



HAL
open science

Analyzing cycling sensors data through ordinal logistic regression with functional covariates

Julien Jacques, Sanja Samardžić

► **To cite this version:**

Julien Jacques, Sanja Samardžić. Analyzing cycling sensors data through ordinal logistic regression with functional covariates. *Journal of the Royal Statistical Society: Series C Applied Statistics*, 2022, 71 (4), pp.969-986. 10.1111/rssc.12563 . hal-03107427v2

HAL Id: hal-03107427

<https://hal.science/hal-03107427v2>

Submitted on 23 Nov 2021

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.

Analyzing cycling sensors data through ordinal logistic regression with functional covariates

Julien JACQUES

Université de Lyon, Lyon 2, ERIC UR 3083, Lyon, France.

E-mail: julien.jacques@univ-lyon2.fr

Sanja Samardžić

Université de Lyon, Lyon 2, ERIC UR 3083, Lyon, France.

Summary. With the emergence of digital sensors in sports, all cyclists can now measure many parameters during their effort, such as speed, slope, altitude, heart rate or pedalling cadence. The present work studies the effect of these parameters on the average developed power, which is the best indicator of cyclist performance. For this, a cumulative logistic model for ordinal response with functional covariate is proposed. This model is shown to outperform competitors on a benchmark study, and its application on cyclist data confirms that pedalling cadence is a key performance indicator. However, maintaining a high cadence during long effort is a typical characteristic of high-level cyclists, which is something on which amateur cyclists can work to increase their performance.

1. Data and motivation

With the emergence of digital sensors in sports (Halson, 2014), there is an increasing need for tools and methods to analyse the produced data. Cycling is no exception, with its many professional but low cost devices available to amateur cyclists (Bini et al., 2014). In cycling, the most frequently produced and used data are cyclist's speed, slope, altitude, heart rate and pedalling cadence. The left panel of Figure 1 illustrates such data for a one-hour bike session. At a slightly higher cost, power sensors also make it possible to measure the instantaneous power developed during the activity (right panel of Figure 1). The `trackeR` package (Frick and Kosmidis, 2017) for R proposes interesting visualisation tools and descriptive statistics for such data. Thus, any cyclist can contemplate the data produced following a training session or a competition. Assuming that these data are reliable indicators of the internal and external loads of the cyclist (refer to Cardinale and Varley (2017) for a discussion on this aspect), is it possible to use them to improve performance? There is a large volume of literature on the subject (Grappe, 2018; van Dijk et al., 2017) that is not always accessible to uninitiated amateur cyclists, who may nevertheless be interested in improving their performance.

If, in the last century, the reference indicator was the cyclist's heart rate, this has been dethroned in recent years by the power developed by the cyclist (Passfield et al., 2017). Indeed, while heart rate can be distorted by external elements such as the weather, heart rate being positively correlated with the temperature, this is not the case with the power, which is the best indicator of cyclist performance to date (Beattie et al., 2016; Grappe, 2012; Passfield et al., 2017). In this work, the cyclist performance refer to the final

result at a cycling competition. The winner of the competition is the cyclist who has the highest average speed, but since the speed depends on many external parameter (slope, altitude, wind direction and power, group aspiration phenomenon, cyclist morphology, material, etc), we focus on the developed power. Indeed, all external parameters fixed, the average speed directly depends on the developed power. The right panel of Figure 1 plots the power developed by the cyclist during the same cycling session as the one corresponding to the left panel. We can observed in this figure that the power data are highly irregular, and cyclists rarely use the precise value of the power developed during the effort. Cyclists are used to working with *power zones*, defined as a set range of watts, calculated on the basis of percentages of the Functional Threshold Power (FTP) (Borszcz et al., 2018; Grappe, 2012; Allen and A., 2010; Passfield et al., 2017). Several definitions of theses ranges exists, and the one used in this study is the Coggan scale (Allen and A., 2010), integrated in the Garmin device used for collecting the data. The limits of the 7 power zones for the cyclist whose data are plotted on Figure 1 are represented by the horizontal lines on the right panel of the figure.

If power is the best indicator of the cyclist's performance, it is necessary to seek to optimise it during the effort. For a cyclist with a fixed and limited capacity, several parameters could help them to optimise the power during the effort. In particular, the cyclist can easily act on their pedalling cadence, which is known to be a parameter influencing significantly the developed power (Faria et al., 2005). However, here again, knowledge in terms of cadence to be developed has changed a lot in recent years. You just have to watch videos of climbing a pass in the 1980s and now to see that the pedalling cadence is absolutely not the same. While in the 1980s cyclists sought to use the biggest gear, by working essentially on the force that they were able to develop, the paradigm evolved during the 2000s. Indeed, using a smaller force combined with a higher pedalling cadence results in a much better final performance. In addition, the final performance indicator that brought this to light is the developed power (Abbiss et al., 2009). Thus, all professional cyclists now use very high pedalling cadences, especially during long climbs where it is essential to maintain the highest possible power during the entire climb (Nimmerichter et al., 2011). However, what about amateur cyclists? Is it also possible for them to reproduce these cadences? Will a higher cadence increase the power that a cyclist is able to develop, and therefore their performance? And what is the effort length?

The goal of the present study is to give some answers to these questions, at least for the cyclist who produced the analysed data. In particular, we want to exhibit which levers of action are available for the cyclist in their practice in order to optimise the average power that they are able to maintain during the activity. And this, for different duration of the activity. For this, the proposed approach relies on a modelling of the mean power zone according to speed, slope, altitude, heart rate, pedalling cadence and outdoor temperature. Different models will be established according to the activity duration, from very short efforts, involving the cyclists' lactic capacities, to longer efforts, involving aerobic channels: 2 minutes, 10 minutes and 30 minutes. For each duration of effort, a variable selection is carried out in order to select which features are the most discriminating for the power zone. This would exhibit which levers of action which the cyclist could act on in order to optimise the mean power zone.

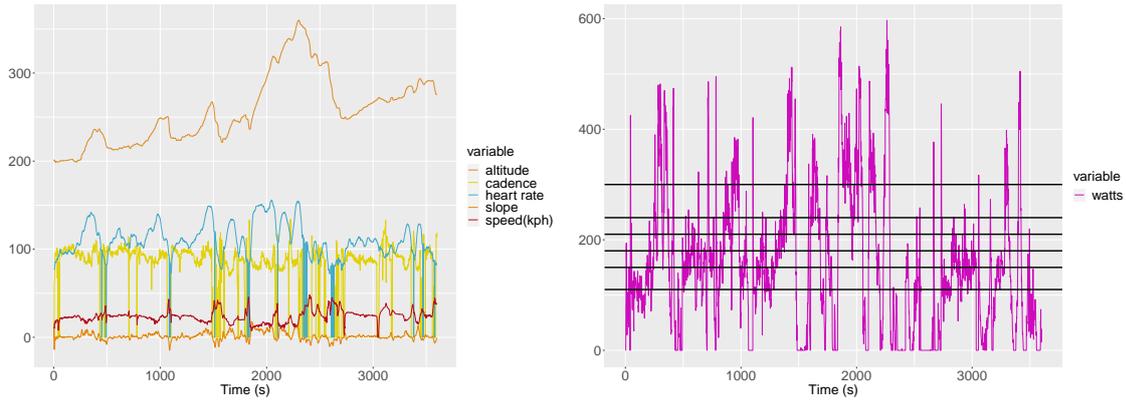


Fig. 1. Raw cycling data (left) and power (right)

From a statistical point of view, the power zones are ordinal data, whereas the other features (speed, slope, altitude, heart rate, cadence and temperature) are functional data (quantitative measures evolving over a continuum). The next section presents the existing models for ordinal and functional data and concludes that no model exists in the literature for predicting an ordinal variable from functional ones. The Functional Ordinal Logistic Regression (FOLR) model is thus proposed in Section 3, as well as its maximum likelihood inference. Section 4 presents a comparison of FOLR with alternative approaches on the basis of a real data set from the literature. Cycling data are then analysed in Section 5. Some conclusions and perspectives are given in Section 6.

