

# D2.5a : Mise en œuvre de simulations haute résolution du modèle de l'IPSL

Frédéric Hourdin, Abdoul Khadre Traoré, Josefina Ghattas, Yann Meurdesoif,  
Laurent Fairhead, Arnaud Caubel

9 novembre 2018

## 1 Introduction

Le modèle de l'IPSL est habituellement utilisé avec des maillages horizontaux assez grossiers dans l'atmosphère, avec une maille de typiquement 2 à 3° de résolution, correspondant à environ une centaine de points répartis régulièrement en longitude et en latitudes. Plus précisément, on est passé pour la version dite basse résolution (LR en anglais) d'une grille 96 x 95 longitude x latitude pour la 5ème édition de l'exercice international d'inter-comparaison des modèles couplés, CMIP5 en 2012, à une grille de 144x142 en 2018. En parallèle, la résolution de la grille verticale s'est accrue pour passer de 39 niveaux à 79 niveaux. Le choix des grilles horizontale et verticale répond à des enjeux et des ajustements très différents. La grille verticale est intimement liée aux choix et aux réglages des "paramétrisations physiques", qui représentent l'effet collectif sous mailles des processus non résolus (turbulence, nuages, convection, états de surface ...). L'accroissement de la grille horizontale jusqu'à quelques dizaines de km de résolution, remet moins fondamentalement en cause le choix des paramétrisations physiques. Il permet en revanche de mieux résoudre les phénomènes météorologiques les plus violents, et de mieux prendre en compte l'effet des contrastes à la surface sur le climat : îles, reliefs, contrastes d'albédo ...

Quand on approche de quelques dizaines de km de résolution cependant, le comportement de la paramétrisation des orages (ou convection profonde) peut être modifié. Une partie de cette pluie orageuse peut se transformer en "orages point de grille", modes physico-numériques dans lesquels une colonne atmosphérique se met à monter en masse à des vitesses de l'ordre du m/s, et précipite toute son eau à la surface. Autant cette convection explicite est souhaitable avec des mailles horizontales kilométriques ou hectométriques (on commence alors à résoudre la convection orageuse explicitement), autant elle peut devenir un problème numérique avec des mailles de quelques dizaines de km.

L'optimisation réalisée quant à la parallélisation de LMDZ, et l'activation de XIOS et du workflow largement portées par l'ANR convergence, a permis de se projeter dans l'exercice CMIP6 vers la réalisation d'un petit jeu de simulations avec une grille beaucoup plus fine que la résolution nominale LR. Au lieu des 144 x 142 points, on utilise une grille de 512 x 360 points en longitude x latitude dite HR, ce qui permet d'atteindre une résolution horizontale isotrope ( $\delta x = \delta y$ ) de 50 km ou 1/2° à 45° de latitude.

Il était initialement prévu d'effectuer des simulations au 1/3 ou au 1/4 de degré. Mais l'initialisation des simulations s'est avérée difficile pour des problèmes de mémoire. En revanche, un accomplissement majeur du projet a été de réaliser les simulations de production pour la partie HighResmip du projet CMIP6, en utilisant la configuration HR, la configuration XIOS et le workflow de CMIP6. Par ailleurs, l'amélioration des paramétrisations physiques du modèle a permis d'obtenir un bien meilleur comportement du schéma convectif, avec des paramétrisations relativement "scale aware", c'est à dire qu'elle s'adaptent à l'échelle du maillage. Cette partie devait faire l'objet du livrable 2.1b qui a été remplacé finalement par un livrable sur le test de la nouvelle dynamique icosahédrique. Cet aspect est du coup également documenté dans le présent

