



Sujet de stage de fin d'études à EDF R&D :

Exploitation des méthodes à noyaux dans un espace de projections par shapelets pour le traitement de séries temporelles

Entreprise	EDF R&D (EDF SA)
Entité d'accueil	Département PRISME ¹ / Groupe P17 GAIA ²
Lieu principal	EDF Lab Chatou, 6 Quai Watier, 78401 Chatou
Co-encadrement	Julien PELAMATTI (Ing.-Chercheur) Nicolas BOUSQUET (Chercheur expert)
Partenariat académique	Antoine GUILLAUME (ENSTA, chercheur en post-doc)
Date de début et durée du stage	Début possible > 01/03/2024 Durée de 6 mois
Mots-clés	Shapelets, méthodes à noyaux, hydrogrammes, XAI, Apprentissage supervisée, classification

Environnement de recherche industrielle à EDF R&D / PRISME. Au sein d'EDF R&D, le département PRISME a pour mission de proposer des solutions innovantes pour une exploitation plus performante des différents moyens de production d'électricité du groupe EDF : maîtrise des risques, amélioration de la sûreté, optimisation des performances technico-économiques et maîtrise de la durée de vie des installations. La thématique du **traitement des incertitudes** (*uncertainty quantification*) désigne l'ensemble des méthodes probabilistes et statistiques dédiées à la modélisation des incertitudes, à leur quantification et à leur propagation dans les codes de simulation numérique. En complément de cette thématique, dans le cas où des données d'observations (mesures expérimentales, données de fonctionnement) sont à disposition, des approches de **traitement de la donnée** (apprentissage statistique ou *machine learning*) peuvent être utilisées afin de conforter la modélisation physique et la modélisation probabiliste des variables d'entrée. Ces deux thématiques, traitement des incertitudes et traitement de la donnée, font l'objet de travaux de recherche intenses au sein du département PRISME et jouent un rôle central et transverse au sein de nombreuses activités du groupe EDF.

Contexte & enjeux Dans le contexte du traitement de séries temporelles, aussi bien comme données d'entrée que comme quantités d'intérêt, l'explicabilité des modèles numériques et/ou des mo-

1. PRISME : Performance, Risque Industriel et Surveillance pour la Maintenance et l'Exploitation

2. GAIA : Gestion d'Actifs, Incertitudes et Apprentissage statistique

dèles de machine learning joue un rôle clé dans une optique d'industrialisation de ces modèles et de leur justification auprès des autorités de sûreté. Des travaux récents sur le sujet ont été menés au sein de EDF R&D (et à travers différents partenariats), cependant, de nombreux verrous restent présents, ainsi que de nombreuses pistes à creuser. Le cas d'application considéré pendant ce stage sera l'analyse et le traitement d'hydrogrammes de crues en collaboration avec la Direction Technique Générale (DTG) de EDF, dans un contexte d'identification automatique d'hydrogrammes non-crédibles et dans un second temps de génération d'hydrogrammes synthétiques.

Ce stage s'inscrit également dans le contexte du Groupement d'Intérêt Scientifique LARTISSTE ([site](#)), ce qui permettra des échanges et potentiellement des collaborations méthodologiques avec les acteurs industriels et académiques du Plateau de Saclay travaillant sur des sujets en lien avec la quantification des incertitudes dans les simulations numériques.

Contenu technique L'objectif de ce stage est d'étudier les potentielles applications et les performances des méthodes permettant de traiter des séries temporelles se basant sur la projection de ces séries sur des espaces définis par des shapelets (ceci fait suite, entre autres, aux travaux de thèse de Antoine Guillaume (GUILLAUME et al., 2022)). En pratique, l'idée est de considérer non pas les séries temporelles entières, mais leurs patterns les plus représentatifs et / ou les plus fréquents, représentés sous forme de shapelet. Dans ces espaces de projection, il est possible d'appliquer facilement des méthodes de machine learning classiques, afin de définir des modèles d'apprentissage supervisés ou non-supervisés traitant ces séries temporelles. Un des avantages principaux de ces approches est que la projection sur une base de shapelets permet de se placer dans un espace qui est intrinsèquement explicable, en termes de patterns représentatifs et récurrents, permettant ainsi de comprendre et justifier le comportement des modèles de machine learning construits. Lors de ce stage, un intérêt particulier sera porté aux méthodes de machine learning dites à *noyaux*, peu étudiées dans le contexte spécifique des shapelets, et pour lesquelles il serait intéressant de travailler de façon jointe sur la définition de la projection des séries temporelles et sur la définition des noyaux utilisés, permettant idéalement d'en améliorer les performances et la robustesse. Cette dernière piste étant relativement peu explorée dans la littérature scientifique

La première partie du stage se concentrerait sur la définition d'un modèle de classification permettant de distinguer les hydrogrammes crédibles de ceux non crédibles, de façon à automatiser le traitement des données. Dans un second temps, on se concentrera sur la définition et l'application de mesures d'importance à base de noyau dans l'espace de projection par shapelet, avec pour but d'identifier les patterns les plus influents sur la prédiction du classifieur, i.e., les patterns les plus influents sur l'identification des hydrogrammes non-crédibles. Un travail sur la définition jointe de la projections des séries temporelles et des noyaux de covariance dans l'espace de projection sera sans doute nécessaire. Finalement, une confrontation avec une expertise métier pourra potentiellement permettre de valider et de justifier l'explication du comportement du modèle ainsi obtenu.

Profil recherché. Étudiant.e de M2 (mathématiques appliquées / probabilités-statistiques) ou d'écoles d'ingénieur.e.s (avec majeure en mathématiques appliquées / probabilités-statistiques).

Comment postuler ? Envoyer un CV détaillé et une lettre de motivation sur <https://www.edf.fr/edf-recrute> ou via les adresses suivantes, en précisant l'objet de votre envoi :

- julien dot pelamatti at edf dot fr
- nicolas dot bousquet at edf dot fr

Références

GUILLAUME, Antoine, Christel VRAIN et Wael ELLOUMI (2022). "Random dilated shapelet transform: A new approach for time series shapelets". In : *International Conference on Pattern Recognition and Artificial Intelligence*. Springer, p. 653-664.