Electronic Journal of Statistics
Vol. 8 (2014) 1793–1796
ISSN: 1935-7524
DOI: 10.1214/14-EJS865D

Analysis of spike train data: Comparison between the real and the simulated data*

Xiaosun Lu and J. S. Marron

Department of Statistics and Operations Research
University of North Carolina at Chapel Hill NC 27599 USA
e-mail: xiaosun@live.unc.edu*;* marron@unc.edu

Abstract: This paper compares the real experimental data with the data simulated by Wu and Srivastava (2011). It turns out that the real data exhibit a different composition of the phase and the amplitude variation from the simulated data, where these two types of variation are separated using the Fisher Rao curve registration. As a result, for the real data the original functions are a better choice of data objects for path discrimination, while for the simulated data the domain warping functions are better.

Keywords and phrases: Curve registration, distance weighted discrimination, functional data analysis.

Received August 2013.

This analysis focuses on comparing the real experimental data presented by Wu et al. (2013) with the data simulated by Wu and Srivastava (2011). Our goal is to understand how well the model performs in mimicking the real data, with an eye towards providing useful diagnostics for possible improvement of the simulation model.

For each of these two data sets, the *horizontal* (or phase) variation and the *vertical* (or amplitude) variation are separated through the domain warping approach proposed by Srivastava et al. (2011). The horizontal variation is captured by the resulting warping functions, and the vertical variation is captured by the aligned functions. Our study shows that the simulated data and the real data exhibit quite different types of variability. In particular, the path difference in the real data is significant both horizontally and vertically, while in the simulated data it is not significant for one vertical comparison.

Object Oriented Data Analysis, first introduced by Wang and Marron (2007), provides useful terminology for this discussion. For both the real and the simulated data, three different types of data objects are considered:

(1) raw functions f,
(2) warping functions h,
(3) aligned functions $\tilde{f}$.

*Main article 10.1214/14-EJS865.

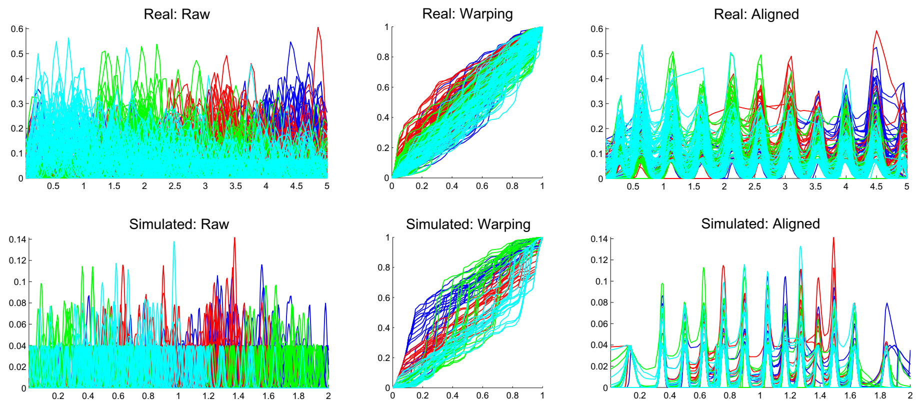

FIG 1. *Fisher Rao alignment for both the real (top) and the simulated (bottom) data. Left: Raw functions. Middle: Domain warping functions. Right: Aligned functions. Colors indicate different types of paths.*

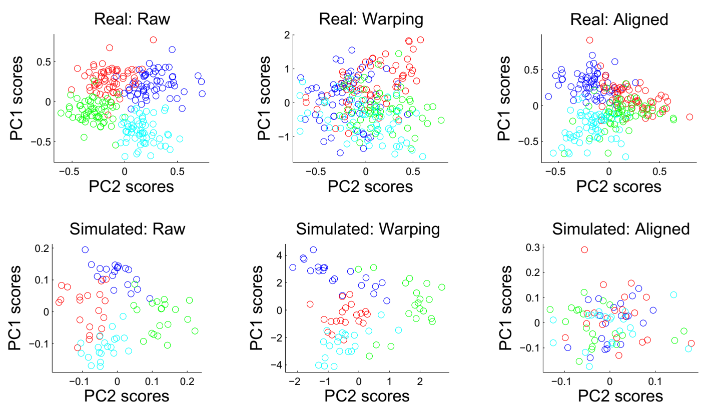

FIG 2. *Scores scatterplots of the first two components obtained by performing FPCA on the data objects (1)–(3) (from left to right) for both the real (top) and the simulated (bottom) data, respectively. In the panels* $(1, 1)$*,* $(2, 1)$ *and* $(2, 2)$*, these four types of paths can be well separated; in the panels* $(1, 2)$ *and* $(1, 3)$*, they are more overlapped with each other; in the panel* $(2, 3)$*, the difference between paths can be hardly seen.*

See Figure 1 for these functions. The latter two types of data objects lead to *horizontal* analysis (i.e. the analysis of the phase or tempo variation) and *vertical* analysis (i.e. the analysis of the amplitude variation), respectively. See Lu (2013) and Lu and Marron (2013) for more discussion.

