

Centre National de la Recherche Scientifique Section 19

Rapport d'activité à 5 ans

Frédéric Hourdin

26 mars 2024

Table des matières

1	Curriculum Vitae abrégé	3
2	Résumé de l'activité de recherche (1/2p)	4
3	Rapport d'activité	5
3.1	Introduction	5
3.2	Paramétrisation de la convection et des nuages	5
3.3	Tuning des modèles de climat	7
3.4	Interactions nuages rayonnement	10
3.5	Anticipation du changement climatique et incertitudes	11
3.6	Quantification des incertitudes sur l'ECS	13
3.7	Calculs sonde Monte Carlo	15
4	Enseignement et formation	18
4.1	M2	19
4.2	Cours de thèse	19
4.3	Le travail en ateliers	20
4.4	L'écriture de textes pour l'enseignement	20
4.5	L'écriture de textes pour la communauté	20
4.6	Conférences grand public et supports	21
5	Transfert technologique, relations industrielles et valorisation	23
6	Encadrement, animation	23
6.1	Le pilotage du projet LMDZ	23
6.2	Le co-portage du projet Dephy	24
6.3	Le rôle dans la modélisation du climat à l'IPSL	25
6.4	Participation à des conseils scientifiques	25

7	Objectifs de recherche	25
7.1	Développement de paramétrisations	25
7.2	La révolution de l’history matching	26
7.3	Recherche sur le développement informatique	27
7.4	Recherche pluridisciplinaire	28

1 Curriculum Vitae abrégé

Frédéric Hourdin

Né le 21 Février 1966

1985-1988 : Ecole Polytechnique

1988-1989 : DEA d'astrophysique à Meudon (Paris-7, major, mention TB)

1989-1992 : Thèse de l'université Paris-7 (Très honorable avec félicitations).

Sujet : Etude et simulation numérique de la circulation générale des atmosphères planétaires.

Laboratoire : Laboratoire de Météorologie Dynamique

Directeur : Olivier Talagrand

1992-1994 : Post-doc CNES au Laboratoire de Météorologie Dynamique incluant deux séjours de 5 semaines à NASA-Ames (Californie, USA) et à l'Université d'Oxford (Grande Bretagne).

1993 : Obtention du prix Guinier, jeune chercheur, de la Société Française de Physique.

1994 : Recrutement en CR2 au CNRS

1997 : Médaille de bronze du CNRS

1998 : CR1.

2005 : Obtention de l'Habilitation à diriger des Recherches sur la base d'un mémoire intitulé *"Représentation du transport direct et inverse dans les modèles globaux de climat et étude des couplages entre composition et dynamique atmosphérique sur Titan"*.

2006-2007 : Mis à disposition du Laboratoire de Physique de l'Atmosphère et de l'Océan Siméon Fongang de Dakar.

2009 : DR2

2008- : directeur adjoint du LMD

2 Résumé de l'activité de recherche (1/2p)

Mes recherches portent sur la modélisation du climat, sur l'anticipation du réchauffement global et de ses conséquences. Elles visent notamment à traduire, sous forme de paramétrisations, le comportement collectif des mouvements convectifs et nuageux qui émergent par exemple de simulations explicites dites LES. Ces paramétrisations sont mises en œuvre dans le modèle LMDZ dont je coordonne le développement. Lors du projet ANR HighTune, avec des collègues du CNRM et un mathématicien statisticien anglais, Daniel Williamson, nous avons introduit dans la communauté une approche de tuning ou calibration des paramètres libres des modèles basée sur l'history matching. Au delà de la semi automatisation du tuning de versions de référence du modèle global, cette approche change assez radicalement le regard porté sur le modèle et la façon de le développer. Elle permet également de calibrer les paramètres libres du modèle tout en quantifiant l'incertitude paramétrique associée, par exemple concernant la sensibilité climatique à l'augmentation de la concentration en CO₂. Cette ANR a été également l'occasion de renforcer un lien historique avec le réseau EDstar qui porte une recherche extrêmement originale sur l'utilisation de méthodes Monte Carlo, basées sur une reformulation "en chemins" de la physique. Cette collaboration a abouti à développer un outil de calcul de référence de rayonnement sur des scènes LES, permettant de combler un vide entre le tuning qui était fait en comparant des simulations uni-colonnes avec des LES sur les propriétés des nuages, et le tuning du modèle global 3D qui était fait prioritairement en utilisant des métriques portant sur les flux radiatifs au sommet de l'atmosphère. L'approche Monte Carlo portée par EDstar nous a aussi permis d'initier des recherches sur l'impact du changement climatique sur le climat urbain et la thermique des bâtiments, développées dans le cadre du projet ANR MC2 dont je suis coporteur, et d'une réponse à un appel d'offre PEPR VDBI ("Ville Durable et Bâtiment Innovant") en cours d'évaluation. Autour de ces recherches, je consacre également une partie significative de mon activité à développer des enseignements et à produire des écrits sur la modélisation du climat et à concevoir et mettre place de nouvelles modalités d'organisation du travail et de développement informatique du modèle.

3 Rapport d'activité

Comme le principe est de présenter le rapport autour de 10 productions scientifiques, j'ai choisi d'insérer les productions au fil du texte sans inclure d'autres éléments bibliographiques, qui auraient en grande partie repris ceux contenus dans les articles mis en avant.

Tous les pdf des articles sont accessibles à <https://web.lmd.jussieu.fr/hourdin/PUBLIS/publis.html>

Je suis désolé d'avoir totalement laissé passer la date butoir d'envoi de ce rapport malgré les nombreuses relances par mail.

3.1 Introduction

Depuis ma thèse sur la modélisation des climats de Mars et de Titan, mon activité de recherche tourne autour de la modélisation du climat. Avec la montée en puissance de la question du changement climatique (déjà présente à mon arrivée au laboratoire dans les années 90) je me suis de plus en plus investi dans la question de l'utilisation des simulations climatiques (qu'il faut distinguer de la modélisation du climat) pour l'anticipation du réchauffement global et de ses conséquences qui pilote la plupart de mes choix de recherche depuis.

3.2 Paramétrisation de la convection et des nuages

Le cadre offert par la séparation entre cœur dynamique et paramétrisations, propre aux modèles de circulation générale fondés dans les années 70, représente bien plus qu'un outil pour simuler les climats passés, présents et futurs ou de prévoir le temps à courte échéance.

Ces modèles reposent sur un découpage entre des équations représentant la circulation atmosphérique aux échelles météorologiques, et les paramétrisations qui rendent compte de transferts verticaux d'énergie, d'humidité et de quantité de mouvement dans un monde supposé localement homogène horizontalement d'un point de vue statistique (pas de variations horizontales des moments des distributions). Ce découpage est un cadre fondamental de compréhension du fonctionnement du système. En ce sens, les modèles de circulation générale à physique paramétrée constituent un cadre de pensée et de construction d'une physique. Au même titre que Navier-Stokes, et sa traduction macroscopique du mouvement des molécules, constituent un cadre de pensée de la mécanique des fluides.

Depuis une vingtaine d'années, nous sommes engagés avec des collègues du LMD et du CNRM dans un travail au long cours de développement et amélioration des paramétrisations de la couche limite, de la turbulence et des nuages, en utilisant massivement des simulations LES (réalisées notamment avec le modèle Meso-NH), simulations à résolution de 10 à 200m, particulièrement adaptées à l'étude de la convection (profonde ou peu profonde) et des nuages associés. En résumant le comportement des LES à quelques équations et en vérifiant qu'on arrive à reproduire les résultats des LES avec une seule colonne d'un modèle de climat, on gagne non seulement un facteur 10^8 à 10^9 en temps de calcul, mais surtout on effectue un saut de compréhension. Les propriétés qui émergent des LES, sous forme de panaches convectifs, cumulus ou autres, sont le résultat, dans les simulations uni-colonne (ou SCM pour Single Column Model en anglais), d'un modèle conceptuel de ces processus. La mise en œuvre de ces paramétrisations dans les modèles de circulation générale permet ensuite de comprendre le rôle de ces processus dans le climat ou les phénomènes météorologiques, en analysant

les flux, tendances ou transferts exprimés au travers des paramétrisations. L'analyse de simulations LES globales ne rend pas directement accessible ce rôle du fait de l'absence de découpage.

Au cours des 20 dernières années, nous avons construit en particulier des paramétrisations de la couche limite convective et des nuages associés basées sur une représentation en flux de masse des panaches convectif (le "modèle du thermique"), combinée à une approche en diffusion turbulente classique, active principalement dans la couche de surface (typiquement 100-300m d'épaisseur). Cette approche mise en œuvre et publiée pour la première fois dans le modèle du LMD, a depuis été popularisée sous le nom d'approche EDMF pour Eddy Diffusion Mass Flux. Le modèle du thermique, couplé à une représentation bi-modale de la distribution sous maille horizontale de l'eau (un mode pour les panaches thermiques et un mode pour leur environnement), a permis une représentation très satisfaisante des cumulus de couche limite.

Au début du présent exercice j'ai finalisé et publié un travail sur le modèle du thermique conduit pendant la thèse d'Arnaud Jam et impliquant Catherine Rio et Fleur Couvreur du CNRM. Ce travail est venu résoudre une impasse des paramétrisations en flux de masse existant précédemment, à savoir leur capacité à représenter simultanément les cumulus de couches limites et les bancs de stratocumulus qui se développent sur les bords Est des océans tropicaux, dans des régions de fortes subsidances. Ce travail a à la fois permis un gain énorme en termes de représentation des couvertures nuageuses dans les simulations climatiques, mais également d'améliorer la représentation des flux d'évaporations à la surface, qui contrôlent en grande partie les températures de surface de l'océan.

En effet, le nouveau schéma permet simultanément de maintenir une couverture nuageuses très élevée (100% pendant plusieurs mois dans certaines régions), ombrageant la surface, et de maintenir grâce au schéma en flux de masse une ventilation suffisante à la surface d'air sec provenant de la troposphère libre. Cette ventilation d'air sec renforce le refroidissement par évaporation très sensible à l'humidité relative de l'air à la surface. Ceci a permis de diminuer le biais chaud des températures océaniques sur les bords Est des océans tropicaux, biais systématique très connu des modèles couplés de climat.

Les transformations nécessaires du modèle du thermique permettant d'arriver à ces résultats concernaient le détrainement latéral d'air en haut des panaches thermiques. Avec la version originale du schéma, les panaches pénétraient trop haut au dessus de l'inversion de sommet de couche limite, qui se joue sur quelques dizaines de mètres verticalement. Il a fallu rendre le schéma de détrainement sensible à l'environnement à une certaine distance au dessus, ce qui revient à le rendre sensible à la force de l'inversion (mesurée notamment par les gradients des grandeurs thermodynamiques). Ce traitement vient sans doute en partie palier le fait que la résolution verticale du modèle est de beaucoup trop grossière (100 à 200 m à 1km d'altitude suivant les configurations, sachant qu'il faut typiquement une résolution verticale d'une dizaine de mètres pour représenter correctement ces nuages dans une LES). Ce traitement a aussi un rapport avec le fait que, en pratique, les panaches overshootent autour de l'inversion, l'air détrainé retournant en grande partie sous l'inversion sous forme de fontaines facilement identifiables au sommet des nuages. Enfin, ce travail a permis de réconcilier des visions très différentes du contrôle de ces nuages au sommet de la couche limite : celle des flux de masse et des paramétrisations "bulk" de l'entraînement (vertical) sommital, formulé comme une vitesse d'entraînement fois un contraste des valeurs de la quantité transportée q à l'inversion : $\overline{w'q'} = \alpha \Delta q$. Dans l'article, on montre comment on peut faire un lien très direct entre cette vision bulk et le transport par la subsidence compensatoire des panaches convectifs, elle même contrôlée par le détrainement. Le modèle du thermique offre finalement un modèle de la vitesse d'entraînement

sommitale w' , qui dépend à la fois des flux de masse dans la couche limite mélangée et de la force de l'inversion, au travers de la paramétrisation du déentraînement latéral. Ceci devrait permettre de tirer bénéfice de la co-existence de deux visions complémentaires de cette même question.

