Bounding Uncertainty in Functional Data: A Case Study

Caleb King
JMP Division
SAS Institute

Nevin Martin
Statistical Sciences
Sandia National Laboratories

J. Derek Tucker
Statistical Sciences
Sandia National Laboratories

Abstract

Functional data is becoming more prevalent as a primary source of information across a wide range of industries. As such, new methods for handling such data are constantly being developed and improved. A particularly challenging aspect of functional data is the ability to quantify uncertainty, specifically within the context of bounding that uncertainty to assess product behavior. In this case study, we discuss several settings in which novel or improved methods were used to create data-driven bounds on functional data. An initial attempt at a viable solution did not render the intended results. This prompted the development of a new methodology for quantifying uncertainty in functional data, one which has continued to be promoted at Sandia National Laboratories.

Keywords: Functional Data Analysis, Functional Principal Components, Reliability Specifications, Tolerance Bounds, Uncertainty Quantification

1 Introduction

Functional data have become more prevalent in as a primary source of information in a variety of industries, thanks in part to the rise of improved sensor technology and sampling methods. Additionally, there has been a large amount of research to develop better methods for analyzing such data, such as functional principal component analysis (fPCA), functional alignment, and functional regression models (see Ramsay

and Silverman (2005); Srivastava and Klassen (2016) for an introduction to the most common methods). Even so, there is still space in this area for additional research in developing methods to answer a wide range of engineering and statistical problems. One such area is the construction of bounding curves, such as confidence or tolerance bounds. The purpose for creating such bounding curves can be varied, but one specific purpose we discuss here is for use in defining and/or assessing product specifications.

Constructing bounding curves can be a difficult task since functional data do not exhibit random variation in the same way as univariate data. Initial attempts have involved converting the functional data to univariate data through summary metrics, such as a mean or maximum value. However, these data reduction techniques obfuscate the true nature of the functional variation and so may lead to specifications that are either insufficient or grossly conservative. While there are times where a univariate metric may be appropriate (i.e., a maximum threshold for a voltage curve), in general a bounding curve that respects the functional variation of the data, however complex it may be, will always prove superior to univariate metrics in capturing the different types of variation that may occur.

Recently, Storlie et al. (2013) developed a method to test the shape of a population of curves using a B-Spline basis, and a hierarchical Gaussian process approach to form confidence intervals. Rathnayake and Choudhary (2016) developed tolerance bounds for functional data using functional principal component analysis (fPCA). Sun and Genton (2011) developed a boxplot display for functional data, which provides a nice visualization technique for a sample of functions and is also useful for detecting functional outliers. However, these methods do not take into account the amplitude and phase variability present in functional data, i.e., they assume that the data (1) does not need to be aligned or (2) has already been aligned using some unrelated criterion. The first assumption is unrealistic in process control applications while the second approach results in suboptimal solutions due to the distjointedness of the two sources of variation. A more systematic approach is to develop methods that build the alignment step into the statistical procedure.

To illustrate the difficulty of assessing functional uncertainty as well as highlight some of the novel approaches to overcome this difficulty, we present a selection of applications from Sandia National Laboratories (SNL) in which functional data analysis (FDA) techniques were used to help assess and bound uncertainty in functional data. Our case study will proceed as follows. First, we will discuss a motivating functional data application as well as a brief discussion of initial techniques used to bound the functional responses. Then, we will present an initial approach to solving the problem and then discuss why this approach failed to achieve the desired goals. Next, we will present an improved approach developed to address the concerns of the initial attempt.

Although this improved method was not implemented in the motivating application, it has and continues to be implemented in similar settings across SNL, which we illustrate through two additional applications. In presenting each methodology, we will provide an overview of each technique, but will avoid too many details on the first method as the intent is not to promote it as a viable option.

This case study may be novel in presenting an initial failed solution in addition to the successful one. By doing so, we intend to shed light on the entire problem-solving process, rather than focus solely on the solution. Success in research is almost never gained in a single step, but often through long bouts of missteps and stumbles. In addition, showing a failed result can instruct other researchers on practices to avoid, eliminating repeated publications on methods that are clearly inappropriate. Publication bias is a pre-iminent issue as of late and has given rise to a replication crisis in scientific research (Ioannidis, 2005). Among the many proposed solutions, one approach is to encourage the publication of null or even negative results (Jena, 2017). While the case study presented here may not necessarily be classified as presenting a "null" result, the idea of being open to publishing case studies in which the outcome was unexpected may be one worth considering. As this very case study illustrates, new and impactful research can emerge from the ashes of a previous attempt.

2 Initial Application

The initial motivating application involved the need for a more data-driven methodology in specifying environmental requirements for a munitions system. Over the course of their lifetimes, military munition systems are exposed to a variety of environments. Of particular interest are environments encountered during transportation, most notably shocks and vibrations. In this application, the focus was on mechanical vibrations experienced by munitions systems carried by aircraft during flight. Engineers wished to characterize the nature of these mechanical vibrations as a function of air speed and pressure, among other factors. From this, they could then compute a bounding curve to serve as the basis for creating product specifications for the munitions system.

The data used in the computation were collected using a carefully designed series of test flights at selected speeds and altitudes. A mock munitions system is carried on the wings of the aircraft, with accelerometers placed at key points across the test object. Once a flight test is completed, the data are downloaded and converted to Acceleration Spectral Densities (ASDs), which express the energy per unit time of the vibration as a function of frequency (Lalanne, 2014). An example of the data are shown in Figure 1. Note that this data is different from more common forms of functional data in that the domain of the data is not time or location.

