

REF : DIT 05-2023

Pour postuler : [rim.kaddah@irt-systemx.fr](mailto:rim.kaddah@irt-systemx.fr), [jesse.read@polytechnique.edu](mailto:jesse.read@polytechnique.edu), Merci de mentionner le numéro de référence

## Thèse de Doctorat

### Titre de la thèse

Estimation probabiliste des indicateurs de santé de systèmes complexes

### Contexte de la thèse

Au sein de l'Institut de Recherche Technologique SystemX, situé au cœur du campus scientifique d'excellence mondiale de Paris-Saclay, vous prendrez une part active au développement d'un centre de recherche technologique de niveau international dans le domaine de l'ingénierie numérique des systèmes. Adossé aux meilleurs organismes de recherche français du domaine et constitué par des équipes mixtes d'industriels et d'académiques, ce centre a pour mission de générer de nouvelles connaissances et solutions technologiques en s'appuyant sur les percées de l'ingénierie numérique et de diffuser ses compétences dans tous les secteurs économiques.

Plus particulièrement au sein de l'IRT SystemX, le doctorant sera rattaché au domaine scientifique « Sciences des données et Intelligence Artificielle ». Le sujet de thèse a été initié par le consortium réuni dans le cadre du projet « [Jumeaux Numériques pour la Maintenance Prédictive et Indicateurs de Santé \(JNI3\)](#) » du programme « [Jumeaux Numériques pour la résilience et la durabilité des systèmes Industriels](#) » (JNI) de l'IRT SystemX. Ce projet vise à développer et mettre en œuvre des Jumeaux Numériques (JNI) pour le suivi de l'évolution de l'état de santé des systèmes industriels. Il se focalise sur la définition de jumeaux numériques capables d'acquérir des données en temps réel, et de les exploiter pour fournir une estimation de l'état de santé du système industriel.

La direction de la thèse sera assurée par Jesse Read du laboratoire d'informatique de l'École Polytechnique (LIX), UMR 7161 et la thèse sera inscrite à l'école doctorale IP Paris (#626). Le doctorant bénéficiera également d'un encadrement scientifique à l'IRT par Rim Kaddah. De plus, il pourra interagir avec des industriels qui proposent des cas d'usage concrets.

Le poste est basé sur le site de l'IRT SystemX à Paris-Saclay, avec des déplacements réguliers au LIX. Hormis les déplacements en conférences internationales, un séjour de longue durée dans un laboratoire à l'étranger est envisagé.

La date souhaitée de démarrage de la thèse est le 01/01/2024. La rémunération de la thèse est de 2784€ brut mensuel sur 3 ans.

### Sujet de thèse

#### CONTEXTE SCIENTIFIQUE

Cette thèse porte sur l'estimation de l'état de santé de systèmes industriels complexes à partir de données de capteurs récoltées au cours du temps pour le suivi de tels systèmes, tout en prenant en compte la physique régissant leur fonctionnement. Elle se focalise sur l'estimation probabiliste pour l'apprentissage d'indicateurs de santé pouvant permettre d'avoir une estimation continue, idéalement en temps réel, de l'état de santé du système, passant d'un état fonctionnel vers un état possiblement dégradé, tout en pouvant intégrer un ensemble d'états intermédiaires.

La prise en compte de la physique régissant le système étudié permet de combiner des caractéristiques invariantes avec des informations apprises à partir de données mesurées/observées. L'objectif de cette hybridation est de quantifier et limiter l'incertitude liée à l'estimation des indicateurs de santé au cours du cycle de vie du système réel. La construction de tels indicateurs de santé du système via la modélisation proposée devra pouvoir mettre en évidence, via de propriétés statistiques quant à la capacité de détection de variables ou de co-variables pertinentes à la prédiction, la pertinence de considérer séparément des indicateurs préconstruits ou via une « combinaison » et ou une « sélection » de sous-ensembles de ceux-ci, selon la pertinence finale sur l'indication de l'état de santé du système étudié.

