

## ORIGINAL ARTICLE

# A general framework for multivariate functional principal component analysis of amplitude and phase variation

Clara Happ<sup>1</sup>  | Fabian Scheipl<sup>1</sup> | Alice-Agnes Gabriel<sup>2</sup> | Sonja Greven<sup>1</sup>

<sup>1</sup>Department of Statistics, LMU Munich, Munich, Germany

<sup>2</sup>Department of Geophysics, LMU Munich, Munich, Germany

## Correspondence

Clara Happ, Department of Statistics, LMU Munich, D-80539 Munich, Germany.  
Email: clara.happ@stat.uni-muenchen.de

## Present Address

Clara Happ, Ludwigstr. 33, D-80539 Munich, Germany.

## Funding information

German Research Foundation (DFG), Grant/Award Number: KA 2281/4-1, GA 2465/2-1 and GA 2465/3-1; BaCaTec, Grant/Award Number: A4; KONWIHR - the Bavarian Competence Network for Technical and Scientific High Performance Computing (project NewWave), Volkswagen Foundation (ASCETE), Grant/Award Number: 88479; KAUST-CRG (GAST), Grant/Award Number: ORS-2016-CRG5-3027; FRAGEN, Grant/Award Number: ORS-2017-CRG6 3389.02; European Union's Horizon 2020 research and innovation program (ExaHyPE), Grant/Award Number: 671698; ChEESE, Grant/Award Number: 823844; Institute of Geophysics of LMU Munich (Oeser et al., 2006), Leibniz Supercomputing Centre (LRZ), Grant/Award Number: h019z, pr63qo and pr45fi; SuperMUC, German Federal Ministry of Education and Research (BMBF), Grant/Award Number: 01IS18036A

Functional data typically contain amplitude and phase variation. In many data situations, phase variation is treated as a nuisance effect and is removed during preprocessing, although it may contain valuable information. In this note, we focus on joint principal component analysis (PCA) of amplitude and phase variation. As the space of warping functions has a complex geometric structure, one key element of the analysis is transforming the warping functions to  $L^2(\mathcal{T})$ . We present different transformation approaches and show how they fit into a general class of transformations. This allows us to compare their strengths and limitations. In the context of PCA, our results offer arguments in favour of the centred log-ratio transformation. We further embed two existing approaches from the literature for joint PCA of amplitude and phase variation into the framework of multivariate functional PCA, where we study the properties of the estimators based on an appropriate metric. The approach is illustrated through an application from seismology.

## KEYWORDS

Bayes Hilbert space, Fréchet variance, functional data analysis, registration, seismology, transformation of warping functions

## 1 | INTRODUCTION

Functional data analysis (Ramsay & Silverman, 2005) is concerned with the broad field of data that comes in the form of functions  $x_1, \dots, x_N$ . For simplicity, consider here  $x_i \in L^2(\mathcal{T})$ ,  $i = 1, \dots, N$ , with an interval  $\mathcal{T} = [a, b] \subset \mathbb{R}$  of length  $\eta = b - a > 0$ . Warping approaches (see, e.g., Marron, Ramsay, Sangalli, & Srivastava, 2015, and references therein) aim at decomposing the observed functions  $x_i$  into warping functions  $\gamma_i$  that account for the phase variation in  $x_i$  and registered functions  $w_i$  that account for the amplitude variation in the data, via concatenation:  $x_i(t) = (w_i \circ \gamma_i)(t) = w_i(\gamma_i(t))$ . As a result, the main features of  $w_i$  such as maximum and minimum points will typically be aligned, which reduces the variation and makes it easier to find common structures. The alignment is accomplished by the warping functions  $\gamma_i$  that map the individual time of each observed curve  $x_i$  to the global, absolute time of the registered curves. Throughout this paper, we will consider the warping, that is, the decomposition of  $x_i$  into  $w_i$  and  $\gamma_i$ , as given. In practical analyses, one would therefore have to choose an appropriate warping algorithm for the data before using our results. For a discussion on possible warping approaches, see, for example, Marron et al. (2015).

In many data situations, the main interest lies in the amplitude variation, meaning that warping is considered as part of the preprocessing and the subsequent analysis is based on the registered functions  $w_i$  only. In others, phase variation is an integral part of the data that carries important information (Kneip & Ramsay, 2008; Sangalli, Secchi, Vantini, & Vitelli, 2010). It is thus worth to be incorporated into the analysis to gain deeper insights into the mechanisms generating the data. As an example, in our seismological application, the seismic waves arrive not only with different intensity but also with different time delays depending on their type and the geographical relation between the hypocentre and each seismometer. Moreover, in a recent result, it was shown that warping may be affected by severe nonidentifiability in the sense that there may exist different representations  $x_i = w_i \circ \gamma_i = \tilde{w}_i \circ \tilde{\gamma}_i$  with  $w_i \neq \tilde{w}_i$  and  $\gamma_i \neq \tilde{\gamma}_i$  (Chakraborty and Panaretos, 2017; to see this, set, e.g.,  $\tilde{w}_i = x_i$  and  $\tilde{\gamma}_i = id$ , i.e., no warping). In order to keep the original information in  $x_i$ , it seems crucial to analyse both the warping and registered functions.

