

# Détection et diagnostic d'anomalies sur des structures de génie civil, par le couplage de techniques de Machine Learning et de simulation numérique

M.S. Diallo<sup>1</sup>, G. Daniel<sup>2</sup>, G. Folzan<sup>1</sup>, L. Jason<sup>1</sup>

<sup>1</sup> Université Paris-Saclay, CEA, Service d'Études Mécaniques et Thermiques, Gif-sur-Yvette 91191, France

<sup>2</sup> Université Paris-Saclay, CEA, Service de Génie Logiciel pour la Simulation, Gif-sur-Yvette 91191, France

**Résumé** — Les structures en génie civil sont de plus en plus complexes, rendant indispensable leur suivi continu. Ce suivi est assuré par l'ajout de nombreux capteurs intégrés directement dans les structures, générant une grande quantité de données. Dans ce contexte, nous proposons une méthode d'analyse automatique de ces données afin d'y détecter de potentielles anomalies et identifier des comportements anormaux dans les structures de génie civil. Cette approche se fonde sur une méthode de Machine Learning non supervisée : l'Analyse en Composantes Principales (ACP).

Dans un premier temps, nous mettons en place une maquette numérique instrumentée par des capteurs virtuels pour fournir des données maîtrisées sur un système simple afin de mettre au point notre approche. Puis, des tests sont réalisés pour valider cette méthodologie, démontrant sa capacité à détecter des anomalies et à fournir des informations pertinentes pour la maintenance prédictive des structures.

**Mots clés** — Machine Learning, SHM, Cast3M, Simulation numérique, ACP, Capteurs, Génie civil.

## 1 Introduction

Les méthodes traditionnelles du suivi et du diagnostic de structures de génie civil sont basées sur des méthodes destructives, non destructives [1] ou des analyses manuelles, souvent contraintes par leur coût en temps de travail. Ces méthodes se limitent alors à des diagnostics ponctuels. Or, le suivi continu des structures représente un enjeu majeur afin de pouvoir détecter des défaillances au plus tôt.

Dans ce contexte, l'intégration croissante de capteurs [2, 3], mesurant par exemple les déformations ou les déplacements au sein des structures, s'inscrit dans le domaine du « Structural Health Monitoring » (SHM) [3, 4, 5], qui a connu une évolution significative ces dernières années. L'un des principaux défis réside dans la gestion et le traitement des importants volumes de données générées par ces capteurs et leur traitement automatique afin de garantir un diagnostic continu des structures, dans le but de renforcer leur sécurité et leur durabilité. Il est donc devenu nécessaire de disposer d'outils pour répondre à ce défi.

Le Machine Learning (ML) a apporté de nouvelles méthodes pour améliorer la précision, l'efficacité, et la robustesse des diagnostics, à travers des algorithmes capables d'analyser et de prédire les comportements structurels à partir des données recueillies. Dans ce travail, nous proposons de mettre en oeuvre une méthodologie de détection d'anomalies dans des relevés de données de capteurs, afin d'alerter sur de possibles comportements anormaux de structures. Cette méthodologie repose sur l'Analyse en Composantes Principales (ACP) que nous présentons en section 2. Nous l'appliquons à des données issues d'une maquette numérique, décrite en section 3, et nous discutons ses performances en section 4.

## 2 Détection d'anomalies par Analyse en Composantes Principales

### 2.1 Définition du problème

Nous considérons une structure instrumentée avec des capteurs mesurant des grandeurs caractérisant le comportement mécanique, par exemple un déplacement ou une déformation. Nous notons  $u_{ij}$ , la mesure à un instant  $i$  d'un capteur  $j$  et  $U_i$ , le vecteur rassemblant l'ensemble des mesures à un instant  $i$ . Lorsque la structure a un comportement normal, un lien est attendu entre les mesures des différents cap-

teurs. Les valeurs de tous les capteurs vont évoluer ensemble selon des modes liés aux caractéristiques de la structure. Différentes sollicitations de la structure activent différents modes de déformation. En particulier, si la structure est dans un régime de comportement élastique, les relations existantes entre les capteurs sont linéaires.

On considérera qu’une anomalie est présente à un instant  $i^*$  si les valeurs mesurées au sein du vecteur  $U_{i^*}$  ne respectent pas les relations attendues entre les capteurs. La méthode que nous proposons consiste à extraire ces relations à partir de données de mesure de la structure lorsque celle-ci est en comportement normal puis à proposer un critère de vérification qui permettra de qualifier si chaque nouvelle mesure présente une anomalie ou non.

