A functional analysis of speed profiles: Smoothing using derivative information, registration, and speed corridors

Cindie Andrieu, Guillaume Saint Pierre, Xavier Bressaud

A functional analysis of speed profiles: Smoothing using derivative information, registration, and speed corridors

C. Andrieu[a,1,*], G. Saint Pierre[a], X. Bressaud[b]

[a] *French Institute of Science and Technology for Transport, Development and Networks, Laboratory for Vehicle Infrastructure Driver Interactions (IFSTTAR/LIVIC), 77, rue des Chantiers, 78000 Versailles, France*
[b] *Université Paul Sabatier, Institut de Mathématiques de Toulouse, F-31062 Toulouse Cedex 9, France*

Abstract

The knowledge of the actual vehicle speeds on the road network is a very informative component of drivers behavior and their road usage. This information become available with the generalization of probe vehicles and the development of GPS-equipped smartphones which has increased the number of "tracers" likely to refer their position and speed data in real time. In addition, advances in sensors technology enable to collect these data with high sampling rate which leads to large volume of individual space-speed profiles (i.e. speed as function of space) with increasingly fine grids measurements, which requires appropriate methods. Indeed, if in practice, data collected are discretized, and then treated as vectors, the classical multivariate statistical methods become inadequate to analyze such high dimensional vectors. The originality of the approach presented in the current paper is to propose a functional analysis of speed profiles, i.e. to treat these objects as functions rather than vectors. This approach takes inspiration from Functional Data Analysis, a statistical domain that has developed recently, and allows to preserve the functional nature of such data that initially derive from a continuous underlying process. This functional analysis begins with a functional modeling of space-speed profiles and the study of mathematical properties of these functions. Then, a smoothing procedure based on spline smoothing is developed in order to convert the raw data into functional objets and to filter out the measurement noise as efficiently as possible. It is shown that this smoothing step leads to a complex nonparametric regression problem that needs to take into account two constraints: the use of the derivative information, and a monotonicity constraint. This paper also presents a methodology to build an aggregated speed profile (average or median), and also speed corridors in order to summarize the information contained in a large set of individual space-speed profiles. This methodology is illustrated on a real data set and is divided in the three following steps: the smoothing step, the registration of speed profiles by the method of landmarks alignment, and the construction of speed corridors by an extension of the classical boxplots to functional data.

Keywords:
Speed profiles, Functional Data Analysis, Smoothing spline, Landmarks alignment, Speed corridors, Probe vehicles

1. Introduction

To operate the road network, it is necessary to take into account various criteria such as speed, safety, comfort, flexibility, energy savings... It is then important to best optimize the use of network in order to increase efficiency, which requires a detailed knowledge of its actual use. Thus, it becomes essential to have efficient and reliable methods to measure and evaluate the actual use of the network that take into account the complex links between these criteria.

The tools used to observe and measure the use of the road network evolved recently. They shifted from the collection of aggregated data obtained through static devices (e.g. magnetic loops, radars) to much finer measurements

*Corresponding author, URL: https://sites.google.com/site/andrieucindie/
Email addresses: cindie.andrieu@ifsttar.fr (C. Andrieu), guillaume.saintpierre@ifsttar.fr (G. Saint Pierre), bressaud@math.univ-toulouse.fr (X. Bressaud)
[1] Affiliation during redaction of the article.

obtained by data recording equipments installed in vehicles, which have contributed to the probe vehicles concept definition. Until now, probe vehicles, that can be seen as mobile sensors exploring continuously the road network, were reduced to private fleets (e.g. buses, taxis). But the development of smartphones has increased the number of digital "traces" left by vehicles, and allows to study the road user's behavior on the complete network. However, if probe vehicles can allow to collect a lot of information about the vehicle parameters or the driver actions, the association of position and speed measurements is already very informative about traffic and driver behavior, and it is easily available with a GPS.

The knowledge of the actual vehicle speeds on roads is essential from several points of view: to locate blackspot in the network, to improve the knowledge of travel time and to evaluate the effects of the modification of the infrastructure (addition of speed bumps, roundabouts, ...). The speed choice of drivers is one of the most important components of their behavior and also their road usage (Ericsson (2000), Laureshyn (2005)). This continuous information of road user's speed, available with the development of smartphones, leads to the obtention of individual speed profiles that can be analyzed as functions of time (time-speed profiles) or functions of space (space-speed profiles) depending on the purpose of the study. Speed profiles allow a more detailed analysis of driver behavior and variations between road users, contrary to aggregated indicators (such as average or median speed), and their study appears in many applications: travel time forecasting (Quiroga and Bullock (1998)), impact of the modification of the infrastructure (Barbosa et al. (2000)), identification of safety critical events (Boonsiripant (2009), Nygard (1999)), effects of Intelligent Speed Adaptation (ISA) systems (Várhelyi et al. (2004))...

The collection of individual space-speed profiles can leads to large volume of data that require the use of appropriate methods. Indeed, since in practice space-speed profiles are composed of time-stamped measurements of speed and position, most studies consider them as $\mathbb{R}^n$ vectors where n is the number of measurements. However, advances in sensors technology enable to collect data with high sampling rate that leads to high dimensional vectors (n is very large), for which classical multivariate statistical methods become inadequate because of problems related to the so-called "curse of dimensionality" and the significant correlation between close observations. A solution is to preserve the functional nature of such data derived from a continuous underlying process and then to view them as functions instead of vectors. This approach is based on the "Functional Data Analysis" (FDA), a statistical domain that has developed considerably over the last twenty years and that appears in several domains such as meteorology, chemometrics, economics... An overview of the theory of statistics with functional data can be found in the reference books of Ramsay (Ramsay and Silverman, 2005, 2002), the book of Ferraty and Vieu (2006), and the papers of Levitin et al. (2007) and Cuevas (2013). The functional approach is particularly suitable for the analysis of speed profiles since it allows to preserve the physical consistency between speed and position (and implicitly time), and their functional characteristics: computation of derivatives (that leads to acceleration or jerk profiles), regularity, shape constraints... In addition, most of the classical multivariate statistical methods can be adapted to functional data (linear models, principal component analysis, canonical and discriminant analysis...), while some methods are specific to functional data such as curve registration.

In this paper, we propose a methodology suitable for the functional analysis of space-speed profiles. This paper is structured as follows. Section 2 presents a functional modeling of space-speed profiles with a definition of the corresponding functional space and the study of some mathematical properties (continuity and differentiability). In Section 3, we propose a smoothing method using spline functions in order to convert the raw data into functional objets and to filter out the measurement noise as efficiently as possible. A validation of the smoothing method is then proposed using real-world GPS data. Section 4 describes a methodology of construction of an aggregated speed profile (average and median) and also speed corridors. Three steps of this methodology are detailed and illustrated with a new data set: the construction of individual speed profiles from raw data by the use of the smoothing procedure, the registration of speed profiles by the method of landmarks alignment in order to obtain an aggregated profile representative of the set of individual speed profiles, and the construction of speed corridors by an extension of the classical boxplots to functional data. Finally, Section 5 provides a discussion and the main conclusions of the present study.

2. Functional modeling of space-speed profiles

2.1. Definition of space-speed profiles

Before beginning a functional analysis of speed profiles, it is necessary to define the functional space of such objects. Indeed, any function $f : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ is not a space-speed profile (e.g. a constant function equal to zero). In practice, a space-speed profile is a sequence of time-stamped measurements of position (from GPS or odometer) and speed, so it can be studied in the three following study areas : `distance` $\times$ `time`, `speed` $\times$ `time` and `speed` $\times$ `distance` (see Fig. 1).

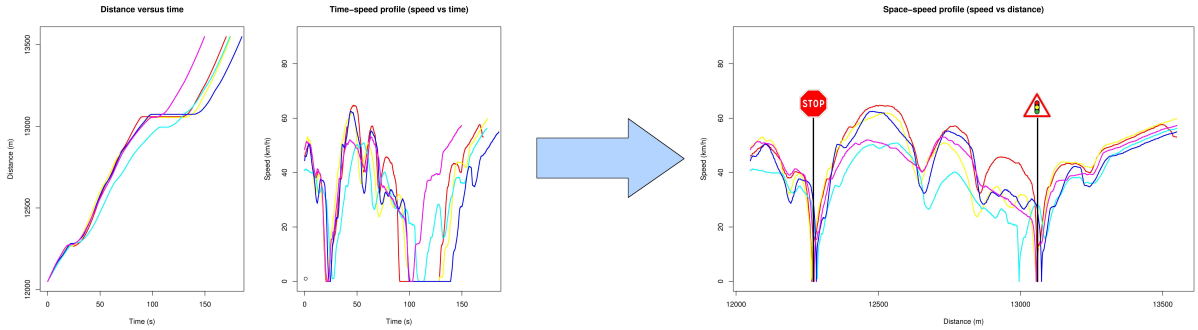


Figure 1: Link between the three study areas : [`distance` $\times$ `time`, `speed` $\times$ `time`] and [`speed` $\times$ `distance`].

The functions defined in each of these three study areas are related mathematically: If we denote $F(t)$ a function defined in the study area `distance` $\times$ `time` that represents the distance traveled as function of time, the derivative function $F'(t)$ represents the speed as function of time and is defined in the study area `speed` $\times$ `time`. So, by definition, the function F must be increasing and at least of class C^1. In order to define an acceleration profile, we propose to require that F is at least C^2, and so we propose the following definition of the functional space of space-speed profiles :

Definition 1. *Let $x_f \in \mathbb{R}^+$. Then the space of space-speed profiles, denoted $\mathcal{E}_{SSP}$, is defined as follows :*
$\mathcal{E}_{SSP} = \{v_S : [0, x_f] \longrightarrow \mathbb{R}^+$ such that there exists a positive real T and an increasing function $F : [0, T] \longrightarrow [0, x_f]$ of class C^2 with $F(0) = 0$ such that $v_S(x) = F' \circ F^{-1}(x))$, $x \in [0, x_f]\}$,
where F^{-1} is the generalized inverse of F defined by $F^{-1}(x) = inf\{t \in [0, T], F(t) = x\}$.

The positive real numbers x_f and T represent respectively the length and the travel time of the studied section. Fig. 2 illustrates Definition 1 by showing the functional link between distance, speed and implicitly time.

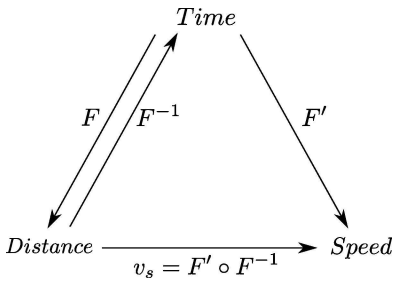


Figure 2: Functional diagram illustrating the definition of space-speed profiles.

