Landmark and elastic registration of aircraft trajectories

Rémi Perrichon [1] & Xavier Gendre [2,3] & Thierry Klein [1,2]

[1] *École Nationale de l'Aviation Civile, Université de Toulouse, Toulouse, France.*
remi.perrichon@enac.fr
thierry01.klein@enac.fr
[2] *Institut de Mathématiques de Toulouse (UMR5219), Université de Toulouse, Toulouse, France.*
xavier.gendre@math.univ-toulouse.fr
[3] *Pathway.com, Paris, France.*

Résumé. L'étude statistique de trajectoires d'avions dans le cadre de l'analyse des données fonctionnelles nécessite un ensemble de pré-traitements bien connus de la littérature. On s'intéresse ici au problème d'alignement de données de trajectoires, fréquent dans de nombreuses applications. Pour les données de trajectoires, la présence de variations de phase est généralement inévitable car les vols étudiés ne se déroulent jamais simultanément, ont des durées différentes et sont opérés avec de grandes variabilités selon les compagnies aériennes. Nous proposons une comparaison de deux méthodes d'alignement : un alignement dit "par landmarks" et un alignement dit "élastique". A partir d'un échantillon de vols, l'objectif est de constituer un profil moyen d'altitude. Ce profil moyen doit être le plus informatif possible au sens où on souhaite qu'il caractérise les amplitudes moyennes de l'altitude. Une bonne procédure d'alignement devrait donc permettre de résumer les variations de l'altitude pour des phases de vol similaires. Dans le cas idéal où les phases de vol sont renseignées dans les données brutes, l'alignement par landmarks est naturel : il suffit d'identifier les landmarks aux changements de phase. Si ce n'est pas le cas, on peut préalablement segmenter les phases de vol. Moyennant un cadre conceptuel plus avancé reposant sur la géométrie différentielle, nous montrons que l'alignement élastique produit un profil d'altitude moyen plus pertinent. Grâce à une métrique riemannienne bien choisie, l'alignement permet de bien distinguer les plateaux de la phase d'approche, alors même que l'information sur les phases de vol n'est pas explicitement utilisée.

Mots-clés. Analyse de données fonctionnelles, alignement, métrique élastique, trajectoires.

Abstract. The statistical analysis of aircraft trajectories within the framework of Functional Data Analysis (FDA) requires a set of preprocessing steps that are well known in the literature. We are interested in the problem of aligning trajectory data, which is common in many applications. For trajectory data, the presence of phase variations is generally inevitable because the studied flights are never simultaneous, have different durations, and are operated with significant variabilities across airlines. We propose a comparison of two registration methods: a landmark-based registration and an elastic registration. From a sample of flights, the objective is to construct an average altitude profile. This average profile should be as informative as possible in the sense that it should characterize the average altitude amplitude. A good registration procedure should therefore summarize altitude variations for similar flight phases. In the ideal scenario where flight phases are provided in the raw data, landmark registration is natural: it suffices to identify landmarks at phase changes. If this is not the case, flight phases can be segmented beforehand. Leveraging a more advanced conceptual framework based on differential geometry, we highlight that elastic registration produces a more relevant average altitude profile. Thanks to a well-chosen Riemannian metric, the registration enables the clear distinction of plateaus in the approach phase, even when information about flight phases is not explicitly used.

Keywords. Functional data analysis, registration, elastic metric, trajectories.

1 Introduction

Thanks to advances in data storage, the growing use of sensors and the development of advanced computational techniques, it is not uncommon for statisticians to manipulate statistical units such as images, sounds, or curves. These new objects have prompted the emergence of a new terminology in statistics. The phrase Object Oriented Data Analysis (OODA), was defined by Wang and Marron (2007) to be "the statistical analysis of populations of complex objects". Part of OODA is Functional Data Analysis (FDA) for which the atoms of the statistical analysis are functions. The term FDA was coined by Ramsay (1982) and Ramsay and Dalzell (1991), even though the origin of FDA can be traced back much earlier as explained by Müller (2016). The foundational monograph of Ramsay and Silverman (2005), the work of Kokoszka and Reimherr (2021) and the review of FDA techniques written by Wang, Chiou, and Müller (2016) may serve as good introductions to the topic.

