D2.1b Comparaison des performances climatiques et computationelles du modèle LMDZ avec l'ancien noyau dynamique longitude-altitude et le nouveau noyau Dynamico

Frédérich Hourdin, Abdoul Khadre Traoré, Thomas Dubos, Yann Meurdesoif, Laurent Fairhead, Ehouarn Millourd

6 novembre 2018

1 Introduction

La mise au point d'un nouveau noyau Dynamique Dynamico pour la modélisation globale du climat, basé sur une discrétisation icosahédrique de la sphère, a été motivée en grande partie par des questions d'efficacité numérique. En effet, la grille basée sur ne discrétisation indépendante des longitudes et des latitudes, appelée grille lonXlat par la suite, possède dues singularités aux pôles. Du fait du rapprochement des méridiens au niveau du pôle, les mailles deviennent de plus petites en latitude. Sans procédure particulière, les critères CFL liés à l'advection ou à la propagation des ondes, contraindrait à utiliser des pas de temps ridiculement petit par rapport à la résolution spatiale effective plus bas en latitude. Depuis longtemps, il existe une solution à ce problème qui consiste à filtrer certaines quantités en longitude quand on s'approche des pôles, pour ne retenir que les échelles spatiales effectivement résolues à plus basse latitude.

Or cet opérateur de filtrage est global en longitude et consomme une partie significative (jusqu'à 30% suivant les configuration) du temps de calcul de la dynamique. De ce fait, il a été déraisonnable de baser la parallélisation du noyau dynamique traditionnel sur un découpage de domaine en longitude. Le découpage est donc fait en mémoire distribué sur les latitudes (avec au minimum deux bandes de latitude complète par processeur) en utilisant la libraire de passage de message MPI. Une parallélisation OpenMP du code est également réalisée sur la verticale, en mode mémoire partagée.

Le modèle de climat, dans sa partie atmosphérique, est constitué fondamentalement de deux parties : le noyau dynamique, et un jeu de paramétrisations physiques qui calcul les transferts verticaux associés avec les processus radiatifs, turbulents, convectifs ou nuageux. Cette partie du code est constituée informatiquement d'une juxtaposition de colonnes verticales 1D. La parallélisation dans cette partie du code est faite fondamentalement en mémoire distribuée, en découpant le vecteur des points de la grille horizontale et autant de segment qu'on n'a de process MIPxOpenMP dans la partie dynamique. Cette parallélisation complètement différente dans les partie dynamique et physique est rendue possible par le coût numérique de ces deux composantes à chaque itération en temps.

A noter que cette indépendance entre le découpage de domaine pour la physique et de la dynamique a été facilité par le fait que le codage de ces deux parties a depuis longtemps été développé dans cette idée d'indépendance ou d'orthogonalité, à l'origine pour permettre le couplage d'un même noyau dynamique avec plusieurs jeux de paramétrisations pour différentes planètes, puis plus récemment pour travailler avec des version 1D de la physique, des versions 2D latitude-altitude du code dynamique (cf. Fig. **??**).

Les premières versions du code Dynamico commençaient tout juste à tourner au début du projet ANR et l'équipe ne s'était pas engagée dessus pour le projet ANR convergence. Mais les travaux ayant bien avancé sur ce point, et moins sur un autre aspect proposé sur le projet (à savoir le caractère "informé sur l'échelle" des paramétrisation de la convection), il a été décidé de remplacer