**RESULTATS ATTENDUS ET APPROCHE SCIENTIFIQUE DE LA THESE**

Les travaux attendus de cette thèse ont pour objectif de développer des méthodes d'apprentissage automatique dotées de constructions probabilistes pour l'estimation d'indicateurs de santé d'un système industriel physique (eg. Moteur aéronautique générique, Centrale d'hydrogène) capables de :

- Exploiter des données acquises à partir du système industriel physique (moteur, centrale d'hydrogène), et fournir une estimation probabiliste de l'état de santé de celui-ci en fonction de différents indicateurs de santé préalablement identifiés par le secteur métier ou sélectionnés par la méthodologie développée;
- Prendre en compte la physique régissant les systèmes étudiés à partir des lois sous-jacentes au système (e.g., lois d'émission d'hydrogène en milieu fermé en vue de la détection de fuite) ou à partir des données simulées de celui-ci (eg. moteur générique d'avion), pour enrichir le modèle d'apprentissage à partir de données observées
- Sélectionner et/ou construire les indicateurs de santé les plus pertinents à l'élaboration du modèle de santé à partir des flux de données collectées sur le système industriel, en tenant compte non seulement de la prédictivité du modèle mais également de son interprétabilité (au sens de la cohérence avec la physique).
- Implémenter une méthode d'estimation de RUL (*Remaining Useful Life*, durée de vie restante) détectant les signes précurseurs de dégradation du système en vue d'optimiser sa durée de vie.

Les directions de recherches qui seront explorées dans cette thèse sont :

- Recours à des approches d'apprentissage statistique permettant de caractériser de façon continue l'évolution incertaine du système étudié à partir de séries temporelles de mesures. Parmi les pistes à explorer durant la phase d'étude bibliographique, l'inférence bayésienne et l'apprentissage par renforcement (éventuellement profond) sont envisagés.
- Sélection / fusion de variables permettant de mettre en évidence les indicateurs (variables) les plus pertinentes à la prédiction ; Typiquement la prédiction à l'instant  $t$  devrait être optimisée pour les différentes variables d'entrée (corrélation, redondance et fusion d'indicateurs éventuellement ... etc), par exemple en optimisant l'espace d'entrée (sélection de variables) et ou l'espace latent et ce par régularisation par exemple ; ...
- Prise en compte de l'aspect temporel des observations et de la prédiction : Il ne s'agit pas seulement d'estimer l'état du système à l'instant  $t$ , mais de prédire l'évolution continue de cet état jusqu'à l'instant  $t+h$  avec  $h > 1$ . Cela implique nécessairement d'implémenter une méthode de quantification d'incertitude à chaque instant, dans une optique de certifiabilité et d'interprétabilité. L'aspect physique du système peut être représenté par les paramètres simulés à partir d'une version générique du système réel (eg. pour le cas de moteur générique) ou par les paramètres des lois simulés ou estimés pour le cas d'application de propagation d'hydrogène dans un milieu fermé.

**Cas d'usage :** Les cas d'usage qui seront étudiés dans cette thèse sont :

- Pour l'estimation de l'état de santé des moteurs aéronautiques (Safran) : La surveillance des moteurs est un sujet bien connu et largement étudié, qui s'avère essentiel pour l'industrie aéronautique. Une approche populaire consiste à construire des indicateurs liés à la physique (par exemple, rendement et débit d'air des modules) reflétant les états de santé des systèmes en exploitant les mesures opérationnelles (c'est-à-dire les données des capteurs pendant les vols comme la température de sortie de gaz, débit du carburant, etc.). Cette approche relève de la catégorie des problèmes inverses, appliquée ici à un modèle du moteur. Estimer l'évolution de ces indicateurs de santé et exploiter/interpréter cette estimation est crucial pour concevoir des stratégies de maintenance prédictive des moteurs aéronautiques. Deux des défis principaux sont d'une part la rareté des données (y compris les limitations dans le choix des capteurs ainsi que le manque d'étiquettes de vérité terrain), et d'autre part la complexité du système (impliquant souvent un grand nombre d'équations thermodynamiques, et des non-linéarités).
- Pour l'estimation en vue de détection de fuites d'hydrogène : Données prélevées à partir de sondes IoT sur banc d'essais

