

PRÉDICTION DE LA VITESSE DU CHEVAL DE SPORT À PARTIR DE DONNÉES ACCÉLÉROMÉTRIQUES ET GYROSCOPIQUES

Amandine Schmutz ^{1,2,3}, Julien Jacques ¹, Laurence Chèze ² & Pauline Martin ³

¹ *Entrepôts Représentation et Ingénierie des Connaissances, 5 avenue Pierre Mendès France, 69500 Bron et julien.jacques@univ-lyon2.fr*

² *Laboratoire de Biomécanique et Mécanique des Chocs, 43 boulevard du 11 Novembre 1918, 69622 Villeurbanne et laurence.cheze@univ-lyon1.fr*

³ *Entreprise CWD, Rue Fontaine de Fanny, 24300 Nontron et pmartin@lim-group.com & aschmutz@lim-group.com*

Résumé. Nous présentons une méthode de prédiction de la vitesse du cheval par un ensemble de variables fonctionnelles. Ces dernières présentent la particularité que le nombre de mesures par individu est variable. La méthode utilisée est la régression non paramétrique multivariée via un estimateur à noyau pour données fonctionnelles proposée par Ferraty et Vieu (2009).

Mots-clés. Modèles semi- et non-paramétriques, Données fonctionnelles

Abstract. We present a predictive method for horse speed with functional predictors. The dataset characteristic is that the number of measurements per individual is variable. The method used is the multivariate nonparametric regression via the functional kernel estimator proposed by Ferraty and Vieu (2009).

Keywords. Semi and Nonparametric statistics, Functional data

1 Introduction

Avec l'essor des nouvelles technologies, le monitoring des activités physiques est depuis plusieurs années en pleine expansion. Ainsi en 2003 est apparue la première montre Garmin disposant d'un GPS pour la course à pied. Ce type d'objet connecté est en train d'être décliné à de nombreuses disciplines dont les sports équestres.

CWD a positionné sur le garrot de chevaux un accéléromètre et un gyroscope. Ainsi à chaque foulée, l'accélération et la vitesse angulaire selon les 3 axes du déplacement (x, y et z) sont mesurées, ce qui permet de calculer différents paramètres comme l'énergie du mouvement, la fréquence de galop, la rotation du garrot... La fréquence d'échantillonnage élevée permet de récupérer un signal quasi continu tout au long de la performance sportive. Ce dispositif ne comporte pas de GPS car la méthode de référence d'estimation de la vitesse du cheval basée sur des acquisitions vidéos rapides, nécessite d'être en intérieur, il fallait donc un dispositif pouvant fonctionner dans des lieux couverts.

L'objectif de ce travail est d'estimer à partir de ces données, la vitesse du cheval à chaque foulée, avec une précision d'au moins 0.6 m/s afin de la retranscrire au cavalier.

2 Les données

Les données sont récoltées à l'aide d'un accéléromètre et d'un gyroscope positionnés sur le garrot du cheval. Leur fréquence d'échantillonnage est de 100 Hertz, ils récoltent des données selon les 3 axes du mouvement : x, y et z.

Une base de données a été constituée en couplant les mesures de l'accéléromètre et du gyroscope avec celles utilisant une méthode de référence : des acquisitions vidéos rapides. Cette méthode permet de mesurer la vitesse à chaque foulée dans un champ d'étude de 26 mètres. Actuellement 1906 foulées de galop en ligne droite ont été mesurées sur 34 chevaux équipés de ce dispositif. Un individu statistique est une foulée. La vitesse d'une foulée de cheval peut varier de 43 à 115 centièmes de seconde : une foulée peut durer moins d'une seconde si le cheval galope vite, ou plus d'une seconde dans le cas où le cheval galope lentement ; on a donc des longueurs de signaux qui peuvent varier d'un individu à l'autre (cf. Figure 1).