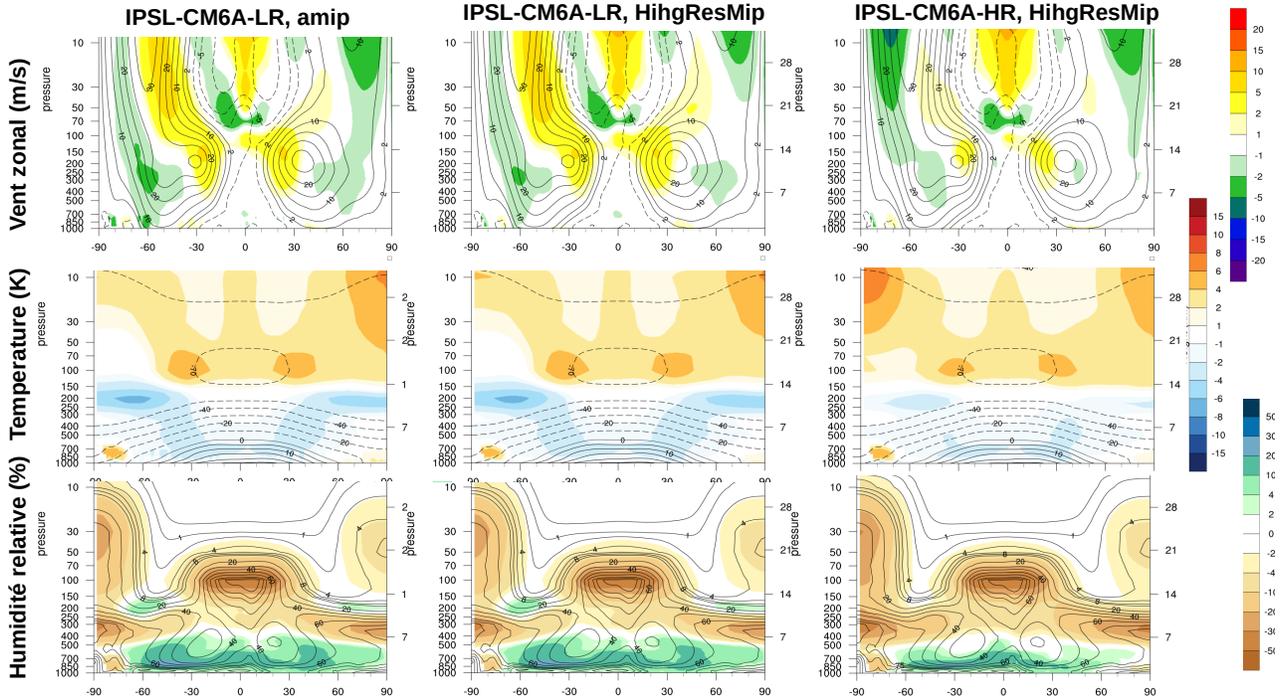


FIGURE 1 – Moyenne zonale et annuelle de la température (K, en haut), du vent zonal (m/s, au milieu) et de l’humidité relative (% en bas) dans des simulations standard IPSL-CM6A-LR (à gauche), HighResMip-LR (au centre) et HighResMip-HR (à droite). Sur chaque graphe, l’axe des ordonnées montre à gauche la pression en hPa et à droite l’altitude correspondante. Les contours correspondent aux champs bruts et les aplats de couleurs à la différences aux réanalyses ERA Interim.

livrable en même temps que les performances physiques et numérique de la configuration HR, comparée à la configuration nominale LR.

## 2 Comparaison des climatologies des simulations LR et HR

### 2.1 Le protocole HighResMip

Les simulations présentées ici suivent le protocole HighResMip de CMIP6 décrit sur :

[http://collab.knmi.nl/public\\_pages/102](http://collab.knmi.nl/public_pages/102)

Il s’agit de simulations de la période actuelle avec le modèle atmosphérique LMDZ6 couplé avec le modèle de surfaces continentales orchidee, mais sans composante océanique. A la place les températures de surfaces sont imposées pour la période 1950-2014 en suivant le protocole HighResMIP. Ce protocole est différent du protocole AMIP sur quelques points et notamment en ce qui concerne la prescription des températures et concentrations de glace sur océan qui viennent des réanalyses journalières HadISST2.2 et les conditions de surface continentales qui sont simplifiées avec une végétation constante et un leaf area index (LAI) imposé par une climatologie mensuelle. Les forçages aérosols ont été d’autre part un peu simplifiés par rapport au protocole AMIP de référence de CMIP6.

Les forçages aérosols ont été d’autre part un peu modifiés par rapport au protocole AMIP de référence de CMIP6. De ce fait, on ne peut pas comparer directement les simulations AMIP-LR et HighResMip-HR, et il faut donc effectuer une simulation complémentaire avec le protocole HighResMip mais la grille LR.

## 2.2 Structure méridienne

Sur la »merid on compare en moyenne annuelle et pour la décennie 1989-1990 les trois simulations, IPSL-CM6A-LR (amip) à gauche, HighResmip-LR au centre et HighResmip-HR à droite. On constate d'abord par une lecture globale que les différences liées au changement de protocole (entre colonne de droite et centrale) sont nettement plus faibles que celles liées au changement de résolution.

La comparaison des colonnes centrale et de droite montre une amélioration significative de la climatologie du modèle avec la montée en résolution entre les configurations LR et HR. Les améliorations sont très proches dans leur nature des améliorations observées dans une exploration systématique de la sensibilité à la résolution horizontale avec une version pré-CMIP5 du modèle (?). On observe notamment, dans la troposphère (entre la surface et 21 km, axe de droite), on observe une diminution systématique des biais de vent zonal (en haut) pour lesquels on perd typiquement un niveau de couleur du côté positif comme négatif. Le dipôle entre des biais négatifs dans les moyennes latitudes (verts) et positifs entre les tropiques (jaunes) visible en configuration LR correspondent à un positionnement des jets d'Ouest (contours positifs) trop proche de l'équateur. La diminution de ces biais correspond à un déplacement de ces jets vers les pôles, en meilleur accord avec l'observation et de façon conforme avec ce qui était observé dans l'étude sus-mentionnée. A ce déplacement des jets est associé une diminution des biais froids (ligne du milieu) et humide (ligne du bas) des moyennes latitudes dans la troposphère. La réduction du biais froid est particulièrement marquée à la tropopause dans les hautes latitudes, où elle est sans doute liée à une diminution de la diffusion numérique qui a tendance à amener trop de vapeur d'eau dans cette région en configuration LR, aboutissant à des biais froids plus forts que -6 K au dessus de l'Antarctique. Cette interprétation est corroborée par le léger biais positif en humidité dans la version LR dans ces hautes altitudes vers 60S et 60 à 90N qui se transforme en biais positif dans la simulation HR.

