

Prédiction du comportement en dynamique rapide de structures assemblées intégrant des entrées fonctionnelles incertaines

T. Y. Abidi¹, F. Béchet^{1,2}, P. Moreau¹, A. Snoun¹, N. Leconte³, F. Massa^{1,2}, T. Delot^{1,2}

¹ Univ. Polytechnique Hauts-de-France, CNRS, UMR 8201 - LAMIH, F-59313 Valenciennes, France.

² INSA Hauts-de-France, F-59313 Valenciennes, France.

³ ONERA, DMAS, Lille, 59000, France

Résumé — Cette communication se focalise sur la prédiction du comportement en dynamique rapide d'une structure mécanique assemblée, pour lequel les réponses effort-déplacement des rivets sont considérées comme incertaines. La stratégie de propagation des incertitudes repose sur la définition d'un métamodèle de type Transformer permettant de lier les entrées fonctionnelles des liaisons, pour différents modes de sollicitation (traction, cisaillement, mixte), au comportement global de l'assemblage. Les résultats obtenus mettent en évidence une bonne capacité prédictive malgré un ensemble d'entraînement de taille réduite.

Mots clés — Modèles de substitution, Entrées fonctionnelles, Incertitudes, Apprentissage automatique, Liaisons rivetées, Simulation explicite.

1 Introduction

Les assemblages rivetés sont largement utilisés dans l'industrie [1] pour leur capacité à relier des composants structurels de manière efficace et fiable. Cependant, la prédiction du comportement en dynamique rapide de ces assemblages demeure complexe à cause des incertitudes associées aux liaisons. Ces dernières ont un impact significatif sur la réponse globale de la structure. La propagation des incertitudes dans ce contexte de simulations explicites requiert nécessairement l'exploitation de modèles de substitution [2], compte tenu des temps de calcul élevés pour chaque simulation. L'évolution de l'apprentissage automatique, et notamment les architectures de type Transformer [3], initialement développées pour le traitement du langage naturel, offrent une capacité remarquable à capturer les dépendances complexes de données séquentielles ou fonctionnelles [4].

Dans ce contexte, cette communication présente une stratégie basée sur un métamodèle Transformer enrichi d'un mécanisme de cross-attention résiduelle [5], capable de traiter les entrées fonctionnelles des liaisons et de prédire le comportement global de l'assemblage avec un ensemble d'entraînement de taille réduite.

2 Stratégie numérique

2.1 Propagation des incertitudes

La stratégie de propagation des incertitudes proposée repose sur trois étapes successives schématisées en Figure 1. Cette approche s'inscrit dans le cadre des méthodes de propagation d'incertitudes par modèles de substitution, largement utilisées en mécanique des structures [2]. Les entrées du métamodèle sont constituées de trois courbes effort-déplacement caractérisant le comportement individuel d'une liaison générique selon différents modes de sollicitation : traction pure, cisaillement pur, et mode mixte. En fait, ces courbes sont construites sur la base de la modélisation d'essais Arcan. Des liaisons de propriétés identiques sont ensuite insérées au sein d'un composant multi-matériau assemblé par neuf rangées de deux rivets. Les sorties du métamodèle correspondent à quatre fonctionnelles traduisant respectivement la réponse globale de l'assemblage ainsi que les courbes effort-déplacement des trois premières rangées de rivets, constituant ainsi un problème de prédiction multi-réponses.

La première étape consiste à définir une base de données d'apprentissage en générant des variantes des courbes d'entrée par simulation explicite sur un assemblage comportant une unique liaison, puis en réalisant pour chaque configuration une simulation explicite du composant complet. Les différentes trajectoires des données fonctionnelles sont générées à partir d'un plan d'expériences de type Latin HyperCube sampling.