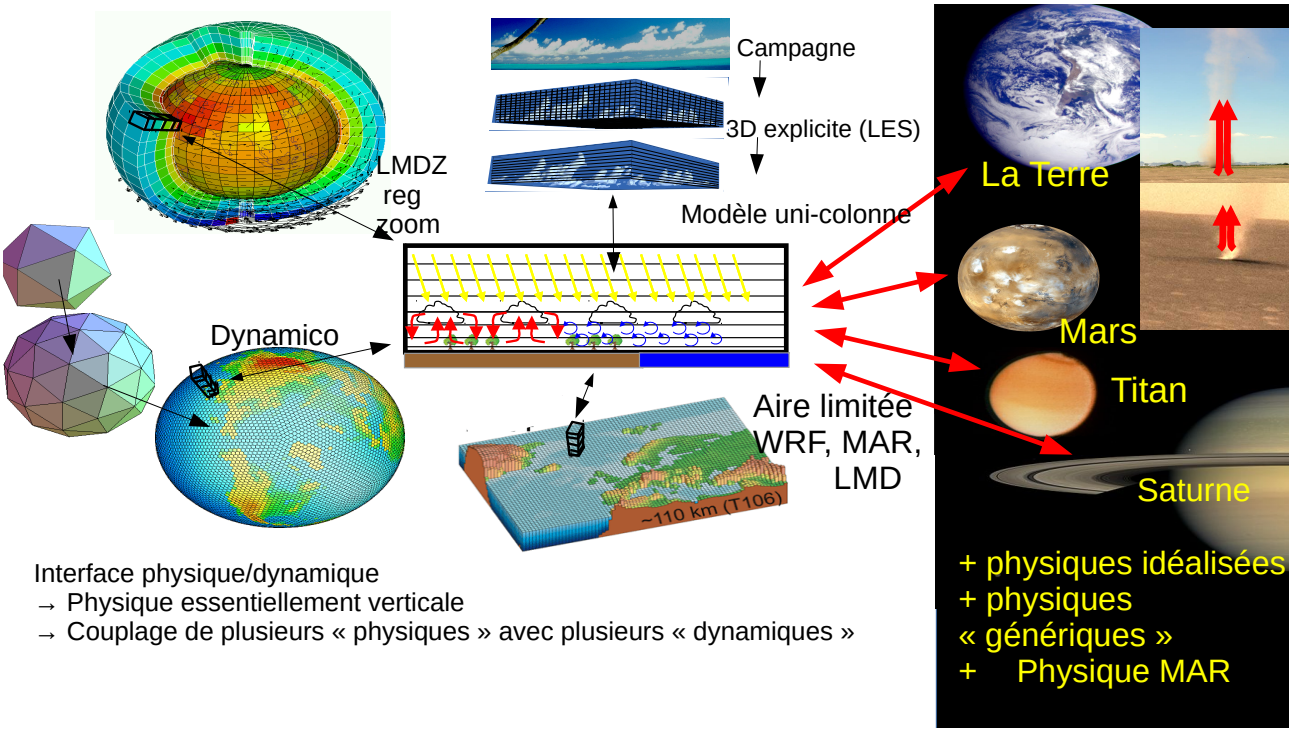

FIGURE 1 – Vision schématique de l'articulation des différents noyaux ou configurations dynamiques et différents jeux de paramétrisations physiques autour du même découpage physique/dynamique.

le Livrable sur cet aspect par un livrable sur la convergence entre les version lon-lat et dynamico du modèle LMDZ, avec la toute dernière version de la "physique" de LMDZ, appelée 6A, développée pour le modèle couplé IPSL-CM, et utilisée pour l'exercice international d'inter-comparaison de modèles CMIP6. Ils s'avère que finalement, des résultats intéressants ont également été obtenus en ce qui concerne le premier aspect, grâce aux avancées des paramétrisations physiques, qui ont permis une bien meilleur comportement du modèle avec les changements de résolution comme on l'explique dans le WP2.5a.

On présente donc ci-dessous, les performances comparées climatiques et informatiques des deux options. Pour la partie physique, le but est essentiellement de reproduire des résultats équivalents. En effet, si la discrétisation des deux noyaux dynamique est différentes, les équations qui sont résolues sont les mêmes et beaucoup des choix numériques (grandeurs conservées, position des variables sur la grille) sont les mêmes.