Second-generation functional data have recently been defined by Koner and Staicu (2023) as "functional data acquired in a multivariate, longitudinal, time series, or spatial design". Multivariate functional data are typically defined on the unit interval as vector-valued functions in $\mathbb{R}^d$ $(d > 1)$. Aircraft trajectories have traditionally been studied as such. To be precise, they have often been modeled as parametric paths in the sense of differential geometry, as originally proposed by Puechmorel and Delahaye (2007) and Delahaye et al. (2014). It resulted in several promising applications. Based on a sample of aircraft landing trajectories at Toulouse-Blagnac airport, Suyundykov, Puechmorel, and Ferré (2010) have identified major flows around an airport. A detection of bad runway conditions has been developed by Andrieu et al. (2016).

When considering flights in their entirety, that is to say from takeoff to landing, it is necessary to register trajectories before any statistical analysis. Registration is a standard pre-processing step in FDA that aims at separating phase variations from amplitude variations. Without this preliminary transformation, Marron et al. (2015) highlighted that a statistical analysis as basic as averaging may not offer an effective data summary. For example, the shifted betas example developed by Marron and Dryden (2021) (Section 9.1, p.176) illustrates the limitations of Functional Principal Component Analysis (FPCA) in the presence of phase variations. These same limitations have been identified for observational data, notably by Nicol (2013), who demonstrated the importance of registration for the Functional Principal Component Analysis (FPCA) of aircraft trajectories.

Phase variations are inevitable as flights are neither simultaneous nor of the same duration and exhibit strong operational variations. The key is to find an effective method to compare them at similar time points, meaning, literally, for the same flight phases. In this work, we compare two commonly used approaches in the alignment of multivariate functional data: a landmark-based approach and an elastic registration approach. The goal is to construct an effective data summary of the average altitude profile. We extend the discussion initiated by Marron et al. (2014) and companion papers to the case of commercial aviation. We show that even in the ideal scenario where the landmark approach fully takes advantage of a known segmentation of flight phases, elastic registration yields a more comprehensible average altitude profile provided a wise choice of component.

In the following, we consider a sample of $n = 5$ flights over the United States made available by the National Aeronautics and Space Administration (NASA). Each flight is observed at a finite number of moments, that is we observe

$$(t_{ij}, \boldsymbol{y}_{ij}),\ t_{ij} \in [0,1],\ \boldsymbol{y}_{ij} \in \mathbb{R}^d,\ i = 1, ..., n,\ j = 1, ..., J_i \tag{1}$$

where $\boldsymbol{y}_{ij} \equiv (y_{ij}^{[1]}, ..., y_{ij}^{[d]})$ $(d > 1)$. For a given flight i, J_i values are observed for all d components. Typically, the first three components describe the position of the aircraft (longitude, latitude, altitude). Speed, acceleration and weather values are classic examples of the other components. Since we are dealing with domestic flights, we do not consider the angular nature of longitude and latitude in this work. Time has been scaled such that the first point of each trajectory is associated with $t = 0$ and the last point with $t = 1$. The high sampling rate enables individual smoothing of each trajectory.

2 A landmark approach to register aircraft trajectories

Originally introduced by Kneip and Gasser (1992) and Gasser and Kneip (1995), landmark registration is a popular approach to align functional data. It easily adapts to the case of multivariate functional data and relies only on a handful of steps. For each flight, some structural features must first be identified as well as their timings. A template is then chosen (often, the mean of the timings). A strictly monotonic time-warping function is computed so that, for each trajectory, landmarks occur at the same time as in the template. Finally, registered values are obtained using the inverse warping function. All details may be found in Ramsay and Silverman (2005) (Section 7.3, p.132). The alignment's accuracy depends on clearly defining the features. Two cases arise in aviation depending on whether the flight phases are already labeled in the raw data or not.