La deuxième étape vise à construire et optimiser le modèle de substitution de type Transformer capable d'établir la relation entre les entrées et les sorties fonctionnelles du système. L'utilisation de l'architecture Transformer pour la modélisation de systèmes physiques s'appuie sur les travaux récents démontrant leur efficacité pour capturer les dépendances complexes dans les données séquentielles et fonctionnelles. Cette architecture, initialement développée pour le traitement du langage naturel, a montré des performances prometteuses pour l'émulation de systèmes physiques et la quantification d'incertitudes. Des approches alternatives basées sur les réseaux récurrents [6] ou convolutionnels [7] ont également été explorées dans la littérature.

Enfin, la troisième étape exploite le métamodèle entraîné pour propager rapidement les incertitudes et obtenir les enveloppes des réponses sans recourir à de nouvelles simulations. Cette stratégie permet d'explorer l'espace de variation à un coût calculatoire réduit tout en préservant la précision des prédictions.

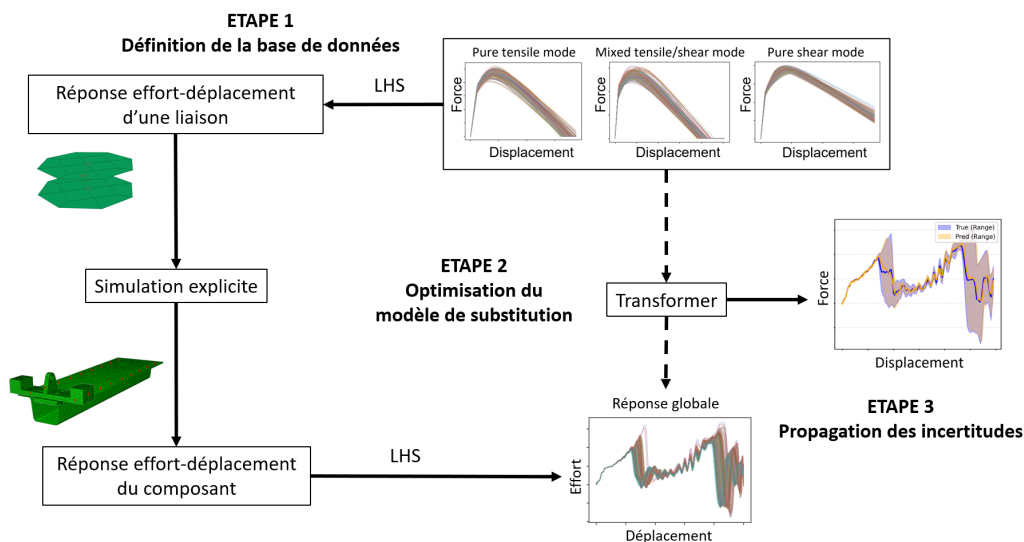


FIGURE 1 – Stratégie de propagation des incertitudes.

2.2 Modèles de substitution

Pour lier les entrées fonctionnelles aux sorties fonctionnelles, deux familles de modèles d'apprentissage profond sont investiguées :

- Les modèles récurrents LSTM (Long Short-Term Memory) de type standard [6], bidirectionnel [8] et hybride CNN-LSTM [9],
- Les modèles Transformers de type encoder seul [10] et encoder avec cross-attention résiduelle directe [5].

2.2.1 Modèles LSTM

La famille LSTM offre une approche séquentielle où des cellules de mémoire spécialisées apprennent les dépendances temporelles et les patterns d'évolution présents dans les séquences de données. Contrairement aux réseaux de neurones MLP classiques qui traitent chaque observation indépendamment, les LSTM possèdent une architecture récurrente capable de maintenir un état interne persistant, permettant ainsi de capturer des dépendances à long terme dans les données séquentielles. Trois variantes architecturales sont évaluées et illustrées à la Figure 2, chacune exploitant différemment la structure temporelle

des séquences :

