

Chapitre 1

Introduction

1 Contexte

1.1 Qu'est-ce qu'un modèle de climat et à quoi sert-il ?

Un modèle de climat est un modèle numérique dont l'objectif est la compréhension du climat et sa simulation. Il est souvent composé de plusieurs modèles qui communiquent entre eux : modèle d'atmosphère, d'océan et de surface des continents. Les modèles d'atmosphère et d'océans sont des GCMs (*General Circulation Model*), c'est-à-dire des modèles qui simulent la circulation du fluide (eau ou air) sur l'ensemble du globe. Les GCMs atmosphériques sont nés dans les années 1950, en même temps que les premiers ordinateurs, d'abord pour prévoir le temps, puis pour simuler et étudier le climat. Que ce soit pour une application en météorologie ou en climat, le GCM va toujours simuler l'état de l'atmosphère (température, humidité, nuages etc) toutes les quelques ~~secondes~~ à minutes, c'est-à-dire qu'il va toujours simuler la météorologie. La différence consiste en la manière dont est fait l'expérience. Pour une application en prévision du temps, l'état initial de la simulation va être très important, et les simulations vont être très courtes, typiquement d'une quinzaine de jours. Pour une application climatique, ce sont les équilibres du système sur des échelles à minima décennales qui vont importer. Le climat d'une planète est alors défini comme les statistiques des grandeurs atmosphériques (température, humidité, nuages etc) sur ces échelles de temps. Autrement dit, le climat est une propriété emergente de la météorologie simulée par le modèle, mais il n'est pas modélisé directement : ce sont les processus météorologiques qui le sont. *et en particulier*

Dans le contexte du changement climatique actuel, les modèles de climat ont deux objectifs principaux. (1) Fournir les projections climatiques les plus fiables et utiles possibles pour informer la société du changement climatique (2) Nous aider à comprendre le fonctionnement du climat et du changement climatique, c'est-à-dire comprendre l'impact des processus atmosphériques sur le climat et leur évolution avec le changement climatique. Le développement des modèles de climat est toujours sous-tendue par ces deux objectifs et le calendrier de la recherche est rythmé par les exercices d'intercomparaison des modèles de climat (CMIP pour *Coupled Model Intercomparison Project*) en phase avec les rapports du GIEC (Groupe d'experts Inter-gouvernemental sur l'Evolution du Climat). La place des modèles de climat dans la recherche est d'autant plus importante que c'est le seul moyen d'expérimentation disponible pour comprendre le climat dans sa globalité, puisqu'il n'existe pas de méthode permettant à ce jour de reproduire l'ensemble des processus responsables du climat dans un laboratoire.

états de régimes stables

Une fois que le modèle a affiné un état de régime respectif

Depuis une trentaine d'années,

milieu ?

1.2 Concrètement, qu'est-ce qu'il y a dans un modèle de climat ?

Les modèles de climat sont basés sur une représentation discrétisée de l'atmosphère, des océans et des surfaces des continents. Ces éléments sont découpés spatialement en maille, qui font typiquement $100 \times 100 \text{ km}^2$ sur l'horizontal, pour quelques mètres à kilomètre de haut. Le modèle calcule alors dans chaque maille et à chaque pas de temps (toutes les secondes à minutes) les valeurs prises par les variables d'états du modèle, qui servent à décrire le milieu. Ces calculs suivent un ensemble d'équation décrivant la physique et parfois la chimie et la biologie du système Terre. ~~modèle~~ une version modifiée du

Dans les GCMs atmosphériques, les mouvements de l'air sont décrit par les équations primitives, qui correspondent aux équations de Navier-Stokes appliquées à l'atmosphère (fluide parfait s'écoulant dans une mince couche sphérique en rotation). Ces équations sont discrétisées en temps et en espace et résolues explicitement par le modèle : c'est ce que l'on appelle le cœur dynamique du modèle. De nombreux processus atmosphériques ne sont pas décrits par ces équations, soit parce qu'ils ont lieu à des échelles spatiales inférieures à la taille des mailles du modèle (c'est le cas de la formation des nuages par exemple), soit parce qu'ils sont d'une nature différente (comme pour le rayonnement et la thermodynamique). Ces deux catégories de processus sont alors pris en compte par un ensemble de sous-modèles appelés paramétrisations : c'est ce que l'on appelle la physique du modèle. Ce découpage en deux parties distinctes, cœur dynamique et paramétrisations physique, rejoint un découpage spatial des équations. La dynamique assure les transferts horizontaux et la partie résolue ou grande échelle des échanges verticaux et les paramétrisations physique assurent les transferts verticaux non résolus dans l'atmosphère. Le même découpage existe dans les GCMs océaniques. ~~ainsi que~~

