Functional Statistical Process Control using Elastic Methods

J. Derek Tucker*

Abstract

The problem of statistical analysis and modeling of data in process control is important in determining when a production has moved beyond a baseline. Recently, the use of functional data has become an interest in statistical process control. The observations here are time samples of real-valued functions on an observation interval, and to perform effective data analysis it is desirable to have a generative, probabilistic model for these observations. The model is expected to properly and parsimoniously characterize the nature and variability in the baseline data. It should also lead to efficient procedures for conducting hypothesis tests, performing bootstraps, and making decisions. We wish to perform statistical process control (SPC) using functional data. We present a technique that takes into account both amplitude and phase variability in the data using the square-root slope framework. In this work we present metrics that are defined to measure both types of variability and show demonstration when the data has gone beyond the control limits.

Key Words: Compositional noise, functional data analysis, functional statistical process control, functional principal component analysis

1. Introduction

A significant amount of data collected in process monitoring is in the form of "functional data"; that is, a collection of data over some index (e.g., time, frequency). It is often the case that, for the purpose of analysis, this functional data is discretized in some manner, with only the key features of the curve being retained. Some examples include the maximum peak, minimum peak, number of peaks, or rate of change at a particular point in the curve. There has been a recent push across industrial statistics for ways to analyze functional data in its entirety. Methodology that involves such analysis is known as functional data analysis. An excellent introduction and reference to this field is the book by Ramsay and Silverman (2005) and we refer interested readers there for more details. Some data analysts may be averse to updating their analysis methodology into working with functional data due to the complicated dependence structure often present in such data. However, ignoring this dependence structure can result in faulty conclusions as it is an integral part of the data and the process that generated it.

One of the methods to perform effective data analysis is to have a generative, probabilistic model for these observations. The model is expected to properly and parsimoniously characterize the nature and variability in the data. It should also lead to efficient procedures for conducting hypothesis tests, performing bootstraps, and making forecasts. An interesting aspect of functional data is that underlying variability can be ascribed to two sources. In a sample data the given functions may not be perfectly aligned and the mechanism for alignment is an important topic of research. The variability exhibited in functions after alignment is termed the amplitude (or y or vertical) variability and the warping functions that are used in the alignment are said to capture the phase (or x or horizontal) variability. Capturing this variability is crucial when modeling and monitoring functional data in a process control architecture.

Hypothesis testing and inference methods have been developed for functional data (Horváth and Kokoszka, 2012). Specifically, Horváth developed a number of statistical

*Sandia National Laboratories, PO Box 5800 MS 1202, Albuquerque, NM 87185

tests, with corresponding distributions, for testing shifts from a mean function and other statistical tests. James and Sood (2006) developed a hypothesis testing procedure for testing the shape of functions by testing the permuted residuals when a smooth function is fitted to the data. Hall and Van Keilegom (2007) developed a two-sample test for smoothing the discrete measurements to form the functional. Recently, Storlie et al. (2013) developed a method to tests the shape of population of curves using a B-Spline basis test and hierarchical Gaussian process approach.

Most of these methods for functional statistical process control (SPC) do not take into account the amplitude and phase variability. Most assume there is a pre-alignment step of the data and then the tests are conducted. We believe that it is imperative that any technique for SPC using functional data should take both these variabilities into account as part of the testing procedure. In order to make a decision when the data has moved outside the control limits, we need to account for this variability. A prominent example of the situation of when this variability is ignored is in modeling functional data using function principal component analysis (fPCA) (see Tucker et al. (2013)). Moreover, we can use metrics defined to measure this variability in order to determine when we have gone beyond the control limit.

In this paper, we present two methods for implementing statistical process control using functional data while accounting for both of these variabilities. We base our tests on the elastic functional data analysis framework presented in (Kurtek et al., 2011; Srivastava et al., 2011b; Tucker et al., 2013). The first method is based on the elastic amplitude and phase distance (Tucker et al., 2014) where the distances are calculated from a known template function (mean) and any distance above a certain control limit (threshold) the sample is flagged. The second method uses *vertical* and *horizontal* fPCA (Tucker et al., 2013) where a model is constructed taking the variability into account. The principal scores are then monitored and any sample above or below the control limit is flagged for each of the fPCA models.

The rest of this paper is organized as follows: Section 2 presents the elastic distance approach. Section 3 presents the functional principal component analysis approach. Results using simulated and real data are provided in Section 4. Finally, conclusions and observations are offered in Section 5.