Nous soulignons ici qu’aucune contrainte spécifique n’est imposée sur le positionnement des capteurs. Nous faisons deux hypothèses principales :

- des liens sont attendus entre les capteurs mais ils n’ont pas besoin d’être explicités a priori ;
- un nombre de capteurs suffisant est placé sur la structure. Ce nombre est dépendant de la structure, des comportements attendus et de la variabilité des sollicitations imposées à la structure.

## 2.2 Approche proposée par Analyse en Composantes Principales

### 2.2.1 Intérêt de l’ACP

L’Analyse en Composantes Principales (ACP) [6] est une technique de Machine Learning non supervisée dont l’objectif est la réduction de dimensionnalité des données. Elle est largement utilisée pour analyser et visualiser des données massives à haute dimension dans un espace réduit de plus petite dimension. La méthode consiste à identifier les directions, appelées composantes principales, qui maximisent la variance des données et à les projeter dans le sous-espace engendré par ces directions. Cette approche permet de réduire le nombre de variables permettant de décrire les données tout en minimisant la perte d’information, en exploitant les relations linéaires entre les différentes variables.

À travers cette réduction de dimension, l’ACP est aussi utilisée pour la détection d’anomalies sur des structures [7]. Par exemple, l’article [8] diagnostique des dommages structurels dans des conditions environnementales variables, dans un cas où le comportement attendu de la structure est linéaire. Les auteurs séparent les effets environnementaux des indicateurs de dommage dans les signaux vibratoires avec différentes méthodes d’analyse statistique, dont l’ACP. L’article [9] prolonge cette analyse en considérant l’influence spécifique des variations environnementales, telles que la température et l’humidité, sur les mesures de vibration pour des cas d’étude avec des non-linéarités. En appliquant plusieurs ACP locales à des données expérimentales incluant des comportements non linéaires, les auteurs sont en mesure de détecter l’apparition de dommages au sein de la structure.

L’intérêt de l’ACP pour notre problématique réside dans le fait que c’est une méthode non supervisée : elle ne nécessite pas l’utilisation de données labellisées mais des données supposées représentatives du comportement normal de la structure. Ces données peuvent par exemple correspondre à des acquisitions lors des premiers temps de mise en service de la structure, lors desquels elle est considérée comme intègre.

Elles constituent les données d’entraînement que nous noterons dans la suite  $\mathbf{U} \in \mathbb{R}^{n_s \times n_f}$ .  $U_{ij}$  représente une observation donnée  $i$  associée à un capteur  $j$ . Si les instrumentations fournissent plusieurs valeurs (par exemple, les trois composantes de déplacement), chaque valeur sera associée à un capteur  $j$  différent. Nous faisons ici l’hypothèse que les mesures des capteurs sont synchronisées, leur nombre est donc identique pour tous les capteurs et que  $n_s \geq n_f$ , de sorte à avoir suffisamment d’observations pour caractériser les liens entre les différents capteurs sans sur-apprentissage. Ainsi, dans la matrice  $\mathbf{U}$ , une ligne  $i$  représente une vue de l’état global du système pour une observation donnée. Une colonne  $j$  regroupe l’ensemble des observations d’un capteur  $j$  donné. Dans le cas de séries temporelles, cette colonne  $j$  est donc la succession des valeurs mesurées par le capteur  $j$  pour les instants considérés. Notons que la méthode présentée dans ce travail ne tient pas compte d’une potentielle corrélation temporelle entre les valeurs successives d’un capteur. L’ordre des colonnes peut être changé sans impact sur le résultat. Ceci est cohérent avec les hypothèses quasi-statiques de nos simulations numériques.

## 2.2.2 Entraînement de l'ACP

Formellement, l'ACP consiste à appliquer une décomposition en valeurs singulières de la matrice  $\mathbf{U}$  préalablement centrée, ce qui correspond à l'identification des vecteurs propres et valeurs propres de la matrice de variance-covariance  $\mathbf{U}^\top \mathbf{U}$ . Cette décomposition, calculée lors de l'entraînement de l'ACP, s'exprime

$$\mathbf{U} = \hat{\mathbf{U}} \mathbf{S} \mathbf{W}^\top \quad (1)$$

avec :

- $\mathbf{W}$  : matrice des vecteurs propres de  $\mathbf{U}^\top \mathbf{U}$ . Ces vecteurs propres sont orthonormés et caractérisent le sous-espace propre dans lequel sont situées les données ;
- $\hat{\mathbf{U}}$  : matrice des composantes principales dans le sous-espace propre ;
- $\mathbf{S}$  : matrice diagonale contenant les valeurs singulières correspondant à la racine carrée des valeurs propres de  $\mathbf{U}^\top \mathbf{U}$ , ordonnées par valeur décroissante. L'importance d'une composante  $k$  est caractérisée par sa valeur singulière associée  $S_{kk}$  : plus cette valeur est grande, plus la composante porte d'information.

