

Proposition de Sujet de thèse 2019

(1 page recto maximum)

Laboratoire (et n° de l'unité) dans lequel se déroulera la thèse :
CNRM - UMR 3589

Titre du sujet proposé : Approche physique des erreurs de modélisation en prévision d'ensemble atmosphérique

**Nom et statut du responsable de thèse : Bouttier François, IGPEF et HDR, équipe CNRM/GMME/PRECIP
coencadrant: Thierry Bergot, équipe CNRM/GMME/PHY-NH**

Coordonnées du responsable de thèse : francois.bouttier@meteo.fr (tél:05 6107 9825)

Résumé du sujet de la thèse

Les erreurs de modélisation contribuent aux erreurs de prévision que cherchent à décrire les systèmes de prévision d'ensemble. Les techniques actuelles de représentation de ces erreurs sont très empiriques. L'objectif de cette thèse est de développer une approche plus scientifique de la représentation des erreurs de modélisation, par identification des processus en cause dans des simulations simplifiées.

La méthodologie proposée repose sur le système de prévision d'ensemble PE AROME, opérationnel à Météo-France. Dans un premier temps, des statistiques d'erreurs de prévisions passées seront utilisées pour identifier des situations archétypales où la dispersion de la PEArome est insuffisante par rapport aux observations, ce qui est un marqueur d'erreurs modèle non représentées dans le système de prévision. On suspecte notamment les situations de brouillard, de couche limite stable, de transition pluie/non pluie et pluie/neige.

Dans un second temps, des simulations ensemblistes dans un cadre simplifié (petit domaine Arome) seront exploitées pour identifier précisément les structures probables de ces erreurs, et proposer des termes de perturbation stochastique pertinents dans la physique du modèle AROME (schémas de turbulence, microphysique, nébulosité, flux de surface). On s'intéressera à l'impact de ces perturbations sur les équilibres physiques et numériques interne du modèle, notamment les bilans thermodynamiques, le couplage surface/atmosphère, et le cycle de l'eau. Les développements seront testés en vraie grandeur grâce aux outils de validation des collaborateurs de l'équipe.

La motivation pratique de cette thèse est le développement de schéma de physique stochastiques pour améliorer les performances de la prévision d'ensemble PEArome opérationnelle.

Nature du travail attendu et compétences souhaitées: travail de recherche amont en modélisation numérique, nécessitant des compétences en physique atmosphérique, travail en équipe, expression scientifique écrite et orale, programmation python.

Références bibliographiques

Bouttier, F., B. Vié, O. Nuissier and L. Raynaud, 2012: Impact of stochastic physics in a convection-permitting ensemble. *Mon. Wea. Rev.*, **140**: 3706-3721. doi:10.1175/MWR-D-12-00031.1

Kober, K. and G. Craig, 2016: Physically Based Stochastic Perturbations (PSP) in the Boundary Layer to Represent Uncertainty in Convective Initiation. *J. Atmos. Sci.*, **73**, 2893-2911, doi:10.1175/JAS-D-15-0144.1

Ollinaho, P. , Lock, S. , Leutbecher, M. , Bechtold, P. , Beljaars, A. , Bozzo, A. , Forbes, R. M., Haiden, T. , Hogan, R. J. and Sandu, I. (2017), Towards process-level representation of model uncertainties: stochastically perturbed parametrizations in the ECMWF ensemble. *Q.J.R. Meteorol. Soc.*, **143**: 408-422. doi:10.1002/qj.2931

Précisions sur le sujet 'Approche physique des erreurs de modélisation en prévision d'ensemble atmosphérique':

Contexte du sujet

La représentation des erreurs de modélisation en prévision d'ensemble est un thème actif dans la communauté de recherche en prévision numérique du temps, qui rassemble les principaux centres météorologiques mondiaux (notamment Centre Européen CEPMMT, services météorologiques nationaux européens, des USA, Canada, Japon, Chine, Corée du Sud), les universités et laboratoires qui y sont associés (notamment CNRS, NCAR, les universités d'Exeter, Oxford, Washington, Munich, etc). Le sujet intéresse la communauté de **modélisation du climat** où ce composant influe sur la variabilité du système océan-atmosphère en zone tropicale. Il concerne aussi les secteurs hydrométéorologiques (prévision des crues) et énergétiques, avec des applications importantes sur la prévision des événements à fort impact et l'**optimisation des activités météosensibles** (par ex. énergies renouvelables, transport aérien). En effet, ces erreurs impactent directement l'anticipation des incertitudes météorologiques.