Two methods that account for both amplitude and phase variation have been proposed for principal component analysis (PCA), which is often used in functional data analysis for data exploration and dimension reduction and as a building block for methods such as regression: Hadjipantelis, Aston, Müller, and Evans (2015) transform the warping functions to the space of square-integrable functions,  $L^2(\mathcal{T})$ , by differentiating them and taking the log. Next, they calculate a separate functional PCA for the transformed warping functions and the registered functions. Dependencies between the warping functions and the registered functions are then modelled via a linear mixed-effects model of the principal component scores. Lee and Jung (2016) also proposed to transform the warping functions to  $L^2(\mathcal{T})$  but then studied a combined PCA analysis by “glueing” the discretized functions together. They describe an optimal weighting scheme for the transformed warping functions based on dimension reduction and optimal reconstruction. The main reason for transforming the warping functions in both approaches is that the space of warping functions on  $\mathcal{T}$ ,  $\Gamma(\mathcal{T})$ , has a complex geometric structure, as it is, for example, not closed under addition or scalar multiplication and is not equipped with a natural scalar product, which is essential to define principal components. On the other hand, PCA in  $L^2(\mathcal{T})$  is well studied and can easily be applied to the transformed warping functions.

In this note, we discuss the role of the transformation of the warping functions to  $L^2(\mathcal{T})$ . We define a general class of transformations that includes existing approaches from the literature as well as alternative transformations based on density functions on  $\mathcal{T}$  and allows to compare their strengths and limitations. We offer theoretical arguments in favour of the centred log-ratio (clr) transformation, which relates to the Bayes space of densities (see, e.g., Egozcue, Díaz-Barrero, & Pawłowsky-Glahn, 2006; Hron, Menafoglio, Templ, Hrušová, & Filzmoser, 2016) and illustrate this by means of a simulated toy example. Second, we embed the methods of Hadjipantelis et al. (2015) and Lee and Jung (2016) for the joint analysis of amplitude and phase variation into a multivariate functional principal component framework, which is based on univariate functional PCA and takes correlations of warping functions and registered functions into account. For general transformations of the warping functions to  $L^2(\mathcal{T})$ , we study properties of resulting joint principal components using an appropriate metric. The proposed method based on the clr transformation is applied to data from a seismological computer experiment yielding spatially referenced high-resolution time series of ground velocity measurements. In this context, the joint analysis of amplitude and phase variation is of particular interest as both contain relevant information on the propagation of seismic waves. The new method is shown to give promising results with a meaningful interpretation.

The manuscript proceeds as follows: In Section 2, we review several transformations from  $\Gamma(\mathcal{T})$  or the space of probability density functions to  $L^2(\mathcal{T})$  proposed in the literature and give reasons why the clr approach should be preferred to other transformations when defining PCA for warping functions. Section 3 embeds the existing methods for joint analysis of amplitude and phase variation into the framework of multivariate functional principal component analysis (MFPCA; Happ and Greven, 2018) and gives new insights into the properties of the joint principal components. The theoretical results are illustrated in Section 4 by means of data from a from a multiphysics computational seismology experiment.

## 2 | TRANSFORMATION APPROACHES FOR WARPING FUNCTIONS

The space of warping functions

$$\Gamma(\mathcal{T}) = \{\gamma : \mathcal{T} \rightarrow \mathcal{T} : \gamma \text{ is a diffeomorphism}^1, \gamma(a) = a, \gamma(b) = b\}$$

has a complex, non-Euclidean geometric structure (Lee & Jung, 2016; Srivastava & Klassen, 2016), as discussed before. In order to simplify the geometry of  $\Gamma(\mathcal{T})$  and to derive principal components, it is natural to consider mappings  $\Psi : \Gamma(\mathcal{T}) \rightarrow L^2(\mathcal{T})$ , which allow to transform warping functions to square-integrable functions. In a second step, one can calculate principal components in the well-studied space  $L^2(\mathcal{T})$  and transform the results back by the inverse map  $\Psi^{-1} : L^2(\mathcal{T}) \rightarrow \Gamma(\mathcal{T})$ . In this section, we present different approaches for this transformation and discuss their strengths and limitations within a new, general framework. Note that, here, we focus on PCA for the warping functions only and will return to the joint approach in the next section.

### 2.1 | Square-root velocity transformation

The square-root velocity framework (see, e.g., Srivastava & Klassen, 2016; Srivastava, Jermyn, & Joshi, 2007) has been shown to be advantageous in statistical shape analysis, particularly when combined with the Fisher–Rao metric, which is invariant under warping (Srivastava & Klassen, 2016).

<sup>1</sup>A *diffeomorphism* describes a smooth and strictly increasing function (Marron et al., 2015). In other words, the warping function  $\gamma$  is required to be invertible and should not allow for abrupt jumps or for “travelling back in time.”