A noter cependant que la structure générale du biais en humidité est assez peu affecté, ce qui reflète sans doute des améliorations possibles dans les paramétrisations physiques, mais doit être relativisé cependant du fait de la moindre fiabilité des analyses utilisées ici comme référence, pour l'humidité que pour les autres champs.

Les moyennes zonales de la pression de surface, de la précipitation et du rayonnement total au sommet de l'atmosphère (colonne de gauche de la Fig. 2) ne montrent également que peu de différences. En revanche, si on détaille un peu la contribution des forçages radiatifs des nuages au bilan radiatif (colonne de droite), on se rend compte que les nuages sont légèrement plus réfléchissants dans la simulations HR que dans les deux autres, dans les tropiques (alors que les différences dues au changement de protocole sont à peine visible à l'échelle du trait, comme on pouvait s'y attendre). Le forçage radiatif dans l'infrarouge thermique (LW) est également quelque peu affecté, avec des nuages fournissant un peu moins d'effet de serre dans les moyennes latitudes notamment. Au total (panneau en bas à gauche), les nuages sont légèrement plus refroidissant dans la configuration LR, ce qui se traduit par une différence globale au sommet de l'atmosphère d'environ -2 W/m<sup>2</sup>.

Les biais saisonniers des températures à 2m sont comparés sur la Fig. 3. On voit une tendance générale à une augmentation des températures sur les continents dans la configuration HR, cohérente avec ce qu'on voit sur les moyennes zonales (Fig. 1). Cette augmentation se traduit par une augmentation d'un biais plutôt chaud en été dans les moyennes latitudes nord (en haut) et un amoindrissement relatif d'un biais froid dans les mêmes régions l'hiver (en bas). Mais c'est finalement à nouveau la similarité des biais dans les différentes simulations qui l'emporte. A noter que les biais très froids sur les calottes groenlandaise et Antarctique sont sans doute davantage un biais chaud dans les réanalyses utilisées ici comme référence qu'un problème du modèle.

## 2.3 Précipitations

On analyse à présent les changements de distribution de précipitation.

On présente dans un premier temps sur la Fig. 4 les cartes de précipitations moyennes de Juin-

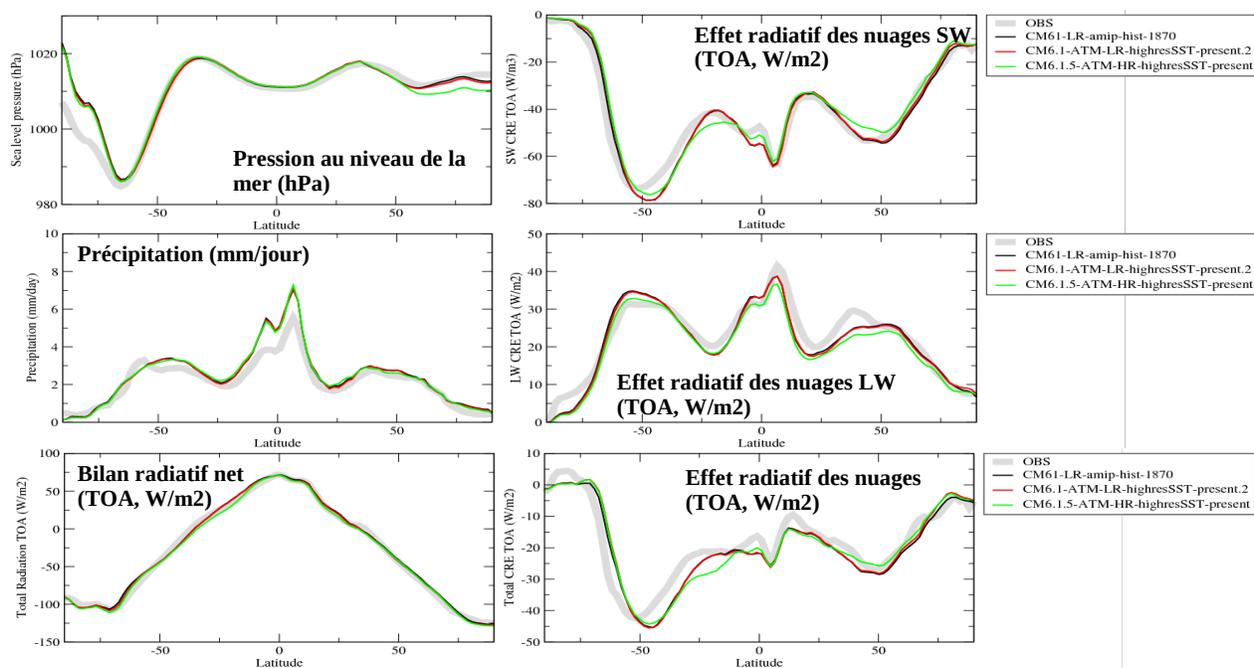


FIGURE 2 – Moyenne zonale et annuelle pour les observations (gris) et les simulations IPSL-CM6A-LR (noir), HighResmip-LR (rouge) et -HR(vert) de différentes variables. Pour la colonne de gauche et de haut en bas : la pression au niveau de la mer, la précipitation (en mm/jour) et le bilan énergétique 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