2. Related work

Ordinal data is one particular type of categorical data, occurring when the categories are ordered. Such data are very frequent in practice, for instance in marketing studies where people are asked to evaluate some products or services on an ordinal scale using questionnaires. Nevertheless, it is not seldom that practitioners either consider them as quantitative integer data, assimilating the indexes of categories to integers, or even as nominal data, neglecting the order among the categories. In supervised learning, when the task is to predict an ordinal response variable, historical models are based on the modelling of cumulative probabilities that the ordinal variable is lower than a given category (Agresti, 2010). More recent research on ordinal data is essentially in classification or clustering, without using covariate. In the classification context, Cardoso and Pinto da Costa (2007) convert the problem of ordinal prediction into a binary classification problem, whereas Chu and Keerthi (2007) adapt the support vector machine paradigm to the ordinal case. In the clustering context, Jacques and Biernacki (2018) propose a mixture model based on a new distribution for ordinal data, whereas McParland and Gormley (2013) propose a latent variable approach. However, when the goal is to predict an ordinal variable using covariate, the reference models remain those modelling the cumulative probabilities with a *link*-linear model (Agresti, 2010).

In these latter, the ordinal response is predicted from the observations of scalar covariates. In the present work, we are interested in functional covariates, occurring when covariates are curves. Functional data (Ramsay and Silverman, 2005) become ubiquitous since modern technologies facilitate the collection of high frequency data. The cycling sport devices discussed in the introduction are a good example. Using the functional data approach for modeling such high frequency data has several advantages. First, it allows a parsimonious representation of high-dimensional data, which is essential in a modeling perspective. Second, it allows to take into account the inherent measurement errors, by introducing a smoothing step, which can be an answer to the reliability problem of such data identified by Cardinale and Varley (2017) and Passfield et al. (2017).

In the literature, regression with functional covariates has been developed for many types of responses. The most common is the regression model for continuous scalar response, which has been proposed either in a parametric (Ramsay and Silverman, 2005) or non-parametric way (Ferraty and Vieu, 2006). Several models have been proposed for categorical nominal response: Ratcliffe et al. (2002) propose a binary logistic regression model for functional covariate, whereas Escabias et al. (2005) propose a model based on functional principal components. A partial least square (PLS) approach has also been considered in Preda et al. (2007). A model for a functional response is also available in Ramsay and Silverman (2005). However, to the best of our knowledge, no functional regression model has been developed for ordinal response. In Preda et al. (2007), the PLS model is proposed to predict the quality of cookies from observation of the resistance of dough during the kneading process. The quality, good, adjustable or bad, is clearly expressed on an ordinal scale, but has been considered as a nominal one, more precisely as a binary one removing the adjustable category.

The present work aims to provide a prediction model for an ordered categorical response variable on the basis of functional covariates. The next section presents the Functional Ordinal Logistic Regression (FOLR) model. Section 3.2 focuses on the specificity of functional data and their modelling, whereas Section 3.3 proposes an estimation algorithm for the FOLR model.

3. The Functional Ordinal Logistic Regression model

3.1. The model

Let Y be an ordinal categorical variable, with C categories, quoted by 1 to C . Let X_j be a functional random variable ($1 \leq j \leq p$) with values in $L_2[0, T]$, $T > 0$, and assume that X_j is a L_2 -continuous stochastic process, $X_j = \{X_j(t), t \in [0, T]\}$. Cumulative logit models aim to model

$$\text{logit } p(Y \leq c | X = x) = \log \frac{p(Y \leq c | X = x)}{p(Y > c | X = x)}$$

for $c = 1, \dots, C - 1$, with a linear combination of predictors. For functional predictors, the Functional Ordinal Logistic Functional Regression (FOLR) model proposed in this paper can be written:

$$\text{logit } p(Y \leq c | X = x) = \alpha_c - \sum_j^p \int_0^T \beta_j(t) x_j(t) dt, \quad (1)$$

where $\beta_j(t)$, $t \in [0, T]$, are the functional regression coefficients, $\alpha_1 \leq \dots \leq \alpha_{C-1}$ and $1 \leq c \leq C - 1$. With this model, each cumulative logit (1) has its own intercept, whereas the effect of the covariates $X_j(t)$ is shared by all of them. The minus sign for the covariates effect is chosen in order that, for small values of $\sum_j^p \int_0^T \beta_j(t)x_j(t)dt$ the response is likely to fall in the first category and for large values the response is likely to fall in the last category. Figure 2 illustrates the corresponding probabilities (for one covariate, $p = 1$).

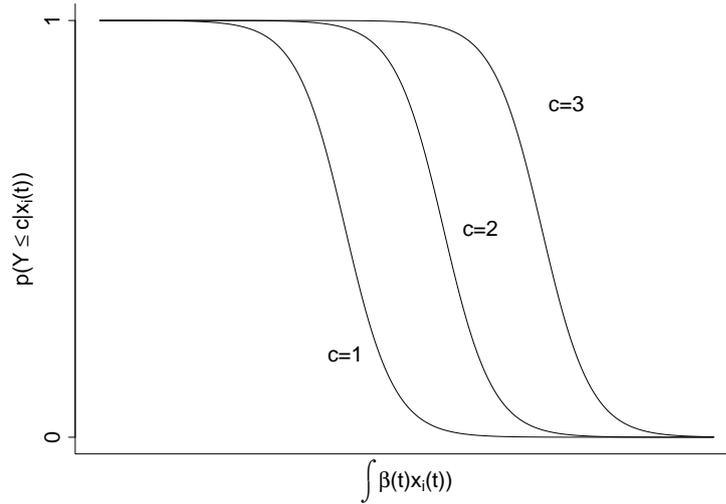


Fig. 2. Illustration of the FOLR model probabilities

In the following, only one functional covariate is considered ($p = 1$) for simplicity, but extension is straightforward. Cycling data analysed at the end of the paper are multivariate ($p = 6$).

3.2. From discrete observation to functional data

Let us consider a data set $(y_i, x_i(t))_{1 \leq i \leq n}$ of joint observations of the ordinal response Y and the functional covariate X . In practice, the functional expression of the $x_i(t)$ are not known, and we only have access to their observation at some discrete time points $0 \leq t_1 \leq \dots \leq t_S \leq T$. For simplicity of presentation, the same number of time points is considered for every $x_i(t)$, but the contrary case can in practice easily be considered. The first task, when working with functional data, is therefore to convert these discretely observed values to a function $x_i(t)$, computable for any desired argument value $t \in [0, T]$. One way to do that is interpolation, which is used if the observed values are assumed to be errorless. However, if there is some noise that needs to be removed, a common way to reconstruct the functional form is to assume that the curves $x_i(t)$ can be decomposed into a finite dimensional space, spanned by a basis of functions (Ramsay and Silverman,

2005):

$$x_i(t) = \sum_{r=1}^R a_{ir} \phi_r(t) = \mathbf{a}'_i \boldsymbol{\phi}(t) \quad (2)$$

where $\boldsymbol{\phi}(t) = (\phi_1(t), \dots, \phi_R(t))'$ is the basis of functions, R the number of basis functions, and $\mathbf{a}_i = (a_{i1}, \dots, a_{iR})'$ the basis expansion coefficients.

