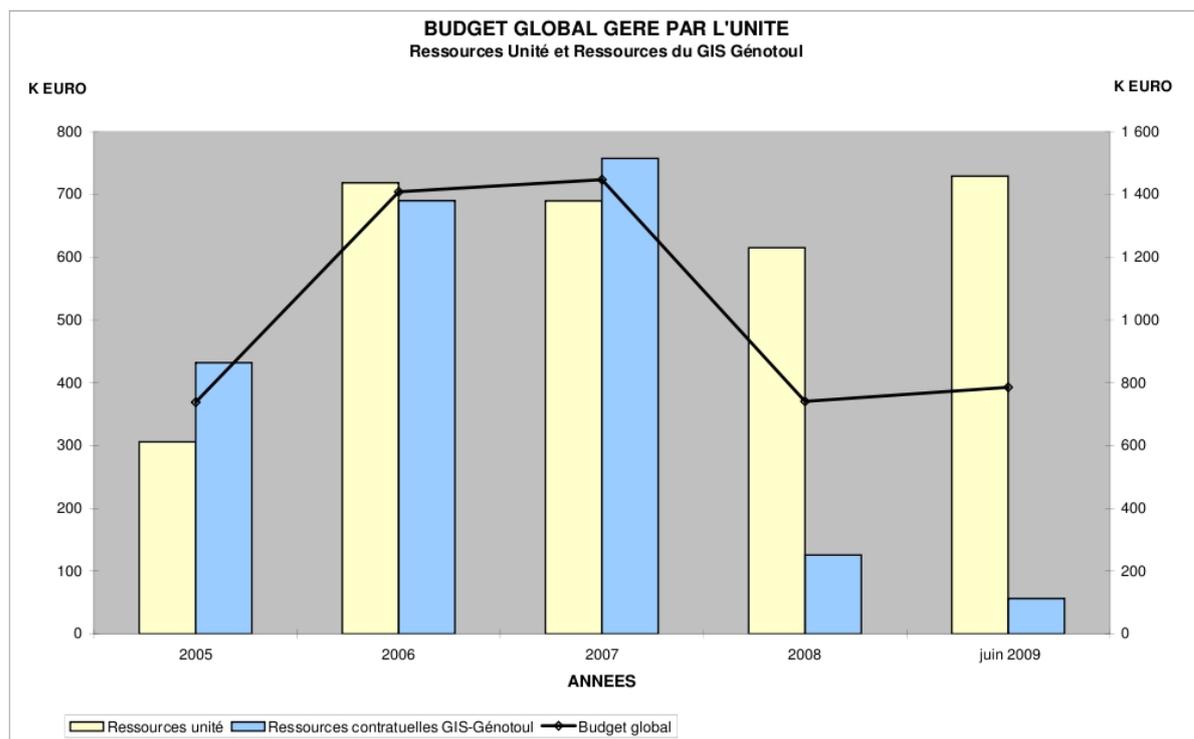


FIG. B.2 – Budget global géré par l'unité

Annexe C

Complément sur les enseignements dispensés

La figure suivante donne le volume des enseignements (en h) dispensés par les chercheurs et ingénieurs de l'unité, ainsi que le nombre d'agents impliqués.