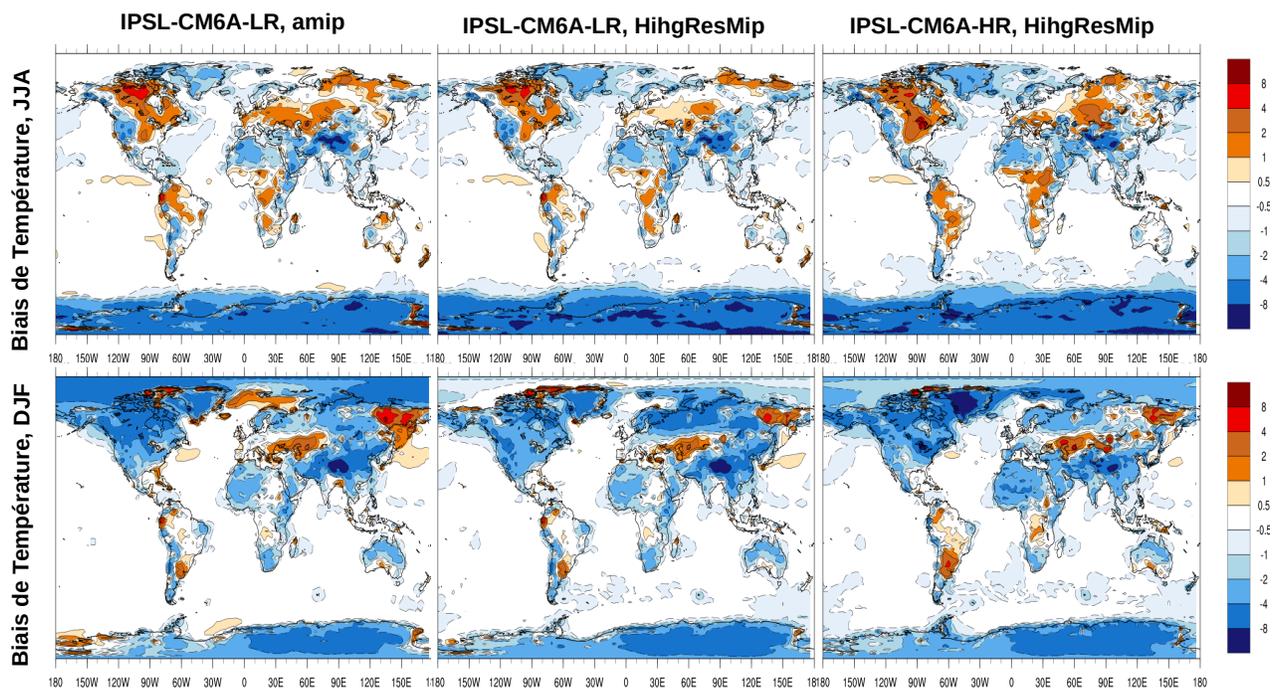


FIGURE 3 – Moyenne des biais de températures de l’air à 2m calculé par rapport aux réanalyses ERA Interim, pour dJuin-Juillet-Août (en haut) et Décembre-Janvier-Février (en bas) et pour les simulations IPSL-CM6A-LR (à gauche), HighResmip-LR (au centre) et -HR (à droite).