La réduction de dimension consiste à ne retenir qu'une partie des composantes principales  $\hat{\mathbf{U}}$ . Cela revient à tronquer les matrices  $\mathbf{W}$  et  $\mathbf{S}$  en ne conservant que les  $m'$  composantes principales, correspondant aux plus grandes valeurs singulières

$$\hat{\mathbf{U}} = \mathbf{U} \mathbf{W}_{n_f \times m'} \mathbf{S}_{m' \times m'}^{-1} \quad (2)$$

La valeur de  $m'$  est généralement choisie de sorte que  $S_{m'm'} \gg S_{(m'+1)(m'+1)}$ , c'est-à-dire que toutes les autres composantes principales au-delà l'indice  $m'$  ne porte plus d'information significative.

## 2.2.3 Erreur de reconstruction par ACP et détection d'anomalies

Une fois les matrices  $\mathbf{S}$  et  $\mathbf{W}$  déterminées, nous pouvons appliquer les résultats de l'ACP à un vecteur quelconque  $V$  pour le projeter dans le sous-espace vectoriel et obtenir ses composantes principales  $\hat{V}$  dans ce sous-espace puis les reprojeter dans l'espace d'origine pour obtenir un vecteur reconstruit  $V_r$ . Cette opération est décrite par l'équation

$$V_r = \hat{V} \mathbf{S}_{m' \times m'} \mathbf{W}_{n_f \times m'}^\top = V \mathbf{W}_{n_f \times m'} \mathbf{W}_{n_f \times m'}^\top \quad (3)$$

On définit alors l'erreur de reconstruction  $\varepsilon(V)$  comme étant la différence entre le vecteur reconstruit  $V_r$  et le vecteur d'origine  $V$ ,

$$\varepsilon(V) = V - V_r = V \left( \mathbf{I} - \mathbf{W}_{n_f \times m'} \mathbf{W}_{n_f \times m'}^\top \right) \quad (4)$$

Si l'erreur de reconstruction  $\varepsilon$  est faible, cela signifie que l'ACP est capable de reconstruire le vecteur  $V$ . Il est donc situé dans le même sous-espace vectoriel que les données d'entraînement. *A contrario*, si cette erreur est élevée, dépassant un certain seuil à définir, cela signifie que l'ACP n'est pas capable de reconstruire le vecteur. Les liens de linéarité entre les composantes du vecteur  $V$  ne sont pas les mêmes que dans les données d'entraînement. Le changement de relation entre les données des capteurs signale un changement de régime de comportement de la structure et, par conséquent, une anomalie.

Dans ce travail, nous proposons d'analyser l'erreur de reconstruction sur la base d'apprentissage afin de décider du nombre de composantes à conserver et de choisir le seuil de détection d'anomalie. Le nombre de composantes est choisi pour présenter des erreurs suffisamment faibles sur la base d'apprentissage. Puis, nous étudions la distribution des erreurs de reconstruction sur les données d'entraînement. Nous utilisons cette distribution pour définir un critère statistique de détection d'anomalie lors de l'application sur des données à tester. Nous détaillons cette méthode sur le cas présenté en section 3.3.

# 3 Maquette numérique et entraînement de l'ACP

## 3.1 Présentation de la maquette

La maquette choisie pour générer nos données est une poutre en béton armé de longueur  $L_0 = 5$  m. Ses armatures sont disposées comme présentées en figure 1.a . Cette structure est modélisée sur Cast3M

2024 [10] avec un maillage hexaédrique à 8 noeuds d’une taille fixée à 2,5 cm (fig. 1.b). Les simulations sont réalisées avec un modèle de comportement de Mazars régularisé en traction.

