

CEMRACS 2013

22 Juillet 2013 - 30 Août 2013

Modéliser et simuler la complexité: approches stochastiques et déterministes.

Nicolas Champagnat, Tony Lelièvre et Anthony Nouy

1 Généralités sur le CEMRACS

Le CEMRACS (Centre d'Eté Mathématique de Recherche Avancée en Calcul Scientifique) a été initié par Yvon Maday et Frédéric Coquel en 1996. Il se déroule chaque été en deux phases. La première phase consiste en une semaine de cours et la deuxième en une période de cinq semaines de recherche sur projet au CIRM, à Luminy. Lors de la deuxième phase, chaque participant travaille en équipe sur un projet proposé soit par un industriel soit par une équipe universitaire. Les équipes sont constituées de jeunes chercheurs encadrés par un ou deux chercheurs confirmés. L'expérience des années précédentes a prouvé l'impact positif important du CEMRACS, non seulement sur l'évolution de ces projets, mais aussi sur l'interaction entre universitaires et industriels. Un séminaire quotidien est organisé afin de susciter des échanges entre les participants et de permettre l'acquisition de nouvelles connaissances dans le thème considéré.

Les thèmes abordés lors des CEMRACS précédents ont été :

- CEMRACS 2012 (S. Descombes, B. Dussoubs, S. Faure, L. Gouarin, V. Louvet, M. Massot, V. Miele) Méthodes numériques et algorithmes pour architectures hautes performances ;
- CEMRACS 2011 (F. Coquel, M. Gutnic, P. Helluy, F. Lagoutière, C. Rohde, N. Seguin) Multiscale Coupling of Complex Models ;
- CEMRACS 2010 (N. Crouseille, E. Sonnendrücker) Modèles numériques pour la fusion contrôlée ;
- CEMRACS 2009 (V. Calvez, M.-A. Dronne, T. Dumont, V. Louvet, P. Vigneaux) Modélisation Mathématique en Médecine ;
- CEMRACS 2008 (J.-B. Apoung-Kamga, L. Boudin, M. Ismaïl, S. Martin, B. Maury, T. Takahashi) Modélisation et Simulation de Fluides Complexes ;
- CEMRACS 2007 (P. Frey, C. Dobrzynski, P. Pébay) Pre and Post Processing in Scientific Computing ;
- CEMRACS 2006 (G. Bal, J. Garnier, D. Lucor) Modélisation de l'aléatoire et propagation d'incertitudes ;
- CEMRACS 2005 (E. Sonnendrücker, C.-D. Munz, M. Dumbser) Aéroacoustique numérique et CFD pour les écoulements turbulents ;
- CEMRACS 2004 (E. Cancès, J.-F. Gerbeau) Mathématiques et applications en biologie et médecine ;
- CEMRACS 2003 (S. Cordier, T. Goudon, M. Gutnic, E. Sonnendrücker) Méthodes numériques pour les problèmes hyperboliques et cinétiques ;

- CEMRACS 2002 (A. Cohen, P.L. Combettes) Méthodes mathématiques pour le traitement d'images ;
- CEMRACS 2001 (Y. Achdou, C. Le Bris, F. Nataf) Modélisation et simulation numérique de problèmes multi-échelles ;
- CEMRACS 2000 (R. Abgrall) Problèmes liés à l'environnement : combustion, stockage des déchets nucléaires ;
- CEMRACS 1999 (F. Coquel, S. Cordier) Approximation numérique d'écoulements complexes de leur description cinétique à leur limite hydrodynamique ;
- CEMRACS 1998 (Y. Maday, F. Coquel) Estimations d'erreur et adaptivité, ondelettes pour le calcul scientifique, méthodes de décomposition de domaine et calcul parallèle, langages orientés objets pour le calcul scientifique ;
- CEMRACS 1997 (Y. Maday, F. Coquel) Optimisation de formes ;
- CEMRACS 1996 (Y. Maday, F. Coquel) Couplage d'équations.