Le développement des paramétrisations est porté par très peu de chercheur.e.s (ou enseignant.e.s-chercheur.e.s), et les résultats marquants concernant les paramétrisations physiques au LMD ont été associés à typiquement 4 thèses sur 15 ans (celles de Catherine Rio, Nicolas Rochetin, Arnaud Jam et Etienne Vignon). Or on sent depuis deux ou trois ans que la situation est en train de changer, sans doute pour différentes raisons : 1) aboutissement d'une version performante sur le plan climatique incluant ces avancées (la "nouvelle physique") dans la versions 6A du modèle LMDZ, composante atmosphérique du modèle IPSL-CM, utilisée pour l'exercice CMIP6; 2) gain en organisation, souplesse et facilité d'intervention dans le modèle LMDZ, 3) recrutement CNRS d'Etienne Vignon, 4) reconnaissance et montée en puissance du programme Dephy. J'encadre actuellement 3 thésitifs et une ingénieure dont les travaux contribuent pour tout ou partie au développement de paramétrisations concernant : 1) la construction d'un modèle pronostique de l'eau sous-maille (thèse de Louis d'Alençon), 2) l'amélioration du modèle des poches froides créées par réévaporation des pluies sous les orages (thèse de Lamine Thiam), 3) la paramétrisation des bourrasques associées aux poches froides (dans la même thèse) et aux panaches thermiques (travail mené avec Adriana Sima, IR CNRS), et 4) une paramétrisation de l'hétérogénéité sous-maille des nuages pour prise en compte dans le nouveau code radiatif ECrad (Maëlle Coulon-Decorzens).

Production scientifique 1 : Frédéric Hourdin, Arnaud Jam, Catherine Rio, Fleur Couvreur, Irina Sandu, Marie-Pierre Lefebvre, Florent Briant, and Abderrahmane Idelkadi, Unified Parameterization of Convective Boundary Layer Transport and Clouds With the Thermal Plume Model, James, 2019, <https://doi.org/10.1029/2019MS001666>

3.3 Tuning des modèles de climat

La question de la calibration des paramètres libres des modèles de climat, souvent désignée comme "tuning" (ce terme pouvant être selon acception un peu plus large que la seule calibration des paramètres, et inclure par exemple des choix de grilles, d'activation de paramétrisations ou autres) a été identifiée dès les travaux pionniers de modélisation du climat. Il a pourtant fallu attendre la fin de la dernière décennie pour que cette question soit identifiée comme une question scientifique à part entière de la modélisation du climat. La conférence internationale que nous avons organisée avec Torsten Mauritsen et l'article dans BAMS ("The art and Science of climate model tuning") qui en a résulté a sans doute largement contribué cette évolution. A l'occasion de cette conférence, j'ai fait la rencontre de Daniel Williamson, mathématicien statisticien, qui promouvait une technique particulière d'apprentissage machine, appelée "history matching", pour la calibration des modèles de climat.

Cette approche se détache des approches classiques par minimisation de fonction coût (avec typiquement des calculs de descente de gradient), en refusant fondamentalement de voir le problème de la calibration comme un problème d'optimisation. Cette calibration se définit ici par l'identification du sous-espace des paramètres (l'espace initial étant défini a priori comme un hyper-cube dont les arrêtes sont définies par les valeurs min et max a priori de chaque paramètre) qu'on ne peut pas rejeter. Le rejet se fait sur la base de la distance d'un certain nombre d'observables calculées sur les sorties du modèle à des valeurs cibles, en tenant compte des différentes sources d'incertitudes :

incertitude sur les cibles des métriques (typiquement des observations ou des LES pour les tests de paramétrisations en mode uni-colonne), incertitude structurelle du modèle à calibrer, éventuellement une incertitude liée à la variabilité interne du système étudié. Pour explorer l'hypercube de dimensions P (le nombre de paramètres, typiquement de 10 à 30 dans les applications actuelles), on réalise $10P$ simulations, correspondant à $10P$ tirages différents des P paramètres libres, sur lesquelles on va calculer les M métriques. Une fois ces métriques calculées sur les simulations réellement effectuées avec le modèle, on développe pour chaque métrique un outil statistique appelé émulateur, capable de calculer en un temps extrêmement rapide une métrique en n'importe quel point de l'hypercube¹. Sur la base de cet émulateur, on décide de rejeter un jeu de paramètres si et seulement si la distance de la métrique émulée à la cible est plus grande que typiquement 3σ où σ est la somme quadratique des sources d'incertitudes, qui inclue l'incertitude de l'émulateur (en plus des sources précédemment mentionnées). L'espace des paramètres étant très vaste initialement, on procède à sa réduction itérative au cours de "vagues" successives. A chaque vague, on tire les nouvelles simulations dans l'espace NROY des paramètres non encore éliminés (Not Ruled Out Yet). Ceci permet de densifier petit à petit le nombre de simulations réellement réalisées dans l'espace des paramètres pertinents, et de réduire ainsi l'incertitude de l'émulateur qui devient au fur et à mesure de plus en plus prédictif (moins incertain). Il faut remarquer que les émulateurs utilisés, des processus gaussiens, sont choisis pour deux raisons qui les caractérisent : un apprentissage sur une base faible (très peu de vraies simulations au regard de la dimension de l'espace exploré) et l'accès à l'incertitude de l'estimateur (un processus aléatoire gaussien).

Cette rencontre avec Daniel Williamson a été le début d'une collaboration fructueuse, avec notamment le montage de l'ANR High-Tune (2016-2020). Cette ANR visait en particulier à passer un cap dans l'utilisation du cadre SCM/LES (comparaison de simulations effectuées avec une version uni-colonne et des LES prises comme référence) en développant des outils basés sur l'history matching pour permettre de re-régler automatiquement une version d'une paramétrisation après modification.

Le logiciel HighTune-Explorer (ou `htexplo`) développé lors de cette ANR, en partie lors de coding sprint collectifs avec Daniel Williamson, et qui s'appuie pour toute la partie statistique directement sur les outils développés à Exeter par DW et ses collègues, est aujourd'hui devenu un outil de base du développement des paramétrisations dans l'équipe. Avec au moins une dizaine d'utilisateurs qui utilisent cet outil au coeur de leur recherche dans l'équipe LMDZ ou pour les autres composantes du système climatique.

Les résultats principaux de cette ANR et la méthode de tuning ont été publiés dans un article en trois parties dont le titre générique a été un choix longuement discuté et réfléchi : "Physics based model development harnessing machine learning". Ce titre voulu avant tout par DW met en avant que c'est bien la physique qui tient les rennes de l'apprentissage machine. Bien sûr il y a des vraies questions de recherches sur les méthodes d'apprentissage utilisées. Elles tirent une grande partie de la recherche en mathématique de DW et de ses collègues. Cependant l'utilisation de ces méthodes est pensée au service de la modélisation physique du climat et de l'amélioration de paramétrisations basées sur la

1. Daniel Williamson, grand spécialiste de ces méthodes, à du mal à parler d'apprentissage machine parce qu'il dit que ce sont simplement des méthodes statistiques. Personnellement, apprentissage machine ça me parle. Mais je trouve très important de ne pas parler d'Intelligence Artificielle à propos de ces méthodes. J'ai peur malheureusement que le combat soit perdu, là comme sur d'autres terrains ... La précision du vocabulaire, voire sa définition, devrait pourtant être une priorité du monde universitaire auquel nous appartenons.

physique.

La première des trois publications est consacrée à la présentation de la méthode de l'history matching appliquée aux comparaisons SCM/LES. J'ai plus particulièrement coordonnée le travail et la rédaction de la seconde partie (l'article retenu ci-dessous). Il s'agit en grande partie d'une preuve de concept de la méthode en grandeur réel et de l'intérêt d'enchaîner une phase mutli-vague de tuning en mode SCM/LES (avec des simulations très peu coûteuses numériquement) avant d'effectuer le tuning complet du modèle global. Une dizaine à quinzaine de métriques ont été définies pour le mode SCM et environ autant pour le modèle global. Les métriques globales portaient sur les flux radiatifs au sommet de l'atmosphère. Leur choix s'inspiraient largement du tuning "à la main" effectuée précédemment pour la version LMDZ6A du modèle pour CMIP6. Ces métriques consistent en des moyennes sur de grandes régions (ou différences de moyennes entre deux régions) de flux au sommet de l'atmosphère, découpées suivant les cas entre rayonnement à source solaire (dit ondes courtes) et rayonnement à source terrestre (dit ondes longues), et entre ciel clair et effet radiatif des nuages. Le résultat sans doute le plus spectaculaire est d'avoir retrouvé par cette méthode dans une procédure quasi automatique, trois valeurs de paramètres avec une gamme possible relativement petites autour des valeurs sélectionnées à la main pour la version LMDZ6A, confirmant à la fois l'efficacité de la méthode et le caractère assez optimale du réglage à la main de ces trois paramètres pour CMIP6. Dans un second temps nous avons montré comment on pouvait retuner automatiquement une version du modèle après une amélioration de la discrétisation de l'épaisseur des mailles dans la couche limite (typiquement un raffinement d'un facteur 2).

L'utilisation de `htexplo` est devenu systématique dans tous les travaux de modélisation que j'encadre. Que ce soit en accompagnement du développement de nouvelles paramétrisations ou élément de paramétrisations. C'est peut-être le cas où on réalise le plus à quel point ne pas avoir un tel outil manquait. En effet, on était précédemment très régulièrement coincé dans le fait de savoir si une simulations avec une paramétrisation modifiée marchait mieux ou moins bien qu'une autre, ou si simplement l'une ou l'autre était mieux réglée. Or même avec 5 ou 6 paramètres l'espace à explorer pour connaitre les zones de "bon comportement" de la simulation modifiée ou de contrôle est gigantesque. Le fait de disposer d'un outil permettant une exploration automatique et à un coût raisonnable de tels espaces, permet de se libérer de cette question et refocaliser le travail sur les équations et l'amélioration du contenu physique du modèle. Concrètement, cette utilisation concerne le travail de développement mentionné précédemment dans les thèses de Louis d'Alençon et Lamine Thiam. Par ailleurs, nous explorons dans le cadre de la thèse de Maëlle Coulon-Decorzens comment exploiter les observations in situ comme celles du Sirta pour le tuning des modèles et selon quelle méthodologie.

Des travaux sont également en cours pour essayer d'étendre l'utilisation de l'history matching aux autres composantes du systèmes et au modèle couplé. Un des grands enjeux est d'identifier dans des simulations courtes des métriques qui nous informe sur les dérives lentes et les biais du système couplé une fois ramené à l'équilibre. Une des stratégies que j'ai identifiée et promue pour ce faire est de générer des ensembles paramétriques avec différentes configurations du modèle mais avec le même ensemble de vecteurs de paramètres au travers des configurations. De tels ensembles ont été générés pour le modèles couplés avec des configurations couplées à l'océan et forcées en SST à différentes résolutions, également en mode forcé avec l'ancienne dynamique longitude-latitude et la nouvelle dynamique icosahédrique du modèle. On peut alors faire des diagrammes de dispersion comparant une métrique d'une simulation forcée par les températures de surface de l'océan par exemple, avec

une métrique de température de surface océanique en couplé. C'est typiquement cette approche qui nous avait permis de démontrer l'importance du refroidissement évaporatif dans le biais chaud de bord Est mentionné plus haut, en utilisant les ensembles CMIP en comparant les simulations `historical` en climat actuel avec des simulations `amip` forcées par des températures de surface de l'océan.

