



ÉDITORIAL
Denis Girou

L'année 2014 a été la deuxième dans l'exploitation des calculateurs Turing (Blue Gene/Q) et Ada (x3750).

À nouveau, ils ont fait preuve d'une fiabilité remarquable, avec pour chacun une disponibilité, hors les arrêts programmés pour des mises à niveau des systèmes, de l'ordre de 99,5 %. De plus, suite à la décision prise par GENCI et le CNRS à la fin de l'année, c'est d'une configuration augmentée de moitié de la machine Turing dont nous disposons à présent, ce qui a fait passer celle-ci à quasiment 100 000 cœurs de calcul et 100 To de mémoire centrale, et ce qui permet d'attribuer désormais chaque année 655 Mh sur cette machine. Après une installation réalisée dans le temps record d'un mois, l'ensemble de la nouvelle configuration est opérationnel depuis la fin novembre.

Mais disposer d'une puissance de calcul ne suffit pas et, afin d'offrir un ensemble équilibré, il est notamment impératif d'y associer des ressources de stockage adaptées. La première opération d'extension de nos capacités de stockage, qui avait été annoncée dans le numéro précédent de cette lettre, a donc vu l'ancienne machine d'archives Gaya remplacée l'été dernier par la machine Ergon de nouvelle génération, aux capacités fortement étendues, et la migration, transparente pour nos utilisateurs, des 4 pétaoctets de données précédemment stockées sur cartouches magnétiques a été achevée à la fin de l'année. La seconde opération, dont la procédure d'acquisition a été finalisée en décembre dernier, est en cours de mise en œuvre

et permettra un accroissement notable, en faisant plus que la doubler, de notre capacité de stockage disques, dont une partie correspond à l'extension de la machine Turing et l'autre vise à faire face à la croissance des besoins des quelques prochaines années. Ces disques de nouvelle génération auront également une performance doublée par rapport à ceux de la configuration actuelle.

Nous faisons par ailleurs figurer plus loin dans ces pages, comme tous les ans, le bilan des consommations de ressources pour l'année passée, ainsi que la répartition des ressources attribuées lors de la première campagne DARI de cette année.

Du côté des projets européens, alors que se termine la troisième phase d'implémentation du projet PRACE, dite 3IP, la quatrième phase 4IP démarre. En même temps, se mettra en place dans le courant de l'année la seconde phase du projet PRACE proprement dit, dite PRACE 2.0, qui couvrira les années 2015 à 2020. D'autre part, ainsi qu'il est expliqué dans ces pages, le projet E-Biothon, quant à lui, a identifié de nouvelles applications scientifiques du domaine des sciences de la vie qui sont en cours de portage et pourront s'exécuter sur la plate-forme associée, une partie de notre ancien calculateur Blue Gene/P, à la suite des trois premières applications sélectionnées l'an passé.

Enfin, c'est dans un long et très pédagogique article que Jean-Louis Dufresne, le directeur adjoint de l'Institut Pierre Simon Laplace, nous explique en quoi consiste la modélisation numérique du climat, avec ses diverses composantes scientifiques, en exposant également ses difficultés, ses progrès pendant ces dernières décennies et aujourd'hui ses enjeux et ses perspectives. En cette année de préparation du sixième exercice CMIP (*Coupled Model Intercomparison Project*) du programme de recherche mondial sur le climat, qui s'étendra sur les années 2016 à 2018, et alors que Paris se prépare à accueillir avec éclat, au mois de décembre, la 21^e conférence mondiale sur le climat, les enjeux, certes scientifiques, mais aussi politiques et sociétaux de ces questions sont plus que jamais d'actualité.

sommaire

- 1 • Éditorial
- 2 à 7 • Modélisation numérique du climat
- 8 et 9 • Bilan des consommations 2014 et allocations 2015
• Bilan des formations IDRIS 2014

- 10 • Extension de la configuration Blue Gene/Q
• Migration de Gaya vers Ergon et extension du stockage
- 11 • Remise de la médaille de Cristal
• Conférence Supercomputing 2014
• Activités de l'IDRIS dans le projet européen PRACE en 2014
• Projet E-Biothon
- 12 • Actualités
• Calendrier des formations

LA MODÉLISATION NUMÉRIQUE DU CLIMAT



Jean-Louis Dufresne : Directeur de recherche CNRS au Laboratoire de météorologie dynamique et à l'Institut Pierre Simon Laplace

Introduction

Le climat est étroitement lié à la météorologie et pour les différencier on peut dire de façon schématique que la météorologie est l'étude déterministe du temps qu'il fait, avec le détail de sa chronologie, alors que le climat est l'étude des caractéristiques statistiques du temps qu'il a fait ou qu'il pourrait faire. Le climat comporte plusieurs particularités, certaines étant communes avec la météorologie, d'autres étant spécifiques : on ne peut en isoler une sous-partie, on ne peut pas faire d'expériences contrôlées, c'est un système chaotique où de nombreux phénomènes interagissent avec des échelles caractéristiques temporelles et spatiales qui varient continûment sur plusieurs ordres de grandeurs, etc. La modélisation numérique permet d'aborder ces différentes difficultés et joue un rôle central dans l'étude du climat, l'interprétation et la mise en cohérence de nombreux résultats aussi bien théoriques qu'issus d'observations.

Le développement des modèles climatiques est intimement lié à celui des ordinateurs. Dans les années 1950, John von Neumann et ses collègues ont conçu le premier modèle numérique de météorologie sur le premier ordinateur programmable existant. Quelques années plus tard, Norman Phillips et son équipe ont réalisé une simulation de l'atmosphère de l'hémisphère Nord sur une période de 31 jours, qui peut être considérée comme la première simulation « climatique ». La philosophie du développement de ces modèles, proposée par Jules Charney, était de simplifier au maximum le problème pour pouvoir le traiter dans sa globalité, puis de complexifier progressivement les modèles afin d'améliorer les points identifiés comme les plus critiques.

Les modèles numériques de climat sont en premier lieu des modèles de circulation générale (MCG) de l'atmosphère et de l'océan : ils sont basés sur les équations de la mécanique des fluides (équations de Navier-Stokes) qui expriment la conservation de la masse, de la quantité de mouvement et de l'énergie pour un fluide.