2 Présentation scientifique du CEMRACS 2013

2.1 Biographie des organisateurs

Nicolas Champagnat est Chargé de Recherche à l'INRIA Nancy - Grand Est et membre de l'Institut Élie Cartan de Nancy (IECN). Il est membre de l'équipe-projet INRIA TOSCA. Il travaille principalement sur la modélisation pour la biologie de l'évolution, l'écologie et la génétique des populations. Il travaille également sur les méthodes numériques probabilistes de type Monte Carlo pour les équations aux dérivées partielles. Il a publié une quinzaine d'articles dans des revues à comité de lecture.

Tony Lelièvre est Ingénieur en Chef des Ponts, des Eaux et des Forêts, professeur à l'École des Ponts ParisTech et chercheur au centre de mathématiques appliquées (CERMICS) de l'École des Ponts ParisTech. Il est responsable du groupe de simulation moléculaire et multi-échelles au CERMICS et membre du projet MicMac de l'INRIA Rocquencourt. Il travaille principalement sur les modèles multi-échelles pour les matériaux, la simulation moléculaire et la simulation des écoulements à surface libre. Il a publié deux livres et une quarantaine d'articles dans des revues à comité de lecture.

Anthony Nouy est Professeur des Universités à l'École Centrale de Nantes, au département de Mathématiques et Informatique. Il est membre de l'Institut de recherche en génie civil et mécanique (GeM). Il travaille principalement sur les méthodes numériques pour la quantification d'incertitudes et les méthodes de réduction de modèle pour les problèmes de grande dimension. Il a publié une vingtaine d'articles dans des revues à comité de lecture.

2.2 Thèmes scientifiques

L'objectif de ce CEMRACS est de s'intéresser aux modélisations et aux méthodes numériques qui *combinent des approches déterministes et des approches stochastiques*. Il apparaît en effet que les problèmes relevant de cette double compétence (probabilités, et analyse numérique déterministe) sont de plus en plus nombreux, et les formations des mathématiciens appliqués n'y sont souvent pas bien adaptées. Un des buts du CEMRACS est donc de favoriser des échanges entre diverses communautés de mathématiciens appliqués, autour de problèmes industriels ou de problèmes provenant d'autres champs scientifiques (physique, chimie, biologie).

Cette thématique est actuellement au coeur de nombreuses problématiques industrielles (la propagation des incertitudes, la modélisation à l'échelle moléculaire, etc...), et semble aussi prendre une place de plus en plus importante dans plusieurs champs scientifiques en dehors des mathématiques appliquées (modélisation en biologie, chimie moléculaire, physique statistique computationnelle, etc...).

Nous présentons ci-dessous dans un premier temps, les thématiques applicatives qui seront typiquement au coeur de ce CEMRACS, puis dans un deuxième temps, les aspects méthodologiques que nous aimerions aborder.

2.2.1 Modélisation déterministe et stochastique

Propagation des incertitudes Du point de vue de la modélisation, de nombreux problèmes qui sont posés naturellement de manière stochastique peuvent être efficacement abordés par des approches déterministes. Un exemple typique est *la modélisation et la propagation des incertitudes*. Le calcul de la propagation des erreurs dans un modèle est devenu un enjeu important en calcul scientifique. C'est un sujet qui intéresse indéniablement les industriels et les scientifiques de domaines très différents (climatologie, mécanique, finance, ...), et qui ouvre de nouvelles thématiques de recherche pour les mathématiciens appliqués, à l'intersection des méthodes numériques pour les équations aux dérivées partielles, des méthodes de Monte Carlo, et des approches statistiques. Il s'agit aujourd'hui d'un domaine de recherche extrêmement actif, qui permet de confronter divers domaines des mathématiques appliquées traditionnellement disjoints.

La question posée est très simple: les modèles utilisés actuellement sont de plus en plus complexes, souvent non-linéaires, et il est donc difficile de prédire *a priori* comment des erreurs ou incertitudes sur les paramètres du modèle (les données en entrée) influent sur le résultat (la sortie). Cette problématique générale a été abordée par diverses communautés, en insistant sur différents aspects: l'estimation des événements rares (la fiabilité [11]), les méthodes de réductions de modèle (surface de réponse) pour diminuer les temps de calcul d'évaluation de la sortie et utiliser efficacement des méthodes de Monte Carlo, la modélisation de l'aléa sur les entrées (estimation de la corrélation entre les paramètres aléatoires), le calcul déterministe de la sortie en fonction des données [18] (méthodes issues de la résolution numériques des équations aux dérivées partielles, avec une problématique naturelle de discrétisation des fonctions en grande dimension), pour ne citer que quelques exemples. On renvoie à la revue [28] pour plus de détails sur les approches numériques standards.

