



**HAL**  
open science

# Quantification d'incertitude dans un contexte de maintenance et d'évaluation des risques industriels à l'aide d'apprentissage automatique

Martin Gagnon, Luc Vouligny, Luc Cauchon, Quang Hung Pham, Vincent Mai

## ► To cite this version:

Martin Gagnon, Luc Vouligny, Luc Cauchon, Quang Hung Pham, Vincent Mai. Quantification d'incertitude dans un contexte de maintenance et d'évaluation des risques industriels à l'aide d'apprentissage automatique. Congrès Lambda Mu 24 " Les métiers du risque : clés de la réindustrialisation et de la transition écologique ", Institut pour la Maîtrise des Risques (IMdR), Oct 2024, Bourges, France. hal-04881372

**HAL Id: hal-04881372**

**<https://hal.science/hal-04881372v1>**

Submitted on 12 Jan 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Quantification d'incertitude dans un contexte de maintenance et d'évaluation des risques industriels à l'aide d'apprentissage automatique

## Machine learning uncertainty quantification in an industrial maintenance and risk assessment context

GAGNON Martin

*Institut de recherche d'Hydro-Québec*  
Varenes (Québec), Canada  
gagnon.martin11@hydroquebec.com

VOULIGNY Luc

*Institut de recherche d'Hydro-Québec*  
Varenes (Québec), Canada  
vouligny.luc@hydroquebec.com

CAUCHON Luc

*Institut de recherche d'Hydro-Québec*  
Varenes (Québec), Canada  
cauchon.luc@hydroquebec.com

PHAM Quang Hung

*Institut de recherche d'Hydro-Québec*  
Varenes (Québec), Canada  
pham.quanghung2@hydroquebec.com

MAI Vincent

*Institut de recherche d'Hydro-Québec*  
Varenes (Québec), Canada  
mai.vincent@hydroquebec.com

**Résumé** — L'apprentissage automatique à partir de données provenant de campagnes de mesure ou de systèmes de surveillance permet d'entraîner un modèle à estimer les signaux provenant de capteurs d'intérêts lorsque ceux-ci ne sont plus disponibles. L'utilisation de ce type de méthodologie soulève cependant plusieurs défis. L'une des principales problématiques est la quantification de l'incertitude. Par exemple, pour les déformations des aubes d'une turbine de groupe turbine-alternateur qui sont difficiles et coûteuses à mesurer, ces valeurs sont directement liées aux dommages de fatigue et leur incertitude au risque de fissuration des pales. Ces valeurs constituent donc une information importante pour réduire la probabilité de pannes imprévues. Cependant, ces mesures ne sont pas disponibles pour la surveillance à long terme; elles ne sont disponibles que lors de campagnes de mesure effectuées à la mise en service de l'équipement. Afin d'utiliser cette information pour la surveillance et la planification de la maintenance, nous proposons la création de capteurs virtuels avec une architecture de type *long short-term memory* (LSTM) combiné à une méthodologie permettant d'évaluer les incertitudes adaptées à la maintenance et l'évaluation des risques industriels. Pour ce faire, nous proposons une méthode d'ensemble de modèles combinés avec une approche d'estimation conforme pour obtenir une incertitude calibrée sur des données d'entraînement. De plus, un auto-encodeur est utilisé pour détecter lorsque les mesures d'intrants sont significativement en dehors des données d'entraînement. La méthodologie est appliquée à la fissuration des aubes de la turbine d'un groupe turbine-alternateur afin de mettre en évidence les difficultés liées à l'utilisation de l'apprentissage automatique pour la maintenance et l'évaluation des risques dans un contexte industriel.

**Mots-clefs** — *Quantification de l'incertitude, capteur virtuel, maintenance industrielle, apprentissage automatique*

**Abstract** — Machine learning can be used to train a model from measurement campaign or surveillance data to estimate sensor values of interest when these are not available anymore. However, one of the main difficulties related to the practical use of such technology is uncertainty quantification. For repairable systems like hydroelectric turbines, the long-term concern is that such system inevitably undergoes maintenance and repair leading to behavior changes with time. As an example, turbine blades strains, which are difficult and costly to measure are directly related to fatigue damage and the risk of blade cracking. Hence, they are an important information to plan maintenance and minimize the probability of unexpected outages. Such values are not available for long-term monitoring following the initial measurement campaign made after commissioning. To tackle this, we propose the use of virtual sensors. A virtual sensor is a software-based model of a physical sensor which provides sensor data without the need for the physical one. In this case, we used a long short-term memory (LSTM) architecture and focused on the development of a methodology to assess uncertainty suitable for the study case. To estimate the uncertainty, we combined ensemble methods, conformal prediction, and an auto-encoder. The methodology is used on an industrial study case to highlight the difficulties and challenges related to the practical use of such methodology where it is necessary to quantify uncertainties to plan maintenance and assess risk of failure.