2 Comparaison des performances physiques

Il s'agit ici de qualifier le nouveau noyau dynamique dynamico en vérifiant qu'il est capable de reproduire un "climat" comparable à celui obtenu dans la configuration standard du modèle avec la grille lon-lat. On utilise pour ce faire la toute dernière version des paramétrisations physiques développées pour le projet CMIP6. On utilise en lon-lat la grille standard de IPSL-CM6A, avec 144 par 142 points, soit une résolution de 2.54^opar 1.25^o, ou encore une taille de maille de 250 par 125 km à l'équateur et 125 par 125 à 60^ode latitude. La résolution équivalente pour l'icosahère est basée sur une discrétisation en 40 points de chaque arrête des triangles de base de l'icoshèdre, soit au total $10 \times 40 \times 40 = 16000$ points de grille. Cette résolution est dite NB40.

Le modèle de surface continentales Orchidee n'étant pas encore complètement adapté au couplage avec Dynamico, on utilise ici un modèle très simplifié dit "de saut d'eau" ou bucket pour l'hydrologie, couplé à une équation de conduction thermique sous la surface. Les simulations sont intégrées sur années consécutives, avec une température de surface de l'océan imposée, qui varie saisonnièrement sans variations interannuelles. On considère dans l'analyse une moyenne sur 5 années.

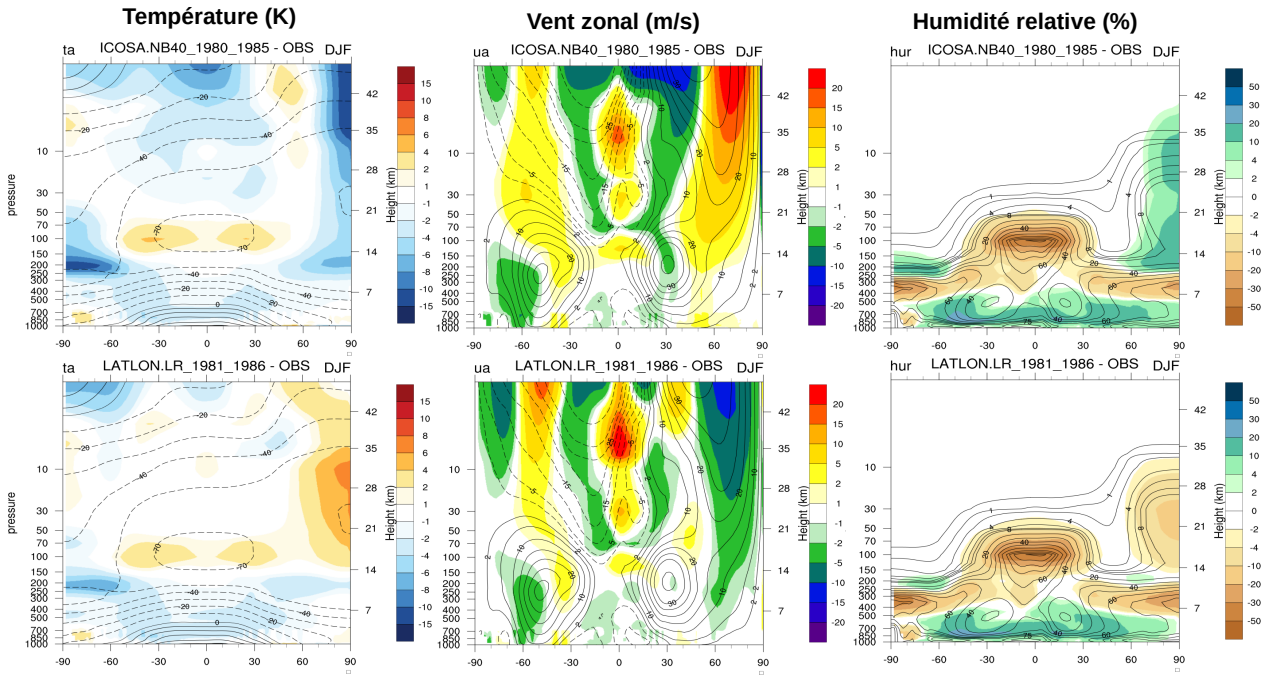

FIGURE 2 – Moyenne zonale pour les mois de Décembre-Janvier-Février de la température (K, à gauche), du vent zonal (m/s, au milieu) et de l'humidité relative (%, à droite) dans des simulations avec le noyau dynamico en haut et lon-lat en bas. Les contours correspondent au champs bruts et les aplats de couleurs à la différences aux réanalyses ERA Interim.