**CALENDRIER PREVISIONNEL DE LA THESE**

Le calendrier prévisionnel de la thèse est le suivant :

- Les 6 premiers mois de la thèse seront consacrés à une étude bibliographique portant sur les principaux axes mentionnés dans le sujet, visant à préciser les premières directions des travaux ;
- De premières expérimentations de compréhension du contexte et de prise en main des questions appliquées et industrielles sur des données provenant des cas d'usages identifiés par le projet JNI3 seront aussi effectuées en cette période de début de thèse ;
- Ensuite, les directions envisagées à la suite de l'étude bibliographique seront approfondies, développées, implémentées et appliquées aux données du projet JNI3 ainsi que sur les données de la communauté ;
- Les travaux effectués seront diffusés sous la forme de publications scientifiques dans les grandes conférences et revues du domaine des sciences de données et de l'IA, mais aussi sous la forme d'outils et de démonstrateurs logiciels dans le contexte du projet JNI3.

#### QUELQUES REFERENCES BIBLIOGRAPHIQUES

- Blanchard, Serge & Cottrell, Marie & Lacaille, Jérôme. (2009). Health monitoring des moteurs d'avions. [Lire en ligne](#)
- Chao, MA, Lilley, DS, Mathé, P, & Schloßhauer, V. "Calibration and Uncertainty Quantification of Gas Turbine Performance Models." *Proceedings of the ASME Turbo Expo 2015: Turbine Technical Conference and Exposition. Volume 7A: Structures and Dynamics*. Montreal, Quebec, Canada. June 15–19, (2015). [Lire en ligne](#)
- Tsirizo Rabenoro. Outils statistiques de traitement d'indicateurs pour le diagnostic et le pronostic des moteurs d'avions. Thèse de doctorant de l'Université Paris 1 Panthéon Sorbonne, 2015. [Lire en ligne](#)
- Bardet, Jean-Marc & Faure, Cynthia & Lacaille, Jérôme & Olteanu, Madalina. (2017). Design Aircraft Engine Bivariate Data Phases using Change-Point Detection Method and Self-Organizing Maps. [Lire en ligne](#)
- Greciet, F. (2020). Régression polynomiale par morceaux pour la propagation de fissures (Doctoral dissertation, Université de Lorraine). [Lire en ligne](#)
- Langhendries, R., & Lacaille, J. (2022). Turbofan Exhaust Gas Temperature Forecasting and Performance Monitoring with a Neural Network Model. In European Conference on Safety and Reliability (ESREL). [Lire en ligne](#)
- Tian, Y., Chao, M., Kulkarni, C., Goebel, K., & Fink, O. (2022). Real-time model calibration with deep reinforcement learning. *Mechanical Systems and Signal Processing*. [Lire en Ligne](#)

#### Profil recherché

De formation bac +5 (Master Recherche ou Ingénieur avec un intérêt avéré pour la recherche) en Informatique, Mathématiques appliquées, Science des données, Apprentissage Automatique, ou équivalent.

##### Connaissances et savoir-faire essentiels :

- Maîtrise des concepts, méthodes et algorithmes de Machine Learning ; données sensorielles, flux de données, séries temporelles, traitement de signaux, inférence probabiliste ;
- Maîtrise des environnements logiciels pour le Machine Learning, et de Python et ses bibliothèques scientifiques par exemple, numpy, pandas, pytorch (ou similaires);

##### Qualités professionnelles :

- Capacité d'analyse, autonomie, esprit d'équipe ;
- Aptitude à communiquer à l'oral et à l'écrit en français et en anglais ;
- Intérêt avéré pour la recherche partenariale et appliquée

#### Contact et comment postuler

Pour postuler, merci de renvoyer les éléments suivants à : [rim.kaddah@irt-systemx.fr](mailto:rim.kaddah@irt-systemx.fr), [jesse.read@polytechnique.edu](mailto:jesse.read@polytechnique.edu)

- CV incluant les coordonnées d'une ou deux référence.s académique.s
- Lettre de motivation
- Relevés de notes des deux années du Master (ou des deux dernières années du cycle ingénieur)
- Au moins une lettre de recommandation