Pour la météorologie et le climat, ces équations sont généralement simplifiées en supposant que la pression est hydrostatique, et leur résolution numérique sur une sphère en rotation (la Terre ou une autre planète) constitue le « noyau dynamique » des modèles. Cette résolution repose sur un maillage qui peut avoir différentes structures avec des mailles de dimension allant de quelques dizaines à quelques centaines de kilomètres sur l'horizontale et de quelques dizaines à quelques centaines de mètres sur la verticale. Le temps est lui-même découpé en pas élémentaires d'une durée typique de quelques minutes à 30'. Ce noyau dynamique est complété par une modélisation d'autres phénomènes (rayonnement, changement d'état, équation d'état, etc.) d'une part, par une modélisation des effets qui ne sont pas explicitement résolus à cause de la discrétisation spatiale et temporelle limitée d'autre part. Ces modèles de circulation générale sont complétés par des modèles de surface continentale, de la glace de mer, des glaciers et des calottes de glace. En plus des phénomènes physiques, les phénomènes chimiques ou biogéochimiques ont été progressivement ajoutés dans de nombreux modèles climatiques.

Évolution des questions scientifiques

Les modèles de climat sont largement utilisés aussi bien pour l'étude du climat actuel que de son évolution récente, passée ou future. Parmi les résultats les plus visibles, il y a les projections de l'évolution du climat régulièrement publiées dans les rapports successifs du GIEC (Groupe d'experts intergouvernemental sur l'évolution du climat). Si on prend l'exemple de l'étude des effets des activités humaines sur le climat, les questions et les besoins ont fortement évolué ces dernières années : nous sommes passés d'une phase d'alerte où il s'agissait avant tout d'estimer s'il y avait ou non un risque de changement climatique de grande amplitude du fait des activités humaines à une phase où il s'agit de fournir des informations pertinentes pour anticiper et prévenir les effets de ces changements, certains d'entre eux étant considérés comme désormais inévitables. Le niveau d'exigence sur le type de résultat et leur fiabilité est alors beaucoup plus élevé, ce qui nécessite de répondre notamment aux questions suivantes :

- être plus quantitatif, plus précis ;
- mieux décrire la distribution géographique et la structure temporelle de ces changements ;
- déterminer à partir de quand un phénomène climatique particulier est dû aux activités humaines et non à la variabilité naturelle du climat ;
- donner des informations qui permettent d'alimenter des questions au-delà de celles climatiques.

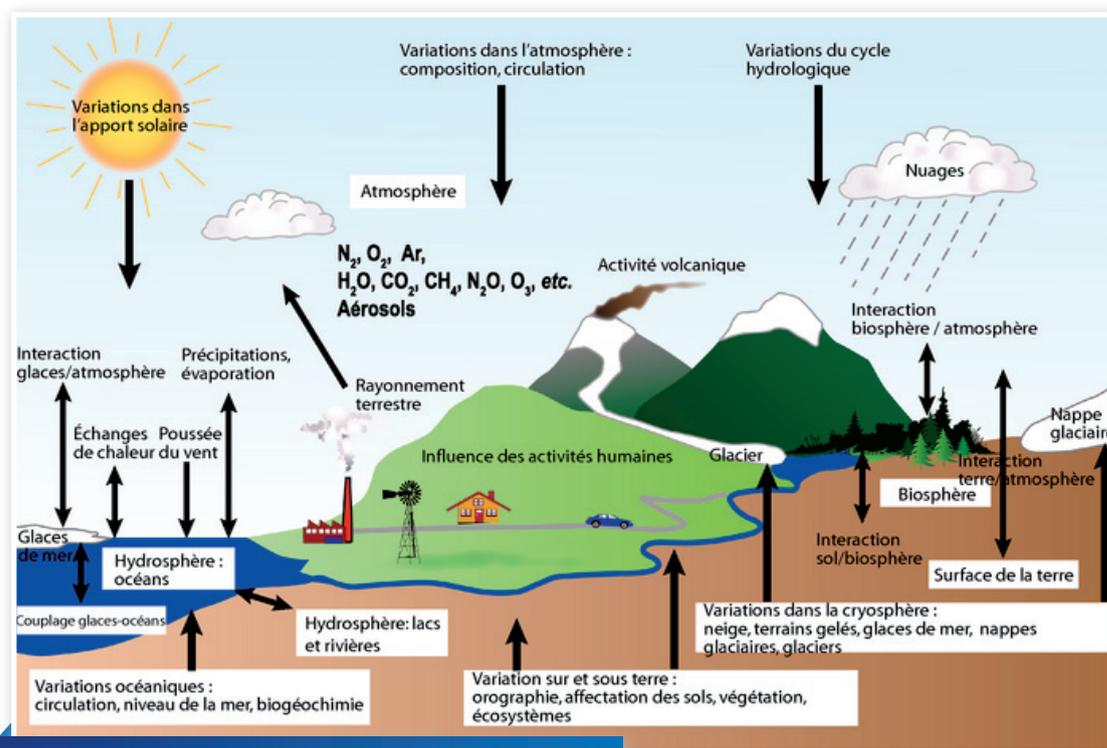


Figure 1 – Représentation schématique du système climatique

Accroissement de la résolution spatiale et temporelle

Les processus qui doivent être pris en compte pour simuler les phénomènes atmosphériques et océaniques s'étalent sur plusieurs ordres de grandeurs en termes d'échelle spatiale et temporelle (cf. Fig. 2 pour l'atmosphère). Dans les années 1980, le maillage avait typiquement une cinquantaine de points en latitude et longitude (typiquement 500 à 1000 km de côté) et une dizaine de mailles sur la verticale. Ceci permettait de simuler les caractéristiques très générales de la circulation atmosphérique et océanique à grande échelle. L'utilisation de mailles de quelques centaines de kilomètres de côté dans les années 90 a permis de bien mieux résoudre les gros tourbillons atmosphériques tels que les dépressions des moyennes latitudes (nos tempêtes hivernales). Un maillage de quelques dizaines de kilomètres permet de commencer à résoudre les gros tourbillons dans l'océan et les dépressions de plus petites tailles telles que les cyclones tropicaux dans l'atmosphère, mais permet surtout de mieux prendre en compte les effets du relief. Par exemple, simuler un phénomène de type mistral méditerranéenne nécessite de reproduire la topographie de la vallée du Rhône, c'est-à-dire d'avoir des résolutions inférieures à 50 km. Au-delà de l'état moyen, cet accroissement de résolution améliore la dynamique temporelle des phénomènes, la distribution géographique dans les régions avec du relief, les caractéristiques

des événements météorologiques et notamment des événements extrêmes. La résolution horizontale actuelle des modèles atmosphériques est typiquement de 50-150 km pour les modèles de climat et de 20-100 km pour les modèles de prévisions météorologiques. Dans l'océan, la dimension horizontale des tourbillons étant plus petite que dans l'atmosphère, la topographie joue un rôle encore plus important, ce qui fait que leur résolution horizontale est généralement plus élevée que pour l'atmosphère.