Historiquement, la recherche en prévision d'ensemble a privilégié la perturbation des conditions initiales, et l'interprétation des produits ensemblistes. La représentation des erreurs de modélisation a été moins étudiée, parce que les premiers ensembles étaient caractérisés par un manque de dispersion inévitablement lié à la mauvaise résolution des modèles utilisés. Aujourd'hui, la dispersion des ensembles est à peu près correcte en moyenne, mais s'avère insuffisante dans certaines situations où les membres 'ratent' tous de la même manière l'évolution de l'atmosphère réelle: c'est le signe que l'erreur de modélisation d'un processus, important à ces moments-là, n'a pas été représentée par une perturbation correspondante dans l'ensemble. Si la dispersion moyenne de l'ensemble est statistiquement correcte, cela signifie aussi que l'on a surestimé d'autres sources de perturbations dans l'ensemble, autrement dit on a créé des **erreurs compensatoires**.

Les représentations actuellement opérationnelles de l'erreur de modélisation sont fondées sur des approches pragmatiques: injection de bruit aléatoire dans les prévisions (= '**physique stochastique**') ou utilisation de modèles variés, dans l'espoir que la diversité des modèles sélectionnés soit un prédicteur de l'incertitude de leurs propres prévisions. Dans cette dernière technique, une diversité de modèles est obtenue empiriquement en regroupant dans un même ensemble, soit des modèles météorologiques indépendants (technique **multimodèle**), soit un modèle dans lequel diverses paramétrisations physiques sont activées (technique **multiphysique**), soit un modèle avec un groupe unique de paramétrisations, mais dans lesquelles on **perturbe des paramètres** de réglage (longueur de mélange de la turbulence, albedo, constantes microphysiques, etc.). Dans cette dernière approche, on postule que n'importe quel processus atmosphérique peut être correctement simulé avec un bon choix des paramètres de réglage, ce qui est illusoire lorsque des biais résultent de l'architecture même du modèle numérique: en effet, par construction, la méthode de perturbation des paramètres n'agit que dans un petit sous-espace prédéfini des erreurs possibles.

Ce foisonnement d'idées a conduit au cours des 20 dernières années à beaucoup d'efforts pour intercomparer différents algorithmes de représentation d'erreur modèle. Ces travaux sont handicapés par leur coût numérique, puisqu'ils impliquent de rejouer des prévisions d'ensemble dans différentes configurations sur de longues périodes. Ils se heurtent aussi à la relative rareté des observations, qui interdit d'établir une relation claire entre performance des ensembles et changements de représentation d'erreur modèle, comme expliqué ci-dessous.

En résumé, les approches pragmatiques et simples de représentation d'erreur modèle ont certes été utiles à l'aube des prévisions d'ensemble, mais leur manque de fondement scientifique les empêche aujourd'hui de progresser durablement. Le but de cette thèse est d'étudier la faisabilité d'une stratégie de modélisation plus solide, mettant en relation étroite les processus physiques actifs dans différents types de temps, et les observations permettant de les valider.

Etat des recherches en cours sur le sujet:

La plupart des études publiées sur les erreurs de modélisation ensemblistes comparent l'impact de différents algorithmes au moyen de scores probabilistes moyennés sur de longues périodes et sur de grands domaines géographiques (scores CRPS, Brier, relation erreur-dispersion). Ces études optimisent la qualité globale des probabilités prévues par les ensembles, mais elles tendent à **confondre amélioration de la dispersion moyenne des ensembles et amélioration de leur valeur prédictive**. En effet, une condition nécessaire (mais non suffisante) d'optimalité d'une prévision d'ensemble est que sa dispersion soit égale à l'erreur de prévision de l'état moyen de l'ensemble (c'est la relation erreur-dispersion). Les scores usuels sont très sensibles à la réalisation de cette condition en moyenne (appelée **fiabilité** de l'ensemble) , mais ils ne garantissent pas qu'elle soit vérifiée dans chaque situation météorologique.