Les paramétrisations modélisent l'effet des processus atmosphériques non résolus par le cœur dynamique sur les variables d'état du modèle (température, humidité, vent etc) via des équations heuristiques alliant théorie, phénoménologie et statistiques. Elles sont appelées paramétrisation par la communauté scientifique car elles contiennent un certain nombre de paramètres. Les processus sous-mailles paramétrisés dans les GCMs atmosphériques sont généralement : la turbulence de petite échelle, la convection peu profonde à l'origine des nuages de cumulus et stratocumulus, la convection profonde à l'origine des cumulonimbus et des orages, la condensation et la précipitation de l'eau liquide et glacée, et les ondes d'échelle inférieure à la taille des mailles. S'ajoute à ces paramétrisations, les paramétrisations permettant de lier les différents modèles entre eux (les échanges d'eau et de chaleur entre la surface océanique ou continentale et l'atmosphère et le freinage de l'atmosphère par la surface) ainsi que les paramétrisations du transfert radiatif. Ces dernières modélisent le transport du rayonnement solaire à travers l'atmosphère et le transport et l'émission du rayonnement infra-rouge par la surface terrestre et l'atmosphère. Les théories physiques fondamentales régissant la dynamique, la thermodynamique et le transfert radiatif dans l'atmosphère sont bien établies. En revanche, les paramétrisations des processus sous-maille et du transfert radiatif en présence de nuage font l'objet de recherches actives. En effet, un certain nombre de ces processus ne peuvent être facilement dérivés des théories physiques, conduisant à une diversité d'approche pour modéliser le même processus. Les panaches convectifs conduisant à la formation des cumulus et stratocumulus dans la couche limite atmosphérique peuvent par exemple être modélisé comme un panache convectif moyen dans une maille ou comme un ensemble de panache décrit statistiquement. La précipitation de l'eau des nuages est paramétrisé de manière plus ou moins complexes : certains modèles ne la font dépendre que de la quantité d'eau condensé dans la maille quand d'autres modèles paramétrisent les processus de coalescence, de collection ou d'effet bergeron *[je dis ça mais je sais pas si c'est vrai]*. Cette diversité d'approche possible impacte significativement la simulation du climat et de la sensibilité climatique (hausse des températures globales en réponse à une énergie rajoutée dans le système Terre). Cela a motivé la création de projet d'intercomparaison des modèles de

des modèles

Chapitre 1. Introduction

climats, dans lesquels un protocole de simulation strict est suivi par un ensemble de modèles. Ces protocoles permettent ainsi d'isoler l'impact des choix de paramétrisation et de leur valeurs de paramètre associés sur la simulation du climat et du changement climatique et de quantifier ainsi l'incertitude des modèles.

1.3 Comment comprend-on le climat avec un modèle ?

Les modèles de climats sont devenus des modèles très complexes au fur et à mesure de leur développement. Leur complexité est le reflet de la complexité du climat terrestre et de cette caractéristique : des processus de très petites échelles peuvent avoir une influence sur le climat d'une région voire du globe. Cette imbrication de processus de différentes échelles spatiales et temporelle rend à la fois les modèles de climat globaux nécessaires - ce sont les seuls modèles pouvant imbriquer ces différentes échelles - mais les rendent aussi très complexes. Le climat simulé par les modèles est le résultats d'équilibres [Fred : Embêtant de dire "résultats d'équilibre"]. On n'y est jamais vraiment entre les différents processus modélisés, qui rétro-agissent entre eux ; et il peut être très difficile de savoir finalement quel processus est responsable de quelle caractéristique simulée par le modèle. Une grande partie de la recherche en modélisation consiste alors à trouver des stratégies pour contourner les difficultés inhérentes à cette intrication des processus. Par exemple, pour savoir si les précipitations du Brésil viennent de l'évaporation de la forêt amazonienne ou non, on pourra raser cette forêt dans le modèle et regarder l'impact sur les pluies brésiliennes. Pour savoir si un processus a une influence sur le climat, on peut le paramétriser et comparer des simulations sur des échelles climatiques avec ou sans la paramétrisation de ce processus. S'il y a des changements significatifs entre les deux simulations, ils pourront être attribués au processus paramétrisé. Le poids de certains processus sur le climat peuvent également être étudiés en faisant des simulations avec différentes valeurs de poids sur les paramétrisations, contrôlées par leur paramètres. On pourra par exemple comparer les climats de simulations ayant une dépendance de la taille des cristaux de glace à la température plus ou moins fortes. C'est ce que l'on appelle les études de sensibilité du climat aux paramètres libres du modèle.

