



Stratégie calcul en modélisation du climat

Mars 2007

Synthèse

Document rédigé par le Pôle de modélisation de l'IPSL

Cette note présente l'état d'avancement de la modélisation du climat et ses besoins en moyens de calcul intensif sur les machines du CNRS et du CEA pour les années à venir, dans le contexte de la préparation du prochain rapport du GIEC¹. Elle fait le point sur les capacités présentes et futures d'utilisation des différentes architectures de calculateur. Elle est constituée de cette synthèse et d'un document détaillé.

Les éléments s'appuient sur l'activité Pôle de modélisation de l'IPSL, qui rassemble 80 chercheurs et 20 ingénieurs de différentes institutions (CEA, CNRS, Ecole Polytechnique, ENS, IRD, UPMC, UVSQ, ...). L'autre groupe Français de modélisation climatique, à Météo France, dispose de moyens de calcul dédiés.

Le modèle de climat décrit dans ce document résulte du couplage de 7 modèles. Il sert de base pour dimensionner les besoins dans ce document car son développement, sa mise au point et son utilisation cumulent les difficultés de chacun des modèles.

Cette note montre que :

1. Pour les 4 prochaines années, nous avons besoin de continuité des moyens de calcul pour la réalisation des prochaines simulations pour le GIEC. Les ressources nécessaires pour réaliser ces futures simulations GIEC sont estimées à 80 processeurs de type NEC-SX8 sur une année. Cette puissance de calcul ne couvre que la moitié des besoins de la communauté climat. Les équipes françaises ne pourront rester au meilleur niveau international que si l'équivalent des ressources vectorielles actuelles de l'IDRIS et du CCRT sont dédiées à la communauté climat.
2. Nous développons une stratégie de parallélisation des modèles permettant d'être indépendants du type d'architecture des calculateurs. Ce parallélisme sera effectif sur le NEC-SX8 en 2007 sur une dizaine de processeurs. Pour utiliser des machines scalaires (~100 processeurs) nous devons résoudre de nombreuses difficultés concernant la gestion des fichiers de données et le traitement des résultats. Pour accélérer le portage, une cellule de 4 personnes supplémentaires à l'IPSL et des relais identifiés dans les centres de calcul sont nécessaires.
3. Les modèles de climat ont des limitations intrinsèques qui ne permettent pas d'utiliser de façon aisée des milliers de processeurs. Quel que soit le type de calculateur massivement parallèle envisagé dans le futur, il faut qu'il dispose de processeurs rapides, d'un réseau mémoire efficace, d'une mémoire suffisante par processeur, et de connexions rapides pour que nos applications soient performantes. Le stockage des données doit aussi être dimensionné en conséquence.
4. Le passage à la parallélisation massive (>1000 processeurs) demande de considérer de nouvelles grilles et de nouveaux types de modèles. Des prototypes existent en Europe, au Japon et aux Etats-Unis. Une dizaine d'années est encore nécessaire pour que ces modèles puissent être utilisés pour le climat. Pour relever ce défi, il faut prévoir la création d'un groupe de travail approprié dans les 2 années qui viennent.

¹ GIEC : Groupe Intergouvernemental d'Expert sur l'Évolution du Climat (<http://www.ipcc.ch>)

Enjeux scientifiques

La recherche sur le climat progresse rapidement. La question du changement climatique est au cœur du débat de société. Néanmoins, s'il est acquis que le climat se réchauffe sous l'action de l'activité humaine, de nombreuses incertitudes demeurent. Mieux comprendre le rôle des différents gaz à effet de serre, des aérosols, de l'agriculture et de la déforestation, ainsi que leur traduction sous forme d'événements climatiques extrêmes demande de développer la modélisation sur plusieurs fronts. Il faut tout à la fois améliorer le contenu physique des modèles, mieux représenter les phénomènes de petite échelle, comprendre les interactions complexes entre le climat et les cycles biogéochimiques en augmentant la complexité et les couplages entre les processus, étudier les incertitudes liées à la nature chaotique du système en réalisant des ensembles de simulations, et tester la possibilité d'événements peu probables, mais à haut risque pour la société.