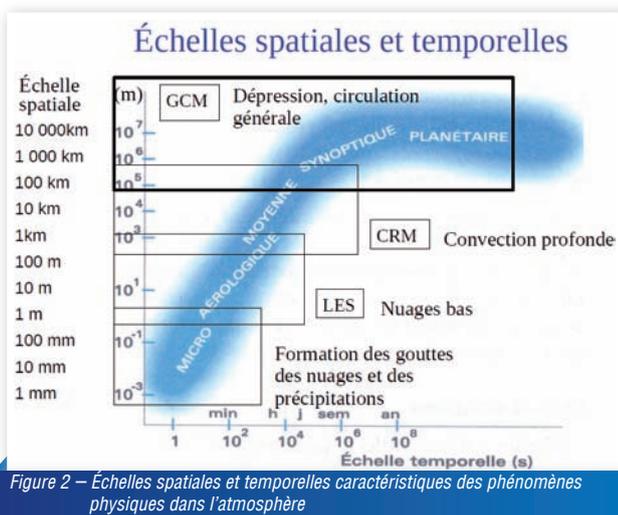


Figure 2 – Échelles spatiales et temporelles caractéristiques des phénomènes physiques dans l'atmosphère

Une meilleure représentation des processus

Les résultats montrent que ce raffinement de maillage permet de réduire d'importants biais des modèles, mais pas tous. En effet, on est encore très loin d'avoir des maillages assez fins pour résoudre toutes les échelles (Fig. 2) et il faudra probablement encore attendre quelques dizaines d'années pour que de telles simulations soient possibles sur de longues périodes. Ainsi, il demeurera nécessaire pendant encore longtemps de modéliser ces phénomènes sous-maillages (c'est-à-dire non explicitement résolus) et de les coupler au noyau dynamique.

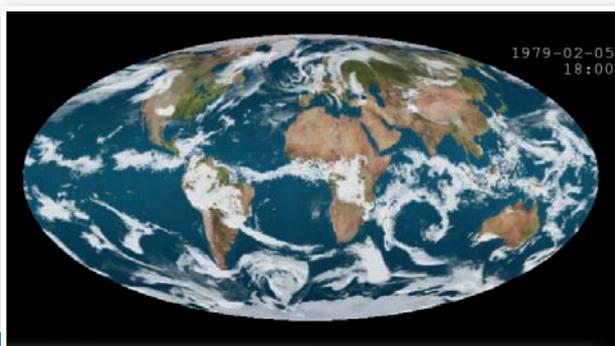


Figure 3 – Représentation des précipitations (en blanc) simulées par le modèle atmosphérique LMDZ à un instant donné du mois. On distingue les grandes étendues de pluies associées aux dépressions des moyennes latitudes, dans les deux hémisphères, ainsi que la bande de pluie de la convergence intertropicale, proche de l'équateur. Les modèles de climat simulent l'évolution de plusieurs centaines de variables, toutes les 30', et ce pendant des dizaines jusqu'à des milliers d'années.

Cette modélisation doit tout d'abord représenter la turbulence tridimensionnelle et les écoulements dont les dimensions caractéristiques sont inférieures à celles de la maille. Pour l'atmosphère, cette modélisation doit également représenter la condensation de la vapeur d'eau en gouttelettes d'eau ou de glace, leur organisation sous forme de nuages, la formation de précipitations éventuelles, etc. Ces phénomènes jouent un rôle clé pour les études climatiques et les prévisions météorologiques. Le développement de cette modélisation repose beaucoup sur des résultats de modèles numériques à très haute résolution mais ne couvrant qu'une surface très limitée de la terre.

Ce sont par exemple les simulations des grands tourbillons (*Large Eddy Simulation, LES*) avec un maillage de la dizaine de mètres ou les « modèles d'orages » (*Cloud Resolving Models, CRM*) avec un maillage de la centaine de mètres ou du kilomètre et qui peuvent résoudre les circulations méso-échelles associées aux nuages convectifs. La simulation numérique directe (*Direct Numerical Simulation, DNS*) ne permet de couvrir que des échelles très limitées. Jusqu'à peu, les échelles spatiales de ces trois familles de modèles, GCM, CRM et LES, étaient disjointes. Grâce à l'accroissement des puissances de calcul, les configurations d'étude de ces modèles commencent à se recouvrir, ce qui facilite l'utilisation de modèles à très haute résolution pour développer la modélisation des phénomènes sous-maillages des modèles à résolution plus faible.

Au-delà des processus physiques, les modèles climatiques peuvent prendre en compte les phénomènes chimiques pour le calcul de la composition chimique de l'atmosphère et des aérosols, petites particules en suspension dans l'air qui interagissent avec le rayonnement et peuvent modifier la formation des gouttelettes des nuages. Certains incluent également des cycles biogéochimiques (cycle du carbone, de l'azote, etc.) avec notamment une modélisation simplifiée du monde vivant (végétation, plancton, etc.). Tous ces développements sont par exemple inclus dans le modèle intégré du climat de l'IPSL qui tourne sur les calculateurs de l'IDRIS et du TGCC.

Des simulations plus nombreuses et plus longues

Le système climatique est un système chaotique, avec une très forte sensibilité à l'état initial et sans état d'équilibre stable. Même sans aucune variation des conditions aux limites, l'état de l'atmosphère et de l'océan fluctue d'une année à l'autre, d'une décennie à l'autre, etc. Étudier le climat consiste à étudier les caractéristiques statiques des variables des différents milieux (atmosphère, océan, continent, cryosphère) et nécessite de réaliser, selon les cas, des ensembles de simulations avec des états initiaux différents ou des simulations très longues.

Ces études de variabilité du climat sont aujourd'hui complétées par des études de prévisibilité aux échelles saisonnières à pluriannuelles.