2.1 Registration when flight phases are labeled in the raw data

When flight phases are already identified in the raw data, the structural features naturally correspond to the beginning (or end) of each flight phase. Their timings are explicitly available, making this situation an ideal scenario. In practice, it is the case for Flight Data Recorder (FDR) data because flight phases are automatically determined based on the monitoring and recording of many flight parameters. These are the data we use. The chosen landmarks are the takeoff, the last point of the climb phase, the first and last point of the approach. Figure 1 illustrates the impact of landmark registration on the determination of an average altitude profile. Note that the use of monotone cubic Hermite spline interpolation instead of linear interpolation for constructing the time warping functions seems that have a little effect on the obtained average altitude profile. In either case, the pronounced plateaus of the approach phase observed in trajectories n°1 and n°4 are not reflected in the average altitude profile.

2.2 Registration when flight phases are not labeled in the raw data

In the vast majority of cases, flight phases are not labeled in raw data. Several approaches are thus possible. First, we can select features based on peaks, points of inflection, and threshold crossings of one or more components of the trajectory and/or their derivatives. We then hope to *implicitly* retrieve the different flight phases and apply the registration steps as usual. A second approach consists of *explicitly* identifying flight phases using algorithms present in the literature of aviation transportation. Note that segmenting a flight into different phases is generally a complex task. Indeed, there are substantial operational differences attributable to weather conditions and/or air traffic control. Even within the same phase, aircraft may climb at different rates or fly at different cruise altitudes. As a consequence, specifying universal thresholds for flight phase segmentation is not the most effective approach. Commonly used approaches in the literature have recently been reviewed by Fala, Georgalis, and Arzamani (2023). Among statistical methods, Perrichon, Gendre, and Klein (2024) have recently demonstrated the usefulness of using Hidden Markov Models (HMMs) to identify the main flight phases of commercial aviation with high precision.

In all cases, landmark registration is only discrete evidence concerning intrinsically continuous warping functions. It ignores what happens in between landmarks, which is why it is customary to adopt a continuous fitting criterion for registration. In our case, it would be desirable to identify certain prominent plateaus.

3 Elastic registration of aircraft trajectories

Elastic registration relies on a continuous fitting criterion. It has been developed by Anuj Srivastava et al. (2011) to tackle several limitations of the $\mathbb{L}^2$ distance registration. It provides a natural template for the alignment and allows for rigorously defining the concepts of phase and amplitude variations. It is based on differential geometry. In the same perspective as