Production scientifique 2 : Hourdin, F., Williamson, D., Rio, C., Couvreur, F., Roehrig, E., Villefranque, N., Musat, I., Fairhead, L., Diallo, F. B. and Volodina, V., Process-based climate model development harnessing machine learning : II. model calibration from single column to global, *James*, vol. 13, no. 6, 2021. doi :10.1029/2020MS002225.

3.4 Interactions nuages rayonnement

Un autre objectif important du projet HighTune était de combler le vide entre le tuning à l'échelle des processus, réalisé avec des comparaisons LES/CSM, et le tuning des modèles globaux visant souvent (directement ou indirectement) une bonne représentation des flux radiatifs au sommet de l'atmosphère, essentielle notamment pour la représentation des températures de surface et des grands éléments de la circulation générale atmosphérique. Or nous ne disposions pas avant le projet HighTune de simulations de référence du transfert radiatifs sur des cas LES. Il a fallu en fait attendre la rédaction du projet HighTune pour que nous réalisions à quel point les simulations LES étaient grossières quand elles représentaient les interactions entre nuages et rayonnement. En effet les codes utilisés alors, et encore aujourd'hui, dans les modèles LES ou les modèles à résolution kilométrique de la communauté, utilisent les mêmes codes que les modèles de climat. Ces codes font une hypothèse fondamentale : les photons sont piégée dans chaque colonne (empilement vertical de mailles) du modèle, n'échangeant aucun photon avec les colonnes voisines. Cette hypothèse, valide pour une maille horizontale de quelques dizaines de km de côté, devient absurde avec des mailles de 20m quand on essaie d'étudier les couplages entre rayonnement et nuages à l'échelle du nuage.

Une idée du projet était donc de se doter de calculs de rayonnement de référence sur les LES en utilisant la méthode de référence du rayonnement, à savoir la méthode Monte Carlo en "chemins". Cette méthode qui a une portée beaucoup plus générale (cf. plus bas), peut être décrite, pour le rayonnement, par analogie comme une méthode de suivi de photons. Cette méthode est celle qui est utilisée aujourd'hui dans l'industrie cinématographique de la synthèse d'image. Ce travail a été mené en collaboration avec le collectif de recherche EDstar, qui porte depuis une vingtaine d'années une recherche de pointe sur les méthodes de MonteCarlo pour des questions principalement de physique de l'ingénieur, pour beaucoup liées à la transition énergétique. Ce travail a été porté par Najda Villefranque pendant sa thèse à cheval entre le CNRM et le Laplace (laboratoire pilote du consortium EDstar). Un des gros enjeu était le coût informatique des calculs radiatifs au travers des milieux semi-transparentes et très diffusants que sont les nuages. Najda Villefranque a étendu à ces milieux dits "participants" des techniques d'accélération de la synthèse d'image, reprises depuis dans les recherches des laboratoires de l'industrie cinématographique. Ceci permet que la communauté du climat dispose aujourd'hui d'un logiciel `htexplo` (ou Hightune Explorer) pour à la fois réaliser des calculs de référence de rayonnement et un rendu physiquement réaliste de scènes LES.

Grâce à cet outil, il a été possible de réaliser des calculs de référence pour toute une série de scènes LES de nuages de couches limites du type de celles utilisées pour le tuning des nuages et de la couche limite en mode SCM/LES. Ces calculs de référence ont permis d'évaluer et ajuster les paramètres

libres de différentes approches proposées aujourd'hui dans le code ECrad du centre Européen pour les Prévisions Météorologiques à Moyen Terme (porté par Robin Hoggan) pour représenter les hétérogénéités nuageuses (méthode MCICA ou Tripleclouds) ou les effets 3D (Spartacus). Ce travail a notamment permis de montrer l'importance de prendre en compte les effets 3D du rayonnement (au travers de la version TripleClouds-Spartacus de ECrad) qui montre notamment une bien moindre dépendance aux angles d'incidence solaire des erreurs sur l'estimation des flux radiatifs.

Production scientifique 3 : Villefranque, N., Blanco, S., Couvreur, F., Fournier, R., Gautrais, J., Hogan, R. J., Hourdin, F., Volodina, V. and Williamson, D., Process-based climate model development harnessing machine learning : I. The Representation of Cumulus Geometry and Their 3D Radiative Effects, *James*, vol. 13, no. 4, 2021. doi :10.1029/2020MS002423.

Ce travail est poursuivi en ce moment dans une partie de la thèse de Maëlle Coulon-Decorzens, co-encadrée avec Najda Villefranque pour étudier notamment les compensations entre nuages et rayonnement, en effectuant un tuning sur la base de cibles purement radiatives. On retrouve dans cette étude l'apport de la représentation des effets 3D par Spartacus. Avec Spartacus en effet, on arrive via des cibles radiatives à ajuster très correctement les profils verticaux de fraction nuageuse et contenus en eau des nuages pour des cas de cumulus et strato-cumulus. L'absence de l'activation de Spartacus aboutit au contraire à la nécessité d'utiliser des fractions nuageuses plus grandes, notamment pour compenser la sous-estimation des effets d'ombres aux angles solaires rasants. C'est la première fois qu'on démontre la possibilité d'une compensation d'une erreur dans un code radiatif par une détérioration de la représentation des nuages, suggérée par le passé dans des articles portant notamment sur une tendance assez systématique des modèles de climat à représenter des nuages trop peu couvrants et trop brillants (too few to bright).

3.5 Anticipation du changement climatique et quantification des incertitudes

La question de l'incertitudes des projections climatiques, tant en termes d'amplitude du réchauffement global que de conséquences climatiques est un des moteurs des projets d'intercomparaisons CMIP organisés par la communauté scientifique en amont des rapports du Giec, et dans lesquels tous les modèles partagent des protocoles extrêmement précis pour pouvoir documenter les différentes sources d'incertitude. Les incertitudes dans les projections climatiques peuvent être décomposées en trois contributions dues (a) à la variabilité interne - définie comme la variabilité intrinsèque non forcée du système climatique -, (b) aux incertitudes dans la physique du modèle - résultant notamment de la paramétrisation des processus sous-maille, et (c) à la gamme des forçages externes futurs possibles, qui dépendent à leur tour des scénarios d'émission et des forçages naturels. Le point (c) est l'objet de la constructions de "scenarios" en émissions ou en concentrations, développés dans d'autres communautés et utilisés en entrée des simulations climatiques, notamment à l'occasion des exercices CMIP. Durant le dernier exercice je me suis plus particulièrement impliqué dans la documentation où l'étude des incertitudes liées à la variabilité interne (a), et la quantification de la contribution de l'incertitude paramétriques à l'incertitude modèle (b).

La variabilité interne ne rentre pas en compte fondamentalement dans la question de la sensibilité climatique. Si on considère par exemple la "Sensibilité Climatique à l'Equilibre" ECS, définie comme le réchauffement global de la terre une fois revenue à l'équilibre après un doublement de la concentration en concentration de gaz carbonique dans l'atmosphère, on voit bien que cette définition est pensée

comme la différence entre deux états climatiques "moyens". Avec un modèle de climat couplé, on peut obtenir une estimation de cet ECS en réalisant une simulation suffisamment longue du climat avec une certaine concentration des gaz à effet de serre, puis une seconde dans laquelle on multiplie la concentration de CO₂ par deux. Il faut typiquement attendre une 30 aine d'années pour l'ajustement des température de surface de l'océan, mais l'océan profond vient se réchauffer encore pendant des centaines d'années. Avec le modèle, il suffit que les simulations soient suffisamment longues pour à la fois absorber ce spin-up de la simulation perturbée et gommer la variabilité interne. Si la variabilité n'intervient donc pas dans la définition du changement climatique une fois le système revenu à l'équilibre, elle peut moduler fortement la réponse des prochaines décennies. Elle est également très importante quand on cherche à évaluer la capacité des modèles à prévoir le climat futur sur la base de la capacité de ces modèles à reproduire l'évolution du climat sur les dernières décennies. Cet enjeu conduit les groupes de modélisation à réaliser de façon de plus en plus systématique des ensembles de simulations "historiques" (depuis le préindustriel jusqu'à l'actuel) ou futures, en ne changeant que les conditions initiales (ensembles perturbés aux conditions initiales).

A l'occasion du dernier exercice CMIP6, L'IPSL a ainsi produit un ensemble de simulations historiques étendues à l'aide du modèle climatique IPSL-CM6A-LR. Cet ensemble (appelé IPSL-EHS) est composé de 32 membres sur la période 1850-2059 qui partagent les mêmes forçages externes mais qui diffèrent par leurs conditions initiales. Cet ensemble a été utilisé pour évaluer la variabilité climatique simulée entre la décennie et la multidécennie dans l'ensemble IPSL-EHS. En particulier, nous avons examiné l'évolution de la température globale et les tendances récentes du réchauffement, ainsi que leur cohérence avec le contenu thermique des océans et la couverture de glace de mer. Le modèle présente une importante variabilité climatique interne à basse fréquence. En particulier, un mode quasi-bicentenaire de variabilité climatique interne est présent dans le modèle et est associé à la circulation méridienne de retournement de l'Atlantique. Cette variabilité module les changements de la température moyenne globale de l'air en surface sur la période historique d'environ 0,1K. Cette modulation est liée à la phase présente dans l'état initial de chaque membre du modèle. Dans le choix des 32 états initiaux de l'ensemble, on est parti de points régulièrement espacés dans une période de cette oscillations. Avec une période de 200 ans, le signal d'évolution vers le chaud ou le froid sur les 100 dernières années peut venir réduire ou amplifier le réchauffement global, avec typiquement 0.2 degrés de différence entre les simulations qui se réchauffent le plus ou le moins. Parmi les 32 simulations nous avons pu montrer que 3 présentaient un réchauffement global tout à fait compatible avec les observations, alors même que le modèle de l'IPSL est considéré comme ayant une ECS à la limite haute des valeurs admissibles (on y revient plus loin). Ces résultats sont présentés dans l'article Bonnet et al.. Nous montrons que les membres ayant les taux les plus faibles de réchauffement global au cours des 6-7 dernières décennies sont également ceux qui présentent un affaiblissement important de la circulation méridienne de retournement de l'Atlantique (AMOC). Ce sous-ensemble de membres correspond semble être en accord avec plusieurs signatures observationnelles de l'évolution de l'AMOC, qui vont dans le sens d'un tel affaiblissement. Cela suggère que la variabilité interne de l'océan Atlantique a pu atténuer l'ampleur du réchauffement climatique au cours de la période 1900-2020. La prise en compte de cet affaiblissement de l'AMOC au cours des dernières décennies pourrait avoir ainsi masqué une partie du réchauffement global récent. Ce résultat est un peu une alerte face aux études qui ont essayé de réduire, sur une base partiellement observationnelle, la plage d'ECS admissibles.

Production scientifique 4 : Rémy Bonnet, Didier Swingedouw, Guillaume Gastineau, Olivier Boucher, Julie Dehayes, Frédéric Hourdin, Juliette Mignot, Jérôme servonnat, Adriana Sima Increased risk of near term global warming due to a recent AMOC weakening , Nature, <https://doi.org/10.1038/s41467-021-26370-0>

3.6 Quantification de l'incertitude paramétrique sur la sensibilité climatique

Documenter l'incertitude modèle dans les projections du changement climatique est un objectif fondamental des exercices CMIP. En comparant la réponse différente des différents modèles pour un même scénario d'évolution des émissions ou de concentration des gaz à effet de serre, on isole effectivement la part d'incertitude liée aux choix de modélisation du climat faits dans les différents groupes. Beaucoup d'études ont montré par le passé qu'une grande partie de la dispersion en termes de réchauffement global, pour un même scénario, était due à la paramétrisation de la convection et des nuages. Or, habituellement, chaque centre de modélisation contribue à ces exercices avec une ou deux configurations de son modèle climatique, correspondant à un choix particulier de valeurs de paramètres libres, résultant d'une phase longue et souvent fastidieuse de réglage du modèle (tuning, calibration, cf. plus haut). Quelle est la part d'incertitude omise par cette sélection et comment les lecteurs des rapports du GIEC et les utilisateurs des projections climatiques peuvent-ils être induits en erreur par cette omission ? Une part de la réponse à ces questions est apportée par un autre type d'ensembles appelés "Ensembles aux Paramètres Perturbés" ou PPE. On effectue typiquement des dizaines ou centaines de simulations en perturbant un certain nombre des paramètres libres, aléatoirement ou pas. Cependant, ces simulations n'étant pas retunées, on explore des climats simulés potentiellement assez éloignés des observations, en termes d'équilibre énergétique global (ou de température globale en mode couplé) ou de structures spatiales.

