

Analyse et prédiction de séries temporelles multivariées pour la surveillance de la santé d'ouvrages d'art

Fatiha CHABI¹ Franziska SCHMIDT¹ Jean-François BERCHER²

¹Univ Gustave Eiffel, MAST-EMGCU, F-77454 Marne-la-Vallée, France

²Univ Gustave Eiffel, ESIEE Paris, CNRS, LIGM, F-77454 Marne-la-Vallée, France

Résumé – La surveillance de l'état des structures comme les ponts est de plus en plus nécessaire et utilisée. Dans cet article, nous utilisons un ensemble de données réelles et proposons une méthode de surveillance basée sur un modèle statistique des données fournies par un ensemble de capteurs déployés sur la structure. Le modèle statistique est un modèle linéaire multivarié de type VAR structurel, régularisé L_1 , qui permet l'utilisation conjointe de nombreux capteurs. En comparant les prédictions du modèle avec les valeurs réelles, nous montrons qu'un tel modèle permettrait de détecter à la fois des anomalies soudaines et des dérives lentes, que ce soit sur la structure ou sur les capteurs.

Abstract – Due to the general aging of our bridge stock, Structural Health Monitoring is increasingly used to monitor the structural behavior of these structures or a portion of their components, and to provide data for bridge structural monitoring. In this paper, we analyze and build on a real-world data set to propose a monitoring method based on a joint statistical model of a set of sensors deployed on the structure. The statistical model is a multivariate linear VAR-like model, regularized by LASSO, which allows the joint use of many sensors. By comparing model's predictions with the real values, we show that such a model would allow detecting of both sudden anomalies and slow drifts, either on the structure or on the sensors.

1 Introduction

La durée de vie théorique d'un ouvrage d'art est estimée à 100 ans et une grande partie des structures françaises, construites dans l'après-guerre devraient atteindre leur fin de vie prochainement. Cette durée de vie varie en fonction de plusieurs facteurs tels que le type de pont mais aussi les conditions environnementales, des conditions d'utilisation, ou encore la qualité et la régularité de l'entretien. Souvent, le suivi des dégradations est uniquement visuel, ce qui est insuffisant pour identifier certains types de pathologies, comme la corrosion, qui se développe à l'intérieur des matériaux. Une solution aujourd'hui largement proposée est la surveillance intensive des structures, aussi appelée « *Structural Health Monitoring* ». Néanmoins, cette surveillance peut parfois conduire à une (très) grande quantité de données, qu'il peut donc être difficile de traiter. Il peut aussi être difficile d'extraire des informations pertinentes qui aideraient le gestionnaire de l'infrastructure à prendre des décisions relatives à la maintenance, à la réparation ou aux actions de renforcement [8].

La détection des dommages est traditionnellement réalisée par l'analyse des vibrations [4, 2]. Le lien entre le comportement dynamique et l'endommagement est dû au fait que la présence d'un endommagement diminue la rigidité globale de la structure, et conduit donc à une réduction de ses fréquences naturelles.

Avec le développement de volumes de données plus importants, de nouvelles méthodes de détection des dommages utilisant des méthodes issues du domaine de l'apprentissage automatique ont également été proposées [7, 9, 11]. Récemment des réseaux neuronaux ont été proposés pour prédire des séries temporelles et évaluer l'écart entre ces données prédites et les données mesurées [6, 10].