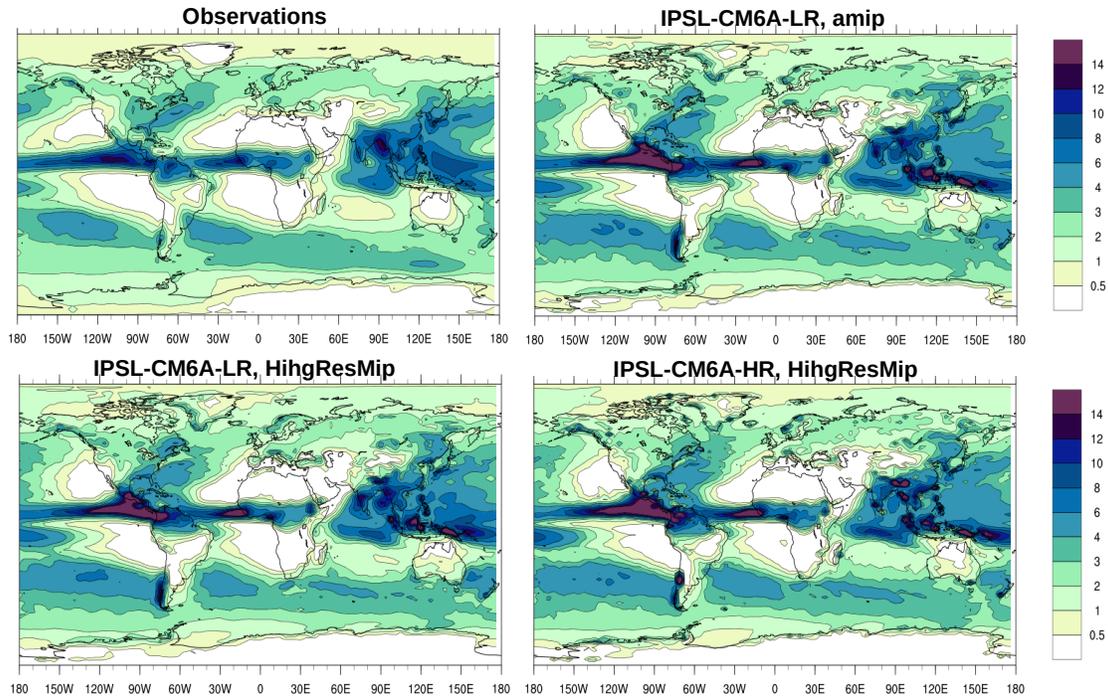


FIGURE 4 – Moyenne des précipitation moyennes de Juin-Juillet-Août (mm/jour) dans les observations GPCP (en haut), la simulation avec Dynamico et la simulation dyn-lon-lat (en bas).

Juillet-Août (la saison des moussons dans l’hémisphère nord). Comme pour les figures précédentes, on constate des modifications relativement faible, tant en changeant le protocole que la résolution. Les principaux défauts sont conservés, à savoir une tendance à exagérer les pluies dans la Zone de Convergence Inter Tropicale (ZCIT) sur les océans atlantique et pacifique et les précipitations sur les îles du continent maritime. La structure des pluies de mousson sur l’Afrique (avec une transition entre zone pluvieuse et désert vers 15N) et l’Asie est relativement bien représentée, malgré une tendance robuste du modèle à situer la zone de transition un tout petit peu trop au sud sur l’Afrique et à sous-estimer les précipitations au Nord Ouest du sous-continent indien. Cette représentation n’est que peu affectée par la résolution. On constate principalement comme différence quelques spots de précipitations, un sur la cordillère des Andes et deux autres au Nord Est du sous continent indien, probablement associés à des aspects particuliers du relief, mieux résolu dans la simulation HR. Mais le comportement est remarquablement proche étant donné les changements de résolution.

La Fig. 5 détaille ce comportement des précipitations du modèle en comparant directement les sorties des différentes simulations à des observations obtenus sur un pluviomètre sur un site d’observation intensive du réseau AMMA-Catch, situé au Mali. On voit sur l’évolution des humidité spécifiques (traits fins) l’alternance entre un hiver très sec, correspondant à l’advection d’air depuis le Sahara par l’Harmatan, et une saison humide (de mai à octobre) au cours de laquelle se réalise l’intégralité des précipitations (barres). L’ensemble des simulations parvient à reproduire dans les grandes lignes cette saisonnalité. Les 3 simulations ont des défauts en commun comme une sous-estimation des maxima de températures (courbes épaisses) en mai, qui fait que la décroissance des températures observée pendant la saison des pluies est moins marquée.

Si on s’intéresse plus spécifiquement à la distribution des pluies, on voit un saut qualitatif par rapport à la simulations IPSL-CM5A-MR qui utilisait la même grille que la simulation CM6A-LR mais un jeu de paramétrisation très différents. Avec cette configuration, le modèle avait tendance à simuler les pluies de façon continue au Sahel, alors que les observations montrent des événements de pluies bien distincts, avec souvent plusieurs jours sec entre deux événements.

Cet aspect est grandement amélioré dans les simulation avec la physique 6A dans laquelle on active le modèle des thermiques (paramétrisation en flux de masse des structures organisées de la couche limite convective), un modèle de poches froides créées sous les cumulonimbus par réévaporation des pluies convective, et une fermeture et un déclenchement modifiés pour la