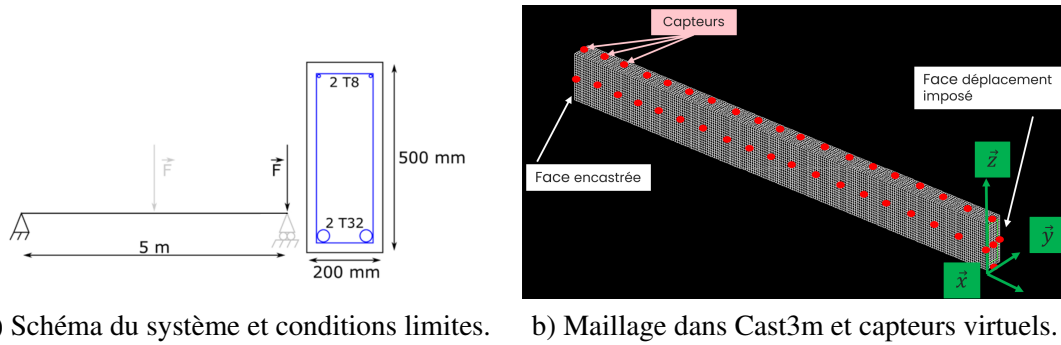


FIGURE 1 – Aperçu de la maquette numérique. En fonction de la sollicitation appliquée, le point d’application de la force, sa direction ou les blocages peuvent être modifiés.

### 3.2 Génération des données d’entraînement

Afin de générer des données d’entraînement, la maquette est soumise à plusieurs types de sollicitation :

- traction d’amplitude 0,015 mm ;
- compression d’amplitude 0,015 mm ;
- flexion suivant l’axe  $\vec{y}$  d’amplitude 0,7 mm ;
- flexion suivant l’axe  $\vec{z}$  d’amplitude 0,5 mm ;
- flexion 3 points suivant l’axe  $\vec{z}$ , appliquée grâce à 2 plaques d’appuis aux extrémités de la poutre et une plaque d’application de la force au milieu de la poutre, d’amplitude 0,7 mm.

L’amplitude de ces sollicitations est choisie de sorte que la structure reste dans le domaine élastique. Chaque sollicitation est appliquée progressivement en 601 pas de chargement. 86 capteurs virtuels sont considérés : 21 capteurs espacés régulièrement sur l’axe médian de chacune des faces longitudinales et un capteur unique au centre des faces avant et arrière suivant l’axe  $x$  de la poutre, comme illustré sur la figure 1.b. La valeur du déplacement est relevée et sauvegardée pour chaque capteur virtuel à chaque pas de chargement. Chaque pas est utilisé ensuite comme autant de réalisation indépendante pour alimenter l’ACP sans notion de temporalité.

Nous appliquons ensuite aux résultats de Cast3M, un bruit aléatoire tiré pour chaque réalisation afin de simuler un bruit de mesure. Le bruit ajouté suit une loi normale centrée d’écart-type  $7 \times 10^{-8}m$ . La matrice résultante  $U$ , de taille  $3005 \times 258$ , constitue les données d’entraînement de l’ACP.

### 3.3 Application de l’ACP sur les données générées et mise en place du critère de détection

Nous utilisons le module *ACP* de la bibliothèque *scikit-learn* [11] pour réaliser l’entraînement à partir des données de notre maquette. Le module calcule directement les matrices  $W$  et  $S$  de la décomposition en valeur singulière de l’équation (3).

Comme présenté en section 2.2.3, afin de déterminer le nombre de composantes, nous analysons l’histogramme des erreurs de reconstruction et conservons les composantes qui permettent d’aboutir à des erreurs suffisamment faibles. Dans notre cas, l’erreur de reconstruction est considérée comme faible si elle est de l’ordre du bruit de mesure introduit en section 3.2.

La figure 2 montre l’erreur de reconstruction produite en se restreignant à 4 composantes. Nous constatons que la répartition des erreurs de reconstruction suit une loi normale centrée, d’écart-type  $\sigma_r \approx 6.94 \times 10^{-8}$ . Un test de Kolmogorov-Smirnov [12] permet en effet de montrer que l’hypothèse de loi normale ne peut pas être rejetée, avec une  $p$ -valeur d’environ 0,93. Le reste des composantes rassemble donc la part due au bruit de mesure simulé dans les données. Le nombre de 4 composantes dominantes s’explique par le choix des sollicitations : la traction et la compression activent le même mode de déformation, une seule composante permet donc de les expliquer. Les trois autres sollicitations

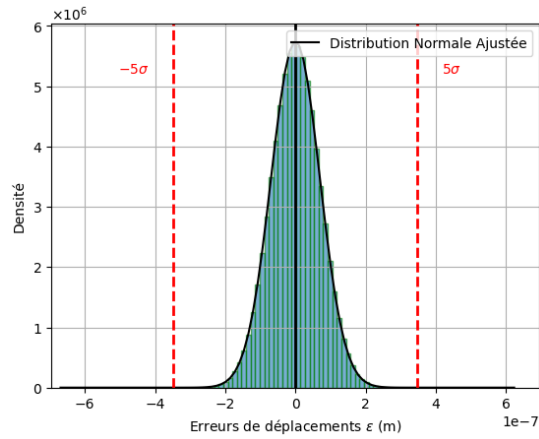


FIGURE 2 – Histogramme de l’erreur de reconstruction.

activent chacune de nouveaux modes ce qui se traduit par l’ajout de composantes significatives dans l’ACP. Par la suite, nous tronquerons alors notre ACP aux 4 premières composantes.