Dans cet article, on considère les données issues d'un ou-

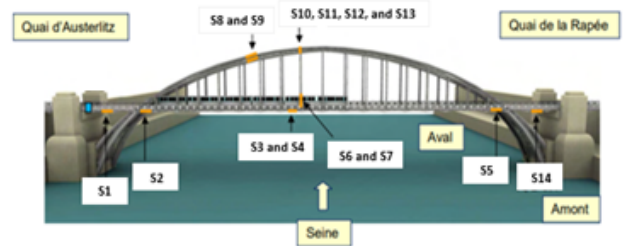


FIGURE 1 : Le viaduc d'Austerlitz, avec les capteurs déployés.

vrage particulier, le Viaduc d'Austerlitz. Dans la section 2, on décrit l'ouvrage, les caractéristiques des données et la préparation des données brutes, étape importante s'il en est. On se propose ensuite, section 3 d'établir un modèle joint des séries temporelles issues d'un ensemble de capteurs déployés sur la structure. Le modèle statistique est un modèle linéaire multivarié de type VAR, régularisé par L_1 , qui permet l'utilisation conjointe de nombreux capteurs. Dans la section 4, en envisageant différents types d'anomalies, nous montrons qu'un tel modèle permettrait de détecter à la fois des anomalies soudaines et des dérives lentes, que ce soit sur la structure ou sur les capteurs. Enfin nous concluons sur les perspectives de ce travail.

2 Description et préparation des données

Le viaduc d'Austerlitz est un pont en acier qui permet le passage de la ligne 5 du métro entre les gares "Austerlitz" et "Quai de la Rapée" à Paris (Figure 1). Pour contrôler l'état de sa structure, des capteurs de déformation à fibre optique

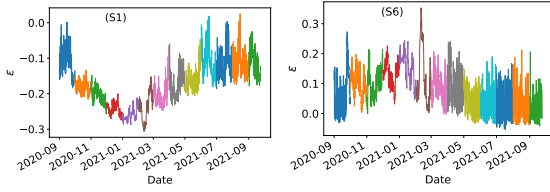


FIGURE 2 : Évolution temporelle des sorties des capteurs S1 (tablier) et S6 (structure verticale).

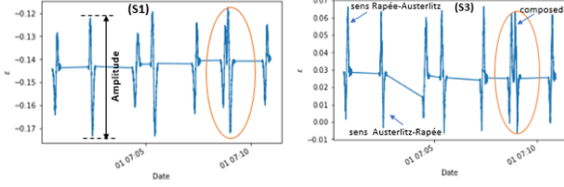


FIGURE 3 : Données brutes recueillies en S1, S3 le 01/09/2020 de 7h00 à 7h10

(14 capteurs) et de température (3 capteurs) ont été installés. Parmi les 14 capteurs de contrainte, 8 sont installés au niveau du tablier, 2 au niveau du cintre, et 6 capteurs sont fixés au niveau de l’arche du pont.

- Pour les capteurs de contrainte, une année entière de mesure est disponible, du 01/09/2020 au 23/09/2021. L’acquisition de ces capteurs est déclenchée au passage du métro (l’enregistrement n’est donc pas continu), avec une fréquence de 100 Hz. Pour chaque capteur, on dispose d’environ 500 millions de points de données.
- Pour les 3 sondes de température, l’acquisition est réalisée toutes les 10 minutes.

Au total, combiné aux données de service (date et heure de l’enregistrement), ces enregistrements représentent plusieurs dizaines de Go de données. Quelques exemples de l’évolution temporelle des données brutes sont fournis Figure 2.

Effet de la température – Les données des capteurs sur le tablier comme sur les structures verticales sont très corrélées à la température, tandis que les structures en arc y semblent moins sensibles. Comme nous disposons de capteurs de température, il est aisé de compenser cet effet en soustrayant la partie corrélée de chacun des capteurs avec la température moyenne.

Résumé des données et sens de passage des trains – Des exemples de signaux bruts sont donnés Figure 3. On y distingue des variations de forme et d’amplitude, qui dépendent du sens de passage du métro ; ainsi que des signaux composites (passage proche ou simultané de deux trains). Un petit algorithme de classification a été mis en oeuvre afin de séparer les enregistrements en fonction des sens de passage et supprimer les composites. Afin de pouvoir exploiter les données (500 millions de points par capteur), nous avons choisi, dans un premier temps, de ne retenir que quelques caractéristiques simples, à savoir, pour chaque enregistrement lors du passage d’un train : le niveau moyen μ , la dynamique A (cf Figure 3-S1), et la durée T_s du signal considéré.