**Keywords** — *Uncertainty quantification, virtual sensor, industrial maintenance, machine learning*

## I. INTRODUCTION

Le risque associé à un dommage potentiel est généralement caractérisé par deux paramètres : sa gravité proportionnelle à l'ampleur des dommages envisagés et sa probabilité d'occurrence. Dans un contexte de maintenance industrielle, si on exclut les événements rares ou extrêmes, les conséquences peuvent souvent être relativement faciles à estimer à partir d'événements similaires étant déjà survenus par le passé. En revanche, il est beaucoup plus compliqué de déterminer la probabilité d'occurrence dans le cas d'événements provenant d'une dégradation non observée qui dépend de données d'opération. Prenons comme exemple la fissuration des aubes de la turbine d'un groupe turbine-alternateur (GTA) tel qu'illustré à la Fig. 1, la fiabilité en fatigue dans un contexte de tolérance au dommage dépend de la dimension des défauts, des fluctuations de contraintes de faibles amplitudes à haute fréquence qui doivent rester sous un seuil et des fluctuations de contraintes de plus grandes amplitudes qui font propager les défauts avec une fréquence beaucoup plus faible (Byrne et al., 2003 ; Gagnon et al., 2013; Gagnon et al., 2014). Un exemple de chargement mesuré pour un cycle démarrage (*Startup*), marche à vide (SNL), montée en charge jusqu'à la puissance maximale (*Maximum Opening*) et l'arrêt (*Shutdown*) est présenté à la Fig. 2a. Ce cycle de chargement est caractérisé par des déformations de grandes amplitudes générées par le régime transitoire du démarrage, le changement de puissance et l'arrêt combiné à des déformations de petites amplitudes à haute fréquence dans les régimes permanents d'opération ce qui forme des blocs de chargement tel que illustrés à la Fig. 2b. La probabilité d'apparition d'une fissure majeure demandant réparation est la probabilité que pour une dimension de défaut donnée les fluctuations de faibles amplitudes à haute fréquence dans les régimes permanents d'opération soient au-dessus du seuil représenté par le diagramme de Kitagawa présenté à la Fig. 3.

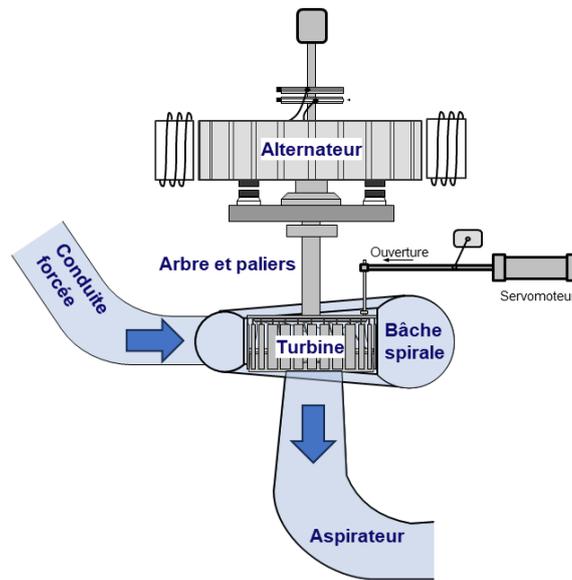


Fig. 1. Schéma d'un groupe turbine-alternateur (GTA)

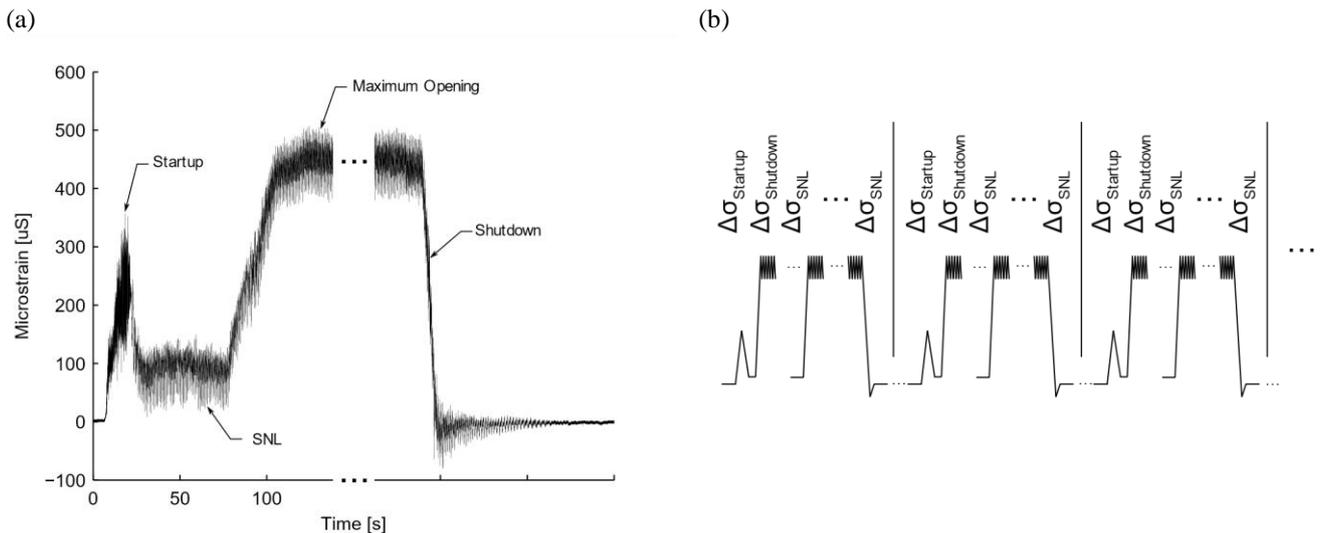


Fig. 2. (a) Exemple de chargement pour un cycle démarrage, montée en charge et arrêt. (b) Blocs de chargement que subit l'aube de turbine.

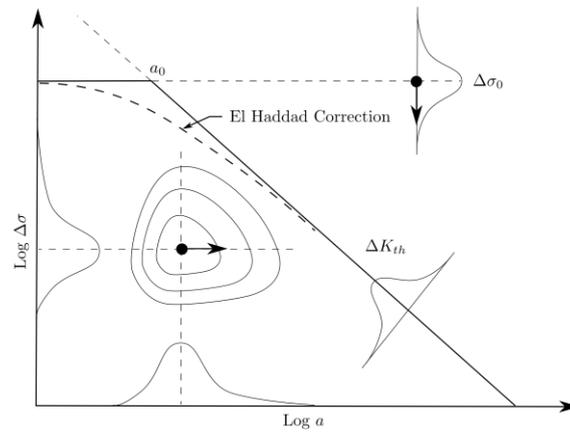


Fig. 3. Diagramme de Kitagawa.