2.2. Mathematical properties of space-speed profiles

We studied some properties of the space-speed profiles, i.e. functions in the space $\mathcal{E}_{SSP}$ (as defined in Definition 1). The continuity property is given by the following theorem whose proof is deferred to Appendix A:

Theorem 1. *All functions* $v_S : [0, x_f] \longrightarrow \mathbb{R}^+$ *belonging to the space of space-speed profiles* $\mathcal{E}_{SSP}$ *(as defined in Definition 1) are continuous on* $[0, x_f]$.

If the continuity property of space-speed profiles is obvious, the differentiability property is less intuitive as shown in the following theorem whose proof is also deferred to Appendix A:

Theorem 2. *Assume that* $v_S : [0, x_f] \longrightarrow \mathbb{R}^+$ *belongs to the space of space-speed profiles* $\mathcal{E}_{SSP}$ *(as defined in Definition 1). Let* $H_0 = \{x \in [0, x_f], v_S(x) = 0\}$*, all points for which the speed is zero. The two following assumptions are added:*

- (H_1) *Assume that* F *is of class* C^2 *on* $[0, T]$ *and strictly increasing, and* $\exists t_0 \in]0, T[$ *such that* $F'(t_0) = 0$ *and* $F'''(t_0)$ *exists with* $F'''(t_0) \neq 0$.
- (H_2) *Assume that* F *is of class* C^2 *on* $[0, T]$ *and increasing,* $\exists t_0, t_1 \in]0, T[$*,* $t_0 \neq t_1$ *such that* $F'(t) = 0$ *on* $[t_0, t_1]$*, and the function* G *defined on* $[0, T - (t_1 - t_0)]$ *by:*

$$\begin{cases} \text{for } t \leq t_0, \;\; G(t) = F(t), \\ \text{for } t \geq t_0, \;\; G(t) = F(t + t_1 - t_0), \end{cases}$$

  satisfies the assumptions (H_1).

If F *satisfies the assumptions* (H_1) *or* (H_2)*, then* $v_S = F' \circ F^{-1}$ *is not differentiable on* H_0.

The assumptions (H_1) and (H_2) are not restrictive and are satisfied in most cases. So, this theorem shows that space-speed profiles are not differentiable at points for which the speed is zero, i.e. when the vehicle is stopped. From a geometrical point of view, if we assume that $x_0 \in H_0$ (i.e. $v_S(x_0) = 0$), it is easily shown that the graph of a space-speed profile v_S has a half-tangent parallel to the axis of y in x_0, i.e. a cusp at the point $(x_0, 0)$.

This property of non differentiability at points for which the speed is zero, implies some difficulties in the calculation of an average profile, particularly in the case of stops. Indeed, if a space-speed profile v_1 is equal to zero at a point x_0, and a space-speed profile v_2 is strictly positive at x_0, then the sum $v_1 + v_2$ is not a space-speed profile as defined in the Definition 1 since $v_1 + v_2$ is not differentiable at x_0 but $(v_1 + v_2)(x_0) > 0$. Thus, the calculation of an average profile is meaningful only in this two cases:

- when all space-speed profiles are strictly positive (no stops) ;
- when all space-speed profiles are equal to zero at the same points (i.e. all vehicles stop at the same location).

This second case raises the issue of registration of speed profiles that will be discussed in Section 4.3.

3. Estimation of a space-speed profile from noisy data using smoothing splines

The data accuracy is very important to analyse speed profiles. However, sensors can provide inaccurate measurements and it is necessary to use a smoothing procedure in order to reduce the errors that affect the raw data. Several studies proposed the use of various smoothing methods to estimate speed or acceleration profiles. For example, Rakha et al. (2001) and more recently Jun et al. (2006) propose a comparaison of different smoothing methods such as the kernel smoothing method or the Kalman filter. The authors show that most of these methods are not robust to the presence of outliers and are not suitable for the calculation of derivatives. Indeed, these discrete smoothing methods provide an estimation of speed profiles at sampling points, and then the estimation of the derivatives (acceleration or jerk profiles) is performed by using numerical differentiation techniques (backward, forward or central difference) which can increase the errors depending on the choice of the method and the data sampling frequency. citeBratt1999 propose to use the local polynomial regression technique which presents the advantage to being simple but tends to overestimate or underestimate the speed and the derivative values (acceleration). In addition, this method requires the determination of many parameters (kernel, bandwidth parameter) that are not easy to choose. In this section, we propose another smoothing procedure based on the spline smoothing method which allows to keep the functional nature of speed profiles, and then to facilitate the calculation of the derivative profiles.

3.1. Smoothing using splines

3.1.1. From discrete data to smooth curves

The first step in a functional data analysis is to convert the raw data into functional objects. In our study, we look for an estimator of the "true" space-speed profile which belongs to the space $\mathcal{E}_{SSP}$ defined in Definition 1. Due to uncertainties in the raw data including speed and position measurements, there is need of use a smoothing procedure in an attempt to filter out this noise as efficiently as possible. This problem can be written as a nonparametric regression model:

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, \dots, n, \tag{1}$$

where (x_i, y_i) are given observations, ε_i are uncorrelated errors with zero mean and σ^2 variance, and f is the regression function.

A classical method for representing discrete data as a smooth function is the use of a *basis expansion*, which is to express a function as a weighted sum or linear combination of elementary functional building blocks called basis functions as follows:

$$f(x) \approx \sum_{k=1}^{K} c_k \phi_k(x), \tag{2}$$

where ϕ_k are K known basis functions, and c_k are the basis coefficients estimated from observations and which represent the weight of each basis function in the construction of the function f. There are many different types of basis function systems such as powers of x (for polynomials), Fourier basis (for periodic functions), spline basis (for nonperiodic and complex fonctions) or wavelets basis. The main advantage of the basis expansion method is to represent the function f that potentially belongs to an infinite-dimensional space within a finite-dimensional framework. Note that this method can also be used in terms of dimension reduction since the number of basis functions K is usually much lower than the number of observations n. However, the difficulty is the choice of the number K of basis functions often framed in terms of the bias-variance trade-off common in many statistical analyses. Using a high number of basis functions will over-fit the data and will produce a curve that is generally not smooth (low bias and high variance), whereas a small number of basis fails to capture interesting features of the curves (high bias and low variance).

Another way to build a smooth curve from discrete data is the use of a *roughness penalty*, also called *regularization method*. The idea is to control the smoothness of the fitting curve by adding a penalty term in the least squares criterion as follows:

$$\min_{f \in \mathcal{E}} \sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda J(f), \tag{3}$$

where $\mathcal{E}$ is a functional space, $J(f)$ is a penalty term, and $\lambda > 0$ is the smoothing parameter which controls the trade-off between the goodness of fit (measured by the first term in Eq. (3)) and the regularity of the regression function (measured by the second term in Eq. (3)). Generally, we use the penalty $J(f) = \int_a^b (f^{(m)}(x))^2 dx$ where $f \in C^m[a, b]$ as a measure of curvature of the smoothing function, and more particularly we use $J(f) = \int_a^b (f^{(2)}(x))^2 dx$, called the *roughness penalty* (Green and Silverman (1994)), which reflects the notion of average curvature of a function $f \in C^2[a, b]$. The bias-variance trade-off is directly quantify through the choice of λ: the more λ is high, the more the curve is smooth, and vice versa. In particular, when $\lambda \to 0$, the fitted curve approaches an interpolant to the data, and when $\lambda \to \infty$, the fitted curve approaches the standard linear regression to the observed data. The parameter λ can be determined by automatic methods like cross-validation (CV) or general cross-validation criterion (GCV) which are based on the minimization of a specific criterion, or by a subjective choice (trial and error method).

3.1.2. Smoothing spline

Assume that the regression function f of the model in Eq. (1) is regular, and more specifically that $f \in W^m[a, b]$, $m \in \mathbb{N}$, where $W^m[a, b]$ is the Sobolev space of order m defined by:

$$W^m[a, b] = \{f : f^{(j)} \text{ absolutely continuous } j = 0, \dots, m-1;\ f^{(m)} \in \mathcal{L}_2[a, b]\}, \tag{4}$$

where $\mathcal{L}_2[a,b]$ is the set of square integrable functions on the interval $[a,b]$. In this space, we can "measure" the regularity of a function f by $J(f) = \int_a^b (f^{(m)}(x))^2 dx$. A remarkable theorem states that if $m \leq n$, the penalized least squares criterion:

$$\sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda \int_a^b (f^{(m)}(x))^2 dx \tag{5}$$

has an explicit, finite-dimensional, unique minimizer in the Sobolev space $W^m[a,b]$, which is a natural spline of order $2m$ with knots at the design points $x_1, \ldots, x_n$ in $[a,b]$. A demonstration of this theorem can be found in Eubank (1999). We recall that if $x_1 < \ldots < x_n$ are n points of an interval $[a,b]$, then a function $f\colon [a,b] \to \mathbb{R}$ is a polynomial spline of order m (where m is an integer ≥ 1) with simple knots at the points $x_1, \ldots, x_n$, if f satisfies the following properties:

- f is a piecewise-polynomial of order m (i.e. of degree $m-1$),
- f has continuous derivatives up to order $m-2$ (if $m \geq 2$).

The term "natural" means that we add the following boundary conditions : $f^{(j)}(a) = f^{(j)}(b),\ j = m, \ldots, 2m-1$. When $m = 2$, the solution of the penalized least squares criterion is a spline of degree 3 (and order 4) called cubic spline. This smoothing method using a roughness penalty and resulting from an optimization problem is called "smoothing spline". By contrast, the "regression spline" is a method that assumes that the regression function is a spline represented as a spline basis expansion and where the basis coefficients are estimated by the minimization of the least squares criterion. However, in regression spline, the choice of the number and location of the knots is generally difficult (Hastie and Tibshirani (1990)).
The most common computational technique for smoothing spline is to represent the solution with a B-spline basis function expansion (De Boor (2001)) with knots at the sampling points :

$$\widehat{f}(x) = \sum_{j=1}^{n} \widehat{\beta}_j B_j(x) \tag{6}$$

and to minimize the criterion given in Eq. (5) with respect to the coefficients of the expansion. The solution vector $\widehat{\beta}$ of the coefficients $\widehat{\beta}_j$ is then written as follows :

$$\widehat{\beta} = (B^T B + \lambda\Omega)^{-1} B^T Y, \tag{7}$$

where $\{B\}_{ij} = B_j(x_i)$ and $\{\Omega\}_{jk} = \int B_j^{(m)}(x) B_k^{(m)}(x) dx$.