Plus on s'intéresse à des périodes longues (supérieures à quelques dizaines d'années), plus ces études de variabilité ont recours à des modèles du fait du peu d'observations précises sur de longues périodes et de la présence de perturbations extérieures, naturelles (volcans, ensoleillement) ou dues aux activités humaines. Depuis de nombreuses années, on étudie les climats des passés lointains (paléoclimat) dont les caractéristiques sont très différentes de celles du climat d'aujourd'hui, et plus récemment on étudie également l'évolution du climat au cours du dernier millénaire, avec un intérêt particulier pour l'optimum climatique médiéval et le « petit âge glaciaire » qui lui succède. Dans un futur proche, on espère pouvoir réaliser des simulations encore plus longues et notamment étudier des périodes critiques comme la sortie de la dernière glaciation, entre 22 000 et 15 000 ans avant la période actuelle.

Des simulations plus proches des observations

Des développements méthodologiques, l'accroissement du réalisme des modèles, la plus grande qualité et variété des observations permettent de renforcer considérablement les liens entre modélisation et observation. Pour permettre une comparaison quantitative, on utilise de plus en plus souvent des simulateurs d'observations qui font le lien direct entre variables du modèle de climat et variables observées. Un premier exemple est la distribution verticale des nuages qui est forcément incomplète lorsqu'elle est mesurée par des méthodes optiques depuis les satellites, puisque les nuages hauts cachent les nuages bas. Le simulateur d'observations permet de simuler ce que mesurerait un satellite s'il observait une Terre dont les caractéristiques sont celles correspondant aux résultats du modèle. Dans un domaine très différent, les indicateurs pour reconstituer les climats anciens dépendent de plusieurs variables climatiques et la modélisation de ces indicateurs (sédiments, bulles de gaz, pollens, etc.) permet une meilleure interprétation des mesures. Enfin, des techniques

issues de la prévision météorologique sont couramment utilisées pour les études climatiques (assimilation de données, simulations d'ensembles, etc.), de même que les résultats des prévisions elles-mêmes, par exemple pour analyser les résultats de sites instrumentés ou de campagnes d'observations intensives.

L'important rôle des projets d'intercomparaison de modèles

De nombreuses questions ouvertes demeurent, tant sur la façon de modéliser les phénomènes qui ne sont pas explicitement résolus à l'échelle des mailles du modèle que sur certains phénomènes encore mal connus, et les groupes qui développent les modèles dans les grands organismes de recherche font des choix de modélisation différents. On a pu montrer que les principaux désaccords entre les résultats des modèles de climat proviennent de ces choix, sans qu'il soit néanmoins possible de faire un lien direct entre les résultats des modèles et une hypothèse particulière de modélisation, du fait des interactions non linéaires multiples entre les différents processus. Ainsi a émergé toute une activité sur l'analyse comparée des résultats issus de simulations réalisées avec des modèles climatiques différents mais avec un même protocole expérimental. Par exemple, tous les modèles sont utilisés pour faire une simulation de contrôle avec des conditions aux limites (ou « forçages ») identiques et constantes dans le temps (constante solaire, concentration des gaz à effet de serre, etc.). Cette simulation sert ensuite de référence pour des expériences de sensibilité comme celle idéalisée où la concentration de CO₂ augmente de 1 % par an. Les analyses couvrent des objectifs très divers comme par exemple d'évaluer les modèles, d'identifier les raisons de leurs comportements différents ou au contraire similaires, d'étudier la variabilité interne du climat (c'est-à-dire à conditions aux limites fixes) ou la réponse à certaines perturbations, etc. Le plus connu et le plus emblématique de ces projets est CMIP (*Coupled Model Intercomparison Project*) dont les différentes phases ont produit de très grandes quantités de données et de très nombreux résultats scientifiques, notamment ceux synthétisés dans les différents rapports du GIEC.

Des collaborations très fortes, à différents niveaux

Le développement des modèles climatiques, la réalisation des simulations et la mise à disposition des résultats nécessitent des moyens importants, aussi bien informatiques qu'humains. Si on prend l'exemple du modèle de l'IPSL, il est constitué de six modèles, d'un « coupleur » (Fig. 4), d'une infrastructure et d'outils logiciels dédiés à la gestion de ces modèles, à leur assemblage, à la réalisation des simulations et à la mise à disposition de leurs résultats.

Le développement de chacun de ces modèles, du modèle couplé et de l'infrastructure logicielle est réalisé par des groupes généralement constitués de 10 à 20 personnes. Pour l'IPSL, ce sont au total 70 à 80 personnes qui travaillent sur ces développements, avec un équilibre important à maintenir entre chercheurs et ingénieurs. Tous ces développements sont réalisés par des équipes de plusieurs laboratoires qui collaborent très étroitement entre elles

dans le cadre du centre de modélisation du climat de l'IPSL. Ces collaborations s'étendent au niveau national et international, et portent sur le co-développement de modèles, de logiciels ou sur le partage de certains d'entre eux. Des standards sont adoptés, comme celui d'un format auto-documenté des données qui permet de partager de nombreux outils de traitement, d'analyse ou de visualisation. La réalisation d'expériences numériques coordonnées mentionnées précédemment a été accompagnée du développement par plusieurs partenaires d'un système et d'une infrastructure distribuée internationale pour l'échange des données (ESGF, *Earth System Grid Federation*). Dans sa première phase (2010-2013), cette infrastructure a permis de mettre à disposition 2 pétaoctets de données réparties en 4,5 millions de fichiers sur une quarantaine de sites de distribution à travers le monde et utilisées par plusieurs milliers de personnes. Elle est maintenant adoptée de plus en plus largement et le volume et le nombre de données mis à disposition continuent de croître.