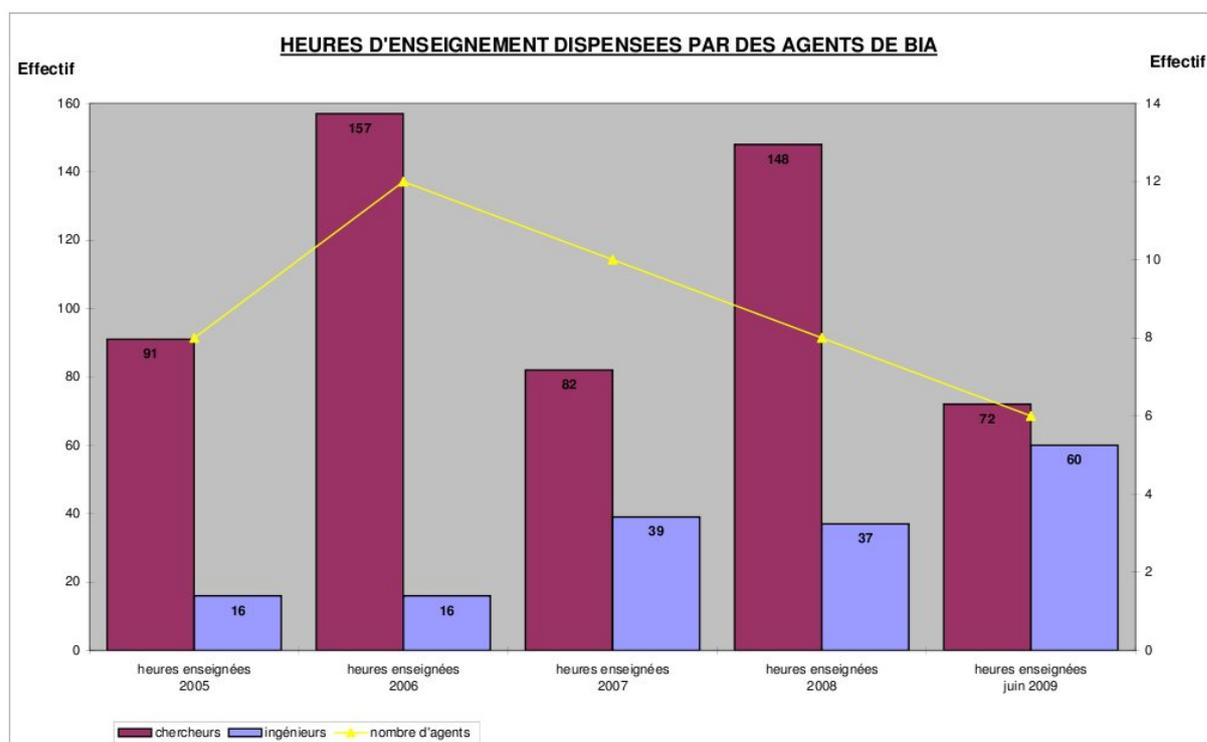


FIG. C.1 – Volume d'enseignement

Annexe D

Implication des agents par activité

Les deux tableaux suivants concernent les agents présents dans l'unité fin juin 2009.

Activités	Agents
Modélisation et simulation de systèmes pilotés avec DIESE	R. Martin-Clouaire, J.-P. Rellier
RECORD - VLE	P. Chabrier, F. Garcia, G. Quesnel, H. Raynal, N. Rouse
Méthodes d'exploration de systèmes complexes	F. Garcia, R. Faivre
Analyse et décision dans les processus sur graphe	N. Peyrard, R. Sabbadin
Approche PDM en écologie, épidémiologie, foresterie	F. Garcia, R. Sabbadin

TAB. D.1 – Répartition des agents par activités dans MAD.

Activités	Agents
Localisation de QTL	C. Cierco, S. Jasson, B. Mangin
Cartographie Comparative et assemblage	S. de Givry, T. Schiex
Réseaux de contraintes pondérés	C. Gaspin, S. de Givry, T. Schiex, D. Allouche
Localisation de gènes d'ARN	C. Cierco, MJ. Cros, C. Gaspin, A. Moisan, T. Schiex
Prédiction de gènes de protéines	T. Schiex

TAB. D.2 – Répartition des agents par activités dans SaAB. M. Vignes, arrivé dans l'équipe en juillet 2009, est positionné dans l'activité « Détection d'interactions entre gènes », détaillée dans la partie projet.

Annexe E

Rapport du comité pour l'évaluation de la période 2003-2006

Rapport du comité d'évaluation sur l'UBIA "Biométrie et Intelligence artificielle"

Dir. Roger Martin-Clouaire

Le comité s'est réuni le 23 mars 2007 à Toulouse pour faire son évaluation.