- **LSTM Standard** : Cette architecture traite les séquences de manière unidirectionnelle, de l'instant initial t_0 jusqu'à l'instant final T . À chaque pas temporel, le réseau reçoit un vecteur d'entrée et met à jour deux états internes : l'état caché, qui sert de sortie à ce pas, et l'état de cellule, qui joue le rôle de mémoire à long terme. Le mécanisme central repose sur trois portes différentiables qui contrôlent le flux d'information : la porte d'oubli décide quelles informations de l'état de cellule précédent doivent être conservées ou oubliées, la porte d'entrée régule l'intégration de nouvelles informations dans l'état de cellule, et la porte de sortie contrôle quelles informations de l'état de cellule sont transmises à l'état caché. Cette propagation séquentielle permet au modèle de construire progressivement une représentation interne qui capture l'historique complet de la séquence observée. Le mécanisme de mémoire interne préserve efficacement les gradients lors de l'apprentissage, résolvant le problème classique de disparition du gradient rencontré dans les RNN standards. Cette architecture est particulièrement adaptée pour modéliser des processus où l'état futur dépend fortement de l'historique, comme les séries temporelles présentant des tendances progressives, des accumulations d'effets ou des phénomènes de dérive.
- **LSTM Bidirectionnel (BiLSTM)** : L'architecture bidirectionnelle étend le concept du LSTM standard en traitant simultanément la séquence dans les deux directions temporelles. Elle repose sur deux couches LSTM indépendantes : une couche *forward* qui progresse de l'instant initial vers l'instant final et une couche *backward* qui parcourt la séquence en sens inverse. À chaque pas temporel, les représentations issues des deux directions sont combinées pour former une représentation enrichie intégrant à la fois le contexte passé et futur.
- **Hybrid CNN–LSTM** : Cette architecture combine les capacités d'extraction de motifs locaux des réseaux convolutionnels (CNN) avec la modélisation des dépendances temporelles des LSTM. Le traitement s'effectue en deux étapes complémentaires :
 - **Module convolutionnel** : Des filtres 1D glissent le long de la séquence d'entrée pour détecter automatiquement des motifs locaux tels que pics, transitions abruptes, variations de pente, oscillations périodiques ou discontinuités. Plusieurs filtres parallèles permettent d'apprendre différents motifs simultanément, et des couches de pooling réduisent la dimensionnalité temporelle tout en conservant les caractéristiques essentielles, créant ainsi une représentation hiérarchique multi-échelle des motifs locaux.
 - **Module LSTM** : Les représentations extraites par le CNN sont transmises séquentiellement à une ou plusieurs couches LSTM qui capturent les dépendances à plus longue portée et apprennent les corrélations entre motifs locaux. Cela permet au modèle de modéliser la dynamique globale de la séquence.

Le couplage CNN–LSTM est particulièrement adapté aux séquences présentant une structure hiérarchique, combinant des micro-patterns locaux à haute fréquence et des macro-tendances à basse fréquence. De plus, la réduction de dimensionnalité opérée par le CNN et le pooling diminue la longueur de séquence à traiter par le LSTM, ce qui réduit le coût computationnel et améliore la stabilité de l'apprentissage.

