

Statistics of time warpings and phase variations

J. S. Marron

*Department of Statistics and O.R., University of North Carolina
Chapel Hill, NC, 27599-3260, USA
e-mail: marron@unc.edu*

James O. Ramsay

*Department of Psychology, McGill University
1205 Dr. Penfield Avenue Montreal, Quebec, H3A 1B1, Canada
e-mail: ramsay@psych.mcgill.ca*

Laura M. Sangalli

*MOX – Department of Mathematics, Politecnico di Milano
Piazza Leonardo da Vinci 32, 20133, Milano, Italy
e-mail: laura.sangalli@polimi.it*

and

Anuj Srivastava

*Department of Statistics, Florida State University
Tallahassee, Florida, FL 32306, USA
e-mail: anuj@stat.fsu.edu*

Abstract: Many methods exist for one dimensional curve registration, and how methods compare has not been made clear in the literature. This special section is a summary of a detailed comparison of a number of major methods, done during a recent workshop. The basis of the comparison was simultaneous analysis of a set of four real data sets, which engendered a high level of informative discussion. Most research groups in this area were represented, and many insights were gained, which are discussed here. The format of this special section is four papers introducing the data, each accompanied by a number of analyses by different groups, plus a discussion summary of the lessons learned.

Keywords and phrases: Functional data analysis, phase and amplitude variation, registration, warping.

Received August 2013.

1. Introduction

Functional Data Analysis (FDA) is a popular statistical area, which is maturing both in practice and in theory. An important challenge in this area, is to separate *amplitude variation* from *phase variation*. These concepts are illustrated using a simulated example in Figure 1. The left panel shows a sample of curves,

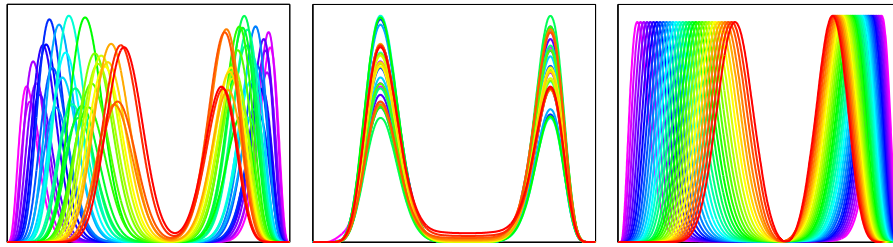


FIG 1. Simulated example, based on sample of curves (left panel), decomposed into amplitude variation (center panel) and phase variation (right panel).

with both types of variation, as seen from the locations and heights of each pair of peaks. The amplitude variation is shown in the center panel, which shows the same peaks after an alignment process. The right panel displays the phase variation in this example, in terms of warps of a notion of the mean of these curves. While most approaches to FDA ignore this type of decomposition, it has become clear from a range of important real data analysis contexts, including growth curves, motion tracking data, chemical spectra, anatomical data and neuroscience data, that such ignorance can entail very substantial loss in statistical efficiency and interpretation.

This realization has motivated a number of efforts to extract separate phase and amplitude modes of variation. Note that there are often quite different statistical contexts where such registration of curves is useful. In some situations phase variation is a nuisance, and the focus of the analysis is on amplitude variation (this is sometimes also called “vertical variation”, leaning perhaps too heavily on the classical Cartesian representation of functions). This happens for a number of spectral examples, where peaks (representing given substances) need to be aligned (so the peaks correctly correspond), but it is the amounts of substances, reflected by the heights of the peaks, that is the main goal of the analysis. In other situations, phase variation (sometimes called “horizontal variation”, but when the independent variable is time, this is usefully thought of as “tempo”) can be the main focus, and amplitude should be considered as a nuisance. In still other situations both amplitude and phase modes of variation are important, and should be studied jointly. Human movement data are generally of this type. But even when both modes are important, decomposition is still very useful, because their concatenation makes the information much more accessible to statistical analysis.

Each of the previous efforts at decomposing amplitude and phase variation typically involves the analysis of a challenging real data set where curve registration is important, and good results from use of the proposed method on that data set are shown. While that format is typical for publication in the statistical literature, it fails to provide useful comparison of methods, in particular not elucidating their relative strengths and weaknesses. This need for comparison could be met by requiring authors to do comparisons themselves, but this is generally an unreasonable request, due to lack of availability of general purpose

software implementations of existing methods, and perhaps more importantly due to the need for large amounts of expert-level tuning that is usually needed to get a good result.