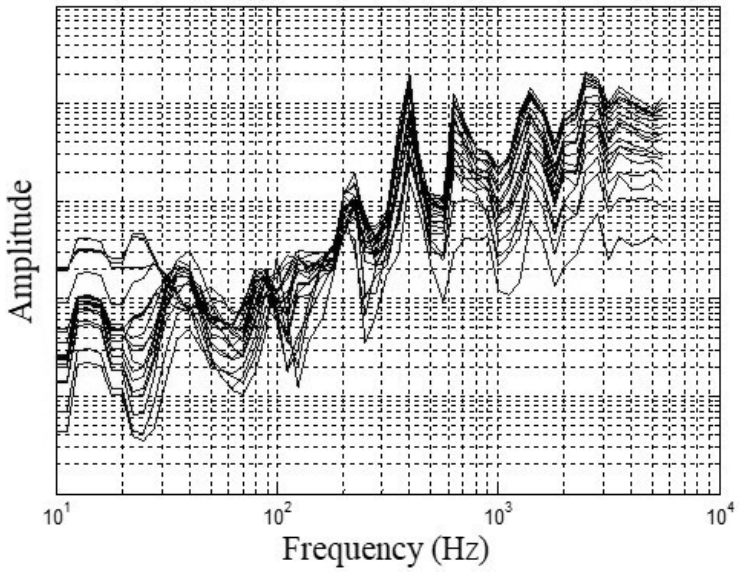


Figure 1: Example of a sampling of ASD curves.

Common practice for the engineers was to create specifications by taking this data and marking a series of lines that bound the maximum values observed at each frequency. These lines would be drawn either right at the maximum values or some specified value above them. There was no explicit formulation for how these bounds incorporated the functional variation into their computation. Furthermore, the resulting specifications were often meant to hold at environments not seen during test flights and so extrapolation was incredibly difficult, if not impossible, using this approach. Thus, an alternative approach was to compute the square root of the area under the curve, which gives the root mean square acceleration (denoted by G_{rms}, see Lalanne (2014) for further details), which could then be used as a univariate response in a standard regression model. While it solves the extrapolation issue, the resulting specifications could only be in terms of G_{rms}, resulting in a loss of functional information.

Realizing these limitations, the engineers sought out a more data-driven methodology. The goal was a methodology that would compute a general data-driven bound of on the ASDs while also allowing that bound to be explicitly tied to environmental factors, so that it could be applied to untested regions.

3 First Attempt

The initial approach to generating the required methodology was to extend the work of Rathnayake and Choudhary (2016), which uses functional principal component analy-

sis (fPCA) to break down the mean-centered functional data into principal component vectors. Specifically, for a collection of functional curves $\mathbf{Y}_i(t), i = 1, 2, \ldots, n$, where t represents a general input, and mean $\boldsymbol{\mu}(t) = 1/n \sum_{i=1}^{n} \mathbf{Y}_i(t)$,

$$\mathbf{Y}_i(t) - \boldsymbol{\mu}(t) \approx \sum_{k=1}^{K} s_{ki}\phi_k(t), \tag{3.1}$$

where $\phi_k(t)$ is a principal component vector representing a principal mode of functional variation and s_{ki} is the ith *loading* or *score* representing an individual scaling of the kth principal component vector for the ith functional response curve Ramsay and Silverman (2005). The number of principal components K is specified by the user and can be determined by how many components are needed to capture a specified percentage of the variation in the data. Rathnayake and Choudhary (2016) applied a bootstrap procedure on the scores which could then be used to construct functional tolerance bounds through a simple rearrangement of Equation (3.1) and then taking the required quantiles from the bootstrap sample of functional curves.

A limitation of this method was that it did not consider the presence of experimental factors. At the time, we felt that it could easily be extended by treating some of the scores as responses in linear models of the factor inputs. In fact, some of the principal component scores showed clear trends with one ore more factors, as illustrated in Figure 2. There, the first principal component showed a strong linear relationship with dynamic pressure (Q), a quantity related to both aircraft speed and altitude. Typically, these relationships were found among the first two principal components, the latter ones being considered random noise.

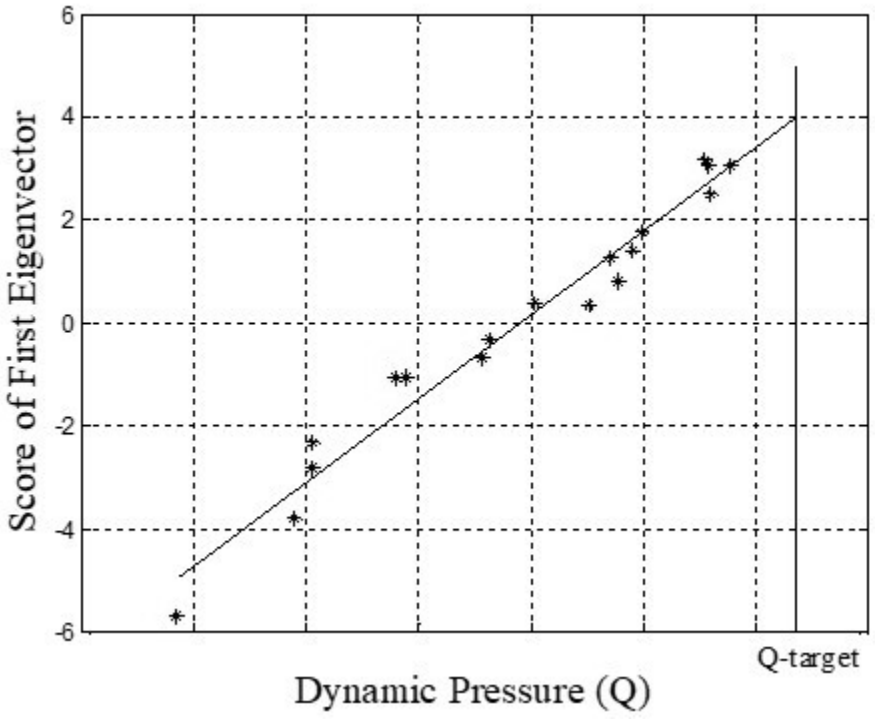


Figure 2: Example of the initial modeling approach. Here, the scores from the first eigenvector hold a linear relationship with dynamic pressure, Q. Q-target represents the specific environment for which tolerance bounds and ultimately product specifications need to be computed.

In addition to regressing scores on experimental factors, we also decided on using a parametric bootstrap to generate the tolerance bounds. By construction, the scores have mean 0 and variance equal to the kth eigenvalue of the principal component decomposition of the functional covariance matrix $\mathrm{E}\{(\mathbf{Y}(t) - \boldsymbol{\mu}(t))(\mathbf{Y}(t) - \boldsymbol{\mu}(t))'\}$. Furthermore, if we denote by $\mathbf{S}$ the matrix of scores $\{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_K\}$, the columns of $\mathbf{S}$ are orthogonal, again by property of fPCA. Loosely based on these properties, we considered treating the scores as realizations of a multivariate normal distribution for ease of computation. In this way, we could rely on traditional normal theory to generate parametric bootstrap score samples of the given quantile p. From these scores, we could then rebuild the functional curves and choose the appropriate confidence percentile $1 - \alpha$ to yield the tolerance bound.