Désinfiltrer l'effet de certains processus sur le climat peut également passer par la désactivation de certaines boucles de rétro-action. Le modèle de climat peut ainsi être utilisé sans le modèle d'océan, ou en imposant la dynamique atmosphérique. Des modèles idéalisés peuvent aussi être utilisés pour comprendre certains phénomènes. Le modèle peut par exemple être tourné en mode aqua-planète (sans continent). Il est aussi possible de ne tourner qu'une seule colonne de modèle, dont les conditions aux bords sont forcées afin de reproduire des situations typiques ou idéalisées, tel que le développement de cumulus de beau temps en journée sur les continents par exemple. Les chercheuses s'appuient alors sur ce que l'on appelle une hiérarchie de modèle, allant de la colonne atmosphérique (ou océanique ou continentale) isolée au système climatique complet.

1.4 Comment savoir qu'un modèle de climat simule un climat réaliste ?

Une partie de la confiance accordée aux projections climatiques faites par les GCMs est basée sur leur capacité à bien simuler le climat terrestre présent et passé. Comment est-ce que l'on fait, concrètement, pour juger que la simulation du climat par le modèle est bonne ou mauvaise ? Cette question est la question de l'évaluation des modèles, qui est toujours une question de recherche active.

Les GCMs sont tout d'abord jugés sur leur capacité à simuler correctement les climats du passé, sensiblement différents du climat actuel, et qui sont connus grâce à des observations indirectes appelées proxy climatiques. Les proxy climatiques les plus connus sont les carottes de

glaces, qui permettent de connaître la composition de l'atmosphère, c'est-à-dire les gaz qu'il contient et leur isotopes, sur plusieurs centaines de milliers d'années. La reconstruction des climats du passé à partir des proxy climatiques et leur simulation par les GCMs sont un domaine de recherche à part entière.

Les GCMs sont également jugés sur leur capacité à simuler correctement le climat ~~actuel~~, pour lequel on dispose d'un ensemble assez vaste d'observations. Tout d'abord, un ensemble de satellites observe la Terre en permanence. Ces satellites embarquent des instruments qui captent le rayonnement émis par l'ensemble du système Terre. Cela permet d'avoir une observation de ce rayonnement, de première importance pour le climat, mais aussi de remonter à d'autres variables atmosphériques comme l'humidité ou la couverture nuageuse. Ces dernières années, des instruments d'observation actifs envoyant des pulses de lasers (LIDAR) ont été embarqués par des satellites. Ces instruments permettent d'observer plus profondément dans les couches atmosphériques opaques au rayonnement et d'avoir accès aux variations verticales d'eau dans les nuages par exemple. Des simulateurs d'observables ont été développés pour transformer les sorties des GCMs et les rendre comparables aux signaux observés par les satellites.

L'atmosphère est aussi ~~pas~~ observé in situ. Des stations météorologiques sont installées dans tous les pays du monde et mesurent un certain nombre de grandeurs en continu (température, pression, humidité, vent etc) nécessaires à la prévision du temps. Deux fois par jour, des capteurs sont embarqués dans des ballons qui montent jusqu'à une trentaine de kilomètre dans l'atmosphère afin de mesurer les variations de ces grandeurs sur la verticale. Des sites de mesures atmosphériques sont également développés à des fins de recherche pour l'étude du climat. Ces sites, moins nombreux, possèdent plus d'instruments, permettant d'observer par exemple la composition des nuages, le rayonnement arrivant à la surface, la chimie de l'atmosphère, etc. Ces sites s'organisent en réseau pour partager des protocoles d'observations communs et fournir ainsi des observations comparables entre les sites. Ils sont également utilisés pour la calibration des instruments embarqués dans les satellites. L'atmosphère est aussi observé depuis la surface des océans, via des bouées météorologiques fixes ou dérivantes. Enfin, des campagnes de mesures sont régulièrement organisées à des fins de recherches, avec des objectifs d'observation et de compréhension des processus physique, chimique ou biogéochimique précis. La campagne STRATEOLE-2 a par exemple envoyé en 2021 une vingtaine de ballons pressurisés à dériver pendant plusieurs mois entre 18 et 20 km d'altitude, afin de comprendre les processus d'échange de vapeur d'eau entre la troposphère et la stratosphère. La campagne de mesure AWACA déploie un certain nombre de stations de mesure atmosphérique en Antarctique, continent très peu observé, afin de mieux comprendre les processus atmosphériques spécifiques des régions polaires.