Lors de ce CEMRACS, nous aimerions étudier des problèmes industriels autour de cette thématique, notamment en combinant des approches déterministes et stochastiques. On pense par exemple à l'utilisation de modèles réduits obtenus par des approches déterministes pour rendre les méthodes de Monte Carlo plus efficaces.

Dynamique moléculaire et modélisation multi-échelles des matériaux La dynamique moléculaire a pris une place importante dans de nombreux domaines: biologie, physique statistique, sciences des matériaux, chimie, etc... Il s'agit de déduire des propriétés macroscopiques (constantes de réactions, paramètres dans des lois de comportement, états de transition dans une réaction chimique, etc...) à partir de modèles à l'échelle moléculaire [25].

L'approche soulève de nombreux problèmes méthodologiques, notamment en terme des échelles de temps, et des échelles en espace. Pour obtenir des résultats significatifs, il faut simuler des échantillons suffisamment gros, sur des échelles de temps suffisamment longues. De nombreuses méthodes ont été proposées par les praticiens des différents domaines,

mais ces thématiques n'ont pas beaucoup mobilisé la communauté des mathématiciens appliqués jusqu'à maintenant. L'apport d'une approche mathématique sur ces techniques serait pourtant très certainement décisif, pour analyser, classifier et valider les méthodes proposées, et pour proposer des améliorations.

Lors de ce CEMRACS, en collaboration avec des praticiens de divers domaines et des industriels, nous aimerions notamment aborder les questions suivantes:

- l'accélération des méthodes de type kinetic Monte Carlo (utilisées par exemple pour simuler la migration des défauts dans les matériaux sous irradiation au CEA, à EDF ou à l'IRSN),
- la génération efficace de dynamiques métastables [8] (repliement des peptides en biologie),
- l'étude des processus hors-équilibres (par exemple les configurations d'une chaîne de polymère sous écoulement),
- les méthodes numériques efficaces en homogénéisation stochastique (avec notamment de nouvelles approches proposées récemment par A. Gloria [19] ou F. Legoll et C. Le Bris [1]).

Modèles en dynamique des populations Les modèles hybrides, combinant une composante déterministe et une composante aléatoire, interviennent très naturellement en dynamique des populations, du fait notamment de la diversité des échelles impliquées dans les problèmes de biologie des populations. La composante aléatoire peut intervenir au niveau de l'individu (modèles individu-centrés, ou multi-agents), soit au niveau des densités de populations dans le cas de grands effectifs (EDP stochastiques, ou diffusions à valeur mesure). On peut citer trois contextes applicatifs parmi beaucoup d'autres:

- Afin d'étudier la dynamique adaptative de populations sujettes à l'évolution darwinienne, il est naturel de décrire la dynamique des effectifs de façon déterministe, et la dynamique des mutations et de leur invasion de façon aléatoire [26, 10].
- En écologie des communautés, la modélisation des interactions entre espèces et des individus avec leur environnement (climat, ressources, prédation, niches écologiques...), en tenant compte des variabilités spatiales ou temporelles, est par exemple indispensable à l'étude de la colonisation spatiale, l'évolution adaptative ou le maintien de la biodiversité [27]. Dans ce contexte, les espèces abondantes peuvent être modélisées de façon déterministe, tandis que les espèces en faible effectif ou dont la répartition spatiale est soumise à de fortes variations, sont modélisées de façon aléatoire.
- En épidémiologie, lorsque l'on souhaite modéliser la propagation d'une épidémie humaine à grande échelle un rôle crucial est joué par les déplacements à longue distance par avion, qui ne concernent qu'une petite fraction de la population. Dans ce contexte, il est nécessaire de modéliser de façon déterministe les déplacements et les contacts locaux, et de façon aléatoire les déplacements à grande échelle [2].