On constate sur le diagramme à la Fig. 3 que l'incertitude sur l'amplitude des déformations  $\Delta\sigma$  est un facteur important dans l'évaluation de la probabilité de franchir le seuil où cette déformation contribuera à la propagation d'une fissure de longueur  $a$  définie par le seuil de propagation pour les fissures longues  $\Delta K_{th}$  et la limite d'endurance en fatigue  $\Delta\sigma_0$ . Donc, l'évaluation de cette incertitude est primordiale pour l'évaluation du risque de fissuration ainsi que la prise de décision de maintenance comme la planification des inspections, les réparations et les campagnes de mesure. Dans ce contexte, un risque de fissuration élevé avec une faible incertitude sur l'amplitude des contraintes n'a pas forcément la même signification qu'un même risque généré par une forte incertitude sur les contraintes. Dans le cas d'un risque élevé généré par une forte incertitude, le décideur aurait le choix de prendre des actions pour diminuer l'incertitude comme la collecte d'information additionnelle pour réduire le risque. Dans le contexte où nous proposons d'utiliser l'apprentissage automatique pour estimer les déformations des aubes en continu à l'aide d'un modèle entraîné sur les données de mise en service, la quantification de l'incertitude est donc primordiale. Cependant, elle soulève plusieurs difficultés. La méthodologie choisie dans cette étude pour quantifier l'incertitude utilise la méthode des ensembles (Lakshminarayanan et al., 2017) calibrés avec une approche conformelle (Jensen et al., 2024; Zeni et al., 2023; Lei et al., 2017) combinée avec un auto-encodeur pour évaluer à quel point nous sommes à l'extérieur de la distribution des données d'entraînement « out of distribution » (OOD) (Markou and Singh, 2003) afin d'évaluer s'il est possible de faire confiance aux valeurs estimées et leurs incertitudes. Le modèle obtenu agit comme capteur virtuel permettant de combiner les signaux de plusieurs capteurs pour estimer les signaux lorsque ceux-ci ne sont plus disponibles. L'approche sera évaluée sur les données d'une campagne de mesure sur un GTA lors de sa mise en service chez Hydro-Québec. Les résultats seront utilisés pour étudier la performance de cette méthodologie et en discuter les limitations.

Dans cet article, nous présentons premièrement le cas d'étude et les données utilisées. Ensuite, la méthodologie est décrite et les principaux résultats obtenus sont présentés. Puis, la performance et l'applicabilité de l'approche proposée dans un contexte où le risque de fissuration doit être évalué pour prendre des décisions de maintenance sont discutées afin de tirer des conclusions sur l'applicabilité de la méthodologie pour quantifier adéquatement l'incertitude et informer l'utilisateur sur ses limites dans le contexte industriel présenté.

## II. CAS D'ÉTUDES

Cette étude reprend le même cas utilisé précédemment par Gagnon et al. (2022). L'arrangement typique pour ce type de campagne de mesure est présenté à la Fig. 4 (Marcouiller and Thibault, 2015). Cet arrangement est composé de multiples systèmes d'acquisition (eDAQ). Au minimum, un système d'acquisition est dédié aux jauges de déformation sur les aubes de la turbine et un pour les capteurs sur la partie fixe collectant des données telles que l'ouverture des directrices, la puissance, la vitesse de rotation, les pressions, les vibrations, etc. Notons qu'il y a des difficultés liées à la synchronisation des signaux provenant des différents systèmes d'acquisitions utilisés. Dans ce cas-ci, les systèmes d'acquisition ont un signal commun permettant d'aligner les signaux. Toutefois, chaque système d'acquisition possède sa propre horloge pouvant générer des décalages dans le temps qui sont jugés négligeables dans cette étude.

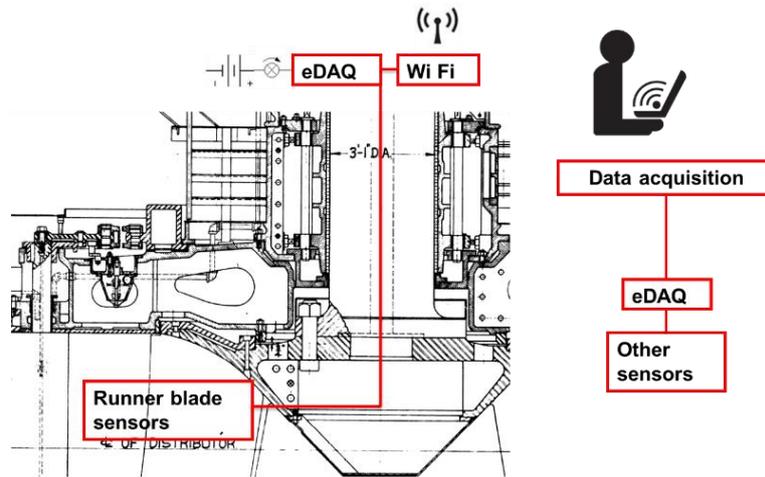


Fig. 4. Typical data acquisition setup.

Dans ce type de campagne de mesure, un large ensemble de valeurs sont mesurées simultanément pour vérifier et valider le bon fonctionnement du GTA. Un sous-ensemble de ces valeurs a été sélectionné afin de limiter le jeu de données à des valeurs disponibles dans le système de surveillance en continu du GTA. Les valeurs d'entrées sélectionnées sont les déplacements en x et y au palier alternateur, au palier turbine et au joint labyrinthe à la ceinture de la turbine, combinés avec l'ouverture des directrices contrôlant le débit entrant dans la turbine. La sortie est la déformation mesurée sur les aubes de la turbine. Afin de faciliter la présentation des résultats, notons que nous présenterons uniquement les résultats d'une des 24 jauges de déformations disponibles.