It is also used as a building block in Lee and Jung (2016) for transforming the warping functions to  $L^2(\mathcal{T})$ . The mapping  $\Psi$  here consists of two steps: First, the warping functions are mapped to the positive orthant of the (scaled) unit sphere in  $L^2(\mathcal{T})$ ,  $S_+^\infty(\mathcal{T}) = \{s \in L^2(\mathcal{T}) : \|s\|_2^2 = \eta, s \geq 0\}$ , where  $\eta$  denotes again the length of the interval  $\mathcal{T}$ . This is realized by the square-root velocity function SRVF :  $\gamma \mapsto \sqrt{\gamma'}$ , where  $\gamma'$  denotes the first derivative of  $\gamma$ . It can be shown that this is a bijection from  $\Gamma(\mathcal{T})$  to  $S_+^\infty(\mathcal{T})$ . In order to map the transformed warping functions  $q = \sqrt{\gamma'}$  to  $L^2(\mathcal{T})$ , choose some  $\mu \in S_+^\infty(\mathcal{T})$  and approximate  $q$  by functions in the tangent space associated with  $\mu$ ,  $T_\mu(\mathcal{T}) = \{v \in L^2(\mathcal{T}) : \langle v, \mu \rangle_2 = 0\}$ . This second transformation step is done via the mapping

$$\tilde{\psi}_{S,\mu} : S_+^\infty(\mathcal{T}) \rightarrow T_\mu(\mathcal{T}), \quad q \mapsto \frac{\theta}{\eta^{1/2} \sin(\theta)}(q - \cos(\theta)\mu), \quad \theta = \cos^{-1}\left(\frac{\langle q, \mu \rangle_2}{\eta}\right). \quad (1)$$

A natural choice for  $\mu$  is the Karcher or Fréchet mean of  $q$  (Tucker, 2014). For the special case of warping functions, one may also choose the constant function  $q_0(t) \equiv 1$ , which is associated with the identity warping  $\gamma_0(t) = t$  (Lee & Jung, 2016). The back transformation to  $\Gamma(\mathcal{T})$  is again in two steps: First, apply

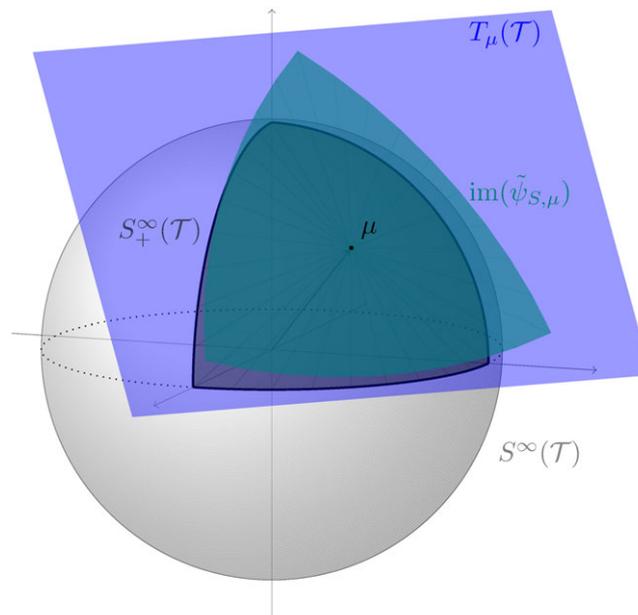
$$\tilde{\psi}_{S,\mu}^{-1} : T_\mu(\mathcal{T}) \rightarrow S^\infty(\mathcal{T}), \quad v \mapsto \cos(\|v\|_2)\mu + \eta^{1/2} \sin(\|v\|_2) \frac{v}{\|v\|_2} \quad (2)$$

to map  $v$  to the sphere  $S^\infty(\mathcal{T}) = \{s \in L^2(\mathcal{T}) : \|s\|_2^2 = \eta\}$ . Then, the results are mapped back to the space of warping functions via  $\text{SRVF}^{-1} : S^\infty(\mathcal{T}) \rightarrow \Gamma(\mathcal{T})$  with  $\text{SRVF}^{-1}(s)(t) = a + \int_0^t s(u)^2 du$ . The overall mapping from  $\Gamma(\mathcal{T})$  to  $L^2(\mathcal{T})$  is hence given by  $\Psi = \tilde{\psi}_{S,\mu} \circ \text{SRVF}$  with the inverse  $\Psi^{-1} = \text{SRVF}^{-1} \circ \tilde{\psi}_{S,\mu}^{-1}$ .