De plus, les incertitudes sur les paramètres des modèles, voire sur la nature des interactions entre individus, dues à la difficulté d'obtention de données ou à la complexité des phénomènes biologiques, appellent une prise en compte de la propagation d'incertitudes dans les modèles. Quoiqu'encore peu utilisée aujourd'hui, cette approche est certainement amenée à se développer fortement dans le futur.

Lors de ce CEMRACS, l'étude de tels modèles pourrait par exemple être menée dans le cadre du contrôle de procédés d'épuration d'eau, où des bactéries sont utilisées pour l'élimination des composés carbonés. Ces techniques impliquent un grand nombre d'espèces bactériennes d'abondances variables.

2.2.2 Approches mathématiques et méthodes numériques

Méthodes de Monte Carlo avancées Nous avons évoqué dans le paragraphe précédent de nombreuses questions qui nécessitent le développement de nouvelles techniques d'échantillonnage. Par exemple, la problématique de la simulation des événements rares [4] est au coeur des problèmes de fiabilité, ou des difficultés liées à la métastabilité en dynamique moléculaire [24]. Il nous semble qu'aborder cette problématique de manière transversale pourrait permettre un échange fructueux entre les différents domaines où ce problème a été traité, par des approches très différentes.

Il reste par ailleurs de nombreuses questions méthodologiques qui sont cruciales, et essentiellement ouvertes. Par exemple, comment échantillonner efficacement un état stationnaire *multimodal* d'une dynamique markovienne *non réversible* ? Les approches classiques (biaisage et échantillonnage d'importance, ou bien conditionnement et échantillonnage sous contraintes) ne s'appliquent plus. Un autre exemple est donné par l'échantillonnage de *trajectoires*: comment générer efficacement des trajectoires d'un état métastable à un autre ? Il s'agit ici de comprendre la dynamique de phénomènes métastables. Un dernier exemple concerne les techniques de *réduction de modèle* [23]. Comment obtenir des modèles réduits sur quelques degrés de liberté, à partir d'un modèle complet en très grande dimension ? Cette problématique est intéressante du point de vue de la modélisation d'une part (un modèle en petite dimension est plus simple à comprendre) et du point de vue numérique d'autre part (on peut utiliser le modèle réduit pour accélérer la simulation du modèle complet).

Equations aux dérivées partielles stochastiques La thématique des équations aux dérivées partielles stochastiques (EDPS) nécessite, par essence, des compétences croisées entre équations aux dérivées partielles et probabilités. Il s'agit d'un domaine très actif, à la fois sur les aspects théoriques (ergodicité, cadre fonctionnel adéquat pour les EDPS non-linéaires) et sur les aspects numériques (discrétisation efficace, analyse numérique des schémas) [22].

Lors de ce CEMRACS, nous aimerions attirer quelques chercheurs sur cette thématique, car ils pourraient naturellement interagir avec les autres participants. Les EDPS apparaissent naturellement dans de nombreux problèmes de modélisation, par exemple en dynamique des populations (étude de la colonisation ou de l'extinction d'une population spatialement structurée en écologie [17], ou bien populations sujettes à l'évolution de traits continus en génétique des populations [14, 9]) ou en dynamique moléculaire (génération de trajectoires réactives selon la distribution stationnaire d'une EDPS par exemple).

Problèmes en grande dimension Des problèmes de grande dimension apparaissent naturellement dans la modélisation et la quantification des incertitudes (équations aux dérivées partielles en grande dimension, équations aux dérivées partielles paramétrées) mais aussi dans les problèmes d'apprentissage statistique. Récemment, un intérêt croissant a été porté aux méthodes de construction de bases d'approximation adaptées à la représentation de fonctions de grande dimension. Les méthodes d'approximation creuse partent classiquement d'une base d'approximation dans laquelle la fonction à représenter admet potentiellement une représentation creuse [6, 15, 30, 13]. Elles utilisent alors des définitions

des approximations qui permettent leur calcul avec une information limitée sur le modèle (par régression, interpolation, ...) ou des méthodes de projection adaptées présentant une complexité réduite. De nombreux récents développements s'intéressent à l'utilisation de méthodes d'approximation nonlinéaire pour la construction (greedy ou non greedy) de bases réduites choisies dans des ensembles particuliers de fonctions [3, 29, 12, 5, 31]. Ces ensembles de fonctions doivent avoir de bonnes propriétés d'approximation et doivent être paramétrables par un faible nombre de paramètres afin de traiter des problèmes de grande dimension. Des ensembles de fonctions exploitant la structure produit tensoriel des espaces de fonction ont récemment reçu une attention particulière [7, 21, 16, 20]. L'intérêt de ces approches est de permettre l'approximation de fonctions en très grande dimension avec une complexité relativement faible.