On commence par présenter sur la Fig. **??** les comparaisons des biais sur la structure méridienne moyenne. On constate de grande similitudes entre les biais des deux simulations. Les seules différences notables concernent les hautes latitudes nord, dans la stratosphère. On voit que le biais chaud dans lon-lat se transforme en un biais froid avec Dynamico. Le biais sec en humidité relative, se transforme en biais humide, sans doute simplement parce que le calcul de l'humidité à saturation est affecté par la température. Dans la troposhère en revanche, on remarque très peu de différence. Moins par exemple que celles qui viendraient d'un changement de résolution comme on le voit dans le livrables D2.5a.

Les biais froids des moyennes latitudes sont présents dans les deux simulations. Le dipôle jaune/vert, positif/négatif, autour du maximum du jet d'ouest des 40èmes rugissants, traduit un positionnement trop près de l'équateur bien connu, très proche dans les deux versions. Le biais humide des basses couches est également très similaire.

On présente ensuite sur la Fig. **??** les moyennes zonales et annuelles de la pression au niveau de la mer, de la précipitation, du bilan radiatif au sommet de l'atmosphère, et de la contribution des nuages à ce bilan (SW, LW et total). On voit généralement que les différence entre les deux modèles sont très inférieures à la différence avec les observations. On sait par ailleurs qu'elles sont inférieures à des petites perturbations des paramètres d'ajustement. La seule différence notable est sans doute la pression dans les hautes latitudes Nord, sans doute associée aux modifications déjà notées pour la structure méridienne. Cet aspect devra être confirmé avec des simulations plus longues du fait de l'importance de la variabilité interannuelle, mais on peut s'attendre à ce que les différences principales de comportement se passent près de la singularité des pôles.

Les cartes moyennes des précipitations montrent des défauts et qualités très proches entre les deux modèles. Avec la version "saut d'eau" de l'hydrologie, les pluies de moussons sont plutôt trop fortes alors qu'elles sont en général sous-estimées avec orchidee. On voit aussi la très forte surestimation de la précipitation sur la zone de convergence inter-tropicale sur les océans Pacifique et Atlantique Est, proches de l'équateur. Mais aucune différence remarquable entre les deux dynamiques.

On regarde enfin les cartes de biais de température de l'air près de la surface à la fois pour l'hiver et l'été. Là encore les différences sont faibles. La seule différence notable concerne le contraste entre