2. Statistical Process Control using Elastic Distances

Let f be a real-valued function with the domain $[0, 1]$; the domain can easily be transformed to any other interval. For concreteness, only functions that are absolutely continuous on $[0, 1]$ will be considered; let $\mathcal{F}$ denote the set of all such functions. In practice, since the observed data are discrete, this assumption is not a restriction. Also, let Γ be the set of boundary-preserving diffeomorphisms of the unit interval $[0, 1]$: $\Gamma = \{\gamma : [0, 1] \to [0, 1] | \, \gamma(0) = 0, \, \gamma(1) = 1, \gamma \text{ is a diffeomorphism}\}$. Elements of Γ play the role of warping functions. For any $f \in \mathcal{F}$ and $\gamma \in \Gamma$, the composition $f \circ \gamma$ denotes the time-warping of f by γ. With the composition operation, the set Γ is a group with the identity element $\gamma_{id}(t) = t$. This is an important observation since the group structure of Γ is seldom utilized in past papers on functional data analysis.

We know from Tucker et al. (2013) that there are two metrics to measure the amplitude and phase variability of functions. These metrics are proper distances, which means they are symmetric, isometric, and obey the triangle inequality. The amplitude or y-distance for any two functions $f_1, \ f_2 \in \mathcal{F}$ is defined to be

$$d_a(f_1, f_2) = \inf_{\gamma \in \Gamma} \|q_1 - (q_2 \circ \gamma)\sqrt{\dot{\gamma}}\|, \tag{2.1}$$

where $q(t) = \text{sign}(\dot{f}(t))\sqrt{|\dot{f}(t)|}$ and is known as the square-root slope function (SRSF) and $q \in \mathbb{L}^2$. For the properties of the SRSF and the reason for its use the reader is referred to Srivastava et al. (2011a). Moreover, it can be shown that for any $\gamma_1, \gamma_2 \in \Gamma$, we have $d_a(f_1 \circ \gamma_1, f_2 \circ \gamma_2) = d_a(f_1, f_2)$.

Optimization Over Γ: The minimization over Γ can be performed in many ways. In case Γ is represented by a parametric family, then one can use the parameter space to perform the estimation as Kneip and Ramsay (2008). However, since Γ is a nonlinear manifold, it is impossible to express it completely in a parametric vector space. One method is to use the standard Dynamic Programming (DP) algorithm (Bertsekas, 1995) to solve for an optimal γ. It should be noted that for any fixed partition of the interval $[0, 1]$, this algorithm provides the exact optimal γ that is restricted to the graph of this partition. Recently, Huang (2014) developed an optimization method of performing optimization methods on Riemannian manifolds, with one method being the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. The RBFGS algorithm was developed to solve for γ and is a faster alternative to DP. Additionally, Cheng et al. (2013) has developed a solution using a Bayesian approach where they use a Dirichlet prior on the warping functions and a Markov chain Monte Carlo algorithm.

The second metric from Tucker et al. (2013) is a measure on the phase-variability which is captured by γ using

$$d_p(\gamma) = \cos^{-1}\left(\int_0^1 \psi^2(t)\, dt\right) \tag{2.2}$$

where $\psi = \sqrt{\dot{\gamma}}$. One important advantage of the above transformation is that since $\|\psi\|^2 = \int_0^1 \psi(t)^2 dt = \int_0^1 \dot{\gamma}(t) dt = \gamma(1) - \gamma(0) = 1$, the set of all such ψs is a Hilbert sphere $\mathbb{S}_\infty$, a unit sphere in the Hilbert space $\mathbb{L}^2$. In other words, the square-root representation simplifies the complicated geometry of Γ to a unit sphere.

We can then use these distances to monitor samples that have deviated from the Karcher Mean or centroid function. We will assume that the Karcher mean has been established on a baseline data set using Algorithm 1 from Tucker et al. (2013) and will denote it as μ_q. Then for each new sample function f_i, we convert to its corresponding q_i and calculate d_a using (2.1) and d_p using (2.2). If either d_a, d_p, or both cross a pre-determined threshold we mark the test function as out of limits. One could also use the weighted combination of

$$d_\tau = (1 - \tau)d_a + \tau d_p, \quad \tau \in [0, 1]$$

which, depending on the value of τ, would give preference to either amplitude or phase, or both. By monitoring d_τ one number can be monitored for when a sample is in-control, depending on a pre-determined threshold.

3. Statistical Process Control using Elastic fPCA

The next method we can use for fSPC is by performing vertical and horizontal fPCA (Tucker et al., 2013). Vertical fPCA is performed using the aligned SRSF's which are product of the Karcher Mean calculation.