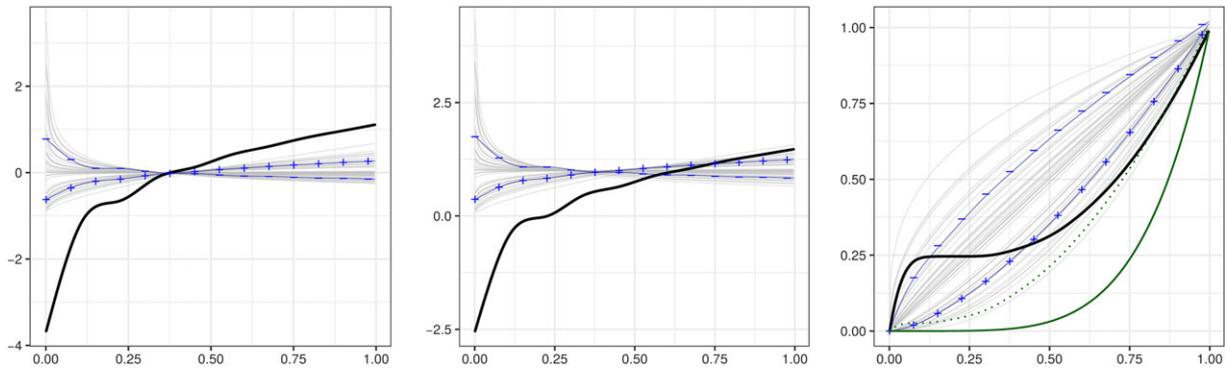
In this framework, an approach for warping-invariant functional PCA has been proposed in Tucker (2014). For the calculation of principal components for warping functions as described above, however, the square-root velocity transformation has two serious shortcomings. First,  $\tilde{\psi}_{S,\mu}$  and  $\tilde{\psi}_{S,\mu}^{-1}$  are undefined for the rather interesting points  $\mu \in S_+^\infty(\mathcal{T})$ , as then  $\theta = \sin(\theta) = 0$ , and  $v_0 \equiv 0 \in T_\mu(\mathcal{T})$ , which implies  $\|v_0\|_2 = 0$ . Although, theoretically, they can easily be completed by applying L'Hôpital's rule, computational instabilities for functions close to  $\mu$  and  $v_0$  often occur in practice. At the same time, the projection from the positive orthant of the sphere  $S_+^\infty(\mathcal{T})$  to the tangent space  $T_\mu(\mathcal{T})$  is a local approximation that works best in the vicinity of  $\mu$ . On the one hand, this shows that the choice of  $\mu$  is important and, on the other hand, requires the data to be neither too close nor too far from  $\mu$ . Second, and even more importantly, the mappings  $\tilde{\psi}_{S,\mu}$  and  $\tilde{\psi}_{S,\mu}^{-1}$  are not inverse to each other as

$$\text{dom}(\tilde{\psi}_{S,\mu}) = S_+^\infty(\mathcal{T}) \subsetneq S^\infty(\mathcal{T}) = \text{im}(\tilde{\psi}_{S,\mu}^{-1}) \quad \text{and} \quad \text{im}(\tilde{\psi}_{S,\mu}) \subsetneq T_\mu(\mathcal{T}) = \text{dom}(\tilde{\psi}_{S,\mu}^{-1}). \quad (3)$$

Although a theoretical proof is given in the Appendix, the main problem can be seen at a glance in Figure 1: The domain  $S_+^\infty(\mathcal{T})$  and the image  $\text{im}(\tilde{\psi}_{S,\mu})$  of the projection  $\tilde{\psi}_{S,\mu}$  are proper subsets and therefore "smaller" than the full sphere  $S^\infty(\mathcal{T})$  and the tangent space  $T_\mu(\mathcal{T})$ , which form the image and domain of the back transformation  $\tilde{\psi}_{S,\mu}^{-1}$ . For all SRVFs  $q \in S_+^\infty(\mathcal{T})$ , which have been projected to  $T_\mu(\mathcal{T})$  by  $\tilde{\psi}_{S,\mu}$ , there is a unique mapping back to  $S_+^\infty(\mathcal{T})$ , because the restriction  $\tilde{\psi}_{S,\mu}^{-1}|_{\text{im}(\tilde{\psi}_{S,\mu})}$  is a bijection (Srivastava & Klassen, 2016). As soon as one leaves the subset  $\text{im}(\tilde{\psi}_{S,\mu})$ , however, there is no guarantee that the image of an arbitrary function  $v \in T_\mu$  under  $\tilde{\psi}_{S,\mu}^{-1}$  is again in the positive orthant of  $S^\infty(\mathcal{T})$  and thus represents a valid SRVF. As  $\text{SRVF}^{-1}$  is not injective, the final transformation back to  $\Gamma(\mathcal{T})$  will always yield a valid warping function, but potentially with atypical structure. In order to see that leaving  $\text{im}(\tilde{\psi}_{S,\mu})$  is likely to happen in practice, consider the following toy example.



**FIGURE 1** Illustration of the sphere  $S^\infty(\mathcal{T})$  and its positive orthant  $S_+^\infty(\mathcal{T})$  (dark grey) and the projection of this orthant to the tangent space  $T_\mu(\mathcal{T})$  via  $\tilde{\psi}_{S,\mu}$  (green)



**FIGURE 2** Principal component analysis of warping functions based on  $\tilde{\psi}_{S,\mu}$  for simulated warping functions and different stages of transformation. The grey functions represent the (transformed) data, and the black functions the first principal component  $\hat{\phi}_1$ . Blue lines visualize the effect of the first principal component as perturbation from the mean ( $\hat{\mu} \pm \hat{\lambda}_1^{1/2} \hat{\phi}_1$ ). Left: Tangent space  $T_\mu(\mathcal{T})$ . Centre: Sphere  $S^\infty(\mathcal{T})$ . Right: Space of warping functions  $\Gamma(\mathcal{T})$ . The green curves correspond to a new observation (solid) and its reconstruction (dotted) based on the first principal component