Nous avons mené une étude pour répondre spécifiquement à la question de savoir dans quelle mesure un modèle de climat particulier, ici le modèle couplé de l'IPSL, pouvait produire des sensibilités climatiques très différentes (ECS), tout en conservant une représentation raisonnable du climat actuel. Pour ce faire, nous avons effectué 3 vagues d'history matching avec la configuration standard du modèle (6A, avec 144 points en longitude, 142 en latitude et 79 niveaux) utilisée pour CMIP6 (3 fois 300 simulations de 2 ans environ), en réduisant petit à petit l'espace des paramètres (contrairement à l'étude décrite plus haut, on n'a pas utilisé le pré-conditionnement 1D du tuning dans ce papier). Par cette technique, on génère également des PPEs mais sous contrainte d'un accord aux données "historique" (d'où le nom d'history matching tire son origine). Parmi les 900 simulations, une quinzaine présentait des scores comparables à la version 6A sur les métriques sélectionnées. Pour ces simulations, nous avons effectué des simulations dites SST+4K, en ajoutant 4 degrés uniformément à la température de surface de l'océan. Dans ces simulations, on force une évolution de la température de surface sans changer la concentration en CO₂. L'intensité du changement d'équilibre énergétique global au sommet de l'atmosphère qui en résulte permet d'avoir une estimation grossière de l'ECS. Cette estimation a été utilisée ici pour sélectionner quatre jeux de paramètres libres parmi ceux qui produisaient les meilleures simulations (en termes de métriques climatiques pour le climat actuel, comme le font les groupes de modélisation quand ils sélectionnent les versions de leur modèle pour les exercices CMIP) et en privilégiant les simulations avec des estimation de l'ECS les plus faibles et les plus fortes. Avec ces jeux de paramètres, nous avons réalisé des simulations couplées

préindustrielles (dans lesquelles les concentrations de gaz à effet de serre sont gardés constant, à leur valeur préindustrielle) et abrupt4CO₂, dans laquelle on multiplie subitement la concentration du CO₂ par 4. La relation souvent quasi linéaire entre le retour à l'équilibre du bilan radiatif net au sommet et le changement de température globale dans ces simulations est la façon de référence pour calculer l'ECS (méthode dite de Gregory). On estime l'ECS à partir de la droite de régression de cette relation, quand le bilan radiatif atteint zéro (ici le changement de température doit être divisé par deux car on effectue un quadruplement plutôt qu'un doublement de la concentration de CO₂, afin de maximiser le rapport signal sur variabilité interne).

La fourchette d'ECS était de 1,8 à 5,6 K dans l'ensemble multi-modèle CMIP6, contre 2,1 à 4,7 K dans CMIP5 (non loin de l'estimation de 1,5 à 4,5 K du rapport Charney en 1979). Avec la configuration standard IPSLCM6A et les 4 nouvelles configurations sélectionnées selon la procédure semi-automatique décrite au-dessus et détaillée dans un article, on obtient des ECS variant de 3,7 à 5,3 K pour une valeur de 4,6 K pour la configuration IPSLCM6A-LR officielle. Parmi ces 4 configurations, au moins deux auraient sans doute pu être mises indifféremment dans la base de données CMIP6 à la place de celles de la configuration IPSLCM6A-LR, tant leurs résultats sont comparable au regard des métriques retenues. Il se trouve qu'il s'agit des deux simulations avec les ECS de 3,7 et 5,3 K. Il est à noter que malgré la différence de sensibilité, les différentes configurations simulent des cartes de réchauffement ou de changement des précipitations en moyenne sur l'année très similaires, à condition qu'on les regarde non pas à un horizon temporel donné mais à un changement de température global donné (qui arrive plus ou moins tôt dans la phase transitoire en fonction de la valeur de l'ECS). Cette séparation entre une forte sensibilité aux paramètres de l'ECS, et une relative insensibilité de la structure des changements climatiques à un degré de réchauffement connu, en partie présente dans les ensembles multi-modèles CMIP est très intéressante d'un point de vue de l'éclairage des questions de société.

Il est sans doute dommage qu'on n'arrive pas à dégager davantage de temps pour comprendre les mécanismes qui amènent à une telle dispersion des ECS juste par ajustement des paramètres libre, sous contrainte d'un réajustement des flux radiatifs au sommet.

Ces résultats ont été publiés dans *Science Advances*. Au delà du fait d'avoir fourni un élément (évidemment très partiel) de quantification de l'incertitude paramétrique pour un modèle particulier, pouvant suggérer une sousestimation de l'incertitude de l'ECS dans un modèle global. On voit ici l'importance d'avoir défini le tuning non pas comme un acte d'optimisation mais comme la détermination de l'espace de paramètres compatible avec un certain nombre de cibles à une tolérance près. On montre dans cette publication comme ce changement de paradigme et la proposition particulière de l'history matching permet d'effectuer dans un même acte une calibration semi-automatique de paramètres libres et une quantification des incertitudes paramétriques associées.

Cette étude nous incite fortement à systématiser la réalisation des petits PPEs sous contrainte (petits du fait de leur coût numérique) dans l'avenir pour quantifier systématiquement l'incertitude paramétrique de la réponse à l'augmentation du CO₂ dans les futures versions du modèle.

À noter que cette approche est une alternative sans doute mieux posée à l'approche des contraintes émergentes qui consiste à identifier des relations entre des propriétés des simulations climatiques sur la période passée et des tendances climatiques, pour sélectionner ou pondérer, dans les projections, les modèles qui représentent le mieux la période historique.

Production scientifique 5 : Hourdin, F , Ferster, B., Deshayes, J., Mignot, J., Musat, I.

and Williamson, D., Toward machine-assisted tuning avoiding the underestimation of uncertainty in climate change projections Toward machine-assisted tuning avoiding the underestimation of uncertainty in climate change projections, Science Advances, 2023, Vol 9, Issue 29, DOI : 10.1126/sciadv.adf2758.

3.7 Calculs sonde Monte Carlo

Le projet HighTune et la thèse de Najda Villefranque ont permis de renforcer des liens historiques entre le collectif EDstar et l'équipe de modélisation du climat du LMD. Jean-Yves Grandpeix, Jean-Louis Dufresnes et Alain Lahellec ont en fait formé il y a 25 ans Richard Fournier et Stéphane Blanco dans un laboratoire de thermique de l'habitat, le LESETH, avant de rejoindre le LMD dans les années 90.

Le collectif EDstar a ces 20 dernières années poussé toujours plus loin les possibilités d'utilisation des méthodes de Monte Carlo pour la modélisation physique, avec un lien très forts avec les calculs en intégrales de chemins et la pensée de Green qui ont été à l'origine du prix Nobel de Feynman et Kac. En fait cette pensée est déjà présente dans des travaux de modélisation menés dans l'équipe : la formulation des transferts radiatifs en puissances nettes échangées qu'on avait utilisée pour paramétriser le transfert radiatif dans des atmosphères d'atmosphères planétaires (avec notamment Jean-Louis Dufresne et des chercheurs aujourd'hui dans le collectif EDstar). Dans cette approche (appliquée ici au besoin d'un modèle de circulation générale), on voit le taux de chauffage dans chaque couche du modèle comme la somme d'échanges nets avec chacune des autres couches. Cette puissance nette échangée peut être fondamentalement exprimée comme le produit la différence des fonctions de Planck calculées à la température des deux couches, et un coefficient d'échange. Ce coefficient d'échange ne dépend que de l'atténuation calculée le long de l'ensemble des chemins de photons possibles entre les deux couches. Si les propriétés des couches sont supposées connues une fois pour toute (ce qui est loin d'être possible pour les nuages sur Terre), on peut calculer ce coefficient d'échange une fois pour toute en tirant par exemple aléatoirement un grand nombre de photons entre les deux couches. Cette symétrie du transfert radiatif est celle qui est utilisée dans la synthèse d'image pour tirer des photons non pas depuis les sources (le soleil ou des ampoules par exemple) mais depuis les pixels de la caméra.

J'avais également travaillé avec Olivier Talagrand et Jean-Pierre Issartel (CEA) sur la détection des essais nucléaires à partir de radio-éléments transportés dans l'atmosphère et mesurés à des stations. Nous avons alors réalisé que les calculs pouvaient aussi bien se faire par transport direct en émettant un traceur depuis le lieu connu d'une source, et mesurer la concentration aux stations, ou, de façon symétrique, en émettant les radio éléments aux stations et en les mesurant aux lieux possibles d'émission avec un transport inverse. A noter que lors de ce transport inverse, calculé en remontant des trajectoires lagrangiennes dans le temps, les concentrations vont diffuser sous l'effet de la dispersion des trajectoires fluides, comme lors du transport direct. On peut montrer que l'échange entre sources possibles et stations peut se voir comme un coefficient d'échange qu'on peut exprimer également comme une intégrale de chemins (les "trajectoires" fluides lagrangiennes). On peut également montrer que la description eulérienne du même transport, appliquée au transport inverse, aboutit à une équation similaire à l'équation directe, avec un changement du signe du temps d'intégration, du vent (on remonte les trajectoires) mais le même coefficient de diffusion sans changement de signe (même intensité de la diffusion dans le passé et le futur). On peut également montrer que l'équation inverse

est en fait l'adjoint de l'équation directe pour un produit scalaire particulier. Dans le cas de la détection des essais nucléaires, on voit bien l'intérêt de l'approche inverse. Si des stations détectent un essai nucléaire, il faudrait avec la méthode directe tester des émissions de concentration en tout points de la grille et les transporter jusqu'au détecteur. Avec le rétro transport, on émet le traceurs aux (80) stations du réseau et on obtient en une simulation pour chaque stations sa "visibilité", à savoir l'ensemble des coefficients d'échange avec toutes les sources potentielles dans le passé et sur l'ensemble du globe.

Cette idée de calcul sonde est très importante dans l'utilisation des reformulations en chemins promue par le réseau EDstar. On va chercher typiquement à évaluer la température d'une pièce à un instant donné. Si on veut effectuer un calcul direct de cette température, il faudra connaître à un certain moment l'état de l'environnement du bâtiment et du bâtiment lui même, puis effectuer un calcul direct pour finalement obtenir la température de la pièce. Beaucoup de calculs seront inutiles dans ce calcul direct si c'est bien la seule température de la pièce qu'on vise. Grâce à la reformulation "en chemin" de la physique des échanges conductifs, radiatifs et convectifs (dans le vocabulaire des sciences de l'ingénieur, on appelle transport convectif tout ce qui est lié au transport par l'air), on peut calculer la température de la pièce en remontant des chemins dans le passé. Dans cette approche, la température est exprimée comme la moyenne sur un grand nombre de réalisations passées, de "chemins" qui peuvent être soit radiatifs, soit convectifs, soit conductifs.