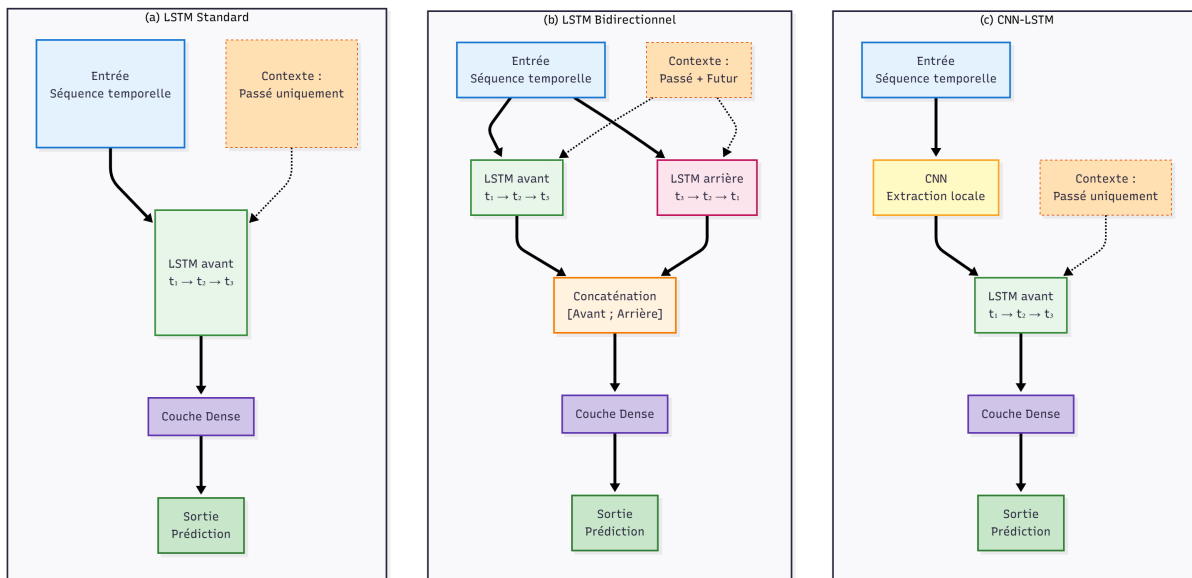
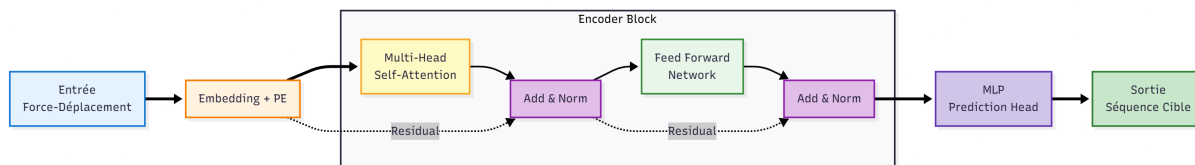


FIGURE 2 – Architectures des modèles LSTM

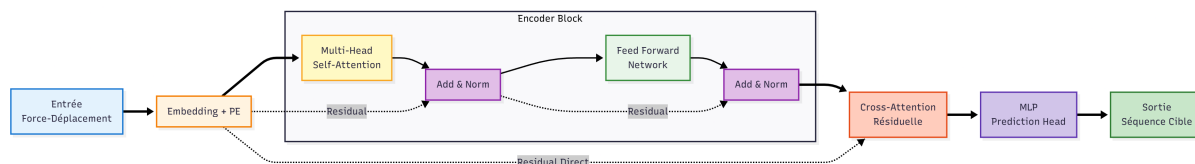
2.2.2 Modèles Transformers

Les modèles Transformers représentent une alternative moderne aux architectures récurrentes. Contrairement aux LSTM, qui traitent les données séquentiellement et maintiennent un état caché propagé dans le temps, les Transformers reposent exclusivement sur des mécanismes d'attention permettant de capturer simultanément les dépendances locales et globales dans la séquence, sans contrainte de proximité temporelle. Cette approche favorise la parallélisation du calcul, l'accès direct à n'importe quelle position de la séquence et la modélisation d'interactions complexes entre éléments distants. Deux variantes sont étudiées et présentées à la Figure 3.