Plan de route et stratégie calcul du Pôle de modélisation du climat de l'IPSL

C'est avec ces éléments à l'esprit que nous abordons dès à présent la préparation du prochain rapport du GIEC : l'AR5², qui implique la réalisation de simulations climatiques selon un protocole international. Les enjeux scientifiques se traduisent pour l'IPSL par le développement de deux nouvelles versions du modèle couplé qui incluent les progrès scientifiques des 5 dernières années. Le premier, basé sur le modèle utilisé pour l'AR4, permettra d'étudier les couplages entre le climat et les cycles biogéochimiques, tandis que le second incorporera les nouveaux développements physiques nécessaires pour traiter avec plus de précision les évolutions du cycle diurne et les événements extrêmes. La mise au point des modèles et les premiers tests sont déjà en cours. **Les versions définitives des modèles seront figées mi-2008** pour une réalisation des nouvelles simulations en 2009. L'exercice à réaliser pour l'AR5 représentera, en 2009, une puissance calcul équivalente à **80 processeurs NEC SX-8**, l'ensemble étant réservé pendant une année entière. Cette estimation ne concerne que **la moitié des besoins pour le climat**, car nous devons aussi pouvoir poursuivre en parallèle les simulations complémentaires qui permettent d'asseoir les résultats des projections climatiques.

Nous avons besoin de la continuité des moyens de calcul sur 4 ans. Conformément à la prospective établie au moment du renouvellement de la machine vectorielle de l'IDRIS, nous avons besoin de l'augmentation des ressources initialement prévues (facteur 10). L'estimation actuelle des besoins **correspond à l'utilisation pour la communauté climat de l'ensemble de la machine IDRIS actuellement disponible**, dont nous représentons la moitié des utilisateurs. Cette demande tient compte de la répartition de nos calculs sur les SX8 de l'IDRIS et du CCRT. Afin de mener le prochain exercice GIEC en tant que groupe leader au niveau international nous avons besoin de l'assurance de bénéficier de cette augmentation **dès mi-2008** pour réaliser les tests préliminaires. Cette estimation se base sur l'expérience acquise pour l'AR4. Le développement du modèle utilisé a demandé **4 ans de développement en amont**. Pour mener à bien ce travail, **nous avons aussi besoin de personnel** efficace, compétent et connaissant en détail nos applications et les calculateurs disponibles **dans les centres de calcul** de l'IDRIS (CNRS) et du CCRT (CEA).

Effort de parallélisation

Un important effort de parallélisation de nos codes a été mené ces dernières années. Les différentes composantes sont désormais toutes utilisables en **version parallèle à mémoire distribuée**. **Une version parallèle du modèle couplé sera opérationnelle sur les NEC SX8 courant 2007**. Le projet CICLE³ (ANR calcul intensif) a permis d'affecter des moyens supplémentaires à ce travail. Cette version bénéficie à la fois des avantages de la parallélisation en mémoire distribuée et de l'utilisation de la vectorisation. L'expertise acquise ces dernières années, y compris sur le « Earth Simulator » au Japon dans le cadre d'un accord de collaboration mis en place par le CNRS, nous permet d'affirmer

² AR5 : 5^{ème} rapport d'évaluation (Assessment Report) du GIEC, prévu pour 2012.

³ CICLE : <http://dods.ipsl.jussieu.fr/CICLE>.

que **nos modèles sont configurés pour utiliser de façon optimale les machines de type NEC-SX8** actuellement disponibles à l'IDRIS (CNRS) et au CCRT (CEA).