Le comité a attentivement examiné le bilan scientifique du laboratoire, pour la période 2003-2006, ainsi que les perspectives pour 2007-2010, en incluant tous les aspects qui entourent le bilan passé et le projet futur. Le comité a noté la bonne qualité des documents fournis, le rapport écrit, clair et concis, et apprécié les prestations orales pourtant assez denses.

L'unité examinée est jeune, et montre un dynamisme indéniable. Ce dynamisme est le résultat d'un niveau scientifique élevé, et se traduit par des thématiques de recherche innovantes, un enthousiasme évident et des publications au plus haut niveau en terme de qualité et quantité.

On peut souligner l'équilibre intéressant qui existe entre la modélisation et l'analyse des données issues de programmes de collaboration avec d'autres laboratoires. Il y aussi un investissement conséquent dans l'enseignement, comme le montre la participation à des enseignements en Masters, et également une contribution à la vie des Ecoles Doctorales du site de Toulouse.

Le comité souhaite souligner les points suivants :

1) Qualité de publication : Le niveau global des publications est remarquable, avec des publications dans les meilleures revues internationales. Cela est associé à l'excellente réputation et au rayonnement de ses membres sur le plan national et international, y compris ceux qui ont rejoint récemment le laboratoire. Cela devrait amener dans le futur à une réputation accrue attachée au laboratoire.

Il est quantitativement très honorable, avec 73 articles dans des périodiques, 15 chapitres d'ouvrage, 44 communications majeures + 90 autres communications, 15 logiciels. En rapport au nombre de chercheurs (18, y compris des ingénieurs), l'unité produit une moyenne de 1 article par an et par chercheur.

Sur le plan qualitatif, les publications sont très bonnes dans chaque discipline. Le ratio entre publications méthodologiques versus publications à vocation finalisée est de 2/3. L'aspect méthodologique est très marqué et de bon niveau pour les 2 équipes de l'Unité. L'aspect finalisé de l'équipe SaAB est de très bon niveau et sa thématique très porteuse. On note également de nombreux projets et contrats obtenus par les membres de l'équipe. Les publications de l'équipe MAD représentent pour une bonne moitié des articles dans des revues. Le point fort est l'équilibre entre les revues d'application et de méthodologie et un point faible est la rareté des articles écrits par des collectifs internes (à l'équipe et à l'UMR). Il serait sans doute souhaitable d'élargir les thématiques de l'équipe aux revues d'écologie dans les publications d'application. L'ouverture récente vers cette discipline et l'installation de la plateforme RECORD, devrait permettre à l'équipe MAD de travailler ces deux derniers points.

L'équilibre entre les aspects méthodologiques, finalisés, logiciels et publications scientifiques représente un atout important de l'UBIA qui faut veiller à maintenir.

2) Impacts socio-économiques : Les travaux de l'unité représentent un grand enjeu socio-économique. Le comité recommande de mettre encore plus en avant les aspects finalisés, d'avoir une stratégie d'application et de favoriser les terrains qui permettraient de croiser certains enjeux pour éviter la dispersion : par exemple les enjeux sur l'épidémiologie et les changements globaux.

3) Les enjeux scientifiques et valorisation :

Les thématiques scientifiques des équipes SaAB et MAD sont au cœur d'enjeux finalisés globaux liés à l'agriculture, l'alimentation, l'environnement et la gestion des territoires. Ils répondent de manière claire aux missions de production des connaissances scientifiques, de conception des innovations dans les domaines de la bio-informatique, de la génétique statistique, de la modélisation et de la prise de décision en agronomie.