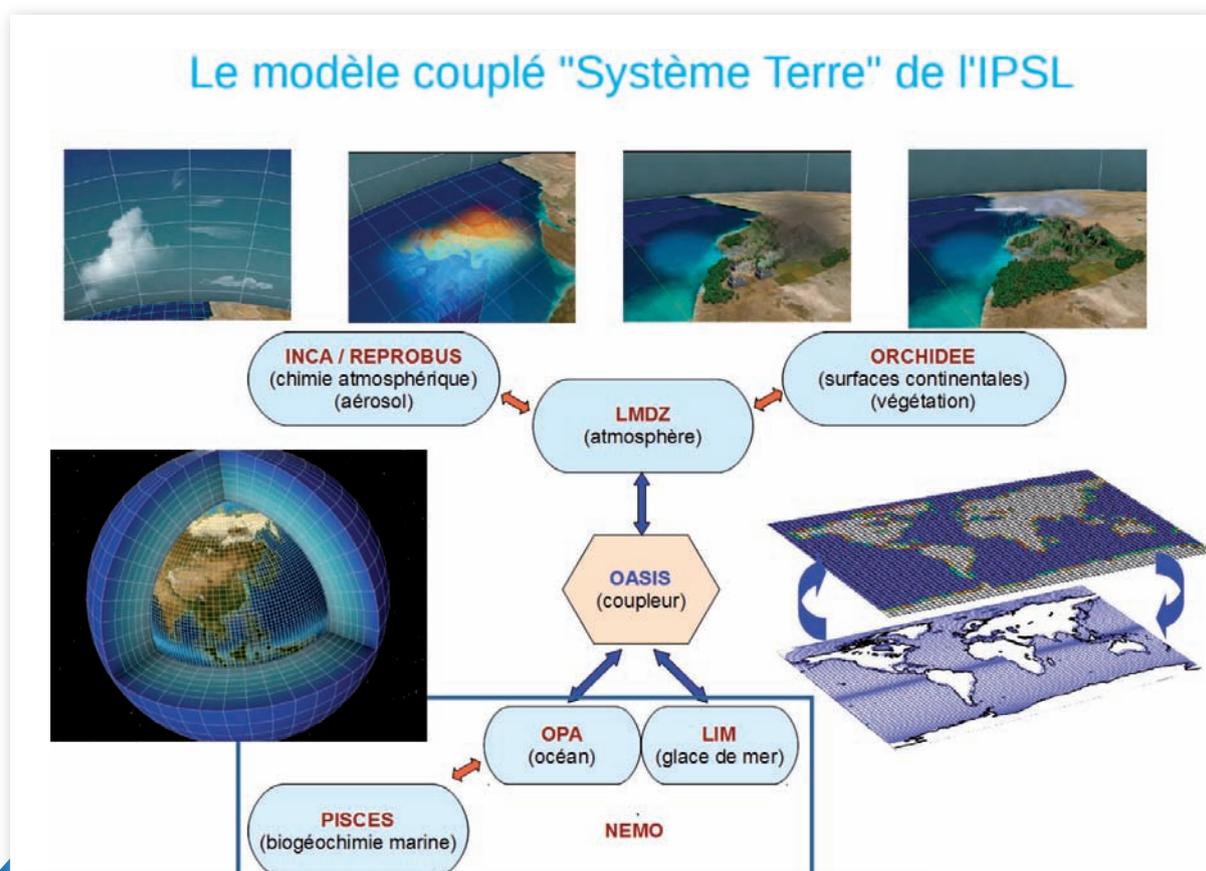


Figure 4 – Le modèle de climat, tel celui développé par l'IPSL, est constitué de plusieurs « sous-modèles » afin de prendre en compte de façon cohérente les processus physiques, chimiques et biogéochimiques des différents milieux (atmosphère, océan, continents, glace de mer, ...) qui interagissent entre eux.

D'importants moyens informatiques pour une communauté très large

L'utilisation des moyens de calcul pour la recherche sur le climat peut se décomposer schématiquement en deux parties à peu près égales : la réalisation de simulations dans le cadre de projets coordonnés, comme mentionné ci-dessus, d'une part, et la réalisation de simulations hors de ce cadre pour des études particulières, d'autre part. Ces dernières portent aussi bien sur le développement et la mise au point des modèles que sur l'utilisation de ces modèles pour étudier le système climatique. La réalisation de simulations coordonnées se fait de façon régulière et avec des calendriers précis. Par exemple, la réalisation des simulations pour le projet CMIP mentionné précédemment s'effectue sur environ trois ans tous les 6 ans. Ceci nécessite d'importants moyens dédiés pour pouvoir être menée à bien, aussi bien en termes de moyens de calcul que de stockage des données. Entre 2009 et 2012, un ordinateur vectoriel NEC SX-9 de 3 nœuds de 16 processeurs chacun a été entièrement dédié à la contribution de l'IPSL à CMIP5. Financé par GENCI, il a été géré par le CCRT. Il représentait alors environ 6 % de l'ensemble¹ des ressources disponibles sur les calculateurs GENCI en 2009 (Brodie, Vargas et Babel à l'IDRIS, Titane et Mercure au CCRT, Jade au CINES). Pour CMIP6, nous projetons d'utiliser pendant 3 ans l'équivalent de 12 000 cœurs de calcul Curie nœuds fins, soit 13 % de Curie. Cela représente environ 5 % de l'ensemble¹ des ressources disponibles sur les calculateurs GENCI en 2015 (Ada et Turing à l'IDRIS,

Curie au TGCC, Occigen au CINES). Concernant le volume de fichiers créés, pour CMIP5 nous avons produit 2 Po de résultats et nous prévoyons 14 Po de résultats pour le projet CMIP6.

Ces simulations intéressent en premier lieu les membres de l'IPSL et on estime que 200 à 300 d'entre eux utilisent ces résultats. Au niveau national et international, les données sont distribuées via le réseau ESGF et sont extrêmement utilisées. Les 250 To de données provenant des simulations réalisées avec les trois versions du modèle de l'IPSL pour le projet CMIP5 ont été téléchargées par plusieurs centaines d'utilisateurs différents, avec un débit moyen sur 2 ans de 30 To de données par mois, ou encore 30 000 fichiers/mois. Chacune de ces trois versions a été utilisée par plus de 300 publications dans les revues scientifiques. Au-delà de la communauté des climatologues, ces données sont utilisées par d'autres communautés scientifiques (agriculture, santé, énergie, etc.) et au-delà du monde académique dans le cadre de l'émergence de « services climatiques ».

La recherche sur le climat repose depuis 50 ans sur l'utilisation de moyens informatiques massifs. Cette tendance va se prolonger dans les prochaines années pour les raisons rapidement évoquées dans ce texte et l'accès à des calculateurs, des moyens de stockage et d'analyse des données à la pointe de ce qui existe demeurera une condition nécessaire pour pouvoir continuer à apporter des contributions significatives à la recherche sur le climat.

¹Les architectures et les caractéristiques de ces machines étant très différentes, les chiffres donnés ici le sont avant tout à titre illustratif.

