Curve clustering based on second order information: application to bad runway condition detection

Stéphane Puechmorel, Florence Nicol, Baptiste Gregorutti, Cindie Andrieu

▶ **To cite this version:**

Stéphane Puechmorel, Florence Nicol, Baptiste Gregorutti, Cindie Andrieu. Curve clustering based on second order information: application to bad runway condition detection. 2017. hal-01799419v1

HAL Id: hal-01799419

https://enac.hal.science/hal-01799419v1

Preprint submitted on 24 May 2018 (v1), last revised 10 Dec 2023 (v2)

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.

Curve clustering based on second order information: application to bad runway condition detection

Stéphane Puechmorel[a], Florence Nicol[a], Baptiste Gregorutti[b], Cindie Andrieu[b]

[a] *Université Fédérale de Toulouse, ENAC, 31055 TOULOUSE, France*
[b] *Safety Line, 75015 PARIS, France*

Abstract

In air transportation, a huge amount of data is continuously recorded such as radar tracks that may be used for improving flight as well as airport safety. However, all known statistical algorithms, even those based on functional data, are unable to distinguish between a safety critical flight and another one departing from standard behavior, but otherwise safe. It is the case in airport safety when radar measurements are used for detecting incidents on airport surface. In this paper, we propose a change of paradigm by switching from a functional data framework to a geometrical one by representing curves as points in a shape manifold. In this way, any intrinsic structure of the data that is amenable to geometry can be directly encoded in the representation space. Based on an extension of a classical distance between shapes, a new one is defined, that explicitly takes into account the second derivative and can be related to slippery. Its properties are investigated in a first part, then some results on datasets of synthetic and real trajectories are presented.

Keywords: curve clustering, outlier detection, similarity measure, shape manifold, functional data analysis, air traffic management, airport safety.
PACS:

1. Introduction

Based on recent studies, European air traffic is expected to grow, yielding an increase of 1.2 million flights in 2023, 14% more than in 2016 [1]. Major framework programs such as SESAR (Single European Sky Air traffic management Research) in Europe aim to improve air traffic management (ATM) by investigating innovative solutions on management and traffic flow analysis. In the context of aviation safety, a crucial issue consists in assessing runway adherence condition and detecting incidents on airport surface by observing only the radar tracks of landing aircraft, e.g. trajectory deviation and abnormal deceleration may be good indicators of runway bad state. Indeed, landing aircraft must brake in a quite short time, putting a stringent condition on the

adherence coefficient of the runways. Bad weather conditions like rain, snow and icing can dramatically lower it, increasing the landing distance and making maneuvers more difficult to perform. Rubber deposited by the wheels during braking may also impair the friction coefficient of the runway making it more slippery. To estimate the runway adherence condition, a direct measurement of it on the pavement is usually performed: it implies sending a vehicle with a dedicated tool, which will interfere with the ongoing traffic. The objective is to develop spatio-temporal similarity measures for trajectories that will be used in clustering or outliers detection methods.

Aircraft trajectories are functional objects mapping time to position, even if we will most of the time observe discretized samples of trajectories, such as radar measurements. The mathematical problem arising from the application presented above falls within the general framework of curve clustering, which is quite common in functional data statistics. Several works were dedicated to the extension of multivariate algorithms to sample paths of Hilbert processes. As a starting point, data is first expanded on a truncated Hilbert basis [2] then the vectors of expansion coefficients enters a standard finite dimensional analysis. A clever choice of the representation space and basis allows to take into account the a priori knowledge about the studied process. Unfortunately, the dimension of the samples produced that way may be high, and varies with the geometric features of the sample paths. In particular, the presence of high curvature values will increase the number of expansion coefficients needed to keep a good approximation of the original function. In [3], an EM functional clustering algorithm is presented with adaptive basis in each group, yielding an efficient numerical method to deal with this issue. Another class of methods relies on a non-parametric approach [4, 5]. A recent work [6] pertaining to this approach presents a hierarchical clustering principle, with application to electric power consumption. Moreover, the choice of the L^p metric as a measure of similarity between trajectories may be problematic in many situations and does not reflect the high internal structure of aircraft trajectories.

For overcoming these difficulties, a solution consists in representing functional data in a Riemannian framework. Indeed, curves can be made amenable to functional data statistics by representing them as points on the so-called shape manifold, that is formally defined as a quotient of the manifold of immersions. It is not a Hilbert space, but it can be provided with Riemannian metrics that allow geodesic distance computation, although some care must be taken in order to avoid degeneracy of the metric [7]. In the shape space literature, curves are considered as geometric objects and are represented in the so-called shape space, that is the set of immersions quotiented by the group of smooth diffeomorphisms. Some algorithms use a shape manifold [8] representation in order to derive a metric between sample paths. Note that this kind of metric verify the reparametrization invariance property. Having it at hand, clustering may be performed in a standard way. It worth mention that most of the time a mean of computing the centroid of a set of curves is mandatory.

Considering methods falling in this category, it appears that the require-

ments of the landing trajectory analysis are not fulfilled: enforcing a full parametrization invariance as in the case of shape based approaches prevents the use of longitudinal acceleration that enters the non-slip condition. On the other side, Sobolev-like metrics are designed to be tailored to specific needs. Motivated by a real use case where one wants to assess runway adherence condition by observing only the radar tracks of landing aircraft, we propose in this paper to use a Riemannian framework for curves with velocity information. The purpose is to introduce a new kind of Riemannian metric that is especially adapted to the study of curves where the velocity is a discriminating feature. In such a case, the original shape space approach cannot be used since the parametrization invariance will wipe out the velocity information. A partial parametrization invariance is introduced, yielding a bundle shape space model on which a relevant metric can be defined. The design of the metric was based on the equations of motion and reflects the internal structure of the data. The equation of geodesics will be given, along with a practical computation algorithm based on a shooting method.

The paper is organized as follows. In Section 2, we formally describe the theoretical framework and define an adapted metric taking into account the velocity information. In Section 3, the numerical implementation of the new metric and the K-medoids clustering method [9, 10] is described in a general setting. This algorithm allows to implement another metric than the usual Euclidean distance. Next, Section 4 illustrates the performance of the K-medoids algorithm with the new metric for low adherence detection on a set of simulated and real trajectories. These results are also compared with competing distances. Finally, some comments and future works are drawn in Section 5.

2. Theoretical framework

This section is dedicated to the definition of a relevant metric for curves taking into account velocity information. The framework presented here will overcome the limitations of the shape approach through the use of a metric in which the slip characteristic of the landing path is explicitly considered.

2.1. Slip detection

Landing aircraft may experience slip during deceleration phase when the runway is in degraded conditions. It may result from icing, snow, bad runway surface state but also from pilot's actions, namely a too high braking action or a sharp turn. In this last case, it is not related to runway condition and must not trigger a maintenance action from the airport services.