In response to the need for a more global and useful comparison of these methods, the Mathematical Biosciences Institute hosted a workshop, with representation from most of the principal research groups in this area. The initial basis of the workshop was for each group to do some analysis on a common set of four data sets. The results from the various groups then formed the basis of a large amount of overall discussion, which led to many new insights. This special section is aimed at conveying the interesting lessons learned at this workshop on curve registration to the larger statistical community.

Following the data centric orientation of the workshop, this special section is oriented around these four data sets:

- Proteomic data, spectral data collected for the study of Acute Myeloid Leukemia by the Adelaide Proteomics Center.
- Juggling Data, digitized traces of human movement during a juggling exercise, recorded from infra-red emitting diodes at McGill University.
- Spike Train Data, recording the electrical activity of a movement-encoded neuron in the primary motor cortex, collected in the Hatsopoulos Lab at the University of Chicago.
- AneuRisk65, datasets of three-dimensional vascular geometries, obtained from 3D angiographies collected within the AneuRisk project for the study of cerebral aneurysms pathogenesis.

All data sets are available at the MBI website: <http://mbi.osu.edu/2012/stwdescription.html>.

The main ideas behind each of these data sets, as well as data analytic goals of interest, together with discussion of all preprocessing that was done in each case, is presented in the four main papers in this special section. The analyses presented by the various research groups are included in a format similar to discussions in other contexts. Next, additional relevant comments together with a summary of insights gained, are given by the original providers of the data.

2. Review

This section contains a brief discussion of a few highlights of the curve registration literature. This is not intended to be comprehensive, but instead to eliminate the need for (perhaps too repetitive) re-discussion of these papers in several of the contributions to this special section.

Important early work in this area was from the viewpoint of landmark registration, where a few anchor points, which correspond across the family of curves, form the basis of the registration. See Gasser and Kneip (1995) and Ramsay and Silverman (2005) for recent results of this type, and for access to the earlier literature.

While landmarks are useful when they exist in a natural way, in many situations these cannot be found. Hence, various landmark free approaches, which

treat curves as continuous data objects, have been subject of more recent studies. Important recent results (containing many earlier references) include: Ramsay and Li (1998); Wang and Gasser (1999); Gervini and Gasser (2004); Ramsay and Silverman (2005); Kaziska and Srivastava (2007); Sangalli et al. (2009), Kneip et al. (2000); Liu and Müller (2004); James (2007).

Registration can be performed jointly with modelling and analysis of data, as in the registration to principal components method described in Kneip and Ramsay (2008).

The issue of registration can also be combined with the one of clustering functional data. Some works considering this aspect are Sangalli et al. (2010); Tang and Müller (2009); Liu and Yang (2009); Boudaoud, Rix and Meste (2010).

Research providing insightful theoretical frameworks can be found in Srivastava et al. (2011a); Vantini (2012).

Important related work can also be found in the context of longitudinal data, where semiparametric non-linear mixed-effects models are proven to be useful Lawton, Sylvestre and Maggio (1972); Lindstrom and Bates (1990); Ke and Wang (2001); Altman and Villarreal (2004); Brumback and Lindstrom (2004).

The issue of registration has been considered also in the shape analysis field. In particular, the earliest work on elastic shape analysis of planar curves is by Younes Younes (1999) who introduced an elastic metric and a complex square-root representation for enabling Euclidean analysis. This was followed by more elaborate studies of such representations, including Younes et al. (2008); Michor and Mumford (2006); Mio, Srivastava and Joshi (2007) and Srivastava et al. (2011b). The last paper extended this elastic shape analysis from planar curves to curves in arbitrary Euclidean spaces.

Finally, although the workshop focus has been on the registration of curves, possibly multidimensional, it is important to cite the work on registration of surfaces, in imaging. For refer the interested reader to the book of Modersitzki (2003), and references therein.

Acknowledgements

All participants are very grateful to the Mathematical Biosciences Institute (<http://mbi.osu.edu/>) for hosting many aspects of this effort, including data accessibility, travel funding, and an ideal venue for the personal interaction that has been the key to success.