3.1.3. Generalization of the spline smoothing problem

In this section, we give a generalization of the spline smoothing problem described in the previous section in the case where the dimension of the study area is greater than 1. We will show in the next section that the studied model can be viewed as a particular case of this general spline smoothing problem. A function f of d variables $x_1, \ldots, x_d$ is "smooth" if the penalty functional $J_m^d(f)$ defined by

$$J_m^d(f) = \sum_{\alpha_1 + \ldots + \alpha_d = m} \frac{m!}{\alpha_1! \ldots \alpha_d!} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left(\frac{\partial^m f}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}\right)^2 dx_1 \ldots dx_d \tag{8}$$

is small.

Let $t = (x_1, \ldots, x_d)$ and $t_i = (x_1(i), \ldots, x_d(i))$. Consider the following data model:

$$y_i = L_i f + \varepsilon_i\ ,\quad i = 1, \ldots, n, \tag{9}$$

where the L_i are continuous linear functionals of f and the ε_i are independent zero mean errors with $E[\varepsilon_i^2] = \sigma_i^2$. A useful result is that the operator L defined by $Lf = \frac{\partial^k f}{\partial^{\alpha_1} x_1 \ldots \partial^{\alpha_d} x_d}$ for $\alpha_1 + \ldots + \alpha_d = k$ $(k, \alpha_1, \ldots, \alpha_d \in \mathbb{N})$ is a continuous linear form if and only if $2m - 2k - d > 0$ (see Berlinet and Thomas-Agnan (2004), Th. 133, and Wahba

and Wendelberger (1980)). A generalization of the spline smoothing problem as defined in Section 3.1.2 is given by the following variational problem: Find f in a suitable Hilbert space $\mathcal{H}$ for which $J_m^d(f)$ is finite, to minimize

$$\frac{1}{n}\sum_{i=1}^{n}\sigma_i^{-2}(L_i f - y_i)^2 + \lambda J_m^d(f). \tag{10}$$

The null space of the penalty functional $J_m^d(f)$, i.e. $\mathcal{H}_0 = \{f : J_m^d(f) = 0\}$, is the $p = \binom{d+m-1}{d}$-dimensional space spanned by the polynomials in d variables of total degree less than or equal to $m-1$. We will denote $\phi_1, \ldots, \phi_p$ these polynomials. Then the following theorem gives the form of the minimizer of Eq. (10):

Theorem 3 (Wahba et Wendelberger, 1980)**.** *Let $L_1, \ldots, L_n$ be n linearly independent continuous linear functionals and suppose $L_k \sum_{\nu=1}^{p} a_\nu \phi_\nu = 0$ implies that all the a_ν are 0.*

Then Eq. (10) has a unique minimizer $\widehat{f_\lambda}$ in $\mathcal{H}$, with representation

$$\widehat{f_\lambda}(t) = \sum_{\nu=1}^{p} d_\nu \phi_\nu(t) + \sum_{i=1}^{n} c_i \xi_i(t), \tag{11}$$

where

$$\xi_i(t) = L_{i(s)} E_m(t, s), \tag{12}$$

and $L_{i(s)}$ means the linear functional L_i applied to what follows considered as a function of s, and

$$E_m(s, t) = E(|s - t|), \tag{13}$$

$|s - t|$ being the Euclidean distance, with

$$E(u) = \begin{cases} \theta_{m,d}|u|^{2m-d} log|u|, & d \text{ even} \\ \theta_{m,d}|u|^{2m-d}, & d \text{ odd} \end{cases} \tag{14}$$

where

$$\theta_{m,d} = \begin{cases} \frac{(-1)^{d/2+m+1}}{2^{2m-1}\pi^{d/2}(m-1)!(m-d/2)!}, & d \text{ even} \\ \frac{\Gamma(d/2-m)}{2^{2m}\pi^{d/2}(m-1)!}, & d \text{ odd.} \end{cases} \tag{15}$$

The coefficients $c = (c_1, \ldots, c_n)^T$ and $d = (d_1, \ldots, d_p)^T$ are solutions of the linear system

$$(K + n\lambda W^{-1})c + Td = y, \tag{16}$$

$$T^T c = 0, \tag{17}$$

where

$$\begin{aligned} K &= \{L_{i(s)} L_{j(t)} E_m(s, t)\}_{i,j=1}^{n}, \\ T &= \{L_i \phi_\nu\}_{i=1\ \nu=1}^{n\ \ p}, \\ W &= diag(\sigma_1^{-2}, \ldots, \sigma_n^{-2}). \end{aligned}$$

The proof of this theorem is given in Wahba and Wendelberger (1980) and use the theory of reproducing kernels (see Kimeldorf and Wahba (1971) and Wahba (1990)). Note that this theorem indicates that the smoothing spline estimate $\widehat{f_\lambda}$ falls in a finite dimensional space. When the bounded linear functional are the evaluation functionals, i.e. $L_i f = f(t_i)$, $i = 1, \ldots, n$, the minimizer f_λ is a thin-plate spline (see Wahba (1990), Section 2.4, and Gu (2002), Section 4.4). The smoothing spline model described in Section 3.1.2 is a particular case of the model of Eq. (9) where $d = 1$ and the bounded linear functional are the evaluation functionals.

3.2. Estimation of a space-speed profile from noisy data: A smoothing problem under constraints

The estimation of a space-speed profile from noisy measurements of position and speed is a complex nonparametric regression problem (see for example Andrieu et al. (2013)). Indeed, on the one hand, both the response variable (corresponding to speed) and the explanatory variable (corresponding to vehicle position) are noisy. And on the other hand, the regression function must belong to the space $\mathcal{E}_{SSP}$ defined in Definition 1 and then check its properties, in particular the non differentiability when the speed is zero. To overcome these difficulties, we propose to change to a more suitable study area and start by estimating the function F representing the distance traveled as function of time (study area `distance` $\times$ `time` in Fig. 1). Then the new nonparametric model is

$$y_i = F(t_i) + \varepsilon_{x,i}, \quad i = 1, \dots, n, \tag{18}$$

where y_i are noisy observations of the distance traveled, $\varepsilon_{x,i}$ are uncorrelated errors with zero mean and σ_x^2 variance, and $F(t)$ is the regression function. This change of study area leads to take into account the two following constraints:

1. Use the derivative information, i.e. estimate the regression function $F(t)$ of the model (18) from both noisy observations of F corresponding to position measurements, and noisy observations of its derivative F' corresponding to speed measurements.
2. A monotonicity constraint since the function F representing the distance traveled as function of time must be increasing.

The consideration of these two constraints in the smoothing step is the subject of the two following subsections. Then, once we have obtained an estimator $\widehat{F}$ of the function F, it is easy to deduce by differentiation an estimator $\widehat{F'}$ of the time-speed profile, and finally to deduce an estimator $\widehat{v_S}$ of the space-speed profile v_S by the transformation $\widehat{v_S} = \widehat{F'} \circ \widehat{F}^{-1}$.

3.2.1. Smoothing using derivative information

The first constraint in the nonparametric model (18) is to use the derivative information, i.e. to estimate the regression function $F(t)$ from both noisy measurements of position and speed. Assume that the domain of $F(t)$ is $\mathcal{X} = [0, T]$, where T is a positive real, and $F \in W^m[0, T]$ with $m > 1$. Then the use of the derivative information leads to consider the following data model:

$$\begin{cases} y_i = F(t_i) + \varepsilon_{x,i}, & i = 1, \dots, n \\ v_i = F'(t_i) + \varepsilon_{v,i}, & i = 1, \dots, n \end{cases} \tag{19}$$

where y_i and v_i are noisy measurements of distance traveled and speed respectively at each sampling time t_i, and $\varepsilon_{x,i}$ and $\varepsilon_{v,i}$ are independent zero mean errors with variance σ_x^2 and σ_v^2 respectively. We also assume that for all $i = 1, \dots, n$, $\varepsilon_{x,i}$ and $\varepsilon_{v,i}$ are independent. Note that we have assumed that the observations y_i and v_i are obtained at the same times t_i. Otherwise, a data resampling will lead to this case.

The problem of smoothing with derivative information appears in various applications such as economy (Hall and Yatchew (2007)), molecular biology (Calderon et al. (2010)) or image analysis (Mardia et al. (1996)). We propose to solve this problem by using smoothing splines, which have the advantage of requiring the estimation of a single smoothing parameter λ, contrary to penalized splines used by Calderon et al. (2010), for which the choice of the number of knots can be difficult. Thus, this problem can be seen as a special case of the general spline smoothing problem (see Wahba (1990) and Wang (2011)) and can be solved by using the theory of reproducing kernel Hilbert spaces (see Cox (1988)). Then, the estimator can be written as a linear combination of basis functions and kernel functions. However, to compute the estimator, it is necessary to choose a norm associated with the design space (in this study, the Sobolev space $W^m[0, T]$) that is suitable. Indeed, the expression of the reproducing kernels used for calculating the estimator depends on the choice of this norm, and can leads to difficulties in the numerical computation (see Andrieu (2013), Chap. 4, for more details). So, we propose to use the theory of thin-plate spline which leads to similar results for the form of the estimator (see Duchon (1977), Meinguet (1979) and Wahba and Wendelberger (1980)) but that greatly simplifies the computation of the estimator.

The model (19) is a particular case of the general spline smoothing model (9) with the dimension $d = 1$, since it can be rewritten as follows:

$$y_j = \mathcal{L}_j F + \varepsilon_j \,, \quad j = 1, \ldots, 2n, \tag{20}$$

where

- the observations y_j are defined by:
  $\begin{cases} y_j = y_i \text{ with } i = j \text{ for } j = 1, \ldots, n \\ y_j = v_i \text{ with } i = j - n \text{ for } j = n+1, \ldots, 2n \end{cases}$,
- the bounded linear functionals $\mathcal{L}_j$ on $W^m[0,T]$ are defined by:
  $\begin{cases} \mathcal{L}_j F = F(t_i) \text{ with } i = j \text{ for } j = 1, \ldots, n \\ \mathcal{L}_j F = F'(t_i) \text{ with } i = j - n \text{ for } j = n+1, \ldots, 2n \end{cases}$,
- the errors ε_j are defined by:
  $\begin{cases} \varepsilon_j = \varepsilon_{x,i} \text{ avec } i = j \text{ pour } j = 1, \ldots, n \\ \varepsilon_j = \varepsilon_{v,i} \text{ avec } i = j - n \text{ pour } j = n+1, \ldots, 2n \end{cases}$.