Slip can be detected on-board by comparing wheel rotation rate with aircraft velocity and computing the so-called wheel slip factor:

$$\lambda = \frac{\omega_w - \omega_a}{\max(\omega_w, \omega_a)} \tag{1}$$

where ω_w is the wheel angular velocity and $\omega_a = V_a/R_w$ is the expected angular velocity that can be computed as the ratio of the aircraft velocity to the wheel radius. Please note that on the real vehicle, several wheels are used, and the λ coefficient has to be understood as a mean value. Furthermore, due to tire elasticity, λ is not zero even if there is no actual slip: this is due to the fact that when a traction or a braking force is applied, the rubber will stretch, resulting in the tire outer part actually traveling more or less than expected from rigid body dynamics. This information is not yet downlinked in real time to ground centers and thus cannot be used in the intended application. From the ground standpoint, λ cannot be observed without on-board information, but some aspects of the landing or taxiing aircraft behavior may still be inferred. It is assumed in the sequel that Coulomb's law for friction [11] is applicable, so that the contact force F_c depends only on aircraft weight and tire/runway conditions:

$$F_c \leq \mu g M \tag{2}$$

with M the aircraft mass, g the gravity of Earth and μ the adhesion coefficient. Without slip, μ is equal to the static friction coefficient μ_s and F_c can be increased until it reaches the upper bound in 2. At that point, slip occurs and μ drops to the value of the dynamic friction coefficient μ_d. F_c remains constant until it falls below $\mu_d g M$. In real world experiments, this simple behavior is no longer valid and one has to expression μ as a function of λ [12]. Within this frame, the expression of the contact force is $F_c = \mu(\lambda) g M$, which is valid for both non-slip and slip case. Furthermore, in the case of aircraft, aerodynamics forces are exerted, with a net result of a braking force F_a that adds to the actual brakes action, but does not contribute to the friction analysis. Putting things together, the equation of motion along the aircraft trajectory γ can be expressed as:

$$\ddot{\gamma}(t) = \frac{F_a(t)}{M} + \mu(\lambda(t)) g \vec{u} \tag{3}$$

where $\vec{u}$ is a unit vector in the direction of the contact force F_c. Without making additional assumptions, it is not possible to use (3) for slip detection. However, if actions taken are assumed to be optimal, then F_a and $\vec{u}$ will be collinear so as to maximize the net braking effect. The expression of the aircraft dynamics becomes:

$$\ddot{\gamma}(t) = \left(K(t) + \mu(\lambda(t)) g\right) \vec{u} \tag{4}$$

where the coefficient $K(t)$ accounts for the aerodynamic braking force intensity. As aircraft must loose speed fast, μ will be close to the maximum at least during the landing and the beginning of taxi. The same applies for K, as it will not impair adherence. It can then be deduced that aircraft will try to keep the ratio between longitudinal and normal acceleration as high as possible. An observable measurement of slip condition can be deduced from the previous remark by considering the angle θ between the acceleration and speed vectors:

$$\sin(\theta(s)) = \frac{\kappa(s) \|D_s\gamma(s)\|^2}{\|D_{ss}\gamma(s)\|} = \frac{\det\left(D_s\gamma(s), D_{ss}\gamma(s)\right)}{\|D_s\gamma(s)\| \|D_{ss}\gamma\|} \tag{5}$$

where γ is the aircraft trajectory and κ its curvature. In the above expression and through all the document, the symbol D_s stands for the partial derivative with respect to variable s. Higher order derivatives are written similarly as $D_{s_1 \ldots s_1, s_2 \ldots s_2, \ldots}$ by repeating the variable p times to indicate a partial derivative of order p.

In good runway conditions, the longitudinal acceleration will be high and nearly constant, at least in the first part of the landing trajectory. As a consequence, one can expect θ to be relatively small and be proportional to $\det(D_s\gamma, D_{ss}\gamma)$. Reciprocally, under slip conditions, a trade off has to be made between path following and deceleration: the angle θ will thus increases towards the limiting value $\pm\pi/2$.

Please note that θ is not defined in the portions of the trajectory where the acceleration or the velocity vanish. While the second situation is highly uncommon, as it will indicate a stop during landing or taxiing, the first one can be encountered when the aircraft is not braking nor turning. In such parts of the trajectory, it is not possible to infer an adherence condition, and curve comparison can be done only on a geometrical basis. In the addressed application, aircraft in the observed part of the landing trajectory are decelerating, so that the above issue is not a concern.

2.2. An adapted metric in the space of trajectories

Recall that a smooth curve $\gamma\colon [0,1] \to \mathbb{R}^2$ will be an immersion when the derivative $D_s\gamma$ is everywhere non vanishing in $]0,1[$. The set of such curves will be denoted by $\mathbf{Imm}([0,1],\mathbb{R}^2)$. It has the structure of Banach manifold, its tangent space at $\gamma \in \mathbf{Imm}([0,1],\mathbb{R}^2)$ being the vector space $C^\infty([0,1],\mathbb{R}^2)$. A tangent vector is thus a couple (γ, h) with γ the base curve and h an element of $C^\infty([0,1],\mathbb{R}^2)$ that is interpreted as an infinitesimal displacement field along γ. This point can be formalized using the notion of admissible variation:

Definition 1. Let γ be a smooth curve. An admissible variation of γ is a smooth mapping $\Phi\colon]-\epsilon,\epsilon[\to \mathbb{R}^2, \epsilon > 0$, such that $\Phi(0,\bullet) = \gamma(\bullet)$ and $\forall t \in]-\epsilon,\epsilon[, \Phi(t,0) = \gamma(0), \Phi(t,1) = \gamma(1)$.

An admissible variation defines a tangent vector $(\gamma, D_t\Phi|t=0)$. The extension to more general immersions is quite straightforward [7]. In the same reference, the variation formula is used to derive a Riemannian metric on the quotient space $\mathbf{Imm}(\mathbb{S}^1,\mathbb{R}^2)/\mathbf{Diff}([0,1],\mathbb{R}^2)$.

In the present work, a similar approach will be taken. However, due to the fact that the slip condition must come into play, it is not meaningful to keep invariance under change of parametrization. Instead, a weaker invariance by affine change of parametrization will be obtained at the end. Furthermore, curves with vanishing second derivative must be excluded since the slip angle θ in (5) is not defined at points where $D_{ss}\gamma(s) = 0$. The last condition boils down to the requirement that the curve $s \in [0,1] \mapsto (\gamma, D_s\gamma)$ be an immersion. The space of such objects will be denoted by $\mathbf{Imm}^*([0,1],\mathbb{R}^2)$. An mentioned above,

this is not an issue for the application, due to the restriction of the braking part of the trajectory. In a more general setting, it may be necessary to segment curves into non vanishing second derivative parts and compare only them.