Saisonnalité - Les données présentent des fortes variations à l’échelle du jour, liées à la variation de la température jour-nuit. Pour éliminer cet effet, nous avons d’abord ré-échantillonné les données à une période heure, puis filtré avec une moyenne glissante sur 24h. Une illustration est donnée Figure 4 pour le capteur S1.

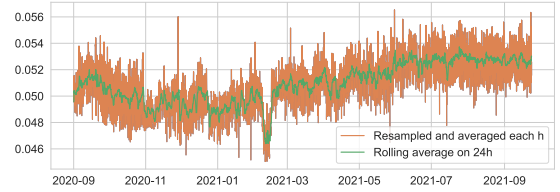


FIGURE 4 : Lissage de l’amplitude sur S1 pour supprimer l’effet de saisonnalité à l’échelle jour.

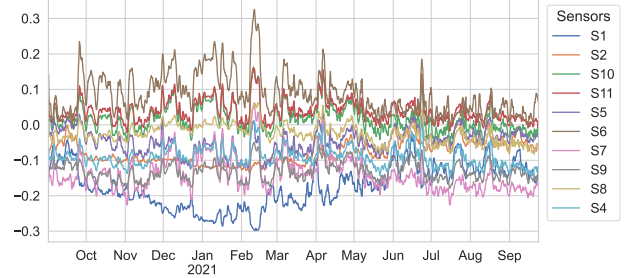


FIGURE 5 : Niveau moyen μ sur les capteurs (avec moyenne glissante sur 24h)

La Figure 5 montre les moyennes μ résultantes dans la direction Austerlitz-Rapée.

Enfin, les différents signaux sont centrés et normalisés à une variance unité.

3 Modèle statistique

À partir de la connaissance détaillée de la structure et de la physique des capteurs, nous pourrions établir un modèle direct qui nous permettrait d’analyser les signaux recueillis. Ici, nous nous concentrons sur un modèle statistique pour décrire et prédire l’évolution des signaux. L’intérêt d’un tel modèle est de prédire l’évolution d’un signal, ce qui peut être pertinent, par exemple, si un capteur est défectueux. L’analyse des résidus de prédiction, qui peut être effectuée en ligne au fur et à mesure des acquisitions, permettra également de détecter des anomalies, que ce soit dans les mesures, le comportement des capteurs ou la structure. Bien entendu, une interprétation physique des paramètres du modèle peut être recherchée.

3.1 Modèle multivarié

Afin de combiner les capteurs pour tirer profit des informations et des interactions entre eux, nous considérons un modèle multivarié et une variante d’un modèle VAR (Vector Autoregressive). Nous utilisons une régression linéaire régularisée multivariée, à la fois entre les capteurs et en fonction du temps. Notons $x(t)$, le vecteur de dimension d rassemblant toutes les mesures des d capteurs à l’instant t . La régression multivariée revient à estimer $x(t)$ selon

$$\hat{x}(t) = A_0 x(t) + \sum_{k=1}^p A_k x(t-k) + c, \quad (1)$$

où les matrices A_k sont $d \times d$, c est un vecteur de décalages. L’erreur d’estimation est notée $\epsilon(t) = \hat{x}(t) - x(t)$. Ainsi

$$x(t) = A_0 x(t) + \sum_{k=1}^p A_k x(t-k) + c + \epsilon(t). \quad (2)$$

Ce modèle ressemble à un modèle VAR structurel (SVAR) [5], si ce n'est que nous n'exigeons pas ici que les composantes de $\epsilon(t)$ soient non corrélées.

