
Extension temporelle et robustesse instrumentale des modèles de Deep Learning pour la météorologie de l'espace : étude du modèle SERENADE

Guillaume Bernoux*¹, Maria Tahtouh², Antoine Brunet², and Gautier Nguyen³

¹DPHY, ONERA, Université de Toulouse [Toulouse] – ONERA – France

²DPHY, ONERA, Université de Toulouse [Toulouse] – ONERA, Communauté d'universités et établissements de Toulouse – France

³DPHY, ONERA, Université de Toulouse [Toulouse] – ONERA – France

Résumé

Cette présentation traitera de SERENADE, un pipeline d'apprentissage profond permettant de prédire l'indice géomagnétique Kp maximal quotidien plusieurs jours à l'avance à partir d'images SDO/AIA EUV. Les premiers prototypes, publiés en 2022 (Bernoux et al., 2022, JGR : SP), avaient déjà atteint des performances comparables à celles des meilleurs modèles pour les événements liés au vent solaire rapide, mais plusieurs limites avaient été identifiées. Des travaux récents (Tahtouh et al., 2025, JGR : MLC) ont remplacé l'extracteur de vecteurs latents générique GoogLeNet par un Auto-Encodeur Variationnel spécialement entraîné sur des images solaires, ce qui a permis d'obtenir un espace latent plus significatif sur le plan physique, et donc des prévisions plus stables.

L'un de nos principaux axes de travail actuels est l'enrichissement de la base de données d'apprentissage de SERENADE. Pour dépasser les limites imposées par la base de données AIA " ML-ready " d'origine (Galvez et al., 2019, ApJS), couvrant la période 2010-2020, nous avons constitué un nouvel ensemble de données destiné à l'apprentissage automatique, couvrant la période 2010-mi-2025 (qui inclut donc les phases ascendante et maximale du cycle solaire actuel, plus actif), ainsi qu'un ensemble de données SOHO/EIT complémentaire, qui remonte à 1996. Nous disposons ainsi de près de 30 ans d'observations prétraitées de façon homogène, au lieu des 11 ans initiaux.

Nous analysons l'impact de cette extension temporelle et évaluons dans quelle mesure notre modèle précédent a pu être limité non pas en raison de son architecture, mais des données utilisées. Étant donné que de nombreuses études s'appuient sur l'utilisation de l'ensemble de données SDOML, nos résultats sont potentiellement généralisables à d'autres modèles et pourraient indiquer que des gains de performance pourraient être obtenus sans modifier l'architecture de ces modèles, mais simplement en étendant la base de données.

De plus, nous profitons du fait de disposer de deux ensembles de données provenant d'instruments différents pour mener une étude préliminaire visant à caractériser la capacité du modèle SERENADE à généraliser à des données non-vues provenant d'autres instruments que ceux utilisés pour l'entraînement initial. Cette capacité deviendra cruciale une fois que les missions SDO et SOHO auront pris fin et que le modèle ne pourra plus être alimenté par ces données.

*Intervenant