The slip angle in (5) has a nice variational interpretation as indicated in the next lemma.

Lemma 1. *Let $\gamma\colon [0,1] \to \mathbb{R}^2$ be a smooth path and Φ an admissible variation of it. Let ϕ be a smooth path such that $\phi(0) = \gamma(1)$, $\phi(1) = \gamma(0)$. Then:*

$$D_t A(0) = \int_0^1 \det\left(D_t\Phi(0,s), D_s(0,s)\right) ds \tag{6}$$

where $A(t)$ is the net area enclosed by the loop $\Phi(t,\bullet), \phi$ for $t \in]-\epsilon,\epsilon[$.

Proof. For any $t \in]-\epsilon,\epsilon[$ the loop $\Phi(t,\bullet), \phi$ is piecewise smooth so that the net area $A(t)$ can be computed using Stoke's formula:

$$A(t) = \frac{1}{2}\int_0^1 \det(\Phi(t,s), D_s\Phi(t,s)ds + \frac{1}{2}\int_0^1 \det(\phi(s), D_s\phi(s)ds \tag{7}$$

Taking the derivative with respect to t yields:

$$D_t A(t) = \frac{1}{2}\left(\int_0^1 \det(D_t\Phi(t,s), D_s\Phi(t,s)ds + \int_0^1 \det(\Phi(t,s), D_{st}\Phi(t,s)ds\right). \tag{8}$$

Recalling that the end points of $\Phi(t,\bullet)$ are fixed, one can use integration by parts to get:

$$D_t A(t) = \frac{1}{2}\left(\int_0^1 \det(D_t\Phi(t,s), D_s\Phi(t,s)ds + -\int_0^1 \det(D_s\Phi(t,s), D_t\Phi(t,s)ds\right) \tag{9}$$

and finally, using the alternating property of the determinant:

$$D_t A(t) = \int_0^1 \det(D_t\Phi(t,s), D_s\Phi(t,s)ds \tag{10}$$

and the conclusion follows by letting $t = 0$. □

Using Lemma 1, the integral :

$$\int_0^1 \frac{|\det\left(D_s\gamma(s), D_{ss}\gamma(s)\right)|}{\|D_s\gamma(s)\|\|D_{ss}\gamma\|}\|D_s\gamma(s)\|ds \tag{11}$$

may be interpreted as the total infinitesimal area swept by the curve γ when moved in the direction $D_{ss}\gamma$.

The integral in (11) is invariant under affine change of parametrization. Given a curve γ defined on an arbitrary interval $[a,b]$, it is thus possible to go back to fixed interval $[0,1]$. As a consequence, any curves will be assumed to

be defined on $[0,1]$. Please note however that general invariance under change of parametrization cannot be obtained, and will not be meaningful as the information sought after is clearly dependent on the velocity.

The slip angle θ can be obtained in a convenient way using the fact that the curves take their values in $\mathbb{R}^2$ and can thus also be considered as complex valued mappings. Within this frame, it comes for the expression of the slip angle θ if curve γ:

$$e^{i\theta} = \frac{D_s\gamma}{|D_s\gamma|}\frac{\overline{D_{ss}\gamma}}{|D_{ss}\gamma|}. \tag{12}$$

Using the complex representation, the following proposition gives access to the variation of the slip angle.

Proposition 1. *Let $\Phi\colon]-\epsilon,\epsilon[\times[0,1]to\mathbb{C}$ be an admissible variation of γ. Let $\theta(t,s)$ be the slip angle at s of the curve $s\mapsto\Phi(t,s)$. Then:*

$$D_te^{i\theta} = \frac{1}{|D_s\Phi|}D_sD_t\Phi|_N\frac{D_{ss}\overline{\Phi}}{|D_{ss}\Phi|} + \frac{1}{|D_{ss}\Phi|}\overline{D_{ss}D_t\Phi|_{\tilde{N}}}\frac{D_s\Phi}{|D_s\Phi|}$$

where the notation $|_N$ (resp. $|_{\tilde{N}}$) denotes the component normal to the curve $s\mapsto\Phi(t,s)$ (resp. $s\mapsto D_s\Phi(t,s)$).

Proof. The proposition is a direct computation of the derivative. The only point to note is that:

$$D_t\left(\frac{D_s\Phi}{|D_s\Phi|}\right) = \frac{D_sD_t\Phi}{|D_s\Phi|} - \frac{D_s\Phi}{|D_s\Phi|}\frac{\Re\left(D_sD_t\Phi.D_s\overline{\Phi}\right)}{|D_s\Phi|^2} \tag{13}$$

$$= \frac{1}{|D_s\Phi|}D_sD_t\Phi|_N\frac{D_{ss}\overline{\Phi}}{|D_{ss}\Phi|} \tag{14}$$

and the equivalent derivation for $D_{ss}\Phi$. □

Proposition 2. *Under the same assumptions as in Proposition 1:*

$$(D_t\theta)^2 = \left(\frac{|D_sD_t\Phi|_N|}{|D_s\Phi|} - \frac{|D_{ss}D_t\Phi|_{\tilde{N}}|}{|D_{ss}\Phi|}\right)^2$$

Proof. Using the fact that $|e^{i\theta}| = 1$ and Proposition 1, it comes:

$$\begin{aligned}(D_t\theta)^2 = {} & \frac{1}{|D_s\Phi|^2}|D_sD_t\Phi|_N|^2 + \frac{1}{|D_{ss}\Phi|^2}|D_{ss}D_t\Phi_{\tilde{N}}|^2 \\ & + 2\frac{1}{|D_s\Phi||D_{ss}\Phi|}\Re\left(D_sD_t\Phi|_N.T.D_{ss}D_t\Phi|_{\tilde{N}}\overline{\tilde{T}}\right)\end{aligned}$$

with:

$$T = \frac{D_s\Phi}{|D_s\Phi|}, \qquad \tilde{T} = \frac{D_{ss}\overline{\Phi}}{|D_{ss}\Phi|}.$$

Since the curve is planar, $N = iT$ (resp. $\tilde{N} = i\tilde{T}$). This combined with the expression of the normal component $D_s D_t \Phi|_N = |D_s D_t \Phi|_N | N$ (resp. $D_{ss} D_t \Phi|_{\tilde{N}} = |D_{ss} D_t \Phi|_{\tilde{N}} | \tilde{N}$) yields:

$$\Re \left(D_s D_t \Phi|_N . T . D_{ss} D_t \Phi|_{\tilde{N}} \overline{\tilde{T}} \right) = -|D_s D_t \Phi|_N || D_{ss} D_t \Phi|_{\tilde{N}}|$$

and the result follows. □

An admissible variation Φ of a curve γ defines a tangent vector to γ in the space $\mathbf{Imm}^*([0,1], \mathbb{R}^2)$ by $D_t \Phi|_{t=0}$. Such a tangent vector is thus a couple (γ, u) where $u \in C^\infty([0,1], \mathbb{R}^2$. The above expression for $(D_t \theta)^2$ allows to define a semi-norm in the tangent space to γ.