Par exemple, la température de l'air de la pièce (supposée uniforme pour simplifier l'exposé) sera influencée par la température des planchers et murs. Si les 6 surfaces (4 murs, plancher et plafond) ont la même aire, et dans un cas stationnaire, la température de la pièce sera la moyenne de la température des 6 parois, qui pourra être calculée par une grand nombre de chemins en en tirant 1/6 sur chacune des parois. Si les parois sont chacune à température uniforme, la température finale de l'air dans la pièce sera la moyenne arithmétique de la température des parois rencontrées sur l'ensemble des tirages. Si la température des murs est connue, le calcul est fini et donne une estimation non biaisée de la température et dont l'erreur statistique gaussienne décroît en $1/\sqrt{N}$ où N est le nombre de tirages. Dans le cas où les températures des parois ne sont pas connues, on peut aussi les calculer comme des moyennes. Si on se donne par exemple pour modèle que la température est la même des deux côtés du mur, on voit bien que la température de ce mur elle même va être la température moyenne entre l'air des deux pièces, si on ne prend en compte que les échanges convectifs. Si on prend en compte en plus des échanges radiatifs entre les différents murs, alors la moyenne fera intervenir la température de l'air des deux pièces et la température des 10 autres parois des deux pièces, mais pourra toujours s'exprimer comme une somme pondérée des différentes températures qui pourra donc à nouveau s'estimer comme un tirage aléatoire.

Une chose importante, c'est qu'on n'est pas obligé de faire converger l'estimation statistique de la température d'un mur autant que celle de la cible qu'on visait ici : la température de l'air d'une pièce. En pratique, pour un tirage d'un mur, on peut ne tirer qu'une qu'une réalisation de la température des autres pièce ou autres mur, et ainsi de suite. Cette propriété baptisée "double randomisation" dans la communauté EDstar, dans laquelle dans ce cas simple, on ne tire qu'une réalisation de la variable aléatoire quand on tombe sur une nouvelle variable d'état du système non prescrite (qui doit en théorie être calculée elle aussi comme une espérance d'une variable aléatoire), définit des chemins qui vont passer de pièce en pièce, possiblement revenir dans la même, puis finir par sortir de la maison jusqu'à s'arrêter à un point où la température est connue (par exemple la température du sol et de l'air autour de la maison si on les impose). La température de la pièce s'exprimera finalement comme la moyenne sur l'ensemble des chemins des températures au bout des chemins, appelées issues.

La méthode s'étend facilement à un échange avec un gaz participatif ou à la conduction dans un mur qu'on modéliserait dans son épaisseur en tenant compte de variations de température. Dans le cas de la conduction, l'approche s'étend également à des calculs non stationnaires, le suivi des chemins étant alors accompagné d'une remontée du temps dans le passé, jusqu'à des conditions initiales où les températures sont connues.

La double randomisation peut être vue comme découlant directement du fait qu'on peut permuter deux espérances ou deux intégrales quand les grandeurs à calculer sont linéaires. Cette remontée en chemin, où on a supposé ici des pièces brassées et des murs isothermes, peut très bien se faire dans un cas plus général sans avoir besoin de prédéfinir un maillage. Dans le cas de la pièce brassée, on peut en effet calculer la température de la pièce comme la moyenne des températures en tous points des murs, et donc la calculer par un triage de positions au hasards sur tous les murs. La température résultante sera simplement la moyenne de la température obtenue pour chacun des tirages.

Le cas linéaire recouvre déjà un grand nombre d'applications pertinents. C'est le cas de la synthèse d'image, de la conduction thermique avec des coefficients supposés connus. Dans un article que nous avons porté avec Najda Villefranque comme une perspective dans Science Advances pour essayer de donner plus de lumière aux résultats des recherches du consortium EDstar, nous avons proposé un exemple pédagogique très proche de la description intuitive donnée ci-dessous.

Ces idées font la base du travail du consortium EDstar depuis une vingtaine d'années à savoir : la possibilité de réécrire certaines physiques et leurs couplages en espaces de chemins ; le calcul sonde qui économise potentiellement énormément de calcul en permettant d'estimer une grandeur cible sans avoir besoin de calculer l'ensemble du champ (de température dans l'exemple de la température de la pièce, de concentration pour la détection des essais nucléaires) ; la double randomisation valable dans le cas linéaire ; l'intégration sans maillage.

L'intégration sans maillage permet d'aborder des géométrie d'une complexité quasi infinie. C'est elle qui a fait basculer en dix ans l'industrie du cinéma d'animation vers le rendu physiquement réaliste par lancé de rayon (ou en suivi de chemins dans le vocabulaire EDstar). Dans un milieu non participatif (pour simplifier), les photons vont rétro partir de la caméra pour taper une surface. Celle de la peau d'un personnage par exemple. La description de la peau aura pu être définie indépendamment par le créateur de l'animation, jusqu'à décrire les poils ou les pores de la peau. La caméra pourra donc tenir compte de cette information, définie totalement indépendamment de la nature du rendu. Cet objectif de pouvoir rendre de grandes images qui tiennent le compte du plus petit détail, cette capacité à traverser la complexité spatiale, sont connus dans le monde de l'informatique graphique comme le problème de la théière dans un stade de foot ou "teapot in the stadium". Avec cette technique, l'enjeu principal en terme d'efficacité de calcul se situe dans l'accès à la donnée de la scène, au point d'intersection avec un lancer de rayon. C'est en faisant tomber des verrous informatiques en termes d'accès à la donnée que l'industrie de l'animation a pu basculer sur le rendu physiquement réaliste par lancer de rayons. La thèse de Najda Villefranque a permis d'étendre ses techniques d'accélération d'accès à la données à ces milieux 3D participatifs, hétérogènes et fortement diffusant que sont les nuages.

Nous avons dans le papier montré une illustration de la puissance de la méthode en réalisant un film entièrement calculatoire de la plongée d'une caméra (on peut penser qu'elle est à bord d'un avion en train d'atterrir) sous une couche de cumulus, et arrivant sur des villes elles aussi créées par un modèle, pour finir par dévoiler une théière sur une table dans un jardin, présente dès le début du film. En est issu le titre du film repris dans le titre du papier : The teapot in the city.

Dans l'article, nous montrons comment la dimension statistique et le côté calcul sonde permettent

de penser d'une façon complètement nouvelle les calculs d'impact du changement climatique. En effet, on peut dans un même calcul, plutôt que de calculer la température à un instant donné dans une pièce (comme on l'expliquait plus haut dans un cas simple), la calculer en moyenne sur toutes les pièces d'un immeuble, en moyenne sur 30 ans sur les mois d'été, et même en moyenne sur un ensemble de simulations du changement climatique si on veut avoir une réponse moyenne multi-modèle. De façon assez systématique, dans le cas de calculs linéaires, ajouter d'autres dimensions statistiques dans ce calcul n'alourdit pas le temps de calcul. Dans l'article, on montre des exemples de l'évolution de la température ou de la consommation de la climatisation de bâtiments au Sahel, plongés dans un climat changeant.

Dans le cas où des non linéarités apparaissent (par exemple si on a dans le bâtiment un système de chauffage réversible qu'on doit prendre en compte pour calculer la température), le calcul devient souvent plus compliqué et plus coûteux. Trouver des solutions à ce type de questions est une des activités fortes du collectif EDstar. Très souvent des solutions existent.

Production scientifique 6 : Villefranque, Najda ; Hourdin, Frédéric ; d'Alençon, Louis ; Blanco, Stéphane ; Boucher, Olivier ; Caliot, Cyril ; Coustet, Christophe ; Dauchet, Jérémie ; El Hafi, Mouna ; Farges, Olivier ; Forest, Vincent ; Fournier, Richard ; Masson, Valéry ; Piaud, Benjamin ; Schoetter, Robert The "teapot in a city" : a paradigm shift in urban climate modeling, Science advances, <https://www.science.org/doi/10.1126/sciadv.abp8934>

Sur la base de ces idées, nous avons obtenu sous la direction de Cyril Caillot une ANR MC2 (on est à mi parcours) sur la ville dont je coordonne une partie. Dans cette partie, on veut étendre les calculs de l'article en passant à une géométrie de ville réaliste (ce qui ne pose pas de problème) et également intégrer le calcul de l'îlot de chaleur urbain dans nos estimations. En effet, dans les résultats publiés dans l'article ci-dessus, on suppose que les températures et que le rayonnement sont imposés. En particulier, la ville ou le bâtiment sont plongés dans un air dont la température est imposée. L'étape suivante (en cours de travail) est de supposer que la température de l'air elle-même est inconnue. On va suivre des chemins qui remontent depuis la surface du bâtiments dans les premières couches atmosphériques, et être en même temps rétro-transportées horizontalement, jusqu'à atteindre l'extérieur de la ville, où on supposera la température imposée par la simulation climatique globale ou régionale. Cette question est travaillée dans le cadre de la thèse de Louis d'Alençon que je co-encadre avec Stéphane Blanco (EDstar).

Dans la suite de MC2, le réseau EDstar a obtenu une autre ANR (MCmet, qui commence tout juste), davantage sur les méthodes et dans laquelle je ne joue pas de rôle de coordination. Nous avons enfin soumis récemment une réponse à un appel d'offre du PEPR Ville Durable et Bâtiment Innovant, en cours d'évaluation.

4 Enseignement, formation et diffusion de la culture scientifique

Depuis une quinzaine d'années, je consacre une partie significative de mon énergie à l'enseignement de la modélisation du climat par différents biais et à différents niveaux. J'ai en fait réalisé que la modélisation du climat s'apprenait sur le tas. Sans réelle formation. Les enjeux de cet enseignement sont pourtant multiples. D'abord pour un souci d'efficacité pour les collègues. Ensuite parce qu'un minimum d'appropriation de ce qu'est la modélisation du climat me semble essentiel à plein d'endroits, dans des domaines de recherches connexes (ceux qui utilisent les sorties de modèles

pour éclairer les nécessaires choix de société), mais aussi dans des enseignements plus généralistes. Une autre raison pour consacrer de l'énergie à l'enseignement de la modélisation, c'est de construire la science de la modélisation. Je pense que si cette science était davantage construite, il n'y aurait pas eu de place pour la polémique sur les jumeaux numériques (basés sur des simulations globales à maille kilométrique).

4.1 M2

Je porte depuis 2010 le module d'initiation à la modélisation du M2 Mocis et Wape réunis. Suivant les années, ce module se déroule sur 8 à 10 demi journées. Il réunit une vingtaine à une trentaine d'étudiant.

Dans la première moitié du cours, je présente les grand principe de la modélisation, avec ses 5 couches hiérarchiques. Cette partie comprend une grande part de travail pratique dans laquelle les étudiants développent en fortran et pas à pas, en codant tout, des éléments d'un modèle d'advection diffusion d'un panache de cheminée. C'est l'occasion d'une initiation aux questions numériques, mais aussi aux grands principes qui sous-tendent la modélisation du climat : séparation entre dynamique 3D et physique 1D, décomposition de Reynolds comme cadre de la séparation d'échelle, différentes façons d'aborder le transport par les mouvements sous-maille au travers de paramétrisations.

Dans la seconde partie, les étudiants se regroupent par 2 à 4 pour mener des mini projets en utilisant les modèles de la communauté. Depuis le début, cet enseignement a été pour moi un moteur de l'automatisation du déploiement de configurations légères du codes LMDZ, typiquement sur des ordinateurs portables. J'avais par ailleurs fait acheter par l'IPSL (sur le budget du Labex) et par une ANR (convergence) une quarantaine d'ordinateurs portables, qu'on distribue depuis 8 ans à tous les étudiants du cours qui le souhaitent, et qu'ils utilisent largement pour les autres cours de janvier à mars. Depuis 4 ou 5 ans, j'ai proposé à Sébastien Masson (Locéan) de compléter l'offre de mini projets LMDZ avec une offre équivalente basée sur le modèle Nemo.