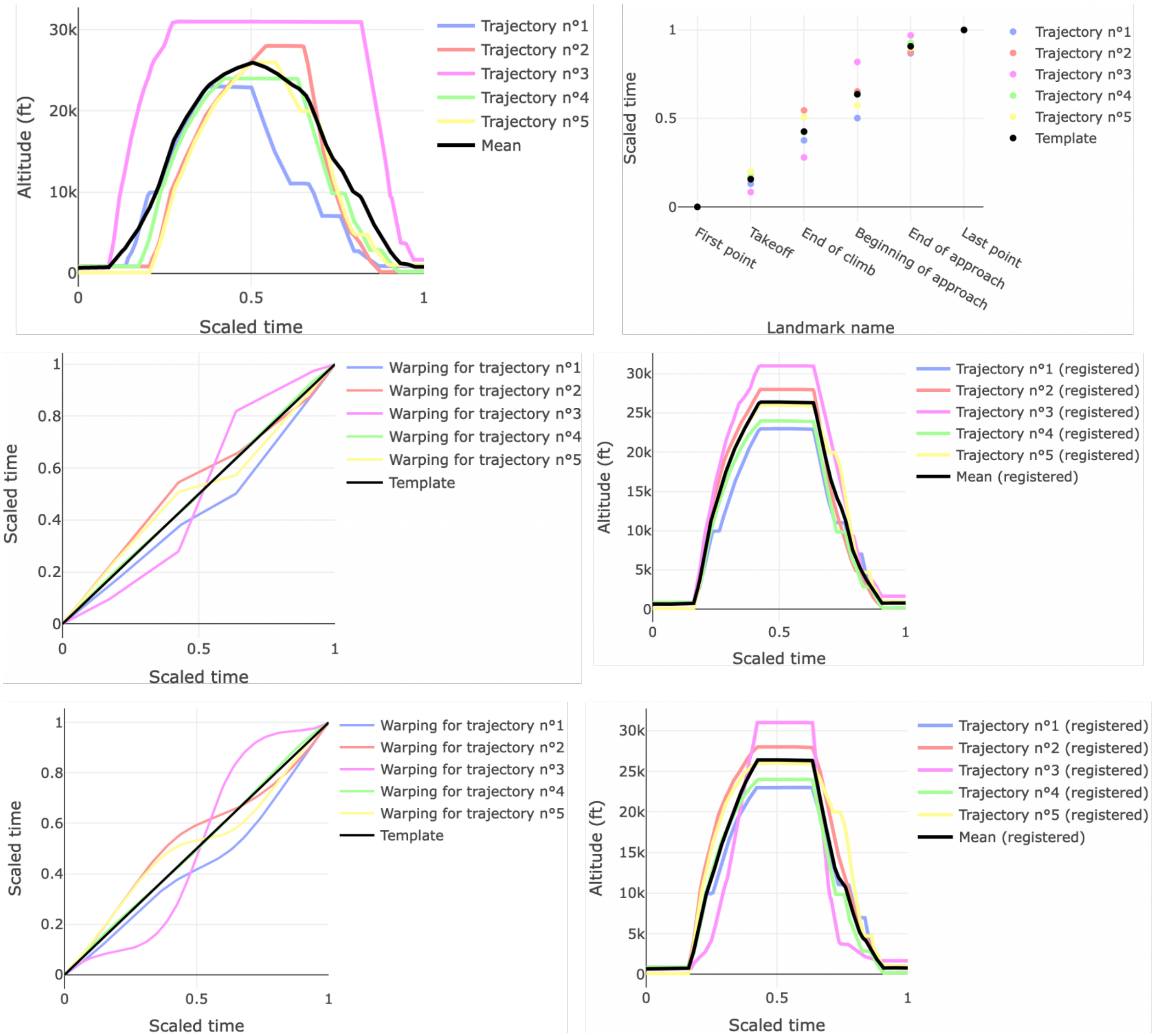

Figure 1: Altitude profiles and empirical average for raw data [top left], identification of landmarks, their timings, and a template based on the average [top right], calculation of time warping functions using linear interpolation [middle left], registered altitude profiles and the obtained registered empirical average when warping functions have been constructed with linear interpolation [middle right], time warping functions using monotone cubic Hermite spline interpolation [bottom left], registered altitude profiles and the obtained registered empirical average when warping functions have been constructed with monotone cubic Hermite spline interpolation [bottom right].

for statistical shape analysis, the foundational idea is to consider a relevant quotient space for the registration task.

Let $\mathcal{F}$ be the space of real-valued, absolutely continuous functions on $[0, 1]$ equipped with the so-called Fisher-Rao Riemannian metric. A priori, the metric is difficult to calculate. The Square-Root Velocity Function (SRVF) q of $f \in \mathcal{F}$ is defined as $q : [0, 1] \to \mathbb{R}$ where $\forall t \in [0, 1]$,

$$q(t) \equiv \begin{cases} \frac{\dot{f}(t)}{\sqrt{\left|\dot{f}(t)\right|}} & \text{if } \left|\dot{f}(t)\right| \neq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Up to a translation, it is a one-to-one map. As f is absolutely continuous, the resulting SRVF is square integrable ($\mathbb{L}^2$ is thus defined as the space of all SRVFs). Remarkably, it can be demonstrated that under the SRVF representation, the Fisher-Rao metric becomes the standard $\mathbb{L}^2$ metric. The next step is to consider an equivalence class of $q \in \mathbb{L}^2$, denoted $[q]$. Any two elements of $[q]$ represent functions which have the same amplitude variability but different phase variability. The quotient space of $\mathbb{L}^2$ under this equivalence relation is denoted $\mathcal{S} = \mathbb{L}^2/\Gamma$ where Γ is the set of orientation-preserving diffeomorphisms of the unit interval $[0, 1]$. An elastic distance is defined on $\mathcal{S}$. For $f_1, f_2 \in \mathcal{F}$, their corresponding SRVFs $q_1, q_2 \in \mathbb{L}^2$, the elastic distance d is defined as

$$d([q_1], [q_2]) = \inf_{\gamma \in \Gamma} \left\| q_1 - (q_2 \circ \gamma)\sqrt{\dot{\gamma}} \right\|. \tag{3}$$