Definition 2. Let (γ, u) be a tangent vector to γ. The semi-norm $M_{\gamma(u)}$ is defined by:

$$M^2_{\gamma(u)} = \int_0^1 \left(\frac{|D_s u|_N|}{|D_s \gamma|} - \frac{|D_{ss} u|_{\tilde{N}}|}{|D_{ss} \gamma|} \right)^2 ds$$

with N (resp. $\tilde{N}$) the normal vector to γ (resp. $D_s \gamma$).

Given two curves γ_1, γ_2, a path between them in $\mathbf{Imm}^*([0,1], \mathbb{R}^2)$ is smooth mapping $\Phi \colon [0,1] \times [0,1] \to \mathbb{R}^2$ such that:

- $\Phi(0, \bullet) = \gamma_1(\bullet), \Phi(0, \bullet) = \gamma_2(\bullet)$
- For all t in $]0,1[$, $\Phi(t, \bullet) \in \mathbf{Imm}^*([0,1], \mathbb{R}^2)$.

Definition 3. Let Φ be a path between γ_1 and γ_2. Its θ-deformation energy is defined as:

$$\begin{aligned} E(\Phi) = & \int_0^1 |D_t \Phi(t,0)|^2 + |D_t \Phi(t,1)|^2 dt \\ & + \int_0^1 \int_0^1 \left(\frac{|D_s D_t \Phi|_N|}{|D_s \Phi|} - \frac{|D_{ss} D_t \Phi|_{\tilde{N}}|}{|D_{ss} \Phi|} \right)^2 ds dt \end{aligned}$$

The first term is needed as the second part involves only derivatives of the curve and will not take endpoints into account. The above definition of the energy is valid for curves in $\mathbf{Imm}^*([0,1], \mathbb{R}^2)$, but is not invariant by affine change of parametrization. Furthermore, it is based on a semi-metric, resulting in possible degeneracies. The path energy is thus modified according to Definition 4.

Definition 4. Let Φ be a path between γ_1 and γ_2. Its total deformation energy is defined as:

$$\begin{aligned} \tilde{E}(\Phi) = & \int_0^1 |D_t \Phi(t,0)|^2 + |D_t \Phi(t,1)|^2 dt \\ & + \int_0^1 \int_0^1 |D_s D_t \Phi|^2 |D_s \Phi| ds dt \\ & + \lambda \int_0^1 \int_0^1 \left(\frac{|D_s D_t \Phi|_N|}{|D_s \Phi|} - \frac{|D_{ss} D_t \Phi|_{\tilde{N}}|}{|D_{ss} \Phi|} \right)^2 |D_s \Phi| ds dt \end{aligned}$$

where $\lambda > 0$ tunes the relative importance of shape and slip angle.

The first double integral is a Sobolev Riemannian metric that accounts for shape deformation. The free parameter λ must be tuned in applications. The energy allows the computation of geodesic path between curves.

Definition 5. The geodesic path Φ_0 between γ_1 and γ_2 is, when it exists, defined by:

$$\Phi_0 = argmin_{\Phi \in I} \tilde{E}(\Phi)$$

where I is the set of paths from γ_1 to γ_2.

The energy of a geodesic path defines a similarity measure between curves. It will be used in the sequel within a clustering algorithm.

3. Numerical implementation

This section describes the numerical implementation of the path computation of the new metric described in Section 2.2. Then, the clustering procedure will be presented.

3.1. Geodesic path computation

Two approaches can be used for computing a geodesic path Φ_0 minimizing the total energy. The first one relies on a system of differential equations involving $D_t\Phi$ that is solved starting at $t = 0$ from the first curve. An iterative shooting procedure reduces gradually the distance at $t = 1$ to the second curve, yielding an approximate geodesic path. This algorithm is appealing as it does not requires an approximation to the entire path Φ_0, but just of the curve $s \mapsto \Phi(t, s)$ for a fixed t. Its two main drawbacks are first the need to establish the equation of geodesics, that may be quite complicated, and second the numerical instabilities in the shooting phase. It will nevertheless be investigated in a future work.

The second approach is to approximate the full path Φ by its samples $\Phi_{ij} = \Phi(x_{ij}), i = 0 \dots N, j = 0 \dots M$ on a grid defined by the points $x_{ij} = (t_j, s_i)$. In the current implementation, the grid is evenly spaced. A numerical differentiation procedure is applied to obtain the derivatives D_s, D_{ss}, D_t and takes the form of matrices DS, DSS, DT that are applied to the matrix $M = (\Phi_{ij})$ of path sample points. Since the derivative with respect to s must be applied columnwise, the derivative $D_s\Phi$ is approximated by $DS.M$. Conversely, derivation is performed rowwise for $D_t\Phi$, so that the approximation is $M.D_t$. Finally, the first and the last column of Φ are held constant, since they represent sampled version of the endpoint curves γ_1, γ_2.

Using the discrete approximation, the total energy of the sampled path described by matrix M is:

$$\begin{aligned}\tilde{E}(M) = &\sum_{i=0}^{M} |(M.DT)_{0i}|^2 + |M.DT_{Ni}|^2 \\ &+ \sum_{i=0}^{M}\sum_{j=0}^{N} |(DS.M.DT)_{ij}|^2 |(DS.M)_{ij}| \\ &+ \lambda \sum_{i=0}^{M}\sum_{j=0}^{N} \left(\frac{|(DS.M.DT|_N)_{ij}|}{|(DS.M)_{ij}|} - \frac{|(DSS.M.DT|_{\tilde{N}})_{ij}||}{|(DSS.M)_{ij}|} \right)^2 |(DS.M)_{ij}|\end{aligned}$$

The optimal approximate path is obtained by a standard numerical optimization algorithm, chosen here to be LM-BFGS [13] that is well suited to high dimensional problem. Initializing Φ with a linear homotopy ensure convergence in a few iterations on most problems. The minimal energy obtained is used in the following clustering phase.

3.2. K-medoids clustering

K-medoids algorithm [9, 10] is a clustering technique that aims to partition unlabelled data into clusters. Unlike the well known K-means algorithm [14] which is sensitive to outliers, K-medoids algorithm is more robust and use an actual point in the cluster to represent it. As a matter of fact, it is possible to perform the clustering with a pairwise distance matrix as input data instead of the raw data. Thereby, we can derive the algorithm with the metric defined in Section 2 in order to classify a set of n spatiotemporal trajectories $\{x_1, \ldots, x_n\}$ (Algorithm 1).