The useful result stated in Section 3.1.3 leads to the conclusion that the linear functionals are bounded on $W^m[0,T]$ if $m > 1$ since $d = 1$, $k = 0$ for $j = 1, \ldots, n$ and $k = 1$ for $j = n+1, \ldots, 2n$. Then, an estimator of $F(t)$ is the minimizer of the following criterion in $W^m[0,T]$:

$$\frac{1}{2n}\{\sigma_x^{-2} \sum_{i=1}^{n} (y_i - F(t_i))^2 + \sigma_v^{-2} \sum_{i=1}^{n} (v_i - F'(t_i))^2\} + \lambda \int_0^T (F^{(m)}(t))^2 dt. \tag{21}$$

Provided that the two hypothesis on the linear functionals are satisfied, which it's the case if the sampling points $t_1, \ldots, t_n$ are distincts, we can apply Theorem 3 of Wahba and Wendelberger. So, under this assumption, we deduce from Theorem 3 that the minimizer of (21) can be written as:

$$\widehat{F}_\lambda(t) = \sum_{\nu=1}^{m} d_\nu \phi_\nu(t) + \sum_{i=1}^{n} c_i E_m(t_i, t) + \sum_{i=1}^{n} c'_i \frac{\partial}{\partial s} E_m(s,t)|_{s=t_i}, \tag{22}$$

where

$$\begin{aligned} E_m(s,t) &= \theta_{m,1} |s-t|^{2m-1}, \\ \theta_{m,1} &= \frac{\Gamma(1/2 - m)}{2^{2m} \pi^{1/2} (m-1)!}. \end{aligned}$$

The coefficients $c = (c_1, \ldots, c_n, c'_1, \ldots, c'_n)^T$ and $d = (d_1, \ldots, d_m)^T$ are solutions of the linear system defined by Eqs. (16) and (17) where

$$\begin{aligned} T &= \{\mathcal{L}_j \phi_\nu\}_{j=1}^{2n} {}_{\nu=1}^{m}, \\ K &= \{\mathcal{L}_{j(s)} \mathcal{L}_{k(t)} E_m(s,t)\}_{j,k=1}^{2n}, \\ W &= diag(\sigma_x^{-2}, \ldots, \sigma_x^{-2}, \sigma_v^{-2}, \ldots, \sigma_v^{-2}). \end{aligned}$$

Then, the solution $\widehat{F}_\lambda(t)$ is a polynomial spline of order m with knots at the sampling time $t_1, \ldots, t_n$.

Note that the error variances σ_x^2 and σ_v^2 are usually unknown in practice. In general, we use an estimator of the error variance corresponding to the criterion used for the selection of the smoothing parameter λ. Three scores are commonly used:

- the UBR score ("Unbiased Risk") which is an extension of the Mallow's C_p criterion ;
- the GCV score ("Generalized Cross-Validation") which is a weighted version of the standard cross-validation ;
- the GML score ("Generalized Maximum Likelihood") based on a Bayesian model, and that required a normality assumption on the errors.

The selection of the smoothing parameter results from the minimization of one of these criteria, and the error variances estimates depend on the smoothing parameter obtained, and therefore on the criterion chosen. Thus, if in a first time, we consider only the position measurements, i.e. the data model (18), and if we denote $A(\lambda)$ the hat matrix defined by

$$(\widehat{F}_\lambda(t_1), \ldots, \widehat{F}_\lambda(t_n))^T = A(\lambda)y, \tag{23}$$

where $\widehat{F}_\lambda$ is the smoothing spline estimate of F for the smoothing parameter λ (that is a polynomial spline of order $2m$ by results stated in Section 3.1.2), then the variance estimate of σ_x^2 associated with the GCV criterion is

$$\widehat{\sigma}^2_{gcv} = \frac{y^T(I - A(\lambda_v))^2 y}{tr(I - A(\lambda_v))}, \tag{24}$$

and the variance estimate associated with the GLM criterion is

$$\widehat{\sigma}^2_{gml} = \frac{y^T(I - A(\lambda_m))y}{n - m}. \tag{25}$$

Similarly, in a second time, we consider only the speed measurements and calculate the smoothing spline estimate of F' for the smoothing parameter λ, and then deduce an estimate of the variance error σ_v^2.

3.2.2. Smoothing under monotonicity constraint

The second constraint in the nonparametric model (18) is a monotonicity constraint since the function F representing the distance traveled as a function of time must be increasing. Various methods of smoothing under monotonicity constraint have been developed. The main approaches are based on kernel smoothers and splines. An overview of these methods can be found in Delecroix and Thomas-Agnan (2000). Among the main methods, we can cite the isotonic regression introduced by Brunk (1955), the monotone splines (for example, Ramsay introduces the I-splines basis in Ramsay (1988) for monotone regression splines) or the projection methods (e.g. Delecroix et al. (1996) or Mammen et al. (2001)).

In a previous study (Andrieu et al. (2012)), the method of homeomorphic splines developed by Bigot and Gadat (2010) have been tested. However, if the monotonization step presented good results, we had difficulties in the implementation of the derivative. So, we propose to use a method developed by Ramsay (1998) which has the advantage of being relatively simple to implement. The principle of this method is to transform the constrained smoothing problem to an unconstrained one. A monotone function has a positive first derivative. So the main idea is that any strictly monotonic function f satisfies the following differential equation:

$$D^2 f = wDf, \tag{26}$$

where Df and $D^2 f$ are respectively the first and second derivative of the function f, and w is an unconstrained function. So any strictly monotonic function f can be written as following (as solution of the equation (26)):

$$f(t) = \beta_0 + \beta_1 \int_0^t exp[\int_0^u w(v)dv]du, \tag{27}$$

where β_0 and β_1 are arbitrary constants such that $f(0) = \beta_0$ and $f'(0) = \beta_1$. Then, the problem is to estimate the coefficients β_0 and β_1 and the unconstrained function w by minimizing the following criterion:

$$\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 h(t_i))^2 + \lambda \int_0^T (w^m(t))^2 dt, \tag{28}$$

where

$$h(t) = \int_0^t exp[\int_0^v w(v)dv]du. \tag{29}$$

The unconstrained function w is computed using an appropriate basis expansion (e.g. B-splines) and the coefficients β_0 and β_1 are estimated by numerical algorithms. However, due to a numerical optimization of the criterion (28), monotone smoothing spline involves considerably more computation than the usual smoothing spline process.

Thus, the monotonicity constraint is consider in a second smoothing step by applying the method of Ramsay described above, to the estimated values $\widehat{F}(t_i)$ obtained at the first smoothing step for which we have used the speed measurements (Section 3.2.1). Therefore, this second smoothing step can be seen as a monotonization step and then is similar to the projection step in projection methods (see Mammen et al. (2001)). The new monotone estimator of $F(t)$ is then denoted $\widehat{F}_{mono}$, and finally we deduce an estimator $\widehat{v_S}$ of the space-speed profile v_S by the transformation $\widehat{v_S} = \widehat{F}'_{mono} \circ \widehat{F}^{-1}_{mono}$.

3.3. Validation of the smoothing method

We propose to test the developed smoothing method using real-world GPS data collected from test track in Versailles-Satory, France. The studied trip corresponds to two laps, each lap having a length of about 2 km. The vehicle is a Renault Clio III equipped with a data recorder which monitors at 1Hz distance traveled and speed from the vehicle network (CAN bus), and vehicle position from two GPS: a GlobalSat BR-355 SiRF III GPS receiver to collect non accurate measurements, and a Thales Sagitta Real Time Kinematic GPS (RTK) with high accuracy to collect reference data in our test. The GlobalSat BR-355 GPS also estimates vehicle speed using the Doppler effect, which is independent of vehicle location, but not the RTK-GPS. The objective of this test is to estimate the space-speed profile v_S of the vehicle during this run from position and speed measurements obtained with the GlobalSat BR-355 GPS. Note that in this study, position measurements correspond to the distance traveled calculated from map-matched GPS.

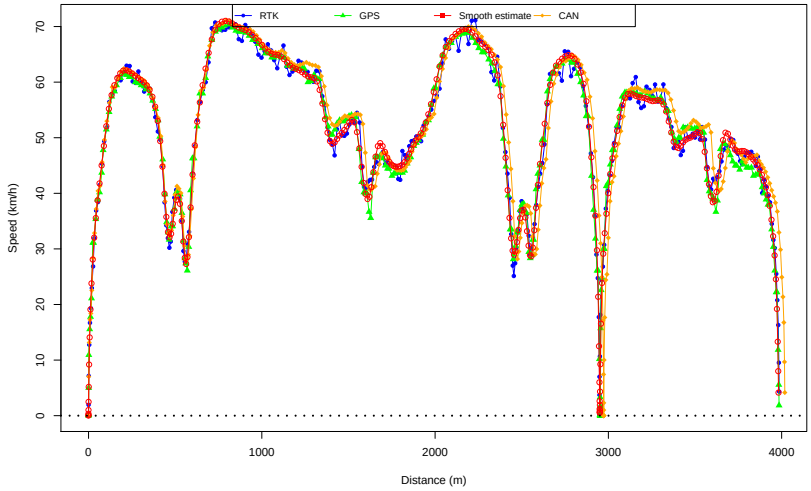
Before applying the smoothing method, the error variances σ_x^2 and σ_v^2 are estimated using a smoothing spline estimate of degree 5 (quintic spline) with a smoothing parameter calculated by minimization of the GML criterion. Results are $\widehat{\sigma}_x = 2.21\ m$ and $\widehat{\sigma}_v = 0.42\ m/s$.

For, the first step of the developed smoothing method, that is the use of the derivative information (speed measurements), we have chosen $m = 3$ (i.e. a quintic spline) and the GML criterion for the selection of the smoothing parameter ($\lambda_{gml} = 0.0143$). The estimator $\widehat{F}_\lambda(t)$ (Eq. (22)) is computed with the function `ssr` in the R package `assist` (R Development Core Team (2008) available at `http://cran.r-project.org`).

For the second step, that is the monotonization step, we have chosen $m = 3$ for the degree of the penalty in Eq. (28) and $\lambda = 10^{-1}$ for the smoothing parameter (trial and error method). The resulting estimator $\widehat{F}_{mono}(t)$ is computed with the function `smooth.monotone` in the R package `fda`. Fig. 3.a shows the smoothing results with the estimated space-speed profiles (in red) obtained from speed and position GPS measurements (in green). Position measurements from RTK-GPS are used as reference data of position of the vehicle, and derivative values of this accuracy position measurements (calculated by central-difference approach) are used as reference data of speed (in blue). This figure shows that the proposed estimator of the space-speed profile is good and corrects some measurement errors. It is also important to note that the proposed estimator (red curve) is a function belonging to the space $\mathcal{E}_{SSP}$ defined in Definition 1, contrary to other curves that are linear interpolations of observed points. The advantage of this functional approach will be illustrated in more details at Section 4. Fig. 3.b that is a focus on the stop located about 3000 m, confirms these good results even if the estimated speed is over-estimates at the stop (estimated speed is non zero). This problem results from the mononization step, since the method of Ramsay which is used, provides a strictly increasing estimator $\widehat{F}_{mono}(t)$ whereas the real function $F(t)$ is constant when the vehicle is stopped. So, the main difficulty of the estimation of the distance traveled $F(t)$ is the estimation of the points for which the speed is zero (i.e. stops). However, we have observed that in practice, only short stops (less than 5 s) are not very good estimated. In Fig. 3.b, the stop lasts exactly 5 s and the estimated speed is less than 2 km/h, which is acceptable.

