

PROJET DE FIN D'ETUDES

INGENIEURS DE L'ECOLE NATIONALE DE LA METEOROLOGIE

FICHE DE PROPOSITION DE SUJET

Titre du sujet proposé : Évaluation et consolidation d'un algorithme d'apprentissage supervisé pour estimer la taille maximale des grêlons associés à un objet orageux.

Organisme ou service proposant le sujet : DESR/GMME/PRECIP et Département Prévision Immédiate (Météo-France/DirOP/PI)

Responsable principal du stage :

NOM : AUGROS
téléphone : 05 61 07 84 89

Prénom : Clotilde
Mél : clotilde.augros@meteo.fr

Autres responsables :
Thibaut MONTMERLE, DirOP/PI/D, thibaut.montmerle@meteo.fr

Le stage présente-t-il un caractère de confidentialité ? : Non

Le stage peut-il être effectué à distance ? : Non

1) Objectifs scientifiques et/ou techniques poursuivis – livrable(s) attendu(s)

La grêle contribue de manière importante aux pertes assurées associées aux orages violents dans le monde chaque année. Elle a représenté le phénomène naturel le plus coûteux en France en 2022. La prévision et la détection en temps réel de grêle et de sa taille sont également essentielles afin d'émettre des alertes, et minimiser les dommages potentiels.

À Météo-France, deux algorithmes sont utilisés en opérationnel pour la détection de la grêle par radar : l'algorithme de logique floue Al-Sakka et al (2013) qui prend en entrée des variables radar polarimétriques et restitue un type de précipitation (dont la grêle) au niveau de chaque point de mesure radar, et l'algorithme HYDRE qui estime le type de précipitations au sol, et combine pour la grêle le diagnostic issu de l'algo Al-Sakka et une probabilité de grêle (POH) à partir de la hauteur maximale d'une certaine valeur de réflectivité radar (echotop35 ou 45 dBZ) au-dessus de l'isotherme 0°C. Les performances de ces algorithmes sont assez satisfaisantes en termes de détection de l'occurrence de grêle mais sont très imprécises sur la détection de la taille des grêlons. La classe maximale estimée inclue tous les grêlons supérieurs à 2 cm, et ne distingue pas les tailles supérieures qui pourtant ont le plus de conséquences en termes de dommages, notamment à partir de 5 cm.

On dispose à Météo-France depuis 2023 de l'accès à la base de données ESWD (European Severe Weather Database) qui collecte et valide les données des réseaux sociaux de taille de grêle, ce qui nous permet de pouvoir évaluer un diagnostic de taille de grêle amélioré. Des travaux sur la détection de l'occurrence et de la taille de grêle à partir d'un algorithme IA basé

sur un réseau de neurones convolutionnels (CNN) appliqués à des données radar ont ainsi pu être menés dans l'équipe PRECIP au CNRM, dans le cadre de la thèse de Vincent Forcadell (2021-2024) et du stage de fin d'étude ENM de Maxandre Ouradou (2024). Ces travaux ont permis de constituer une base de données soigneusement validée de rapports de grêle sévère de l'ESWD (et de rapports de pluie ou grêle non-sévère). Un CNN prenant en entrée différentes variables et diagnostics radar dans un voisinage restreint ciblant les réflectivités maximales a ensuite été implémenté. Ses performances ont été comparées à d'autres algorithmes plus classiques, tels que MESH (Maximum Estimated Size of Hail : Witt 1998, Morillo et Homeyer, 2019) qui est un algorithme s'appuyant sur des données de réflectivité 3D et des isothermes de température 0°C et -20°C, et qui donne directement une taille de grêlon. Les résultats sont prometteurs et ont été publiés (Forcadell et al, 2024).

Le stage proposé s'inscrit dans la continuité de la thèse de Vincent Forcadell et du stage de Maxandre Ouradou. **Le premier objectif est de consolider les travaux effectués en complétant la base de données avec tous les cas d'orage 2024.** Il sera intéressant en particulier d'évaluer l'algorithme sur les nouveaux cas de grêle, mais également de vérifier si les prédicteurs les plus informatifs du CNN sont bien les mêmes quand on ajoute des données supplémentaires.

Parallèlement à ces travaux prospectifs sur la grêle, le département DirOP/PI produit opérationnellement les objets orageux OPIC (Objets pour la Prévision Immédiate de la Convection), identifiés et suivis toutes les 5 min dans les images de réflectivités radar. La lecture objet synthétise l'information 2D de l'image composite et facilite le suivi et l'anticipation en temps réel des cellules convectives en France, tout en permettant le calcul de nombreux attributs. Ces attributs caractérisent la géométrie de l'objet, par exemple l'aire de sa surface. Ils caractérisent également son activité convective (réflectivité maximale), activité qui peut également être déduite d'observations co-localisées, par exemple les précipitations au sol, la densité d'éclairs, la présence ou non d'un mésocyclone défini par la mesure Doppler, etc. L'information sur la grêle (pourcentage de petits, moyens ou gros grêlons, surfaces associées) est actuellement fournie par le produit de fusion de données HYDRE.