Let $f_1, \ldots, f_n$ be a given set of functions, and $q_1, \ldots, q_n$ be the corresponding SRSFs, μ_q be their Karcher Mean, and let $\tilde{q}_i$s be the corresponding aligned SRSFs using Algorithm 1 in Tucker et al. (2013). In performing vertical fPCA, one should not forget about the variability associated with the initial values, i.e., $\{f_i(0)\}$, of the given functions. Since representing functions by their SRSFs ignores this initial value, this variable is treated separately. That is, a functional variable f is analyzed using the pair $(q, f(0))$ rather than just q. This way, the mapping from the function space $\mathcal{F}$ to $\mathbb{L}^2 \times \mathbb{R}$ is a bijection. In

practice, where q is represented using a finite partition of $[0,1]$, say with cardinality T, the combined vector $h_i = [q_i \;\; f_i(0)]$ simply has dimension $(T+1)$ for fPCA. We can define a sample covariance operator for the aligned combined vector $\tilde{h} = [\tilde{q}_1 \;\; f_i(0)]$ as

$$K_h = \frac{1}{n-1}\sum_{i=1}^{n} E[(\tilde{h}_i - \mu_h)(\tilde{h}_i - \mu_h)^{\mathsf{T}}] \in \mathbb{R}^{(T+1)\times(T+1)} \,, \tag{3.1}$$

where $\mu_h = [\mu_q \;\; \bar{f}(0)]$. Taking the SVD, $K_h = U_h \Sigma_h V_h^{\mathsf{T}}$ we have the singular vectors and singular values $\{\sigma_{h,k}\}$.

We then can set the control limits to be the following

$$\begin{aligned} UCL &= 3\sqrt{\sigma_{h,k}} \\ CL &= 0 \\ LCL &= -3\sqrt{\sigma_{h,k}} \end{aligned}$$

for a chosen singular value. The scores for each observation function are computed and those functions beyond the control limits are marked.

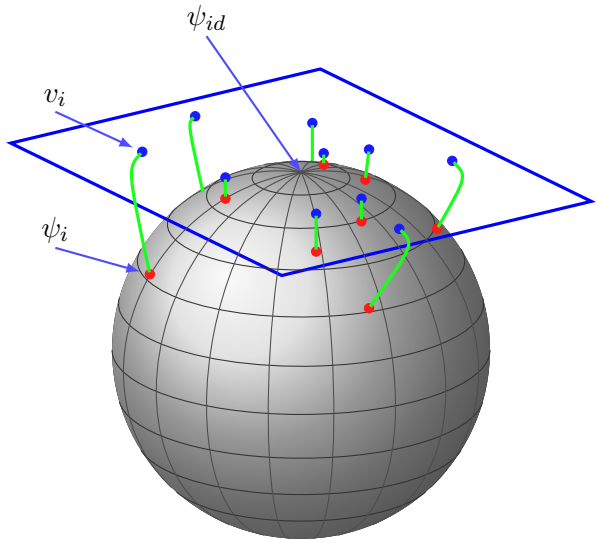


Figure 1: Depiction of the SRSF space of warping functions as a sphere and a tangent space at identity ψ_{id}.

Additionally, we should monitor the scores produced from horizontal fPCA. First, we find the Karcher mean of the set of warping functions as described in Algorithm 2 (Tucker et al., 2013). These are the warping functions produced from aligning the baseline set of functions. Horizontal fPCA is performed using ψ which the space of all $\{\psi_i\}$ is a Hilbert sphere, $\mathbb{S}_\infty$. However, since $\mathbb{S}_\infty$ is a nonlinear space (a sphere), one cannot perform principal component analysis on it directly. Instead, we choose a vector space tangent to the space, at a certain fixed point, for analysis. The tangent space at any point $\psi \in \mathbb{S}_\infty$ is given by: $T_\psi(\mathbb{S}_\infty) = \{v \in \mathbb{L}^2 | \int_0^1 v(t)\psi(t)dt = 0\}$. In the following, we will use the tangent space at μ_ψ to perform analysis. Note that the outcomes of Algorithm 2 include the Karcher mean μ_ψ and the tangent vectors $\{v_i\} \in T_{\mu_\psi}(\mathbb{S}_\infty)$. These tangent vectors, also called the *shooting vectors*, are the mappings of ψ_is into the tangent space $T_{\mu_\psi}(\mathbb{S}_\infty)$, as depicted in Figure 1. In this tangent space we can define a sample covariance function: $(t_1, t_2) \mapsto \frac{1}{n-1}\sum_{i=1}^{n} v_i(t_1)v_i(t_2)$. In practice, this covariance is computed using a finite number of

points, say T, on these functions and one obtains a $T \times T$ sample covariance matrix instead, denoted by K_ψ. The singular value decomposition (SVD) of $K_\psi = U_\psi \Sigma_\psi V_\psi^\mathsf{T}$ provides the estimated principal components of $\{\psi_i\}$: the principal directions $U_{\psi,j}$ and the observed principal coefficients $\langle v_i, U_{\psi,j} \rangle$.