En notant $[A_k]_i$ la i^e ligne de A_k , la i^e composante s'écrit

$$x_i(t) = [A_0]_i x(t) + \sum_{k=1}^p [A_k]_i x(t-k) + c_i + \epsilon_i(t). \quad (3)$$

Deux cas sont envisagés : pour un capteur, (1) l'estimée à l'instant t est calculée en fonction des autres capteurs au même instant et de l'ensemble du passé et (2) l'estimée ne dépend que des autres capteurs (pas de lui même). Le second cas impose que les diagonales des A_i sont contraintes à zéro quand seule A_0 est contrainte dans le premier cas. Le second cas permettrait d'estimer les sorties d'un capteur quand bien même il serait défectueux. Faute de place, seul ce second cas est explicité dans la suite.

3.2 Estimation des paramètres

En collectant les observations de $x_i(t)$, pour $t = 0..T-1$ dans un vecteur x_i , en définissant de la même manière un vecteur ϵ_i et en notant

$$a_i^T = [[A_0]_i^T, [A_1]_i^T \dots [A_p]_i^T, c_i]$$

où $[A_k]_i^T$ dénote la i^e ligne de A_k privée de sa i^e composante, alors on voit que (3) peut être ré-écrite comme

$$x_i = X_i a_i + \epsilon$$

où X_i est $T \times ((d-1)(p+1) + 1)$, a_i est de dimension $(d-1)(p+1) + 1$; et x_i et ϵ sont de dimension T . X_i est essentiellement une matrice bloc Toeplitz construite à partir des $x(t)$, complétée par une colonne de 1; si ce n'est qu'on supprime les $(kd+i)^e$ colonnes, ceci afin de satisfaire implicitement la contrainte de diagonale nulle sur les A_k . Les paramètres a_i , pour $i = 1..d$, peuvent bien entendu être recherchés par la méthode des moindres carrés ordinaires, sous réserve que le nombre d'observations soit suffisant. Cependant, le nombre total de paramètres, égal à $d(d-1)(p+1) + d$, peut être très grand – e.g. pour $d = 10$ et $p = 4$ nous avons 460 paramètres libres ! Dans de telles conditions, et nous devons utiliser une méthode de sélection de variables ou une méthode de régularisation; et ce d'autant plus que le problème peut même être sous déterminé.

Nous choisissons d'utiliser une régularisation L_1 , qui effectue une sorte de sélection de variables lors du calcul de la régression. En effet, elle a pour effet connu d'assurer une parcimonie en forçant certains coefficients à zéro [3]. Pour une composante, le critère correspondant s'écrit :

$$\|x_i - X_i a_i\|_2^2 + \alpha_i \|a_i\|_1$$

où $\|\cdot\|_2^2$ et $\|\cdot\|_1$ désignent respectivement les normes L_2 (somme des carrés) et L_1 (somme des valeurs absolues). Le second terme est le terme de régularisation, avec un poids donné par l'hyperparamètre α_i .

Bien entendu, il faut également choisir une valeur approchée de l'ordre p et de α_i . Une valeur optimale de ces hyperparamètres pourrait être sélectionnée en évaluant l'erreur sur un ensemble de test ou par validation croisée. Du fait de la régularisation, le choix de p n'est pas critique. Compte tenu du grand

nombre de paramètres et du nombre limité de données d'entraînement (387 jours), les valeurs α et p sont sélectionnées à l'aide du critère AIC bien connu [1], qui peut être calculé directement sur l'ensemble d'apprentissage.

4 Expérimentations et résultats

4.1 Dispositif expérimental

Les données réelles ont été collectées du 1er septembre 2020 au 23 septembre 2021. Nous n'avons pas connaissance d'anomalies dans la structure durant cette période (ce qui est probablement une chance pour les voyageurs quotidiens !). Pour illustrer et valider notre approche, nous apprendrons le modèle sur une partie des données (du 2 septembre au 17 mai : deux tiers des données). Pour tester la capacité du modèle à détecter des anomalies, nous avons introduit artificiellement des « anomalies » dans les données suivantes. Des expérimentations ont été menées sur les données de dynamique et d'évolution moyenne, mais seuls des résultats sur les données de moyennes sont présentés ici.