L'objectif est d'estimer les signaux de déformations en régime permanent pour évaluer le risque de fissuration dans ces régimes d'opération. Le jeu de données comprend plusieurs conditions d'opérations sous la forme de paliers de puissance. Un aperçu de la plage d'opération complète est présenté à la Fig. 5. Cette figure permet de mettre en évidence les changements de comportement dynamique du GTA sur l'ensemble de la plage d'opération pour les conditions de niveaux des réservoirs amont et aval au moment de la campagne d'essais. Les conditions d'opération présentées vont de la marche à vide (MAV) à la Puissance maximale (Pmax) et sont identifiées en % de Pmax. Les conditions d'opération sélectionnées pour l'étude: 67%, 77%, 92%, 94%, 99%, Pmax. Les données à 92% de Pmax (en orange sur la Fig. 4) sont utilisées pour l'entraînement et la performance de la méthodologie a été évaluée sur les 5 autres conditions d'opération (en bleu sur la Fig. 4). Ces conditions d'opération ont été sélectionnées, car elles représentent des changements graduels de comportement afin d'évaluer la sensibilité de la méthodologie.

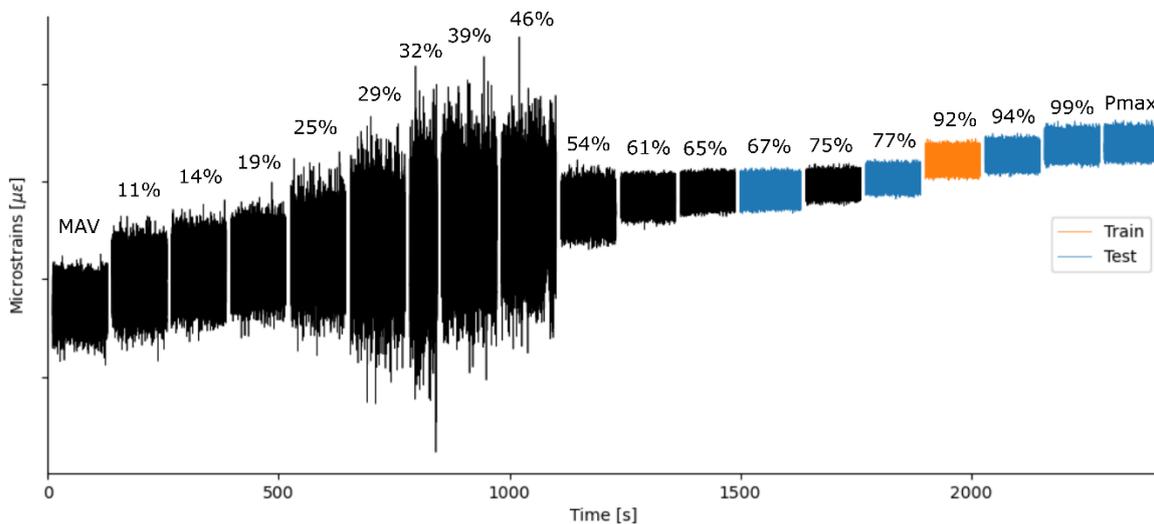


Fig 5. Données en fonction de la puissance pour une jauge de déformation. (Orange) Données d'entraînement. (Bleu) Données d'évaluation.

### III. MÉTHODOLOGIE

Les réseaux de neurones de type récurrent (RNN) sont particulièrement bien adaptés pour le traitement des séries temporelles telles que les signaux en provenance de capteurs. En effet, ils ont la capacité de conserver une mémoire de l'information passée dans les données d'apprentissage afin de prédire de nouvelles séquences de données. Parmi les diverses variantes des RNN

pouvant être utilisées, l'architecture la plus connue est le LSTM (Long Short-Term Memory), ou sa variante bidirectionnelle, Bi-LSTM, que nous avons utilisée et qui est représentée à la Fig. 6.

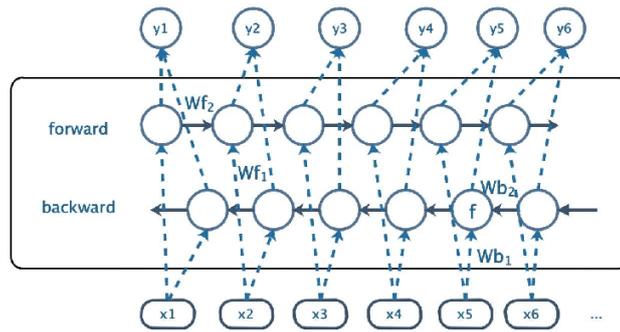


Fig. 6. Réseau Bi-LSTM.

Toutefois, bien que l'utilisation de réseau Bi-LSTM permette d'obtenir de bonnes estimations, nous avons besoin aussi de l'incertitude d'estimation. Pour ce faire, l'approche des ensembles profonds de Lakshminarayanan et al. (2017) a été utilisée. Cette approche consiste à utiliser plusieurs modèles indépendants (au moins 5 étant suggérés) dont les estimations différentes approximent un échantillonnage des prédictions possibles. Afin de générer plusieurs modèles indépendants à partir des mêmes données d'apprentissage, l'approche consiste à initialiser les paramètres de chacun des réseaux de neurones avec des valeurs aléatoires. Étant donné le grand nombre de paramètres, il est peu probable que l'apprentissage converge vers des valeurs identiques. De plus, pour rendre les modèles le plus indépendants possible, un ensemble d'apprentissages pour chaque modèle est généré en mélangeant l'ordre des échantillons d'intrants et extrants correspondants. Des prédictions de cet ensemble, on extrait la moyenne ou la médiane si l'on souhaite un estimateur robuste moins sensible aux valeurs extrêmes, et l'écart type qui est utilisé pour établir un intervalle de confiance.