Using the singular values, $\{\sigma_{\psi,k}\}$ and the following control limits,

$$\begin{aligned} UCL &= 3\sqrt{\sigma_{\psi,k}} \\ CL &= 0 \\ LCL &= -3\sqrt{\sigma_{\psi,k}} \end{aligned}$$

for a specified singular value. For each new observation, the function is aligned to μ_q forming the corresponding warping function, γ. The SRSF of γ (ψ) is formed and the corresponding shooting vector is found. From the shooting vector, the score is formed and checked against the control limits and those shooting vectors with scores beyond the control limits are marked.

4. Results

In this section we test the distance and fPCA approach to functional statistical process control on a simulated data and a real data set. The real data is the shockwave curves, also known as "onionskin" curves, from Storlie et al. (2013). In both data sets using the provided fSPC approaches we measure the *shape* of the functions and if they change from a specified mean.

4.1 Numerical Simulation

To illustrate the developed methods for functional statistical process control, we evaluated each method on a simulated data set constructed from a mean function

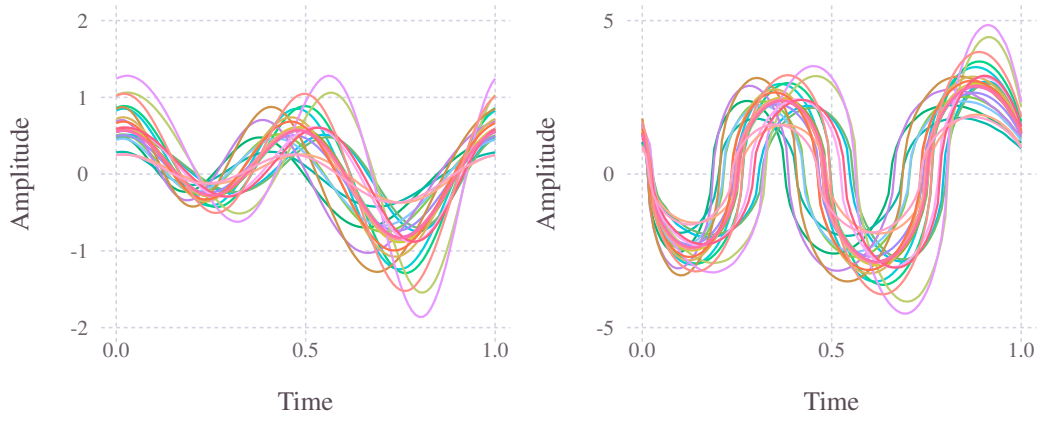
$$\mu_f(t) = 0.2\sqrt{2}\sin(t) + 0.4\sqrt{2}\cos(t)$$

where $t \in [0, 2\pi]$. The data then was constructed from μ_f using

$$f_i(t) = c_i(\mu_f \circ \gamma_i)$$

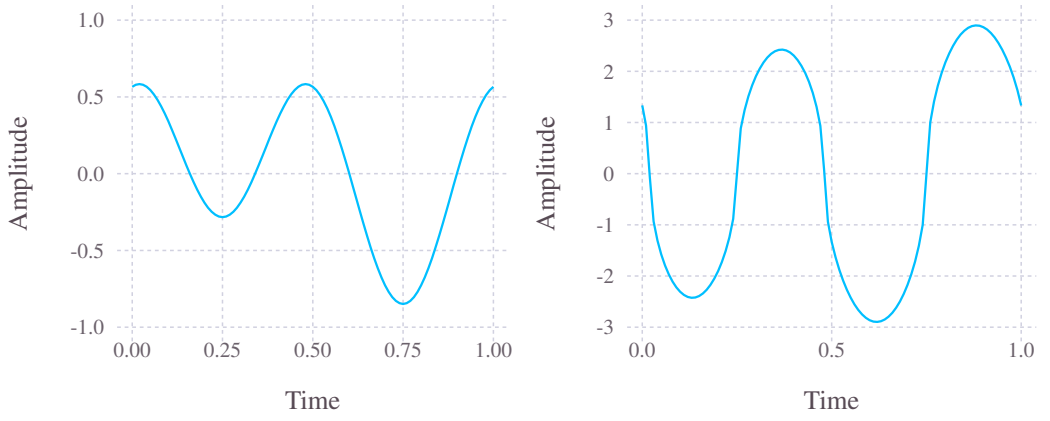
where $c_i \sim \mathcal{N}(1, 0.05)$ for in-control data and $c_i \sim \mathcal{N}(1, 0.4)$ for out of control data. For both cases the data was randomly warped as described in Tucker et al. (2014) where the warping was increased for the out of control case. There are a total of 40 functions generated with 20 from both cases. The functions $\{f_i\}$ and the corresponding SRSF's $\{q_i\}$ are presented in Figures 2a and b, respectively. The mean function μ_f and mean SRSF μ_q which were found from the 20 in-control samples using elastic functional alignment algorithm (Srivastava et al., 2010) are presented in in Figures 3a and b, respectively.