En savoir plus :

- Le climat à découvert, CNRS Éditions, sous la direction de Catherine Jeandel et Rémy Mosseri, ISBN 978-2-271-07198-9
- Numéro spécial *Climat* de la revue *La météorologie*, N°88, février 2015, <http://documents.irevues.inist.fr/handle/2042/56350>
- Portail du projet CMIP: <http://www.wcrp-climate.org/index.php/wgcm-cmip/about-cmip>
- Le cinquième rapport du GIEC : http://www.ipcc.ch/home_languages_main_french.shtml
- Peter Lynch (2008), *The origins of computer weather prediction and climate modeling*, *Journal of Computational Physics*, 227(7), pp. 3431-3444, doi:10.1016/j.jcp.2007.02.03

CONSOMMATION DES HEURES DE CALCUL EN 2014 ET ATTRIBUTIONS EN 2015

L'année 2014 a été marquée par l'installation fin novembre d'une extension de deux cabinets supplémentaires pour Turing, augmentant ainsi ses capacités de 50 %.



Voici le bilan des heures consommées en 2014 ainsi que celui des heures attribuées pour la première campagne DARI de l'année 2015 sur Turing et Ada.

Ada

Sa disponibilité (hors maintenances programmées) en 2014 a été de 99,6%. Plus de 65 millions d'heures de calcul ont été consommées par 259 projets sur les 74 millions allouées.

Bilan des heures consommées en 2014

Turing

La disponibilité de Turing en 2014 (en dehors des maintenances programmées) a été de 99,5 %. Au total, 426 millions d'heures auront été utilisées par 127 projets des différents comités thématiques sur les 470 millions qui avaient été attribuées.

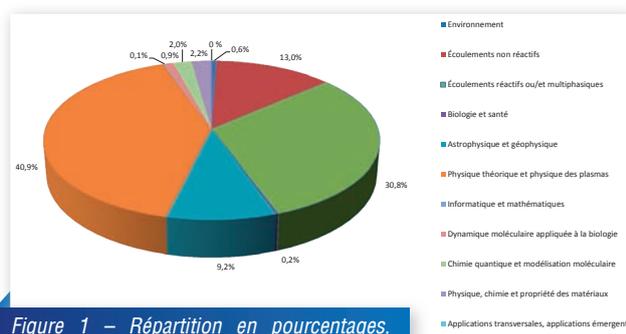


Figure 1 – Répartition en pourcentages, par comités thématiques, des heures de calcul consommées en 2014 sur Turing, IBM Blue Gene/Q.

Comme on peut le constater sur la figure 1, les comités thématiques les plus représentés ont été le CT2b (écoulements réactifs ou/et multiphasiques), le CT5 (physique théorique et physique des plasmas) et le CT2a (écoulements non réactifs).

On peut constater sur la figure 2 que les comités dont les projets ont été les plus consommateurs sont le CT2b (écoulements réactifs ou/et multiphasiques), le CT8 (chimie quantique et modélisation moléculaire), le CT1 (environnement) et le CT2a (écoulements non réactifs).

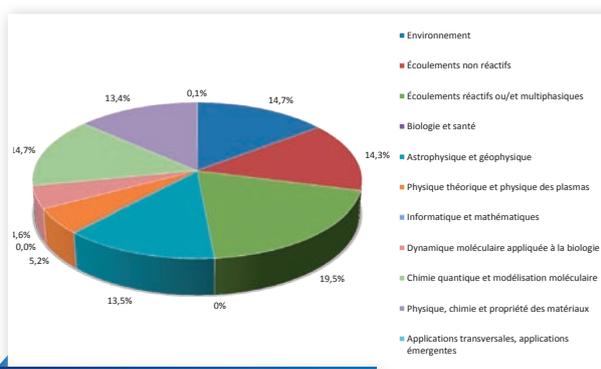


Figure 2 – Répartition en pourcentages, par comités thématiques, des heures de calcul consommées en 2014 sur Ada.

De plus, ce sont 28 dossiers d'accès préparatoires (16 sur Ada et 12 sur Turing) qui ont été acceptés en 2014, allouant des ressources (respectivement 15 000 h sur Ada et 50 000 h sur Turing) pour favoriser le portage et l'optimisation des applications sur ces calculateurs.

Bilan des actions de formation 2014

Les activités de formation sont une composante essentielle de notre rôle de pôle de compétences en calcul de haute performance.

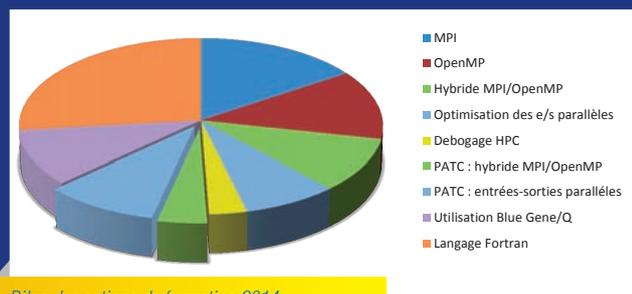
Une dizaine d'ingénieurs de l'équipe Support aux utilisateurs (SU) donnent des formations principalement orientées vers les méthodes de programmation parallèles (MPI, OpenMP, programmation hybride MPI/OpenMP) et les langages de programmation scientifiques généraux (Fortran 95, 2003 et C). Ces formations dispensées à l'IDRIS ou à l'extérieur sont accessibles à toute la communauté académique et industrielle utilisatrice du calcul intensif.

Tous les supports de cours IDRIS et les travaux pratiques associés sont disponibles librement par téléchargement depuis le site de l'IDRIS : www.idris.fr/formations/supports_de_cours.html

L'IDRIS participe également aux formations PATC (*Prace Advanced Training Center*) françaises coordonnées par la Maison de la simulation. En 2014, les ingénieurs de l'IDRIS ont ainsi animé 2 sessions de formation PATC concernant la programmation hybride et les entrées-sorties parallèles.

Fin juin 2014, l'IDRIS a organisé, en collaboration avec IBM, une session de portage et d'optimisation de codes sur Blue Gene/Q. Cette session a associé des experts de l'IDRIS et d'IBM qui ont travaillé directement avec les développeurs des applications présents pendant ces 3 jours. Devant le succès rencontré par ce *workshop*, il a été décidé de l'organiser de manière régulière et de l'inscrire au catalogue des formations données à l'IDRIS.