Functional Principal Component Analysis (FPCA) is used here to analyze these three types of data objects separately. The corresponding scores scatterplots of the first two components are shown in Figure 2. The top three panels show the real data, and the bottom panels show the simulated data. Colors indicate the four types of paths. It is seen, in the analysis of the raw functions

TABLE 1

Summarized results of the DiProPerm t-tests for investigating the path difference in both the real and the simulated data. The blue background color of a cell indicates insignificant results, while the pink/red color indicates significant results (the red color highlights the data objects that lead to the most significant results)

	Data Objects	Classes	p-values		Z-scores	
			Real	Simulated	Real	Simulated
(1)	Raw f	1 2 vs. 3 4	0.00	0.00	33.81	8.59
		1 vs. 2	0.00	0.00	13.50	5.28
		3 vs. 4	0.00	0.00	20.41	5.80
(2)	Warping h	1 2 vs. 3 4	0.00	0.00	11.50	10.05
		1 vs. 2	0.00	0.00	5.70	13.51
		3 vs. 4	0.00	0.00	6.15	11.21
(3)	Aligned $\tilde{f}$	1 2 vs. 3 4	0.00	0.00	18.13	4.46
		1 vs. 2	0.00	0.00	10.39	3.86
		3 vs. 4	0.00	0.83	12.26	-0.96

for both the real and the simulated data, i.e. the two panels in the first column, that these four types of paths are visually well separated. Such a good separation is also seen in the horizontal analysis of the simulated data in Panel $(2, 2)$. In the horizontal and the vertical analyses of the real data, Panel $(1, 2)$ and $(1, 3)$ respectively, one can still tell the difference between these four types of paths, although there is more overlap. The vertical separation shown in Panel $(1, 3)$ seems to be better than the horizontal separation shown in Panel $(1, 2)$. However, in the vertical analysis of the simulated data, i.e Panel $(2, 3)$, there is no obvious difference among the four types of paths. This shows an important difference between the simulated data (where the classes only differ in phase) and the real data, which also have class differences in amplitude.

Further analysis was conducted to more deeply investigate this path difference. For each of the two data sets and each type of data objects, we first used Distance Weighted Discrimination (DWD), proposed by Marron et al. (2007), to find a useful direction in function space to separate the union of Path Types 1 and 2 from the union of Path Types 3 and 4. Then we studied the differences between 1 and 2, and also 3 and 4, separately. For each of these separating steps, the DiProPerm (Direction Projection Permutation) t-test based on the DWD direction (See Wei et al. (2013) for details of this approach) was used to test the mean difference between two classes. The results are summarized in Table 1. It is seen from the p-values that the path difference is significant for the real data, both horizontally and vertically. However, in the simulated data, some paths are significantly different only in a horizontal way, but not in a vertical way (highlighted in blue background color). Since all the significant p-values are zero, no information is available for comparison. Such comparisons can be made in DiProPerm by studying the Z-scores of the statistics with respect to the null population (i.e. with no class difference). A higher score value indicates a bigger class difference. In the present case, the Z-scores (last two columns in the table) show that the path differences in the real data are more vertical than horizontal, while the path differences in the simulated data are mainly horizontal. This provides rigorous verification of the visual observation comparing the simulated and the real data score scatterplots in Figure 2. It is also seen that, among these three choices of data objects, the best one for classifying the four types of paths in the real data is the raw functions, which lead to the highest Z-scores in all of the three DiProPerm tests (highlighted in dark red color). However, for the simulated data, the best choice for classifying the paths is the warping functions. This shows that there are strong differences between the real and simulated data sets, suggesting more work could improve the quality of the simulated data sets, and showing that this can be done by creating a bigger simulated difference in amplitude.

Acknowledgements

The Mathematical Biosciences Institute provided travel funding which led to many useful interactions leading to this work.

References

Lu, X. (2013). Object oriented data analysis of cell images and analysis of elastic functions. *Ph.D. Dissertation*, University of North Carolina, Chapel Hill, NC, USA. MR3153472

Lu, X. and Marron, J. S. (2013). Principal nested spheres for time warped functional data analysis. arXiv:1304.6789.

Marron, J. S., Todd, M. J., and Ahn, J. (2007). Distance weighted discrimination. *Journal of the ASA*, 102(480):1267–1271. MR2412548

Srivastava, A., Wu, W., Kurtek, S., Klassen, E., and Marron, J. S. (2011). Statistical analysis and modeling of elastic functions. arXiv:1103.3817.

Wang, H. and Marron, J. S. (2007). Object oriented data analysis: Sets of trees. *The Annals of Statistics*, 35(5):1849–1873. MR2363955

Wei, S., Lee, C., Wichers, L., Li, G., and Marron, J. S. (2013). Direction-projection-permutation for high dimensional hypothesis tests. arXiv:1304.0796.

Wu, W., Hatsopoulos, N. G., and Srivastava, A. (2014). Introduction to neural spike train data for phase-amplitude analysis. *Electronic Journal of Statistics*, 8:1759–1768, Special Section on Statistics of Time Warpings and Phase Variations.

Wu, W. and Srivastava, A. (2011). An information-geometric framework for statistical inferences in the neural spike train space. *Journal of Computational Neuroscience*, 31(3):725–748. MR2864743