Le plan de route du pôle de modélisation du climat de l'IPSL inclut le développement d'une version utilisable sur différents types de calculateurs parallèles (scalaires et vectoriels). Nous nous préparons aussi à réaliser des simulations climatiques à très haute résolution sur un ordinateur de classe exceptionnelle qui pourrait voir le jour au niveau européen (projet HPC Europe). L'ensemble de ce travail est bien avancé, mais fait apparaître des difficultés de « scalabilité » et de performances. **Deux points sont particulièrement critiques.** D'une part, le **portage et l'optimisation du modèle couplé**, qui demande l'optimisation de 5 à 8 modèles suivant les configurations, est une étape lourde, car elle cumule l'ensemble des difficultés des différentes composantes auxquelles s'ajoutent les difficultés propres aux couplages. D'autre part, chaque machine exige un travail particulier **pour traiter les entrées/sorties** (optimisation, recombinaison des fichiers produits en parallèle, création des séries temporelles, atlas de résultats, ...), plus complexe encore sur des machines massivement parallèles dont l'environnement matériel et logiciel est encore peu standardisé. **Les délais pour la réalisation des simulations GIEC pour l'AR5 sont incompatibles avec la bascule sur une nouvelle machine massivement parallèle.**

Afin d'assurer la pérennité de nos applications, nous devons maintenir la portabilité de nos outils en généralisant les solutions optimales. **Les adaptations nécessaires ne doivent pas dépendre de la machine utilisée et doivent rester transparentes pour les utilisateurs.** Pour pouvoir utiliser une machine scalaire parallèle plus rapidement, il est nécessaire d'anticiper toutes les difficultés prévisibles et **du personnel supplémentaire** s'avère nécessaire. **Une cellule de 4 personnes à l'IPSL** et un support au coupleur OASIS permettrait de traiter plus efficacement la validation des développements et des versions successives des modèles, la gestion des entrées/sorties, le post-traitement efficace des résultats produits en masse, la diffusion des méthodes de travail et la formation des utilisateurs chercheurs et étudiants.

Contraintes spécifiques à la modélisation du climat

Les simulations climatiques ont aussi des limitations particulières. **Elles impliquent de choisir le bon compromis entre l'augmentation de la résolution spatiale et l'augmentation de la durée des simulations** et ce en fonction des ressources en calcul disponibles. Ainsi il est envisageable d'utiliser une très haute résolution pour une simulation météorologique à quelques jours, mais cette résolution devient prohibitive en temps de calcul si l'on doit simuler le climat pendant 300 ans. L'utilisation d'un plus grand nombre de processeurs s'avère cependant précieuse pour certaines études scientifiques (cas météorologiques, 'snapshots', océan à haute résolution). Nous n'envisageons d'aborder la très haute résolution pour le climat qu'avec la perspective d'une machine européenne dimensionnée en conséquence.

L'utilisation de plusieurs processeurs permet de réduire globalement les temps d'exécution à une résolution donnée, mais les grilles que nous utilisons et la représentation géographique intrinsèque de la Terre, qui oblige de filtrer l'écoulement aux pôles pour éviter les instabilités numériques, limitent les possibilités d'utiliser de façon optimale un grand nombre de processeurs. Nous sommes actuellement limités par le nombre minimal de points à garder par sous-domaine de calcul dans l'atmosphère. Ainsi, **à la résolution actuelle de 3,75°x2,5°x19 niveaux, le nombre maximal de processeurs est de 16.** Une simulation réalisable en 5 jours sur 4 processeurs de SX8 (3 jours sur 8 processeurs) ne l'est qu'en 6 à 13 jours sur 16 processeurs scalaires (selon le type de processeurs). A cette résolution l'utilisation d'une machine scalaire/parallèle est envisageable pour réaliser des ensembles de simulations au prix d'un important travail d'organisation. Il faut cependant garder à l'esprit que **le temps de restitution de chaque simulation sera plus long sur une machine scalaire, ce qui se traduit par une perte de réactivité et d'efficacité pour l'exploitation scientifique.**