Finding the optimal registration for f_1 and f_2 is actually the same as computing the elastic distance (see Srivastava and Klassen (2016), Definition 4.7, p.99). From Equation 3, it is clear that elastic registration is not simply a least-square alignment of SRVFs. The Karcher mean (also known as the Fréchet mean) on $\mathcal{S}$ is used to derive a template for the registration. All details are provided by Anuj Srivastava et al. 2011.

To execute elastic alignment, it entails selecting a component d of the trajectory characterized by distinct inflections signifying the transition between flight phases. Unlike the longitude profile, the altitude profile happens to exhibit the appropriate characteristics as the succession of flight phases is delineated by distinct breaks, as illustrated on Figure 2. The altitude profile obtained through elastic alignment is presented in Figure 3. Time warping functions are obtained using the Dynamic Programming (DP) algorithm presented by Srivastava and Klassen (2016) (Appendix B, p.435) and implemented by Tucker (2024). The average altitude profile now reveals the plateaus of the approach phase. Interestingly, as shown in Figure 4, once elastic registration is performed, landmarks almost perfectly coincide with the template chosen in the landmark registration procedure.

4 Conclusion and perspectives

The statistical analysis of aircraft trajectories requires a pre-processing step. This necessity arises from the very fact that, in general, flights are neither simultaneous nor of the same duration. They exhibit strong operational variations. We must ensure to compare comparable

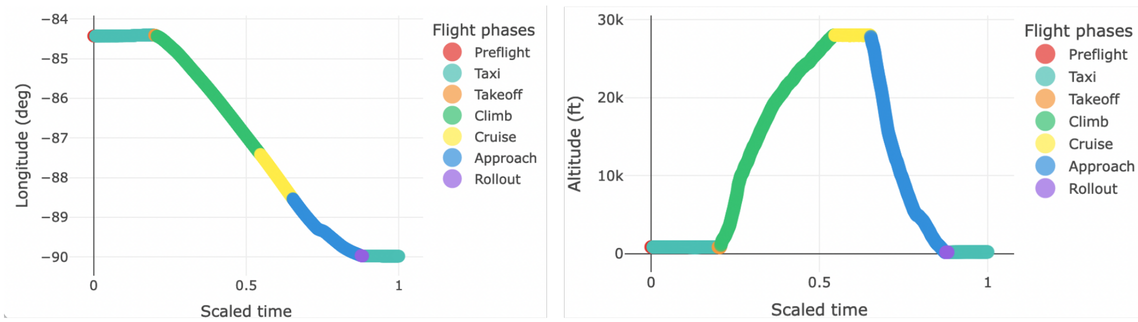

Figure 2: For trajectory n°2, it is evident that the longitude values do not exhibit any inflection points associated with a transition from one flight phase to another. Any elastic alignment procedure relying on longitude values would not yield an average altitude amplitude profile as desired, meaning for similar flight phases. We will then use the altitude profile.