À partir de la distribution des erreurs de reconstruction sur les données d’entraînement, nous choisissons de définir un seuil de détection d’anomalie à  $5\sigma_r$ , qui correspond, sur une loi normale, à une probabilité de  $1 - 6.10^{-7}$  d’obtenir cette valeur sous l’hypothèse qu’elle est produite en comportement normal de la structure.

## 4 Détection d’anomalies sur la maquette numérique

Afin d’illustrer l’utilisation de cette méthodologie, nous mettons en place des simulations de test, certaines conduisant à l’endommagement de la structure, pour vérifier si la détection d’anomalie permet de mettre en évidence la présence ou non d’endommagement. La section 4.1 décrit les scénarios de test puis nous appliquons la méthode de détection d’anomalie en section 4.2. Enfin, nous soulignons l’impact du bruit de la base d’apprentissage sur les performances de détection en section 4.3.

### 4.1 Génération de données de test

Nous avons conduit les simulations de test suivantes :

1. flexion trois points d’amplitude 0,7 mm ;
2. flexion trois points d’amplitude 10 mm ;
3. traction d’amplitude 5 mm ;
4. flexion selon l’axe  $\vec{z}$  d’amplitude 20 mm.

Dans le premier scénario de flexion trois points, la structure reste dans le domaine élastique : il sert à vérifier que la méthode développée ne détecte pas de faux positifs. Pour les autres scénarios, l’amplitude de sollicitation est choisie pour conduire la structure jusqu’à son endommagement, des anomalies doivent alors être détectées.

À l’instar des données d’entraînement, les données de test sont le relevé des déplacements des capteurs virtuels auquel nous avons ajouté un bruit aléatoire de même amplitude que pour la base d’entraînement.

### 4.2 Détection d’anomalie sur les scénarios de test

Pour chaque scénario et pour chaque pas de chargement, nous appliquons l’ACP avec 4 composantes obtenue en section 3.3. Nous calculons ensuite l’erreur de reconstruction  $\epsilon$  sur ces données en utilisant l’équation (4). Enfin, nous vérifions si cette erreur dépasse le seuil de  $5\sigma_r$ , défini en section 3.3.

La figure 3 montre les résultats obtenus pour les différents scénarios. Chaque ligne correspond à un scénario différent. La première colonne représente la déformée de la poutre. La deuxième colonne est l’endommagement maximal renvoyé par Cast3M en fonction du pas de chargement : cette information