In an unsupervised context, there is no straight rules about the choice of the basis functions $\boldsymbol{\phi}(t)$ (Jacques and Preda, 2014). Generally, this choice is empirically done by the user, and it is recommended to use a Fourier basis in the case of data with a repetitive pattern, and B-spline functions in most other cases. In a supervised context, this choice can be optimised by cross-validation according to predictive objective.

The estimation of the coefficients \mathbf{a}_i is usually done through least square smoothing (see Ramsay and Silverman (2005)), as a preliminary step of the estimation of Model (1). If $\mathbf{x}_i = (x_i(t_1), \dots, x_i(t_s))'$ is the vector of discrete observations of $x_i(t)$, and $\boldsymbol{\Phi}$ the $S \times R$ matrix containing the $\phi_r(t_s)$ ($1 \leq r \leq R$), the least square estimation of \mathbf{a}_i is:

$$\hat{\mathbf{a}}_i = (\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\mathbf{x}_i.$$

Similarly, the functional regression coefficient $\beta(t)$ is also assumed to be decomposed into a finite basis of functions. For simplicity, it is assumed to be the same basis as for X :

$$\beta(t) = \sum_{r=1}^R b_r \phi_r(t) = \mathbf{b}'\boldsymbol{\phi}(t) \quad (3)$$

with $\mathbf{b} = (b_1, \dots, b_R)'$.

Under these basis expansion assumptions, the FOLR model is:

$$\begin{aligned} \text{logit } p(y_i \leq c | X = x_i) &= \alpha_c - \int_0^T \sum_{r=1}^R b_r \phi_r(t) \sum_{r'=1}^R a_{ir'} \phi_{r'}(t) dt \\ &= \alpha_c - \sum_{r=1}^R \sum_{r'=1}^R b_r a_{ir'} \int_0^T \phi_r(t) \phi_{r'}(t) dt \\ &= \alpha_c - \mathbf{b}'\boldsymbol{\Psi}\mathbf{a}_i \\ &= \begin{bmatrix} 1 & -\boldsymbol{\Psi}\mathbf{a}_i \end{bmatrix} \begin{bmatrix} \alpha_c \\ \mathbf{b}' \end{bmatrix} \end{aligned}$$

where $\boldsymbol{\Psi}$ is the $R \times R$ matrix of inner products between basis functions $\int_0^T \phi_r(t) \phi_{r'}(t) dt$. Let remark that if different basis are chosen for $\beta(t)$ and for $x_i(t)$, the only change will be in the matrix $\boldsymbol{\Psi}$ which would contains the scalar product between the basis functions of the two bases.

3.3. Model inference

For a data set $(y_i, \mathbf{a}_i)_{1 \leq i \leq n}$ of joint observation of the response and the basis expansion coefficients, there is a need to estimate FOLR model parameters $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \mathbf{b})$ with

$\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{C-1})$. Assuming that the observations are independent, this is done by maximizing the following log-likelihood:

$$\begin{aligned} \ell(\boldsymbol{\theta}) = & \sum_{i=1}^n (1_{y_i=1} \log g(\alpha_1 - \mathbf{b}'\Psi\mathbf{a}_i) + 1_{y_i=C} \log[1 - g(\alpha_{C-1} - \mathbf{b}'\Psi\mathbf{a}_i)]) \\ & + \sum_{i=1}^n \sum_{c=2}^{C-1} 1_{y_i=c} \log[g(\alpha_c - \mathbf{b}'\Psi\mathbf{a}_i) - g(\alpha_{c-1} - \mathbf{b}'\Psi\mathbf{a}_i)] \end{aligned} \quad (4)$$

where $g(t) = 1/(1 + \exp(-t))$ is the standard logistic cumulative density function.

In order to compute the maximum likelihood estimator, the derivative according to \mathbf{b} and $\boldsymbol{\alpha}$ are computed. By denoting $h(t) = \exp(-t)/(1 + \exp(-t))^2$ the derivative of $g(t)$, we have:

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \mathbf{b}} = & \sum_{i=1}^n \left(1_{y_i=1} \Psi\mathbf{a}_i \frac{h(\alpha_1 - \mathbf{b}'\Psi\mathbf{a}_i)}{g(\alpha_1 - \mathbf{b}'\Psi\mathbf{a}_i)} + 1_{y_i=C} \Psi\mathbf{a}_i \frac{-h(\alpha_{C-1} - \mathbf{b}'\Psi\mathbf{a}_i)}{1 - g(\alpha_{C-1} - \mathbf{b}'\Psi\mathbf{a}_i)} \right) \\ & + \sum_{i=1}^n \sum_{c=2}^{C-1} 1_{y_i=c} \Psi\mathbf{a}_i \frac{h(\alpha_c - \mathbf{b}'\Psi\mathbf{a}_i) - h(\alpha_{c-1} - \mathbf{b}'\Psi\mathbf{a}_i)}{g(\alpha_c - \mathbf{b}'\Psi\mathbf{a}_i) - g(\alpha_{c-1} - \mathbf{b}'\Psi\mathbf{a}_i)} \end{aligned} \quad (5)$$

and

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha_c} = & - \sum_{i=1}^n \left(1_{y_i=1} \frac{\delta_{c,1} h(\alpha_1 - \mathbf{b}'\Psi\mathbf{a}_i)}{g(\alpha_1 - \mathbf{b}'\Psi\mathbf{a}_i)} + 1_{y_i=C} \frac{-\delta_{c,C-1} h(\alpha_{C-1} - \mathbf{b}'\Psi\mathbf{a}_i)}{1 - g(\alpha_{C-1} - \mathbf{b}'\Psi\mathbf{a}_i)} \right) \\ & - \sum_{i=1}^n \sum_{k=2}^{C-1} 1_{y_i=c} \frac{\delta_{c,k} h(\alpha_k - \mathbf{b}'\Psi\mathbf{a}_i) - \delta_{c,k} h(\alpha_{k-1} - \mathbf{b}'\Psi\mathbf{a}_i)}{g(\alpha_k - \mathbf{b}'\Psi\mathbf{a}_i) - g(\alpha_{k-1} - \mathbf{b}'\Psi\mathbf{a}_i)} \end{aligned} \quad (6)$$

where $\delta_{c,k}$ is the Kronecker delta, equal to 1 if category c is the same as category k , 0 otherwise.

Since the maximum likelihood equations deriving from these derivatives have no closed form solutions, an iterative optimisation algorithm has to be applied. Here, we have opted for the Fisher scoring algorithm Osborne (1992). Let $\mathcal{V}(\boldsymbol{\theta})$ be the gradient of $\ell(\boldsymbol{\theta})$, composed of terms given in equations (5) and (6), and the $\mathcal{I}(\boldsymbol{\theta})$ be the Fisher Information matrix. Starting from an initialisation $\boldsymbol{\theta}^{(0)}$ of $\boldsymbol{\theta}$, the Fisher scoring algorithm updates the parameter by:

$$\boldsymbol{\theta}^{(q+1)} = \boldsymbol{\theta}^{(q)} + \mathcal{I}(\boldsymbol{\theta}^{(q)})^{-1} \mathcal{V}(\boldsymbol{\theta}^{(q)})$$

until convergence of the parameter values, i.e. when $|\boldsymbol{\theta}^{(q+1)} - \boldsymbol{\theta}^{(q)}| < \epsilon$.

The second derivatives of log-likelihood function required to compute the Fisher Information matrix $\mathcal{I}(\boldsymbol{\theta})$ are given in Appendix A.