First, we calculated the distances d_a and d_p from μ_q as defined in (2.1) and (2.2), respectively. Figure 4 presents the distances for each of the 40 functions with orange representing d_a and the purple representing d_p. A threshold was set at 0.16 for d_a and 0.05 for d_p and any function that was above this threshold was flagged and marked in red in the figure. Almost all of the 40 out of control samples were flagged, there was one in which the d_a did not flag it however, the corresponding d_p was over the threshold. This demonstrates the power of the two metrics and the ability to measure the two types of variability. Depending on the desire of the user we can give priority to one or the other. Moreover, by using the warping function, γ, we can demonstrate where in time the function was out of control as well. The thresholds in this case were set tight and could be relaxed given the users requirement.



(a) $\{f_i\}$ (b) SRSF $\{q_i\}$

Figure 2: Original simulated data which contains variation in amplitude and phase showing both the original functions and SRSFs.



(a) μ_f (b) SRSF μ_q

Figure 3: Mean function and SRSF found using Elastic Functional Alignment for the simulated data

Second, we performed vertical and horizontal fPCA as described in Section 3. The control limit (CL) was determined for the simulated data to be 2.12 for the vertical and 0.269 for the horizontal based on the in-control samples. The first eigenvalue was used in both methods to set the control limits. Figure 5 presents the control plot with the CLs shown as the black horizontal lines for the vertical fPCA. Any function that was found to be outside the control limit is marked in red. Two of the out of control samples were determined to be in control, which is slightly different than the distance measurements. Additionally, one of the in control samples was right at the CL and was flagged. This method is similar to standard SPC methods based PCA, however it incorporates amplitude and phase variability found in functional data.

Figure 6 presents the control plot with CLs shown as the black horizontal lines for the horizontal fPCA. Most of the 20 out-of control samples were flagged as out of limits. Those that were not had a warping similar to the original 20 functions and should not have been flagged as they the phase was similar to the original model. Therefore, depending on what one wishes to monitor, amplitude or phase, would determine the model you would use.

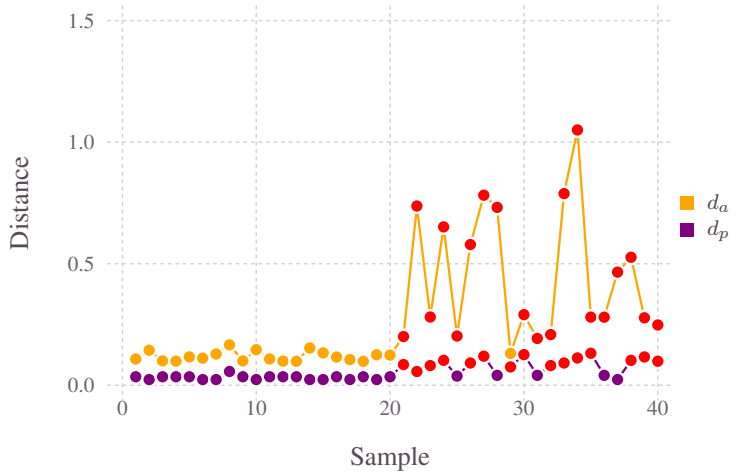


Figure 4: The distances d_a and d_p calculated for the simulated data and out of control samples marked as red.

