

# FICHE DE PROPOSITION DE SUJET POUR UNE CONVENTION CIFRE

**Entreprise :** Chantiers de l'Atlantique.

**Service :** Performance énergétique, Equipe R&D du programme ECORIZON

**Tuteurs proposés :** Matthieu Lorang, responsable du service performance énergétique et Arnaud Jacques, ingénieur efficacité énergétique.

**Direction de thèse :** Romain Billot (Maître de conférences, habilité à diriger des recherches, IMT Atlantique Brest, département LUSSE) et Patrick Meyer (Professeur, habilité à diriger des recherches, IMT Atlantique Brest, département LUSSE).

**TITRE :** Fouille de données et aide à la décision pour la supervision et la prévision court-terme de l'efficacité énergétique d'un paquebot.

**1- Contexte et thème de recherche**(dans les grandes lignes ; si vous le souhaitez, vous pouvez joindre une explication complète en annexe)

## **Contexte**

Les Chantiers de l'Atlantique de Saint-Nazaire bénéficient d'une très longue expertise dans la conception et la réalisation de grands navires complexes à forte valeur ajoutée. Des paquebots prestigieux, des navires militaires ou spéciaux et des équipements lourds pour les Energies Marines Renouvelables ou Oil & Gas y sont construits depuis des décennies.

Les chantiers poursuivent également la diversification de leurs activités dans les domaines de l'ingénierie et des services aux navires en exploitation. Le groupe comprend 2500 salariés en France et réalise un chiffre d'affaires d'environ 1,4 milliard d'euros, majoritairement à l'export. Chantiers de l'Atlantique mène une politique active de R&D (développements techniques, procédés et nouveaux concepts de navire) pour anticiper les besoins de ses clients et conforter sa position de leader du marché. Dans ce cadre, la réduction de l'impact environnemental des navires (efficacité énergétique, émissions dans l'air et dans l'eau, traitement des déchets, écoconception) est un élément déterminant pour ses armateurs. Afin de répondre à cette demande, Chantiers de l'Atlantique doit développer de nouvelles solutions pour continuer à améliorer l'efficacité énergétique des futurs navires, tout en prenant en compte les contraintes d'intégration, de conception et de réglementation.

L'équipe R&D du programme ECORIZON a développé un modèle physique de la consommation énergétique qui permet d'estimer cette consommation à partir de données d'entrée relatives au profil de croisière et des conditions météorologiques. Ensuite, une fois le navire en exploitation par le client, des rapports hebdomadaires de consommation sont envoyés depuis le navire à l'équipe ECORIZON qui prépare des tableaux de bord mis à la disposition du client. Ces outils d'analyse, mis en avant dès

2016 pour la livraison du paquebot *Harmony of the Seas*, tendent à s'améliorer et se complexifier au gré des livraisons successives. Les livraisons successives du *Symphony of the Seas* (Mai 2018) et du *Celebrity Edge* (prévu Octobre 2018) bénéficieront également de ces méthodes d'analyse.

En parallèle, l'équipe réalise depuis plusieurs années des audits énergétiques à bord de paquebots en exploitation et du suivi de navire en opération. L'ensemble de ces prestations nécessite l'acquisition et le traitement de nombreuses données.

