

# Proposition de post-doctorat, projet IMPT

## Missions

Une part importante de l'incertitude dans les modèles climatiques vient des processus de fine échelle, qui ne sont pas directement résolus. Ainsi, ils sont représentés par des paramétrisations, dont le rôle est d'estimer l'impact des processus « sous-maille » sur l'écoulement de grande échelle. Ces paramétrisations, telles qu'elles ont été développées jusqu'à présent, mélangent physique et pragmatisme, et demeurent peu contraintes, ne faisant qu'un usage très limité des observations.

Ce projet de recherche utilisera les méthodes d'apprentissage statistique pour capturer la relation entre l'écoulement de grande échelle et les ondes internes de gravité, qui constituent un processus sous-maille déterminant pour la circulation de la stratosphère. Cette démarche permettra de tirer pleinement parti des données provenant de campagnes récentes d'observation par ballons stratosphériques coordonnées par le Laboratoire de Météorologie Dynamique (LMD), et de guider le développement des paramétrisations des ondes de gravité dans les modèles de climat.

Les principales missions seront de mettre en place des méthodes d'apprentissage pour estimer des quantités clés des ondes de gravité, telles qu'observées par les ballons, à partir d'informations provenant de modèles météorologiques et décrivant l'écoulement de grande échelle. Le travail effectué répondra aux questions suivantes :

- quelle fraction des ondes de gravité peut être déterminée, quelle part doit être considérée comme « stochastique » ?
- quels éléments de l'écoulement de grande échelle apportent le plus d'information ?
- quelles sont les méthodes les plus efficaces dans ce problème ?

## Activités

Nous disposons de jeux d'observations uniques fournissant une quantification précise des ondes de gravité (flux de quantité de mouvement, fréquences, vitesses de phase), provenant de campagnes de ballons stratosphériques de longue durée. Ces observations fourniront les variables cibles. Une description de l'écoulement à grande échelle, tel qu'il pourrait être simulé par un modèle de climat, sera obtenue par un lissage contrôlé de la réanalyse ERA5 du Centre Européen de Prévision Météorologique à Moyen Terme (European Center for Medium-Range Weather Forecasts, ECMWF). Les principales activités seront :

- Extraire de cette description « basse résolution » de l'écoulement des variables explicatives co-localisées avec les observations des ballons. Ce travail préliminaire est déjà entamé.
- Appliquer des méthodes d'apprentissage statistique pour estimer la variable cible à partir des variables explicatives. Des méthodes de régression paramétriques et non-paramétriques seront employées, les premières étant attractives pour leur interprétabilité, les secondes pour leurs performances. Chacune des méthodes devra être calibrée avec soin pour une performance optimale.
- Explorer la sensibilité des résultats à la quantité d'information fournie (résolution, nombre de variables), à la quantité visée (flux de quantité de mouvement, caractéristiques des ondes), et aux méthodes employées (régressions linéaires ou polynomiales, méthodes de k-plus proches voisins, estimateurs à noyau, arbres de régression et méthodes d'ensemble associées, support vector machines...).
- améliorer les performances des méthodes d'apprentissage par des stratégies plus avancées, comme des procédures d'agrégation, ou une étape préalable de clustering sur les entrées.

## Compétences attendues

Cette offre s'adresse à un(e) jeune docteur dans le domaine des mathématiques appliquées et des méthodes d'apprentissage statistique. Une connaissance des sciences de l'atmosphère et du climat, et/ou de la mécanique des fluides, sera un élément d'appréciation favorable, mais pas indispensable. Une certaine aisance avec les expériences numériques et la programmation pour l'analyse de grands jeux de données (par exemple en R ou en Python) sera appréciée. Un goût pour le travail collaboratif, les interactions et l'interdisciplinarité seront des points positifs.