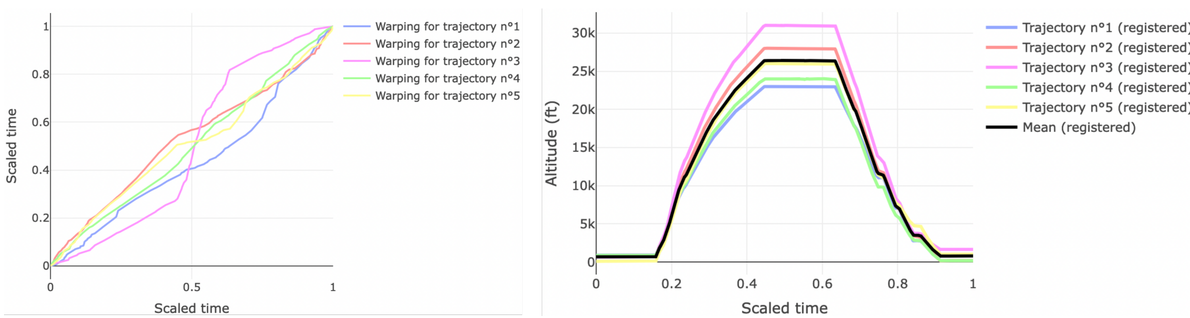

Figure 3: Time warping functions [left] and aligned trajectories [right] when using elastic registration.

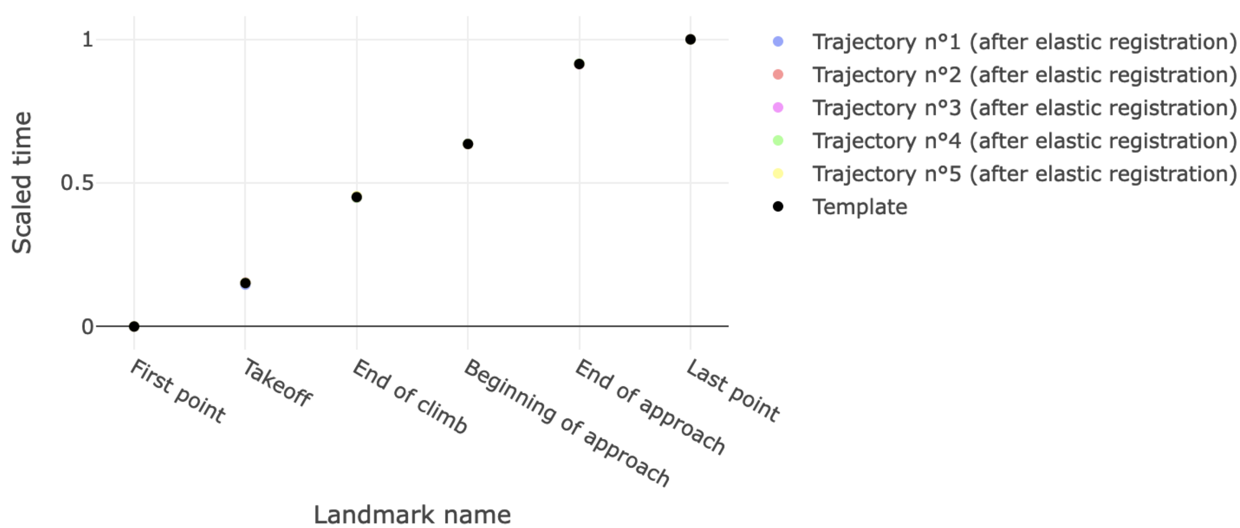

Figure 4: Once elastic alignment is performed, the landmarks almost perfectly coincide with the template chosen in the landmark alignment procedure.