Les différents types d'observations de la Terre sont utilisés pour produire des climatologies du système Terre, c'est-à-dire des données statistiques en tout point du globe décrivant certains aspects du climat (température, précipitations, humidité de l'atmosphère et du sol etc). Ces climatologies regroupent la description du climat actuel et sont largement utilisées pour l'évaluation des modèles de climat.

Des simulations de référence sont également utilisées pour évaluer les modèles de climat. Des LES (*Large Eddies Simulations* ou simulation des grands tourbillons) sont utilisées pour simuler à haute résolution des portions de l'atmosphère ou de l'océan. Une partie des processus sous-maille paramétrisés dans les GCMs sont alors explicitement simulés par la LES, qui sert ainsi de référence pour la simulation de ces processus par les paramétrisations du modèle. Des calculs de transfert radiatif sur les scènes nuageuses simulées par les LES ont récemment été développés en utilisant les méthodes de Monte Carlo. Ces calculs de références permettent de connaître avec fiabilité les flux radiatifs associés aux nuages simulés par la LES et ouvrent la voie à de nouvelles études sur la modélisation des interactions entre nuage et rayonnement.

citation

Le serait bien de mettre quelques citations
mais pas trop.

avec des

Des simulations de CRM (*Cloud Resolving Model* ou modèle résolvant les nuages) régionaux ou globaux, à résolution kilométrique, peuvent également être utilisés comme référence pour la modélisation des systèmes convectifs et leur organisation spatiale.

1.5 La question de l'ajustement des modèles de climat

[pas sur du dernier paragraphe. Il manque éventuellement la question de tuner en ciblant la sensibilité climatique. Peut-être que ça vaut le coût de faire un paragraphe sur pourquoi on pense que tuner le modèle ne change pas ses capacités de projection climatique ?]

1.5.1 Pourquoi a-t-on besoin d'ajuster les modèles de climat ?

Les paramétrisations des modèles sont basées sur un certain nombre de paramètre, appelés paramètres libres, généralement peu contraint par la théorie ou les observations. Ces paramètres peuvent être des facteurs de proportionnalité, des temps de vie, des paramètres géométriques (taille des gouttes par exemple) ou caractéristique d'un élément (inertie thermique, albedo de la neige) etc. Ils dépendent de la formulation des paramétrisations physiques et ne sont donc généralement pas partagés entre les modèles de climat. La valeur de ces paramètres est fixe pendant toute la durée des simulations. Idéalement ces valeurs sont fixes sur tout le globe mais il arrive qu'elles varient régionalement, en fonction du type de surface par exemple. La question de l'ajustement, ou tuning, du modèle est la question de l'attribution d'une valeur à cet ensemble de paramètre libre. Cette question n'est pas spécifique aux modèles de climat. Elle est commune à la modélisation numérique en générale, et est souvent désigné par le terme de calibration. Elle est vu comme une étape à part entière dans le développement d'un modèle numérique, et à foriori, d'un modèle de climat.