The most benefit of K-medoids algorithm compared to the K-means is that there is no need to explicity compute the cluster centers since we just need to store the indices $i_k^\star$. But similarly to K-means and other clustering techniques, the number of clusters K is assumed to be known. In practice, the knowledge of the data can help the user to estimate the value of K. In the case of aviation safety, one could choose $K = 3$ corresponding to three levels of alerts: low, medium, high.

In general, the number of clusters is not known and has to be computed from the data. In [16], the authors introduced the Silhouette criterion as a graphical method for interpreting a validating the results of a clustering algorithm. Broadly speaking, the Silhouette score measures the tightness and the separation of the clusters. More formally, for an observation x_i, it is defined as:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))},$$

where $a(i)$ is the distance between x_i and all other points in the same cluster and $b(i)$ is the distance between x_i and all other points in the next nearest

Algorithm 1 K-medoids algorithm, [15, Chap. 14]

Require: A sample of n trajectories $\{x_1, \ldots, x_n\}$ and the number of clusters K.

1: **repeat**

2: For a given cluster assignment C find the observation in the cluster minimizing the total distance to other points in that cluster:

$$i_k^\star = \underset{\{i:C(i)=k\}}{\arg\min} \sum_{C(j)=k} d(x_i, x_j).$$

Then $m_k = x_{i_k^\star}, k = 1, \ldots, K$ are the current estimates of the cluster centers.

3: Given a current set of cluster centers $\{m_1, \ldots, m_K\}$, minimize the total error by assigning each observation to the closest (current cluster) center:

$$C(i) = \underset{1 \leq k \leq K}{\arg\min}\, d(x_i, m_k).$$

4: **until** the assignments do not change.

cluster. The overall Silhouette score, which ranges between -1 and 1, is then the mean value of the score computed on each sample, i.e.

$$s = \frac{1}{n} \sum_{i=1}^{n} s(i).$$

Hence, the number of clusters K can be chosen by maximizing the Silhouette criterion. It can also be used as a validation metric of a clustering method: the higher value of s is, the better is the clustering.

Note that the mean values of $a(i)$ and $b(i)$ can be interpreted respectively as the intra-cluster variance and the inter-cluster variance. Knowing that, a higher values of s indicates that the clusters are dense and well separated while a lower value of s refers to an incorrect clustering. A score around zero indicates overlapping clusters.

Note also that the K-medoids can be sensitive to the initialization of the algorithm and may find a local minimum. This is actually a classical drawback of the clustering methods. A common solution is to repeat the clustering with different initializations and to retain the best one with respect to the Silhouette score.

4. Numerical experiments

In this section, the performance of the algorithm is validated and compared with other existing methods. The first subsection is dedicated to a dataset of simulated data. In the second subsection, we apply clustering procedures to a real dataset of landing aircraft tracks.

4.1. Simulation procedure

As already mentioned, it is virtually impossible to ensure that an observed trajectory that deviates from the nominal one is the result of a bad adherence condition. To perform such a task, a physical measurement is mandatory and the only way to correlate runway quality to landing tracks is to have recorded trajectories right after or before a measurement.

A more tractable approach is to simulate landing and taxiing in a realistic fashion, so that the detection performance of the methods can be estimated. Very good models of wheeled ground vehicles are available in the literature [12] and the case of braking aircraft has been considered in [17]. For the purpose of the study, a very detailed model is not needed, but a good conformance to the observations is required. The approach taken was to use the model of Figure 1. The aircraft is controlled using the engine thrust and the front wheel steering

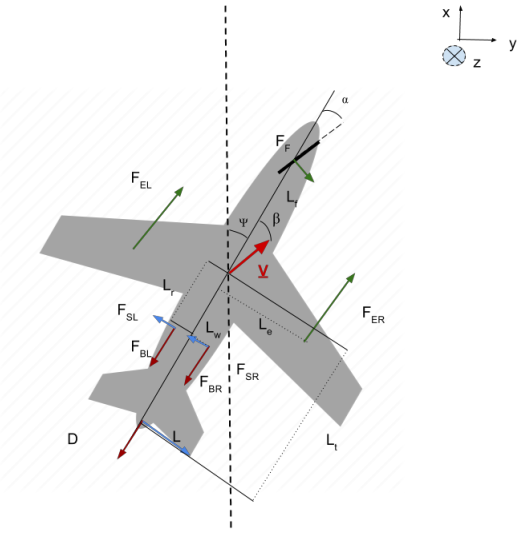


Figure 1: Aircraft physical model

angle. The landing gear tires train is replaced by a single equivalent tire. The simulation is performed by solving the equations of motion using a runge-kutta 45 scheme with adaptive step. The control loop consists of two PID regulators: the inner acts on the steering angle of the front wheel and corrects lateral deviations from the intended path while the outer controls deceleration. When the mobile is skating, the first control loop takes precedence and consumes the apart of the available friction force to minimize the lateral error. The remaining friction is transfered to the outer loop to apply some deceleration. It is clear that this process will increase the slip angle when trajectory is turning.

A simulator was coded in Java and applied to synthetic tracks with tunable adherence coefficient. Since slippery can be detected easily within the simulation loop, all output trajectories are flagged as "skating" or "normal".

4.2. Simulated data and comparison with competitors

Simulated trajectories

Landing and taxiing trajectories are simulated following a circle representing the nominal trajectory. The initial landing point is randomly disturbed as well as the initial velocity. In this way, it may induce a skating at the beginning of the landing trajectory which is then corrected by the regulators as described above. The dataset of simulated trajectories is composed of 16 skating trajectories and 284 normal trajectories.

The K-medoids clustering method described in Section 3.2 is performed on a set of 300 simulated landing trajectories. Two geometric metrics have been first used: the Mumford metric [7] and the slippery metric developed in Section 2.1. The first one is a geometric metric that is invariant by reparametrization and relevant for clustering trajectories with different morphology; this is the case when some aircraft trajectories deviate from the nominal one. However, the Mumford metric does not take into account the second derivative of trajectories. As an alternative, the slippery metric fits this requirement and should be able to detect abnormal deceleration levels that may indicate a potential degraded runway condition. Next, these two metrics are compared with competitor metrics described in [18]: some warping based distances such as DTW [19], LCSS [20], EDR [21] and ERP [22], and some shape-based distances, such as the Hausdorff distance [23], the Fréchet distance [24] and the SSPD distance [18]. The K-medoids clustering method is then performed in order to detect two clusters: the normal trajectories cluster and the skating trajectories one.