## **Contexte de travail**

Le post-doctorat s'insère dans une collaboration existant depuis environ cinq ans entre Aurélie Fischer, du Laboratoire de Probabilités, Statistique et Modélisation (LPSM, Université de Paris), et Riwal Plougonven, du Laboratoire de Météorologie Dynamique (LMD, Ecole Polytechnique). Le post-doctorant ou la post-doctorante travaillera au LPSM. Au début du post-doctorat, il ou elle viendra quelques fois au LMD, à Palaiseau, pour établir un socle minimal de connaissances sur la modélisation du climat et les processus atmosphériques pertinents. Dans la suite, des interactions régulières à trois, par visioconférence, permettront d'effectuer un suivi continu et de discuter de toutes les questions se posant, notamment sur l'interprétation des données géophysiques. Des réunions en personne, plus espacées, pourront être prévues sur le site du LPSM, à Paris, pour compléter ce dispositif. Le post-doctorant ou la post-doctorante sera encouragée à présenter ses résultats, les publier, participer dans la mesure du possible à des réunions, colloques ou conférences qui seront bénéfiques pour son réseau et sa carrière.

# Post-doc proposition, IMPT project

## Missions

An important part of the uncertainty in climate models comes from fine-scale processes, which are not directly resolved. Thus, they are represented by parameterizations, whose role is to estimate the impact of "subgrid" processes on the large-scale flow. These parameterizations, as they have been developed so far, mix physics and pragmatism, and remain unconstrained, making only very limited use of observations.

This research project will use statistical learning methods to capture the relationship between the large-scale flow and internal gravity waves, which are a key subgrid process for stratospheric circulation. This will take full advantage of data from recent stratospheric balloon campaigns coordinated by the Laboratoire de Météorologie Dynamique (LMD), and guide the development of gravity wave parameterizations in climate models.

The main tasks will be to implement learning methods to estimate key quantities of gravity waves, as observed by the balloons, from information from meteorological models describing the large-scale flow. The work performed will answer the following questions:

- what fraction of the gravity waves can be determined, what fraction should be considered as "stochastic" ?
- which elements of the large-scale flow provide the most information ?
- what are the most efficient methods in this problem ?

## Activities

We have unique sets of observations providing precise quantification of gravity waves (momentum fluxes, frequencies, phase velocities), from long duration stratospheric balloon campaigns. These observations will provide the target variables. A description of the large-scale flow, as it could be simulated by a climate model, will be obtained by the European Center for Medium-Range Weather Forecasts (ECMWF). The main activities will be:

- Extract from this "low resolution" description of the flow explanatory variables co-located with the balloon observations. This preliminary work is already underway.
- Apply statistical learning methods to estimate the target variable from the explanatory variables. Parametric and non-parametric regression methods will be used, the former being attractive for their interpretability, the latter for their performance. For each method, a careful calibration will be needed for optimal performance.
- Explore the sensitivity of the results to the amount of information provided (resolution, number of variables), to the target quantity (momentum flux, wave characteristics), and to the methods used (linear or polynomial regressions, k-nearest neighbor methods, kernel estimators, regression trees and associated ensemble methods, support vector machines...).
- Improve the performances of the learning methods by more advanced strategies, such as aggregation procedures, or a preliminary clustering step on the inputs.

## Expected skills

This offer is for a young PhD (M/F) in the field of applied mathematics and statistical learning methods. Knowledge of atmospheric and climate sciences, and/or fluid mechanics, will be an advantage, but not essential. Familiarity with numerical experiments and programming for the

analysis of large datasets (e.g. in R or Python) will be appreciated. A taste for collaborative work, interactions and interdisciplinarity will be positive points.

## **Context of the project**

The context of this postdoctoral fellowship, founded by the Institut des Mathématiques pour la Planète Terre (IMPT), is an approximately five-year collaboration between Aurélie Fischer, from the Laboratoire de Probabilités, Statistique et Modélisation (LPSM, University of Paris), and Riwal Plougonven, from the Laboratoire de Météorologie Dynamique (LMD, Ecole Polytechnique). The postdoctoral fellow will work at LPSM. At the beginning of the post-doc, he or she will come a few times to LMD, in Palaiseau, to establish a minimal knowledge base on climate modeling and relevant atmospheric processes. Thereafter, regular three-way interactions via videoconference will allow for continuous monitoring and discussion of any issues that arise, including the interpretation of geophysical data. More frequent face-to-face meetings may be scheduled at LPSM in Paris to complete this arrangement. The postdoctoral fellow will be encouraged to present his or her results, publish them, and participate in meetings, colloquia or conferences that will be beneficial to his or her network and career.