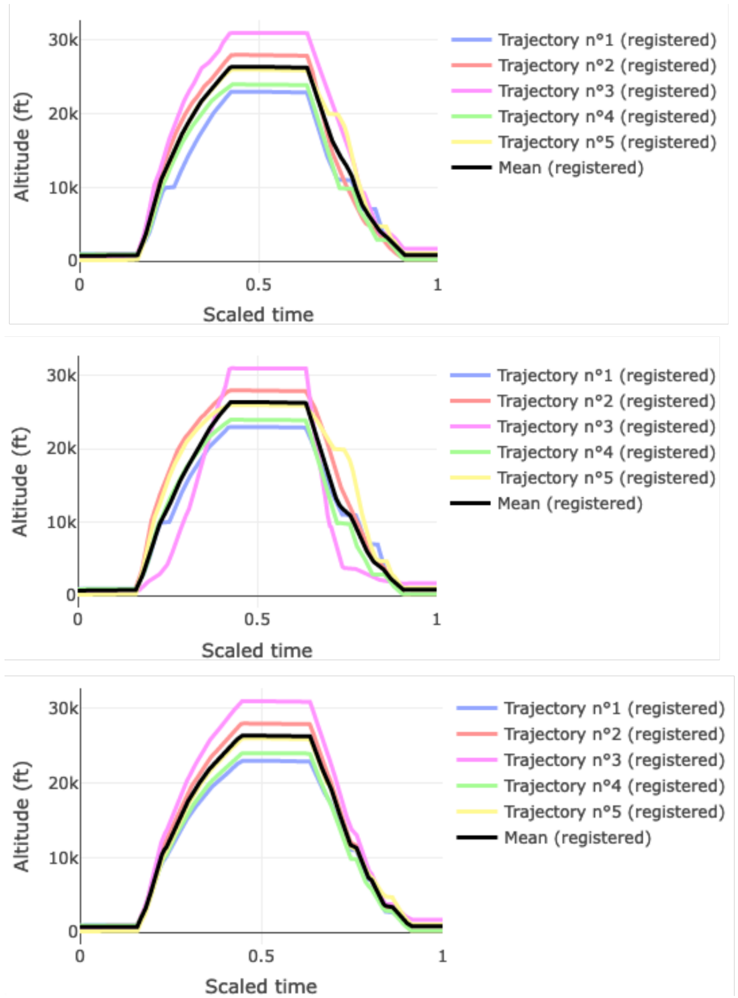

Figure 5: Landmark registration when warping functions are based on linear interpolation [top], monotone cubic Hermite spline interpolation [middle] and elastic registration [bottom].

flight instances, meaning, ensuring that the flight phases coincide after the registration. The simplest approach, when flight phases are known, is to perform landmark-based registration. However, even in this ideal scenario, it remains that landmark registration is only discrete evidence concerning intrinsically continuous warping functions. In the example presented, elastic registration, more complex both conceptually and computationally, allows for obtaining a more detailed average altitude profile, as summarized on Figure 5. Its reliability relies on selecting a component where we have good reason to believe that the breakpoints reflect the transition from one flight phase to another.

This last point is not so obvious for general aviation or drone flights. Several components are usually necessary to characterize flight phases, which are more numerous and more difficult to segment. A drone may, for instance, hover in place while rotating. A natural extension of elastic registration in this case lies in the alignment of parametric curves for which the theoretical framework has been proposed by Srivastava et al. 2011.

References

Andrieu, Cindie, Baptiste Gregorutti, Florence Nicol, and Stéphane Puechmorel. 2016. "Espaces de courbes pour l'analyse de données aéronautiques." 48ème Journées de Statistique, Montpellier.

Delahaye, D., S. Puechmorel, P. Tsiotras, and E. Feron. 2014. "Mathematical Models for Aircraft Trajectory Design: A Survey." In *Air Traffic Management and Systems,* 205–247. Lecture Notes in Electrical Engineering. Tokyo: Springer Japan. ISBN: 978-4-431-54475-3. https://doi.org/10.1007/978-4-431-54475-3_12.

Fala, Nicoletta, Georgios Georgalis, and Nastaran Arzamani. 2023. "Study on Machine Learning Methods for General Aviation Flight Phase Identification." *Journal of Aerospace Information Systems* 20 (10): 636–647. ISSN: 1940-3151. https://doi.org/10.2514/1.I011246.

Gasser, Theo, and Alois Kneip. 1995. "Searching for Structure in Curve Sample." *Journal of the American Statistical Association* 90, no. 432 (December): 1179. ISSN: 01621459. https://doi.org/10.2307/2291510.

Kneip, Alois, and Theo Gasser. 1992. "Statistical Tools to Analyze Data Representing a Sample of Curves." *The Annals of Statistics* 20, no. 3 (September). ISSN: 0090-5364. https://doi.org/10.1214/aos/1176348769.

Kokoszka, Piotr, and Matthew Reimherr. 2021. *Introduction to functional data analysis.* Texts in statistical science series. Boca Raton London New York: CRC Press. ISBN: 978-1-03-209659-9 978-1-4987-4634-2.