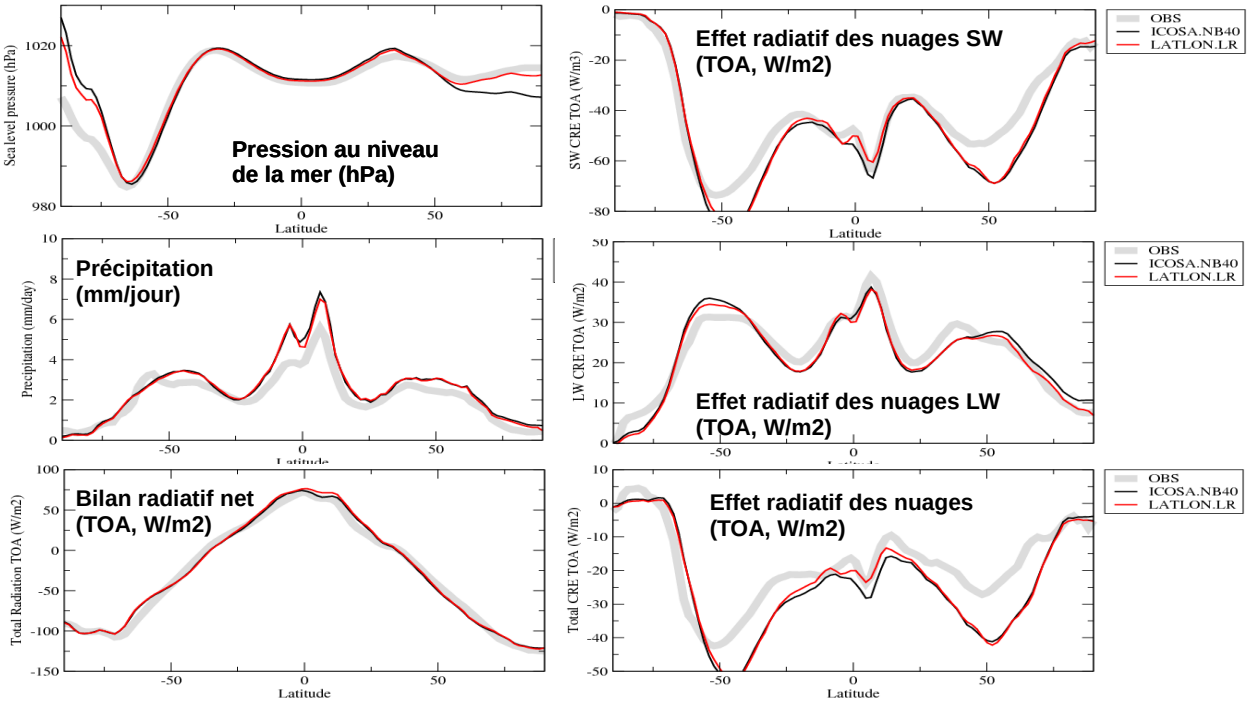

FIGURE 3 – Moyenne zonale et annuelle pour la colonne de gauche et de haut en bas : la pression au niveau de la mer, la précipitation (en mm/jour) et le bilan énergétique total au sommet de l'atmosphère (TOA pour top-of-atmosphère). Colonne de gauche : forçage radiatif des nuages (différence entre le rayonnement avec nuages et le rayonnement sans) dans les courtes longueurs d'ondes (SW pour Sort-Wave en haut), dans l'infra-rouge thermique (LW, au milieu) et total (en bas).

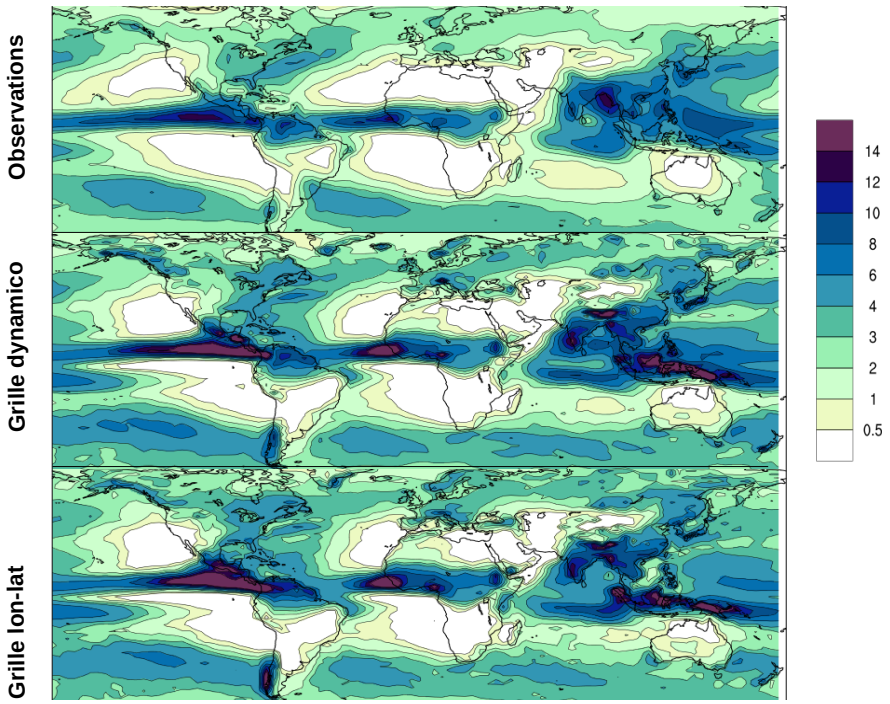

FIGURE 4 – Moyenne des précipitation moyennes de Juin-Juillet-Août (mm/jour) dans les observations GPCP (en haut), la simulation avec Dynamico et la simulation dyn-lon-lat (en bas).