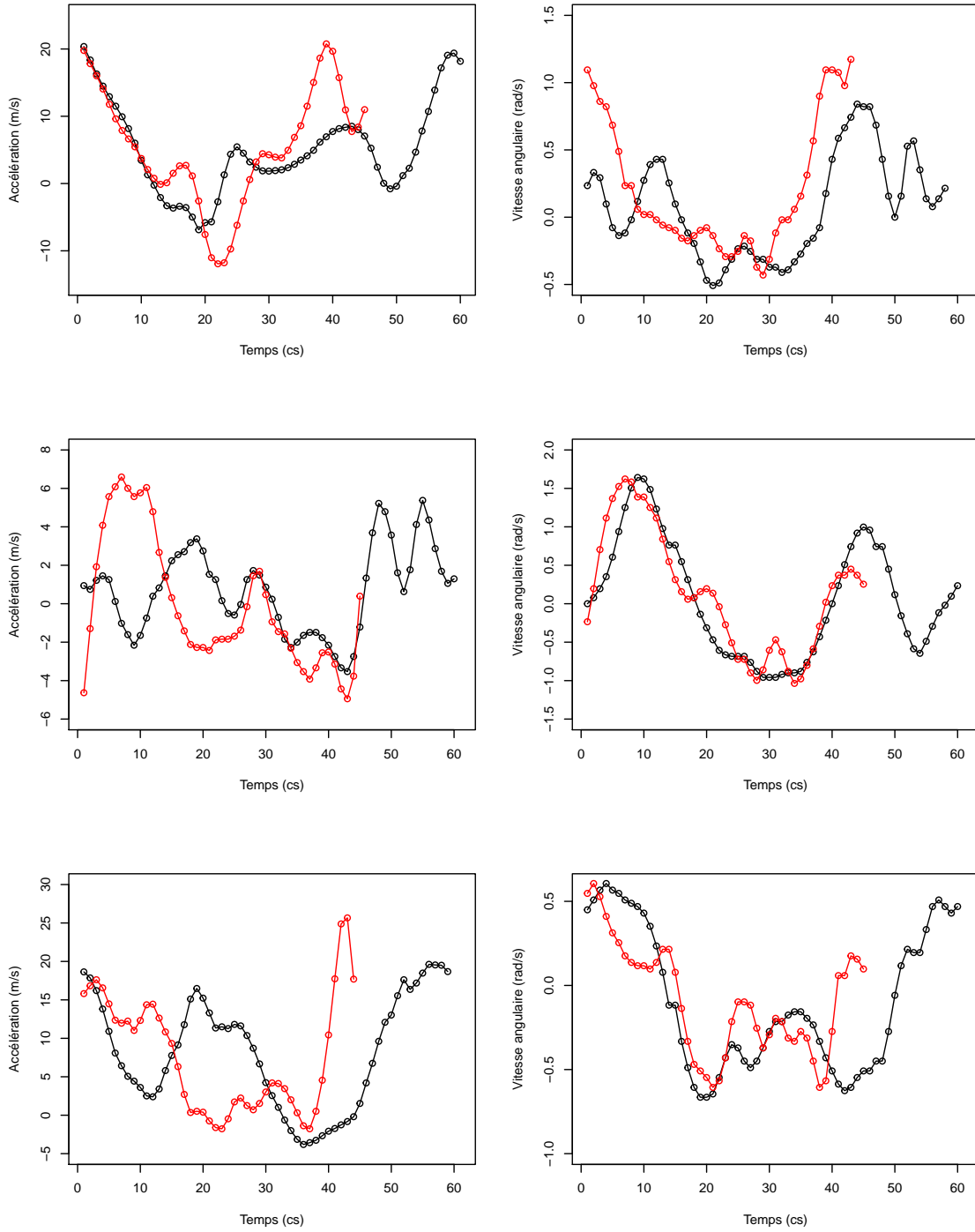


Figure 1: Courbe d'accélération (à gauche) et de vitesse angulaire (à droite) selon x, y et z pour une foulée à 8.6 m/s (rouge) et 5.0 m/s (noire)