L'intervalle de confiance est calibré à l'aide d'une approche conformelle. Nous avons utilisé une méthode de normalisation non-conforme avec ensemble (Jensen et al., 2024; Zeni et al., 2023; Lei et al., 2017). Cette approche utilise l'écart type des estimations de l'ensemble de modèles sur un jeu de données de calibration pour établir un intervalle de confiance proportionnel à l'incertitude de l'estimation du capteur virtuel pour un intervalle de confiance donné (*miscoverage*)  $\alpha$ , une valeur entre 0 et 1, comme suit :

1. Établir les estimations  $y_i$  et les déviations standards  $\sigma_i$  pour les  $n$  échantillons de calibration.
2. Pour un  $\alpha$  donné, établir la bande d'estimation  $d$  :

$$\Sigma_i, R_i = |y_i - y_i'| / \sigma_i, i \in \text{ensemble des données de calibration}$$

$$k = \lfloor (n+1) * (1 - \alpha) \rfloor$$

$$d = k^{\text{ième}} \text{ plus petit des } R_i$$

Ainsi, pour l'estimation  $y'$  d'un  $X$  donné, l'intervalle de confiance sera :

$$C_x = [y' - d \cdot \sigma_{y'}, y' + d \cdot \sigma_{y'}]$$

Cette approche conformelle garantit une incertitude bien calibrée à l'intérieur de la distribution des données d'entraînement, si le jeu de données de calibration est représentatif de cette distribution.

Toutefois, à l'extérieur de la distribution des données d'entraînement, il est difficile de savoir si l'on peut faire confiance à la valeur estimée et à l'intervalle de confiance bien que l'on puisse s'attendre à une augmentation de la variabilité de l'ensemble des modèles de prédiction. Pour détecter les écarts par rapport à la distribution des données d'entraînement, un auto-encodeur entraîné sur le jeu de données d'entraînement est utilisé. C'est une approche relativement classique pour la détection de valeurs considérées à l'extérieur de la distribution des données d'entraînement (Markou and Singh, 2003). On s'attend à ce que l'erreur de reconstruction soit faible à l'intérieur de cette distribution. Ensuite, pour évaluer la sensibilité du modèle à cet écart, les valeurs décodées peuvent être utilisées comme entrées du modèle pour évaluer les écarts générés sur la quantité d'intérêt tel que présenté à la Fig. 7. En effet, il se pourrait très bien que le capteur virtuel fournisse une estimation adéquate malgré la divergence de certains intrants en fonction de l'importance de ceux-ci dans l'estimation offerte par le capteur virtuel. Pour effectuer ce travail, une première approche que nous avons évaluée est d'établir l'écart entre  $y'(X)$  et  $y'(X')$ . L'idée étant que cet écart devrait globalement augmenter plus la distance à la distribution d'entraînement est significative.

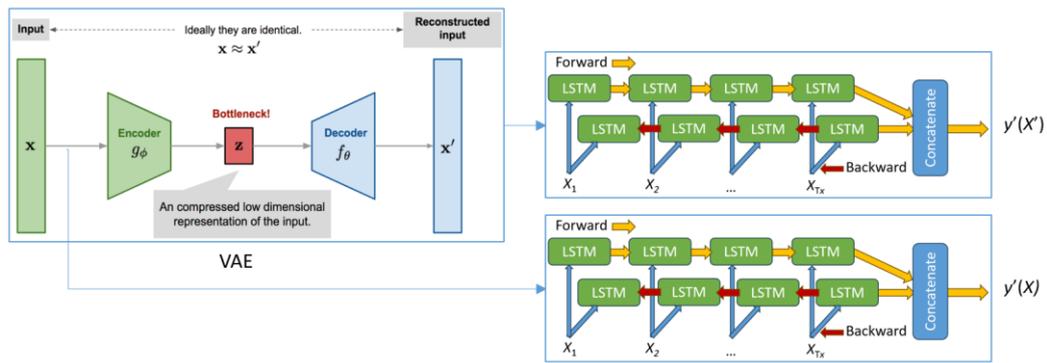


Fig. 7. Architecture des modèles pour établir l'écart d'estimation et la dérive fondée sur la variance d'estimation avec les signaux auto-encodés.

Une seconde approche, liée au fait qu'un ensemble de modèles entraînés de façon indépendante devrait diverger lorsque les intrants s'éloignent de ceux utilisés à l'entraînement (Lakshminarayanan et al., 2017) consiste à regarder la variabilité des écarts entre  $y'(X)$  et  $y'(X')$ . Cela devrait potentiellement aussi être une bonne métrique pour identifier à partir de quel moment on ne peut plus s'attendre à ce que les prédictions et l'intervalle de confiance ne soient plus calibrés, car les entrées seraient trop éloignées de la distribution d'entraînement.

Une troisième approche pour estimer la dérive, illustrée à la Fig. 8, consiste à entraîner un nouvel ensemble de modèles dont le but est de prédire l'erreur d'estimation  $|y-y'|$  en fonction de l'erreur de reconstruction  $|X-X'|$ . Cette approche utilise également un ensemble de modèles (Lakshminarayanan et al., 2017) pour évaluer la variabilité des  $|y-y'|$  estimés afin d'identifier le moment où les prédictions et l'intervalle de confiance ne sont potentiellement plus calibrés. Pour ce faire, nous avons utilisé la variance des valeurs des modèles visant à déterminer l'erreur d'estimation  $|y-y'|$  à partir de l'écart entre les signaux d'origine et auto-encodés  $|X-X'|$ .