On se propose dans le cadre de ce stage de capitaliser sur cette production d'objets orageux pour ne focaliser le calcul par IA de la taille de la grêle que sur l'environnement restreint ciblant l'OPIC. En effet, pour des questions de temps de calcul, cette méthodologie basée sur le CNN n'est pas applicable en chaque point d'une grille couvrant la métropole. **Le deuxième objectif est ainsi d'associer une estimation de taille de grêle faite par le réseau de neurones aux objets orageux OPIC.** Cette estimation viendra compléter les attributs existants des OPIC par une estimation robuste de la taille des grêlons, notamment pour les plus volumineux. Ce nouveau diagnostic participera ainsi directement à la caractérisation de la violence générale de l'orage ainsi que dans celle des aléas qui lui sont associés.

Un rapport présentant l'approche et les différents résultats à diverses étapes sera rédigé. Le stage se conclura par une présentation générale devant jury.

2) Méthodologie envisagée

Une première partie du travail consistera à prendre en main la base de données existante à GMME/PRECIP et à la compléter par des situations de 2024. La base de l'ESWD sera en particulier exploitée, de manière analogue à ce qu'a réalisé V. Forcadell pendant sa thèse. De nouvelles bases d'apprentissage, de validation et de test seront ensuite définies, pour être ensuite exploitées par le CNN existant. L'apport de ces nouvelles données sur la qualité des prévisions du modèle sera documenté et comparé à une première version entraînée sur les bases constituées de cas observés en 2023.

Différentes situations orageuses pour lesquelles des données de référence existent seront ensuite identifiées pour l'année 2024. Les OPIC co-localisés à un moment de leur trajectoire avec ces rapports de grêle seront identifiés pour chacune de ces situations. Pour chaque OPIC et à chaque instant de leur suivi, les données 3D radar locales seront extraites grâce à la projection des scans volumiques des radars du réseau ARAMIS sur une grille régulière

restreintes, centrée sur l'OPIC. Le code existant à GMME/PRECIP sera exploité dans ce cadre.

Les différents prédicteurs utilisés par le CNN seront alors diagnostiqués, et une taille de grêle associé à l'OPIC sera restituée. Différents scores objectifs seront mis en place pour comparer ces résultats à ce que donnent HYDRE et MESH, ainsi que pour évaluer chacune des méthodes aux références observées issues de la base de données de test constituée en début de stage.

Tous les outils à adapter ou à mettre en œuvre seront en langage python et pourront s'appuyer sur la librairie PyTorch pour la partie construction/adaptation du réseau de neurones.

3) Plan de travail prévisionnel

- Deux premiers mois : études bibliographiques, prise en main de l'environnement et des outils utilisés à GMME/PRECIP, notamment du réseau de neurone, constitution des bases d'entraînement, de validation et de test à partir de la base existante complétée par de nouvelles données issues de l'ESWD sur 2024.
- 3^{ème} mois : entraînement et validation du CNN sur cette nouvelle base, comparaison de ses performances par rapport au modèle existant.
- 4^{ème} et 5^{ème} mois : Mise en place du code permettant de fournir au CNN ses prédicteurs dans le voisinage d'un OPIC, application du modèle, estimation de ses performances sur des cas d'OPIC détectés sur les situations de la base de test, comparaisons à MESH et HYDRE
- Derniers mois : documentation de l'évolution de l'activité grêligène tout au long du cycle de vie de différents OPIC, illustration de l'intérêt de la méthode pour des cas sévères, rédaction du rapport, préparation de la soutenance.

4) lieu du stage, durée ou période

Le stage durera 6 mois à partir de février et se déroulera dans les locaux de GMME/PRECIP à Toulouse. Plusieurs jours dans les bureaux de DirOP/PI seront nécessaires pour prendre en main les OPIC.

Références

Al-Sakka, H., Boumahmoud, A.-A., Fradon, B., Frasier, S. J., & Tabary, P. (2013). A new fuzzy logic hydrometeor classification scheme applied to the French X, C and S-band polarimetric radars. *Journal of Applied Meteorology and Climatology*, 52(2013), 2328–2344.
<http://journals.ametsoc.org/doi/abs/10.1175/JAMC-D-12-0236.1>

Forcadell, V., Augros, C., Caumont, O., Dedieu, K., Ouradou, M., David, C., Figueras i Ventura, J., Laurantin, O., & Al-Sakka, H. (2024). Severe hail detection with C-band dual-polarisation radars using convolutional neural networks. *EGUsphere*, 1–43. <https://doi.org/10.5194/egusphere-2024-1336>

Murillo, Elisa M., and Cameron R. Homeyer. "Severe Hail Fall and Hailstorm Detection Using Remote Sensing Observations." *Journal of Applied Meteorology and Climatology*, vol. 58, no. 5, May 2019, pp. 947–70, <https://doi.org/10.1175/JAMC-D-18-0247.1>.

Witt, A., Eilts, M. D., Stumpf, G. J., Johnson, J. T., Mitchell, E. D. W., & Thomas, K. W. (1998). An Enhanced Hail Detection Algorithm for the WSR-88D. *Weather and Forecasting*.
[https://doi.org/10.1175/1520-0434\(1998\)013%3C0286:AEHDAF%3E2.0.CO;2](https://doi.org/10.1175/1520-0434(1998)013%3C0286:AEHDAF%3E2.0.CO;2)

Espace Confluence DIROP/PI : <http://confluence.meteo.fr/pages/viewpage.action?pageId=319220546>