

## Stage de Master 2 - 2020-2021

# Apprentissage automatique pour l'analyse de simulations de rupture de gouttes

**Encadrants :** Jorge César Brändle de Motta<sup>1</sup>, Simon Bernard<sup>2</sup>, Laurent Heutte<sup>2</sup>

**Laboratoires :**

<sup>1</sup> CORIA, Saint Etienne du Rouvray

<sup>2</sup> LITIS, Saint Etienne du Rouvray

**Mots clés :** Apprentissage Automatique - Classification - Simulation numérique - Physique du spray - Atomisation

**Contexte :**

L'atomisation est présente dans de nombreuses applications telles que l'injection de carburant, l'irrigation par aspersion, le séchage par pulvérisation ou l'extinction des incendies. Ce processus est généralement divisé en deux régimes : le régime d'atomisation primaire, où le jet se déstabilise en créant de grandes structures liquides et le régime d'atomisation secondaire, où ces grandes structures se divisent en petites gouttes sphériques. Ces gouttelettes finales restent sphériques à cause de l'effet dominant de la tension de surface. La prédiction précise de la distribution finale de la taille des gouttes est nécessaire pour améliorer les applications concernées par l'atomisation. En particulier, la taille des gouttes est l'un des principaux facteurs des émissions finales des moteurs automobiles.

Afin de réaliser cette prédiction il faut établir des modèles de rupture secondaire qui permettent de savoir, à partir de l'état d'une goutte, la probabilité de rupture. De nombreux modèles ont été proposés dans la littérature, voir [1] pour un examen détaillé. Les modèles actuels sont basés sur des corrélations empiriques prenant en compte des paramètres comme la turbulence environnante ou la vitesse de glissement. La diversité des effets provoquant la rupture fait qu'à ce jour aucun modèle n'a permis une prédiction fiable sur l'ensemble de régimes utilisés dans l'industrie.

**Le but de ce projet est d'utiliser les algorithmes d'apprentissage machine (Machine Learning) pour produire de meilleures prédictions de l'atomisation secondaire.** Pour ce faire, le stagiaire devra explorer les outils de l'apprentissage automatique à partir de la base de données générée à l'aide du code ARCHER suivant la méthodologie présentée dans [2-3], dont nous donnons un exemple en figure 1.



*Figure 1: Evolution d'une goutte allant jusqu'à la rupture*

Cette base de données, et la tâche de prédiction associée, posent plusieurs défis à l'apprentissage automatique. Tout d'abord, le nombre de gouttes étudiées est limité par le coût de génération qui reste non négligeable. On envisage la génération de quelques milliers de gouttes pour réaliser notre étude, ce qui, au regard de la tâche d'apprentissage est relativement faible. En particulier, ce volume de donnée est bien en deçà des volumes généralement nécessaires à l'apprentissage profond (Deep Learning). Bien que ce soit difficile à estimer de façon précise, car dépendant à la fois du problème et de la méthode utilisée, nous savons que la quantité de données nécessaire à l'apprentissage profond est plutôt de l'ordre de plusieurs millions de données. Ensuite, chaque donnée (goutte) de la base d'apprentissage est décrite par un volume

de 64x64x64 voxels, eux-même décrits par 6 descripteurs numériques. Cela implique que la dimension du problème d'apprentissage est de l'ordre du million, ce qui est bien supérieur aux nombre de données disponibles. Ces situations sont connues pour être particulièrement difficiles à traiter en apprentissage machine et nécessite des méthodes dédiées. Finalement, la base d'apprentissage sera constituée d'un nombre très inégal de gouttes qui subissent une rupture au cours du temps et de gouttes stables. Ce déséquilibre est une difficulté supplémentaire, qui nécessite également des méthodes d'apprentissage spécifiques.

Ces trois problématiques d'apprentissage, i.e. l'apprentissage en grande dimension, avec de petits échantillons et en présence de données débalancées ont déjà été étudiés par le passé au Laboratoire LITIS, et solutionnées notamment avec des méthodes de forêts aléatoires (voir [4] à nouveau) et d'ensembles de classifieurs, voir [5-6]. Ces travaux ont démontré l'efficacité de ces approches pour ces types de problèmes, c'est pourquoi nous les envisageons pour ce projet.

Ce stage sera financé par l'ANR DropBreak et peut déboucher sur un contrat d'ingénieur d'études de 6 mois financées dans le cadre d'un projet RIN (Région Normandie - Union Européenne). Le contenu de ce contrat de 6 mois est dans la continuité des tâches proposées dans le stage.

#### **Tâches à réaliser durant le stage :**

- Prise en main de la base de données (analyse statistique, extraction des données)
- Application des algorithmes d'apprentissage machines.

#### **Résultats attendus pour la rédaction du rapport de stage :**

- Étude bibliographique des outils d'apprentissage machine adaptées,
- Préconisation d'une procédure d'apprentissage adaptée pour la problématique physique posée.

#### **Profil du candidat :**

- Étudiant en Master 2 ou dernière année d'école d'ingénieurs.
- Maîtrise des algorithmes d'apprentissage machine.
- Goût pour la physique.

#### **Pour candidater :**

Envoyez un CV et une lettre de motivation à [jorge.brandle@coria.fr](mailto:jorge.brandle@coria.fr), [simon.bernard@univ-rouen.fr](mailto:simon.bernard@univ-rouen.fr) et [laurent.heutte@univ-rouen.fr](mailto:laurent.heutte@univ-rouen.fr). N'hésitez pas à signaler vos références, notamment les enseignants pouvant attester de votre niveau dans les matières concernées par ce stage.

#### **Références bibliographiques :**

- [1] J. Solsvik, S. Tangen, and H. A. Jakobsen, "On the constitutive equations for fluid particle breakage," *Rev. Chem. Eng.*, vol. 29, no. 5, 2013.
- [2] T. Chen, V. Chéron, G. Zhaoli, J. C. Brändle de Motta, T. Ménard, and L.-P. Wang, "Simulation of Immiscible Two-Phase Flows Based on a Kinetic Diffuse Interface Approach," in 10 th International Conference on Multiphase Flow, Brazil, 2019, p. 10. <https://hal.archives-ouvertes.fr/hal-02315240>
- [3] V. Chéron, J. C. Brändle de Motta, G. Vaudor, T. Ménard, and A. Berlemont, "From droplets to particles: Transformation criteria," in 29th European Conference on Liquid Atomization and Spray Systems, Paris, 2019, p. 8 [Online]. Available: <https://hal.archives-ouvertes.fr/hal-02315145>
- [4] L. Breiman, "Random forests," *Machine Learning*, 45: 5, 2001.
- [5] C. Désir, S. Bernard, C. Petitjean, L. Heutte. "One class random forests," *Pattern Recognition* 46 (12), 3490-3506, 2013.
- [6] H. Cao, S. Bernard, R. Sabourin, L. Heutte. "Random forest dissimilarity based multi-view learning for Radiomics application," *Pattern Recognition* 88, 185-197, 2019.