En 2014, l'IDRIS a dispensé au total 20 sessions de formation (dont 6 à l'extérieur) qui ont été suivies par 314 personnes (227 à l'IDRIS, 87 à l'extérieur) pour un total de 60 journées de formation. Tous ces chiffres sont en légère hausse par rapport à ceux de l'année 2013. Il faut noter aussi que 28 stagiaires payants s'étaient inscrits via CNRS Formation Entreprises.



Bilan des actions de formation 2014 (en nombre d'inscrits)

Bilan des heures de calcul attribuées lors de la première session DARI 2015

La mise en exploitation de l'extension de Turing a permis une augmentation substantielle du nombre d'heures mises à la disposition des chercheurs lors de cette première session de la campagne 2015. En effet, ce sont 524 millions d'heures qui ont été attribuées sur Turing à 110 projets des différents comités thématiques.

Concernant Ada, ce sont plus de 64 millions d'heures qui ont été attribuées à 237 projets.

La répartition des heures de calculs attribuées par comités thématiques lors de cette 1^{re} session est résumée dans les figures 3 et 4.

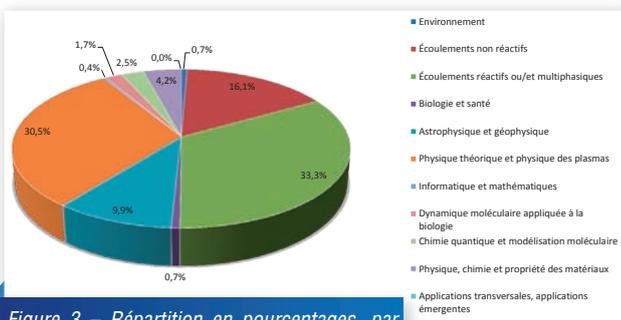


Figure 3 – Répartition en pourcentages, par comités thématiques, des heures de calcul attribuées lors de la 1^{re} session de la campagne DARI 2015 sur Turing.

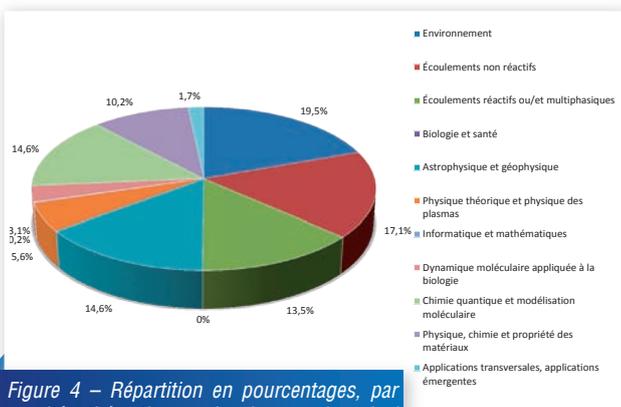


Figure 4 – Répartition en pourcentages, par comités thématiques, des heures de calcul attribuées lors de la 1^{re} session de la campagne DARI 2015 sur Ada.

Extension de la configuration Blue Gene/Q

Ainsi qu'annoncé à la fin de l'année 2014, deux cabinets supplémentaires ont été ajoutés au supercalculateur Turing après que la société GENCI (Grand équipement national de calcul intensif) et le CNRS aient pris la décision d'étendre la configuration de la machine IBM Blue Gene/Q.

Ces deux nouveaux cabinets ont été installés en seulement un mois et ils ont été mis en production le 28 novembre 2014, portant cette configuration à quasiment 100 000 cœurs de calcul, pour une capacité mémoire de près de 100 To et une performance crête de 1,26 PFlop/s.



Les deux nouveaux cabinets en cours d'installation

Ainsi, ce sont désormais 655 Mh d'heures qui seront chaque année mises à la disposition des communautés scientifiques. Cette extension permet de positionner présentement Turing au 42^e rang mondial du classement des supercalculateurs.

Cette configuration, en plus du surcroît de ressources qu'elle offre aux utilisateurs, permet à l'IDRIS de conforter sa position dans le domaine du calcul massivement parallèle, clé pour l'avenir du calcul intensif et offre également un tremplin pour l'accès aux configurations Blue Gene/Q des partenaires européens dans le cadre des appels à projets PRACE.



La nouvelle configuration Blue Gene/Q

Migration Gaya vers Ergon

La migration des fichiers entre le serveur de fichiers Gaya et le serveur d'archives Ergon s'est déroulée du 1^{er} juillet à la fin novembre 2014 pour les données des utilisateurs.

Cela a représenté un volume de 3,1 Po pour environ 23 millions de fichiers. Tout le reste des données (fichiers IDRIS) a été migré au plus tard mi-janvier 2015. Pendant cette période, les utilisateurs ont généré 12 millions de fichiers supplémentaires représentant 1,4 Po de volume sur l'espace du serveur d'archives Ergon.



Le nouveau serveur d'archives, Ergon



Le robot de stockage StorageTek

Remise de la médaille de cristal

Philippe Wautelet, lauréat 2014 de la médaille de cristal du CNRS (voir www.cnrs.fr/fr/recherche/prix/cristal/2014.htm#wautelet), a reçu sa distinction lors d'une cérémonie au siège de la délégation régionale CNRS/IDF Sud, le 14 novembre dernier, en présence de madame Véronique Debisschop déléguée régionale, monsieur Michel Bidoit directeur de l'institut des sciences de l'information et de leurs interactions (INS2I), madame Catherine Rivière présidente du Grand équipement national de calcul intensif (GENCI) et de monsieur Denis Girou directeur de l'IDRIS. Philippe Wautelet est spécialiste du calcul massivement parallèle et à ce titre il est depuis 2008 responsable du support applicatif des machines Blue Gene de l'IDRIS (voir son interview dans le numéro de juin 2014 de la Lettre de l'IDRIS www.idris.fr/media/data/lalettre/la-lettre-de-l-idris-juin-2014.pdf).