4. Comparison with competitors

In this section we show that the proposed FOLR model is competitive compared to the closest competitors, which are the multinomial functional logistic regression and the

following non-functional methods: random forest, ordinal logistic regression and support vector machine. Rather than choosing a subjective simulated data set, we choose to base the comparison on a real data set from the literature, with an ordinal response to forecast from functional features.

4.1. *The Kneading data set*

The Kneading data set is a well-known benchmark in functional data analysis, described in detail in Lévédér et al. (2004). It concerns the quality of cookies and the relationship with the flour kneading process. There are 115 different flours for which dough resistance is measured during the kneading process for 480 seconds. One obtains 115 kneading curves observed at 241 equispaced instants of time in the interval $[0, 480]$. The 115 flours produce cookies of different quality: 50 of them produced cookies of *good* quality, 25 produced *medium* quality and 40 *low* quality. These data have already been studied in a supervised classification context (Lévédér et al., 2004; Preda et al., 2007). They are known to be hard to discriminate, even for supervised classifiers, partly because of the medium quality category. Taking into account that the resistance of dough is a smooth curve measured with error and following previous works on this data (Lévédér et al., 2004; Preda et al., 2007), least squares approximation on a basis of cubic B-spline functions is used to reconstruct the true functional form of each sample curve. If Lévédér et al. (2004); Preda et al. (2007) used cubic B-spline with 18 knots (22 basis functions), we propose to select it according to our prediction purpose using cross-validation. Raw and smoothed data are plotted on Figure 3.

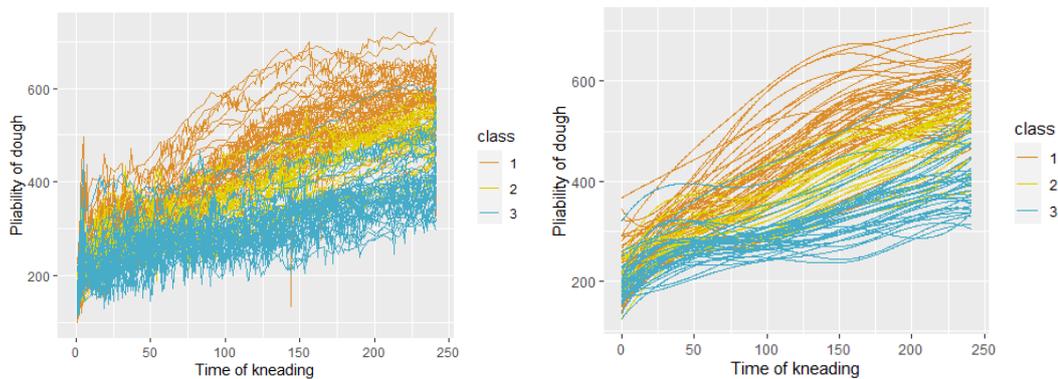


Fig. 3. Raw and smoothed kneading data

4.2. *Models in competition*

All the following models are compared on the basis of the correct classification rate (accuracy ratio) evaluated by 5-fold cross-validation.

Functional Ordinal Logistic Regression (FOLR) As previously mentioned, we have to select how many cubic B-spline basis functions we have to use. The main idea was to test

multiple values to determine the optimal number of basis functions. Different number of B-spline basis functions have been considered in a range of values from 5 (single interior knot) to 15 (11 interior knots).

The global evaluation scheme, used to compare the different models, is a 5-fold cross-validation. Here, we have implemented nested cross-validation, which consists of two cross-validations. First of all, we have divided our data set on 5 folds, in a way in which each of them contains an equal amount of data (20%). Next, at each step, we have retained one fold for testing that would leave 80% of data for training. In order to test different number of basis functions, 10-fold cross-validation is implemented on this training data set. Using this nested cross-validation scheme led to selecting six cubic B-spline basis functions.

Functional Multinomial Logistic Regression (FMLR) FMLR is the multinomial (non-ordinal) version of functional logistic regression, introduced in its binary version in Ratcliffe et al. (2002). We extended this method to more than two categories and used our own implementation in R. This implementation is based on the `multinom` function of the `nnet` package. The same nested cross-validation as for FOLR indicated that nine cubic B-spline basis functions should be selected.

LASSO-Ordinal Logistic Regression (OLR) We also used as a competitor the non functional version of ordinal logistic regression, applied directly on the raw data (241 features corresponding to the 241 time points). Due to the fact that the number of predictors was larger than the number of observations, we performed the LASSO penalised version of the ordinal logistic regression model, with the usage of the `ordinalNet` R package. The choice of the penalty parameter λ is done with the same nested cross-validation scheme as before.

Random Forest (RF) The next competitor is random forest, applied through the `caret` R package. The `mtry` parameter that determines the optimal number of variables that will be used at each random split of the decision tree, has also been selected by nested cross-validation.

Support Vector Machine (SVM) Last but not least, support vector machine was also considered. The cost parameter C is selected by nested cross-validation. The cost parameter determines the width of the margin of classification. For small values of the cost parameter C , observations inside the margin are not penalised and we obtain better fit but larger estimation error. The otherwise is true for larger values of cost parameter, where the estimation error is minimised, but the model may overfit the data.

4.3. Results

Table 1 summarises the obtained results from the nested cross-validation technique for all considered models. Standard deviations across folds are in parenthesis. The highest accuracy ratio is achieved with FOLR model, which can be expected because FMLR is

Table 1. Cross-validated Accuracy Ratio (AR) and Root Mean Square Error (RMSE) for Kneading data set (with standard deviation across folds).

Models	AR	RMSE
FOLR	0.829 (0.070)	0.49 (0.12)
FMLR	0.770 (0.087)	0.612 (0.075)
RF	0.776 (0.097)	0.578 (0.17)
OLR	0.760 (0.123)	0.617 (0.14)
SVM	0.80 (0.090)	0.537 (0.11)

not able to take into account the ordinal nature of the response, whereas RF, OLR and SVM do not consider the functional nature of the covariate.

5. Determination of the factors influencing the average power during cycling session

5.1. The data set

The data set is composed of 216 one-hour bike sessions, during which the speed, slope, altitude, heart rate of the cyclist, pedalling cadence and outdoor temperature are measured every second. These data are measured with a Garmin Edge 520. The power is also recorded every second with a powermeter ROTOR INpower ROAD. These cycling sessions were carried out during the same year (2019), by the same amateur cyclist, and combine training and competition sessions. More precisely, this dataset corresponds to the first hour of all sessions for this cyclist in 2019, after removing some sessions with missing parameters (due to a battery problem or forgotten equipment). The fact that only the first hour is considered is provided an homogeneous data set of sessions of the same duration. Indeed, to treat functional observations of different lengths remains one of the main open challenges when working with functional data, and the solution used here is one of the most common (Schmutz et al., 2020).

Data quality In this work, the data are processed as they are produced by the powermeter. None correction is considered. There is a large literature on the subject of the quality of such data, and we refer to Cardinale and Varley (2017); Halson (2014); Passfield et al. (2017); Frick and Kosmidis (2017) for a discussion on it. Note that the functional data approach which is considered in the present paper allows to remove measurement error from the data thanks to the smoothing step of the functional data reconstruction (Section 3.2).