The initial tolerance bound construction procedure involved two steps. First, relationships between principal components and covariates would be identified and estimated. The models were also treated as specific to their respective principal components, so parameter estimates were component-specific. Once the models were estimated, they could then be used to extrapolate to a target environment (here in terms of

dynamic pressure). At that point, the parametric bootstrap would be used to generate the tolerance bound as briefly described above.

The parametric boostrap algorithm is summarized below. ***Note that this is only for illustrative purposes and should not be used in practice due to the flaws which will be discussed in the next section.***

1. If the kth principal component has a linear model associated with it, compute $\widehat{\mu}_k = \mathbf{x}_0^T \widehat{\boldsymbol{\beta}}$. Otherwise, compute $\widehat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} s_{ki}$.
2. For $b = 1, 2, \ldots, B$,
   (a) Simulate $\widetilde{\sigma}_{kb} \sim \widehat{\sigma}_k \sqrt{\frac{df_k}{\chi^2_{df_k}}}, k = 1, 2, \ldots, p$, where $\widehat{\sigma}_k$ is the estimated standard deviation of the scores for the kth principal component, df_k are the appropriate degrees of freedom, and $\chi^2_{df_k}$ is a Chi-squared random variable with df_k degrees of freedom.
   (b) If the kth principal component has a linear model associated with it, compute $\widetilde{\sigma}^2_{\mu} = \widetilde{\sigma}^2_{kb} \mathbf{x}_0^T (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0$, where $\mathbf{X}$ is the model matrix and $\mathbf{x}_0$ is the target level. Otherwise, compute $\widetilde{\sigma}^2_{\mu} = \widetilde{\sigma}^2_{kb}/n$.
   (c) Simulate $\widetilde{\mu}_{kb} \sim \text{Normal}(\widehat{\mu}_{kb}, \widetilde{\sigma}^2_{\mu})$.
   (d) Compute $\bar{\mathbf{Y}}_b(t) = \bar{\mathbf{Y}}(t) + \sum_{k=1}^{K} \widetilde{\mu}_{kb}\phi_k(t)$ and $\mathbf{V}_b(t) = \sqrt{\sum_{k=1}^{K} \phi_k(t)^2 \widetilde{\sigma}^2_{kb}}$, where $\bar{\mathbf{Y}}$ is the pointwise mean of the original data and $\phi_k(t)$ is the kth eigenfunction.
   (e) Compute $\mathbf{P}_b(t) = \bar{\mathbf{Y}}_b(t) + z_p \mathbf{V}_b(t)$, where z_p is the pth percentile of the standard normal distribution.
3. The pointwise $(1-\alpha)$-level upper tolerance bound is given by the $(1-\alpha)$-percentile of the $P_{bj}, j = 1, 2, \ldots, J$.

4 Flaws of the Initial Approach

To assess the validity of the approach, we conducted simulation studies to determine the accuracy of the tolerance bound coverage. The simulation studies were divided into two main groups: one which simulated the same number of curves as the ASD data and one with 10 replicates of each curve. Simulated curves were created by first simulating scores from a multivariate normal distribution with the estimated model means from the ASD data treated as the true means. The scores are then combined with the eigenfunctions and mean curve from the original ASD data. The range of tolerance bounds considered were 99-90, 99-50, 99-75, 95-90, and 95-50, where the first number represents the percentile of interest and the second number represents

the confidence level. It was here that an unfortunate flaw was discovered. We found that actual coverage could vary widely over the domain of the functional data and the mean actual coverage was inconsistent. For example, in one simulation illustrated in Figure 3, the mean true pointwise coverage for a 99-90 tolerance bound (bound on the 99th percentile with 90% confidence) was actually closer to 85%, fluctuating between 77% and 96% across the function domain. The true coverage only seemed to stabilize with lower confidence levels, as illustrated in Figure 4, making it impractical for reasonable use. The behavior did not seem to improve with the replicated data. These results were quite disturbing and a new approach was desperately needed.

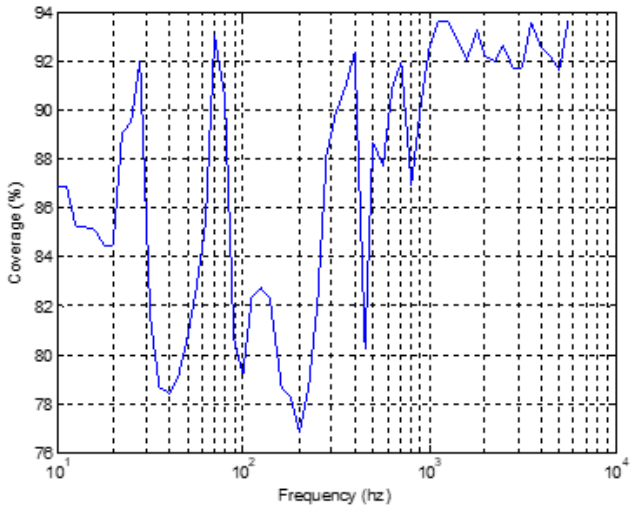


Figure 3: Example of simulation results. The intended coverage on the 95th percentile is 90%. However, as can be seen here, the actual coverage varies significantly.

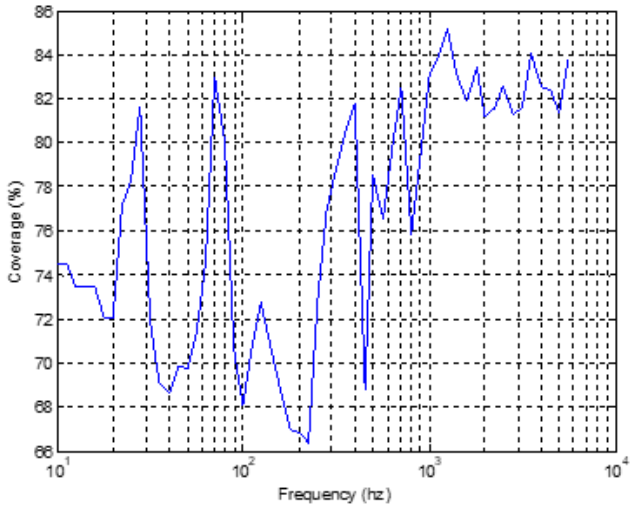


Figure 4: Example of simulation results for lower coverage tolerance bound. The intended coverage on the 95th percentile is 75%. Note that the pointwise bounds are more stable here.