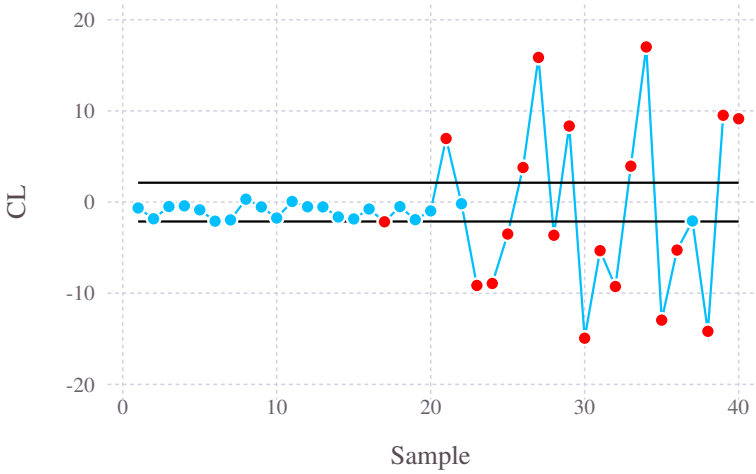


Figure 5: Vertical fPCA scores for the simulated data with the control limits shown and samples outside the control limits marked with red.

4.2 Real Data

Next we demonstrate the elastic fSPC methods on shock wave curves collected at Los Alamos National Laboratory (Storlie et al., 2013). These curves are known as "onionskin" curves and are a recording of a shock wave as it arrives at a streak camera. The challenge with these curves is to characterize the population of the curves and compare them data from tests to a baseline set of curves for changes in shape. Changes in shape of the curves are related to a change in the quality of explosive material.

The curves were smoothed using a 3rd order B-Spline basis and Figure 7a presents 9 onionskin curves that are considered the baseline from historical tests. Figure 7b presents 4 onionskin curves from a different detonator (an LX07 detonator). Because of the change in the detonator, the shape and amplitude are different from the baseline curves and the question is how different are they.

The original onionskin curves were aligned using the elastic function alignment and the computed mean function is presented in Figure 8. Next, we calculated the distances d_a and d_p from μ_q as defined in (2.1) and (2.2), respectively. Figure 9 presents the distances for the 13 onionskin curves with orange representing d_a and the purple representing d_p. The

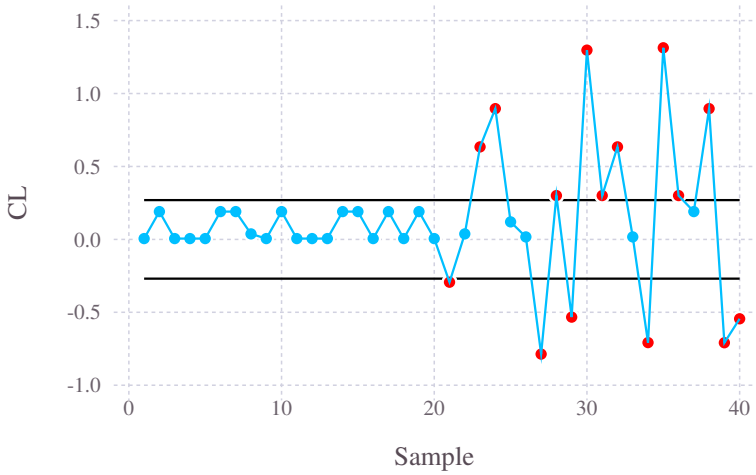
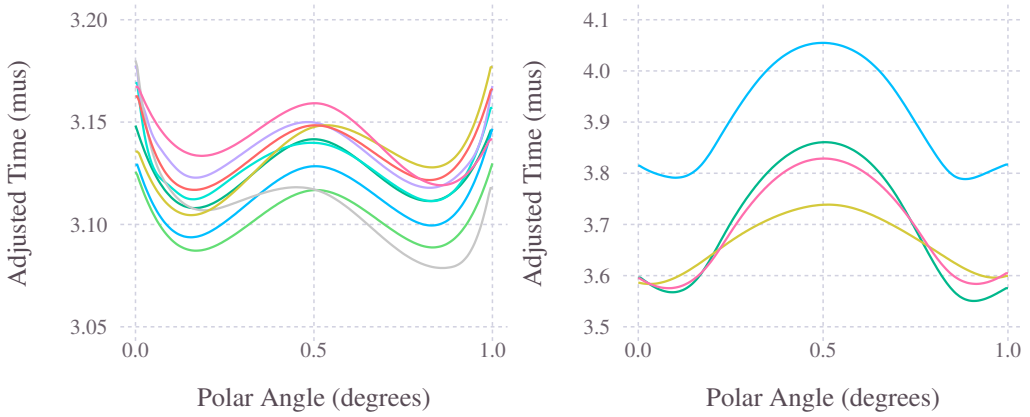


Figure 6: Horizontal fPCA scores for the simulated data with the control limits shown and samples outside the control limits marked with red.



(a) Baseline onionskin shock wave curves from historical tests.

(b) New set of onionskin curves from a new LX07 detonator.