Du côté atmosphérique, l'intérêt des mini projets se renforce d'année en année, comme la diversité des sujets. Certains sujets vont mener de front, sur une question (comme le rôle de la convection et les nuages de couche limite sur les biais chauds de bords Est des océans tropicaux) des simulations 1D, 3D forcées en SST, 3D couplées à un océan "slab" (à une couche). D'autres vont repartir d'une maquette vide de la physique de LMDZ pour coder et tester des paramétrisation de transport sous-maille en flux de masse, en diffusion turbulente, des modèles de conduction dans le sol ou des schémas statistiques de nuages.

4.2 Cours de thèse

Dans la même veine, nous avons monté depuis une dizaine d'années, une école d'une semaine à destination des étudiants de thèse principalement. Cette école réunis des enseignants des Math appliquées et des sciences du climat : Guillaume Roulet, Laurent Debreu, Thomas Dubos et moi même. Elle tourne selon les années entre Grenoble, Brest et Paris, les lieux de travail des trois enseignants. Là aussi nous voyons passer une trentaine d'étudiants chaque année, originaires comme nous soit de la communauté climatique soit des mathématiques appliquées.

4.3 Le travail en ateliers

Au delà de situations d'enseignement à proprement parler, je ressens très fort et depuis longtemps un besoin de transmission aux collègues. C'est une des motivations (pas la seule) qui nous a poussé à travailler de plus en plus souvent en mode ateliers. On se met à 4 à 10. On reprend des équations au tableau. On code. Avec différents niveaux de connaissance sur le sujet ou l'atelier, mais avec l'idée qu'aucune des personnes présentes ne doit être perdue à un moment, ce qui implique bien sûr que seules les personnes motivées sont là. Cette pratique est notamment très mise en avant dans le cadre de semaines en résidence du collectif EDstar, une semaine par an près du Puy-en-Velay, sur le site de Cap Saint Front dont je reparle plus bas. Najda Villefranque (de par sa thèse à cheval entre ce collectif et la communauté climat) et moi même participons régulièrement à cette semaine. Nous y avons amené des collègues (Daniel Williamson, Fleur Couvreur, Jean-Batpiste Madeleine, Etienne Vignon, Fleur Couvreur) avec l'objectif à la fois d'échanger scientifiquement entre les deux communautés mais aussi de pouvoir réfléchir collectivement à la pertinence d'un tel format pour le collectif Dephy. Après une première expérience en 2022 à Oléron, où on avait alterné ateliers et présentations, il a été décidé d'organiser l'année suivante (en 2023 à Fréjus) une semaine Dephy entièrement en mode atelier. Et cet atelier réunissant une bonne trentaine de collègues a été un succès, remplissant toutes ses attentes en termes d'échanges scientifiques, de travail collectif et de transmission. Il est apparu à quasiment tous les participants que ces semaines en résidence, sous forme d'ateliers, devaient devenir annuelles. La prochaine édition en mai 2024 regroupe 42 inscrit.e.s.

4.4 L'écriture de textes pour l'enseignement

En lien avec ce sentiment d'un manque de construction de ce qu'est la modélisation du climat, et d'un manque de supports pour l'enseigner, j'ai été impliqué dans l'écriture de plusieurs textes. J'ai notamment participé à l'écriture d'un volume de livre de cours sur la modélisation du climat piloté par l'IPSL. J'ai passé aussi beaucoup d'énergie à écrire un chapitre en 30 pages pour l'encyclopédie Universalis, devenue pour l'essentiel un support numérique, avec comme cible typique (identifiée par l'entreprise) des lycéens de filières scientifiques. Quand on m'a proposé le projet, j'ai demandé qu'y soit associée Hélène Guillemot, historienne des sciences, qui a notamment soutenu une thèse sur l'histoire de la modélisation du climat en France. En plus de cette volonté de mêler éclairage historique et scientifique, l'écriture de ce chapitre a été l'occasion d'un vrai travail d'interaction avec une editrice et un graphiste.

Production scientifique 7 : Hourdin, F. Guillemot, Modélisation du climat, Encyclopédie Universalis

4.5 L'écriture de textes pour la communauté

La polémique lancée par le "comment" dans "Nature Climate Change" de Bauer, Stevens et Hazeleger en 2021, papier réellement mensonger et qui promettait une rupture dans l'éclairage du futur climatique grâce à des jumeaux numériques, combinant modèle globaux à résolution kilométrique et assimilation de données, a eu le mérite de forcer à redire le rôle de la modélisation du climat et ses spécificités. Un des principaux mensonges est de laisser penser qu'un modèle global kilométrique peut se passer d'une paramétrisation sous maille des nuages. On sait (quantité de

publications à l'appuie) que pour représenter explicitement les cumulus et stratocumulus sans autre paramétrisation que celle de la microphysique et de la turbulence petite échelle (ce qui est fait dans les simulations globales pilotées par Bjorn Stevens) il faut une résolution de quelques dizaines de mètres. D'ailleurs, les simulations globales réalisées à résolution kilométrique avec Icon (sous le pilotage de Bjorn Stevens) montrent dans ces cas là une absence totale de stratocumulus. J'ai écrit un texte d'une dizaine de pages pour la prospective Insu décortiquant tous les glissements de Bauer et al. Au passage, on ne parle pas là d'un débat anodin ou sain. On parle de digues qui ont sauté dans la prééminence de la méthode scientifique dans ce type de débat. Il n'y a pas d'un côté les tenants des modèles globaux et de l'autre ceux des jumeaux numériques. Il y a du côté des porteurs des jumeaux numériques un glissement problématique pour la communauté qui a profondément blessé des collègues proches mais aussi à l'étranger ou d'autres disciplines.

Plus fondamentalement, les modèles de circulation générale ne sont pas un modèle parmi d'autres, dans une hiérarchie. Comme je l'ai dit plus haut, le découpage entre des équations représentant la circulation atmosphérique aux échelles météorologiques, et les paramétrisations qui rendent compte de transfert verticaux d'énergie, d'humidité et de quantité de mouvement, joue un rôle tout à fait particulier en sciences du climat. Les paramétrisations sont un modèle des processus, une série d'équations qui résume des processus émergents comme les organisations observées dans les LES. Testé dans des versions unicolonnes du modèles, on vérifie que le modèle ainsi dérivé a un caractère prédictif de l'évolution de ces processus. Intégrées dans les modèles de climat, les paramétrisations permettent de comprendre le rôles de ces processus dans la machine climatique.

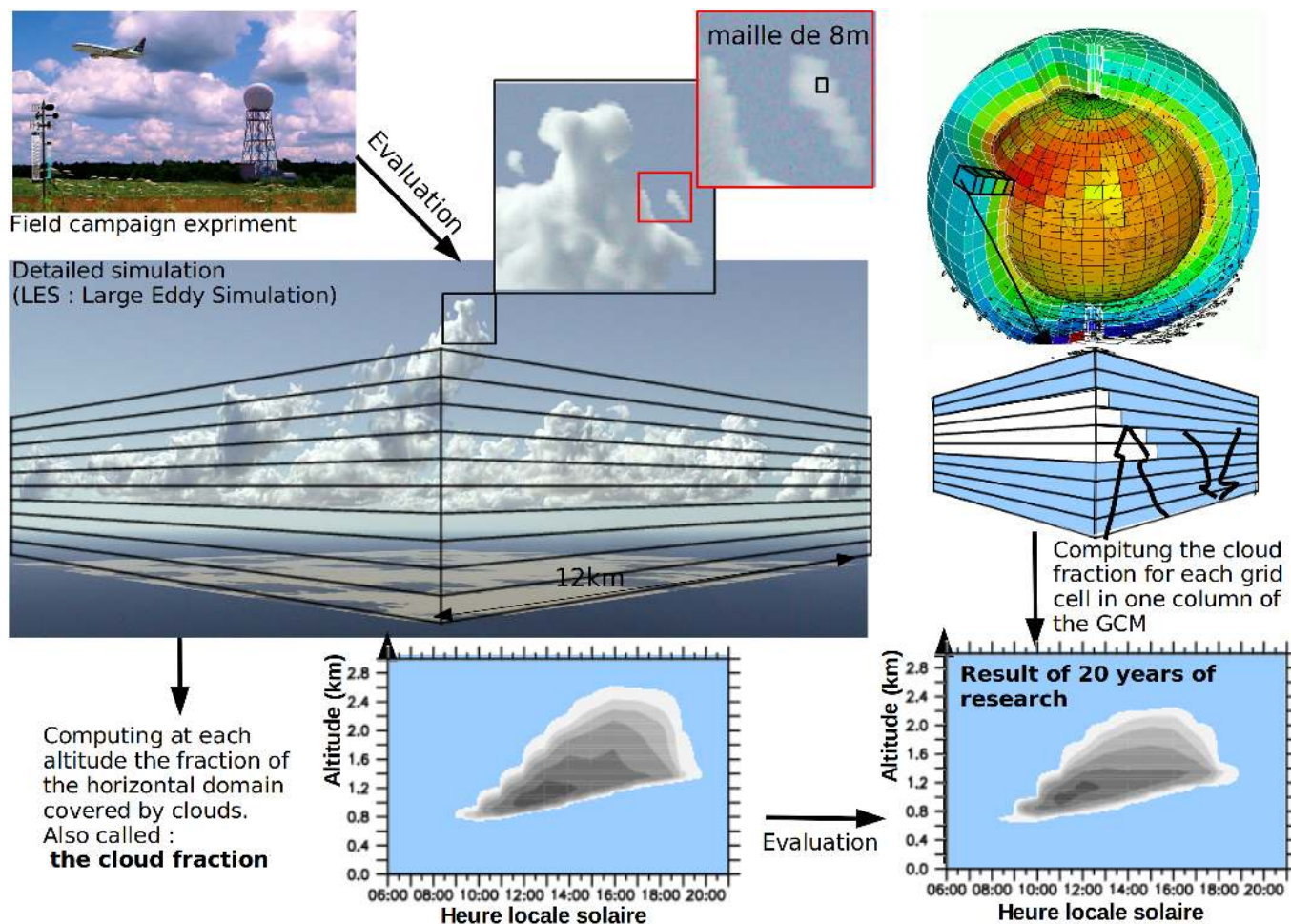
Une LES globale (typiquement à 25 m de résolution, on pourrait effectivement imaginer se passer en grande partie de paramétrisations sous maille) permettrait sans doute de simuler correctement les nuages et leur rôle dans le climat. En revanche, on voit bien qu'en termes de compréhension du rôle des processus nuageux dans le climat, on a remplacé un modèle de cette compréhension (au travers des paramétrisations) par une simulation et qu'on a donc perdu quelque chose en termes de compréhension. Nous avons récemment écrit un texte avec Catherine Rio pour une encyclopédie sur ce rôle des paramétrisations dans l'étude et la compréhension des interactions d'échelle.

Nous avons également écrit un article de point de vue pour PNAS, pour brosser plus largement le panorama de la modélisation du climat, de ces enjeux et limites, soulignant l'importance de garder des modèles relativement légers pour éclairer le changement climatique et pour pouvoir quantifier les incertitudes. Toute ce travail a permis de clarifier les idées et a permis à la communauté française de modélisation du climat d'affirmer l'importance d'un soutien pour le travail d'anticipation du changement climatique avec au cœur les modèles de circulation générale, et d'obtenir le financement du PEPR Traccs qui affirme très clairement ce point de vue.