The initial method was presented at the 2016 Joint Statistical Meetings (Thomas et al., 2016) with the intent of showcasing the inherent flaws in the method and requesting feedback on alternative approaches to improve the accuracy of the coverage. A fellow SNL colleague, James Derek Tucker, was in attendance and was able to immediately diagnose the source of the issues. The problem lay in creating a model for the component scores. By doing so, this actually eliminated that source of variation, leading to the counterintuitive behavior we were seeing. In essence, the model was not using a metric that is a proper distance for functional data where the data contained both variability in time and amplitude (this will be explained in the next section). By failing to use a proper metric, the mean function is no longer representative of the underlying data mechanism which will lead to false conclusions. This motivated a new collaboration and led to an improved methodology for constructing tolerance bounds on functional data, which we will now present.

5 Improved Approach

The basis for the updated tolerance bound methodology stems from the fact that most functional data have two types of variability - amplitude (vertical) and phase (horizontal). Most standard FDA approaches do not account for both types of variability in a principled fashion and so an *elastic* method is needed (Srivastava and Klassen, 2016). These variabilities can be seen in the ASD data presented in Figure 1 by noting that the "peaks" and "valleys" of the functions are not perfectly aligned across

frequency (i.e., phase variability) and that the functions vary across amplitude (i.e., amplitude variability). The previous method applied fPCA on the original data which only considered the amplitude variability, resulting in bounds that were inaccurate and nonphysical. Therefore, through collaboration, a new method was developed that uses a more natural approach to characterizing functional data by using fPCA on a joint representation of the phase and amplitude spaces.

By this point, due to the unfortunate aforementioned coverage problems with the above methodology, the engineers on the military munitions project had decided to revert back to their original method for creating specifications for the ASD data. However, the new method for elastic functional tolerance bounds was published in the Journal of Applied Statistics (Tucker et al., 2020) along with the results after applying the method to a weld residual stress problem at SNL. The weld residual stress data are represented as stress versus normalized weld depth and are an important part of estimating crack-growth in nuclear power plant piping systems. The goal of the work was to develop a method to estimate tolerance bounds which would be assessed by subject matter experts prior to using them in crack-growth calculations. The phase variability component of the data was a particularly crucial aspect to capture as it directly affects the crack-growth calculations. The new method allowed for the estimation of tolerance bounds on the weld residual stress data that accounted for both phase and amplitude variabilities.

5.1 Elastic Functional Data Approach

The new methodology begins by characterizing the phase and amplitude variabilities of the data. This is an important first step as this characterization is used to fit an fPCA model that generates new functions that represent the phase and amplitude variability in the original data. These generated functions are then used in a bootstrap procedure to estimate tolerance bounds on the phase and amplitude spaces separately, resulting in bounds that appropriately handle the variabilities seen in functional data. The remainder of this section describes each step of this approach, with more detail available to the reader in Tucker et al. (2020) and methods used in the results are available in R package `fdasrvf` package available on CRAN[1].

To characterize the phase and amplitude variabilities in the original data, the functions are first aligned along the x-axis using *warping functions* (Srivastava and Klassen, 2016). The resulting aligned functions are used to characterize amplitude variability, while the warping functions are used to characterize the phase variability. These warping functions, Γ, are defined as a set of functions (γ) on the interval $[0, 1]$ that have

[1]MATLAB code is also available at https://github.com/jdtuck/fdaSRSF_MATLAB

the following properties:

- γ has one-to-one correspondence from $[0,1]$ into $[0,1]$,
- $\gamma(0) = 0$ and $\gamma(1) = 1$, and
- both γ and γ^{-1} are differentiable.

For any function $f \in F$, where F is the set of all real-valued functions on $[0,1]$, f is warped by γ through composition: $f \circ \gamma = f(\gamma(t))$. The goal is to align any two functions f_1 and f_2 using the following amplitude distance metric:

$$d_a(f_1, f_2) = \inf_{\gamma \in \Gamma} \|q_1 - (q_2 \circ \gamma)\sqrt{\dot{\gamma}}\|, \tag{5.1}$$

where

$$q_i(t) = \text{sign}(\dot{f}_i(t))\sqrt{|\dot{f}_i(t)|} \tag{5.2}$$

is known as the square-root slope function (SRSF) and $\dot{f}$ is the derivative of f with respect to t. By using the SRSF for alignment (instead of the original functions), the amplitude distance is a proper distance; that is, this distance metric is non-negative, symmetric, and satisfies the property $d_a(f_1, f_3) \leq d_a(f_1, f_2) + d_a(f_2, f_3)$[2]. Because of this, it is now possible to define statistics on the functional space (e.g., mean and variance) that naturally handle the known variabilities. Once the SRSFs are aligned, they can then be mapped back to F to obtain the aligned original functions. Additional details about SRSFs are given in Srivastava et al. (2011); Tucker et al. (2013); Marron et al. (2015) and Srivastava and Klassen (2016).

After aligning the functions, the phase variability can be characterized using the same functions γ that were used to align the data. The space of Γ is a nonlinear manifold which creates some difficulty in defining a distance between two warping functions γ_1 and $\gamma_2 \in \Gamma$. In order to define a distance, we use the square-root of the derivative of γ:

$$\psi = \sqrt{\dot{\gamma}}. \tag{5.3}$$

This representation allows the set of all $\psi \in \Psi$ to be the positive orthant of a Hilbert sphere ($\mathbb{S}_\infty^+$). Although this is still an infinite dimensional manifold, it now has a known geometry which we can utilize. Therefore, the phase distance between γ_1 and γ_2 can then be calculated as the arc-length between ψ_1 and ψ_2 on the Hilbert sphere:

$$d_p(\gamma_1, \gamma_2) = d_\psi(\psi_1, \psi_2) \equiv \cos^{-1}\left(\int_0^1 \psi_1(t)\psi_2(t)dt\right) . \tag{5.4}$$

[2]Mathematical proof provided in Srivastava and Klassen (2016)

To be able to compute standard statistics, the geometry of Ψ is then further simplified by analyzing the warping functions on a tangent space defined as

$$T_\psi(\Psi) = \{v \in \mathbb{L}^2 | \int_0^1 v(t)\psi(t)dt = 0\}. \tag{5.5}$$

Here, v is a tangent space vector and allows us to perform fPCA in a linear space and then map back to Γ. For details see Tucker et al. (2020).