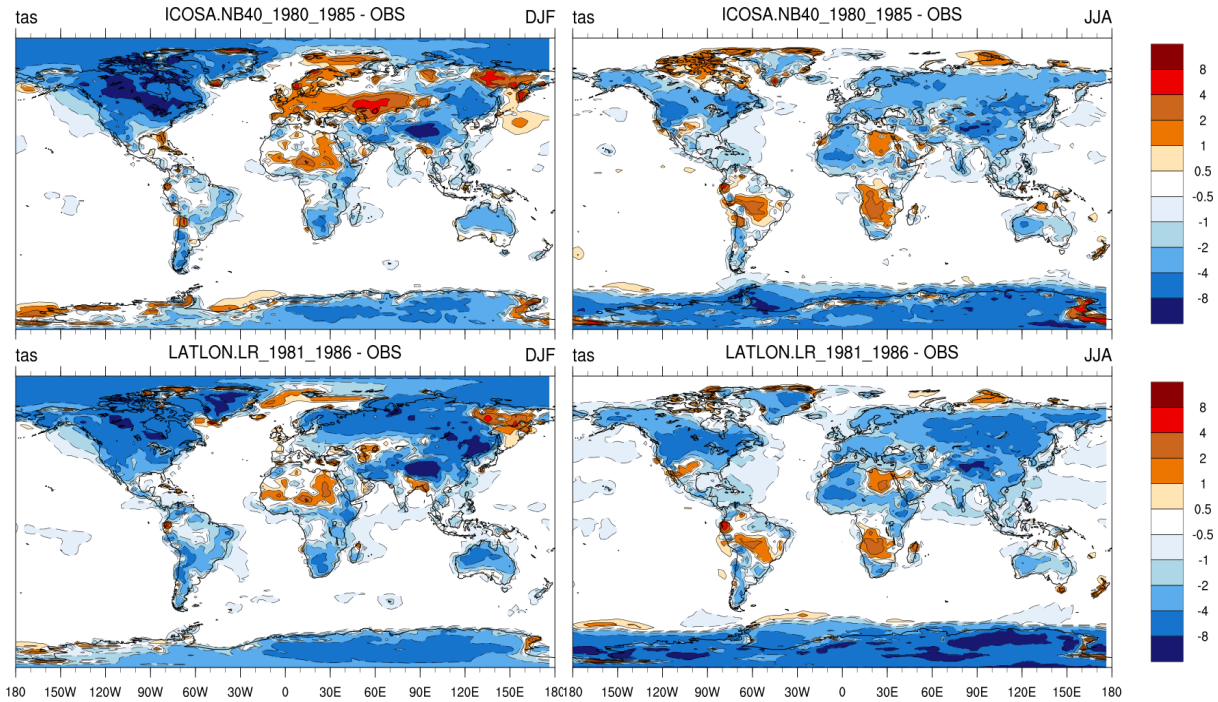

FIGURE 5 – Moyenne des biais de températures de l'air à 2m calculé par rapport aux réanalyses ERA Interim, pour dJuin-Juillet-Août (en haut) et Décembre-Janvier-Février (en bas).

l'Europe et les Etats-Unis en hiver. A nouveau, cette différence pourrait être dans la variabilité interne et devra être confirmée par des simulations plus longues. Mais elle semble tout de même un peu forte pour être due uniquement à la variabilité. Cela pourrait être dû aussi à un effet différent entre les deux modèles de l'orographie sous-maille dans les deux simulations.

