

**Projet 2011-2014 de l'Equipe LSTI / CROCUS
Calcul de Risque, Optimisation et Calage
par Utilisation de Simulateurs**

Auto-analyse

Points forts

Il existe aujourd'hui, dans l'industrie, en ingénierie de l'environnement, et plus généralement en sciences pour l'ingénieur, un réel besoin de méthodes pour l'exploitation de simulateurs. Ce besoin vient notamment des incertitudes sur les données, d'utilisations répétitives, et de couplages entre le simulateur et d'autres sources d'information. Par exploitation, nous entendons des tâches telles que l'analyse de sensibilité, le calage de modèles, l'optimisation, l'estimation de risque, etc. Cette thématique est émergente et continuera à se développer car elle est liée à des défis scientifiques fondamentaux (apprentissage en haute dimension, optimisation de fonctions bruitées).

Points faibles

L'équipe CROCUS est relativement jeune. Elle commence à être connue nationalement dans son domaine d'activité par les industriels et les chercheurs. Par contre, ni l'équipe ni l'UR LSTI à laquelle elle appartient n'est encore reconnue par les instances d'évaluation nationales telles que le CNRS.

Le second point faible de l'équipe tient à sa petite taille (5 enseignants chercheurs permanents) et au fait que plusieurs de ses activités d'enseignement et de valorisation reposent sur du personnel temporaire. Enfin, l'équipe assume une charge d'enseignement importante.

Opportunités

Le projet scientifique développé ci-après est en phase avec le projet d'« Institut Fayol ». Cet Institut est consacré aux méthodes et outils avancés de la performance industrielle et sera créé à l'ENSM-SE en 2010¹. La compatibilité de ces deux projets résulte

- des interfaces que la thématique « exploitation de simulateurs » crée entre les experts de la simulation des autres UR de l'Ecole et les mathématiciens, informaticiens et experts des organisations qui forment notre UR LSTI,
- des applications en sciences de l'environnement prévues. Dans l'équipe, ces applications sont notamment portées par Mireille Batton-Hubert, une spécialiste de l'ingénierie de l'environnement (centre SITE).

Risques

L'équipe CROCUS est de petite taille (5 enseignants chercheurs permanents) et donc sensible à d'éventuels mouvements de personnel.

Projet et objectifs scientifiques de l'équipe

Nous envisageons trois directions de développement, qui viennent affiner ou compléter nos travaux actuels :

- la propagation d'incertitudes et l'optimisation par calcul distribué,
- les mélanges de modèles,
- la méta-modélisation sur bases physiques.

Ces trois points sont repris ci-après. Sur le plan applicatif, nous prévoyons une intensification de nos travaux relatifs aux problématiques environnementales et en particulier à la gestion de l'eau.

Sous -thème 1 : Optimisation et propagation d'incertitudes en calcul distribué

Les ressources en calcul continuent de croître à un rythme très rapide. Depuis 20 ans, les ressources en calcul haute performance doublent chaque année. Récemment, cette croissance repose principalement sur l'essor du calcul distribué (la mise en commun de ressources informatiques). La perspective de plus grandes ressources de calcul permet d'anticiper une utilisation accrue de la simulation dans la prise de décision en sciences pour l'ingénieur.

Cependant, on peut aussi anticiper que, sans le développement de nouvelles méthodes numériques, adaptées au calcul distribué, même l'apparition de grilles de calcul ou d'ordinateurs haute performance atteignant les petaflops changera relativement peu l'utilité de la simulation :

- La précision physique des simulateurs évolue avec les ressources en calcul et tend à les saturer. Ainsi, les thématiques de l'équipe qui concernent le coût des simulateurs resteront d'actualité malgré les progrès de l'informatique.

¹ Cf. lettre de cadrage ORGA 2.41, P. Jamet, 27 mars 2009.

25/11/2009

- Les méthodes actuelles de propagation d'incertitudes et d'optimisation passent mal à l'échelle (i.e., perdent en efficacité) quand le nombre de processeurs croît car elles ont été imaginées pour des processeurs uniques.
- Enfin, pour des raisons de fractionnement des compétences, de confidentialité ou de propriété du savoir, la collaboration d'experts disciplinaires ou d'entreprises autour d'un même projet va de plus en plus impliquer des simulations distribuées chez chaque acteur et connectées par Internet. L'existence de méthodes d'analyse d'incertitudes et d'optimisation aptes à travailler dans un tel cadre devient nécessaire.