Production scientifique 8 : V. Balaji , Fleur Couvreur , Julie Deshayes , Jacques Gautrais , Frédéric Hourdin , and Catherine Rio Are General Circulation Models obsolete?, PNAS, <https://doi.org/10.1073/pnas.2202075119>

4.6 Conférences grand public et supports

Je suis régulièrement invité à faire des présentations sur la modélisation du climat dans différents cercles : lycées, associations, universités. Je construis petit à petit des supports. Cet objectif a été une des grosses motivations pour réaliser avec l'aide de Najda Villefranche une LES à 8m de résolution



Diapositive illustrant l'utilisation des LES dans le travail sur les paramétrisations des modèles de climat. Les figures en bas montrent à gauche pour la LES, l'évolution au cours de la journée du profil vertical de la fraction des points de la grille horizontale qui contiennent de l'eau condensée. La figure de droite montre la même évolution temporelle du profil de fraction nuageuse calculée avec les paramétrisations de LMDZ, dans la configuration la plus récente du modèle.

du cas Arm Cumulus sur laquelle j'ai appliqué le logiciel de rendu physiquement réaliste *htrdr* (High-Tune renderer). J'utilise maintenant ces images devant tous les publics auxquels je présente la modélisation du climat, et le principe des paramétrisations. On voit sur la figure ci-dessus, une illustration de cette simulation pour un cours de modélisation.

J'ai réalisé avec Najda Villefranche la vidéo "Teapot in the city" et une vidéo plus simple pour parler de la physique des nuages.

Production scientifique 9 : Films basés sur la LES à 8 m du cas ARM cumulus, avec rendu physiquement réaliste par le logiciel *htrdr* the teapot in the city (https://web.lmd.jussieu.fr/~nvillefranche/pages/teapot_city), film des cumuls pour présentations pédagogiques (https://web.lmd.jussieu.fr/~hourdin/anim_d/movie_cumulus.mp4)

5 Transfert technologique, relations industrielles et valorisation

RAS malgré des tentatives passées de monter des choses avec les PME autour de la diffusion du modèle LMDZ.

6 Encadrement, animation et management de la recherche

6.1 Le pilotage du projet LMDZ

La coordination du projet LMDZ est une part quotidienne de mon activité depuis une vingtaine d'années. Aujourd'hui, nous effectuons ce pilotage à deux, moi plutôt comme responsable scientifique et Laurent Fairhead comme responsable technique. J'ai passé beaucoup d'énergie et j'en passe beaucoup sur les modalités de fonctionnement de ce groupe, et à œuvrer autant que faire ce peut à ce que chacun.e trouve sa place et se sente partie prenante d'un collectif. Cet aspect collectif s'est accompagné d'une montée en puissance de l'organisation informatique, dans laquelle nous avons mis beaucoup de tête au fil des années avec Laurent Fairhead.

Concrètement il y a aujourd'hui :

- Un POInt Hebdomadaire Lmdz (POIHL) réunissant pendant 1h30 le lundi une quinzaine de collègues;
- Des réunions Pedallons (POIHL Etendu aux Développeurs et Amis de Lmdz, Outil Numérique et Science);
- 3 jours de formation LMDZ;
- Un site web régulièrement mis à jour;
- Un contrôle qualité de plus en plus rationalisé : des scripts de "non régression" lancés automatiquement au milieu de la nuit, pour vérifier que les fonctionnalités vitales du modèle n'ont pas été affectées par les "commissions svn" (commit en anglais) du jour. Ces tests se sont enrichis de tests plus légers effectués à chaque commission.
- Typiquement tous les deux mois, la définition d'une version de référence sur laquelle on passe une plus grande série de tests.
- Depuis récemment, la montée en puissance de l'utilisation de maternmost (forum d'échange) qui a rationalisé les échanges faits auparavant de façon anarchique sur le mail.
- 4 listes mail : `lmdz` pour les annonces larges (utilisée volontairement avec parcimonie), `lmdz-commit` qui annonce chaque commission svn et chaque contrôle qualité, `poihl` pour l'organisation interne aux développeurs, `lmdz-users` une liste forum de discussion des utilisateurs, notamment pour poser des questions naïves ou parler d'un problème apparu, savoir si d'autres ont rencontré la même difficulté, etc.

Depuis 20 ans, j'ai comme souci permanent qu'on pense collectivement cette organisation, essentielle au bon fonctionnement du projet LMDZ.

En 2019, LMDZ a été labellisé SNO (Service National d'Observation) de l'Insu, ce qui nous a permis à la fois une reconnaissance de tout ce travail et aussi de formaliser les attentes. L'ambition principale de LMDZ en tant qu'outil peut se résumer à : développer un outil de modélisation à la pointe de la recherche, permettant à un bout de déployer des configurations très légères pour la recherche sur les paramétrisations et l'enseignement et, à l'autre bout, gérer une grande exigence en termes

de performances climatiques et d'opérationnalité comme composante atmosphérique du modèle de l'IPSL.

L'autre élément important de ce travail de pilotage du projet est de fournir régulièrement des configurations du modèle LMDZ produisant des résultats climatiques aux standards internationaux. Sans en faire un concours de beauté, ces exercices poussent par une saine émulation à l'amélioration continue des performances climatiques des modèles globaux. Un des enjeux est d'articuler ces versions de référence avec l'intégration d'avancées récentes dans la représentation des processus au travers de paramétrisations, avec la nécessaire phase de tuning. Après avoir porté le tuning des configurations de référence de LMDZ pour trois exercices CMIP consécutifs (CMIP4 avec LMDZ4, CMIP5 avec LMDZ5A et 5B et CMIP6 avec LMDZ6A), j'ai dit qu'il n'était plus possible de tuner le modèle "à la main" comme on le faisait précédemment (quelques simulations de sensibilité à certains paramètres, suivies d'analyses de cartes pour faire le choix de nouveaux tests, etc). Tout ce qu'on a travaillé dans HighTune était clairement motivé par ça. Et les outils sont là aujourd'hui.

Porter un tel projet représente une charge mentale permanente ; un souci permanent des outils et des personnes ; mais également une réflexion permanente sur les priorités, l'articulation avec les sollicitations (parfois centrifuges, parfois sous forme d'injonction) de l'environnement de travail local ou international. Il est essentiel de savoir trier et trancher, accélérer sur les sujets importants et savoir en freiner d'autres, pour ne pas faire perdre de temps sur de l'agitation. J'essaie au maximum (à la fois pour le collectif et pour me décharger, et beaucoup d'évolutions récentes du fonctionnement sont travaillées en ce sens) d'assumer toutes ces dimensions collectivement, mais on ne peut pas tout discuter tout le temps ; et certain.e.s autour peuvent avoir une image de protection trop forte. Dans une idée d'une nécessaire et souhaitable transmission de ce rôle, il est nécessaire d'en prendre la mesure et de travailler ces questions.

Ce travail sur l'outil et d'organisation s'accompagne de la rédaction d'articles scientifiques de référence documentant la mise au point et les avancées de ces configurations de référence, comme la publication ci-dessous pour LMDZ6A.

Production scientifique 10 : Hourdin, F., C. Rio, J.-Y. Grandpeix, J.-B. Madeleine, F. Cheruy, N. Rochetin, A. Jam, I. Musat, A. Idelkadi, L. Fairhead, M.-A. Foujols, L. Mellul, A. Traore, J.-L. Dufresne, O. Boucher, M.-P. Lefebvre, E. Millour, E. Vignon, J. Jouhaud, B. Diallo, F. Lott, G. Gastineau, A. Caubel, Y. Meurdesoif, and J. Ghattas, LMDZ6A : the atmospheric component of the IPSL climate model with improved and better tuned physics, James, <http://dx.doi.org/10.1029/2019MS001892>

6.2 Le co-portage du projet Dephy

Depuis une vingtaine d'années, nous œuvrons à la mutualisation du travail sur les paramétrisations physiques des modèles de climat et de prévision du temps au niveau national. Le projet "physique commune" initial (motivé par la convergence de la physique des deux modèles globaux français) a donné naissance par la suite au projet Dephy, d'abord comme projet Lefe, puis comme GdR porté par Catherine Rio. Après avoir dirigé une première mouture du projet Lefe, je reste impliqué dans l'organisation de ce réseau de recherche, dans le cadre de son comité de pilotage.

6.3 Le rôle dans la modélisation du climat à l'IPSL

De la même façon, je continue à être très impliqué dans la coordination de la modélisation du climat à l'échelle de l'IPSL, et participe en particulier à toutes les réunions du conseil scientifique (typiquement tous les deux mois). Et au delà bien sûr, la mise au point du modèle couplé de l'IPSL motive une grande partie des activités que je mets en place autour. Ce rôle a pris une autre forme dans la mesure où, ayant amené les outils de tuning dans cette communauté, je me retrouve très sollicité en ce moment pour la diffusion et la transmission de l'outil et des bonnes modalités de son utilisation.

6.4 Participation à des conseils scientifiques

Je suis membre des conseils scientifiques de MétéoFrance (Comsi, où je représente le CNRS) et du Cerfacs.

7 Objectifs / Projet de recherche

7.1 Développement de paramétrisations

J'ai plusieurs pistes ouvertes de recherche autour de la paramétrisation de la couche limite et des nuages.

La première concerne des raffinements du modèle du thermique, schéma en flux de masse de la couche limite convective. Une partie essentielle du modèle du thermique est sa formulation de l'alimentation latérale à la base des panaches ascendants. Cette alimentation contrôle directement le flux de masse convectif et tient donc une place essentielle dans la "fermeture" du modèle du thermique. Cette alimentation, présente dès la première version du modèle publiée en 2002, s'appuie directement sur l'image de convergence vers la base des panaches sous l'effet de gradients de pression. Ces gradients de pression eux mêmes sont liés au fait que l'air est plus chaud et donc moins dense dans le panache, avec donc une moindre pression en surface que dans les régions de subsidence. Le profil vertical de cette alimentation latérale a été choisis de façon relativement euristique, sur des images de couche limite. Lorsque des paramétrisations plus classiques ont été introduites pour représenter l'entraînement latéral dans la couche mélangée et dans la couche nuageuse, nous avons été tenté à un moment d'étendre ces formulations jusqu'à la surface pour remplacer cette formulation de l'alimentation. Tous ces essais ont été infructueux suggérant qu'il y avait quelque chose d'assez fondamentalement pertinent dans la version proposée en 2002. Le fait de disposer de LES à très haute résolution (8m) va permettre d'interroger les hypothèses faites dans cette couche d'alimentation et éventuellement de mieux asseoir ou modifier cette formulation. Des travaux ont aussi été initiés récemment, aux ateliers Dephy de 2023, concernant les couplages de l'équation cinétique turbulente et le modèle du thermique.

Nous comptons également revisiter le schéma statistique de nuages et son couplage à la convection profonde et peu profonde. Dans la maille d'un modèle de climat (typiquement 25 à 300 km de côté), on ne peut supposer que le nuage se crée d'un coup sur la maille quand on atteint la saturation à l'échelle de cette maille et on introduit donc une paramétrisation de la distribution sous-maille de l'eau, le nuage étant alors défini comme la fraction de la maille sur laquelle l'humidité dépasse sa valeur à saturation. Jusqu'alors, la distribution sous-maille de l'eau était calculée de manière "diagnostique" en

fonction de l'état de la colonne atmosphérique, en tenant compte par exemple du fait que la présence de panaches convectifs tend à accroître la largeur de cette distribution en amenant de l'air humide de la surface vers des niveaux plus secs. Pour différentes raisons, il est apparu que ce modèle diagnostic était insuffisant et un modèle pronostique, plus sophistiqué, est en construction pour le modèle LMDZ. Des premières preuves de l'intérêt de cette proposition ont été faites en mode SCM/LES. Il s'agit maintenant à la fois de développer des paramétrisations manquantes et de prendre en compte la variance de la vapeur d'eau comme une variable d'état du modèle, advectée à grande échelle, ce qui forcera à repenser plus fondamentalement le sens des variables d'état associées à l'eau dans ces différentes phases.