De nombreuses difficultés restent cependant à surmonter pour l'analyse de ces méthodes et le développement d'algorithmes pour une large classe de modèles. Nous aimerions aborder en particulier la problématique de la propagation des incertitudes dans les modèles régis par des EDP. Dans ce contexte, il se pose notamment la question du choix des formats d'approximation optimaux pour une classe d'application donnée, et du développement d'algorithmes pour la construction des algorithmes. Un véritable enjeu industriel consiste à développer des algorithmes de construction utilisant des codes de calcul en boîte noire.

Un des objectifs de ce CEMRACS est de rassembler des chercheurs de différentes sensibilités (apprentissage statistique, signal, analyse numérique) afin d'aborder avec un regard nouveau les problèmes de grande dimension dans le calcul scientifique.

3 Programme de la première semaine

La première semaine sera organisée en cinq cours de cinq heures, et neuf exposés d'une heure. Nous prévoyons d'organiser une table rond sur les thèmes du CEMRACS dans le cadre de l'année spéciale "les Mathématiques pour la Planète Terre".

Pour les cours, les intervenants seront:

- Homogénéisation stochastique et matériaux aléatoires: Félix Otto.
- Evènements rares: modèles et simulations: Josselin Garnier.
- Approximation non-linéaire pour les problèmes en grande dimension: Albert Cohen.
- Quantification et méthodes quasi Monte Carlo: Gilles Pagès.
- Equations aux dérivées partielles stochastiques: Gabriel Lord.

Pour les exposés, les intervenants seront:

- Problèmes en grandes dimensions: Boris Khoromskij
- Assimilation de données: Didier Auroux
- Hydrodynamique statistique: Freddy Bouchet
- Dynamique des populations: Sylvie Méléard
- Classification et apprentissage: Sylvain Arlot
- Bases réduites: Gianluigi Rozza
- Apprentissage statistique: Jean-Marc Martinez (CEA)

- Fiabilité: Régis Lebrun (EADS)
- Aspects mathématiques du stockage des déchets radioactifs: Alexandre Ern

References

- [1] A. Anantharaman, R. Costeaouec, C. Le Bris, F. Legoll, and F. Thomines. Introduction to numerical stochastic homogenization and the related computational challenges: some recent developments. volume 22 of *Lecture Notes Series*, pages 197–272. Institute for Mathematical Sciences, National University of Singapore, 2011.
- [2] P. Bajardi, C. Poletto, J.J. Ramasco, M. Tizzoni, V. Colizza, and A. Vespignani. Human mobility networks, travel restrictions, and the global spread of 2009 h1n1 pandemic. *PLOS One*, 6(1):e16591, 2011.
- [3] A.R. Barron, A. Cohen, W. Dahmen, and R.A. DeVore. Approximation and learning by greedy algorithms. *Ann. Statist.*, 36(1):64–94, 2008.
- [4] J.A. Bucklew. *Introduction to Rare Event Simulation*. Springer-Verlag, 2004.
- [5] A. Buffa, Y. Maday, A. Patera, C. Prud'homme, and G. Turinici. A priori convergence of the greedy algorithm for the parametrized reduced basis method. *ESAIM: Mathematical Modelling and Numerical Analysis*, 46:595–603, 2012.
- [6] H-J. Bungartz and M. Griebel. Sparse grids. *Acta. Numer.*, 13:147–269, 2004.
- [7] E. Cancès, V. Ehrlacher, and T. Lelièvre. Convergence of a greedy algorithm for high-dimensional convex nonlinear problems. *Mathematical Models & Methods In Applied Sciences*, 21(12):2433–2467, 2011.
- [8] F. Cérou, A. Guyader, T. Lelièvre, and D. Pommier. A multiple replica approach to simulate reactive trajectories. *J. Chem. Phys.*, 134:054108, 2011.
- [9] N. Champagnat, R. Ferrière, and S. Méléard. Unifying evolutionary dynamics: From individual stochastic processes to macroscopic models. *Theor. Pop. Biol.*, 69:297–321, 2006.
- [10] Nicolas Champagnat and Sylvie Méléard. Polymorphic evolution sequence and evolutionary branching. *Probab. Theory Related Fields*, 151(1-2):45–94, 2011.
- [11] C. Coccozza-Thivent. *Processus stochastiques et fiabilité des systèmes*. Springer, 1997.
- [12] A. Cohen, R. DeVore, and C. Schwab. Convergence rates of best n-term galerkin approximations for a class of elliptic spdes. *FOUNDATIONS OF COMPUTATIONAL MATHEMATICS*, 10(6):615–646, 2010.
- [13] A. Doostan and H. Owhadi. A non-adapted sparse approximation of pdes with stochastic inputs. *Journal of Computational Physics*, 230(8):3015–3034, 2011.
- [14] A. M. Etheridge. Survival and extinction in a locally regulated population. *Ann. Appl. Probab.*, 14(1):188–214, 2004.
- [15] C.G. Webster F. Nobile, R. Tempone. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2309–2345, 2007.