**Example 1.** Let  $\mathcal{T} = [0, 1]$ . Then  $\gamma(t) = t^k$  clearly is a valid warping function for all  $k > 0$ . The associated square-root velocity function is given by  $q(t) = \sqrt{\gamma'(t)} = \sqrt{k t^{k-1}}$ . Choosing the SRVF of the identity warping function  $\gamma_0$  for  $\mu$  being chosen, the approximation of  $q$  in the tangent space  $T_\mu$  is given by  $v = \tilde{\psi}_{S,\mu}(q)$  using  $\theta = \cos^{-1}(\langle q, \mu \rangle_2) = \cos^{-1}(\frac{2\sqrt{k}}{k+1})$ . Figure 2 illustrates  $N = 50$  warping functions  $\gamma_i(t) = t^{k_i}$ ,  $i = 1, \dots, N$ , together with their SRVFs  $q_i$  and the tangent-space approximations  $v_i$  (grey curves). The values  $k_i$  are generated according to a  $\text{Ga}(5, 5)$  distribution, such that  $\mathbb{E}(k_i) = 1$ , which corresponds to the identity warping, and  $\text{Var}(k_i) = 0.2$  for intermediate variation. The first eigenfunction  $\hat{\phi}_1$  of the tangent-space approximations in  $L^2(\mathcal{T})$  explains more than 95% of the variability in  $v_i$  and is shown in black in Figure 2. Note that  $\hat{\phi}_1$  takes values below  $-1$  in the left part of  $\mathcal{T}$ , meaning that  $\hat{\phi}_1$  is outside  $\text{im}(\tilde{\psi}_{S,\mu})$ . As the eigenvalue  $\hat{\lambda}_1 = 0.036$  is rather small, visualizing the effect of the first principal component as scaled perturbation from the mean ( $\hat{\mu} \pm \hat{\lambda}_1^{1/2} \hat{\phi}_1$ , blue curves in Figure 2) yields valid SRVFs and sensible warping functions after the back transformation. However, transforming the original principal component  $\hat{\phi}_1$  back via  $\tilde{\psi}_{S,\mu}^{-1}$  yields a function  $q_\phi \in S^\infty(\mathcal{T})$ , but outside  $S_+^\infty(\mathcal{T})$ , as it has negative values and thus does not represent typical structures of the  $q_i$ . The back transformation to  $\Gamma(\mathcal{T})$  yields a valid warping function  $\gamma_\phi$ , but again with a very different form than the original data. The green curve in the right of Figure 2 has a more pronounced curvature than the curves in the original sample. Reconstructing the green curve based on the first principal component yields a function that is shifted towards the identity warping function and does not represent the original green curve well.

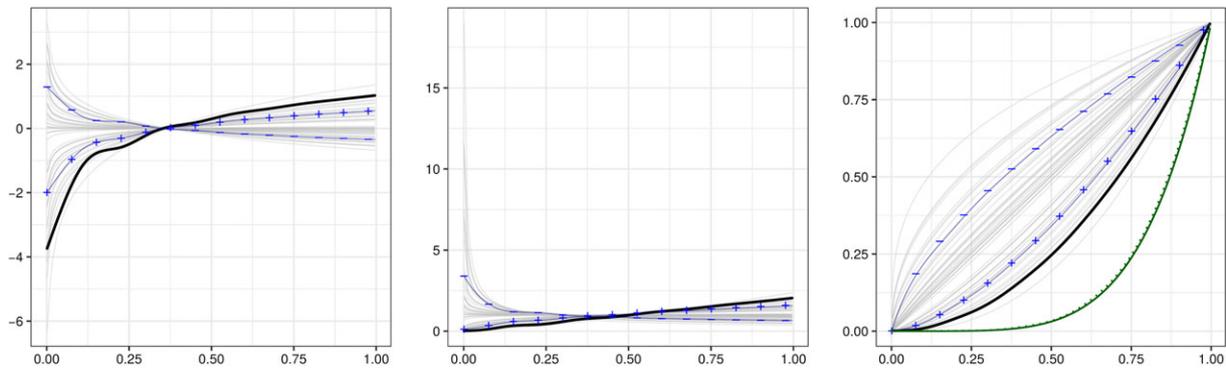
This simple example shows that whenever one uses the projected SRVFs  $v_i$  for statistical analyses in  $T_\mu(\mathcal{T})$  whose results are not guaranteed to stay within  $\text{im}(\tilde{\psi}_{S,\mu})$ , there is a risk of obtaining atypical results on the level of the SRVFs and of the warping functions. As  $\text{im}(\tilde{\psi}_{S,\mu})$  is not closed under vector space operations in  $T_\mu(\mathcal{T})$ , this risk occurs, for example, when calculating principal components of  $v_i$ , representing  $v_i$  in a truncated basis expansion or when using  $v_i$  as covariates in a regression. In all these cases, our results show that it is very likely that transforming the results of the analysis from  $T_\mu(\mathcal{T})$  to  $\Gamma(\mathcal{T})$  yields warping functions with atypical structure, which may be hard to interpret. As  $T_\mu(\mathcal{T})$  is a local approximation, anomalies may even occur if the results are within  $\text{im}(\tilde{\psi}_{S,\mu})$ , but close to the boundary. An example for this can be seen in the prediction of the green curve in Figure 2.