Remise de la médaille de cristal

Conférence Supercomputing 2014

La conférence Supercomputing est l'événement annuel majeur du calcul intensif et du traitement des grandes masses de données. L'édition 2014 s'est tenue à la Nouvelle Orléans du 16 au 21 novembre. Pour la première fois cette année, le CNRS présentait sur son stand l'ensemble de ses activités dans ces domaines à travers la présentation de l'IDRIS et du centre de calcul de l'IN2P3 (CC-IN2P3), les travaux et ressources de ces deux centres étant complémentaires pour le CNRS puisqu'ils se concentrent respectivement sur le calcul haute performance et sur la gestion et le traitement des données en particulier dans les domaines de la physique nucléaire, de la physique des particules et des astroparticules. Voir le communiqué du CNRS à l'adresse : www.cnrs.fr/ins2i/spip.php?article1005



Supercomputing 2014

Activités de l'IDRIS dans le projet européen PRACE en 2014

La mise en place de l'infrastructure européenne de recherche pour le calcul de haute performance PRACE (*Partnership for Advanced Computing in Europe*) a débuté en 2008. Actuellement, alors que 3IP se termine, 4IP va très prochainement démarrer. Courant 2015, la seconde phase du projet PRACE, dite PRACE 2.0, va se mettre en place pour la période 2015 à 2020. Dans 3IP, deux groupes de travail étaient dédiés à l'accès des utilisateurs aux ressources de calcul sur les différents supercalculateurs intégrés dans cette infrastructure, tant Tier-0 que Tier-1, en leur proposant un support technique pour le portage et l'optimisation des applications sélectionnées lors des appels à projets.

L'IDRIS s'est fortement impliqué dans ce support applicatif, grâce à l'expertise de ses ingénieurs. De plus, un important travail de rédaction d'un manuel *Best Practice Guide* dédié aux utilisateurs des machines Blue Gene/Q a été rédigé, conduit par un membre de l'IDRIS. En collaboration avec l'ensemble des partenaires français, coordonnés par la Maison de la Simulation, l'IDRIS s'est aussi activement impliqué dans les formations proposées dans le cadre des PRACE *Advanced Training Centers*. Par ailleurs, l'environnement commun à toutes les plates-formes PRACE est continuellement enrichi, concernant l'accès aux logiciels et bibliothèques, les transferts de fichiers à haute performance, la comptabilité des projets, la gestion des comptes et les accès sécurisés, etc.

Bilan d'activité du projet E-Biothon

Début 2014, le CNRS, IBM, l'INRIA, l'Institut français de bioinformatique et la société SysFera déploient E-Biothon, une plateforme Cloud expérimentale pour accélérer et faire progresser la recherche en biologie, santé et environnement (www2.cnrs.fr/presse/communiqu/3313.htm). L'IDRIS assure, en partenariat avec l'INRIA, le support applicatif. La plateforme, basée sur une partie de notre ancienne configuration Blue Gene/P, a permis l'exécution des premiers projets :

- Projet Synthénie de l'INRA Jouy-en-Josas : 20 millions de travaux ont été exécutés sur 8 mois en 2014, le projet étant terminé mais pouvant être prolongé dans une seconde phase.
- Projet de phylogénie PhyML du LIRMM à Montpellier : 80 utilisateurs du monde entier ont pour le moment demandé à bénéficier de la plateforme, leur consommation ayant représenté plus d'1 Mh sur les trois premiers mois de l'année 2015.
- Projet en Barcoding Disseq de l'INRA Bordeaux : le code initialement séquentiel a été parallélisé avec MPI et il vient d'entrer en production.

Suite à l'appel à propositions lancé dans la communauté bioinformatique fin 2014, 14 nouveaux projets se sont portés candidats. Parmi eux, 7 ont été retenus et devraient s'exécuter dans le courant de l'année :

- 3 projets basés sur le logiciel NAMD et 1 projet basé sur le logiciel LAMMPS, dont les intégrations dans le portail SysFera-DS sont en cours par cette société.
- 3 projets mettant en œuvre des codes utilisateurs qui sont en cours de portage.

ACTUALITÉS

Session Blue Gene/Q des 21-23 janvier 2015

Après celle de juin dernier, l'IDRIS a organisé une deuxième session de portage et d'optimisation de codes sur Blue Gene/Q sur trois jours, avec le concours d'experts d'IBM et des ingénieurs applicatifs de l'IDRIS.

Cette session a réuni les utilisateurs de 7 projets dans des domaines variés appartenant à 5 comités thématiques (écoulements non réactifs ; biologie et santé ; informatique, algorithmique et mathématiques ; physique théorique et physique des plasmas ; physique, chimie et propriétés des matériaux). Chaque projet a bénéficié du support d'une ou de deux personnes de l'assistance de l'IDRIS. Les projets concernés représentaient environ 40 millions d'heures attribuées sur la machine Turing lors de la première session DARI de 2015. Devant le succès de ces journées, l'IDRIS va organiser une nouvelle session début juillet (voir le site web de l'IDRIS www.idris.fr pour plus d'informations) et rendre ces formations semestrielles.



Tutoriels CECAM à l'IDRIS

Après une première session d'échanges accueillie dans nos locaux les 12 et 13 mars derniers, le tutoriel CECAM *deMon2k and deMon-nano* se déroulera lui aussi à l'IDRIS du 15 au 19 juin 2015. Pour plus d'informations consulter la page : www.cecam.org/workshop-1139.html

Formation Débogage HPC

Une formation Débogage HPC, organisée conjointement par la Maison de la simulation et l'IDRIS, aura lieu du 8 au 9 octobre 2015 à l'IDRIS. Pour plus d'informations consulter la page : www.idris.fr/annonces/formation-debogage-hpc-8-9oct2015.html

CALENDRIER DES FORMATIONS IDRIS PROGRAMMÉES AUX 2^e et 3^e TRIMESTRES 2015

TITRE DE LA FORMATION	DATE DE DÉBUT	DURÉE
Fortran base	20/05/2015	3 jours
Langage C	01/06/2015	5 jours
MPI/OpenMP	08/06/2015	5 jours
Fortran avancé	23/06/2015	4 jours
MPI	21/09/2015	4 jours
Fortran base	29/09/2015	3 jours
OpenMP	13/10/2015	2 jours

Ces dates vous sont communiquées à titre d'information et sont susceptibles d'être mises à jour. Pour une information récente et plus complète, n'hésitez pas à consulter le serveur Web des cours de l'IDRIS : <https://cours.idris.fr>. Vous y trouverez le catalogue complet des formations et pourrez vous préinscrire aux sessions annoncées.

Nous vous rappelons que les formations IDRIS sont gratuites pour les personnes appartenant au CNRS ou à une université. Elles sont aussi accessibles au personnel d'entreprises publiques ou privées via CNRS Formation Entreprises : les conditions d'inscription sont alors consultables sur le site Web : cnrsformation.cnrs.fr

cnrs formation
entreprises