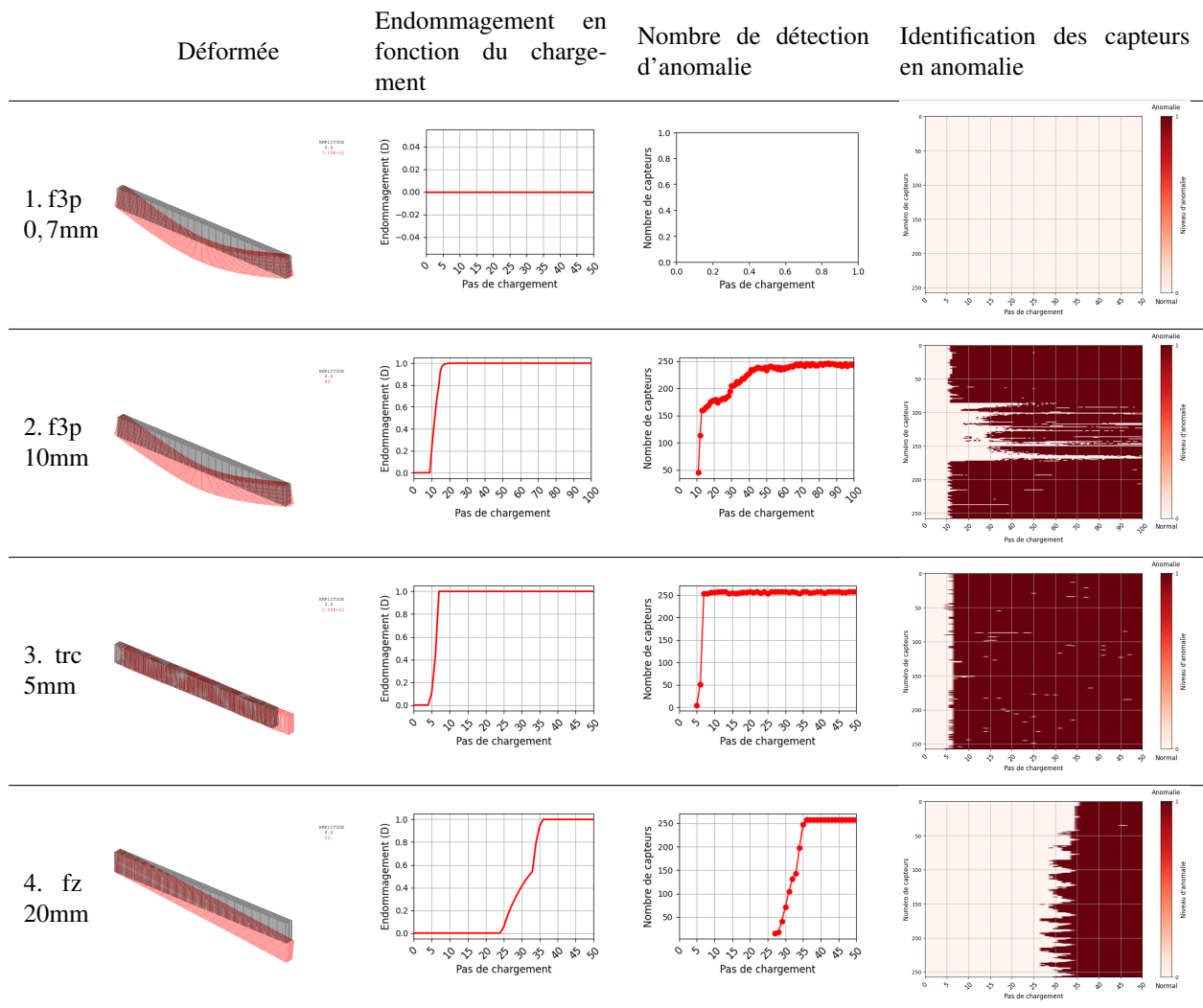


FIGURE 3 – Détection d’anomalie pour les différents scénarios.

est en pratique non accessible et la détection d’anomalie doit justement détecter la présence de l’endommagement. La troisième colonne relève le nombre de capteurs ayant détecté une anomalie en fonction du pas de chargement. La dernière colonne représente les capteurs ayant détecté une anomalie, avec en ordonnée le numéro du capteur et en abscisse le pas de chargement; une anomalie est représentée en rouge sur ce diagramme.

Dans le premier scénario de flexion trois points, sans endommagement (cas 1), nous constatons que notre méthode ne détecte aucune anomalie, ce qui était le résultat attendu. Pour les autres scénarios, notre méthode renvoie effectivement une détection d’anomalie quand des dommages apparaissent sur la structure.

L’endommagement apparait au pas de chargement 10 pour la flexion trois points d’amplitude 10 mm (cas 2) et au pas de chargement 4 pour la traction de 5 mm (cas 3). Au premier pas de chargement d’apparition de l’endommagement, la méthode ne détecte pas immédiatement d’anomalie. Cela s’explique par le fait que la non-linéarité produite par l’endommagement reste faible devant le bruit des capteurs. En revanche, dès les pas de chargement suivants, la non-linéarité devient suffisamment importante pour provoquer des détections d’anomalie sur plusieurs capteurs, ce qui est bien marqueur d’un changement de régime de la structure.

### 4.3 Sensibilité de l’ACP à la base d’apprentissage

Les performances d’un modèle de Machine Learning dépendent de la qualité et de la quantité des données d’apprentissage. Afin de mettre en évidence cet effet sur la détection d’anomalies, nous avons modifié la base d’apprentissage produite au paragraphe 3.3 en divisant l’ensemble des valeurs d’entraîne-

ment et le bruit par 10. Cette opération n'impacte pas les liens attendus a priori entre les capteurs. Nous calculons l'ACP correspondant à cette nouvelle base d'apprentissage et nous l'utilisons pour effectuer de nouveau la détection d'anomalie sur le cas 3 de scénario de traction proposé en section 4.1. Les résultats sont présentés en figure 4.