Pour les données moyennes, nous avons introduit

- La simulation de variations brutales de l'amplitude : une atténuation de 0.75 sur S9 entre le 20 mai et le 20 juillet; suivie d'un facteur de 1.2 du 20 juillet à la fin.
- Une dérive linéaire de l'amplitude de S7 à partir du premier juin.

4.2 Un indicateur de suivi

Le modèle fournit des résidus pour chacun des capteurs. L'écart-type sur les résidus, calculé sur l'apprentissage, permet de construire un intervalle de confiance qui peut être utilisé pour identifier des anomalies potentielles. Il est aussi intéressant de disposer d'un indicateur unique qui permette de résumer la situation. Comme les résidus R sur chaque capteur sont corrélés, nous pouvons chercher à les recombinaison de manière constructive pour identifier de potentielles anomalies. La solution du problème qui consiste à trouver la meilleure combinaison linéaire $r = R w$, de plus forte variance, à partir d'une collection de vecteurs R , est bien connue : il s'agit de sélectionner pour w le vecteur propre associé à la plus grande valeur propre de la matrice de corrélation de R . Le représentant associé est alors la composante principale.

4.3 Résultats

Pour la simulation des « ruptures », les données originales, modifiées ainsi que la prédiction sont présentées Figure 6, calculée avec $p = 4$. L'examen des coefficients du modèle (non montrés ici) montre que la régularisation L_1 permet d'annuler un grand nombre de coefficients, qui apparaissent essentiellement pour un retard nul (interaction entre les capteurs au même instant) ou faible. La dépendance temporelle est plus longue pour les données d'amplitude (dynamique du signal).

Pour la simulation « Dérive », les données originales, modifiées et la prédiction ($p = 4$) sont présentées Figure 6.

On constate sur la figure 6 que la prédiction est très proche des données initiales, ce qui montre que la méthode est capable de corriger les anomalies qui ont été introduites ! Ceci

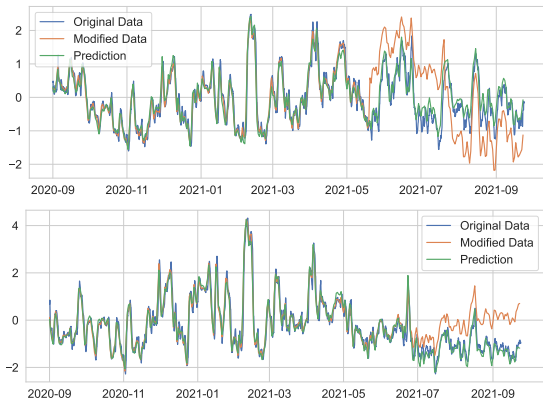


FIGURE 6 : Données initiales, modifiées et prédiction ($p = 4$) pour des ruptures simulées en S9 (haut) et une dérive simulée en S7 (bas).

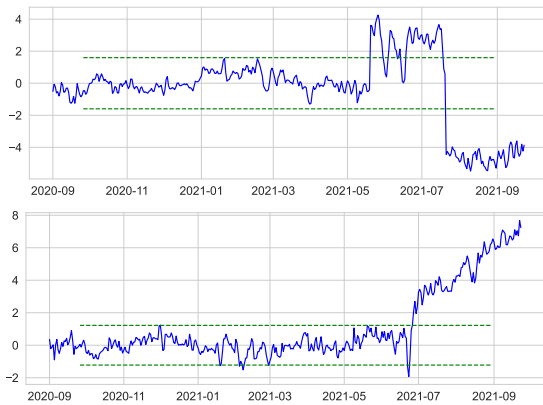


FIGURE 7 : Composante principale des résidus pour les expérimentations « Ruptures » (haut) et « Dérive » (bas).

est possible grâce à la capacité de prédiction basée sur les redondances fournies par les autres capteurs.