3 Le modèle

Un signal est composé au minimum de 43 points et au maximum de 115 points. Afin de ramener chaque signal à la même longueur, nous avons choisi dans un premier temps de ne conserver que les 43 premiers points de chaque signal. Ceci nous permet donc de conserver l'intégralité du signal pour certaines foulées et le début du signal pour d'autres. Dans une étude ultérieure nous chercherons à traiter l'intégralité du signal pour chaque foulée. La base de données est alors composée de 1906 individus et 258 variables.

Une régression non paramétrique fonctionnelle multivariée (Ferraty et Vieu (2006), Ferraty et Vieu (2009)), va ensuite être utilisée pour prédire la vitesse.

Soit $\chi = (\chi^1, \chi^2, \dots, \chi^6)$ notre jeu de variables explicatives fonctionnelles et Y notre variable scalaire à expliquer. Aucune hypothèse n'est faite sur la dimension de chaque covariable χ^j . Le modèle de régression s'écrit alors :

$$\begin{aligned} Y_i &= r(\chi_i^1, \dots, \chi_i^6) + \epsilon_i, \\ E(\epsilon_i | (\chi_i^1, \dots, \chi_i^6)) &= 0. \end{aligned}$$

Où r est l'opérateur fonctionnel.

4 Résultats

Le logiciel R 3.3.2 a été utilisé pour les analyses. La fonction utilisée est `funopare.knn.lcv` disponible sur le site internet NPFDA (Ferraty et Vieu (2006)). Cette fonction permet de réaliser une régression non paramétrique pour une covariable fonctionnelle. L'application de la méthode présentée ci-dessus nécessite plusieurs étapes. Dans un premier temps on réalise une régression de la première covariable fonctionnelle explicative sur Y . On récupère les résidus de ce modèle puis l'on réalise une régression de la deuxième covariable fonctionnelle sur les résidus du premier modèle. On procède ainsi par itération jusqu'à avoir réalisé une régression pour chacune des covariables fonctionnelles.

Un échantillon d'entraînement et un échantillon de test sont constitués par tirage aléatoire des données avec les proportions 80/20. Il est souhaité que la vitesse soit estimée avec une précision d'au moins 0.6 m/s, nous allons donc prendre comme indicateur de qualité de prédiction le pourcentage d'erreur de prédiction au-dessus de 0.6 m/s. La Figure 2 présente les résultats d'une simulation dont le pourcentage d'erreur supérieur à 0.6 m/s est de 16%.

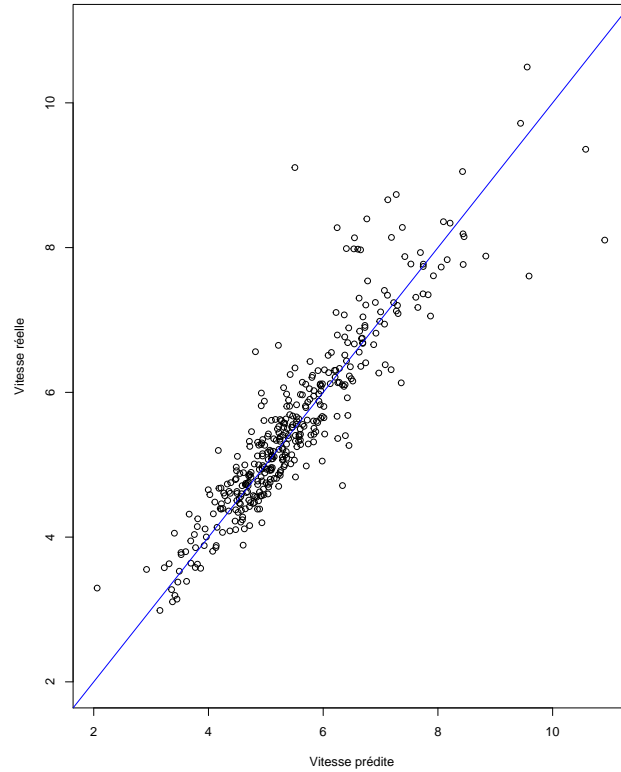


Figure 2: Résultats d’une simulation : valeurs prédites de vitesse vs valeurs mesurées avec la cinématique 2D par rapport à la première bissectrice (bleue)