Ainsi, l'équipe CROCUS a pour perspective de contribuer au développement de méthodes de propagation d'incertitudes et d'optimisation pour des infrastructures informatiques distribuées. Ces nouvelles méthodes devront être

- aussi asynchrones que possible,
- robustes aux pannes de processeurs et de transmissions,
- bénéficier de gains en efficacité superlinéaires avec le nombre de processeurs.

Compétences actuelles dans le domaine : des travaux préliminaires ont été réalisés par l'équipe pour paralléliser un important algorithme d'optimisation globale, la méthode EGO [Amigo, Ginsbourger et al., 2007 (E4.46)]. La parallélisation tire partie des lois jointes d'ensembles de points décrites par les processus aléatoires.

Dans le domaine de la propagation d'incertitudes, l'inférence bayésienne et les méthodes de Monte Carlo améliorées (e.g., l'échantillonnage d'importance, les méthodes MCMC) seront les bases de nos développements.

La mise en œuvre de tels algorithmes dans des grilles de calcul suppose une maîtrise des techniques de l'informatique distribuée. Sur ce point, une montée progressive en puissance est prévue à partir de 2009. Le projet ANR/OMD2 (pour Optimisation MultiDisciplinaire Distribuée) qui commence fin 2009 sera l'occasion de prendre en main le middleware ProActive (développé par Activeon, INRIA Sophia). Plus généralement, le projet 1 sera l'occasion de collaborations avec les informaticiens du LSTI. Par ce projet, l'équipe s'attaque à un des défis du LSTI, celui de la distribution et de la coopération et s'ouvre au second axe transverse du laboratoire portant sur les systèmes distribués et les e-technologies . Un projet ANR/COSINUS, nommé ID4CS (Integrated Design for Complex Systems) a été proposé en partenariat avec G. Picard de l'équipe ISCOD en 2009. ID4CS a été accepté.

Sous-thème 2 : Mélange de modèles

Simulateurs redondants

En général, il n'existe pas un mais plusieurs simulateurs décrivant l'objet d'étude. Ces simulateurs peuvent différer :

- par la physique qu'ils capturent. Par exemple, un simulateur d'écoulement pétrolier peut intégrer la loi de Darcy, mais celle-ci peut aussi être couplée avec la thermique et des transformations chimiques.
- par la finesse avec laquelle ils représentent une même physique : taille des mailles dans les méthodes de résolution d'EDP par discrétisation (éléments finis, différences finies, volumes finis, éléments frontières), hypothèse de symétrie permettant de réduire les dimensions du problème (e.g., hypothèse de symétrie radiale autour de puits en écoulements poreux).
- par les principes de leur construction. On opposera les modèles basés sur la résolution d'équations physiques aux modèles reproduisant des observations sans physique interne, tout en notant que des modèles intermédiaires existent.
- par les ressources numériques (temps de calcul, place mémoire) qu'ils utilisent pour être évalués.

Si plusieurs modèles perdurent, c'est qu'ils réalisent différents compromis entre représentativité physique, qualité de prédiction et consommation de ressources numériques. Des enjeux stratégiques, comme la non-dépendance aux entreprises développant les simulateurs, vont aussi dans le sens d'une diversité de modèles.

Pour propager les incertitudes et optimiser avec ces simulateurs, il est important de tenir compte des différents modèles disponibles. Par exemple, les modèles de faible précision mais rapides peuvent orienter une procédure d'optimisation à son début. Ils doivent être périodiquement complétés par des simulations fines de manière à converger vers des résultats compatibles avec les meilleures simulations. De même, lorsque la simulation sert à dimensionner des normes (de pollution, de sécurité) et qu'il existe plusieurs méthodes de simulation (typiquement, des modèles basés sur la physique et des modèles capitalisant des mesures antérieures), il est important de prendre en compte simultanément les prédictions données par différents simulateurs.

25/11/2009

Pour mélanger les prédictions de simulateurs concurrents, nous nous intéresserons aux approches de fusion de données (Bayes, fonctions de croyance, analyse multicritère). Pour les applications à l'optimisation, la prise en compte de plusieurs modèles simultanément est le sujet de l'optimisation multi-niveau, sujet qui complétera cette direction de recherche.