Table 1: number of trajectories

Metric	*Cluster 1*	*Cluster 2*
slippery	284	16
mumford	154	146
erp	284	16
dtw	143	157
edr	1	299
lcss	1	299
frechet	158	142
hausdorff	157	143
sspd	160	140

Table 1 displays the number of trajectories in each cluster. The three shape based distances (Hausdorff, Fréchet and SSPD), the warping based distance DTW and the geometric Mumford metric produce two relatively balanced classes. The two warping based distances EDR and LCSS badly perform: only one single trajectory is clustered apart from the others. Finally, the geometric slippery metric and the warping based distance ERP are the only one that could be able to correctly detect the two skating and normal trajectories clusters.

In Table 2, the output trajectories are flagged as "skating" or "normal" in each cluster. We remark that the geometric slippery metric and the warping based distance ERP perform well and are both be able to correctly detect the 16 skating trajectories cluster and the 284 normal trajectories one. The three shape based distances (Hausdorff, Fréchet and SSPD) and the geometric Mumford metric produce one cluster containing only normal trajectories. The second one is composed of all the skating trajectories, but also of a large number of normal trajectories: the false detection rate is high around 42%. Finally, by using the last two metrics DTW and SSPD, the false detection rate increases up to 48%.

Table 2: number of skating and normal trajectories

Metric	*Cluster 1*		*Cluster 2*	
	skating	*normal*	*skating*	*normal*
slippery	0	284	16	0
mumford	0	154	16	130
erp	0	284	16	0
dtw	5	138	11	146
edr	0	1	16	283
lcss	0	1	16	283
frechet	0	158	16	126
hausdorff	0	157	16	127
sspd	14	146	2	138

Slightly deformed simulated trajectories

In order to assess the robustness of the results obtained by using the geometric slippery metric and the shape based distance ERP, we have slightly deformed trajectories by using a random coefficient (around 1%) so that landing trajectories are now randomly flattened following an ellipsoid.

Table 3: number of trajectories

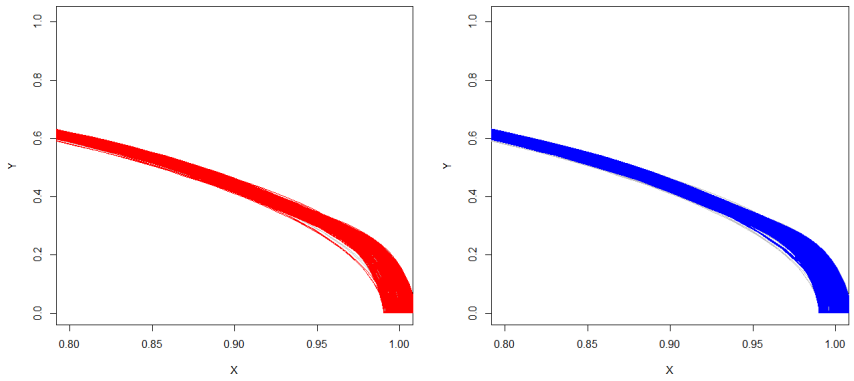
Metric	*Cluster 1*	*Cluster 2*
geometric	284	16
erp	130	170

Table 4: number of skating and normal trajectories

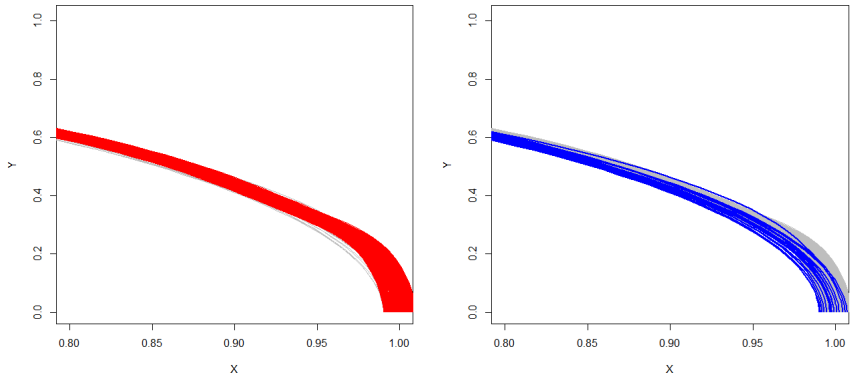
Metric	*Cluster 1*		*Cluster 2*	
	skating	*normal*	*skating*	*normal*
geometric	0	284	16	0
erp	0	130	16	154

Table 3 and Table 4 show that the shape based distance ERP is no longer able

to correctly detect the two clusters: the false detection rate now reaches more than 50%. In Figures 2a and 2b, we can visualize the K-medoids clustering with ERP distance and geometric slippery metric.



(a)



(b)

Figure 2: K-Medoids clustering with (a) ERP distance and (b) slippery metric.

4.3. Application to Radar tracks

In this section, we apply the K-medoids clustering method to a database of radar tracks in order to detect bad runway condition and skid situations.

Description of the data

The Radar tracks used here are a set of landings comming from A-SMGCS[1] which is a system that propose to air traffic controllers services to maintain airport safety and capacity. These data that are generated by sensor fusion of Primary Surveillance Radar (PSR), Secondary Surveillance Radar (SSR) using onboard transponders and multilateration Automatic Dependent Surveillance Broadcast (ADS-B).

Each landing trajectory T is defined as a set of spatiotemporal data, i.e. includes GPS position x_k, y_k and the aircraft speed v_k:

$$\begin{aligned} T &= (p_1, \ldots, p_n) \\ &= (\gamma(t_1), \ldots, \gamma(t_n)), \end{aligned}$$

where $p_k = (x_k, y_k, v_k)$ and $\gamma : [0, 1] \longrightarrow \mathbb{R}^3$. Note that the deceleration values is derived from the aircraft speed as a preprocessing step. The dataset consists of 357 landing trajectories.

We have chosen to focus on the runway exit curve only since slippery is mainly visible in curves. Then, following experts, as a skid situation is a loss of grip with the runway, a degraded runway condition is defined as follows:

- The aircraft deviate from the centreline
- The deceleration value is greater than the runway grip coefficient (which is low when the runway is slippery)

Following this definition, a skid situation is not only reduced to a lateral deviation of the trajectory. The aircraft deceleration value is indeed used as the indicator of the runway state. A low deceleration level indicates that the runway does not respond to the brakes.

Clustering-based bad runway condition detection

The K-medoids clustering method detailed in Section 3.2 is performed on the Radar tracks data. Two metrics are used: the Mumford metric [7], i.e. without taking into account the deceleration of the aircraft and the slippery metric defined in Section 2.

The Figure 3 shows that using the geometric metric (without the deceleration part [7]), the trajectories are well clustered by the algorithm. Indeed, the cluster 4 contains mostly trajectories with a large lateral deviation, while the cluster 3 corresponds to aircraft trajectories following the runway centreline (reference trajectory). Cluster 2 contains mostly standard trajectories even if there is several trajectories with slight lateral deviations. Cluster 1 corresponds to outliers trajectories.