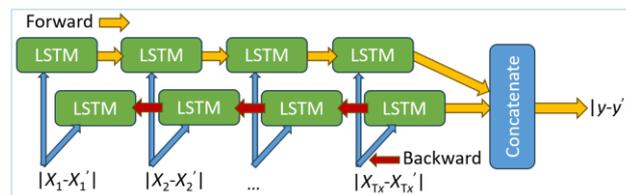


Fig. 8. Architecture d'un second modèle pour détecter la dérive.

#### IV. RESULTATS

Pour évaluer la performance de la méthodologie, 6 conditions d'opération ont été sélectionnées dans la plage d'opération normale du GTA où les changements de comportement ne sont pas trop prononcés. Les conditions d'opération sélectionnées en fonction de  $P_{max}$  sont : 67%, 77%, 92%, 94%, 99%,  $P_{max}$ . L'entraînement a été effectué uniquement sur les données à 92% pour évaluer s'il était possible de détecter des écarts graduels par rapport à la distribution d'entraînement. Les résultats sont présentés dans les Fig. 9 et 10. Ces figures présentent les courbes suivantes :

- $y$ , le signal à estimer, la courbe épaisse en noir
- $y'$ , l'estimation correspondant à la médiane des estimateurs ; la courbe épaisse en vert
- $|y-y'|$ , l'erreur entre le signal à estimer et son estimé ; la courbe en bleu
- $\sigma|y'|_{\alpha(0.1)}$ , l'écart d'estimation avec  $\alpha=0.1$  (90% IC) ; la zone ombragée en beige
- $y'_{AE}$ , l'estimation correspondant à la médiane des estimateurs alimentés avec les signaux auto-encodés  $X'$  ; la courbe épaisse en orange
- $\sigma|y'_{ae}-y'|$ , la variance entre les valeurs estimées des signaux d'origine et auto-encodés ; la courbe en vert
- $\sigma_{ma}|y-y'|$ , la variance des valeurs des modèles visant à déterminer l'erreur d'estimation  $|y-y'|$  à partir de l'écart entre les signaux d'origine et auto-encodés  $|X-X'|$  ; la courbe en rouge.

On observe que l'erreur d'estimation  $|y-y'|$  est faible à 92% de  $P_{max}$  puis augmente lorsqu'on s'écarte de cette condition d'opération utilisée pour l'entraînement. On remarque la même chose pour la calibration de l'incertitude qui se détériore rapidement lorsqu'on s'éloigne de la condition d'opération utilisée pour l'entraînement.

La métrique basée sur la variance entre les valeurs estimées des signaux d'origine et auto-encodés  $\sigma|y'_{ae}-y'|$  augmente aussi, mais elle est très bruitée avec des fluctuations difficiles à interpréter. D'un autre côté, la variance  $\sigma_{ma}|y-y'|$  des valeurs des modèles visant à déterminer l'erreur d'estimation  $|y-y'|$  à partir de l'écart entre les signaux d'origine et auto-encodés  $|X-X'|$  est

beaucoup moins bruitée et les valeurs sont visuellement corrélées à la détérioration des estimés et de la calibration de l'incertitude qui en font une métrique intéressante pour évaluer l'erreur générée par l'écart à la distribution des données d'entraînement.