Cutting and averaging These sessions were cut into three different lengths: 2 minutes, 10 minutes and 30 minutes. Consequently, we have 6480 sessions of 2 minutes, 1296 sessions of 10 minutes and 432 sessions of 30 minutes. The power is averaged during the session and ranked into the seven power zones. Table 2 describes the distribution of the sample over the power zone in function of the session duration. The power zone 1 is the zone of lowest intensity, in which the cyclist can stay a long time. The power zone 7 is

Table 2. Distribution over the power zones in function of the session duration

power zone	1	2	3	4	5	6	7
2'	1618	1431	927	631	493	785	595
10'	163	404	246	159	115	175	34
30'	27	148	117	46	31	59	4

the zone of highest intensity, in which the cyclist can stay for only a few seconds. Figure 4 plots a sample of data, corresponding to a 30-minute session in power zone 1 (lowest one), 3, 5 and 7 (highest one).

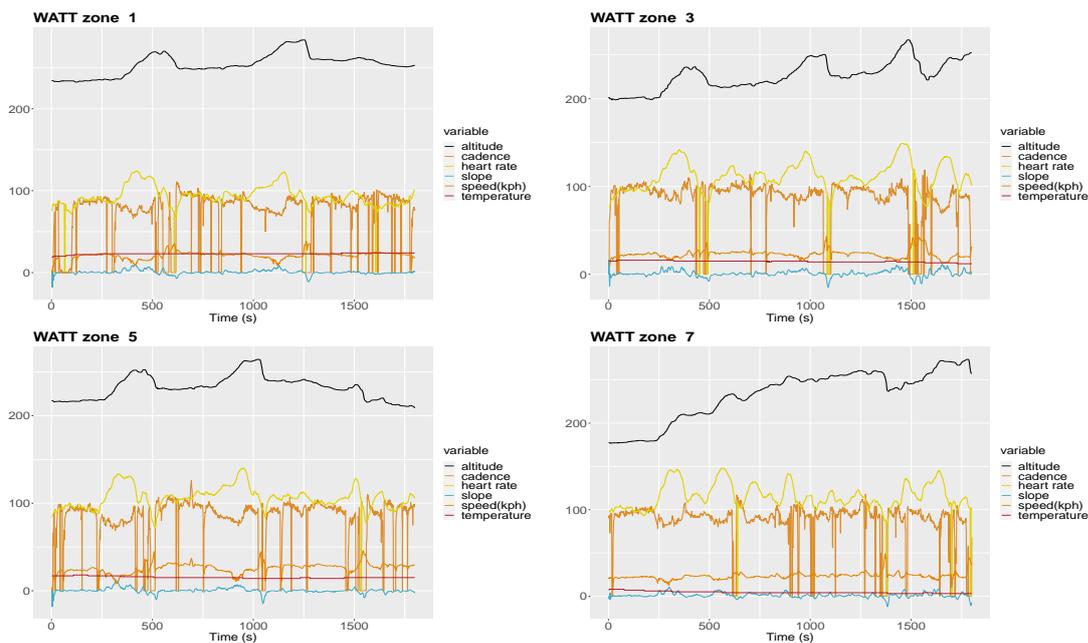


Fig. 4. Cycling data for 30-minute bike session in power zone 1 (top left), 3, 5 and 7 (bottom right)

According to Passfield et al. (2017), using the mean power output for summarising a cycling session is attractive, since analysing a single number is easy, but suffer from the drawback of providing no information about the evolution of the power during the session. In order to keep the simplicity of summarising a session with a unique power mean value and to take into account the variability of the power during the session, a cutting strategy has been employed in order to create shorter sessions, as described above. Other refined approaches would certainly have allowed less information to be lost, and could be the subject of future works. For instance, by working with the time spent within the different power zones (Passfield et al., 2017). It would requires a multivariate output regression model, in which the output are linearly dependent, which is challenging from a statistical modelling point of view, above all when covariates are functional. Another way to answer to this problem is to consider the functional nature

of the power as for the other parameters, and to model it with a functional regression model (Ramsay and Silverman, 2005). But it is possible that the interpretation of the modeling results would be more difficult due to the complexity of the functional nature of the output.

Independence assumption In this work, all the sessions, whatever are their length, are assumed to be independent. This assumption is requested for performing maximum likelihood estimation (Section 3.3) of the FOLR model. Since the sessions have been carried out by the same cyclist, and moreover that the 2-minute, 10-minute and 30-minute sessions are obtained by cutting one hour-session, such assumption should be discussed. From a cycling point of view, this independence assumption means that the level of performance is assumed to be constant throughout the year and also throughout the session. More precisely, this is the link between the mean power and the covariates (speed, slope, altitude, heart rate of the cyclist, pedalling cadence and outdoor temperature) which are assumed to be constant throughout the session and the year. Even if the independence assumption throughout the year may seem reasonable since the data have been provided by the same individual, who has been cycling for a long time, this is more discutable for the independence throughout a session, especially for short session. Indeed, a 2-minute session with a mean power in zone 7 should probably have an impact of the following 2-minute session. Nevertheless, regarding the whole set 6480 2-minute sessions together make this independence assumption not so strong due to the diversity of the sessions throughout the year. At any rate, it would be useful to compare the results of the present paper with those obtained by models taking into account a dependence within data. But such models remain to be developed. Let finally mention that if these data came from a panel of different cyclists, the individual difference between cyclists would strongly violate the independence assumption, and the proposed model should be adapted by introducing a cyclist effect in the model. This extension should not be too complicated, which is not necessarily the case with a model taking into account a dependency between the sessions themselves.

Functional data reconstruction Functional data reconstruction is performed with cubic spline basis with 20 basis functions. This choice was made empirically, so that the main variations of the curves are taken into account. For simplicity, the same number of basis functions is used for every session duration, although we could have used a smaller number for the shorter sessions. Similarly, the same basis is used for all covariates, but it is possible to choose different bases. It should be noted that for a given model, better predictive results could probably be obtained by selecting the basis by cross-validation. However, in the present study, different models using different sets of features will be compared, and we want all of them to use the same basis for better interpretability.

5.2. *Experimental setting*

The goal is to model the mean power zone of the bike session according to the six functional covariates: speed, slope, altitude, heart rate of the cyclist, pedalling cadence and outdoor temperature. Since these covariates are not necessarily relevant for modelling

Table 3. Cross-validated Root Mean Square Error (RMSE) on the cycling data set.

Models	2-minute	10-minute	30-minute
FOLR	0.728 (0.024)	0.650 (0.036)	0.630 (0.091)
FMLR	1.470 (0.048)	1.296 (0.084)	1.298 (0.284)
RF	0.752 (0.029)	0.714 (0.033)	0.650 (0.062)
OLR	0.737 (0.040)	-	-
SVM	0.738 (0.026)	0.691 (0.070)	0.633 (0.108)

the power zone, a variable selection is performed. For this, all the possible subsets of covariates are considered and evaluated by 10-fold cross-validation. Even if this strategy is often avoided because of the exponential combinatorics of the number of subgroups of variables, it is quite feasible here because six variables imply 63 subgroups. Due to the ordinal nature of the power zone, the models are evaluated by the Root Mean Square Error (RMSE), as suggested in Gaudette and Japkowicz (2009). This criterion is preferred to the classification accuracy, which does not take into account the proximity between two ordinal categories.

Model estimation is implemented in the R package **FRM**. This package provides several models for functional predictors (linear regression, logistic regression FMLR and the FOLR model), as well as the cycling data set. The **FRM** package is available from the authors upon request and will be submitted to CRAN after publication of the present paper. With this package, one FRM model estimation is about 10 seconds on a 3.5 GHz Intel Core i7 processor with 16 Go of memory.

5.3. Results

Comparison with competitors Table 3 show the cross-validated RMSE on the cycling data set of FOLR and its competitors. Let notice that for OLR the heavy computation time allow to obtain the results only for the 2-minute data set. These results confirm that the proposed FOLR method over-performs the competitors on this dataset again. It is not surprising since it is the only method which take into account both the ordinal nature of the output and the functional nature of the covariate.