Finally, the proposed smoothing method is validated with a comparison between the *Root Mean Square Error* (RMSE) of the GPS measurements and the RMSE of the estimated values. Table 1 provides the results for distance and speed values. Even if noisy measurements from the GPS are not too bad, the proposed estimator provides smallest errors for both position and speed values.

a. Complete run.



b. Zoom on the stop located about 3000 m.

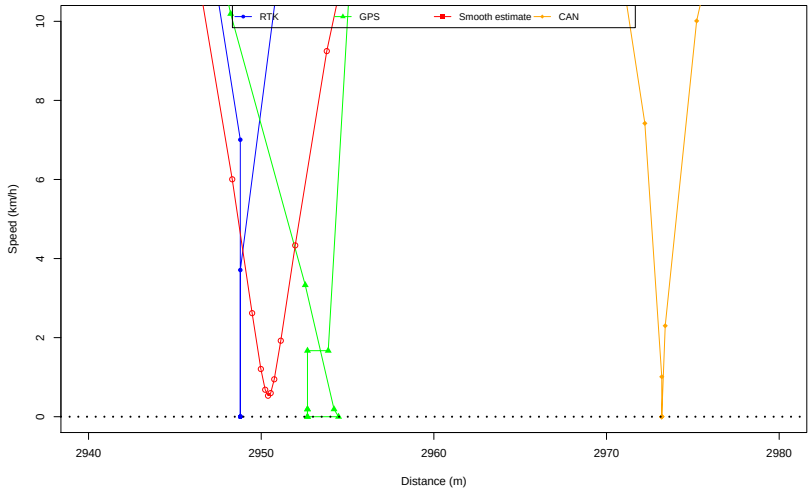


Figure 3: Smoothing results in the study area [**speed** $\times$ **distance**]. Measurements from the RTK-GPS (reference data) and the GlobalSat BR-355 GPS (observations) are in blue and green, respectively. The estimated space-speed profile ("Smooth estimate") is in red. By comparison, measurements from the CAN bus are in orange.

4. Construction of an aggregated speed profile and speed corridors

If the speed profiles are very informative on the individual behavior of road users and especially on their speeds, the estimation of actual speeds from measurements collected by probe vehicles is not easy and requires the use of appropriate methods. Indeed, we have shown at the previous section that a smoothing procedure allows to convert discrete measurements into individual speed profiles of functional nature. However, when the volume of data is

Table 1: Table of the Root Mean Square Error (RMSE) for distance and speed.

	GPS data	Smoothing estimate
RMSE for distance (m): $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{F}_{mono}(t_i) - F(t_i))^2}$	3.13	2.05
RMSE for speed (km/h): $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{F'}_{mono}(t_i) - F'(t_i))^2}$	2.14	1.64

large, it is necessary to summarize the information with the construction of an aggregated speed profile which is representative of the set of the individual speed profiles.

Various studies used the average speed profile such as Hyden and Varhelyi (2000) who study the impact of a modification of the infrastructure by comparing the average speed profile before and after the addition of roundabouts, or Várhelyi et al. (2004) who study the effects of an active accelerator pedal by comparing the average speed profile with and without the system. However, if the average speed profile appears to be a good representative profile, its construction is not always easy. For example, Laureshyn et al. (2009) distinguish different groups associated with driving situations at a signalised intersection and construct the corresponding average profile, but due to phase variation (i.e. horizontal variation) between the individual speed profiles, the average profile is not representative of the corresponding class. This problem shows that it is necessary to synchronize the set of individual speed profiles before calculating an average profile in order to remove these phase variations. Some authors such as citetKerper2012 have approached this problem of alignement of curves, and a method based on landmarks alignement will be proposed in Section 4.3.

In addition, the speed variations between road users are also an important information to take into account. Boonsiripant (2009) considers speed variations as a risk indicator and provides various indicators which summarize the speed variability between drivers with an overall indicator. In Section 4.4, we propose a method to obtain speed corridors that reflect the speeds dispersion on a given road section, and preserve the functional nature of speed profiles.

4.1. The data

The data set used in the next sections of this study is extracted from an experiment conducted by the French laboratory IFSTTAR-LIVIC and that took place in 2012 in Versailles, France. Thirty-nine drivers participated to this experiment and performed twice a road section of urban and inter-urban type with a length of about 1100 m. This road section, illustrated at Fig. 4, corresponds to the path from A to B and is composed of a stop sign, two roundabouts and a traffic light. For logistical reasons, two vehicles were used for this experiment: a Renault Clio III equipped with a Garmin GPS 16x LVC (for 20 drivers), and a Renault Modus with a GPS GlobalSat BR-355 (for 19 drivers). Note that the use of two vehicles and the fact that each driver performed twice the studied section lead to conditions close to naturalistic driving studies where different drivers were observed in a natural setting, in particular during regular travels such as the commute to work. Thus, we do not take into account the correlation between the two paths of the same driver. The aim of the study is to focus on space-speed profiles from this data set, i.e. position and speed measurements from the GPS.

4.2. From raw data to individual speed profiles: the smoothing step

Since the 39 drivers performed twice the studied section, the data set is composed of 78 individual space-speed profiles. However, the raw data, illustrated at Fig. 5.a, are represented as sequences of time-stamped noisy measurements of position and speed from GPS, and then as vectors. So, the first step is to convert them in functional objects by using a smoothing procedure in order to filter out the noise as efficiently as possible and to preserve the continuous underlying process. The smoothing procedure described in Section 3 is applied as follows:

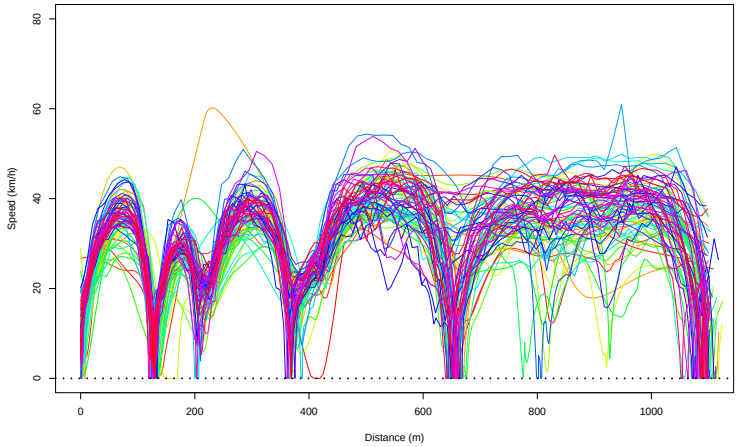
Figure 4: Map of the studied section.

1. A first smoothing step using derivative information, with for each path j, $j = 1, \ldots, 78$, an estimation $\widehat{F}_{\lambda_j}(t)$ of each function $F_j(t)$ (representing the distance traveled as function of time) with the following parameters:
   - an estimation of the variance $\sigma^2_{x,j}$ and $\sigma^2_{v,j}$ for each path j, $j = 1, \ldots, 78$;
   - $m = 3$ (quintic spline) ;
   - an automatic selection of each smoothing parameter λ_j resulting from the minimization of the GML criterion.
2. A second smoothing step under monotonicity constraint, corresponding to a monotonization of each estimate $\widehat{F}_{\lambda_j}(t)$ obtained at the previous step with the following parameters:
   - $m = 3$ for the degree of the penalty ;
   - a selection of each smoothing parameter by trial and errors.

Fig. 5.b illustrates the results of this smoothing procedure and shows the smooth individual space-speed profiles obtained with the transformation $\widehat{F}'_{j,mono}(t) \circ \widehat{F}^{-1}_{j,mono}(t)$. Results are good since some peaks which appear in raw data, and that probably correspond to outliers, are reduced (e.g. blue and orange curves). Missing values are also be corrected by the smoothing procedure. The disadvantage of the step monotonisation mentioned in Section 3.3 which causes an over-estimation of speed, appears mainly at the stop (short stop) but is less important at the traffic light (long stop). Finally, note that the main difference between Fig. 5.a and Fig. 5.b is that the smoothing procedure allows to

reduce the study of these individual speed profiles to a functional framework, and then to take into account functional characteristics such as phase variation (see Section 4.3).

a. Raw data.



b. Smooth data.

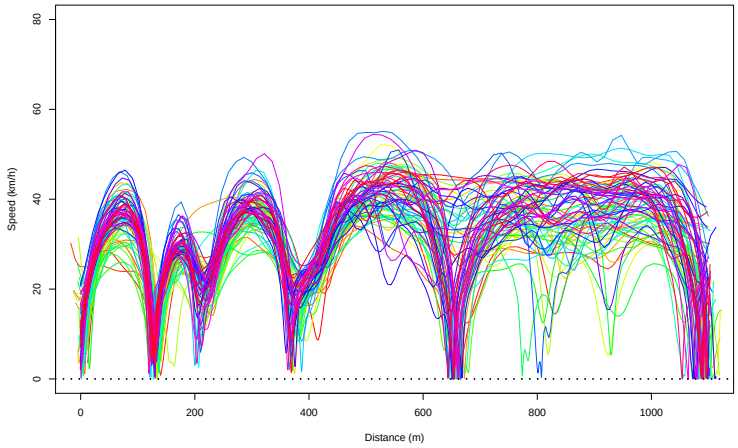


Figure 5: Smoothing step on the 78 individual space-speed profiles.

4.3. Registration of speed profiles by landmarks alignment

The *curve registration* or *curve alignment* problem appears in many areas such as biology, meteorology, pattern recognition... (Ramsay and Li (1998), Bigot (2006)). Indeed, frequently, observed curves exhibit two types of variability: *amplitude variation* which corresponds to vertical variation, and *phase variation* which corresponds to horizontal variation (Ramsay and Silverman (2005)). Then, to build a representative curve of a set of observed curves, it is necessary to correct the phase variation in order to obtain curves with similar features. For example, the calculation of an average speed profile which is representative of a set of speed profiles requires that all vehicles stop at the same location, which is not the case in practice (see Fig. 5.b). The curve registration problem consists in finding, for each curve, a warping function and to deform all the curves in order to align them. If the literature about this problem is

relatively large (e.g. Kneip and Gasser (1992), Wang and Gasser (1997)), it is not treated or treated in a simple way in velocity profiles studies (shift registration is proposed in Kerper et al. (2012) and Violette et al. (2010)).