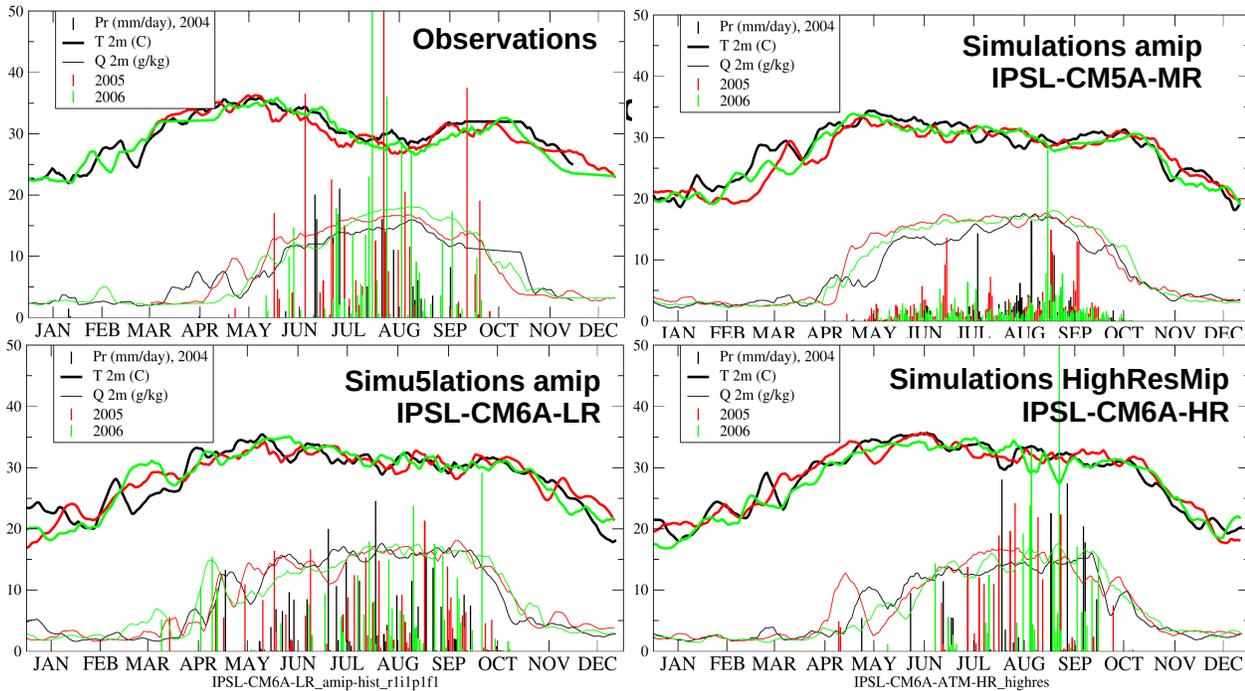


FIGURE 5 – Evolution sur 3 années consécutives (2004, noir, 2005, rouge, 2006, vert) des températures de l’air à 2m (courbes épaisses), de l’humidité spécifique (courbes fines) et de la précipitations (barres verticales) à la station d’Agoufou au Mali, pour les observations (en haut à gauche), la simulation LMDZ5A-MR (en haut à droite) et les simulations HighResmip (en bas) avec la grille LR (à gauche) et HR (à droite).

convection profonde, tenant compte des mouvements verticaux créés par les thermiques et par les fronts de rafales des poches froides. Le déclenchement de la convection est contrôlé par le fait qu’un panache thermique de la couche limite permette de passer une barrière d’inhibition et dépasse un seuil en surface (de quelques km<sup>2</sup>). Ce déclenchement fait intervenir un tirage aléatoire de la taille du plus gros panache thermique dans la maille. Ce tirage permet de définir un déclenchement qui est sensible à la taille de la maille du modèle. Effectivement, si la probabilité d’avoir des panaches dont la taille dépasse le seuil est faible, alors la probabilité de déclencher sera plus grande dans une maille plus grande.

Grâce à ce nouveau jeu de paramétrisations physique, on obtient une bien meilleure représentation de la distribution temporelle de la pluie à la station. Les maxima observés sont un peu plus faibles que dans les observations, mais, les mesures étant faites localement sur un pluviomètre, une variabilité locale vient se surajouter. De façon cohérente, on voit que les événements extrêmes de précipitations s’accroissent avec la résolution en passant de la grille LR (environ 200 km de résolution) à la grille HR (environ 50 km).

Si on regarde pour un mois donné les extrêmes de précipitation obtenus, Fig. 6, pour différentes grilles horizontales, on voit que la précipitation augmente fortement quand on raffine la grille pour l’ancienne physique 5A. On atteint à 50 km de résolution des pluies qui approchent le m sur la journée, ce qui semble a priori très excessif. On voit que la situation était encore bien pire avec la version 5B du modèle, correspondant à la première version de la ”nouvelle physique” de LMDZ mise en oeuvre pour CMIP5. On voit enfin que la dernière version de la physique, avec la paramétrisation du déclenchement qui tient compte de la surface de la maille au travers du déclenchement stochastique, permet de contrôler beaucoup mieux les extrêmes de pluies. On voit des extrêmes qui augmentent un peu quand on passe de 200 environ (LR), puis à 150 (MR) puis à 50 km de résolution, mais dans une bien moindre mesure que les pour les versions précédentes de la physique.

Des études beaucoup plus fines sont nécessaires, à la fois sur les observations et les sorties de modèles, pour vérifier quantitativement la cohérence des observations et de ces simulations, mais l’essentiel semble être présent dans le modèle.