### **Thèmes de recherche de la thèse**

Alors qu'environ 250 variables sont actuellement récoltées chaque semaine à des pas de temps horaires, il est attendu que ce nombre augmente significativement à court terme, allant de pair avec un raffinement du pas d'agrégation, passant de 1h à 5min. Cette évolution fait basculer le problème dans un paradigme *big data*, qui nécessite une modernisation des outils actuels. L'objectif premier de la thèse CIFRE sera de mettre en place un **outil de big data** pour la gestion de l'ensemble des données mesurées à bord, pour ensuite améliorer l'analyse de ces données. Ensuite des techniques récentes de **machine learning** seront mobilisées pour améliorer d'une part la supervision de la performance énergétique du navire (analyses multidimensionnelles des facteurs clés de la consommation énergétique) et d'autre part la prévision court terme de ces performances. En terme de prévision, un verrou scientifique concerne le couplage entre un modèle physique (le modèle STX d'estimation de la consommation) et un modèle statistique (une méthode de *machine learning* fondée sur de la reconnaissance de formes dans les historiques de données) selon les situations et la disponibilité des données. Enfin, un dernier champ de recherche concerne la restitution des résultats issus des analyses prédictives vers les différents acteurs de ce système complexe qu'est l'exploitation d'un navire. Les ingénieurs experts du programme ECORIZON, les exploitants hôteliers du navire, le capitaine du navire, constituent un ensemble d'acteurs hétérogènes qu'un outil **d'aide à la décision** doit intelligemment prendre en compte dans sa construction. L'analyse multicritère des décisions est un outil théorique approprié à cette problématique ambitieuse.

**Mots-clés** : fouille de données, aide à la décision, *big data*, énergie, mer.

### **2- Argumentez l'intérêt du travail de recherche pour l'entreprise**

Le travail de recherche revêt de forts enjeux pour les Chantiers de l'Atlantique. Alors que l'activité de service aux navires en exploitation se développe, l'entreprise s'attend à une explosion du volume de données à laquelle il faut faire face. De plus, certains processus doivent être automatisés car le temps passé par les ingénieurs sur les phases d'analyses est trop important. Enfin, et surtout, l'offre des Chantiers de l'Atlantique doit s'orienter vers une analyse quasi temps réel et une prévision court terme de la performance énergétique du navire. Les méthodes envisagées vont au-delà d'un simple travail d'ingénierie, d'autant plus que l'entreprise souhaite un modèle final qui couple intelligemment l'estimation existante, par composante (approche *model driven*) et l'apprentissage des historiques de données (approche *data driven*). La thèse doit apporter de réelles contributions scientifiques à la problématique de la performance énergétique d'un navire.

### **3- Description des missions**

Le planning prévisionnel de la thèse comporte les étapes suivantes :

1. **Migration vers un outil big data** : actuellement, quelques fichiers Excel sont envoyés chaque semaine à l'ingénieur responsable du suivi énergétique aux chantiers, qui a la seule maîtrise du stockage et de l'analyse des données. Alors que quelques milliers de données et des centaines de variables sont collectées, un fort changement d'échelle est attendu dans les prochaines années. La première étape consiste à effectuer un benchmark des solutions *big data* disponibles (si possible *open source*) afin de migrer vers une architecture distribuée. La sécurité des données dans un contexte industriel concurrentiel et sensible est un élément primordial.
2. **Perfectionnement des méthodes d'analyse à l'aide du *data mining*** : le doctorant explorera les données disponibles à l'aide de techniques de fouille de données qui permettront d'obtenir une vision multidimensionnelle des données de consommation énergétiques. Les conclusions des analyses seront synthétisées dans de nouveaux tableaux de bord qui incluront des résumés statistiques standards et des résultats plus poussés (exemple : *clustering* spatiotemporel des facteurs de consommation d'un navire). Avec l'affinement de la granularité des données, ces tableaux de bord feront évoluer peu à peu la solution des Chantiers de l'Atlantique vers une supervision temps réel de la performance énergétique d'un navire. Les méthodes statistiques développées comprennent l'identification de déviations (anomalies) et la recherche de causes, ainsi que la comparaison des performances du navire avec d'autres navires en exploitation.
3. **Prise en compte des historiques de données pour une prévision court terme de la consommation énergétique** : ce bloc de la thèse consiste à appliquer des techniques récentes de *machine learning* (apprentissage supervisé) pour prévoir à court terme la consommation d'un navire. Ce module d'apprentissage prendra en entrée les prévisions météorologiques et de croisière de la prochaine semaine de navigation afin de prédire au mieux la consommation et les besoins énergétiques. Au niveau méthodologique, des méthodes ensemblistes type *gradient boosting machine* (GBM) sont envisagées (Chen *et al*, 2016).
4. **Couplage entre modèle physique d'estimation de la consommation et machine learning**. L'étape suivante consiste à développer une approche originale qui combine une estimation précise de la consommation de type *model driven* (le module existant des Chantiers de l'Atlantique) avec des techniques de *machine learning* développées par le doctorant. Le doctorant devra effectuer un état de l'art conséquent sur le couplage entre approches *data driven* et *model driven*. En fonction de la qualité des données d'entrée, de la profondeur de l'historique de données par rapport aux diverses prévisions, la méthode la plus efficace doit prendre le pas sur l'autre pour fournir la prévision court terme la plus précise possible.
5. **Modèle d'aide à la décision multicritères (AMCD)** : ce dernier objectif de la thèse vise à développer des stratégies d'aide à la prise de décision pour le comportement d'exploitation du navire (par l'équipe hôtelière), ou de croisière (par le capitaine) à l'aune des sorties des modules de prévision de court terme. Ce module doit prendre en compte les préférences des différents acteurs du système (experts des chantiers, exploitants du navire, capitaine etc.).