En effet, les biais les plus évidents des ensembles, peuvent souvent être corrigés à moindre coût par **calibration statistique** de leurs sorties: ce peut être une homothétie des membres ('dilatation' de l'ensemble), ou un ajout de bruit aléatoire ayant une distribution analytique triviale ('habillage' des membres de l'ensemble). Ces calibrations, qui n'ont pas de base physique, ont sur les scores un impact comparable voire supérieur aux représentations existantes d'erreur modèle. Cela suggère que ces dernières sont souvent sans justification physique, et ne font qu'ajouter un bruit numérique arbitraire aux prévisions. Ce bruit est dans une certaine mesure transformé en structures physiquement cohérentes par les équations du modèle de prévision, mais aucune méthode ne garantit aujourd'hui que l'on a perturbé les vraies sources d'erreur de modélisation, c'est à dire les causes de d'incohérence entre les membres des ensembles et la réalité. Il est possible que, pour obtenir la 'bonne' dispersion dans un ensemble, on ait à la fois trop perturbé (donc dégradé) certaines variables, et trop peu perturbé des processus qui sont mal, ou pas, représentés.

La principale communauté qui a tenté d'attaquer ce problème de manière physique est britannique (UK Met Office, Université d'Oxford et CEPMMT). Leur hypothèse est que la plupart des erreurs de modélisation proviennent d'échelles plus fine que la grille du modèle: on peut donc les modéliser en analysant des statistiques de différences entre un modèle utilisé pour produire un ensemble, et un autre modèle beaucoup plus fin (supposé parfait). Cette technique, appelée **coarse-graining** vise, avec des simulations très coûteuses, à construire rigoureusement des paramétrisations d'erreur de modélisation; plusieurs versions en ont été testées en vraie grandeur, notamment le *Stochastic Kinetic Energy Backscatter* (SKEB). Le problème est que l'hypothèse de base (= que le modèle à haute résolution est parfait), si elle est défendable dans les modèles de climat à basse résolution, est peu convaincante dans les modèles météorologiques modernes visé par cette thèse: la maille horizontale du modèle AROME est de l'ordre du kilomètre, et augmenter sa résolution ne permet pas d'améliorer significativement les prévisions. De nombreux indices suggèrent que les erreurs les plus importantes ne sont pas des fluctuations sous-maille, mais des biais qui se manifestent de manière généralisée dans

certaines situations: transition pluie-neige, brouillard, structuration de cellules convectives, grêle, etc. Le coarse-graining n'a donc aucune chance d'y fonctionner.

Une autre école de pensée a cherché à modéliser les sources d'erreurs à l'intérieur même des schémas de paramétrisation physique, partant de l'hypothèse que ces schémas sont souvent conçus pour représenter simplement des phénomènes compliqués, avec une démarche consciente de simplification. On peut donc identifier dans leur équations les termes (ou les absences de termes) qui sont potentiellement sources d'erreurs. C'est le concept des **paramétrisations physiques stochastiques**, qui est au coeur de ce sujet de thèse. Un bon exemple est donné par le travail de **Kober et Craig (2016)** qui ont identifié, dans le schéma de diffusion verticale, le bruit implicite par l'hypothèse de fermeture turbulente du schéma. Teixeira et Reynolds (2008) ont fait une proposition similaire sur la fermeture d'un schéma de convection profonde.

C'est cette dernière approche que l'on souhaite tester et développer dans la thèse, dans le contexte du modèle atmosphérique AROME.

Ebauche de méthodologie et plan de travail

AROME est un modèle numérique non-hydrostatique de l'atmosphère comportant différentes paramétrisations physiques candidates pour l'injection de perturbations stochastiques représentant l'erreur de modélisation:

- un schéma de diffusion turbulente analogue à celui utilisé par Kober et Craig (2016) et un schéma de convection sous-maille non précipitant (EDKF): on se propose d'y tester les schémas stochastiques simples qui ont été proposés dans la littérature avec des résultats encourageants
- un modèle de surface (SURFEX), et des schémas de microphysiques et de rayonnement, plus complexes, dans lesquels on se limitera pour le moment à injecter des perturbations stochastiques très simples, sur les flux de surface et la représentation statistique de la nébulosité, en complément des schémas de perturbation de paramètres de microphysique et rayonnement qui y ont été déjà développés (thèse en cours dans l'équipe voisine CNRM/RECYF).