La mise en place de ce modèle pronostique de l'eau sous-maille, piloté en partie par la convection, permettra en retour de prendre en compte de cette variabilité dans la représentation du cycle de vie de la convection. En effet, beaucoup de travaux issus d'observations ou de simulation LES suggèrent que le développement des cellules convectives, la transition de la convection peu profonde vers la convection profonde, et l'organisation de la convection sont sensibles aux variations horizontales du contenu en vapeur d'eau de l'atmosphère à méso-échelles, un des mécanismes étant que les panaches se développent préférentiellement dans une atmosphère déjà humidifiée. La variabilité de la vapeur d'eau peut donc à la fois contrôler le déclenchement et l'intensité de la convection mais aussi son organisation. On profitera de l'originalité du modèle LMDZ d'avoir une paramétrisation relativement unique des poches froides créées par réévaporation des pluies sous les cumulo-nimbus, élément essentiel du cycle de vie de la convection et de son organisation. On se basera également sur l'analyse de simulations LES d'événements convectifs réalisées avec le modèle non-hydrostatique MESO-NH pour explorer les relations possibles entre la variabilité de l'eau et les propriétés des panaches convectifs (collaboration Catherine Rio, CNRM). Il y a de bonnes chances que j'encadre à la rentrée prochaine une thèse sur ce sujet.

Un autre aspect déjà abordé dans le cadre de la thèse de Maëlle Coulon-Decorzens et qui se poursuivra les deux années prochaines concerne la paramétrisation des paramètres radiatifs nécessaires aux nouveaux "solvers" radiatifs de ECrad, à savoir : un paramètre d'hétérogénéité, une longueur de décorrélation pour le modèle de recouvrement vertical et une taille moyenne de nuages nécessaire quand on active le solver Spartacus pour simuler l'effet 3D des nuages. L'idée est de profiter de la richesse de la description de la convection et des nuages associés dans LMDZ pour estimer ces coefficients.

7.2 La révolution de l'history matching

L'arrivée de la méthode d'history matching et l'utilisation et le développement de l'outil `htexplo` ouvrent des perspectives multiples, bien au delà de l'idée de tuning qui avait initialement motivé son développement.

C'est d'abord bien sûr la capacité, dans les périodes de développement des paramétrisations dans le cadre SCM/LES de se décharger de la question du réajustement des paramètres. Chaque nouvelle proposition s'accompagne presque automatiquement aujourd'hui d'un tuning permettant de comparer la paramétrisation modifiée avec la version de contrôle, une fois les deux ajustées au mieux.

Il y a plusieurs travaux qui commencent pour étendre la stratégie de tuning du modèle atmosphérique aux autres composantes du modèle et aux versions couplées. Ce tuning pose notamment le problème des constantes lentes du système. Une approche qui s'était avérée fructueuse

pour le tuning "à la main" de la version IPSL-CM6ALR pour CMIP6, était d'avoir identifié des biais de flux en surface, dans les simulations forcées en SST, qui avaient une signature dans les SST, avec une coïncidence des régions qui se refroidissaient insuffisamment en SSTs forcées et qui avaient des biais positifs de SST dans la configuration couplée. Nous avons généré récemment des ensembles paramétriques PPEs, d'une centaine de membres dans lesquels toutes les configurations (forcées, couplées, à différentes résolutions) utilisent les mêmes jeux de paramètres. Cet ensemble devrait permettre d'"apprendre" (au sens de l'apprentissage machine) certaines relations entre des caractéristiques des simulations forcées et couplées. Des relations de ce type pourraient être utilisées pour tuner indirectement les constantes lentes du système à partir de simulations atmosphériques courtes. D'autres stratégies sont actuellement en cours pour les couplages avec les surfaces continentales, utilisant des simulations sur sites et des simulations guidées.

Articuler ces différentes stratégies est à la fois une façon d'accélérer l'amélioration des modèles et l'ajustement de configuration mais également un formidable outil pour comprendre le fonctionnement du système climatique.

7.3 Recherche sur le développement informatique

Le fonctionnement mis petit à petit en place autour du développement de LMDZ, un peu par la force des choses, est en fait un mode particulier de développement connu, appelé "intégration continue" et même "Intégration continue sur branche unique". Ce modèle de développement prône une intégration chaque jour des nouvelles étapes d'un développement, même si le code n'est pas encore opérationnel. Le développement est alors "caché" sous des tests "if". Ceci permet d'éviter que les développeurs courent toujours après les modifications des autres développeurs. L'autre élément absolument essentiel est de garantir un maximum de rétro-compatibilité. Pour la recherche en développement de paramétrisations qui anime une grande partie de l'activité LMDZ, il est essentiel de pouvoir tester, dans la même version des "sources" fortran, plusieurs versions d'une même paramétrisation, et y compris une version vieille d'une dizaine ou vingtaine d'années pour pouvoir s'y comparer. Avec une dynamique très forte de développement de paramétrisations ces deux dernières années, on voit en fait se multiplier les ajouts dans le code, pas forcément finalisés. Bien sûr cette situation a un côté effrayant et amène à beaucoup de complexité dans le code, et n'aide pas à la visibilité. Il m'apparaît cependant de plus en plus clairement que cette façon de faire est essentielle si on veut mener une recherche de pointe sur le développement de paramétrisations. J'en vient même à me dire que de ne pas avoir identifié ce besoin de profusion dans le code comme un enjeu du développement informatique dans la communauté peut expliquer en partie la relative lenteur de l'amélioration des performances climatiques des modèles de climat.

Bien sûr, pour que cette profusion soit profitable, il faut se doter d'outils pour la gérer. Plusieurs éléments sont déjà en place, comme un contrôle qualité avec beaucoup de tests de non régression. A noter que cette nécessité d'automatiser ce type de contrôle qualité est clairement identifiée comme un enjeu dans les textes qui parlent de l'intégration continue. Mais ça ne suffit pas. Il faut se donner les moyens de rendre le code de plus en plus lisible et accessible malgré cette profusion.

Je souhaite dans les années qui viennent creuser ces questions d'organisation du développement informatique comme une question de recherche à part entière de la modélisation du climat. Nous avons déjà réfléchi, avec Laurent Fairhead notamment, à identifier différents niveaux de codage dans la partie du code qui concerne les paramétrisations physiques. A un bout, on a un moniteur dans

lequel on orchestre cette complexité. A l'autre bout des paramétrisations individuelles, qui travaillent sur des colonnes du modèle. Une des propositions est de différencier fortement une informatique avancée sur le moniteur et sur un niveau intermédiaire pour l'orchestration des paramétrisations et d'utiliser un fortran minimaliste à l'échelle des paramétrisations, avec des règles de codage simples et précises, avec une écriture au plus près de l'algorithmie, très explicite, permettant aux physiciens d'intervenir efficacement dans le code. Ce découpage m'a permis de développer récemment un outil "replay" qui permet d'équiper une paramétrisation ou sous paramétrisation automatiquement avec des sorties. Cette fonctionnalité permet automatiquement d'avoir accès à toutes les variables internes de la paramétrisation et également de rejouer une paramétrisation unique avec les entrées qu'elle a vu dans un premier calcul. Cet outil est devenu immédiatement un outil de travail des étudiant.e.s que j'encadre sur le développement du modèle.

Cette formalisation et cet outil `replay` sont également utilisés dans le travail qui commence sur le portage du modèle sur GPU.

7.4 Recherches pluridisciplinaires autour du changement climatique

Dans la suite de l'ANR MC2 et de l'article "Teapot in the city", je compte continuer à m'investir avec les collègues du réseau EDstar dans l'exploitation des approches Monte Carlo en chemin pour éclairer l'anticipation de l'impact du réchauffement global sur des installations thermiques ou des équipements de production d'énergie.

Dans cette perspective, j'ai coordonné à l'automne 2023 la rédaction scientifique d'un projet que nous co-portons avec Marion Carrier, chercheuse CNRS aux Mines d'Albi, en réponse à un appel d'offre du PEPR Ville Durable et Bâtiment Innovant. Le nom du projet, LAVEC pour "LA Ville en Chemin" fait référence à l'utilisation de reformulations de la physique en chemin utilisées pour les calculs. Mais ce nom fait également référence à la question de la circulation des savoir et à l'appropriation par les citoyens des enjeux scientifiques de la transformation énergétique.

Résumé en français du projet soumis : *Le projet LAVEC émane d'un collectif de recherche constitué de modélisateurs du climat impliqués fortement dans l'anticipation des changements climatiques et de leurs conséquences, du collectif EDStar qui porte depuis vingt ans le développement d'un cadre théorique, méthodologique et informatique permettant de modéliser des systèmes complexes multi-physiques et multi-échelles dans un formalisme en espaces de chemins hérité de Feynman et Kac, et enfin de didacticiens des sciences, intégrés dans le collectif depuis une dizaine d'années, et qui s'intéressent à la circulation des savoirs liés à l'énergie et au climat. Le projet s'est construit avec le Centre d'Accueil Permanent (CAP) de Saint-Front en Haute-Loire lui-même engagé dans l'éducation populaire sur les questions d'environnement. Le CAP est un lieu de résidence régulier du collectif qui y expérimente des modalités originales de circulation des savoirs en son sein et avec les habitants du territoire. Bénéficiant des avancées de l'informatique graphique, les méthodes en espaces de chemins permettent d'effectuer la simulation thermique dynamique d'un bâtiment ou d'une ville couplée avec la météorologie ou le climat, quelles que soient les échelles temporelles et spatiales, du pont thermique à la ville. Le caractère intuitif de ces méthodes facilite l'appropriation de la physique nécessaire à la compréhension des enjeux de la transition énergétique. Les preuves de concepts ayant été validées par des publications, l'objectif du projet est à présent le passage à l'échelle pour que les citoyens s'approprient la physique et soient acteurs de l'adaptation des villes au changement climatique. Il vise à concevoir, développer et distribuer une suite de logiciels libres*

(dite “le simulateur”), pour aider à la conception, l’aménagement et la gouvernance des milieux urbains, ainsi qu’à la circulation des savoirs. Du point de vue calculatoire, comme didactique, le simulateur s’adresse à un public large allant d’ingénieurs experts du bâtiment, à des classes de découverte “énergie / climat” au CAP Saint Front. Pour ce faire, la stratégie proposée est de guider sa conception par la nécessité d’en faire “un jeu d’enfants”. Cela pose des questions de recherche avec de fortes interactions entre l’informatique, la physique, le climat, et les didactiques de ces disciplines. Le site du CAP Saint Front sera utilisé comme lieu “pilote” pour expérimenter de nouvelles formes de circulation des savoirs avec différents publics. Le projet est structuré en 4 axes : (1) la production de logiciels permettant le passage d’un dessin d’enfant ou d’architecte à une donnée 3D conforme, nécessaire aux calculs thermiques qui seront réalisés avec le logiciel Stardis largement utilisé dans la communauté nationale de thermiciens ; (2) l’instrumentation du bâtiment du CAP pour confronter les mesures au modèle thermique couplé à la météo ou au climat. Les images infrarouges obtenues par les différents publics seront comparées à des images infrarouges simulées avec le modèle thermique afin de l’affiner ; (3) le couplage de la thermique des bâtiments avec le modèle de climat LMDZ et la quantification des incertitudes associées, une question majeure dans la conception de services climatiques. Le dernier axe (4), transverse au projet, vise à concevoir et expérimenter de nouvelles modalités de circulation des savoirs, et de mesurer la capacité de transformation d’un enseignement usant d’outils de simulations à la pointe des recherches climatiques et d’ingénierie énergétique, en contact étroit avec le monde de la recherche. Le projet résonne avec la forte ambition sociétale du PEPR en visant la massification, ici autour de l’utilisation du simulateur et de nouvelles modalités de transmission des savoirs.

Que ce projet soit accepté ou non, le travail sur ces questions est engagé et guidera une partie de mon activité de recherche pour les 5 ans qui viennent.