#### **4- Verrous scientifiques**

1) La diversité des sources de données en présence place le problème de fusion et qualification de données dans une problématique *big data* (voir billot *et al.*, 2018). Le candidat devra être en mesure de créer une architecture distribuée flexible afin de stocker efficacement les données, puis dans un deuxième temps, qualifier les différentes sources à l'aide de méthodes statistiques ou de fusion de données.

2) Le module de prévision court terme doit mener à un couplage innovant entre méthodes de *machine learning* et modèles physiques pour l'estimation à une semaine n+1 de la consommation énergétique d'un navire. Le verrou scientifique réside dans le couplage entre le modèle physique existant et les techniques ensemblistes d'apprentissage supervisé. La richesse du modèle existant développé par l'équipe ECORIZON permettra d'aller beaucoup plus loin que des approches récentes de l'état de l'art, qui utilisent un nombre restreint de paramètres (voir Besikci *et al.*, 2016, Wang *et al.*, 2017).

4) La multiplicité et l'hétérogénéité des acteurs en présence constituent un exemple d'application ambitieuse de prise de décision dans un système complexe (voir Bisdorff *et al.*, 2015). Le développement d'une aide à la décision multicritères fondée sur les préférences permettra d'aller plus loin qu'une simple somme pondérée de paramètres, à travers une fine analyse des dynamiques préférentielles, en amont de l'envoi de recommandations pour la future semaine de navigation du navire.

#### **5- Compétences requises du futur thésard**

##### ***Connaissances spécifiques :***

- Fouille de données/Machine Learning.
- Outils *big data*.
- Recherche opérationnelle et optimisation.
- Une connaissance de la filière énergétique serait un plus.

##### ***Formation souhaitée :***

- Ingénieur généraliste, Option Génie Industriel ou Data Science.
- Master Recherche en Génie Industriel, Data Science ou Recherche Opérationnelle.

##### ***Aptitudes personnelles souhaitées :***

- Dynamique et force de proposition
- Rigoureux
- Autonome et Transversal
- Attrait pour des problématiques techniques pointues
- Innovant
- Maîtrise de l'anglais indispensable
- Une connaissance du milieu maritime est un plus.

#### **6- Références bibliographiques**

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of

the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794). ACM.

Billot, R., Bothorel, C., & Lenca, P. (2018). Introduction to Big Data and Its Applications in Insurance. *Big Data for Insurance Companies*, 1, 1-25.

Beşikçi, E. B., Arslan, O., Turan, O., & Ölçer, A. I. (2016). An artificial neural network based decision support system for energy efficient ship operations. *Computers & Operations Research*, 66, 393-401.

Wang, S., Ji, B., Zhao, J., Liu, W., & Xu, T. (2017). Predicting ship fuel consumption based on LASSO regression. *Transportation Research Part D: Transport and Environment*.

Bisdorff, R., Dias, L. C., Meyer, P., Mousseau, V., & Pirlot, M. (2015). *Evaluation and Decision Models with Multiple Criteria*. Springer Berlin Heidelberg: Imprint: Springer,.