L'ajustement des modèles de climat est fait de manière à ce que le modèle simule un climat réaliste. Autrement dit, les valeurs des paramètres libres sont choisies de manière à ce que le modèle simule un climat en accord avec les observations ou autre références disponibles, sur un certain nombre de caractéristiques mesurées à l'aide de métriques (les précipitations moyenne sur une région par exemple). Savoir définir ce qu'est un climat réaliste est une question de recherche à part entière, mais une certitude est partagé par tous les centres de modélisation : le modèle climatique doit simuler une température moyenne d'équilibre proche des observations lorsque l'énergie reçue du soleil est proche de sa valeur réelle (environ 340 W.m^{-2}). Le système Terre à l'équilibre doit également renvoyer autant d'énergie vers l'espace (sous forme de réflexion du rayonnement solaire et d'émission de rayonnement thermique) que l'énergie qu'il reçoit du soleil, à condition qu'il conserve numériquement l'énergie, ce qui n'est pas toujours le cas. Le système climatique n'est pas exactement équilibré du fait du changement climatique. Une partie de la chaleur est absorbée par les océans dont les constantes de temps varient de la vingtaine à des milliers d'années. Cette absorption de chaleur par les océans provoque un déséquilibre énergétique du système autour de $0,5$ à 1 W.m^{-2} , bien en dessous des incertitudes d'observations (autour de 4 W.m^{-2}). La précision attendue sur le bilan énergétique des modèles de climat est ainsi de l'ordre du dixième de W.m^{-2} , ce qui correspond à une précision de l'ordre du dixième de degrés sur la simulation de la température globale (un changement d'un W.m^{-2} sur le bilan énergétique provoque un changement de $0,5$ à $1,5^\circ\text{C}$ sur la température moyenne globale). Une partie des processus influençant le bilan d'énergie du système Terre ne pourront jamais être simulés, ni même observés, avec suffisamment de précision pour que le modèle de climat soit "naturellement" à l'équilibre énergétique, c'est-à-dire sans intention délibérée qu'il le soit. C'est le cas des nuages, dont l'effet radiatif global net (-20 W.m^{-2}) n'est connu qu'à quelques W.m^{-2} près. Même les simulations avec les plus hautes résolutions (quelques mètres pour les LES) doivent paramétriser un certain nombre de processus de très fine échelle qui

Cette question de discuter avec le r est claire. La question des modèles, les temps réel c'est à dire horodaté mais dans le cas de l'ordre de 10 à 100 ans de temps, de l'ordre de 10 à 100 mètres de hauteur, de l'ordre de 10 à 100 km de distance.

influencent la réflectivité et l'émissivité des nuages. Les tailles de gouttes de nuages, allant de quelques nanomètre à quelques millimètres, ne pourront par exemple jamais être simulées explicitement par les modèles de climat, et leurs influence sur la réflectivité des nuages est de plusieurs W.m^{-2} , bien au dessus de la précision recherchée. Il n'est donc pas possible d'espérer un jour pouvoir contraindre les valeurs des paramètres libres du modèle uniquement par la théorie ou par l'observation. Les modèles de climat auront toujours besoin d'être ajustés afin qu'ils simulent, à minima, une température moyenne globale en accord avec les observations à quelques dixième de degrés près, et un bilan énergétique à quelques dixième de W.m^{-2} près.

1.5.2 Vers une science de l'ajustement des modèles

Cette question de l'ajustement des modèles a été identifiée dès les premiers travaux de modélisation du climat (Manabe and Wetherald, 1975), mais ~~elle est restée~~ au second plan. Les méthodologies utilisées pour ajuster les modèles de climat ont en effet peu été publiées et n'étaient pas demandées dans la documentation des modèles participants aux exercices d'intercomparaisons internationaux CMIP. Ce manque d'intérêt peut s'expliquer par le fait que la question de l'ajustement des modèles était vu par une partie des chercheuses comme une question purement technique, un bidouillage inévitable pour avoir un modèle "opérationnel" voire même un moyen de masquer les défauts des modèles. ~~Elle peut aussi être expliquée par peur que cet ajustement ne renforce les arguments de ceux qui remettent en cause la validité des projections relatives au changement climatique.~~ La question de l'ajustement des modèles a connu un net regain d'intérêt il y a une dizaine d'année. Un colloque international a été organisé en 2014 afin de partager les pratiques de tuning des différents centre de modélisation et d'expliquer les besoins et les questions scientifiques liées au tuning des modèles. Une volonté forte de ce colloque était de faire de cette question de l'ajustement des modèles une science à part entière de la recherche en modélisation climatique, qui doit donc être étudiée, publiée et débattue dans la communauté scientifique.