We propose to use the method of landmarks alignment which consists to determine, for each curve, a deformation function so that specific points called landmarks of the registered curves are aligned. Specific points defined as landmarks are generally the positions of maxima, minima, inflection points, or zero crossings. Then, the landmarks registration of m signals $f_1, \ldots, f_m$ defined on the same interval $[0, X]$ can be divided into the five following steps (see Bigot (2003)):

1. Definition of characteristic points to be used as landmarks (eg, minimum, maximum, zero crossing ...).
2. Extraction of landmarks $x_{i,1}, \ldots, x_{i,K}$ from an observed sequence of each signal f_i, $i = 1 \ldots, m$. Note that since observed signals are noisy, the landmarks $x_{i,1}, \ldots, x_{i,K}$ are usually extracted from a estimator $\widehat{f_i}$ of the signal f_i.
3. Identify landmarks reference $x_{0,1}, \ldots, x_{0,K}$, i.e. the points at which the curves must match.
4. Determine deformation functions $h_1, \ldots, h_m$ so that corresponding landmarks are matched, i.e. for all $i = 1, \ldots, m$, $h_i(x_{0,j}) = x_{i,j}$, $j = 1, \ldots, K$.
5. Deformation of the signals using transformations obtained in the previous step. The registered functions $\widetilde{f_i}(x) = f_i[h_i^{-1}(x)]$, $i = 1, \ldots, m$, are then aligned at each points $x_{0,1}, \ldots, x_{0,K}$.

The deformation functions $h_i(x)$, $i = 1 \ldots, m$, called *warping functions*, must check the following properties:

- Initial conditions: $h_i(0) = 0$, $h_i(X) = X$.
- Landmarks alignment: $h_i(x_{0,j}) = x_{i,j}$.
- Strict monotonicity: $x_1 < x_2$ implies $h_i(x_1) < h_i(x_2)$ (in order to respect the sequencing of points).

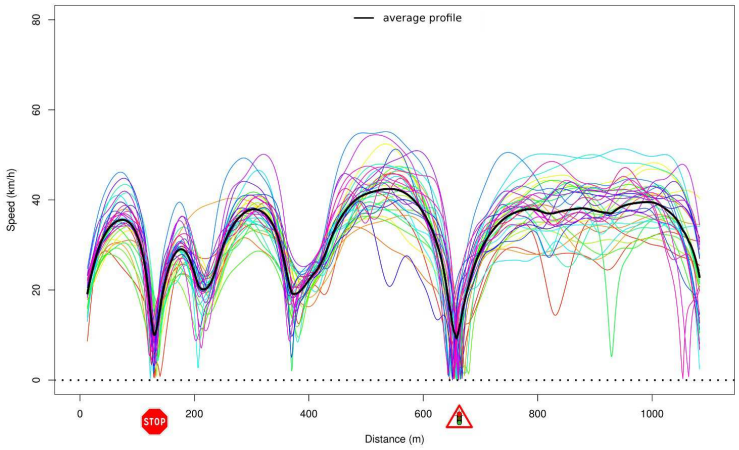
The method of landmarks alignment is applied to the set of speed profiles illustrated at Fig. 5.b. In order to compare similar speed profiles, we distinguish the two driving situations corresponding to the state of the traffic light (red or green light). Only the red light case will be studied in the following, that represents a sample of 36 individual profiles. We have chosen to define landmarks as the positions of the two elements of the infrastructure that require a stop of the vehicle, namely the stop sign and the red light. Thus, the landmarks, corresponding to zero-crossing (or local minima) at the stop sign and the traffic light positions, are extracted from the estimated space-speed profiles obtained with the smoothing procedure, and are matched with the reference landmarks defined by the average position of vehicle stops at this two elements of the infrastructure. Then, monotone cubic spline interpolation have been determined as warping functions and have been computed with the R function `splinefun` and the option "monoH.FC". We also impose the condition that the warping functions are linear with a slope equal to one around the stops (we fix an interval of length 100 m around each stop) in order to not too distort the space-speed profiles in the neighborhood of each stop and to obtain "true" space-speed profiles as defined in Definition 1.

Fig. 6 compares the unregistered (Fig. 6.a) and the registered (Fig. 6.b) speed profiles in the red light case (36 curves). Fig. 6.a illustrates the fact that averaging unregistered profile results in an average profile (black curve) that is not representative of the set of the individual speed profiles. Indeed, this average profile doesn't equal to zero at the red light unlike all individual profiles. In contrast, Fig. 6.b shows that the average of the registered profiles tends to resemble much more closely most of the individual profiles, and then is a good aggregated speed profile of the sample.

4.4. Construction of speed corridors

If the construction of an aggregated speed profile, such as the average, leads to a good representation of the actual speeds on a road network section, such an aggregated profile does not reflect the variability between road users. The boxplot proposed by Tukey (1977) is a graphical method used to represent the distribution of univariate data, and can be used to represent speed variations between individuals at a given point. For example, Fig. 7.a represents pointwise boxplots calculated at a regular interval of 10 m in the red light case, with medians connected by a red line (V50 profile) and 85th percentiles connected by a blue line (V85 profile). However, this representation lost the continuous form of the individual profiles, and then the V50 and V85 profiles are not true space-speed profiles as defined in Definition 1 in contrast to the average speed profile obtained at Fig. 6.b.

a. Unregistered space-speed profiles.



b. Registered space-speed profiles.

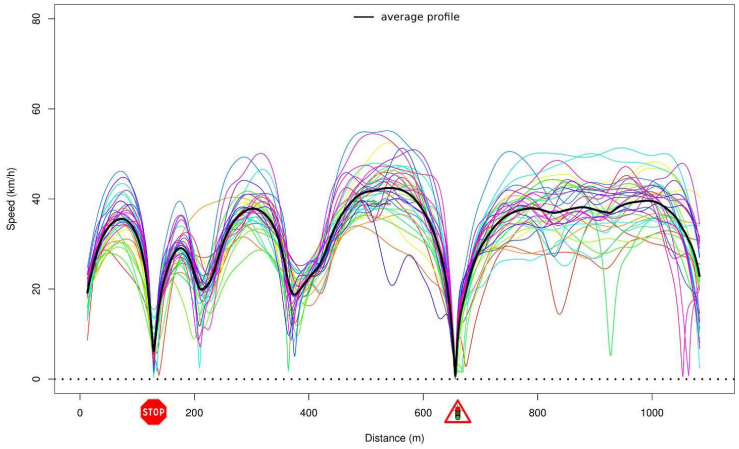


Figure 6: Registration of space-speed profiles in the red light case (36 curves). The black curve is the average profile.

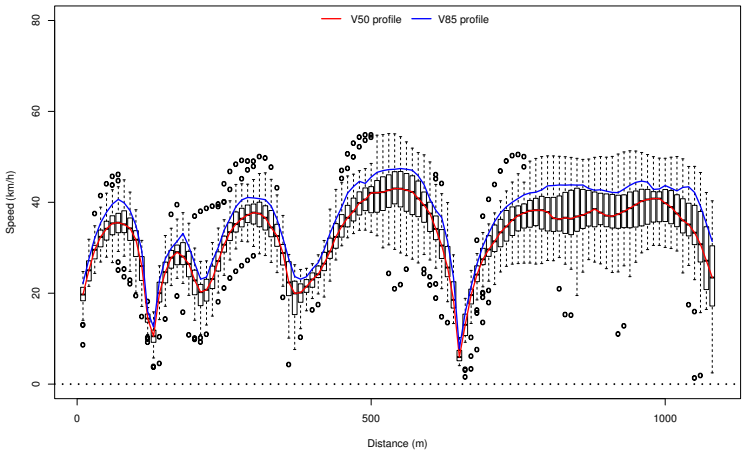
So, we propose to use a graphical tool called functional boxplots, recently developed by Sun and Genton (2011), which extends the notion of boxplots to functional data. This tool is based on the notion of functional depth which generalizes order statistics or ranks to the functional setting. Indeed, the first step to construct a boxplot is the data ordering. But if the notion of order is obvious in the univariate setting, it is much more complicated in the functional setting. This problem has led to the emergence of the concept of functional depth, first introduced for multivariate data (Zuo and Serfling (2000)), that provides a measure of "centrality" and "outlyingness" for a function within a sample of curves and allows to order them from center-outward (López-Pintado and Romo (2009)). The median curve is then the curve with the higher depth. Various examples of functional depth have been proposed in the literature such as the Fraiman and Muniz depth (Fraiman and Muniz (2001)), the random projection depth (Cuevas et al. (2007)) or the band depth (López-Pintado and Romo (2009)). If Sun and Genton (2011) use the band depth and its modified version for the construction of its functional boxplots, a comparison of the results obtained with various functional depth led us to choose the h-mode depth introduced by Cuevas et al. (2006) and based on the concept of mode. The authors defined a functional mode as the curve most densely surrounded by the rest of curves of the dataset. Thus, the h-modal functional depth of a curve x_i with respect the set of curves $x_1, \ldots, x_n$ is given by:

$$MD_n(x_i, h) = \sum_{k=1}^{n} K(\frac{\|x_i - x_k\|}{h}), \tag{30}$$

where $\|.\|$ is an appropriate norm, K is a kernel function, and h is a bandwidth. In practice, the L^2 norm and the truncated Gaussian kernel are used, and the bandwidth taken is the 15th percentile of the empirical distribution of $\{\|x_i - x_k\|,\ i,k = 1,\ldots,n\}$. Functional boxplots create with the h-modal depth are illustrated in Fig. 7.b in the red light case. This functional boxplot is composed of the maximum enveloppe (blue curves), the median profile (black curve) which is the most central curve with the highest h-modal depth, the 25% central region (dark magenta region), the 50% central region (magenta region) and the 75% central region (pink region). The red dashed curves are the outlier candidates detected by the 1.5 times the 50% central region rule (see Sun and Genton (2011)). This functional boxplot have been computed with the function `fdepth` of the R package `rainbow` and the function `fbplot` of the R package `fda` (R Development Core Team (2008)).

The advantage of this graphical tool is that it allows to represent the speed dispersion among individuals on a given road section. These speed corridors allow to distinguish road sections where the speed variability is large and those for which speeds are more homogeneous. Moreover, this tool leads to the extraction of the median profile which depends to the choice of a functional depth, and that can be used as a representative speed profile of the set of the individual speed profiles instead of the average profile.

a. Pointwise boxplots.



b. Functional boxplots.