L'indicateur combiné, la composante principale des résidus, défini en section 4.2 et présenté Figure 7 confirme ceci et permet de détecter clairement les anomalies introduites.

5 Conclusion

Dans cet article, nous avons examiné un riche ensemble de données réelles, décrit leurs caractéristiques et développé modèle conjoint de l'évolution des capteurs déployés sur la structure. La préparation des données comprend le rééchantillonnage et l'interpolation des données obtenues de manière irrégulière, l'élimination des effets liés à la température et la normalisation des données. Le modèle statistique développé est un modèle de type VAR, régularisé L_1 , qui permet l'utilisation conjointe de tous les capteurs en sélectionnant automatiquement les données les plus pertinentes. Le paramètre de régularisation est également sélectionné de manière automatique. Un tel modèle fournit une méthode simple, reposant sur la régression de séries temporelles multivariées, pour détecter des changements dans le comportement des structures.

En introduisant artificiellement des anomalies dans les mesures - qui simulerait soit une véritable anomalie structurelle, soit une anomalie de capteur - nous illustrons qu'un tel modèle statistique VAR régularisé, appris sur une partie des données,

est par la suite capable de :

- (a) détecter des anomalies sur la partie non vue des données,
- (b) corriger l'effet de ces anomalies sur les données ;

ceci permettant ainsi d'alerter sur une anomalie qui pourrait être structurelle, ou de compenser un capteur défectueux.

Les travaux futurs se concentreront sur (i) la possibilité de définir un modèle d'apprentissage « en ligne », qui actualiserait le modèle au fil du temps ; (ii) la possibilité d'introduire des calculs physiques/mécaniques dans l'analyse.

Références

- [1] H. AKAIKE : A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.
- [2] D. A. T. BURGOS, R. C. GOMEZ VARGAS, C. PEDRAZA, D. AGIS et F. POZO : Damage identification in structural health monitoring : A brief review from its implementation to the use of data-driven applications. *Sensors*, 20, 2 2020.
- [3] T. HASTIE, R. TIBSHIRANI et M. WAINWRIGHT : *Statistical Learning with Sparsity : The Lasso and Generalizations*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press, 2015.
- [4] M. LAKHDAR, D. MOHAMMED, L. BOUDJEMÂA, A. RABIÂ et M. BACHIR : Damages detection in a composite structure by vibration analysis. *Energy Procedia*, 36:888–897, 2013.
- [5] H. LÜTKEPOHL : *New Introduction to Multiple Time Series Analysis*. Springer Berlin Heidelberg, 2007.
- [6] N. MANZINI, N. MAR, F. SCHMIDT, J.-F. BERCHER et al. : An automated machine learning-based approach for structural novelty detection based on SHM. *Proceedings of EUROSTRUCT 2021*, pages 1180–1189, 2021.
- [7] M. RASOL, F. SCHMIDT, S. IENTILE, S. SANTOS-ASSUNÇÃO et M. POMONI : Construction and building materials fric-pm : Machine learning-based road surface friction coefficient predictive model using intelligent sensor data... *Construction and Building Materials*, 2023.
- [8] O. S. SALAWU : Detection of structural damage through changes in frequency : A review. *Engineering Structures*, 19:718–723, 1997.
- [9] T. WANG, P. REIFFSTECK, C. CHEVALIER, Z. ZHU, C.-W. CHEN et F. SCHMIDT : A novel extreme gradient boosting algorithm based model for predicting the scour risk around bridge piers. *European Journal of Environmental and Civil Engineering*, 0(0):1–19, 2022.
- [10] Z.-D. XU, X. ZE et S. LI : Damage detection strategy using strain-mode residual trends for long-span bridges. *Journal of Computing in Civil Engineering*, 29, 9 2015.
- [11] Q. ZHOU, Y. NING, Q. ZHOU, L. LUO et J. LEI : Structural damage detection method based on random forests and data fusion. *Structural Health Monitoring*, 12:48–58, 1 2013.