La question des compensations d'erreur est une question centrale dans le tuning des modèles. Lors du réglage du modèle, on souhaite au maximum que le modèle simule le bon climat pour les bonnes raisons. Notre confiance en les projections climatiques repose en effet sur ces deux aspects : la capacité du modèle à simuler correctement le climat présent, et le fait que le modèle soit basé sur la bonne combinaison d'équation physique. Une mauvaise combinaison d'équation physique peut conduire à une bonne simulation du climat actuel. Pire, cette mauvaise combinaison d'équations physique pourrait être provoquée, intentionnellement ou non, pour que le modèle simule correctement le climat actuel. Par exemple, supposons que la formulation des équations conduisant à la simulation des nuages de cumulus contiennent des erreurs fondamentales (dites erreurs structurelles), et supposons que ces erreurs impliquent une sous-estimation moyenne de la couverture nuageuse des cumulus, responsable du renvoi de XX W.m^{-2} de l'énergie solaire. Le bilan énergétique du modèle pourra quand même être correctement simulé, en choisissant par exemple des valeurs de paramètres libres conduisant à une réflectivité de ces nuages irréalistement grande (grâce à une toute petite taille de goutte par exemple). La question de savoir à quel point ces compensations d'erreurs sont désirables ou non, et où est-ce qu'elles sont acceptables ou pas, est aujourd'hui une question ouverte. La seule chose qui est sûre est que le modèle doit simuler un bilan énergétique global au dixième de W.m^{-2} près, au prix de compensation d'erreur si besoin.

On peut tout de même chercher à s'assurer qu'un certain nombre de processus soient correctement simulés par le modèle tout en garantissant une bonne simulation du climat. Cela peut se faire en ajustant le modèle à différentes étapes du développement : pendant le développement de la paramétrisation (au plus proche des processus), puis au niveau de chaque composante du modèle indépendamment les unes des autres, et enfin au niveau du modèle climatique complet.

alors définir comment faire des recherches
le fut

Une idée pour ce faire

L'idée est donc de s'appuyer sur la hiérarchie de modèle existante et utilisée pendant le développement du modèle pour son ajustement. La question de l'arbitrage entre "simuler le bon climat" et "simuler les bons processus" pourra alors se traduire concrètement dans l'importance accordée aux métriques du modèle complet par rapport à celle accordées aux métriques orientées processus (sur des portions découpées du modèle) lors de l'ajustement du modèle.

La question des compensations d'erreur rejoint la question du surajustement (*overfitting* ou *overtuning*) des modèles numériques en général, qui arrive dans des contextes où l'ajustement est fait à un degré plus fin que les incertitudes de comparaison entre le modèle et les références utilisées, ou alors lorsqu'une partie trop partielle du modèle est ajustée (*not tuning the whole model state vector*). Ce surajustement est identifié comme une des causes principales de la lenteur du développement de nouvelles paramétrisations, et donc de l'amélioration des modèles de climat. Introduire une nouvelle paramétrisation dans le modèle conduit en effet souvent à une moins bonne simulation du climat, alors que celui-ci a moins d'erreur structurelle, car cela casserait les compensations d'erreurs introduites lors de l'ajustement du modèle. Il faut alors réajuster le modèle à chaque fois qu'une nouvelle paramétrisation est introduite. Cet ajustement à longtemps été mené sur la base d'essai-erreur, en faisant varier un à deux paramètres à la fois, ce qui le rendait fastidieux et chronophage. Cela explique aussi pourquoi le tuning à longtemps été vu comme un frein au développement des modèles.

1.5.3 Les méthodes d'ajustement objectives

L'ajustement des modèles est un problème à grande dimensions, tant au niveau du nombre de paramètres libres présent dans les modèles (de l'ordre de plusieurs centaines, souvent réduit à quelques dizaines dans le processus d'ajustement), qu'au niveau des caractéristiques climatiques que l'on souhaite que le modèle simule correctement (les métriques). Au vu de cette difficulté, les centres de modélisation ont adopté, à partir des années 2010, des méthodes objectives d'ajustement de leur modèle, c'est-à-dire des méthodes basées sur un cadre mathématique ou statistique bien établi, développées par les communautés des statistiques, de l'ingénierie et de l'informatique. On distingue généralement deux types d'approches pour les méthodes d'ajustements. La première approche utilise un cadre d'optimisation, où l'ajustement consiste à minimiser une fonction des paramètres libres du modèle appelé fonction de coût, qui mesure la distance entre les observations et la simulation. La deuxième approche, connue sous le nom de quantification des incertitudes, utilise des modèles statistiques (ou émulateurs) pour relier le modèle climatique aux observations, en quantifiant explicitement les principales sources d'incertitudes présentes dans la comparaison entre modèle et observations (incertitudes des observations, incertitudes dues aux conditions initiales et aux limites incertaines du modèle, incertitude structurelle du modèle due à sa formulation incomplète ou incorrecte). Ces méthodes sont basées sur des méthodes d'inférences bayesiennes, qui fournissent une incertitude autour des valeurs des paramètres libres. Les deux méthodes ont comme avantage de rendre la question du tuning objective, contrairement aux méthodes d'essai-erreur, et ouvrent ainsi la voie à l'exploration des questions scientifiques liées à l'ajustement des modèles. Elles font toutes les deux face à la problématique très pragmatique du temps de calcul disponible pour l'ajustement des modèles, motivant ainsi à chercher des méthodes et des stratégies de tuning accessibles en terme de coût de calcul.