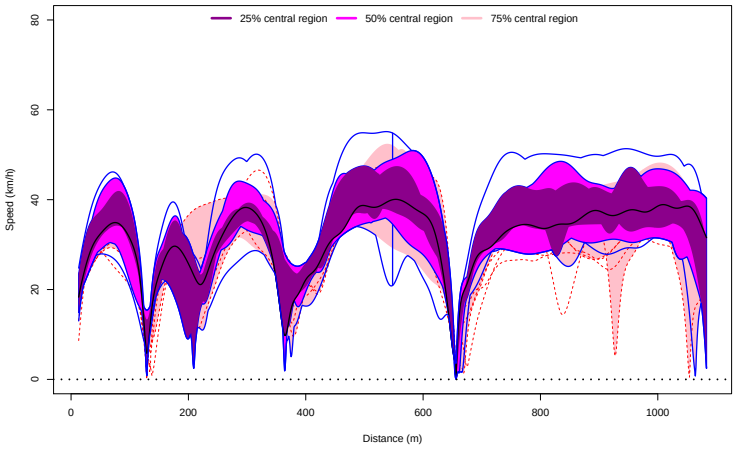


Figure 7: Pointwise boxplots ans functional boxplots in the red light case (36 curves).

5. Conclusion and discussion

The current paper presents a methodology of construction of an aggregated speed profile and speed corridors based on a functional analysis of speed profiles. Indeed, the development of probe vehicles and smartphones, that can turn everyone in data collector ("Floating Car Data"), and the progress in sensors technology leads to collect large volume of data that requires adapted methods. If in practice data collected are discretized and then view as vectors, the fact that grids being increasingly fine makes them inherently functional and leads us to view them as functions. The functional approach proposed in this paper takes inspiration from Functional Data Analysis, a statistical domain that has developed recently, which overcomes the problems obtained with high dimensional vectors and allows to take into account functional characteristics (derivatives, shape constraints, phase variation...).

This paper focus on space-speed profiles, i.e. speed as function of space, which are very informative components of drivers behavior and their road usage. Thus, a definition of the functional space of these objects is proposed and the study of their mathematical properties has shown the remarkable property of non differentiability at points for which speed is zero. Solutions have been proposed to take into account this property that implies some difficulties in the smoothing step and the calculation of an aggregated speed profile such as the average profile.

In a first time, a smoothing procedure has been proposed in order to convert raw data into functional objects and to reduce errors that affected them. We have shown that the estimation of a space-speed profile from noisy measurements of position and speed is a complex nonparametric regression problem that needs to take into account two constraints: the use of the derivative information, and a monotonicity constraint. The studied nonparametric model is viewed as a particular case of a more general spline smoothing model, and the use of known results about thin-plate spline leads us to obtain an estimator of the "true" speed profile. Contrary to the local polynomial regression method commonly used, this method does not require the estimation of many parameters except the smoothing parameter which allows to control the degree of smoothness. The performances of the estimator have been shown using real-world GPS data even if speed tends to be overestimated at short stops.

In a second time, a methodology has been proposed to summarize a set of individual space-speed profiles with an aggregated speed profile. The functional approach allows to use curve registration method in order to correct phase variation, and then to obtain a representative speed profile with similar features of corresponding individual speed profiles. The method of landmarks alignment, which consists to align specific points of the curves, is applied on a data set where two driving situations corresponding to the state of the traffic light (red or green light) are distinguish. A comparison of the unregistered and the registered speed profiles at stops imposed by the infrastructure in the red light case, as well as the corresponding average profiles, illustrates the interest of the method. In addition, the construction of the average profile is extended to the construction of speed corridors that reflect the variability between road users. The construction of speed corridors is based on a graphical tool called functional boxplots, initially proposed by Sun and Genton (2011), which are an extension of the classical boxplots used in the univariate setting. This tool is based on the notion of functional depth that generalizes order statistics to functional data, and allows to order curves from center-outward and to extract the median curve (associated with the highest depth value).

The construction of an aggregated speed profile or speed corridors is very informative about actual operating speed and it could be interesting to incorporate them in a driving assistance system to provide a reference speed or a safe speed corridor to the driver such as the Porsche InnoDrive system (Porsche InnoDrive (2013)). It is also a very useful information for infrastructure manager to study the impact assessment of a modification of the infrastructure with a safety or environmental point of view. In addition, individual speed profiles reflect the adaptation of drivers to the environment and then provide information on the infrastructure itself. Therefore, the present study can be used to characterize some elements of the infrastructure by a "signature" (e.g. a stop sign is characterized by a zero-crossing for all (or most) speed profiles), and to develop a probabilistic model of the presence of various elements of the infrastructure in order to enrich or update digital maps.

Future works will be focused to the development of the smoothing method with the fusion of the two constraints in a single smoothing step, and the addition of spot speed measurements from radars or magnetic loops in order to improve the data accuracy (see for example the article of Louah (2005) on this subject). Future works will also be focused to the development of unsupervised classification methods to distinguish traffic conditions (free vs congestion)

or specific driving conditions (state of traffic lights). Finally, it would be interesting to use the functional approach to detect peaks in acceleration or jerk profiles for the identification of safety critical events. In this case, the use of wavelets seems to be more adapted than splines for the signals denoising.

Acknowledgments

The authors would like to gratefully thank Christine Thomas-Agnan from the University of Toulouse I for discussions about RKHS and the general smoothing problem, and for her help in the use of the derivative information in the smoothing procedure presented in this article. We thank Yuedong Wang from the University of California for his help in the implementation of the smoothing procedure with the R package `assist`, and Jeremie Bigot from the Toulouse Mathematics Institute for discussions about the present study. This work was funded by the French Institute of Science and Technology for Transport, Development and Networks (IFSTTAR).

Appendix A. Proofs of properties given in Section 2.2

Proof of Theorem 1:
Let $x_0 \in [0, x_f]$. There are two distinct cases :
1^{st} case : x_0 is a point of continuity of F^{-1}. Then by composition of two continuous functions, we deduce that $F' \circ F^{-1}$ is continuous at x_0.
2^{nd} case : x_0 is a point of discontinuity of F^{-1}. We begin by demonstrating the following lemma:

Lemma 4. *Let $x_0 \in [0, x_f]$ a point of discontinuity of F^{-1}. Then the speed is zero at this point, i.e. $v_S(x_0) = F' \circ F^{-1}(x_0) = 0$.*

This lemma can be proved easily. Indeed, if x_0 is a point of discontinuity of F^{-1}, then there is a close interval $[t^-, t^+]$ where F is constant and equal to x_0, and by definition of F^{-1}, $F^{-1}(x_0) = t^-$. This implies that $F'_+(t^-) = 0$ where $F'_+(t^-)$ is the right derivative of F at t^-, and as it was assumed that F was differentiable, we also have $F'_-(t^-) = 0$ where $F'_-(t^-)$ is the left derivative of F at t^-. Finally, $F'(t^-) = 0$, and therefore $F' \circ F^{-1}(x_0) = 0$ which ends the proof of the lemma 4.
Now, we study the one-sided limit of $v_S = F' \circ F^{-1}$ at x_0. $(v_S)_-(x_0) = \lim\limits_{\substack{x \to x_0 \\ x<x_0}} v_S(x) = \lim\limits_{\substack{x \to x_0 \\ x<x_0}} F' \circ F^{-1}(x)$. When $x \to x_0$ by lower values, $t \to t^-$ by lower values, and $\lim\limits_{\substack{t \to t^- \\ t<t^-}} F'(t) = 0$ since $F'(t) = 0$ on $[t^-, t^+]$ and F' is continuous at t^-. So, we deduce that $(v_S)_-(x_0) = 0$. Similarly, $(v_S)_+(x_0) = \lim\limits_{\substack{x \to x_0 \\ x>x_0}} v_S(x) = \lim\limits_{\substack{x \to x_0 \\ x>x_0}} F' \circ F^{-1}(x)$. When $x \to x_0$ by upper values, $t \to t^-$ by upper values, and $\lim\limits_{\substack{t \to t^- \\ t>t^-}} F'(t) = 0$ since $F'(t) = 0$ on $[t^-, t^+]$. So, we deduce that $(v_S)_+(x_0) = 0$. Hence, using Lemma 4, we conclude that v_S is continuous at x_0. □

Proof of Theorem 2:

1^{st} case: Assume that F satisfies the assumptions (H_1).
Let x_0 such that $t_0 = F^{-1}(x_0)$. Since $F'(t_0) = 0$, then $v_S(x_0) = 0$, i.e. $x_0 \in H_0$. Under the assumptions (H_1), we can apply the Taylor-Young's formula to F': For all θ in a neighborhood of t_0, $F'(t_0+\theta) = F'(t_0)+\theta F''(t_0)+\frac{\theta^2}{2}F'''(t_0)+\theta^2\varepsilon(\theta)$, where $\varepsilon(\theta) \to 0$ when $\theta \to 0$. But since $F'(t_0) = 0$, if we had $F''(t_0) \neq 0$, then F' would change sign at t_0, which contradicts the strict monotonicity of F. Therefore $F''(t_0) = 0$. So, $F'(t_0 + \theta) \underset{\theta \to 0}{\sim} \frac{\theta^2}{2}F'''(t_0)$ (since it is assumed that $F'''(t_0) \neq 0$).
Let $h = F(t_0 + \theta) - F(t_0)$. We apply the Taylor-Young's formula to F :
$h = F(t_0 + \theta) - F(t_0) = \theta F'(t_0) + \frac{\theta^2}{2}F''(t_0) + \frac{\theta^3}{6}F'''(t_0) + \theta^3\varepsilon'(\theta)$ where $\varepsilon'(\theta) \to 0$ when $\theta \to 0$. In order to study the differentiability of v_s at x_0, we define the following growth rates:

$\frac{v_S(x_0+h)-v_S(x_0)}{h} = \frac{F'(t_0+\theta)-F'(t_0)}{F(t_0+\theta)-F(t_0)} \underset{\theta\to 0}{\sim} \frac{\frac{\theta^2}{2}F'''(t_0)}{\frac{\theta^3}{6}F'''(t_0)} = \frac{3}{\theta}$. This growth rate has no limit when $\theta \to 0$, but this does not prove that it has also no limit when $h \to 0$.
We will prove this by contradiction. Assume that $\frac{v_S(x_0+h)-v_S(x_0)}{h} \underset{h\to 0}{\to} \ell \in \mathbb{R}$. Then, by definition,
$\forall \varepsilon > 0$, $\exists \alpha > 0$ such that $|h| < \alpha \Rightarrow |\frac{v_S(x_0+h)-v_S(x_0)}{h} - \ell| < \varepsilon$.
But since F is continuous at t_0, $\exists \beta > 0$ such that $|t - t_0| < \beta \Rightarrow |F(t) - F(t_0)| < \alpha$, or similarly $|\theta| < \beta \Rightarrow |\underbrace{F(t_0+\theta) - F(t_0)}_{h}| < \alpha$.
Hence, $\forall \varepsilon > 0$, $\exists \beta > 0$ such that $|\theta| < \beta \Rightarrow |\frac{v_S(x_0+h)-v_S(x_0)}{h} - \ell| < \varepsilon$. This means that the growth rate has a limit $\ell \in \mathbb{R}$ when $\theta \to 0$, which is a contradiction. Hence, under the assumptions (H_1), v_S is not differentiable at x_0.