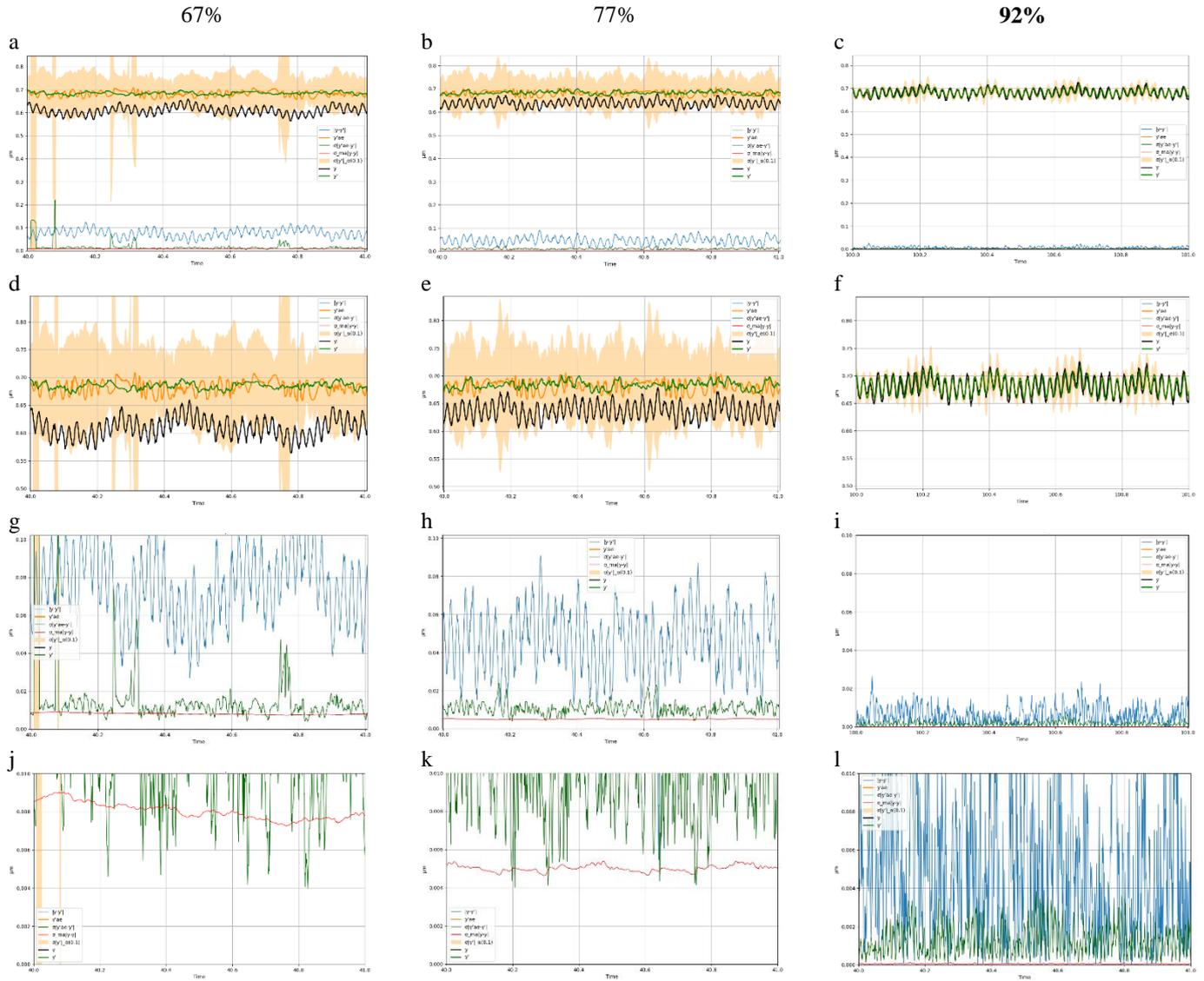


Fig. 9. Résultats pour les conditions d'opération à 67%, 77% et 92% de Pmax. (a, b, c) Vue d'ensemble des résultats. (d, e, f) Agrandissement des estimations. (g, h, i) Agrandissement des métriques d'écart à la distribution d'entraînement. (j, k) Agrandissement de la variance  $\sigma_{ma|y-y'}$  des valeurs des modèles visant à déterminer l'erreur d'estimation (courbe en rouge).

## V. DISCUSSION

On note que la variance d'estimations des modèles visant à déterminer l'erreur d'estimation  $|y-y'|$  à partir de l'écart entre les signaux d'origine et auto-encodés  $|X-X'|$ ,  $\sigma_{ma|y-y'}$ , est effectivement proportionnelle à l'écart de distribution des données d'application du capteur virtuel par rapport à la distribution d'entraînement. Il s'agit d'une métrique dont la valeur est très peu bruitée comparativement à la variance des estimations entre les signaux d'origine et auto-encodés,  $\sigma_{y|ae-y'}$  qui également été évaluée pour déterminer cet écart. Cette métrique comporte cependant un coût computationnel et une complexité non négligeable vu l'ensemble des modèles supplémentaires à générer. Cela soulève une problématique liée au coût de l'acquisition des données d'entraînement et leur représentativité de la distribution des données lors de l'utilisation. Le fait que l'on ne puisse ni faire confiance à la valeur estimée ni à son intervalle de confiance lorsqu'on s'éloigne de la distribution d'entraînement représente un frein au déploiement rapide de ce type d'approche, particulièrement lorsque ce type d'approche est utilisé dans la prise de décisions pouvant entraîner des conséquences importantes. Dans le cas présenté ici, qui est représentatif d'un équipement pour lequel il est impossible d'avoir des données pour l'ensemble des conditions d'exploitation et dont le comportement peut changer à la suite d'une dégradation ou d'une maintenance, cette métrique est une information essentielle malgré le coût computationnel et l'augmentation de la complexité rendant l'architecture encore plus opaque pour les utilisateurs. Notons que même si l'utilisation de chacun des modèles est relativement peu coûteuse après l'entraînement, le déploiement de la méthodologie devient beaucoup plus difficile en raison du nombre de modèles à utiliser.