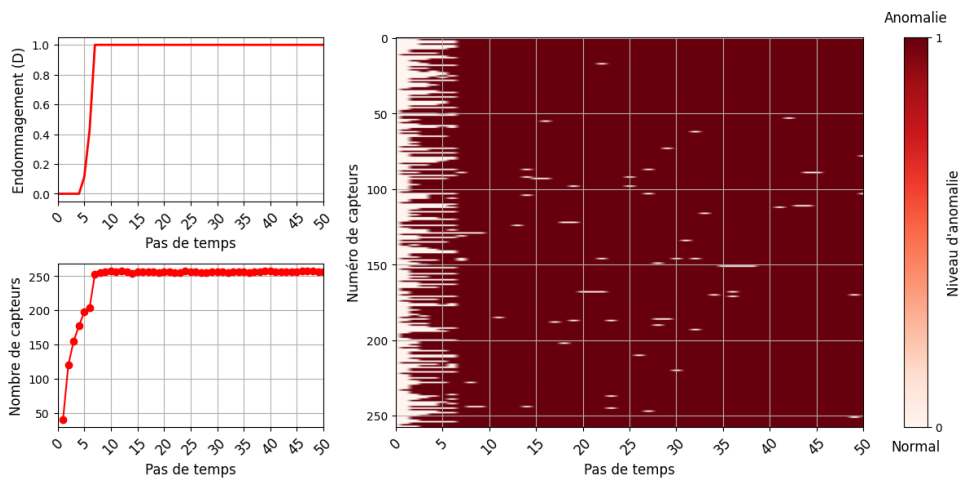


FIGURE 4 – Impact de la base d'apprentissage sur la détection d'anomalie.

On constate que notre méthode détecte des anomalies dès les premiers pas de chargement : cela correspond à des faux positifs car la structure est encore en comportement élastique sur ces pas de chargement (cf endommagement fig. 4). Nous proposons l'explication suivante : le bruit de mesure dans les données d'apprentissage engendre une incertitude sur les vecteurs propres calculés par l'ACP. Cette incertitude se repercute ensuite sur l'erreur de reconstruction.

Formellement, notons  $W$  la matrice des vecteurs propres que l'on obtiendrait sans bruit sur les données d'apprentissage et  $W_b$ , la matrice des vecteurs propres obtenus en considérant les données bruitées. Elle peut s'exprimer sous la forme  $W_b = W + \varepsilon_W$  où  $\varepsilon_W$  est une matrice aléatoire issue de l'incertitude sur les vecteurs propres calculés par l'ACP.

Considérons un vecteur de test  $U_b$  bruité et  $U_p$  ce même vecteur « parfait » que l'on aurait obtenu sans bruit. Nous pouvons exprimer une relation entre les deux vecteurs  $U_b = U_p + b$ , où  $b$  est la composante de bruit de mesure que nous avons dans notre cas choisi comme étant gaussien et homoscédastique. Dans cette situation, l'erreur  $\varepsilon$  entre le vecteur  $U_r$  reconstruit par ACP et le vecteur de test  $U_b$  s'exprime

$$\varepsilon(U_b) = U_b(I - W_b W_b^\top) = U_p(I - W W^\top) + b(I - W W^\top) - U_b(W \varepsilon_W^\top + \varepsilon_W W^\top + \varepsilon_W \varepsilon_W^\top). \quad (5)$$

Cette décomposition fait apparaître trois termes :

- le terme  $U_p(I - W W^\top)$  est l'erreur de reconstruction du vecteur de test sans bruit, par les composantes principales non bruitées. Si la donnée de test est en comportement normal, il est attendu que l'ACP soit capable de reconstruire parfaitement la donnée non bruitée et donc ce terme doit être égal à 0 et s'il est non nul, cela doit correspondre à une anomalie ;
- le terme  $b(I - W W^\top)$  est la part due au bruit de mesure sur le vecteur testé. C'est ce qui explique pourquoi la reconstruction n'est pas parfaite lorsque la donnée de test est bruitée, le bruit n'étant pas reconstruit par l'ACP. Cette composante d'incertitude est ce qui nous a poussé à mettre en place le critère statistique de détection d'anomalie en section 2.2.3 ;
- le terme  $U_b(W \varepsilon_W^\top + \varepsilon_W W^\top + \varepsilon_W \varepsilon_W^\top)$  est le terme d'erreur qui est dû à l'incertitude sur les vecteurs propres calculés par l'ACP à cause du bruit sur les données d'apprentissage. Ce terme s'ajoute donc à l'erreur de reconstruction et peut impliquer de fausses détections s'il n'est pas pris en compte. Il évolue proportionnellement à l'amplitude du déplacement  $U_b$ . Le fait d'avoir réduit d'un ordre de grandeur les valeurs des données d'entraînement a donc augmenté l'impact relatif de l'incertitude  $\varepsilon_W$  sur l'erreur de reconstruction par rapport au terme  $b(I - W W^\top)$ .