Variables selection In the proposed approach, the discriminant variables are selected through a model selection approach. There is a large volume of literature on model selection, and a recurring question is whether we should choose the best model or a set of good models. In order to illustrate this question, Figure 5 plots the cross-validated RMSE value for the 63 models (ordered by increasing RMSE), for the three bike session durations (2, 10 and 30 minutes). On this Figure, the lower the RMSE, the better the model. On the one hand, we can see that there is not one model which is clearly better than the others. On the other hand, we notice that a group of model stands out slightly from the others. We arbitrarily select this set of models, denoted in the following as *good models*, by stopping at the first break in the RMSE values (red vertical lines). This leads to the selection of 24 models for 2 and 10-minute bike session, and 26 models for 30-minute length.

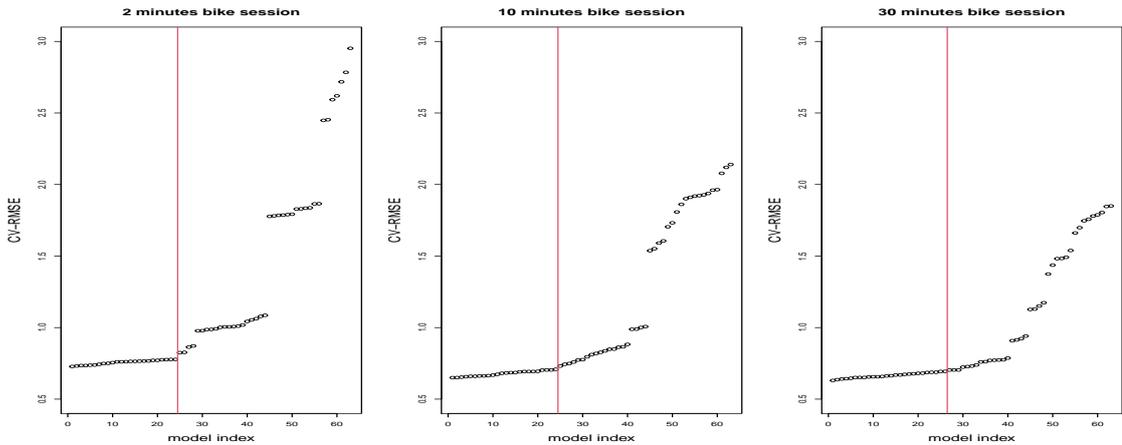


Fig. 5. Cross-validated RMSE values for the 63 models and for the three bike session durations

In order to estimate the importance of each variable in this set of good models, we give a score to each variable occurring in each model. This score depends on the ranking of the model. For instance, for 2-minute bike sessions, variables occurring in the best model obtain a score of 24 (the number of good models), those in the second-best model a score of 23, and so on. Then, the score of the variables is summed over all the good models. Figure 6 plots the resulting variables importance, according to the bike session duration. On this figure, the more outward the indicator, the more important the variable.

Prediction accuracy Table 4 presents the cross-validated prediction results on whole data set. These results are obtained by the best model according to CV-RMSE, for each bike session duration. For 2-minute sessions, the best model uses all variables except the altitude, and has a cross-validated RMSE of 0.728. For 10-minute sessions, the best model uses the heart rate, cadence, slope and speed and has a cross-validated RMSE of 0.650. For 30-minute sessions, the best model uses the heart rate, cadence and slope, and has a cross-validated RMSE of 0.630.

Model coefficients visualization An advantage of the proposed FOLR model, and especially the fact that it respects the functional nature of the covariates, is the possibility to visualize the model coefficients. For instance, Figure 7 plots the regression coefficients for the best model on the 30-minute sessions data set. Even if the coefficients fluctuations are difficult to interpret since the model is learned on a large data set, containing heterogeneous bike sessions, we constate that the effect is mainly positive, what means (according to model coefficients effect explained in Section 3.1) that the higher are these features (heart rate, cadence and slope), the highest is the mean power zone.

Table 4. True power zones versus cross-validated predicted ones by the best model (according to CV-RMSE), for 2-minute (top), 10-minute (middle) and 30-minute (bottom) bike sessions. In bold are the correct predictions.

pred \ true	1	2	3	4	5	6	7
1	1260	250	19	1	1	0	0
2	324	939	329	40	1	0	0
3	33	207	412	229	38	3	0
4	1	32	129	224	149	32	0
5	0	3	34	94	142	99	4
6	0	0	4	43	161	530	195
7	0	0	0	0	1	121	396

pred \ true	1	2	3	4	5	6	7
1	113	25	1	0	0	0	0
2	50	304	92	6	0	0	0
3	0	68	120	50	1	0	0
4	0	7	31	80	22	3	0
5	0	0	2	21	55	31	0
6	0	0	0	2	37	135	21
7	0	0	0	0	0	6	13

pred \ true	1	2	3	4	5	6	7
1	19	7	0	0	0	0	0
2	8	108	34	3	0	0	0
3	0	27	76	17	0	0	0
4	0	6	7	22	4	1	0
5	0	0	0	3	17	8	0
6	0	0	0	1	10	49	4
7	0	0	0	0	0	1	0

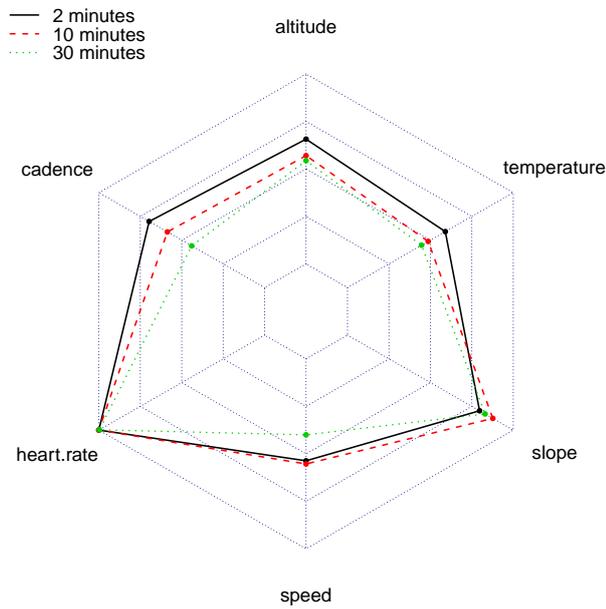


Fig. 6. Variable importance according to the bike session duration

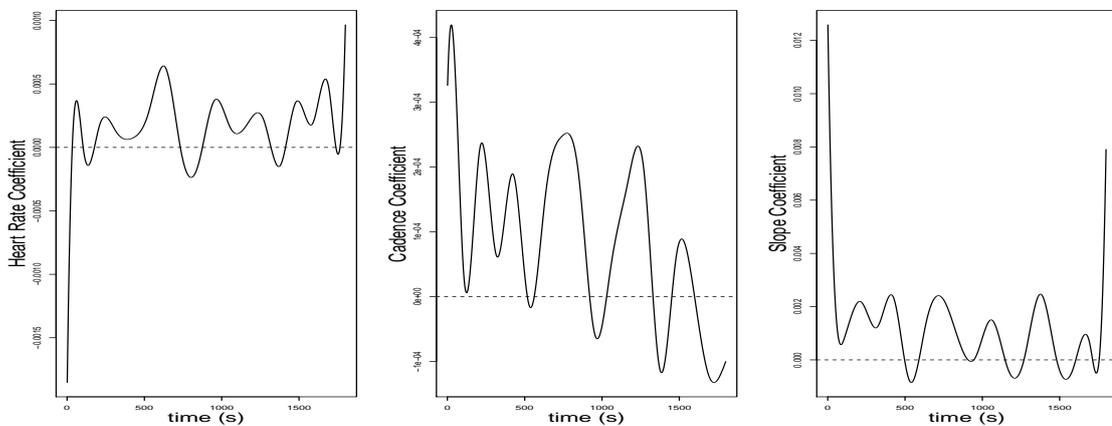


Fig. 7. Functional coefficients for the best 30-minute sessions model

5.4. Analysis

First of all, we can note that the quality of the prediction is relatively correct, with a majority of elements on the diagonal of the confusion matrices (bold numbers in Table 4). Even when a power zone is wrongly predicted, the error is small since this is generally

a contiguous power zone which is being predicted (numbers on the subdiagonal or the superdiagonal). This means that the studied variables reflect the average developed power, i.e. the cyclist's performance. Then, we can notice that the quality of the prediction increases with the duration of the exercise, as RMSE decreases. It seems easier to predict the average power of a longer effort, which is subject to fewer irregularities than the shorter efforts.