[1] http://www.eurocontrol.int/articles/advanced-surface-movement-guidance-and-control-systems-smgcs

Moreover, as explained before, if lateral deviations in landing trajectories correspond to abnormal behavior, it does not necessarily represent a skid situation. In practice, the detection of bad runway conditions has to take into account the deceleration profile (see Section 2).

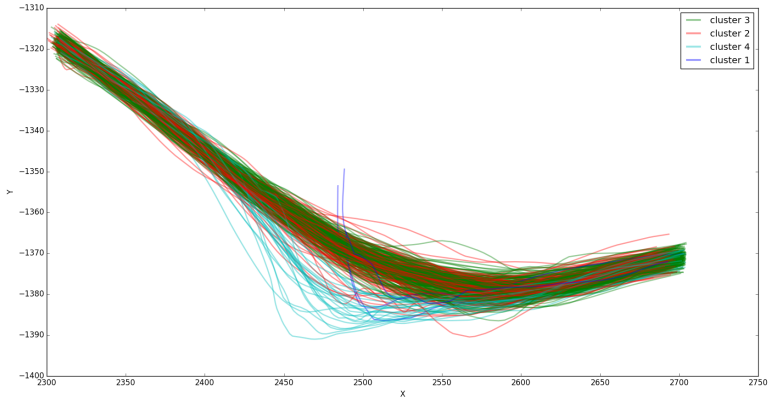
When adding the information of the deceleration in the metric, the results are slightly different. Indeed, in a geometric point of view, results are quite similar to the one described with the Mumford metric. However, if the cluster 2 contains the most deviant trajectories, we observed one deviant trajectory in the cluster 4 (3a).

After a deep investigation, we found that this particular trajectory cannot be considered as a slipped trajectory. Indeed, the figure 5 compares this trajectory and its deceleration profile with the medoid of the cluster 2 (deviant trajectories). Again the blue trajectory deviates from the red one but the deceleration profiles are significantly different: the deceleration level of the blue profile is higher than the red medoid deceleration profile. It means that the runway is not slippery since there is a sufficient grip to obtain a good braking action. From an operational point of view, it seems the aircraft had a high speed when arriving to the runway exit which led to a high braking.

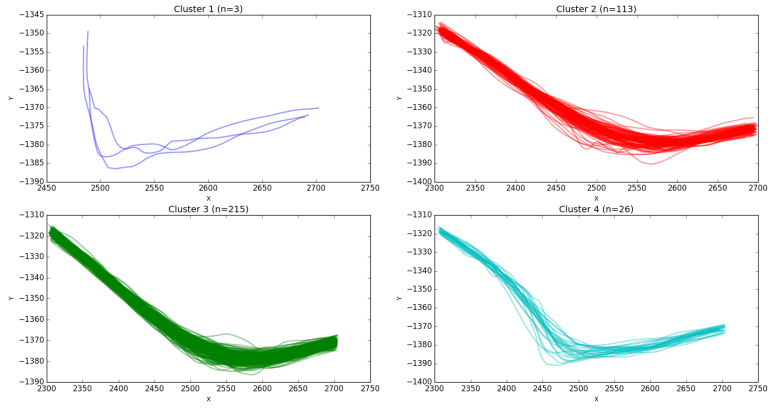
Briefly speaking, unlike the Mumford metric which is only based on the geometric shape, the benefits of the metric with deceleration is twofold: firstly, it detects lateral deviations, and secondly, it removes false positives by taking into account the deceleration part. Finally, an expert study shows that there was 3.7% of slippery alerts in the dataset among which 92.3% are in the cluster 2 (red trajectories). This results confirms the good performance of the proposed metric.

5. Conclusion and future work

The ability to detect bad runway condition without resorting to intrusive on-site measurements will be of major interest both for economic considerations and for passenger experience. In the future, aircraft will be able to downlink their flight and taxi parameters, but until then, only radar tracks may be used. Clustering landing trajectories into skating and normal ones falls within the frame of functional data. However, some extra information must be added to really distinguish between trajectories experiencing slipping and those resulting from late breaking action of the pilot. In this work, a new distance between curves especially tailored for runway bad adherence condition has been introduced. Using the slip angle as a measure of skating, a deformation energy is built, that associates to an admissible variation of a trajectory the square derivative of the slip angle. Using the fact that all curves are planar, its expression can be reduced to a simple semi-metric. Finally, its integration along a smooth homotopy between two curves gives an energy. Minimizing it among all possible such homotopies allows the computation of a geodesic path, with associated minimal distance. Using this new metric within a clustering algorithm yields a

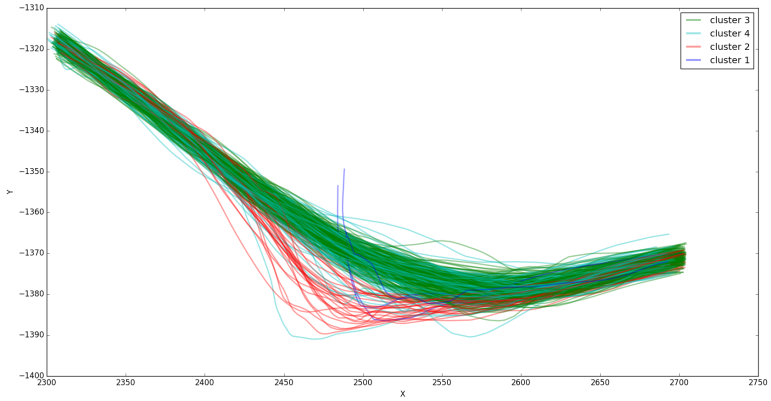


(a)

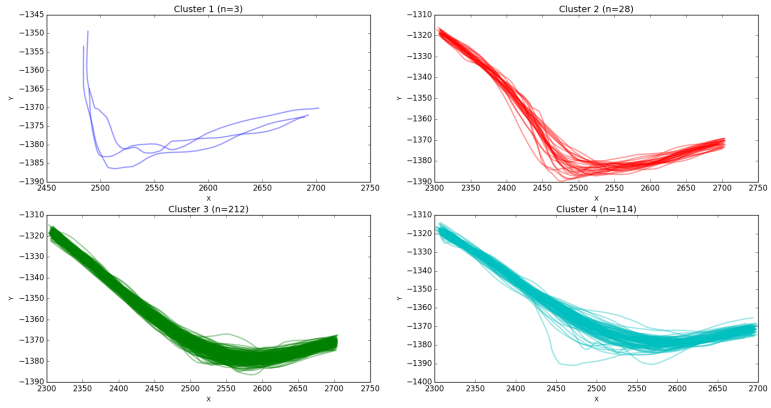


(b)

Figure 3: K-Medoids clustering with Mumford metric [7].



(a)



(b)

Figure 4: K-Medoids clustering with slippery metric (See Section 2).