La diversité des modèles de climat, inhérente à la complexité du système climatique et aux choix faits dans les paramétrisations des modèles, ainsi que la diversité des priorités de chaque centre de modélisation, implique une diversité, une subjectivité et une expertise dans l'ajustement des modèles. Cette diversité se retrouve dans les choix des paramètres libres qui vont être utilisés pour ajuster le modèle, ainsi que dans les métriques utilisées pour l'ajustement. Elle permet d'explorer au mieux les incertitudes des projections climatiques des modèles, évaluées lors

des projets d'intercomparaison des modèles de climat. L'utilisation de méthode d'ajustement objectives ne remplace en aucun cas la subjectivité de l'exercice d'ajustement, et sa diversité entre les différents centre de modélisation. ~~Elle~~ implique que cette expertise soit explicitée. L'arrivée de ces méthodes ouvre la voie à une meilleur connaissance de l'incertitude liée aux choix des valeurs des paramètres libres sur la simulation du climat et du changement climatique. Le tuning pourrait alors même devenir une aide et un guide au développement des modèles, en tant que méthode pour identifier les erreurs structurelles des modèles, indépendantes des valeurs des paramètres libres choisis.

1.5.4 Stratégie d'ajustement du modèle climatique de l'IPSL

Suite à la conférence internationale de 2014, la question de l'ajustement des modèles de climat a été (ré)ouverte dans les deux centres de modélisation climatique français : IPSL en région parisienne et Météo-France à Toulouse. Cette question s'est structurée dans le projet ANR HighTune "Amélioration et calibration des paramétrisations des nuages de couche limite à partir de simulations haute résolution", réunissant, de 2016 à 2020, des statisticien.e.s, des expert.e.s du rayonnement, des modélisateurices du climat et de l'atmosphère et des spécialistes des processus atmosphériques. Ce projet a permis l'arrivée dans la communauté de modélisation du climat, d'une méthode objective d'ajustement semi-automatique des modèles issue de la communauté des quantifications d'incertitudes. Cette méthode, nommée "History Matching with Iterative Refocusing", vise à déterminer le sous-espace des paramètres libres en accord avec un ensemble de métrique à une tolérance à l'erreur prêt, en échantillonant de manière itérative l'espace des paramètres libres et en utilisant des émulateurs gaussiens des métriques pour accélérer cette exploration. Elle est implémentée dans l'outil de tuning htexplo et est au cœur de cette thèse.

Une proposition forte de ce projet est de combiner des métriques orientées processus avec des métriques climatiques. Les métriques orientées processus portent sur des simulations atmosphériques uni-colonne forcées pour reproduire des situations météorologiques typiques, et qui ont comme référence des simulations tri-dimensionnelles haute-résolution (LES) ayant les mêmes forçages au bord de leur domaine. Les métriques climatiques portent sur le rayonnement et le bilan d'énergie au sommet de l'atmosphère du modèle atmosphérique non couplé au modèle océanique. L'objectif de cette proposition est double. Tout d'abord, il s'agit de garantir que la méthode d'ajustement des paramètre conduit à la bonne simulation du climat pour les bons processus. Ensuite, cette stratégie profite de l'aspect itératif de la méthode d'ajustement pour pré-contraindre l'ensemble des paramètres libres avec des simulations uni-colonne très peu couverte, et avoir ainsi un espace de paramètre libre moins grand à explorer pour contraindre l'espace des paramètres avec les métriques portant sur les simulations climatiques, couteuses en temps de calcul. La preuve de concept de cette stratégie d'ajustement du modèle est faite sur l'ajustement des paramètres libres liés aux paramétrisations des couches limites faiblement convective du modèle atmosphérique de l'IPSL LMDZ. La comparaison entre les simulations uni-colonne et les sorties des simulations LES est implémentée de manière automatique dans l'outil de tuning htexplo, permettant un pré-conditionnement rapide de l'espace des paramètres libres.