L'augmentation de la résolution nous permet d'être plus 'scalable' en augmentant le nombre de points de calcul et de bénéficier d'une parallélisation plus massive. Ainsi les mêmes 100 ans de simulation à la haute résolution envisagée pour l'AR5 (1° de résolution en longitude et 40 niveaux verticaux) prendraient au mieux 96 jours sur 100 processeurs de type Itanium. Cette même simulation peut-être réali-

sée en 36 jours sur 56 processeurs SX8. L'écart de performances entre les deux types de machines est encore creusé à cette résolution. Quelque soit le calculateur parallèle envisagé, nos **applications ne peuvent être efficaces** que si la **machine cible dispose de processeurs rapides, d'un réseau mémoire efficace, d'une mémoire suffisante par processeur et de connexions rapides**. Dans l'état actuel du marché, les calculateurs parallèles-vectoriels sont les mieux adaptés à nos problèmes scientifiques, et présentent l'environnement de travail le plus éprouvé.

Solution pour avoir de plus hautes performances à l'échelle de 10 ans

Enfin, la solution pour lever les limitations intrinsèques de nos applications est de travailler sur des nouvelles grilles pour la partie dynamique des codes pour éviter les points singuliers. C'est un travail de longue haleine qui a démarré au Japon, aux Etats-Unis et en Allemagne. Les solutions actuelles sont au niveau de prototypes permettant de traiter les écoulements de l'atmosphère et de l'océan dans des situations simplifiées, mais qui ne permettent pas de considérer toute la physique du système et posent des problèmes de conservation de l'énergie, élément indispensable pour les simulations climatiques. Cette nouvelle génération de modèles ne sera pas en production avant 10 à 15 ans. Nous commençons à mener une réflexion pour développer de tels modèles. Il faut mettre **en place d'un groupe de travail pour aborder les modèles de la prochaine génération**, faisant appel à de nouvelles grilles, ce qui nécessite une approche multidisciplinaire alliant des compétences en mécanique de fluides géophysiques, physique, algorithmique et numérique. Un renforcement à court terme des effectifs pour aborder pleinement ces nouveaux aspects est nécessaire.

Conclusion

La modélisation du climat a besoin d'investissements suivis dans le long terme. Les simulations pour le 5^{ème} rapport du GIEC sont déjà en préparation et nécessitent une continuité des moyens de calcul sur les 4 prochaines années. Le portage des modèles sur des machines scalaires/parallèles nécessite d'assurer une bonne scalabilité et une gestion performante des entrées/sorties et des opérations de post-traitement. Nos modèles ne peuvent être très performants que sur des machines disposant de processeurs rapides, d'un réseau mémoire efficace, d'une mémoire suffisante par processeur, de connexions rapides. Nous avons besoin d'accroître les effectifs pour mener à bien le travail engagé, pouvoir basculer nos applications sur des calculateurs ayant différents types d'architecture parallèle et également assurer le développement scientifique des modèles et de leurs applications.

Les moyens humains nécessaires pour aborder le parallélisme massif peuvent donc se résumer ainsi :

- Dans les centres de calcul, besoin de continuité de services pendant les 4 prochaines années avec doublement vectoriel IDRIS mi-2008.
- A l'IPSL, aide pour la migration vers le scalaire/parallèle :
 - Validation des développements et des versions successives des modèles
 - Gestion des Entrées//Sorties
 - Post-traitement efficace des résultats produits en masse
 - Diffusion des méthodes de travail
 - Formation des utilisateurs chercheurs et étudiants
- Un support au coupleur OASIS.
- Mise en place d'une cellule de développement numérique pour la prochaine génération de modèles.

Notons que ces développements nécessitent aussi un suivi scientifique pointu, car **les modèles doivent répondre à des objectifs scientifiques**. Ce dernier point constitue actuellement **un goulot d'étranglement**. La montée en puissance de l'utilisation de nos codes sur des machines variées et le besoin croissant de mieux comprendre les différents couplages intervenant au sein du système climatique demande aussi d'envisager **le recrutement de chercheurs supplémentaires sur les thématiques ayant recours aux modèles couplés** pour répondre aux questions scientifiques.