- Transformer Encoder Only :** Cette architecture utilise uniquement la pile d'encodage du Transformer, sans décodeur ni mécanisme autoregressif. Chaque élément de la séquence est transformé en un embedding auquel on ajoute un encodage positionnel pour conserver l'information de position. La séquence traverse plusieurs couches d'encodeur identiques, composées de sous-couches de self-attention multi-têtes et de réseaux feedforward position-wise, chacune accompagnée de connexions résiduelles et de normalisation de couche. Après l'encodage, les représentations finales sont transmises à une tête de prédiction légère, générant toutes les sorties simultanément. Cette approche permet un parallélisme maximal, réduisant le temps d'inférence et facilitant l'apprentissage de dépendances à longue portée.
- Transformer Encoder avec Cross-Attention Résiduelle (Encoder + ResCross) :** Cette variante reprend le pipeline d'encodage standard précédent mais ajoute un module de cross-attention après la dernière couche d'encodeur permettant la réinjection des entrées fonctionnelles dans la représentation latente. La séquence d'entrée originale sert de query, tandis que les représentations encodées sont utilisées comme keys et values. La sortie de la cross-attention est combinée aux représentations encodées via une connexion résiduelle et normalisée. Cette architecture hybride préserve les informations de bas niveau, favorise la fidélité aux détails locaux et introduit une connexion à longue portée qui améliore la propagation du gradient et l'optimisation. Le coût computationnel supplémentaire reste modéré, tout en offrant une régularisation implicite et une amélioration de la précision pour les séquences où différentes échelles d'information sont importantes.



(a) Encoder-Only



(b) Encoder + ResCross (cross-attention résiduelle directe)

FIGURE 3 – Architectures des modèles Transformers

3 Prédiction du comportement du composant

3.1 Prédiction des réponses déterministes

La base de données utilisée pour l'apprentissage comprend 251 échantillons, dont la longueur des séquences en déplacement est fixée à 128 valeurs. La base de données est séparée en 166 échantillons pour l'entraînement (incluant la validation interne) et 85 pour le test. Avant l'apprentissage, une standardisation par Z-score est appliquée à l'ensemble données. Les performances des métamodèles sont évaluées au moyen du critère MSE (Mean Squared Error), du coefficient de détermination R^2 , ainsi que de la corrélation entre les enveloppes prédictives et réelles.

La Figure 4 présente les plus mauvaises prédictions de chaque modèle pour l'ensemble de test, permettant d'identifier les limitations respectives.

Le LSTM classique présente principalement un décalage temporel sur certaines portions du signal, en particulier sur les phases tardives, ainsi qu'une sous-estimation de l'amplitude sur les pics les plus marqués. Il a également tendance à produire une réponse trop lissée, omettant une partie des détails à haute fréquence. Ces limitations traduisent une sensibilité aux variations locales et une difficulté à capturer fidèlement les phénomènes rapides et tardifs du signal.

A contrario, l'architecture bidirectionnelle Bi-LSTM permet de mieux capturer les transitions grâce au contexte apporté par les séquences passées et futures. Les oscillations sont atténuées et les décalages temporels réduits par rapport au LSTM classique, même si certaines variations très locales restent moins bien reproduites.

L'architecture hybride CNN-LSTM améliore la stabilité générale et réduit le bruit sur les signaux prédits. Les erreurs sont plus homogènes sur l'ensemble des sorties, bien que les pics les plus prononcés soient légèrement sous-estimés. Cette architecture est la plus performante de la famille des LSTM.

Le Transformer encodeur-only ne parvient pas à reproduire correctement les courbes efforts-déplacement. Les prédictions sont globalement plus approximatives que celles produites par le LSTM standard, avec des décalages importants et une incapacité à suivre fidèlement les phénomènes transitoires. Ces limites suggèrent qu'un enrichissement du mécanisme de traitement des séquences serait nécessaire pour améliorer la précision globale et se rapprocher des performances de référence.

En combinant encodeur et cross-attention, cette architecture profite à la fois de la vision globale du signal et de l'information locale brute. Les prédictions sont plus précises, notamment sur les transitions rapides et les phases critiques, tout en conservant une bonne fidélité globale.

Globalement, tous les modèles parviennent à reproduire la forme générale des réponses du système mécanique, ce qui montre qu'ils ont appris les comportements physiques fondamentaux. L'architecture Transformers avec cross-attention, se distingue par une meilleure gestion des dépendances longues et des transitions rapides, offrant une fidélité globale plus élevée. Les principales erreurs concernent les détails fins et pourraient être améliorées surtout par un enrichissement du jeu d'entraînement plutôt que par des modifications architecturales.