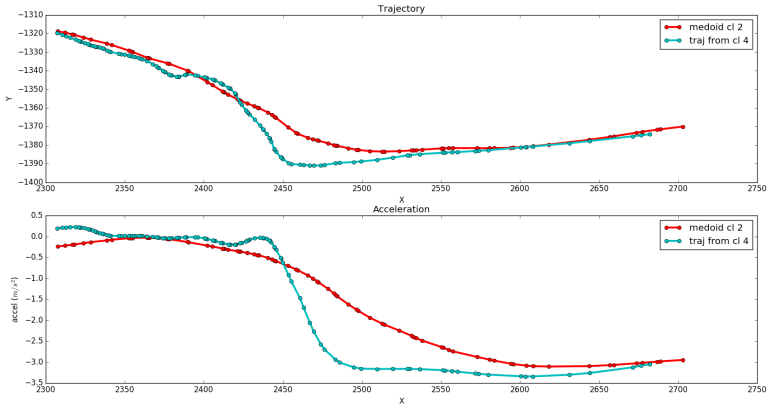


Figure 5: Comparison of the most deviant trajectory of the cluster 4 ant its related deceleration profile (blue curves) with the medoid of the cluster 2 (red curves).

new methods for clustering landing trajectories that outperforms usual shape-based metrics on both simulated and real data. The cost of distance evaluation is still compatible with offline use of the procedure, in the order of ten minutes for 15000 pairs of trajectories. It a future implementation, parallel evaluation of the distance matrix will be used, with an expected speed-up close to the number of available compute nodes.

The metric introduced is of Finsler type. The investigation of some properties of it, like its spray, was not in the scope of the present study that focused on practical applications. It is deferred to future works on the topic.

[1] Annual Analysis of the EU Air Transport Market, European Commision, 2016.
URL `https://ec.europa.eu/transport`

[2] J. Ramsay, B. Silverman, Functional Data Analysis, Springer Series in Statistics, Springer, 2005.
URL `https://books.google.co.uk/books?id=mU3dop5wY_4C`

[3] C. Bouveyron, J. Jacques, Model-based clustering of time series in group-specific functional subspaces, Advances in Data Analysis and Classification 5 (4) (2011) 281–300. doi:10.1007/s11634-011-0095-6.
URL `http://dx.doi.org/10.1007/s11634-011-0095-6`

[4] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis: Theory and Practice, Springer Series in Statistics, Springer New York, 2006.
URL `https://books.google.fr/books?id=lMy6WPFZYFcC`

[5] A. Delaigle, P. Hall, Defining probability density for a distribution of random functions, The Annals of Statistics 38 (2) (2010) 1171–1193.

[6] M. Boullé, R. Guigourès, F. Rossi, Advances in Knowledge Discovery and Management: Volume 4, Springer International Publishing, Cham, 2014, Ch. Nonparametric Hierarchical Clustering of Functional Data, pp. 15–35. doi:10.1007/978-3-319-02999-3_2.
URL http://dx.doi.org/10.1007/978-3-319-02999-3_2

[7] P. W. Michor, D. Mumford, Vanishing geodesic distance on spaces of submanifolds and diffeomorphisms., Documenta Mathematica 10 (2005) 217–245.
URL http://eudml.org/doc/125727

[8] D. Mumford, Colloquium De Giorgi 2009, Scuola Normale Superiore, Pisa, 2012, Ch. The geometry and curvature of shape spaces, pp. 43–53. doi:10.1007/978-88-7642-387-1_4.
URL http://dx.doi.org/10.1007/978-88-7642-387-1_4

[9] L. Kaufman, P. Rousseeuw, Clustering by Means of Medoids, Delft University of Technology : reports of the Faculty of Technical Mathematics and Informatics, Faculty of Mathematics and Informatics, 1987.
URL https://books.google.fr/books?id=HK-4GwAACAAJ

[10] L. Kaufman, P. Rousseeuw, Finding Groups in Data: an introduction to cluster analysis, Wiley, 1990.

[11] V. L. Popov, Contact Mechanics and Friction: Physical Principles and Applications, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, Ch. Coulomb's Law of Friction, pp. 133–154. doi:10.1007/978-3-642-10803-7_10.
URL http://dx.doi.org/10.1007/978-3-642-10803-7_10

[12] R. Rajamani, Vehicle Dynamics and Control, Mechanical Engineering Series, Springer US, 2011.
URL https://books.google.fr/books?id=cZJFDox4KuUC

[13] D. C. Liu, J. Nocedal, On the limited memory bfgs method for large scale optimization, Mathematical Programming 45 (1) (1989) 503–528.

[14] J. MacQueen, Some methods for classification and analysis of multivariate observations, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, University of California Press, 1967, pp. 281–297.
URL https://projecteuclid.org/euclid.bsmsp/1200512992

[15] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Springer Series in Statistics, Springer New York Inc., 2001.

[16] P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, Journal of Computational and Applied Mathematics 20 (Supplement C) (1987) 53 – 65. doi:https://doi.org/10.1016/0377-0427(87)90125-7.

[17] L. Jones, Modélisation des forces de contact entre le pneu d'un avion et la piste, Ph.D. thesis, thèse de doctorat dirigée par Bes, Christian et Boiffier, Jean-Luc Mathématiques appliquées et systèmes industriels Toulouse, ISAE 2012 (2012).
URL http://www.theses.fr/2012ESAE0019

[18] P. C. Besse, B. Guillouet, J.-M. Loubes, F. Royer, Review and perspective for distance-based clustering of vehicle trajectories, IEEE Transactions on Intelligent Transportation Systems 17 (11) (2016) 3306–3317.

[19] D. Berndt, J. Clifford, Using dynamic time warping to find patterns in time series 10 (1994) 359–370.

[20] M. Vlachos, D. Gunopoulos, G. Kollios, Discovering similar multidimensional trajectories, in: Proceedings of the 18th International Conference on Data Engineering, ICDE '02, IEEE Computer Society, Washington, DC, USA, 2002, pp. 673–.
URL http://dl.acm.org/citation.cfm?id=876875.878994

[21] L. Chen, M. Özsu, V. Oria, Robust and fast similarity search for moving object trajectories, in: Proceedings of the 2005 ACM SIGMOD international conference on Management of data, ACM, 2005, pp. 491–502.

[22] L. Chen, R. Ng, On the marriage of lp-norms and edit distance, in: Proceedings of the Thirtieth International Conference on Very Large Data Bases - Volume 30, VLDB '04, VLDB Endowment, 2004, pp. 792–803.
URL http://dl.acm.org/citation.cfm?id=1316689.1316758

[23] F. Hausdorff, Grundz uge der mengenlehre.

[24] M. Fréchet, Sur quelques points du calcul fonctionnel 22 (1906) 1–72.