## 2.2 | Bayes space transformation

Due to their typical characteristics, warping functions  $\gamma$  can also be interpreted as generalized cumulative distribution functions of continuous random variables  $X : \Omega \rightarrow \mathcal{T}$  in the sense that  $\gamma(a) = a, \gamma(b) = b$  and  $\gamma$  is monotonically increasing. Taking the first derivative  $\gamma'$  yields a unique (scaled) probability density function on  $\mathcal{T}$ . Although this has been mentioned in Hadjipantelis et al. (2015), the authors does not exploit the geometric structure of the space of the transformed functions. As shown in Egozcue et al. (2006), density functions, or, more precisely, equivalence classes of these functions, form a vector space, known as Bayes Hilbert space  $B^2(\mathcal{T})$ . We propose to make use of this specific space to define a PCA for warping functions.

Following Egozcue et al. (2006), two functions  $f, g \in B^2(\mathcal{T})$  are equivalent in the Bayes Hilbert space, if they are proportional ( $f = \alpha g$ ). A natural representative for each class is given by the function integrating to  $\eta = b - a$ , which we interpret as the derivative of a warping function. For  $f, g \in B^2(\mathcal{T})$  and  $\alpha \in \mathbb{R}$ , operations on  $B^2(\mathcal{T})$  are defined as (Egozcue et al., 2006)

$$\begin{aligned} (f \oplus g)(t) &= \frac{f(t)g(t)}{\int_{\mathcal{T}} f(s)g(s) ds} & (a \circ f)(t) &= \frac{f(t)^a}{\int_{\mathcal{T}} f(s)^a ds} \\ \langle f, g \rangle_B &= \frac{1}{2\eta} \int_{\mathcal{T}} \int_{\mathcal{T}} \log\left(\frac{f(x)}{f(y)}\right) \log\left(\frac{g(x)}{g(y)}\right) dy dx. \end{aligned}$$



**FIGURE 3** Principal component analysis of warping functions based on  $\psi_B$  for the same warping functions as in Figure 2 and different stages of transformation. The grey functions represent the (transformed) data, and the black functions the first principal component  $\hat{\phi}_1$ . Blue lines visualize the effect of the first principal component as perturbation from the mean ( $\hat{\mu} \pm \hat{\lambda}_1^{1/2} \hat{\phi}_1$ ). Left:  $L^2(\mathcal{T})$ . Centre: Bayes space  $B^2(\mathcal{T})$ . Right: Space of warping functions  $\Gamma(\mathcal{T})$ . The green curves correspond to a new observation (solid) and its reconstruction (dotted) based on the first principal component

The centred log-ratio transform  $\psi_B : B^2(\mathcal{T}) \rightarrow L^2(\mathcal{T})$

$$\psi_B(f)(t) = \log(f(t)) - \frac{1}{\eta} \int_{\mathcal{T}} \log(f(x)) dx \quad \psi_B^{-1}(f)(t) = \eta \cdot \frac{\exp(f(t))}{\int_{\mathcal{T}} \exp(f(s)) ds}$$

is an isometric isomorphism with respect to the norm induced by  $\langle \cdot, \cdot \rangle_B$ . Therefore,  $\psi_B(f \oplus g) = \psi_B(f) + \psi_B(g)$ ,  $\psi_B(\alpha \odot f) = \alpha \cdot \psi_B(f)$  and  $\langle f, g \rangle_B = \langle \psi_B(f), \psi_B(g) \rangle_2$  with the standard operations  $(+, \cdot, \langle \cdot, \cdot \rangle_2)$  on  $L^2(\mathcal{T})$  (Egozcue et al., 2006). One may note that Hadjipantelis et al. (2015) used a similar log transformation for warping functions as  $\psi_B$  but without the integral term.

The geometric structure of  $B^2(\mathcal{T})$  allows to easily transfer PCA to density functions by transforming them to  $L^2(\mathcal{T})$ , calculating the PCA in this space, and transforming the result back to  $B^2(\mathcal{T})$  (Hron et al., 2016). As the mapping between a warping function  $\gamma$  and the associated density is one to one, we can directly define PCA for warping functions based on this Hilbert space transformation. The resulting principal components give more meaningful results in our toy example, as shown in Figure 3. Moreover, the reconstruction of the green curve based on the first principal component is almost perfect.

As shown in Hron et al. (2016), the transformation  $\psi_B$  does also not map into full  $L^2(\mathcal{T})$ , but into the subspace  $U_B(\mathcal{T}) = \{v \in L^2(\mathcal{T}) : \int_{\mathcal{T}} v(s) ds = 0\}$ . This space is equivalent to the space  $T_{\mu}(\mathcal{T})$  in the SRVF approach, if we choose  $\mu$  the SRVF associated with the identity warping function, as proposed in Lee and Jung (2016). An important difference between the SRVF approach and the Bayes space transformation, however, is that the centred log-ratio transformation  $\psi_B$  gives a one-to-one transformation from  $B^2(\mathcal{T})$  to  $U_B(\mathcal{T})$ , meaning that  $\text{im}(\psi_B)$  covers full  $U_B(\mathcal{T})$ , which is closed under vector space operations. Leaving  $\text{im}(\psi_B)$  in PCA or regression is thus not possible. Particularly, this means that the constraint  $\int_{\mathcal{T}} v(s) ds = 0$  is automatically fulfilled for the principal components and does not need to be explicitly imposed in the implementation. Moreover, as an isometric isomorphism,  $\psi_B$  preserves a lot more of the geometric structure of the transformed warping functions than  $\tilde{\psi}_{S,\mu}$ , which only preserves distances.