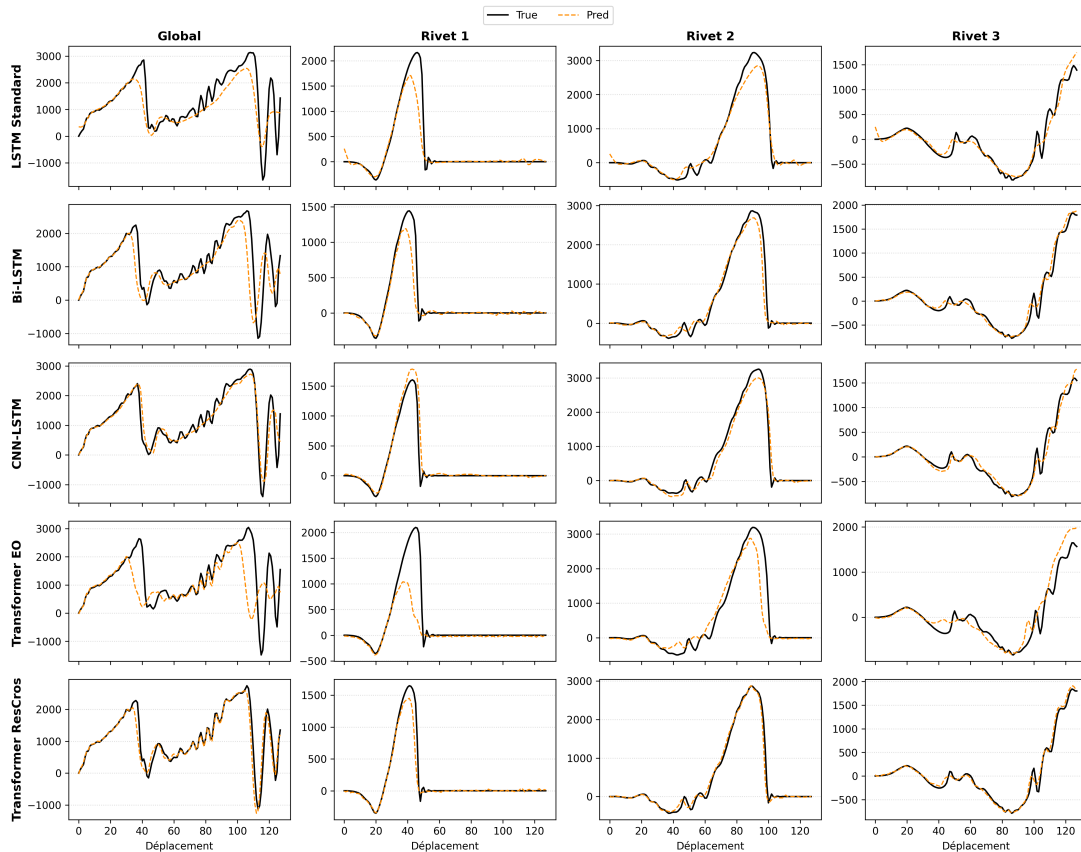


FIGURE 4 – Plus mauvaises prédictions avec différentes familles pour l'ensemble de test

La Figure 5 présente le coefficient de détermination (R^2) pour l'ensemble des modèles. On observe que les architectures Transformer, et plus particulièrement celle avec cross-attention, atteignent les valeurs les plus élevées, confirmant leur capacité à capturer à la fois la tendance globale et les variations locales des signaux. Parmi les modèles récurrents, le CNN-LSTM se distingue en affichant une capacité approximative proche de celle du Transformer cross-attention, tandis que les LSTM et Bi-LSTM classiques restent performants mais présentent une légère perte de précision sur les transitions rapides ou les phénomènes tardifs.