[J'ai l'impression qu'il manque un petit paragraphe ici ; la fin est trop abrupte]

Partageant ^{u1} utilisé tout au long de

*oui : une phrase pour dire que tout ça
est très récent et issue de nombreuses
postes & recherche.*

2 Positionnement de la thèse

L'arrivée très récente de la méthode d'ajustement objective d'History Matching with Iterative Refocusing pour l'ajustement du modèle climatique global à l'IPSL ouvre un nouveau champ de recherche. Cette thèse est la première thèse portant intégralement sur la science de l'ajustement des modèles de climat. Elle vise globalement à explorer et à clarifier certains aspects de ce nouveau domaine de recherche.

La méthode d'HMIR présente l'avantage de pouvoir combiner des métriques portant sur différentes configurations de modèle, offrant une gamme de possibilité de métrique très large pour l'ajustement des paramètres du modèle. Actuellement, l'ajustement se fait d'un côté au plus proche des processus avec des métriques portant sur des cas un-unicolonne de convection peu profonde. Puis dans un second temps, des métriques sont ajoutées sur les flux radiatifs au sommet de l'atmosphère de simulation atmosphérique globale forcée dans l'océan. Le premier jeu de métrique a pour but de garantir la bonne simulation des processus, et le deuxième de garantir la bonne simulation du climat. Ces deux jeux de métriques pourraient être enrichis au sein de cette stratégie en deux parties. Côté simulation du climat global, on pourrait s'imaginer contraindre une plus grande diversité de caractéristique du climat, tel que des caractéristiques du climat à la surface (température, précipitations, humidité, rayonnement etc) qui est très observé et dont l'enjeu sociétal est fort, ou de viser spécifiquement les caractéristiques de phénomènes grande échelle tel que les mousson ou ENSO. On pourrait également enrichir la stratégie avec des métriques visant à réduire les biais systématiques du modèle atmosphérique, comme par exemple la simulation d'un double ITCZ, la trop faible variabilité des précipitations, les trop fortes précipitations sur le continent maritime etc. Côté processus, on pourrait ajouter des métriques sur une plus grande diversité de cas uni-colonne, et contraindre ainsi une plus grande diversité de processus atmosphériques ; tel que les cas de couche limite stable pour la turbulence de petite échelle ou les cas de convection profonde plus ou moins idéalisés. L'enjeu sur ces métriques est souvent d'être capable de quantifier l'incertitude de la simulation LES utilisée comme référence, ce qui n'est pas directe.

Cette stratégie en deux parties pourrait plus largement être enrichie de métriques portant sur des configurations intermédiaires déjà utilisées pour le développement et l'évaluation du modèle, afin de garantir une continuité entre l'ajustement au plus proche des processus et celui sur les métriques climatiques globales. Cela pourrait être fait en utilisant des configurations globales où la dynamique atmosphérique est forcée via des techniques de guidage et est ainsi découpées de la partie physique du modèle. Des configurations guidées et zoomées pourraient être utilisées pour contraindre le climat spécifique d'une région. Ces configurations permettent en particulier de comparer le modèle à des données in-situ de campagne de mesure ou de site instrumentés ; qui pourraient ainsi être utilisés comme référence dans le tuning du modèle. Une partie de l'ajustement pourrait également porter spécifiquement sur les couplages entre les différents modèles (océan-atmosphère et surfaces continentales-atmosphère) avec l'objectif d'ajuster conjointement les paramètres libres des différents modèles (et non plus modèle par modèle). Cette enjeu de continuité est particulièrement fort sur le rayonnement, qui est important de l'échelle des processus à l'échelle la plus globale : c'est le bilan des flux radiatifs au sommet de l'atmosphère, résultant de toute la complexité de la surface et des nuages, qui constitue le bilan énergétique du système Terre et qui est visé à une précision de quelques dixième de W.m^{-2} près lors de l'ajustement final du modèle climatique complet.

La méthode d'HMIR a été choisie pour son potentiel à aborder de nouvelles questions scientifiques. Cette méthode permet une exploration systématique de la sensibilité des métriques aux valeurs possibles des paramètres libres, et ainsi une quantification de l'incertitude paramétrique des métriques. Cette quantification de l'incertitude paramétrique est un énorme bond en avant.

Dans cette thèse, on va financer plus particulièrement on a exploré --