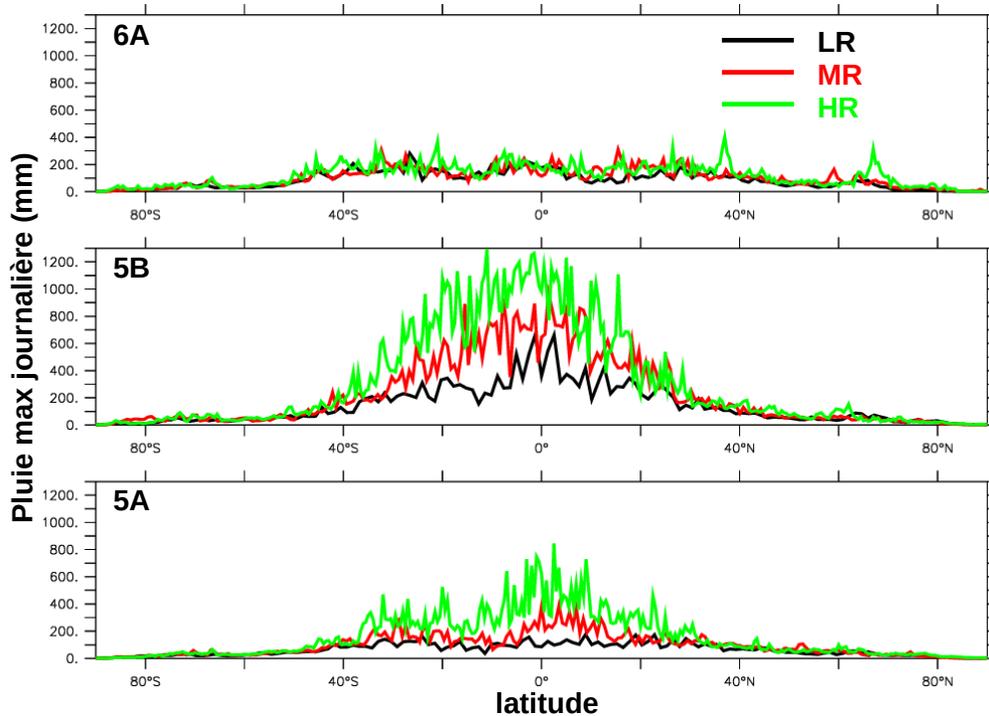


FIGURE 6 – Maximum de la précipitation journalier à une latitude donnée atteinte pour un mois de mars dans des simulations climatiques forcées en SST, avec les trois physiques atmosphériques, 5A, 5B et 6A et trois grilles horizontales, LR, MR et HR.

### 3 Performances numériques

Le modèle LMDZ est doté d’une double parallélisation MIP/OpenMP.

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 des deux parties est faite très différemment. Dans la partie dynamique, la présence d’un opérateur filtrage dans la direction  $x$  (longitudes) fait qu’aucun découpage de domaine n’est réalisé sur cette direction. On a donc ici une double parallélisation : 1/ par bande de latitude en mémoire distribué en utilisant la bibliothèque MPI. 2/ sur la verticale en mémoire partagée sous OpenMP. Le nombre de processeur est limité par cette parallélisation du noyau dynamique. En effet, on ne peut pas avoir moins de deux bandes de latitude par processeur MPI. Sur la verticale, on voit que les performances commencent à saturer à partir d’une dizaine de niveaux verticaux par tâche OpenMP.

La parallélisation de la partie physique 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 composants à chaque itération en temps.

On montre sur la Fig. 7 les performances obtenues en variant à la fois le nombre de processeurs MPI et de threads OpenMP. On montre à la fois le nombre d’années simulées par jour de calcul (à gauche) et le coût en heures CPU (ou milliers d’heures CPU Pour HR) de 10 ans de simulations. Les tests ont été réalisés sur la machine ada de l’IDRIS.

Si on analyse les résultats de la simulation LR, on voit qu’on arrive à monter jusqu’à une vingtaine d’années de simulation par jour si on utilise le maximum autorisé de processus MPI et 8 threads OpenMP. Cette configuration est cependant la plus coûteuse en CPU. Le coût notamment de passer de 4 à 8 threads est sans doute plus important que ce qu’on pourrait attendre puisqu’on a 79 niveaux sur la verticale, donc 10 niveaux par threads dans la version OMP8. On voit cependant que le modèle conserve une certaine scalabilité jusqu’à ces niveaux