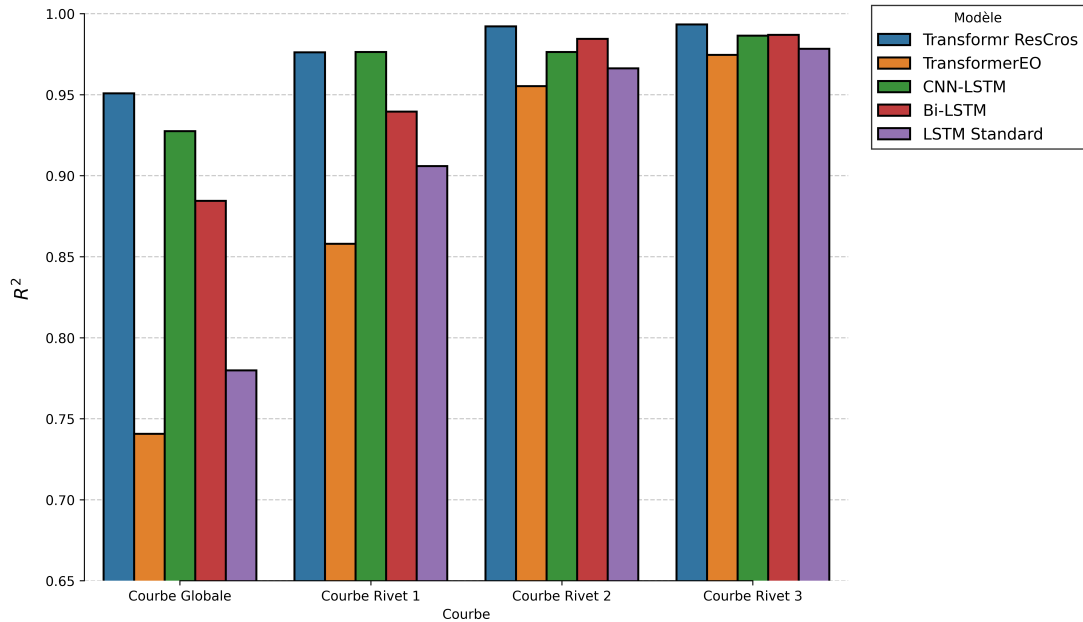


FIGURE 5 – Comparaison du coefficient de détermination (R^2) des différents modèles

3.2 Prédiction des réponses incertaines

La Figure 6 illustre l'enveloppe des réponses prédites, permettant de visualiser la variabilité des sorties face à la diversité des fonctionnelles d'entrées.

L'architecture Transformer ResCross montre une enveloppe plus cohérente et mieux alignée avec la référence du système, traduisant une meilleure capacité de généralisation. Les fluctuations résiduelles se concentrent principalement sur les phases de transition rapide, où les phénomènes locaux varient fortement d'un échantillon à l'autre.