We now initiate here an attempt to analyze the results, based on the postulate that the variables present in the best models are those on which it is necessary to act in order to optimize the power. Sport science specialists could probably carry out a more detailed analysis of these results, in particular by interpreting the values of the functional coefficients (as those plotted on Figure 7). Regarding the importance of the variables, Figure 6 shows that all variables are important. The most important variable, regardless of the duration of the exercise, is the cyclist's heart rate: it is selected in all the good models. This is not surprising, because developing significant power requires significant physical effort. And this confirms that the heart rate, when power was not easily measurable during an exercise, was a good indicator of performance, or at least of the developed power. Then comes the slope, with no significant difference of importance in function of the session duration. The third variable in order of importance is pedalling cadence. This variable is of particular interest because it is the one on which the cyclist can act, unlike the slope, the temperature and the altitude which are external parameter on which the cyclist cannot act. It is very interesting to note that the pedalling cadence is the variable for which the difference in importance according to the length of the effort is the most obvious. Moreover, the importance of the pedalling cadence decreases significantly with the length of the effort. One possible interpretation can be as follows: if the amateur cyclist studied is able to hold high cadences to maintain high power during short efforts, it becomes much more complicated during longer efforts. Maintaining high power using high cadence over a long period of time is the prerogative of high-level cyclists, such as the professional cyclists mentioned in the introduction. Finally, the least important variable is speed. Even if this might seem surprising at first glance, it is because it is closely linked to the nature of the terrain: we will develop significant powers to climb steep climbs, without going very fast; and on the contrary, we will go very quickly downhill at very low or even null power when the cyclist is not pedalling.

6. Conclusion

This work proposes a study of the data that are commonly produced by cyclists during the practice of their sport. In particular, this study is interested in the factors that makes it possible to discriminate the average power developed during the effort. One of the most interesting conclusions from a sports practice point of view is that the pedalling cadence is indeed a lever for optimising the developed power. Nevertheless, this study shows that for the amateur cyclist who provided these data, maintaining a high cadence over a long time is difficult. This maintenance of a high cadence over a long time is one of the typical characteristics of high-level cyclists and is a factor on which the amateur cyclist can work to increase their performance. From a statistical modelling point of view, this study has needed the development of an ordinal logistic regression model

with functional predictors (FOLR). Experimental study on a benchmark has shown the efficiency of this model in comparison to competitors, which either omit the ordinal nature of the response or the functional nature of the covariate.

The perspectives from a cycling data point of view are numerous, and some of them are described in the sequel. The first one would be to complete the data set with data from other cyclists, of various profiles and levels, in order to build a more heterogeneous database and thus to be able to draw more general conclusions. However, that will require adapting the FOLR model to take into account an individual effect, or at least to incorporate the effect of the age, the weight, etc. This requires developing a new cumulative logit model with functional and non functional covariate. Similarly, it would also be interesting to be able to take into account the period of the year at which the data is measured. Cyclists have a level that evolves throughout the year and it would be nice to be able to take that into account. However, that would again require the development of a new model. Even more finely, it could be interesting to model the dependence between sessions that are close in time, whether it is between two sessions spaced a day apart or even between different times of the same session. Another aspect concerns the bike session durations, which have been homogenized in cutting them after one hour. This necessarily generates a loss of information and a bias in the analysis. A challenging perspective would be to be able to take into account functional observations of different lengths, but it currently remains an open question. Analogously, averaging the power during a bike session, lead to loose information about the variation of the power during this session. Modelling the power as a functional data using a functional regression model (Ramsay and Silverman, 2005), or modelling the time spent within the different power zones, could provide a complementary analysis of the present work.

References

- Abbiss, C., Peiffer, J. and Laursen, P. (2009) Optimal cadence selection during cycling. *ECU Publications*, **10**.
- Agresti, A. (2010) *Analysis of ordinal categorical data*. Wiley Series in Probability and Statistics. Wiley-Interscience, New York.
- Allen, H. and A., C. (2010) *Training and Racing With a Power Meter*.
- Beattie, K., Carson, B., Lyons, M. and Kenny, I. (2016) The effect of maximal and explosive strength training on performance indicators in cyclists. *International Journal of Sports Physiology and Performance*, **12**, 1–25.
- Bini, R., Diefenthaler, F. and Carpes, F. (2014) Determining force and power in cycling: A review of methods and instruments for pedal force and crank torque measurements. *International Sportmed Journal*, **15**, 96–112.
- Borszcz, F., Tramontin, A., Bossi, A., Carminatti, L. and Costa, V. (2018) Functional threshold power in cyclists: Validity of the concept and physiological responses. *International Journal of Sports Medicine*, **39**.

- Cardinale, M. and Varley, M. (2017) Wearable training-monitoring technology: applications, challenges, and opportunities. *International journal of sports physiology and performance*, **12**, 2–55.
- Cardoso, J. and Pinto da Costa, J. (2007) Learning to classify ordinal data: The data replication method. *Journal of Machine Learning Research*, **8**, 1393–1429.
- Chu, W. and Keerthi, S. (2007) Support vector ordinal regression. *Neural Computation*, **19**, 792–815.
- van Dijk, H., van Meegen, R. and Vroemen, G. (2017) *The Secret of Cycling: Maximum Performance Gains Through Effective Power Metering and Training Analysis*. Meyer & Meyer Sport.
- Escabias, M., Aguilera, A. and Valderrama, M. (2005) Modeling environmental data by functional principal component logistic regression. *Environmetrics*, **16**, 95–107.
- Faria, E., Parker, D. and Faria, I. (2005) The science of cycling: Factors affecting performance ??? part 2. *Sports medicine (Auckland, N.Z.)*, **35**, 313–37.
- Ferraty, F. and Vieu, P. (2006) *Nonparametric functional data analysis*. Springer Series in Statistics. New York: Springer.
- Frick, H. and Kosmidis, I. (2017) tracker: Infrastructure for running and cycling data from gps-enabled tracking devices in r. *Journal of Statistical Software*, **82**, 1–29.
- Gaudette, L. and Japkowicz, N. (2009) Evaluation methods for ordinal classification. In *Proceedings of the 22nd Canadian Conference on Artificial Intelligence: Advances in Artificial Intelligence*, Canadian AI '09, 207–210. Berlin, Heidelberg: Springer-Verlag.
- Grappe, F. (2012) *Puissance et performance en cyclisme*. De Boeck.
- (2018) *Cyclisme et optimisation de la performance*. De Boeck.
- Halson, S. (2014) Monitoring training load to understand fatigue in athletes. *Sports medicine*, **44**, 139–147.
- Jacques, J. and Biernacki, C. (2018) Model-based co-clustering for ordinal data. *Computational Statistics and Data Analysis*, **123**, 101–115. URL: <http://www.sciencedirect.com/science/article/pii/S0167947318300239>.
- Jacques, J. and Preda, C. (2014) Model based clustering for multivariate functional data. *Computational Statistics and Data Analysis*, **71**, 92–106.
- Lévêder, C., Abraham, P., Cornillon, E., Matzner-Lober, E. and Molinari, N. (2004) Discrimination de courbes de p̄içœtrissage. In *Chimioviçœtrie 2004*, 37–43. Paris.
- McParland, D. and Gormley, C. (2013) *Algorithms from and for Nature and Life: Studies in Classification, Data Analysis, and Knowledge Organization*, chap. Clustering Ordinal Data via Latent Variable Models, 127–135. Switzerland: Springer.