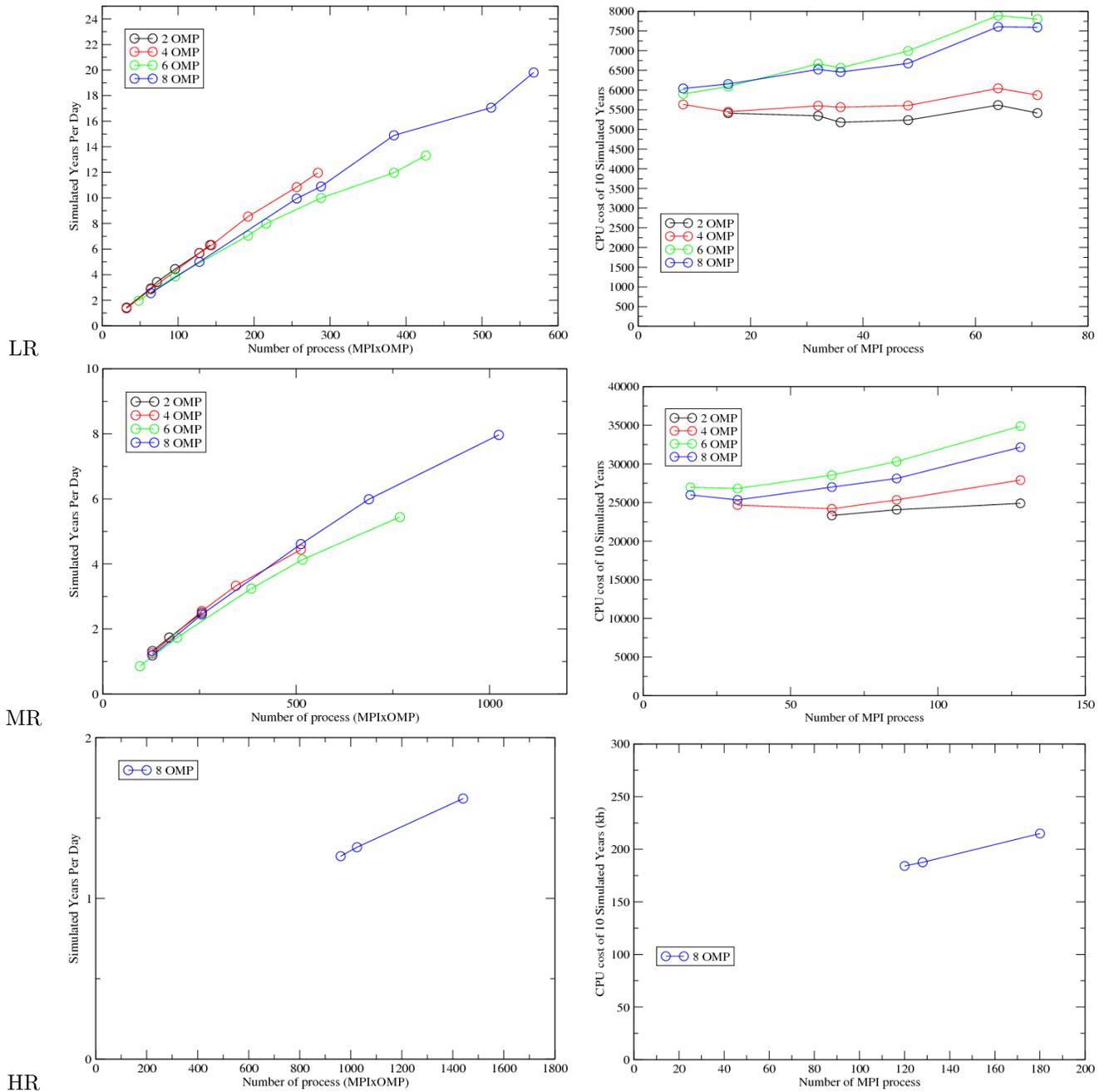


FIGURE 7 – Coût des configurations LR (en haut), MR (au milieu) et HR (en bas) en nombre d'années simulées par jour de calcul (à gauche) et en coût en heures CPU (ou de milliers d'heures CPU pour HR) de 10 ans de simulation (à droite). Les différentes courbes correspondent à différents niveaux de parallélisation OpenMP. L'axe des abscisse au nombre de processus MPI (à droite) ou nombre total de threads (MIPxOpenMP, à gauche).

de découpages extrêmes, sûrement en grande partie avec la scalabilité quasi parfaite de la partie physique (parallélisation en blocs de colonnes verticales indépendantes les une des autres).

Le passage en moyenne résolution MR montre des résultats très similaires.

Pour la haute résolution HR, on commence à buter sur les limites en mémoire de la machine par processus ou threads. Seules des configurations avec 8 threads arrivent à entrer en machine, et seulement avec un nombre de processus MPI supérieur à 120. On voit cependant là aussi qu'on conserve une certaine scalabilité. Pour cette configuration, le découpage le plus rapide permet de s'approcher des deux ans par jour de calcul.

Avec des nombres de points en longitude x latitude de 144 x 142 (LR), 256 x 256 (MR) et 512 x 360 (HR), on a un rapport de nombre de point de grille MR/LR = 3,2 et HR/MR = 2.8. Le pas temps de calcul de la partie dynamique doit aussi être réduit en fonction de la résolution en longitude (critère CFL). Ce qui donne pour la partie dynamique un rapport de coût MR/LR=1,8 et HR/MR=2. En revanche, le pas de temps de la partie physique est indépendant de la résolution horizontale (15 minutes).

Comparer à ces chiffre, le rapport entre les coûts de calcul pour 10 ans est pour MR/LR de 32000/7500=4,3 et pour HR/MR de 6,8.

En temps de retour, le changement du nombre de latitude est sensé être entièrement absorbé par l'augmentation du nombre de processeur MPI, ce qui donne les mêmes ratio que pour le pas de temps. Les ratio des nombres d'années simulés par jour de calcul sont pour LR/MR de 20/8=2,5 (ce qui est bien intermédiaire entre 1,7 et son carré) et pour MR/HR de 8/1,7=4,7 (ce qui est là un peu supérieur au carré de 2.).

Ces résultats en coût de calcul et la limitation en mémoire suggère que la configuration HR est à la limite des capacités de la machine ada.

La situation s'est un peu améliorée sur irene, la nouvelle machine du TGCC.

Avec 180 MPI et 12 OpenMP + 44 x 8 dédiés pour XIOS, il a été possible de tourner en 6h30 par année de simulation, c'est à dire un peu plus de 3,5 années par jour. Les 8 threads OpenMP utilisés pour XIOS n'étaient là que pour accroître la mémoire, à cause d'un problème d'implémentation de XIOS/OpenMP dans LMDZ, résolu depuis.

Il a été ainsi possible de réaliser une simulation de production pour HighResmip sur la machien irene, dans sa période de mise en route, dans le cadre d'un "grand challenge" piloté par Abdoul Khadre Traore, et dont les résultats sont présentés ci-dessus. Comme pour les simulations CMIP6 standard, ces simulations ont profité du logiciel XIOS pour les entrées/sorties et du contrôle des fichiers de sorties par les fichiers xml.