2^{nd} case: Assume that F satisfies the assumptions (H_2).
As in the first case, we define x_0 such that $t_0 = F^{-1}(x_0)$. The graph of G :

- coincides with F on $[0, t_0]$,
- is deduced from the graph of F by the translation vector $(t_0 - t_1)\overrightarrow{i}$ on $[t_0, T - (t_1 - t_0)]$.

Thus, the graph of G is similar to the graph of F but removing the time period $[t_0, t_1]$ for which the function is constant. So, the same growth rate occurs at x_0, and if v_S is not differentiable at x_0 for one, it is not for the other. In other words, the results of the first case where $F' = 0$ at one point t_0 extend to the more general case where F' is zero on an interval $[t_0, t_1]$ $(t_0 \neq t_1)$, subject to the assumptions (H_1) on G. □

References

Andrieu, C., 2013. Modélisation fonctionnelle de profils de vitesse en lien avec l'infrastructure et méthodologie de construction d'un profil agrégé. Ph.D. thesis. Université Toulouse 3 Paul Sabatier.

Andrieu, C., Saint Pierre, G., Bressaud, X., 2012. Modélisation fonctionnelle et lissage sous contraintes d'un profil spatial de vitesse, in: 44ème Journées de Statistique, Bruxelles, Belgique.

Andrieu, C., Saint Pierre, G., Bressaud, X., 2013. Estimation of space-speed profiles: A functional approach using smoothing splines, in: Proceedings of the IEEE Intelligent Vehicle Symposium (IV 2013), Gold Coast, Australie, juin 2013.

Barbosa, H.M., Tight, M.R., May, A.D., 2000. A model of speed profiles for traffic calmed roads. Transportation Research Part A: Policy and Practice 34, 103–123.

Berlinet, A., Thomas-Agnan, C., 2004. Reproducing kernel Hilbert spaces in probability and statistics. Springer.

Bigot, J., 2003. Recalage de signaux et analyse de variance fonctionnelle par ondelettes : application au domaine biomédical. Ph.D. thesis. Universite Joseph Fourier, Grenoble I.

Bigot, J., 2006. Landmark-based registration of curves via the continuous wavelet transform. Journal of Computational and Graphical Statistics 15, 542–564.

Bigot, J., Gadat, S., 2010. Smoothing under diffeomorphic constraints with homeomorphic splines. SIAM Journal on Numerical Analysis 48(1), 224–243.

Boonsiripant, S., 2009. Speed profile variation as a surrogate measure of road safety based on GPS-equipped vehicle data. Ph.D. thesis. School of civil and environmental engineering, Georgia Institute of Technology, Atlanta.

Brunk, H.D., 1955. Maximum likelihood estimates of monotone parameters. Annals of Mathematical Statistics 26, 607–616.

Calderon, C., Martinez, J., Carroll, R., Sorensen, D., 2010. P-splines using derivative information. Multiscale Modeling & Simulation 8, 1562–1580.

Cox, D., 1988. Approximation of method of regularization estimators. The Annals of Statistics , 694–712.

Cuevas, A., 2013. A partial overview of the theory of statistics with functional data. Journal of Statistical Planning and Inference .

Cuevas, A., Febrero, M., Fraiman, R., 2006. On the use of the bootstrap for estimating functions with functional data. Computational statistics & data analysis 51, 1063–1074.

Cuevas, A., Febrero, M., Fraiman, R., 2007. Robust estimation and classification for functional data via projection-based depth notions. Computational Statistics 22, 481–496.

De Boor, C., 2001. A Practical Guide to Splines. Applied mathematical sciences, Springer.

Delecroix, M., Simioni, M., Thomas-Agnan, C., 1996. Functional estimation under shape constraints. Journal of Nonparametric Statistics 6, 69–89.

Delecroix, M., Thomas-Agnan, C., 2000. Spline and kernel regression under shape restrictions, in: Smoothing and Regression: Approaches, Computation, and Application. Wiley Online Library.

Duchon, J., 1977. Splines minimizing rotation-invariant semi-norms in sobolev spaces, in: Constructive theory of functions of several variables. Springer, pp. 85–100.

Ericsson, E., 2000. Variability in urban driving patterns. Transportation Research Part D 5, 337–354.

Eubank, R., 1999. Nonparametric regression and spline smoothing. Statistics, textbooks and monographs, Marcel Dekker.
Ferraty, F., Vieu, P., 2006. Nonparametric Functional Data Analysis: Theory and Practice. Springer Series in Statistics, Springer-Verlag, New York.
Fraiman, R., Muniz, G., 2001. Trimmed means for functional data. Test 10, 419–440.
Green, P., Silverman, B., 1994. Nonparametric regression and generalized linear models: a roughness penalty approach. Monographs on statistics and applied probability, Chapman & Hall.
Gu, C., 2002. Smoothing spline ANOVA models. Springer Verlag.
Hall, P., Yatchew, A., 2007. Nonparametric estimation when data on derivatives are available. The Annals of Statistics 35, 300–323.
Hastie, T., Tibshirani, R., 1990. Generalized additive models. Monographs on statistics and applied probability, Chapman & Hall.
Hyden, C., Varhelyi, A., 2000. The effects on safety, time consumption and environment of large scale use of roundabouts in an urban area: a case study. Accident Analysis & Prevention 32, 11–23.
Jun, J., Guensler, R., Ogle, J., 2006. Smoothing methods to minimize impact of global positioning system random error on travel distance, speed, and acceleration profile estimates. Transportation Research Record: Journal of the Transportation Research Board 1972, 141–150.
Kerper, M., Wewetzer, C., Mauve, M., 2012. Analyzing vehicle traces to find and exploit correlated traffic lights for efficient driving, in: Intelligent Vehicles Symposium (IV), 2012 IEEE, pp. 310–315.
Kimeldorf, G., Wahba, G., 1971. Some results on tchebycheffian spline functions. Journal of Mathematical Analysis and Applications 33, 82–95.
Kneip, A., Gasser, T., 1992. Statistical tools to analyze data representing a sample of curves. The Annals of Statistics , 1266–1305.
Laureshyn, A., 2005. Automated video analysis and behavioural studies based on individual speed profiles, in: Proceedings of 18th ICTCT, Helsinki, October 2005.
Laureshyn, A., Åström, K., Brundell-Freij, K., 2009. From speed profile data to analysis of behaviour. IATSS research 33, 89.
Levitin, D.J., Nuzzo, R.L., Vines, B.W., Ramsay, J.O., 2007. Introduction to functional data analysis. Canadian Psychology 48, 135–155.
López-Pintado, S., Romo, J., 2009. On the concept of depth for functional data. Journal of the American Statistical Association 104, 718–734.
Louah, G., 2005. The accuracy of a speed profile estimation method combining continuous and spot speed measurements, in: Road Safety on Four Continents: 13th International Conference, Warsaw, Poland.
Mammen, E., Marron, J., Turlach, B., Wand, M., 2001. A general projection framework for constrained smoothing. Statistical Science , 232–248.
Mardia, K., Kent, J., Goodall, C., Little, J., 1996. Kriging and splines with derivative information. Biometrika 83, 207–221.
Meinguet, J., 1979. Multivariate interpolation at arbitrary points made simple. Zeitschrift für angewandte Mathematik und Physik ZAMP 30, 292–304.
Nygard, M., 1999. A method for analysing traffic safety with help of speed profiles. Master's thesis. Tampere University of Technology, Department of Civil Engineering, Finlande.
Porsche InnoDrive, 2013. http://www.porsche.com/international/aboutporsche/responsibility/environment/technology/porscheinnodrive/.
Quiroga, C.A., Bullock, D., 1998. Travel time studies with global positioning and geographic information systems: an integrated methodology. Transportation Research Part C: Emerging Technologies 6, 101–127.
R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.
Rakha, H., Dion, F., Sin, H., 2001. Using global positioning system data for field evaluation of energy and emission impact of traffic flow improvement projects: Issues and proposed solutions. Transportation Research Record: Journal of the Transportation Research Board 1768, 210–223.
Ramsay, J., 1988. Monotone regression splines in action. Statistical Science , 425–441.
Ramsay, J., 1998. Estimating smooth monotone functions. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60(2), 365–375.
Ramsay, J., Li, X., 1998. Curve registration. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60, 351–363.
Ramsay, J.O., Silverman, B.W., 2002. Applied Functional Data Analysis: Methods and Case Studies. Springer Series in Statistics, Springer-Verlag, New York.
Ramsay, J.O., Silverman, B.W., 2005. Functional Data Analysis, Second Edition. Springer-Verlag, New York.
Sun, Y., Genton, M., 2011. Functional boxplots. Journal of Computational and Graphical Statistics 20.
Tukey, J., 1977. Exploratory data analysis. Reading, MA 231.
Várhelyi, A., Hjälmdahl, M., Hyden, C., Draskóczy, M., 2004. Effects of an active accelerator pedal on driver behaviour and traffic safety after long-term use in urban areas. Accident Analysis & Prevention 36, 729–737.
Violette, E., Hublart, A., Subirats, P., Louah, G., 2010. Estimation de la vitesse pratiquée sur un itinéraire - Méthode et application. Technical Report. CETE Normandie Centre, DITM, ESM.
Wahba, G., 1990. Spline models for observational data. SIAM.
Wahba, G., Wendelberger, J., 1980. Some new mathematical methods for variational objective analysis using splines and cross-validation. Monthly weather review 108, 1122–1145.
Wang, K., Gasser, T., 1997. Alignment of curves by dynamic time warping. The Annals of Statistics 25, 1251–1276.
Wang, Y., 2011. Smoothing Splines: Methods and Applications. Monographs on Statistics and Applied Probability, Chapman & Hall/CRC Press.
Zuo, Y., Serfling, R., 2000. General notions of statistical depth function. Annals of Statistics , 461–482.