Pour 500 simulations, le pourcentage d’erreur supérieur à 0.6 m/s est compris entre 9 et 24% avec une valeur médiane de 16%. Ces résultats montrent l’importance de tenir compte de plusieurs covariables fonctionnelles dans un modèle de prédiction. En effet lorsque nous avons réalisé une régression non paramétrique avec une seule covariable fonctionnelle, l’erreur de prédiction était comprise entre 14 et 29% avec une valeur médiane de 21% (cf. Figure 3). Par contre il est important de vérifier après l’ajout de chaque nouvelle covariable le pourcentage d’erreur, car selon la simulation, le modèle complet n’est pas toujours le meilleur modèle de prédiction.

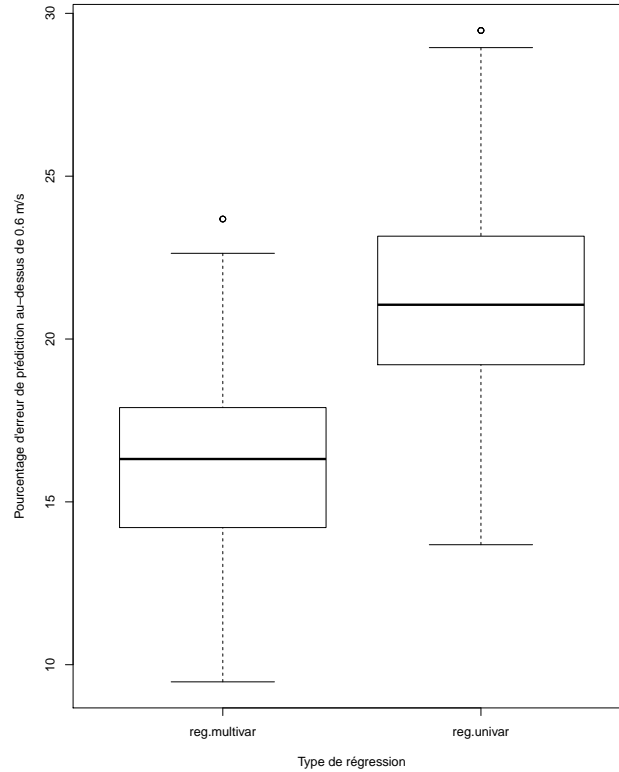


Figure 3: Pourcentage d’erreur de prédiction en fonction du type de régression

Pour conclure, les résultats obtenus en utilisant les 43 premiers points du signal sont encourageants. Nous allons chercher à améliorer ce modèle de façon à prendre en compte l’intégralité du signal. De plus, compte tenu de la grande variabilité inter-individus, une répartition en classes pourrait être envisagée : les observations ayant le même type de comportement seraient classées dans un même groupe. Il pourrait être intéressant d’étudier l’apport que pourrait avoir un modèle de mélange de régressions sur la qualité des prédictions.

Bibliographie

- [1] Ferraty, F. et Vieu, P. (2009), Additive prediction and boosting for functional data, *Computational Statistics and Data Analysis*, 53, 1400–1413.
- [2] Ferraty, F. et Vieu, P. (2006), Nonparametric functional data analysis theory and practice, édition Springer, New York.
- [3] Ferraty, F. et Vieu, P. (2006), NPFDA.[en ligne]. <https://www.math.univ-toulouse.fr/staph/npfda/> [consulté le 01/10/2016]