Figure 7: Onion Data

first 9 are the original curves and the last 4 are thew new curves. A threshold was set at 0.2 for d_a and d_p and any function that was above this threshold was flagged and marked in red in the figure. All four new curves were marked out of control by d_a due to their large amplitude differences and the last 3 were marked out of control by d_p with the last two being on the edge of the threshold. These results coincide with the results obtained by Los Alamos National Laboratory.

We then performed vertical and horizontal fPCA on the onionskin data with models being formed using the baseline 9 curves. The control limit (CL) was determined to be 2.24 for the vertical and 0.9095 for the horizontal. The first eigenvalue was used in both methods to set the control limits. Figure 10 presents the control plot with the CLs shown as the black horizontal lines for the vertical fPCA. The second function of the LX07 was flagged above the control limits, through the first and third are very close the UCL. Overall, the functions are very close to being different in amplitude from the baseline set.

Figure 11 presents the control plot with CLs shown as the black horizontal lines for the horizontal fPCA. The second curve is flagged again as out of the control limits. This would relate to a large change in shape and the other three curves are similar in phase to

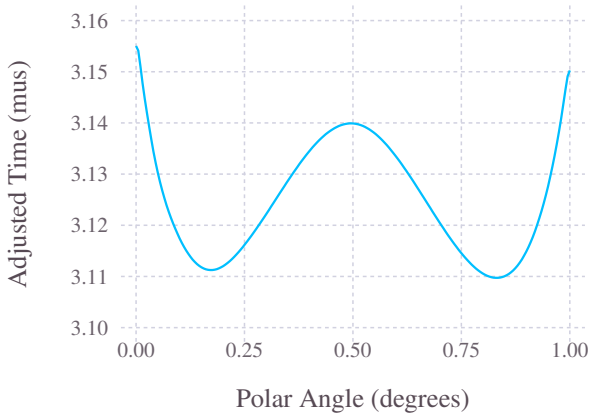


Figure 8: Mean onionskin curve calculated using elastic function alignment.

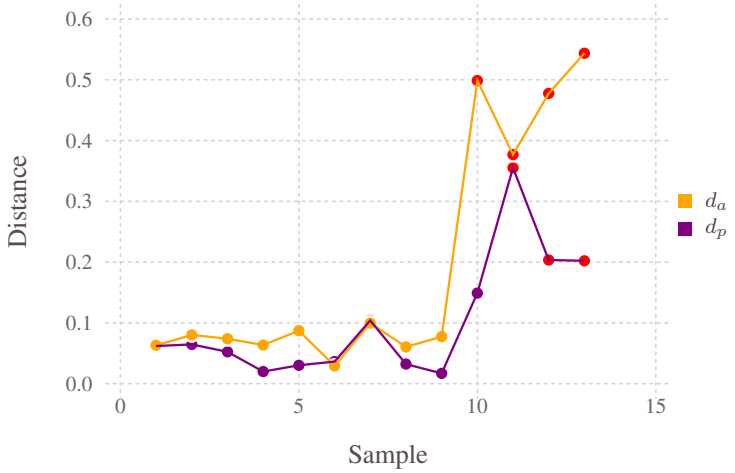


Figure 9: The distances d_a and d_p calculated for the onionskin data and out of control samples marked as red.

the baseline.

The PCA-based approach does not flag as many functions as the distance based approach. This has to do with how high or low we set the CLs. Overall, both methods provide a way to measure the shape of the curves and monitor as curves are sampled from a process.

5. Conclusions

The statistical modeling and monitoring of functional data for process control is difficult task. We have proposed an approach that solves the problem of monitoring a process using functional data that takes into account both amplitude and phase variability of the data. The main idea is to use an elastic distance to separate the given functional data into phase and amplitude components, and to develop individual models for these components. Then based on those models set control limits and monitor the elastic distance or fPCA score for samples whose function go beyond those limits. The strengths of these models are illustrated in two ways: simulated and real data. In both cases, each method was able to identify when there were changes in the data in both amplitude and phase, essentially detecting when the shape of the function has changed. This in turn relates to a change in the process being monitored.