Figure 5 shows an example of how phase and amplitude variabilities are decomposed using simulated data. Figure 5a gives the original data y_i, where amplitude and phase variation are both present, noted by the fact that the "peaks" of the curves do not align along either the x or y axes. Figures 5b and 5c give the warping γ_i and aligned f_i^* functions, respectively, while Figure 5d shows the SRSFs q_i. The f_i^* functions have been aligned such that the only variability remaining is variability along the y-xis (amplitude). The corresponding γ_i provide a measure of the phase variability in the original data.

Once the variabilities have been characterized, an fPCA model can then be fit on a joint representation of the aligned and warping functions. This idea expands on the work of Lee and Jung (2017) by using a concatenated function (g^C) on the extended domain $[0, 2]$ for some $C > 0$, defined as

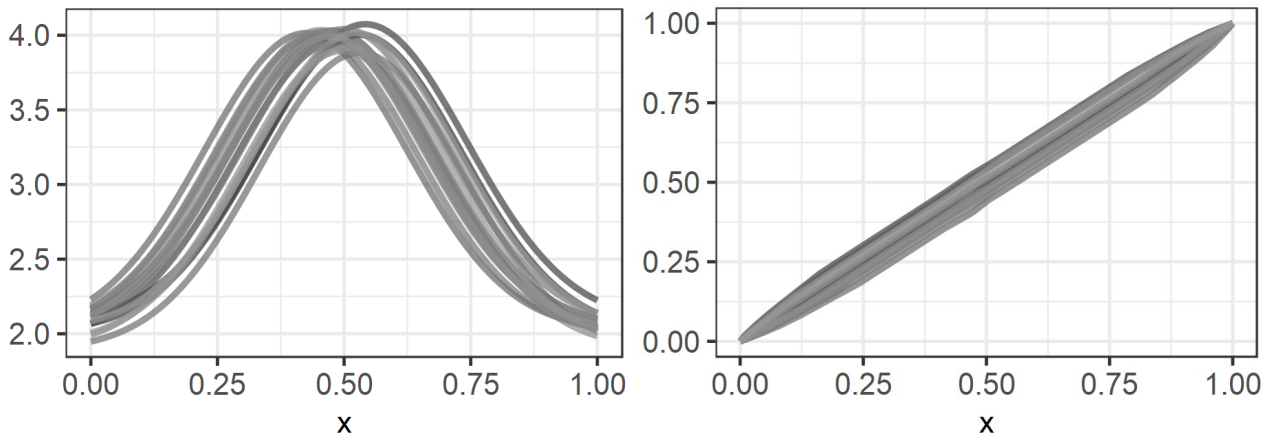
$$g^C(t) = \begin{cases} q^*(t), & t \in [0, 1) \\ Cv(t-1), & t \in [1, 2], \end{cases} \tag{5.6}$$

where C is a scaling factor and q^* is the SRSF of the aligned function f^* and v is the corresponding tangent vector computed from γ. Then, for a sample of amplitude-phase functions $\{g_1^C, \ldots, g_n^C\}$, their sample mean can be defined as $\hat{\mu}_g^C = [\hat{\mu}_{q^*} \quad \hat{\mu}_v^C]$ with a sample covariance matrix of

$$K_g^C = \frac{1}{n-1} \sum_{i=1}^{n} (g_i^C - \hat{\mu}_g^C)(g_i^C - \hat{\mu}_g^C)^\mathsf{T} \in \mathbb{R}^{(2T)\times(2T)} \ . \tag{5.7}$$

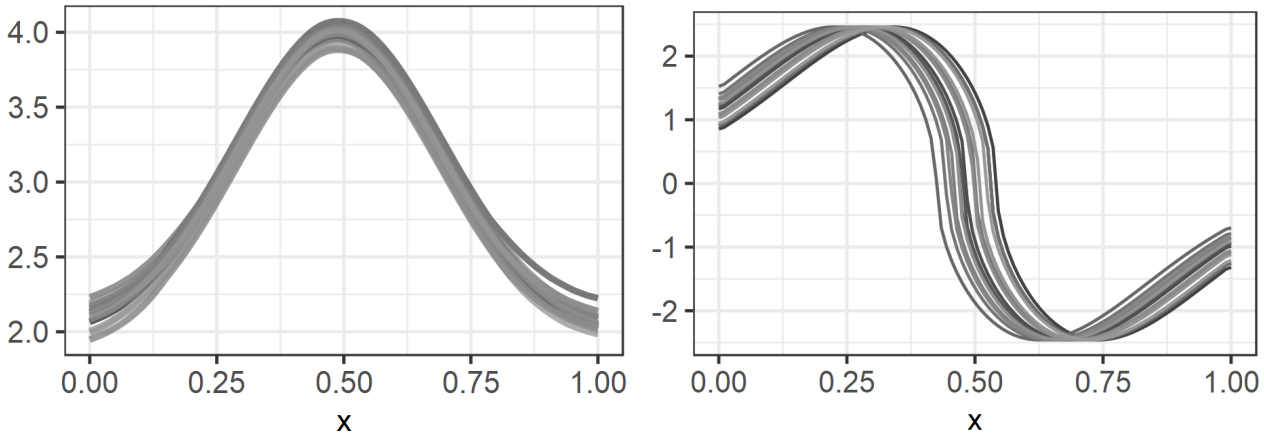
The Singular Value Decomposition $K_g^C = U_g^C \Sigma_g^C (U_g^C)^\mathsf{T}$ provides the joint principal directions of variability in the given amplitude-phase functions as the first $K \leq n$ columns of U_g^C.