- [16] A. Falco and A. Nouy. Proper Generalized Decomposition for nonlinear convex problems in tensor Banach spaces. *Numerische Mathematik*, (2012).
- [17] Nicolas Fournier and Sylvie Méléard. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *Ann. Appl. Probab.*, 14(4):1880–1919, 2004.
- [18] R.G. Ghanem and P.D. Spanos. *Stochastic finite elements: a spectral approach*. Springer-Verlag, 1991.
- [19] A. Gloria and J.-C. Mourrat. Spectral measure and approximation of homogenized coefficients. *Probab. Theory Related Fields*, 2012. To appear.
- [20] W. Hackbusch. *Tensor Spaces and Numerical Tensor Calculus*, volume 42 of *Series in Computational Mathematics*. Springer, 2012.
- [21] Boris N. Khoromskij and Christoph Schwab. Tensor-structured galerkin approximation of parametric and stochastic elliptic pdes. *SIAM Journal on Scientific Computing*, 33(1):364–385, 2011.
- [22] O. P. Le Maître and O. M. Knio. *Spectral Methods for Uncertainty Quantification With Applications to Computational Fluid Dynamics*. Scientific Computation. 2010.
- [23] F. Legoll and T. Lelièvre. Effective dynamics using conditional expectations. *Nonlinearity*, 23:2131–2163, 2010.
- [24] T. Lelièvre. Two mathematical tools to analyze metastable stochastic processes. <http://arxiv.org/abs/1201.3775>.
- [25] T. Lelièvre, M. Rousset, and G. Stoltz. *Free energy computations: A mathematical perspective*. Imperial College Press, 2010.
- [26] J. A. J. Metz, S. A. H. Geritz, G. Meszéna, F. J. A. Jacobs, and J. S. van Heerwaarden. Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction. In *Stochastic and spatial structures of dynamical systems (Amsterdam, 1995)*, Konink. Nederl. Akad. Wetensch. Verh. Afd. Natuurk. Eerste Reeks, 45, pages 183–231. North-Holland, Amsterdam, 1996.
- [27] Peter J. Morin. *Community Ecology*. Wiley-Blackwell Press, Malden, 1999.
- [28] A. Nouy. Recent developments in spectral stochastic methods for the numerical solution of stochastic partial differential equations. *Arch. Comput. Method. E.*, 16(3):251–285, 2009.
- [29] G. Rozza, D. B. P. Huynh, and A. T. Patera. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. *Archives of Computational Methods In Engineering*, 15(3):229–275, 2008.
- [30] C. Schwab and C. Gittelson. Sparse tensor discretizations of high-dimensional parametric and stochastic pdes. *Acta Numerica*, 20:291–467, 2011.
- [31] V.N. Temlyakov. Greedy approximation. *Acta Numerica*, 17:235–409, 2008.