Les perspectives scientifiques et les stratégies mises en jeu en matière d'analyse du génome (équipe SaAB) sont particulièrement pertinentes et leur visibilité se manifeste par une excellente reconnaissance à l'international. La notoriété grandissante de l'équipe se structure autour d'outils informatiques produits récemment (Carthagene, Mendelsoft, MCQTL, EUGENE, etc). Un point pouvant être amélioré concerne la politique de développement de ces nombreux logiciels (environ 2 par chercheurs). Il serait opportun de chercher à renforcer des projets collectifs d'envergure plutôt que de disperser l'activité de l'équipe en travaux individuels, liés aux interactions de chacun, mais donnant l'impression d'une unité de service scientifique auprès d'autres laboratoires. La gestion et la maintenance de nombreux logiciels peut s'avérer coûteuse à moyen terme pour l'équipe, dans une ambiance de profusion internationale où seuls les outils les plus efficaces et adaptés aux besoins des utilisateurs peuvent espérer un succès durable. Cette remarque se généralise au mode de fonctionnement de l'unité BIA qui induit des comportements de travail de nature plutôt individualiste se traduisant par un taux relativement faible de publications co-signées par plusieurs membres de l'unité (autour de 35%). L'unité devrait viser à augmenter ce taux en vue des évaluations futures.

L'équipe MAD, née de la fusion récente de deux anciennes équipes, expérimente une phase de restructuration, et se concentre sur l'étude des agro-systèmes. Il est clair que le domaine de la modélisation pour l'analyse des agro-systèmes n'offre pas la visibilité de domaines tels que la bio-informatique ou l'analyse génétique. Elle n'en demeure pas moins un objet de finalisation important pour la prise de décision en agronomie. Les mouvements récents dans cette équipe, concernant en particulier l'intégration de N. Peyrard, montrent la volonté de développer des projets fédérateurs dans les domaines de l'environnement et de l'écologie, mais aussi dans les axes méthodologiques (modèles markoviens, statistique spatiale). Il paraît bénéfique que l'équipe centre son activité sur la valorisation de quelques projets (par exemple, les projets logiciels RECORD, DIESE) et recherche une interaction systématique entre les chercheurs au sein de l'équipe.

Concernant la finalité de l'unité, il serait souhaitable de veiller à l'équilibre du rapport entre les travaux de nature théorique ou méthodologique et les travaux à vocation finalisée afin d'atteindre un objectif autour de 50%. Le comité recommande de mettre encore plus en avant les aspects finalisés, d'avoir une stratégie d'application et de favoriser les terrains qui

permettraient de croiser certains enjeux à tendance divergente. Les interactions proposées pour l'avenir entre 2 équipes actuelles montrent un souci de cohérence au niveau de la passerelle scientifique. Le comité recommande d'utiliser les moyens de recrutement pour favoriser la cohésion interne de projets fédérateurs au sein de chaque équipe et la cohésion globale de l'unité. Les orientations futures, notamment les potentiels de recherche, doivent favoriser le recentrage thématique et non pas de nouvelles directions. Il recommande par ailleurs que l'embauche de personnels sur Contrats à Durée Déterminée se fasse sur des périodes plus longues que la pratique actuelle dans le but d'augmenter la stabilité de projets durables. Il paraît indispensable que l'unité augmente son recrutement doctoral en cherchant par tous les moyens des financements de thèse (contrats CIFFRE, bourses BDI, bourse régionales), en contribuant ainsi de manière significative à la formation par la recherche, et en valorisant son fort potentiel d'encadrement.

4) Valorisation des moyens :

Le comité apprécie une restructuration réussie qui a pris en compte les conclusions de l'évaluation de 2003. L'organisation globale de l'unité semble performante.

L'équipe collective apporte manifestement satisfaction aux équipes de recherche. Elle contribue à la productivité de l'Unité en déchargeant les chercheurs de nombreuses tâches administratives et il est donc important que les moyens de l'équipe collective soient maintenus (les départs programmés devront être compensés). Par ailleurs, afin de simplifier la gestion de l'Unité, il semble vivement souhaitable que la gestion de la Génopole soit séparée de celle de l'Unité.

Les interactions entre 2 équipes actuelles, proposés aussi pour l'avenir, montrent un souci de cohérence indéniable au niveau de la passerelle scientifique. Cependant, le comité remarque un manque d'interaction entre les chercheurs au sein de l'UMR. Le comité recommande d'utiliser les nouveaux recrutements pour favoriser ce lien. Les orientations futures, notamment les potentiels de recherche, doivent favoriser le recentrage thématique et non pas une nouvelle direction.