### 2.3 | Other transformations

PCA of density functions based on transformations to  $L^2(\mathcal{T})$  has been considered more generally in Petersen and Müller (2016). They present two alternative transformations that we can again extend to warping functions, by applying them to the derivatives of the latter. The log-hazard transformation is

$$\psi_H(f) = \log\left(\frac{f}{1-F}\right)$$

with  $F$  the warping function scaled to  $[0, 1]$  and  $f$  the associated density. As hazard functions are known to diverge at the right endpoint of  $\mathcal{T}$ , the transformation is considered only on a subinterval  $[a, b - \delta\eta]$  with  $\delta$  a threshold parameter. For the back transformation, uniform weight is assigned to  $t \in (b - \delta\eta, b]$ . Alternatively, Petersen and Müller (2016) considered the log-quantile density transformation

$$\psi_Q(f) = -\log(f(Q))$$

with  $Q$  the inverse (quantile) function associated with  $F$ . Of the two, Petersen and Müller (2016) recommended to use the log-quantile transformation.

Regarding the practical usability in the context of warping functions, the log-hazard transformation  $\psi_H$  may be highly influenced by the threshold parameter  $\delta$ , as shown in Figure A1 in the Appendix for the curves of our toy example. There, we choose  $\delta = 0.05$ , which corresponds to cutting the warping functions at the 95% quantile of the identity warping function and results in an abrupt change in the principal component when transformed to  $B^2(\mathcal{T})$  or  $\Gamma(\mathcal{T})$ . On the other hand, the log-quantile density transformation  $\psi_Q$  requires numerical inversion of the warping functions  $\gamma$ , which may also lead to instabilities. Regarding the reconstruction of the new green curve in Figure A1, the PCA based on log-hazard

transformation performs similarly well as the one based on the clr transformation. By contrast, for the log-quantile transformation, the prediction is rather poor, as it is shifted considerably towards the identity warping. This, however, is in contradiction with the recommendation given in Petersen and Müller (2016).

## 2.4 | General framework

All discussed transformations can be subsumed as transformations  $\Psi_* : \Gamma(\mathcal{T}) \rightarrow L^2(\mathcal{T})$  with  $\Psi_* = \psi_* \circ D$ . Here,  $D$  denotes the differential operator that maps a warping function  $\gamma$  to its density  $\gamma'$ . The restriction  $\gamma(a) = a$  and  $\gamma(b) = b$  makes this a one-to-one mapping. The mapping  $\psi_*$  from the space of density functions to  $L^2(\mathcal{T})$  depends on the specific transformation. In particular, the SRVF approach can also be considered as such a transformation with  $\psi_{S,\mu}(q) = \tilde{\psi}_{S,\mu}(\sqrt{q})$  and  $\psi_{S,\mu}^{-1}(v) = [\tilde{\psi}_{S,\mu}^{-1}(v)]^2$ . For the Bayes space approach,  $\psi_*$  corresponds to the centred log-ratio transform  $\psi_B$ ; and for the log-hazard transformation and the log-quantile density transformation proposed in Petersen and Müller (2016),  $\psi_*$  corresponds to  $\psi_H$  and  $\psi_Q$ , respectively. In order to compare the different transformations, one can therefore focus on  $\psi_*$  only.

As discussed before, the SRVF transformation via  $\tilde{\psi}_{S,\mu}$  does not necessarily preserve structure in the transformed warping functions, as the back transformation via  $\tilde{\psi}_{S,\mu}^{-1}$  may result in functions outside  $S_+^\infty(\mathcal{T})$ , having negative values. The alternative formulation  $\Psi_S = \psi_{S,\mu} \circ D$  respects this structure better, as  $\psi_{S,\mu}^{-1}$  always yields valid density functions. However, this reformulation does not change the fact that  $\psi_{S,\mu}$  is not surjective and therefore PCA based on the SRVF transformation does not necessarily return structures that resemble those of the original data. Moreover, the transformation may not be computationally robust for functions close to  $\mu$  (in  $S_+^\infty(\mathcal{T})$ ) or  $v_0$  (in  $L^2(\mathcal{T})$ ), which is likely to happen if  $\mu$  represents some kind of mean function. The centred log-ratio transform  $\psi_B$  is an isometric isomorphism between  $B^2(\mathcal{T})$  and  $L^2(\mathcal{T})$ . It hence preserves the Hilbert space structure and seems particularly suitable for PCA. Due to the log transformation, numerical instabilities may occur in regions in which  $\gamma'(t)$  is close to zero, where the warping function  $\gamma$  is near the degenerate constant case, where warping is not uniquely defined. The alternative transformations proposed in Petersen and Müller (2016) do not represent isometric isomorphisms but preserve the space structures by construction. However, they may be affected by numerical instabilities, as shown in the toy example (Figure A1). Overall, the clr transformation seems to be the most suitable transformation for PCA.