- Nimmerichter, A., Eston, R., Bachl, N. and Williams, C. (2011) Longitudinal monitoring of power output and heart rate profiles in elite cyclists. *Journal of sports sciences*, **29**, 831–40.
- Osborne, M. R. (1992) Fisher’s method of scoring. *International Statistical Review*, **60**, 99–117.
- Passfield, L., Hopker, J., Jobson, S. A., Friel, D. and Zabala, M. (2017) Knowledge is power: Issues of measuring training and performance in cycling. *Journal of sports sciences*, **14**, 1426–1434.
- Preda, C., Saporta, G. and Lévêder, C. (2007) PLS classification of functional data. *Computational Statistics*, **22**, 223–235.
- Ramsay, J. O. and Silverman, B. W. (2005) *Functional data analysis*. Springer Series in Statistics. New York: Springer, second edn.
- Ratcliffe, S. J., Heller, G. Z. and Leader, L. R. (2002) Functional data analysis with application to periodically stimulated foetal heart rate data. ii: Functional logistic regression. *Statistics in medicine*, **21**, 1115–1127.
- Schmutz, A., Chèze, L., Martin, P. and Jacques, J. (2020) A method to estimate horse speed per stride from one imu with machine learning method. *Sensors*, 518.

A. Fisher Information Matrix

Let remind that $g(t) = 1/(1+\exp(-t))$, $h(t) = \exp(-t)/(1+\exp(-t))^2$ the first derivative of $g(t)$ and let introduce $k(t) = \frac{-\exp(-t)}{(1+\exp(-t))^2}$ is the second derivative of $g(t)$.

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \mathbf{b}^2} = & \sum_{i=1}^n \left(1_{y_{i=1}} \Psi^2 \mathbf{a}_i^2 \frac{k(\alpha_1 - \mathbf{b}'\Psi \mathbf{a}_i)g(\alpha_1 - \mathbf{b}'\Psi \mathbf{a}_i) - h(\alpha_1 - \mathbf{b}'\Psi \mathbf{a}_i)^2}{g(\alpha_1 - \mathbf{b}'\Psi \mathbf{a}_i)^2} \right. \\ & + \left. 1_{y_{i=C}} \Psi^2 \mathbf{a}_i^2 \frac{-k(\alpha_{C-1} - \mathbf{b}'\Psi \mathbf{a}_i)(1 - g(\alpha_{C-1} - \mathbf{b}'\Psi \mathbf{a}_i)) - h(\alpha_{C-1} - \mathbf{b}'\Psi \mathbf{a}_i)^2}{(1 - g(\alpha_{C-1} - \mathbf{b}'\Psi \mathbf{a}_i))^2} \right) \\ & + \sum_{i=1}^n \sum_{c=2}^{C-1} 1_{y_{i=c}} \Psi^2 \mathbf{a}_i^2 \frac{(k(\alpha_c - \mathbf{b}'\Psi \mathbf{a}_i) - k(\alpha_{c-1} - \mathbf{b}'\Psi \mathbf{a}_i))(g(\alpha_c - \mathbf{b}'\Psi \mathbf{a}_i) - g(\alpha_{c-1} - \mathbf{b}'\Psi \mathbf{a}_i))}{(g(\alpha_c - \mathbf{b}'\Psi \mathbf{a}_i) - g(\alpha_{c-1} - \mathbf{b}'\Psi \mathbf{a}_i))^2} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \mathbf{b} \partial \alpha_c} = & - \sum_{i=1}^n \left(1_{y_{i=1}} \Psi \mathbf{a}_i \frac{k(\alpha_1 - \mathbf{b}'\Psi \mathbf{a}_i)g(\alpha_1 - \mathbf{b}'\Psi \mathbf{a}_i) - h(\alpha_1 - \mathbf{b}'\Psi \mathbf{a}_i)^2}{g(\alpha_1 - \mathbf{b}'\Psi \mathbf{a}_i)^2} \right. \\ & - \left. 1_{y_{i=C}} \Psi \mathbf{a}_i \frac{-k(\alpha_{C-1} - \mathbf{b}'\Psi \mathbf{a}_i)(1 - g(\alpha_{C-1} - \mathbf{b}'\Psi \mathbf{a}_i)) - h(\alpha_{C-1} - \mathbf{b}'\Psi \mathbf{a}_i)^2}{(1 - g(\alpha_{C-1} - \mathbf{b}'\Psi \mathbf{a}_i))^2} \right) \\ & - \sum_{i=1}^n \sum_{c=2}^{C-1} 1_{y_{i=c}} \Psi \mathbf{a}_i \frac{(k(\alpha_c - \mathbf{b}'\Psi \mathbf{a}_i) - k(\alpha_{c-1} - \mathbf{b}'\Psi \mathbf{a}_i))(g(\alpha_c - \mathbf{b}'\Psi \mathbf{a}_i) - g(\alpha_{c-1} - \mathbf{b}'\Psi \mathbf{a}_i))}{(g(\alpha_c - \mathbf{b}'\Psi \mathbf{a}_i) - g(\alpha_{c-1} - \mathbf{b}'\Psi \mathbf{a}_i))^2} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \alpha_c^2} = & \sum_{i=1}^n \left(1_{y_{i=1}} \frac{k(\alpha_1 - \mathbf{b}'\Psi \mathbf{a}_i)g(\alpha_1 - \mathbf{b}'\Psi \mathbf{a}_i) - h(\alpha_1 - \mathbf{b}'\Psi \mathbf{a}_i)^2}{g(\alpha_1 - \mathbf{b}'\Psi \mathbf{a}_i)^2} \right. \\ & + \left. 1_{y_{i=C}} \frac{-k(\alpha_{C-1} - \mathbf{b}'\Psi \mathbf{a}_i)(1 - g(\alpha_{C-1} - \mathbf{b}'\Psi \mathbf{a}_i)) - h(\alpha_{C-1} - \mathbf{b}'\Psi \mathbf{a}_i)^2}{(1 - g(\alpha_{C-1} - \mathbf{b}'\Psi \mathbf{a}_i))^2} \right) \\ & + \sum_{i=1}^n \sum_{c=2}^{C-1} 1_{y_{i=c}} \frac{(k(\alpha_c - \mathbf{b}'\Psi \mathbf{a}_i) - k(\alpha_{c-1} - \mathbf{b}'\Psi \mathbf{a}_i))(g(\alpha_c - \mathbf{b}'\Psi \mathbf{a}_i) - g(\alpha_{c-1} - \mathbf{b}'\Psi \mathbf{a}_i))}{(g(\alpha_c - \mathbf{b}'\Psi \mathbf{a}_i) - g(\alpha_{c-1} - \mathbf{b}'\Psi \mathbf{a}_i))^2} \end{aligned}$$