References

- ALTMAN, N. S. and VILLARREAL, J. C. (2004). Self-modelling regression for longitudinal data with time-invariant covariates. *Canad. J. Statist.* **32** 251–268. [MR2101755](#)
- BOUDAUD, S., RIX, H. and MESTE, O. (2010). Core shape modelling of a set of curves. *Comput. Statist. Data Anal.* **54** 308–325. [MR2756428](#)
- BRUMBACK, L. C. and LINDSTROM, M. J. (2004). Self modeling with flexible, random time transformations. *Biometrics* **60** 461–470. [MR2066281](#)

- GASSER, T. and KNEIP, A. (1995). Searching for structure in curve samples. *Journal of the American Statistical Association* **90** 1179–1188.
- GERVINI, D. and GASSER, T. (2004). Self-modelling warping functions. *Journal of the Royal Statistical Society, Series B* **66** 959–971. [MR2102475](#)
- JAMES, G. M. (2007). Curve alignment by moments. *The Annals of Applied Statistics* **1** 480–501. [MR2415744](#)
- KAZISKA, D. and SRIVASTAVA, A. (2007). Gait-based human recognition by classification of cyclostationary processes on nonlinear shape manifolds. *Journal of the American Statistical Association* **102** 1114–1128. [MR2412534](#)
- KE, C. and WANG, Y. (2001). Semiparametric nonlinear mixed-effects models and their applications. *J. Amer. Statist. Assoc.* **96** 1272–1298. [MR1946577](#)
- KNEIP, A. and RAMSAY, J. O. (2008). Combining registration and fitting for functional models. *Journal of the American Statistical Association* **103** 1155–1165. [MR2528838](#)
- KNEIP, A., LI, X., MACGIBBON, K. B. and RAMSAY, J. O. (2000). Curve registration by local regression. *Can. J. Stat.* **28** 19–29. [MR1789833](#)
- LAWTON, W. H., SYLVESTRE, E. A. and MAGGIO, M. S. (1972). Self modeling nonlinear regression. *Technometrics* **14** 513–532.
- LINDSTROM, M. J. and BATES, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics* **46** 673–687. [MR1085815](#)
- LIU, X. and MÜLLER, H. G. (2004). Functional convex averaging and synchronization for time-warped random curves. *J. Am. Stat. Assoc.* **99** 687–699. [MR2090903](#)
- LIU, X. and YANG, M. C. K. (2009). Simultaneous curve registration and clustering for functional data. *Comput. Statist. Data Anal.* **53** 1361–1376. [MR2657097](#)
- MICHOR, P. W. and MUMFORD, D. (2006). Riemannian geometries on spaces of plane curves. *J. Eur. Math. Soc.* **8** 1–48. [MR2201275](#)
- MIO, W., SRIVASTAVA, A. and JOSHI, S. H. (2007). On shape of plane elastic curves. *Intl. Journal of Computer Vision* **73** 307–324.
- MODERSITZKI, J. (2003). *Numerical Methods for Image Registration*. Oxford University Press. [MR2035877](#)
- RAMSAY, J. O. and LI, X. (1998). Curve registration. *Journal of the Royal Statistical Society, Series B.* **60** 351–363. [MR1616045](#)
- RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, Second ed., Springer. [MR2168993](#)
- SANGALLI, L. M., SECCHI, P., VANTINI, S. and VENEZIANI, A. (2009). A case study in exploratory functional data analysis: Geometrical features of the internal carotid artery. *J. Amer. Statist. Assoc.* **104** 37–48. [MR2663032](#)
- SANGALLI, L. M., SECCHI, P., VANTINI, S. and VITELLI, V. (2010). K-mean alignment for curve clustering. *Computational Statistics and Data Analysis* **54** 1219–1233. [MR2600827](#)
- SRIVASTAVA, A., WU, W., KURTEK, S., KLASSEN, E. and MARRON, J. S. (2011a). Registration of functional data using fisher-rao metric. *arXiv:1103.3817v2 [math.ST]*.

- SRIVASTAVA, A., KLASSEN, E., JOSHI, S. H. and JERMYN, I. H. (2011b). Shape analysis of elastic curves in Euclidean spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33** 1415–1428.
- TANG, R. and MÜLLER, H. G. (2009). Time-synchronized clustering of gene expression trajectories. *Biostatistics* **10** 32–45.
- VANTINI, S. (2012). On the definition of phase and amplitude variability in functional data analysis. *TEST* **21** 676–696. [MR2992088](#)
- WANG, K. and GASSER, T. (1999). Synchronizing sample curves nonparametrically. *The Annals of Statistics* **27** 439–460. [MR1714722](#)
- YOUNES, L. (1999). Optimal matching between shapes via elastic deformations. *Journal of Image and Vision Computing* **17** 381–389.
- YOUNES, L., MICHOR, P. W., SHAH, J., MUMFORD, D. and LINCEI, R. (2008). A metric on shape space with explicit geodesics. *Matematica E Applicazioni* **19** 25–57. [MR2383560](#)