After the fPCA coefficients have been estimated, a probability model can be applied to the coefficients, allowing for a distribution on F from which to sample new functions. Let $c = (c_1, \ldots, c_K)$ be the K dominant principal coefficients. These c can then be modeled using a multivariate Gaussian probability distribution with zero mean and covariance Σ, i.e., $c \sim \mathcal{N}_k(0, \Sigma)$, where Σ is a $K \times K$ diagonal matrix with diagonal elements estimated using the eigenvalues of the sample covariance matrix, $\hat{\sigma}_1^C, \ldots, \hat{\sigma}_K^C$. The superscript C shows the dependence on the scaling factor.



(a) Original Functions f_i

(b) Warping Functions γ_i



(c) Aligned Functions f_i^*

(d) Square Root Slope Functions q_i

Figure 5: Notional example of warping functional data.

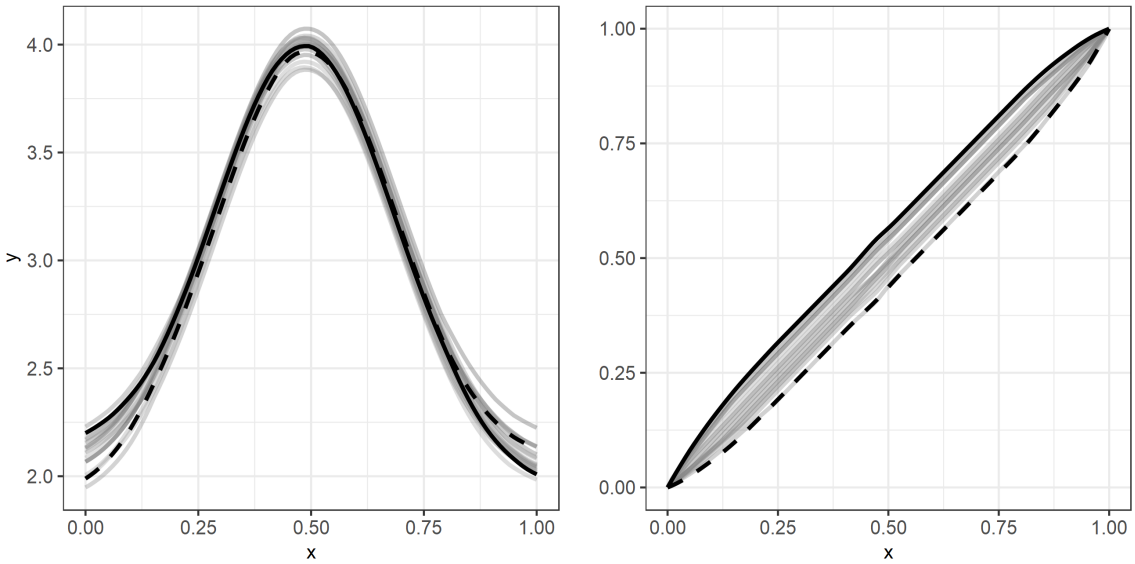
The fPCA model is then used in a bootstrapping procedure to estimate tolerance bounds. For the purposes of illustration, we will describe the construction of a symmetric tolerance interval intended to capture $100p\%$ of the population. An extension to one-sided tolerance bounds is straightforward. The fPCA model is repeatedly sampled using the above described coefficient model and the $(1-p)/2$ and $1-(1-p)/2$ quantiles of the set of random SRSF-based amplitudes and warping functions (denoted as $(q^*_{(1-p)/2}, q^*_{1-(1-p)/2})$ and $(\gamma_{(1-p)/2}, \gamma_{1-(1-p)/2})$, respectively) are estimated. The quantiles are estimated using an extension of the boxplot procedure of Xie et al. (2017) This process is repeated S times such that the $\alpha/2$ and $(1-\alpha/2)$ quantiles of $(q^*_{(1-p)/2}, q^*_{1-(1-p)/2})$ and $(\gamma_{(1-p)/2}, \gamma_{1-(1-)p/2})$ can be estimated. These quantiles form $(1-\alpha)100\%$ bootstrap tolerance intervals with $100p\%$ coverage.

Figure 6 shows the resulting tolerance bounds on aligned functions (left) and the warping functions (right) of the simulated data. While the bounds on the warping functions are intuitive, the bounds on the aligned functions may be unexpected as they overlap near $x = 0.60$. This is due to the geometric approach in which the bounds are estimated. For the functions in Figure 6, the tolerance region indicates that the variation is not solely in amplitude (as is typically expected) but rather as more of a "twisting" variation. The aligned functions represent the overall "shape" of the original data and therefore their tolerance bounds capture a global "shape" tolerance region of the aligned function space, allowing for the possibility of the bounds crossing. Further discussion of this idea is provided in Tucker et al. (2020) and Xie et al. (2017).

Tucker et al. (2020) describe a coverage study that was performed using the new method that showed the tolerance bounds were close to their true coverage levels, albeit somewhat conservative. In particular, for a 90% coverage level with 95% confidence, the approach achieved actual confidence levels of 97.6% and 98.4% for the aligned and warping functions, respectively. The confidence coverage was greatly improved compared to the original method proposed in Section 3 as it accounted for variability in both the amplitude and phase directions. Note that while the elastic functional data approach does not currently account for external factors, the method could be extended using the approach similar to Cardot (2007), by adding factors into the elastic fPCA representation.

6 Futher Extensions

Following the development of the updated method for estimating tolerance bounds for functional data, engineers who were working on compact model calibration for electrical device design at SNL presented a new problem involving functional data. Their data were collected by measuring Zener diode devices and were captured as



(a) Aligned Functions (b) Warping Functions

Figure 6: Tolerance bounds on aligned (left) and warping (right) functions.

log current vs. voltage for approximately 44 nominally identical diodes, as shown in Figure 7. Differences in the log current-voltage behavior across diodes were due to manufacturing process variation and it was of interest to bound a percentage of the population of log current-voltage functions with a given level of confidence. These bounds would then be used in the calibration process of their compact models, i.e., models for individual devices within an electrical circuit design.

