

Laboratoire : Laboratoire d'Optique Atmosphérique

Directeur de Thèse : Frédéric André

Co-directeur de Thèse : Laurent C.-Labonnote, Paul Gautier

Thématique : Physique Atmosphérique

## **Exploration des capacités de l'apprentissage informé par la Physique pour la télédétection nuageuse à partir des mesures instrumentales de la plateforme EPS-SG.**

Un des enjeux majeurs pour améliorer notre compréhension du climat, de sa modélisation et donc de son évolution, réside dans la bonne représentation des processus physique qui décrivent la dynamique multi-échelle des masses d'air. Ceci repose sur la caractérisation détaillée de la répartition spatiale et temporelle des constituants atmosphériques, ainsi que de leurs propriétés optiques dont la connaissance est fondamentale pour comprendre et simuler leurs interactions avec le rayonnement électromagnétique et quantifier leur impact sur le bilan énergétique de l'atmosphère.

La télédétection spatiale est un outil incontournable pour tester et améliorer les modèles climatiques. Elle a aussi un rôle important dans la mise en œuvre d'approches par assimilation de données permettant d'adapter de façon dynamique les modèles de prévision numérique du temps. C'est dans ce contexte que l'agence météorologique européenne (EUMETSAT) investit, via le programme Eumetsat Polar System (EPS), dans le déploiement d'une instrumentation variée permettant, depuis l'espace, de produire de grandes quantités de données visant à maximiser les informations disponibles sur les constituants atmosphériques. Obtenir ces informations à partir des mesures spatiales nécessite un couplage étroit entre modélisation et observations. Il faut ainsi disposer d'un modèle direct, permettant de simuler de façon précise le signal mesuré par les instruments à partir de données d'entrée connues. Les modèles directs reposent généralement sur la modélisation du transfert par rayonnement dans l'atmosphère et des interactions avec les gaz et particules rencontrés. Ce modèle direct est ensuite couplé à des méthodes dites inverses, qui permettent d'estimer l'état de l'atmosphère (le vecteur d'entrée optimal du modèle direct) à partir d'un jeu de données d'observations. Toute la difficulté consiste à extraire de la mesure le maximum d'informations possibles en termes de concentration, taille, forme des particules ou encore leurs répartitions spatiales (verticale mais aussi horizontale dans certains cas).

Les méthodes inverses basées sur la Physique (car utilisant explicitement des modèles physiques afin d'estimer un état atmosphérique permettant d'expliquer au mieux les observations) font référence depuis une vingtaine d'années. On peut notamment citer la méthode d'estimation optimale, qui permet d'approcher de manière itérative l'état le plus probable de la colonne atmosphérique via la minimisation d'une fonction coût (distance entre la simulation et l'observation, pondérée par une matrice d'erreur souvent liée au rapport signal sur bruit de l'instrument) et le calcul de la sensibilité de la mesure au paramètre d'intérêt. Si l'avantage de ces méthodes réside dans l'utilisation de modèles physiques permettant de simuler l'observation à chaque itération du processus d'optimisation (c'est-à-dire à chaque état intermédiaire évoluant vers celui, le plus probable, qui minimise la fonction coût), leur principal inconvénient réside dans leur temps de calcul. Ces méthodes rencontrent aussi parfois des difficultés à extraire de manière optimale toute l'information contenue dans l'observation.

L'avènement, depuis quelques années, de méthodes basées sur la Donnée (apprentissage statistique / machine au sens large) ouvre de nouvelles perspectives dans le domaine de la télédétection, notamment lorsque le nombre d'observations est important. Dans ce cas, les méthodes « data-driven » permettent souvent d'obtenir une estimation très précise de la cible (ici, les propriétés de la colonne atmosphérique) avec des temps de calcul très courts. Dans ces approches, le coût de calcul est en effet déporté vers la phase

d'entraînement, de sorte que le modèle obtenu au terme de la phase d'apprentissage est parfaitement compatible avec un contexte opérationnel.

L'objectif de cette thèse sera donc d'explorer les capacités des méthodes d'apprentissage (« data-driven » et « physics-informed ») pour la restitution de paramètres atmosphériques, principalement nuageux, tels que la position de la couche nuageuse, son extension verticale et/ou horizontale, le profil de vapeur d'eau condensée dans la couche, ainsi que la forme des particules lorsque le nuage est constitué de glace. Nous utiliserons pour cela les mesures de la plateforme EPS-SG1, lancée en août 2025, et plus spécifiquement les instruments IASI-NG (sondeur infrarouge à haute résolution spectrale), MetImage (radiomètre multi-longueurs d'onde, du visible à l'infrarouge) et 3MI (polarimètre multidirectionnel et multi-longueurs d'onde, du visible au proche infrarouge). Les codes de transfert radiatif, ou modèles directs, utilisés lors de la phase d'entraînement seront RTTOV pour simuler les mesures de l'instrument IASI-NG, et ARTDECO pour les deux autres instruments. Le travail de thèse proposé s'effectuera en trois grandes phases. Dans un premier temps, il s'agira de proposer différents prototypes de modèles neuronaux (depuis du pur « data-driven » jusqu'à des approches informées par la Physique). Ces prototypes seront ensuite utilisés, dans un second temps, afin d'estimer un jeu de paramètres « classiques » traditionnellement obtenu par des méthodes physiques. Ceci permettra de tester leurs rapidités d'exécution, mais également leurs précisions. Enfin, dans un troisième temps, nous complexifierons les cibles ou le jeu de paramètres atmosphériques, afin d'évaluer la capacité de ces méthodes à extraire le maximum d'information possible. Les approches seront évaluées sur des jeux de données synthétiques, avant d'être appliquées à des mesures expérimentales.