Koner, Salil, and Ana-Maria Staicu. 2023. "Second-Generation Functional Data." *Annual Review of Statistics and Its Application* 10 (1): 547–572. https://doi.org/10.1146/annurev-statistics-032921-033726.

Marron, J. S., James O. Ramsay, Laura M. Sangalli, and Anuj Srivastava. 2014. "Statistics of time warpings and phase variations." *Electronic Journal of Statistics* 8, no. 2 (January): 1697–1702. ISSN: 1935-7524, 1935-7524. https://doi.org/10.1214/14-EJS901.

———. 2015. "Functional Data Analysis of Amplitude and Phase Variation." *Statistical Science* 30, no. 4 (November). ISSN: 0883-4237. https://doi.org/10.1214/15-STS524.

Marron, J.S., and Ian L. Dryden. 2021. *Object Oriented Data Analysis.* Boca Raton: Chapman / Hall/CRC, October. ISBN: 978-1-351-18967-5.

Müller, Hans-Georg. 2016. "Peter Hall, functional data analysis and random objects." *The Annals of Statistics* 44, no. 5 (October). ISSN: 0090-5364. https://doi.org/10.1214/16-AOS1492.

Nicol, Florence. 2013. "Functional principal component analysis of aircraft trajectories." In *2nd International Conference on Interdisciplinary Science for Innovative Air Traffic Management (ISIATM).*

Perrichon, Rémi, Xavier Gendre, and Thierry Klein. 2024. "Hidden Markov Models and Flight Phase Identification." *Journal of Open Aviation Science* 1, no. 1 (January). ISSN: 2773-1626. https://doi.org/10.59490/joas.2024.7269.

Puechmorel, S., and D. Delahaye. 2007. "4D Trajectories: A functional data perspective." In *2007 IEEE/AIAA 26th Digital Avionics Systems Conference.* October. https://doi.org/10.1109/DASC.2007.4391832.

Ramsay, J. O. 1982. "When the data are functions." *Psychometrika* 47, no. 4 (December): 379–396. ISSN: 1860-0980. https://doi.org/10.1007/BF02293704.

Ramsay, J. O., and C. J. Dalzell. 1991. "Some Tools for Functional Data Analysis." *Journal of the Royal Statistical Society. Series B (Methodological)* 53 (3): 539–572. ISSN: 0035-9246.

Ramsay, J. O., and B. W. Silverman. 2005. *Functional Data Analysis.* Springer Series in Statistics. Springer New York. ISBN: 978-0-387-40080-8 978-0-387-22751-1.

Srivastava, A, E Klassen, S H Joshi, and I H Jermyn. 2011. "Shape Analysis of Elastic Curves in Euclidean Spaces." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, no. 7 (July): 1415–1428. ISSN: 0162-8828. https://doi.org/10.1109/TPAMI.2010.184.

Srivastava, Anuj, and Eric P. Klassen. 2016. *Functional and Shape Data Analysis.* Springer Series in Statistics. Springer New York. ISBN: 978-1-4939-4018-9 978-1-4939-4020-2.

Srivastava, Anuj, Wei Wu, Sebastian Kurtek, Eric Klassen, and J. S. Marron. 2011. *Registration of Functional Data Using Fisher-Rao Metric,* May.

Suyundykov, Ruslan, Stéphane Puechmorel, and Louis Ferré. 2010. "Multivariate Functional Data Clusterization by PCA in Sobolev Space Using Wavelets." 42ème Journées de Statistique, Marseille.

Tucker, J. Derek. 2024. *CRAN package 'fdasrvf' (version 2.2.0).* https://github.com/jdtuck/fdasrvf_R.

Wang, Haonan, and J. S. Marron. 2007. "Object oriented data analysis: Sets of trees." *The Annals of Statistics* 35, no. 5 (October): 1849–1873. ISSN: 0090-5364, 2168-8966. https://doi.org/10.1214/009053607000000217.

Wang, Jane-Ling, Jeng-Min Chiou, and Hans-Georg Müller. 2016. "Functional Data Analysis." *Annual Review of Statistics and Its Application* 3 (1): 257–295. https://doi.org/10.1146/annurev-statistics-041715-033624.