

# Génération automatique de codes performants et fiables pour l'assimilation de données

## Mots clés :

- **Directeur de thèse** : Fabienne Jezequel
- **Co-encadrant(s)** :
- **Unité de recherche** : Laboratoire d'informatique de Paris 6
- **Ecole doctorale** : École Doctorale Informatique, Télécommunications, Électronique de Paris
- **Domaine scientifique principal**: Divers

## Résumé du projet de recherche (Langue 1)

Dans le domaine des sciences de l'environnement, la modélisation des phénomènes climatiques fait appel à des codes numériques de très grande taille. De nombreux paramètres de ces modèles sont estimés par assimilation d'observations in-situ ou satellitaires. Nous nous intéressons particulièrement à l'assimilation des données dans les modèles par méthodes variationnelles. Celles-ci sont basées sur la minimisation d'une fonction de coût dont le gradient est calculé par l'adjoint du modèle direct. Le calcul de l'adjoint d'un gros code est une opération délicate. Elle engendre un long travail sur le modèle direct et nécessite dans la plupart des cas une réécriture de celui-ci. Ce constat a mené au développement d'un logiciel d'aide à la programmation d'un code numérique et de son adjoint. Ce logiciel, appelé YAO [1,2], utilise une modélisation sous forme de graphe pour générer à la fois le code du modèle direct et celui de l'adjoint ainsi qu'une boîte à outil type permettant de réaliser entièrement l'assimilation de données. L'approche YAO s'est montrée pertinente sur des codes de tailles petites à moyennes. Des travaux portent actuellement sur la génération automatique par YAO de codes parallèles afin de traiter des problèmes beaucoup plus vastes (par exemple, la modélisation de la circulation océanique à l'échelle globale). Une version de YAO permettant la génération automatique de codes parallèles pour les architectures à mémoire partagée a été développée [3]. Les codes numériques générés utilisent la bibliothèque OpenMP. Une autre partie du projet concerne la validation numérique des codes générés par YAO. Ce travail consiste à intégrer à YAO la bibliothèque CADNA [4,5] afin que les applications générées soient automatiquement associées à cet outil de contrôle des erreurs d'arrondi. Dans un premier temps, les efforts ont porté sur la validation numérique de codes séquentiels générés par YAO, afin que celle-ci devienne un procédé standard qui ne nécessite pas de modification du code de l'application. La qualité numérique de différentes applications générées par YAO a ainsi pu être étudiée [6]. Cette thèse vise à faire générer par YAO des codes numériques à la fois performants et fiables pour les architectures massivement parallèles, qu'elles soient homogènes ou hétérogènes. Il s'agit de générer automatiquement, dans le cadre du formalisme de YAO et en s'appuyant sur les directives OpenMP générées par YAO, des codes pouvant s'exécuter efficacement sur plusieurs noeuds de calcul multicœurs (via MPI, et grâce à un recouvrement automatique des communications par du calcul) ou sur des architectures parallèles hétérogènes comme les GPU (via des directives de compilation comme par exemple celles du standard OpenACC [7]). Le logiciel YAO sera ensuite modifié afin de générer automatiquement des codes parallèles sur lesquels la propagation d'erreur d'arrondi sera contrôlée.

Références : [1] YAO: a Software for Variational Data Assimilation in Numerical Models, <http://www.locean-ipsl.upmc.fr/~yao>. [2] L. Nardi, C. Sorrow, F. Badran, S. Thiria, YAO: A Software for Variational Data Assimilation Using Numerical Models, LNCS 5593, Computational Science and Its Applications - ICCSA 2009, pp. 621-636 (2009). [3] L. Nardi, F. Badran, P. Fortin, S. Thiria, YAO: a generator of parallel code for variational data assimilation applications, 14th IEEE International Conference on High Performance Computing and Communications (HPCC-2012), Liverpool, UK, 25-27 June 2012. [4] The CADNA library, <http://www.lip6.fr/cadna>. [5] F. Jézéquel, J.-M. Chesneaux, CADNA: a library for estimating round-off error propagation, Computer Physics Communications, Vol. 178, No 12, pp. 933-955 (2008). [6] J. Brajard, P. Li, F. Jézéquel, H.-S. Benavides, S. Thiria, Numerical Validation of Data Assimilation Codes Generated by the Yao Software, SIAM Annual Meeting, San Diego, California, USA, July 8-12 2013. [7] <http://www.openacc-standard.org/>

## Résumé du projet de recherche (Langue 2)

Les défis à résoudre au cours de cette thèse sont les suivants. -\* Le premier défi concerne le passage à l'échelle de YAO. Il s'agit de faire générer par YAO, grâce au graphe sous-jacent de son formalisme, des codes performants sur architecture à mémoire distribuée et sur architecture hybride CPU-GPU. -\* Le passage à l'échelle de YAO ne doit cependant pas se faire sans s'assurer de sa fiabilité. Des travaux ont montré la faisabilité de la validation numérique grâce à CADNA de codes parallèles pour architecture à mémoire distribuée et pour architecture hybride CPU-GPU. Néanmoins des optimisations restent nécessaires pour réduire le coût en temps d'exécution de ces versions de CADNA. D'autre part, une version de CADNA devra être proposée pour la validation de codes sur architectures à mémoire partagée. Une attention particulière sera portée à l'optimisation des performances de cette version multi-cœur de CADNA. -\* Puis le logiciel YAO sera modifié afin de générer automatiquement des codes parallèles utilisant ces versions optimisées de CADNA. Ces travaux rentrent dans le cadre de la validation numérique des calculs à haute performance qui constitue actuellement un défi important. -\* Enfin des problèmes en sciences du climat de taille réaliste seront étudiés, par exemple la modélisation numérique de l'océan par la plate-forme NEMO (<http://nemo-ocean.eu>). La qualité numérique de résultats de codes parallèles générés par YAO sera analysée. En outre, de nouvelles méthodologies de programmation pourront être proposées afin d'optimiser les critères de convergence dans ces codes.

## Informations complémentaires (Langue 1)

Cette thèse entre dans le cadre d'une collaboration avec le LOCEAN (Laboratoire d'Océanographie et du Climat : Expérimentations et Approches Numériques, laboratoire de l'université Paris 6), mais aussi de collaborations avec des équipes aux Etats-Unis et au Royaume-Uni, notamment des collaborations pluri-disciplinaires qui ont pour but la validation numérique de codes de simulation en physique atomique ou en sciences de l'environnement : -\* l'équipe dirigée par Stan Scott à "Queen's University of Belfast" (Belfast, UK) qui développe des applications en physique atomique -\* les équipes dirigées par Philippe Baveye à "Rensselaer Polytechnic Institute" (Troy, New York, USA) et à "Alberty University" (Dundee, UK) qui développent des applications en sciences de l'environnement.