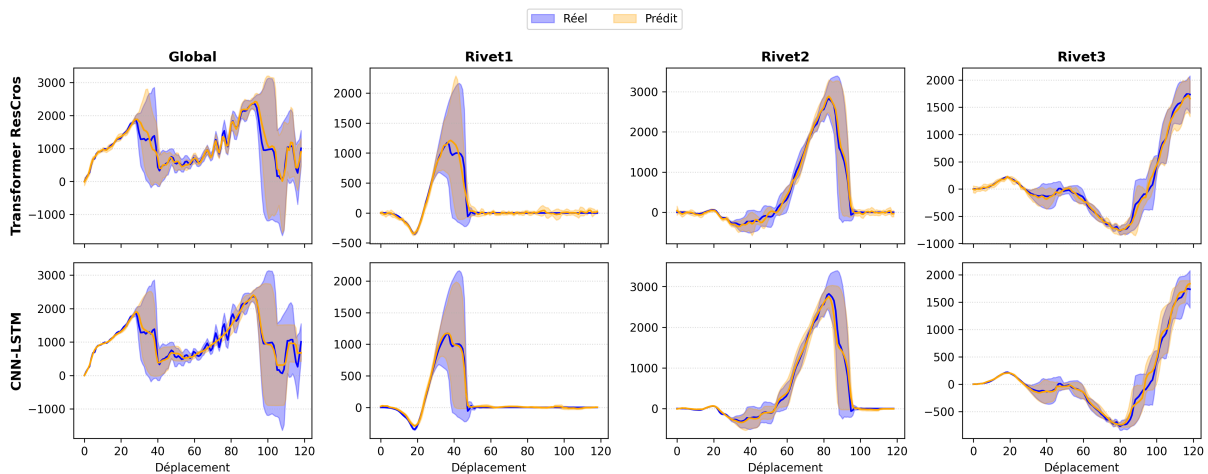


FIGURE 6 – Réponse incertaine pour le modèle CNN-LSTM et Transformer ResCross

L'estimation des réponses incertaines permet de représenter plus fidèlement la réalité et les variabilités observées durant les essais physiques, comme observées dans la référence [1].

L'étape suivante s'oriente tout naturellement vers une hybridation des données expérimentales et numériques pour améliorer la représentativité des modèles développés

4 Conclusion

Cette communication a présenté une stratégie de propagation des incertitudes appliquée au comportement en dynamique rapide d'un composant assemblé, en s'appuyant sur des modèles de substitution fondés sur l'apprentissage profond. Les approches évaluées — LSTM, variantes bidirectionnelles et hybrides, ainsi que Transformers — démontrent que des architectures séquentielles peuvent efficacement remplacer des simulations explicites coûteuses, tout en conservant la structure hystérétique des réponses mécaniques.

L'analyse comparative montre que si les modèles récurrents parviennent à capturer les grandes tendances du système, ils présentent davantage de fluctuations locales et une sensibilité accrue aux transitions rapides. Les architectures Transformer, en particulier la version enrichie par une cross-attention résiduelle directe, offrent les meilleures performances. Elles se distinguent par une capacité supérieure à intégrer simultanément les dépendances globales et les phénomènes locaux, ce qui se traduit par des valeurs élevées de R^2 , une enveloppe prédictive précise et une variabilité réduite par rapport à la référence.

Références

- [1] N. Leconte, B. Bourel, F. Lauro, C. Badulescu, E. Markiewicz. *Strength and failure of an aluminum/PA66 self-piercing riveted assembly at low and moderate loading rates : Experiments and modeling*, International Journal of Mechanical Sciences, 2016.
- [2] A. F. López-Lopera, F. Massa, I. Turpin & N. Leconte. *Modeling complex mechanical computer codes with functional input via Gaussian processes*, Proceedings of the 43rd Ibero-Latin American Congress on Computational Methods in Engineering (CILAMCE), Foz do Iguaçu, Brazil, Nov. 2022. HAL preprint hal-03882473.
- [3] A. Vaswani, N. Shazeer, N. Parmar, et al. *Attention Is All You Need*, Advances in Neural Information Processing Systems (NeurIPS), pp. 5998–6008, 2017.
- [4] N. Geneva, N. Zabarar. *Transformers for modeling physical systems*, Neural Networks, 146, pp. 272–289, 2022. doi :10.1016/j.neunet.2021.11.022.
- [5] C. F. Chen, Q. Fan, R. Panda. *CrossViT : Cross-Attention Multi-Scale Vision Transformer for Image Classification*, Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 357–366, 2021.
- [6] S. Hochreiter, J. Schmidhuber. *Long Short-Term Memory*, Neural Computation, MIT Press, 9(8) :1735–1780, 1997.
- [7] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner. *Gradient-based learning applied to document recognition*, Proceedings of the IEEE, 86(11), pp. 2278–2324, 1998.
- [8] M. Schuster, K. K. Paliwal. *Bidirectional Recurrent Neural Networks*, IEEE Transactions on Signal Processing, 45(11), pp. 2673–2681, 1997.
- [9] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, W.-C. Woo. *Convolutional LSTM Network : A Machine Learning Approach for Precipitation Nowcasting*, Advances in Neural Information Processing Systems (NeurIPS), pp. 802–810, 2015.
- [10] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova. *BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding*, arXiv preprint arXiv :1810.04805, 2018.