## 5 Conclusion

Nous avons mis en place une approche de détection d'anomalies de capteurs par ACP. Nous l'avons validée et testée sur des données synthétiques générées par une maquette numérique. Les données d'apprentissage proviennent du fonctionnement normal de la structure, ici représenté par le comportement élastique du système, tandis que les données à tester sont issues de simulations de la même maquette soumise à des chargement plus importants, provoquant un endommagement de la structure. Nous avons montré qu'avec une base d'apprentissage pertinente, notre méthode est en mesure de détecter correctement l'endommagement de la structure peu après son apparition.

Néanmoins, des améliorations sont envisagées et feront l'objet de prochains travaux. D'une part, il serait intéressant d'approfondir la méthode pour essayer d'obtenir une détection plus précoce de l'endommagement dès son apparition. Pour ce faire, nous envisageons d'analyser les corrélations entre les erreurs de reconstruction sur les capteurs. En effet, une anomalie peut produire un effet collectif sur plusieurs capteurs, provoquant une erreur globale, sans que l'erreur individuelle ne dépasse le seuil de détection prédéfini. L'analyse de cette erreur globale fera l'objet de travaux futurs pour comprendre si cela peut permettre de détecter plus efficacement et précocément les anomalies.

D'autre part, le critère de détection d'anomalie devra prendre en compte l'influence du bruit dans la base d'apprentissage. Nous cherchons à obtenir une estimation théorique de ce facteur dans l'erreur de reconstruction qui soit généralisable quelle que soit la structure considérée. En effet, dans un cas d'application réel, les composantes principales non bruitées  $W$  ne seront pas connues car elles seront estimées à partir de données de mesure au début de la mise en service de la structure, donc bruitées.

Enfin, l'application de la détection d'anomalie constitue la première phase d'une méthodologie en trois étapes visant à assurer le suivi d'un système mécanique. Les deux phases suivantes sont :

1. le diagnostic de l'anomalie détectée permettant de vérifier l'état réel du système à l'aide d'une simulation mécanique en appliquant les données des capteurs comme condition limite ;
2. la prise de décision sur l'anomalie : identifier le type d'anomalie (cause structurelle, défaillance de capteur), afin de planifier les actions de maintenance ou d'intervention adaptées.

Cette démarche intégrée illustre la capacité de l'approche proposée à contribuer à la surveillance et à la maintenance prédictive des structures de génie civil.

## Références

- [1] Davis, A. G. et al. *Nondestructive Test Methods for Evaluation of Concrete in Structures*, American Concrete Institute (ACI) Committee 228, 1998
- [2] Courtois, Alexis et al. *La surveillance en exploitation des enceintes de confinement et des aéroréfrigérants à tirage naturel du parc nucléaire d'EDF*, Revue Générale Nucléaire, 49-59, 2011.
- [3] Bremer, K. et al. *Fibre Optic Sensors for the Structural Health Monitoring of Building Structures*, Procedia Technology, 524-529, 2016.
- [4] Xiong, Q. et al. *Machine learning-driven seismic failure mode identification of reinforced concrete shear walls based on PCA feature extraction*, Structures, 1429-1442, 2022.
- [5] Tan, X. et al. *Monitoring restrained shrinkage and cracks of ultra-high-performance concrete (UHPC) using distributed fiber optic sensors*, Construction and Building Materials, 2024.
- [6] Svante Wold, Kim Esbensen et al. *Principal component analysis*, Chemometrics and Intelligent Laboratory Systems, 37-52, 1987.
- [7] Villegas Berbesi, T. et al. *Principal Component Analysis for Fault Detection and Diagnosis : Experience with a Pilot Plant*, Proceedings of the International Conference on Computational Intelligence, Man-Machine Systems and Cybernetics, 2010.
- [8] Yan, A.-M. et al. *Structural damage diagnosis under varying environmental conditions—Part I : A linear analysis*, Mechanical Systems and Signal Processing, 847-864, 2005.
- [9] Yan, A.-M. et al. *Structural health monitoring, Effects of environmental variations, Local principal component analysis, Bridge diagnosis*, Mechanical Systems and Signal Processing, 865-880, 2005.
- [10] *Cast3M* 2024, <https://www-cast3m.cea.fr/>.
- [11] Pedregosa, Fabian and Varoquaux et al. *Scikit-learn : Machine Learning in Python*, Journal of Machine Learning Research, 2825–2830, 2011.

- [12] P. Férignac. *Test de Kolmogorov-Smirnov sur la validité d'une fonction de distribution*, Revue de Statistique Appliquée, 1962.