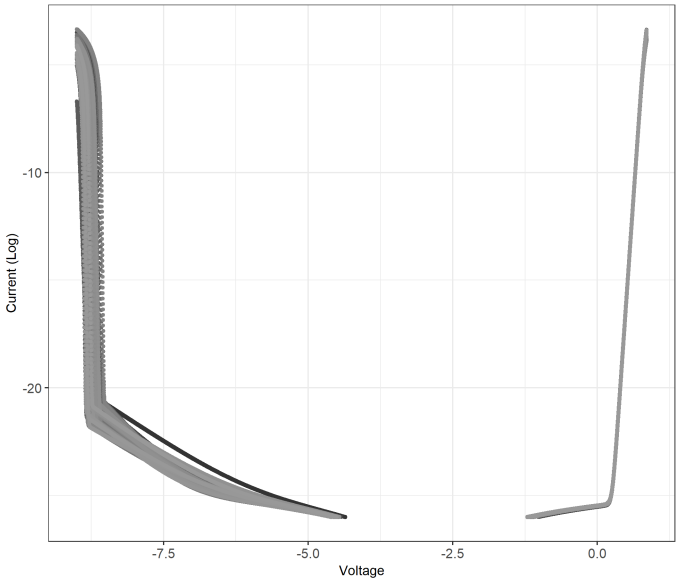


Figure 7: Zener diode electrical data from 44 devices.

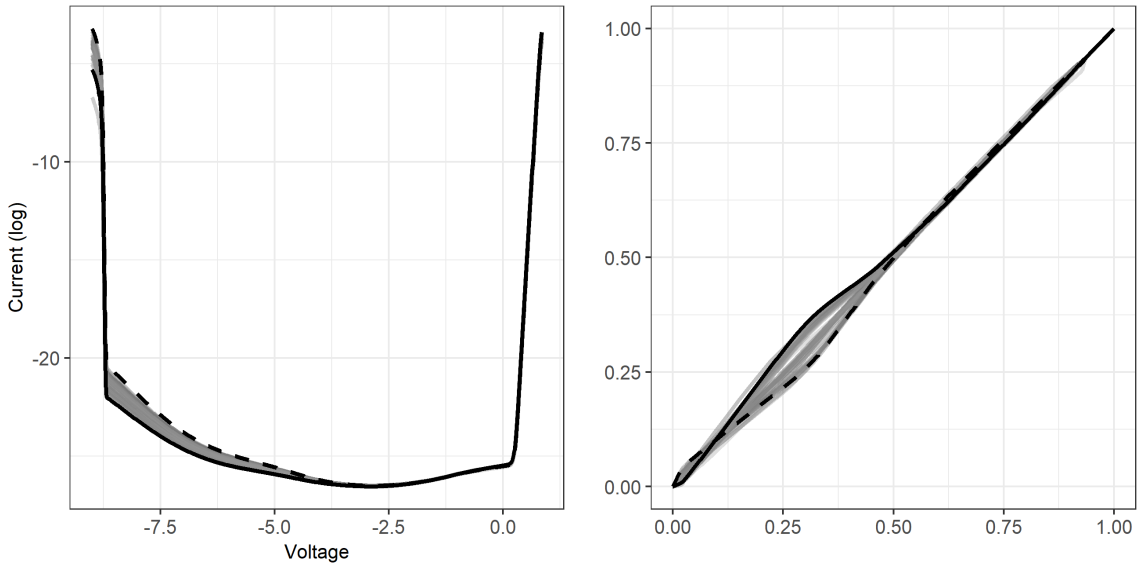
As these bounds were to be used in compact model calibration, they needed to be defined in the original data space (as opposed to the warping and aligned spaces separately) while still accounting for both types of variability. It was determined that the elastic FDA method could be extended to produce bounds in the original data space by composing 4 pairwise combinations of the bounds, as follows:

1. Lower amplitude bound with the inverse of the lower phase bound: $f^*_{\alpha/2} \circ \gamma^{-1}_{\alpha/2}$.
2. Lower amplitude bound with the inverse of the upper phase bound: $f^*_{\alpha/2} \circ \gamma^{-1}_{1-\alpha/2}$.
3. Upper amplitude bound with the inverse of the lower phase bound: $f^*_{1-\alpha/2} \circ \gamma^{-1}_{\alpha/2}$.
4. Upper amplitude bound with the inverse of the upper phase bound: $f^*_{1-\alpha/2} \circ \gamma^{-1}_{1-\alpha/2}$.

While estimating four tolerance bounds may not be intuitive, each bound represents a combination of phase and amplitude variability that is important to capture. Two of the bounds bound lower phase variability, while the other two bound upper phase variability. Similarly, two bound lower amplitude variability, while the other two bound upper amplitude variability. Each combination of upper and lower bound is important to consider as it corresponds to some bounding behavior that may be relevant to capture in compact modeling.

An fPCA model was fit to the Zener diode data and 500 bootstrap samples were generated to calculate tolerance bounds on the aligned and warping functions, as shown in Figures 8a and 8b. These bounds were constructed such that they represent where we would expect 90% ($p = 0.90$) of the devices to fall with 95% confidence ($\alpha = 0.05$). The bounds were then transformed back to the original data space, as shown in Figure 9. Figure 10 provides a zoomed in section of the current-voltage curve that is critical to estimate accurately for compact model calibration. The engineers were pleased with the resulting four bounds, as they represent combinations of high and low current and voltage that are important to characterize. They also recognized that if phase variation had been ignored, the bounds that would have been produced would result in a non-physical representation of Zener diode behavior (i.e., they would lead to physically-impossible parameters in the model calibration).

Additionally, a coverage study was conducted to assess whether these bounds are performing as intended. This study involved sampling a new set of 44 diode devices from the original data with replacement (i.e., the same device can be sampled more than once). The warping functions and aligned SRSFs were then calculated for each device. For amplitude, the 90th quantile of the SRSFs was calculated and compared to the tolerance bound to see if the entire SRSF fell within that bound. This was repeated 500 times to estimate the coverage level of this bound. This same process was performed for the phase tolerance bounds using the warping functions. The actual



(a) Aligned Functions

(b) Warping Functions.

Figure 8: Tolerance bounds on aligned (left) and warping (right) Zener diode functions.

amplitude and phase confidence levels were estimated to be 97.7% and 99.3%, respectively. While somewhat conservative, these values are relatively close to the expected value of 95% confidence, and they are consistent with the results seen in Tucker et al. (2020) and an improvement over the initial approach discussed in this paper.