A bout du compte, les résultats montrent clairement que les deux versions du modèle ont un comportement très similaire et que le couplage entre les noyaux dynamiques et la physique de LMDZ est réalisé correctement.

3 Performances computationnelles

Sur Irene, en poussant au maximum le parallélisme avec 71 processus MPI x 8 threads OpenMP soit 568 coeurs, la configuration CMIP6 standard de LMDZ tourne typiquement à 18 ans par jour CPU (à 18 SYPD, pour Simulated Years Per Day). Sachant que les demandes en sorties des exercices de production CMIP doublent le coût, on obtiendrait en l'absence de surcharge par les IOs environ 35 SYPD.

Ces chiffres peuvent être comparés, moyennant quelques hypothèses, à ceux de configurations DYNAMICO-LMDZ légèrement différentes. En particulier bien que DYNAMICO et LMDZ supportent chacun OpenMP, l'interface DYNAMICO-LMDZ ne le supportait pas au moment où des mesures de performance ont été faites, à la fin de l'été 2018.

Une configuration dynamico NB-40 en configuration CMIP5a tourne à 35 SYPD sur Curie et 50 SYPD sur Irene sur 160 processus MPI. Cette configuration compte 39 niveaux verticaux (contre 79 pour CMIP6) et fait appel toutes les 30 minutes à la physique. L'absence de parallélisme OpenMP implique une forte décomposition spatiale du domaine icosaédrique, conduisant à des domaines locaux de seulement 10×10 colonnes d'atmosphère. Le code passe environ 50% du temps dans la physique et le reste pour l'essentiel dans la dynamique (la dynamique elle-même passe plus de 50% de son temps en communications MPI en raison de la petite tailles des domains locaux). À nombre de coeurs constant, ces performances seraient divisées par 2 en passant à 79 niveaux verticaux et par 1,5-2 en passant à un appel à la physique toutes les 15 minutes. En exploitant le parallélisme OpenMP (réalisé sur des versions de développement et intégré dans les versions officielles de LMDZ d'ici fin 2018) il est envisageable d'atteindre 50 SYPD sur Irene à $160 \times 4 = 640$ coeurs.

Par ailleurs DYNAMICO permet beaucoup plus facilement de monter en résolution. Ainsi une année d'aquaplanète CMIP5a à nbp=320 (1/4 degré) et un pas de temps physique δt_{phys}=30min a réalisé environ 9 SYPD sur 10240 processus MPI. Le modèle passe 10% de son temps dans la physique. La dynamique passe les 3/4 de son temps en MPI car les domaines font seulement 5x5. Une configuration plus raisonnable serait 2560 MPI x 4 OpenMP pour avoir des domaines locaux de 10x10. En imaginant que ça divise par 3 le coût de MPI on passerait à environ 16 SYPD. On passerait à 8 SYPD à llm=79. Le coût de calcul serait dû à 80% à la dynamique, de sorte qu'on réaliserait encore environ 7 SYPD en réduisant δt_{phys}à 15 minutes, et encore 3-4 SYPD en divisant encore par 4 δt_{phys}.

En résumé, LMDZ-CMIP6 réalise sur Irene environ 35 SYPD sur 568 coeurs, sans surcharge par les IOs. Une configuration DYNAMICO-LMDZ équivalente devrait réaliser environ 50 SYPD sur 640 coeurs. Une configuration DYNAMICO-LMDZ au 1/4 de degré devrait réaliser 3-4 SYPD sur 10240 coeurs.

4 Conclusion

Le projet a donc permis de finaliser une première version du modèle atmosphérique, à la fois en termes de climat et de temps de retour. En termes climatiques il faudra quand même regarder plus finement ce qui se passe l'hiver dans l'hémisphère nord. En termes d'efficacité numériques ...

Le couplage avec orchidee est en cours de finalisation ainsi qu'un outils d'interpolation utilisant le logiciel XIOs pour généré les poids liants les grilles atmosphérique et océanique, poids qui sont passés au coupleur OASIS pour le couplage du modèle LMDZ et du modèle océanique Nemo.