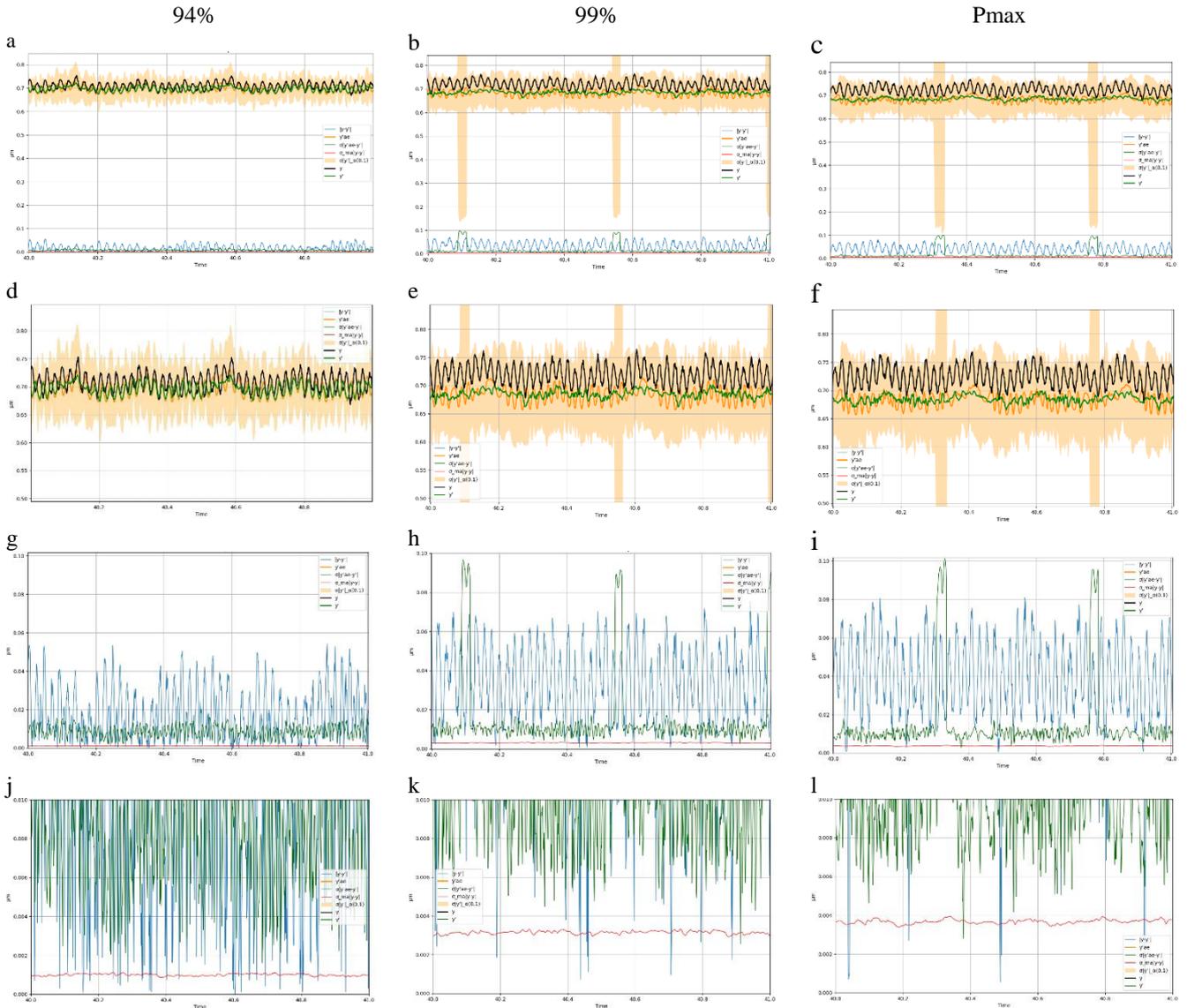


Fig. 10. Résultats pour les conditions d'opération à 94%, 99% de Pmax et Pmax. (a, b, c) Vue d'ensemble des résultats. (d, e, f) Agrandissement des estimations. (g, h, i) Agrandissement des métriques d'écart à la distribution d'entraînement. (j, k, l) Agrandissement de la variance  $\sigma_{maj-y'}$  des valeurs des modèles visant à déterminer l'erreur d'estimation (courbe en rouge).

## VI. CONCLUSION

La difficulté des méthodes d'apprentissage automatique à estimer l'incertitude est bien identifiée dans la littérature et des outils sont disponibles pour calibrer celle-ci à l'intérieur des données d'entraînement. Par contre, si les données d'entraînement ne sont pas représentatives du comportement de l'actif ou si celui-ci opère en dehors de la distribution des données d'entraînement, on peut difficilement avoir confiance aux valeurs estimées ou à la calibration de l'incertitude telle que l'on peut l'observer dans le cas d'étude présenté. Nous avons démontré à l'aide de ce cas d'étude industriel qu'il est possible de générer une métrique corrélée à l'erreur générée avec l'écart à la distribution d'entraînement en utilisant une méthodologie qui s'appuie sur l'ajout de modèles d'apprentissage automatisé pour auto-encoder les intrants du modèle principal. Toutefois, cette approche a une complexité et un coût computationnel non négligeable, car elle nécessite l'entraînement de plusieurs modèles d'apprentissage automatique supplémentaires. Cela augmente significativement l'opacité de la méthodologie dans le but de fournir une information sur les limites des estimations du modèle principal. Dans un contexte de prise de décisions entraînant des conséquences importantes, ce coût semble justifié et permettra d'éviter la prise de décisions basées sur des indicateurs peu crédibles. En revanche, ceci représente un frein important au déploiement d'outils basés sur l'apprentissage automatique pour l'évaluation du risque de bris et planifier la maintenance d'actif critique tel que les turbines de groupes turbines-alternateurs. Nous jugeons que ce constat est applicable à une large variété d'actifs industriels. Notons qu'en plus de l'augmentation du coût computationnel un autre frein à l'adoption de ce type d'approche est la complexité accrue due à la combinaison successive de modèles. Cette complexité limite l'explicabilité de l'approche, ce qui est particulièrement problématique dans les cas où le décideur doit justifier ces décisions.

## REFERENCES

- Byrne, J., Hall, R.F., Powell, B.E. (2003). Influence of LCF overloads on combined HCF/LCF crack growth. *Journal of Fatigue*, 25, 827-834
- Gagnon, M., Tahan, A., Bocher, P. and Thibault, D. (2013). A probabilistic model for the onset of High Cycle Fatigue (HCF) crack propagation: Application to hydroelectric turbine runner. *International Journal of Fatigue*, 47, 300–307.
- Gagnon, M., Tahan, A., Bocher, P. and Thibault, D. (2014). Influence of load spectrum assumptions on the expected reliability of hydroelectric turbines: A case study. *Structural Safety*, 50, 1–8.
- Lakshminarayanan, B., Pritzel, A. and Blundell, C. (2017). Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. *arXiv:1612.01474v3 [stat.ML]*.
- Jensen, V., Bianchi, F. M. and Anfinsen, S. N. (2024). Ensemble Conformalized Quantile Regression for Probabilistic Time Series Forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, 1–12.
- Fontana, M., Zeni, G. and Vantini, S. (2023). Conformal prediction: A unified review of theory and new challenges. *Bernoulli*, 29.
- Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J. and Wasserman, L. (2017). Distribution-Free Predictive Inference For Regression. *arXiv:1604.04173v2 [stat.ME]*.
- Markou, M. and Singh, S. (2003). Novelty detection: a review—part 2: neural network based approaches. *Signal Processing*, 83, 2499–2521.
- Gagnon, M., Vouligny, L., Cauchon, L. and Giroux, A.-M. (2022). Virtual sensors to generate turbine runner blade strains from indirect measurements. *IOP Conference Series: Earth and Environmental Science*, 1079, 12067.
- Marcouiller, L., & Thibault, D. (2015). Obtaining stress measurements on runners as a key contribution to reducing their degradation and improving the reliability of hydroelectric production units. In *Conference: HYDRO*