Compétences actuelles dans le domaine : nous avons commencé à travailler sur ces sujets à travers la méta-modélisation, en particulier avec le krigeage [e.g., E4.2, E4.3, E4.7, E4.29, E4.31]. Il y a dans ce cas deux niveaux de simulateurs, le simulateur physique fin, et le méta-modèle de krigeage.

Simulateurs complémentaires : multidisciplinarité et couplage de modèles

Il arrive aussi souvent que plusieurs simulateurs soient complémentaires. C'est le cas de l'analyse multidisciplinaire et de l'emboîtement de simulateurs où certains modèles produisent des paramètres utilisés par d'autres modèles. Par exemple, les avions sont conçus sous l'angle de l'aérodynamique, du contrôle, du calcul de structures et de la motorisation. Un autre exemple est celui de la réaffectation de territoires post-industriels (e.g., le projet PiLoT) dans lequel intervient la biologie (phyto-rémédiation), l'hydrogéologie, l'urbanisme et la chimie. Les scénarios d'emboîtement de modèles vont se généraliser avec le développement de la simulation et sa distribution à travers Internet.

Sur un plan numérique, la multidisciplinarité pose le problème de la gestion des couplages entre disciplines : les simulations de chaque discipline sont couplées par la physique, mais la prise de décision a besoin d'un certain niveau de découplage pour que les acteurs puissent travailler de manière autonome. Les types de simulations peuvent aussi grandement différer entre disciplines (e.g., calculs aérodynamiques et calculs de structures) et il ne faut pas que les disciplines les plus lentes paralysent l'ensemble du processus de décision.

Les stratégies de gestion de simulations complémentaires forment les domaines de l'analyse multi-physique et de l'optimisation multidisciplinaire. Elles constituent un de nos axes de recherche, qui sur le plan technique a une parenté forte avec le projet 1 (calcul distribué) et le sous-projet précédent (simulateurs redondants). L'équipe possède une connaissance de l'optimisation multidisciplinaire du fait du pilotage des deux projets OMD (ANR et Fondation Cetim, cf. bilan). Elle a également déjà travaillé sur l'interopérabilité entre modèles à échelles différentes (« *multinesting* » pour l'atmosphère), et sur les interfaces entre modélisations dans le cas des écoulements hydrauliques.

Sous-thème 3 : Méta-modélisation prenant en compte la physique et les connaissances d'experts

Le domaine de l'apprentissage (ou, dans des espaces continus, méta-modélisation) a connu un développement important ces dix dernières années avec des méthodes telles que les réseaux de neurones, les réseaux à bases radiales, le krigeage, les machines à supports vectoriels, GAM, Ces méthodes peuvent être considérées comme de la modélisation de boîtes noires dans la mesure où les modèles sont construits uniquement à partir de données d'entrée et de sortie. A l'opposé se trouvent les méthodes de modélisation physique basées en général sur la résolution d'équations aux dérivées partielles (éléments finis, différences finies, volumes finis, éléments frontières).

Nous proposons de travailler au développement de méthodes de méta-modélisation intermédiaires entre les méthodes boîtes noires et les méthodes physiques. Ceci sous-entend une relative spécialisation de nos méta-modèles : nous nous intéresserons à la modélisation d'écoulements en milieux poreux et à ses applications en gestion de l'eau.

Les méthodes pressenties pour bâtir de telles « boîtes grises » sont :

- les méthodes d'extraction de champs pertinents par POD (proper orthogonal decomposition) ou analyse en composantes principales,
- la construction de méta-modèles généralistes sous contraintes de respect de relations physiques,
- la mise à jour de prédictions au moyen de mesures (techniques d'assimilation de données d'échantillonnage).

Compétences actuelles dans le domaine : Eric Touboul (a travaillé chez Schlumberger à la modélisation dans le domaine du pétrole), et Mireille Batton-Hubert. Une thèse sur le sujet de la méta-modélisation en hydrogéologie commence fin 2009.

25/11/2009

Adéquation des moyens humains et financiers de l'unité avec le projet

Volume d'activité prévu par projet

L'équipe est actuellement composée de 5 enseignants-chercheurs permanents, soit 20 hommes/an permanents entre 2011 et 2014. La moitié de ce temps est consacré aux tâches d'enseignement et d'administration. En termes de recherche, le volume d'activité restant se répartira comme suit.