Further details on the application of this method to Zener diode data are described in Reza et al.. This work illustrated a successful application and extension of the new tolerance bound approach. Future work for the engineers on this project involves incorporating this approach into their semi-automated model calibration workflow.

7 Lasting Impacts

While the ASD application did not have an ideal resolution, it prompted the development of a new method for estimating tolerance bounds for functional data that has continued to be used at SNL. The new methodology was applied to weld residual stress data for nuclear power plant systems, as well as Zener diode data for compact model calibration, and it continues to be used across different application areas at the laboratory. There has been a sharp increase in awareness of the importance of FDA in the Statistical Sciences department at SNL which has led to additional research to extend the methodology. Additionally, as more engineers are becoming familiar with the concept of FDA, they are recognizing the added benefit that can come from analyzing functions in their entirety.

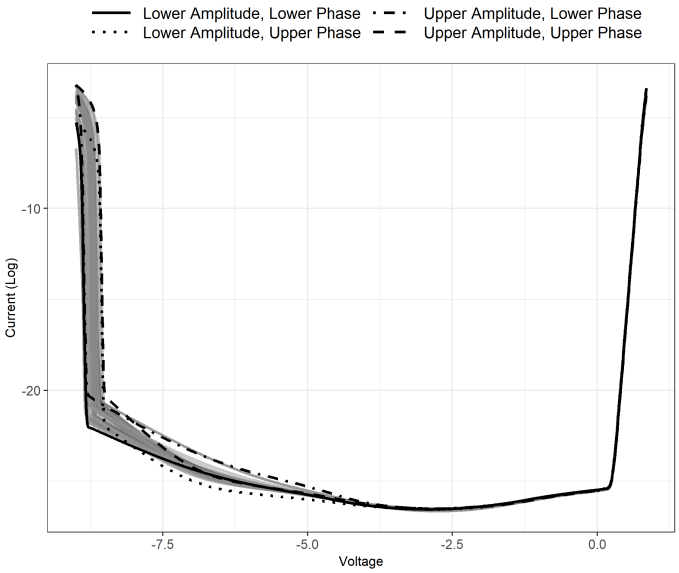


Figure 9: Tolerance bounds on original Zener diode data.

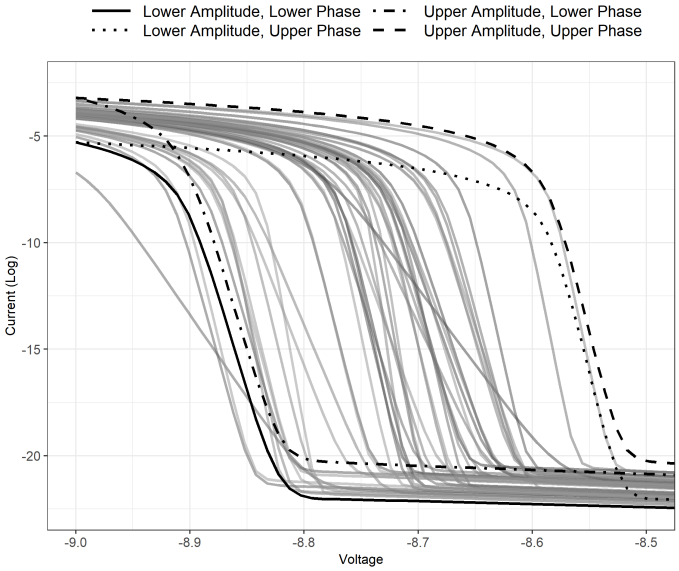


Figure 10: Zoomed-in tolerance bounds on the reverse breakdown region of the log current-voltage curve.

8 Acknowledgements

The authors would like to thank Thomas Buchheit, Shahed Reza, Biliana Paskaleva and Andrew Sandoval for their contribution to the Zener diode work.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DENA0003525.

This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

References

Cardot, H. (2007). Conditional functional principal components analysis. *Scandinavian Journal of Statistics 34*(2), 317–335.

Ioannidis, J. (2005). Why most published research findings are false. *PLOS Med 2*, e124.

Jena, A. B. (2017). 'Null' research findings aren't empty of meaning. Let's publish them. https://www.statnews.com/2017/11/10/null-research-findings/. Accessed: 2019-12-10.

Lalanne, C. (2014). *Random Vibration*. John Wiley & Sons, Ltd.

Lee, S. and S. Jung (2017). Combined analysis of amplitude and phase variations in functional data. *arXiv:1603.01775 [stat.ME]*, 1–21.

Marron, J., J. Ramsay, L. Sangalli, and A. Srivastava (2015). Functional data analysis of amplitude and phase variation. *Statistical Science 30*(4), 468–484.

Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis* (2nd ed.). New York: Springer.

Rathnayake, L. N. and P. K. Choudhary (2016). Tolerance bands for functional data. *Biometrics 72*(2), 503–512.

Reza, S., N. Martin, T. Buchheit, and J. D. Tucker (2020). Tolerance bound calculation for compact model calibration using functional data analysis. *Proc of IEEE Electron Devices Technology and Manufacturing Conference*.

Srivastava, A., E. Klassen, S. Joshi, and I. Jermyn (2011). Shape analysis of elastic curves in Euclidean spaces. *IEEE Trans. Pattern Analysis and Machine Intelligence 33*(7), 1415–1428.

Srivastava, A. and E. P. Klassen (2016). *Functional and Shape Data Analysis*. Springer-Verlag.

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

Sun, Y. and M. G. Genton (2011). Functional boxplots. *Journal of Computational and Graphical Statistics 20*(2), 316–334.

Thomas, E., C. King, J. Cap, and A. Montoya (2016). Construction of tolerance bounds for a multivariate response associated with a covariate: A case study. Paper presented at the Joint Statistical Meetings, Chicago.

Tucker, J. D., J. R. Lewis, C. King, and S. Kurtek (2020). A geometric approach for computing tolerance bounds for elastic functional data. *Journal of Applied Statistics 47*(3), 481–505.

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

Xie, W., S. Kurtek, K. Bharath, and Y. Sun (2017). A geometric approach to visualization of variability in functional data. *Journal of American Statistical Association 112*(519), 979–993.