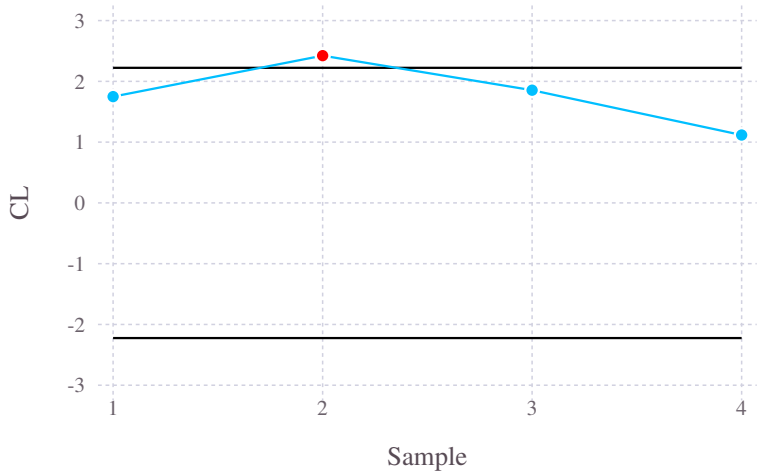


Figure 10: Vertical fPCA scores for the simulated data with the control limits shown and samples outside the control limits marked with red.

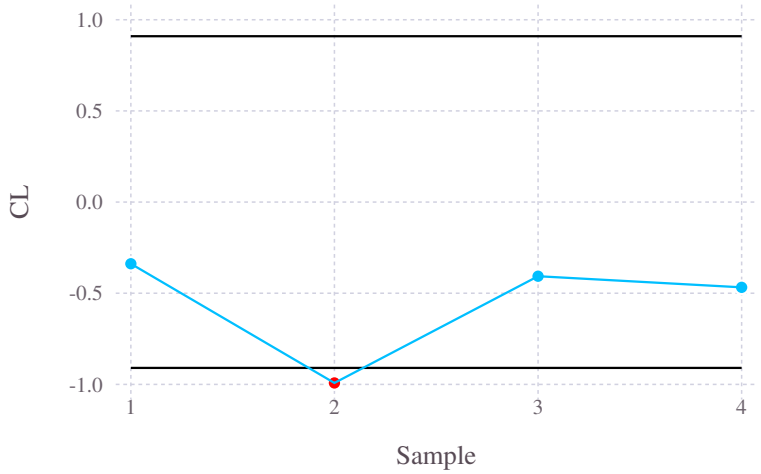


Figure 11: Horizontal fPCA scores for the onionskin data with the control limits shown and samples outside the control limits marked with red.

References

Bertsekas, D. P. (1995), *Dynamic Programming and Optimal Control*, Athena Scientific.

Cheng, W., Dryden, I. L., and Huang, X. (2013), "Bayesian registration of functions and curves," *arXiv:1311.2105 [stat.ME]*.

Hall, P. and Van Keilegom, I. (2007), "Two-sample tests in functional data anlaysis starting from discrete data," *Statistica Sinica*, 17, 1511–1531.

Horváth, L. and Kokoszka, P. (2012), *Inference for Functional Data with Applications*, Springer Science & Business Media.

Huang, W. (2014), "Optimization algorithms on riemannian manifolds with applications," Ph.D. thesis, Florida State University.

James, G. and Sood, A. (2006), "Performing hypothesis tests on the shape of functional data," *Computational Statistics and Data Analysis*, 50, 1774–1792.

Kneip, A. and Ramsay, J. O. (2008), "Combining Registration and Fitting for Functional Models," *Journal of the American Statistical Association*, 103.

Kurtek, S., Srivastava, A., and Wu, W. (2011), "Signal Estimation Under Random Time-Warpings and Nonlinear Signal Alignment," in *Proceedings of Neural Information Processing Systems (NIPS)*.

Ramsay, J. O. and Silverman, B. W. (2005), *Functional Data Analysis*, Springer.

Srivastava, A., Klassen, E., Joshi, S., and Jermyn, I. (2011a), "Shape Analysis of Elastic Curves in Euclidean Spaces," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 33, 1415–1428.

Srivastava, A., Wu, W., Kurtek, S., Klassen, E., and Marron, J. S. (2010), "Statistical Analysis and Modeling of Elastic Functions," *Journal of American Statistical Association*, submitted.

— (2011b), "Registration of Functional Data Using Fisher-Rao Metric," *arXiv:1103.3817v2 [math.ST]*.

Storlie, C. B., Fugate, M. L., Higdon, D. M., Huzurbazar, A. V., Francois, E. G., and McHugh, D. C. (2013), "Methods for characterizing and comparing populations of shock wave curves," *Technometrics*, 55, 436–449.

Tucker, J. D., Wu, W., and Srivastava, A. (2013), "Generative models for functional data using phase and amplitude separation," *Computational Statistics and Data Analysis*, 61, 50–66.

— (2014), "Analysis of signals under compositional noise with applications to SONAR data," *IEEE Journal of Oceanic Engineering*, 39, 318–330.