On souhaite donc à ce stade limiter l'injection de perturbations physiques à des processus dont la nature stochastique des paramétrisations est connue et a été documentée dans la littérature.

Dans un premier temps, **on identifiera des cas d'étude** dans lesquels ces processus ont un impact significatif, ce qui sera étudié au travers de diagnostics de cohérence entre dispersion et erreur de l'ensemble. Un stage effectuant ce type d'étude sur le brouillard a lieu en 2019 (équipe CNRM/GMME/PHY-NH), ce qui permettra de définir un protocole de diagnostic de l'erreur modèle sur le brouillard, et de s'assurer que l'on cible les bons processus (on sait déjà que la représentation de la turbulence y est cruciale). Le papier de Kober et Craig (2016) a montré un impact de l'erreur sur la turbulence sur des situations d'orages sur l'Allemagne, et on identifiera des cas d'étude semblable sur la France métropolitaine afin de bénéficier d'observation de validation abondantes (radar, profileurs, avions et foudre notamment)

Dans un second temps, un **système de prévision d'ensemble AROME simplifié sur des petits domaines** (quelques dizaines de km de taille) sera configuré pour étudier ces cas. Cela qui facilitera les développements de **modifications dans les paramétrisations physiques**, et l'extraction de diagnostics détaillé (par rapport à l'ensemble AROME complet, dont l'exécution est très lourde et coûteuse). Les cas

seront choisis pour bénéficier d'une bonne couverture en observations, et de l'intérêt des collaborateurs physiciens spécialistes des différentes paramétrisations au CNRM. Durant cette thèse, on s'intéressera particulièrement à la campagne SOFOG3D qui étudiera le brouillard dans le sud-ouest de la France, à des cas marquants de convection orageuse, et peut-être à un cas de campagne de mesure ciblant la modélisation de la turbulence (par exemple, BLLAST). L'encadrement vérifiera s'il est intéressant d'utiliser des sites instrumentés comme le SIRTA à Palaiseau, ou Cabaw aux Pays-Bas: cela permettrait éventuellement de compléter les études de cas ponctuelles par des validations sur de longues périodes (plusieurs mois), par exemple pour échantillonner les problèmes connus d'AROME sur les transitions pluie/non-pluie et pluie/neige.

Les études de cas consisteront, avec plusieurs configurations possibles des schémas de représentation d'erreur modèle (bruit SPPT, perturbation de paramètres, physique stochastique) à **documenter les variations d'erreur systématique et de dispersion de l'ensemble en relation avec l'erreur de prévision** (donnée par les écarts aux observations). On examinera notamment les profils verticaux de tendances et de flux liés au différents modules du modèle AROME. Les processus responsables des variations de dispersion seront identifiés, et validés en termes de compatibilité structurelle de cette dispersion produite avec les erreurs observables. On s'intéressera particulièrement aux **perturbations des équilibres physiques du modèle**: conservation de l'énergie et de l'eau notamment. Plusieurs équipes ont en effet indépendamment constaté que les schémas stochastiques existants introduisent des biais non-physiques dans les simulations (assèchement), mais le mécanisme en jeu n'est actuellement pas compris. Notre étude permettra de qualifier les différentes approches de représentation de l'erreur modèle en termes de différents aspects de leur réalisme physique.

La thèse pourra se conclure par une validation, dans le système de prévision d'ensemble AROME en vraie grandeur, d'une nouvelle version du schéma de représentation de l'erreur modèle, à la lumière des résultats obtenus dans le petit modèle. On pourra ainsi comparer les schémas existants (physique stochastique SPPT), celui développé dans la présente thèse (perturbations stochastiques à l'intérieur des paramétrisations physique), et le schéma de perturbation des paramètres physiques d'AROME qui fait actuellement l'objet d'une thèse en cours dans une autre équipe (CNRM/GMAP/RECYF).