Table 1 : Répartition des hommes/an de permanents recherche (hors ressources CDD et nouveaux arrivants)

	M. Batton	X. Bay	R. Le Riche	O. Roustant	E. Touboul	total
P1 : calc. dist.	0	1	1	1	0	3
P2 : mél. mod.	1	0	1	1	1	4
P3 : mm. phys.	1	1	0	0	1	3
total	2	2	2	2	2	10

Ces ressources en permanents seront complétées de ressources d'emplois CDD (post-doctorants, doctorants, ingénieurs, techniciens). En prédiction, on prolonge le volume moyen de CDD de ces trois dernières années (3 CDD chaque année) hors doctorants et on ajoute un doctorant par projet soit $3*3+3*3 = 18$ ho.ans, répartis uniformément sur les trois projets :

Table 2 : Répartition des hommes/an de recherche avec CDD prévus

	Cumul sur 4 ans	Par an
P1 : calc. dist.	$3 + 6 = 9$	$9/4 = 2.25$
P2 : mél. mod.	$4 + 6 = 10$	$10/4=2.5$
P3 : mm. phys.	$3 + 6 = 9$	$9/4 = 2.25$
Total	28	7

Analyse prospective à moyen et long terme des besoins et des compétences nécessaires aux mutations scientifiques et techniques identifiées, la politique de formation

Un projet d'équipe ambitieux

Le projet de l'équipe CROCUS s'attaque à trois points scientifiques difficiles et importants : l'optimisation avec prise en compte des incertitudes, la fusion de modèles et la métamodélisation intégrant les EDPs de la physique. Ce projet est pleinement en phase avec les défis et les axes transverses du LSTI : passage à l'échelle (montée en nombre de variables), intégration (des modèles), coopération et distribution (des calculs), et robustesse (aux incertitudes).

Comme le montre la Table 2 ci-dessus, le projet de l'équipe CROCUS alloue à chacun de ces sous-thèmes près de 10 hommes/an en recherche, ce qui est suffisant et équilibré.

Le plan de travail de CROCUS inclut un recours conséquent aux CDD (les 2/3 du temps de recherche), ce qui implique un certain nombre de tâches administratives additionnelles pour les permanents (e.g., portage de consortium). De plus, il faut rappeler que l'équipe joue un rôle central dans l'enseignement des mathématiques à l'ENSM-SE. Ainsi, le projet aurait toutes les chances de succès avec 3 hommes/an sur chacun des trois thèmes de recherche proposés, ce qui revient à ajouter deux enseignants-chercheurs permanents à l'équipe.

Collaborations

Des collaborations seront nécessaires pour mener à bien ces projets dans les meilleures conditions. On identifie les besoins en collaborations suivants :

- Projet 1, calcul distribué : collaboration avec des informaticiens (internes à l'ENSM-SE, ou collaboration avec l'INRIA) pour l'informatique de distribution.
- Projet 2, mélange de modèles : collaboration avec des utilisateurs de modèles multiples, a priori des ingénieurs et chercheurs en science de l'environnement pour des problèmes de polluants ou de gestion de l'eau.
- Projet 3, méta-modélisation physique : collaboration avec des créateurs de simulateurs sur base physique dans les autres UR de l'Ecole.

La politique de construction de partenariats

1. L'équipe, comme l'ensemble des équipes de l'UR LSTI, postule à l'Institut Fayol.
2. L'équipe participera aux initiatives stéphanoises du type collégium d'ingénierie (Manufacture Innovante).
3. Nous collaborerons avec d'autres mathématiciens du site de St Etienne.
4. Nous allons construire des partenariats dans le cadre des appels d'offre (ANR, 7^e PCRD, etc.).
5. Nous développerons des partenariats internationaux (Etats-Unis, ouverture d'une autre collaboration solide souhaitable).

25/11/2009

Schéma de financement

- ANR / OMD2 (accepté, débute fin 2009, finance 3,5 ho.ans dans l'équipe).
- ANR / ID4CS (projet en collaboration avec l'équipe LSTI / ISCOD, accepté, finance une demi-thèse).
- DICE2 (en projet).
- Ouvrir un projet Européen.

Valorisation, diffusion de l'information scientifique et technique

- Continuer la croissance en nombre de publications.
- Publier un ou deux livres sur nos thèmes de recherche.
- Mieux exister sur internet (projet « optimisation » déposé à UNIT 2009, sites propres).