Le comité recommande d'entretenir des moyens d'avoir des CDD de plus long terme pour favoriser l'efficacité de travail.

Dans le présent contrat, l'UBIA dépendait de plusieurs écoles doctorales : Mathématiques, EDIT, SEVAB, ... Suite à une restructuration des ED du site toulousain, l'unité ne dépendra plus que de 2 ED : SEVAB pour l'aspect finalisé et MITT pour l'aspect théorique.

En matière des thésards, le comité recommande de chercher tous les moyens de financement de thèse, e.g. CIFFRE, Région, allocations ministérielles, qui ne semblent pas avoir été suffisamment explorés.

5) Plateforme :

Le comité considère que la création d'une équipe à part pour individualiser la plateforme bioinformatique de la Génopole est une bonne mesure qui permet une meilleure lisibilité à la fois pour l'équipe SaAB et pour la plateforme de la Génopole. Le comité est un peu étonné par les relations assez compliquées entre SIGENAE et la plateforme bioinformatique de la Génopole : la plateforme Génopole héberge le projet SIGENAE et le responsable du projet SIGENAE est désormais responsable technique de la plateforme. Le comité note cependant

que cette organisation ne semble pas poser de problème aux personnes interrogées (responsable scientifique de la plateforme et responsable du projet SIGENAE). Le comité se demande toutefois si il ne serait pas bénéfique pour tous de considérer simplement que le projet SIGENAE est porté par la plateforme bioinformatique de la Génopôle.

Le comité salue la volonté de l'équipe SaAB de ne pas s'approprier la ressource que constitue la plateforme de la Génopôle. Il note cependant que la valorisation des travaux de recherche fait partie des missions d'une plateforme de bioinformatique et il lui semble donc que la plateforme pourrait apporter un soutien à l'équipe SaAB dans le domaine de la valorisation (exploitation, maintenance et support) des logiciels développés par l'équipe SaAB.

Le comité souligne l'importance de mettre en avant l'impact scientifique de la plateforme. Cela nécessite la mise en place d'une démarche active d'identification des travaux scientifiques ayant fait appel aux prestations de la plateforme.

Enfin, le comité salue la labellisation RIO obtenue par la plateforme et se félicite de l'implication de celle-ci dans les activités du réseau ReNaBi.

6) Le comité a particulièrement apprécié le mode de fonctionnement harmonieux ainsi que la mobilisation de tous les membres du laboratoire. Des discussions de la journée ressort une satisfaction générale sur le fonctionnement interne du laboratoire avec le conseil de direction, le conseil de service avec 6 élus pour 2 ans (permettant un renouvellement rapide) et « café du mardi » qui s'avère efficace pour le transfert des infos. L'ambiance paraît sereine dans l'unité. Ceci traduit le parti pris d'un fonctionnement ouvert et transparent, et de décisions d'ordre collectif largement débattues. Le comité a aussi noté un certain enthousiasme général pour la recherche qui se dégageait des exposés, des conversations, et de l'ambiance de l'unité. Il tient à souligner sa satisfaction à ce propos.

7) **En conclusion**, le comité est extrêmement satisfait du bilan du laboratoire, et cela aussi bien pour les aspects scientifiques que pour les aspects humains. Il approuve et soutient le projet de renouvellement, dans la composition des équipes et dans ses projets scientifiques. Finalement, le comité apprécie l'énergie de la direction à monter, soutenir et développer le laboratoire qu'elle anime, et constate qu'elle lui insuffle par là beaucoup de cohérence et de dynamisme. Le comité lui apporte toute sa confiance.

Composition du comité d'évaluation:

Nom	Fonction
Sovan Lek	Professeur Université Paul Sabatier, Labo. Evolution & Diversité Biologique, UMR 5174 CNRS-UPS
Marianne Le Bail	Professeure Agro Paris Tech, UMR SADapt INRA–Paris Grignon)
Olivier François	Professeur INP Grenoble, Equipe TIMB
François Charpellier	Directeur de Recherche, INRIA, projet MAIA
Antoine de Daruvar	Professeur Université Bordeaux2, Centre de Bioinformatique