Insertion de la thèse dans un cadre collaboratif

La thèse bénéficiera de la proximité de plusieurs chercheurs spécialistes en prévision d'ensemble (équipe CNRM/GMME/PRECIP) et en modélisation avec le code AROME (équipe CNRM/GMME/PHY-NH), incluant des experts sur le brouillard et sur chacune des paramétrisations physiques qui seront abordées. Au CNRM, le groupe CNRM/GMAP comporte des compétences en prévision d'ensemble, et en intégration et diagnostic des paramétrisations physiques AROME, incluant la configuration 'petit domaine' de ce modèle numérique.

Le travail bénéficiera d'interactions avec les projets scientifiques HyMeX (pour les précipitations et la convection orageuse), SOFOG-3D (pour l'étude du brouillard), plus d'autres campagnes dédiées à l'étude de la turbulence et à l'utilisation de sites instrumentés, en fonction des opportunités (BLLAST, SIRTA...). Les validations seront interprétées en termes du potentiel d'amélioration du système de prévision d'ensemble opérationnel PE-AROME, ce qui permettra de valoriser le travail pour les activités de prévision de Météo-France et de ses partenaires (centres météorologiques européens du réseau HIRLAM-ALADIN, réseau européen SRNWP), qui organisent régulièrement des conférences sur la prévision d'ensemble.

De nombreuses autres équipes ont déjà explicitement affiché leur intérêt pour la problématique abordée dans la thèse, notamment l'université de Munich (en collaboration avec le service météorologique allemand), l'université d'Oxford, le CEPMMT (Centre Européen pour les prévisions météorologiques à moyen terme), le NCAR (centre de recherche atmosphérique national des USA). En cas de succès de la thèse, il est donc certain que les compétences acquises seront appréciées dans la communauté météorologique.

Compétences et outils de l'équipe encadrante sur le sujet

F. Bouttier (HDR) a encadré et publié de nombreux travaux de recherche sur la prévision d'ensemble AROME, il a notamment développé le schéma actuel de représentation stochastique d'erreur de modèle SPPT. Il maîtrise un environnement complet de validation des expériences de prévision d'ensemble AROME (scores probabilistes, comparaison aux observations in situ, radar, foudre, satellite, hydrométéorologie). Il entretient des collaborations internationales sur la représentation des erreurs de modélisation (réseau de coopération ALADIN-HIRLAM, projet spécial du CEPMMT). Il a coencadré une thèse sur la représentation des erreurs de modélisation dans la PE AROME (Mihaly Szucs, service météorologique hongrois).

T. Bergot est un expert internationalement reconnu en modélisation du brouillard. Il participera activement à la campagne de mesure SOFOG3D. Il maîtrise l'outil technique central de la thèse, qui est la prévision d'ensemble AROME sur petit domaine. En 2019, il sera l'encadrant principal d'un stage utilisant cet outil.

Le sujet s'inscrit dans le programme de recherches amont du groupe CNRM/GMME pour l'amélioration de la prévision d'ensemble PE AROME et l'étude de la prévisibilité des processus précipitants. Ce programme est complémentaire des activités du CNRM/GMAP qui travaille sur d'autres aspects de la PE AROME. Un travail d'intercomparaison et de coordination des résultats des différentes équipes sur ce thème sera effectué par les encadrants en parallèle de la thèse.

Références complémentaires

Lin, J., and J. Neelin, 2002: Considerations for stochastic convective parameterization. *J. Atmos. Sci.*, 59, 959-975.

Palmer, T., 2001: A nonlinear dynamical perspective on model error: a proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quart. J. Roy. Meteor. Soc.*, 127, 279-304

Plant, R., and G. Craig, 2008: A stochastic parameterization for deep convection based on equilibrium statistics. *J. Atmos. Sci.*, 65, 87-105.

Shutts, G. J., T. N. Palmer, 2007: Convective Forcing Fluctuations in a Cloud-Resolving Model: Relevance to the Stochastic Parameterization Problem. *J. Climate*, 20, 187-202. doi: 10.1175/JCLI3954.1

Teixeira, J. and Reynolds, C. A. 2008. Stochastic nature of physical parameterizations in ensemble prediction: a stochastic convection approach. *Mon. Wea. Rev.* 136, 483-496