## 3 | MODES OF JOINT VARIATION IN AMPLITUDE AND PHASE

In order to study the joint variation of amplitude and phase in the data, Lee and Jung (2016) proposed to concatenate the registered functions  $w_i$  and the transformed warping functions  $v_i = \Psi(\gamma_i)$ , including a weighting parameter  $C$ :

$$g_i^C(t) = \begin{cases} w_i(t) & t \in [a, b) \\ Cv_i(t - (b - a)) & t \in [b, 2b - a) \end{cases}.$$

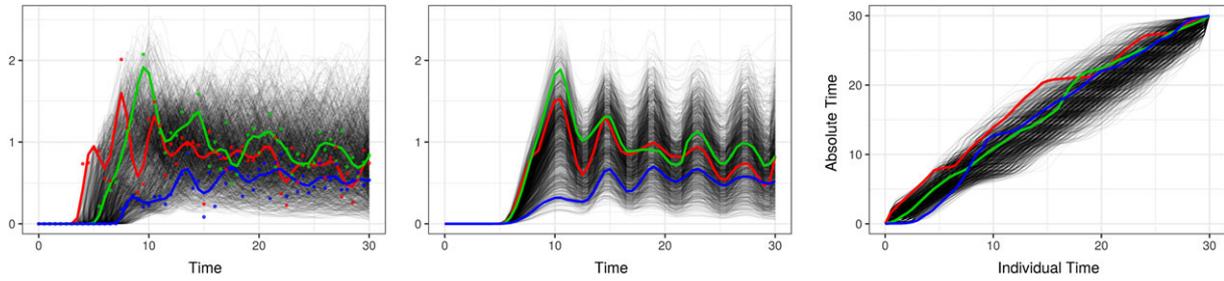
Remember that the warping is assumed to be given; that is, in practical applications,  $w_i$  and  $\gamma_i$  have to be replaced by their estimates obtained from the data. Conditioning on the warping results as in Lee and Jung (2016) allows to jointly analyse warping functions and registered functions. Alternative approaches as in Kneip and Ramsay (2008) have a slightly different focus, as they aim at finding optimal warping functions for parsimonious functional PCA of the registered functions.

Lee and Jung (2016) used  $\Psi = \Psi_S$  to transform the warping functions, but the method can be directly transferred to general transformations  $\Psi : \Gamma(\mathcal{T}) \rightarrow L^2(\mathcal{T})$ . The PCA is based on these concatenated functions,  $g_i^C$ , evaluated on a fine grid. The resulting PCs are then separated and transformed back for the final interpretation.

As noted in Lee and Jung (2016), one may alternatively use methods for MFPCA (Chiou, Yang, & Chen, 2014; Happ and Greven, 2018). These approaches are indeed more appropriate as they better reflect the characteristic nature of the data in terms of bivariate functions  $s_i = (w_i, v_i) \in L^2(\mathcal{T}) \times L^2(\mathcal{T}) =: \mathcal{H}$ , and therefore, we will use them in the following. Moreover, the bivariate formulation ensures that we do not need to deal with potential discontinuities of  $g_i^C$  at  $b$ . The approach in Happ and Greven (2018) estimates the multivariate functional PCA via a univariate functional PCA followed by a PCA of the combined score vectors based on a theoretical equivalence result. In this sense, it is similar to the approach in Hadjipantelis et al. (2015), who also calculated a separate functional PCA for  $w_i$  and  $v_i$  but then used a linear mixed model to capture joint variation in the scores. The two-step approach in Happ and Greven (2018) can also easily be applied to warped curves in  $\mathbb{R}^2$  or  $\mathbb{R}^3$  or warped images. In the case of two-dimensional curves, one would, for example, have a registered function for the  $x$ - and  $y$ -coordinates and a warping function. One would then apply the method to the trivariate vector of all three functions.

The weighting factor  $C$  can directly be included into a weighted scalar product  $\langle\langle z_i, z_j \rangle\rangle_w = \langle w_i, w_j \rangle_2 + C^2 \langle v_i, v_j \rangle_2$ . This corresponds to the standard scalar product on  $\mathcal{H}$  for the bivariate function  $\bar{z}_i = (w_i, Cv_i)$ , which is mimicked by  $g_i^C$ . In analogy to Lee and Jung (2016), an optimal choice for  $C$  can be found by optimizing the reconstruction based on the first  $M$  principal components. More specifically, let the truncated and estimated Karhunen–Loève representation of the bivariate function  $z_i$  be

$$\hat{z}_i^{[M]} = \hat{\mu}_z + \sum_{m=1}^M \hat{\rho}_{im} \hat{\phi}_m$$



**FIGURE 4** Earthquake dataset representing 1,558 observations of simulated ground velocity over time (log transformed). Left: Smoothed curves. Middle: Registered functions. Right: Warping functions. The red (green/blue) curves have minimum (median/maximum) hypocentral distance in the dataset

with  $\hat{\mu}_z = (\bar{w}, \bar{v})$  the bivariate mean function,  $\hat{\phi}_m = (\hat{\phi}_m^{(w)}, \hat{\phi}_m^{(v)})$  the  $m$ th bivariate principal component, and  $\hat{\rho}_{im} = \langle \langle z_i - \hat{\mu}_z, \hat{\phi}_m \rangle \rangle_w$  the observation-specific principal component scores (see Happ and Greven, 2018, for further details). By denoting the elements of  $\hat{z}_i^{[M]}$  by  $\hat{w}_i^{[M]}$  and  $\hat{v}_i^{[M]}$ , respectively, a reconstruction of  $x_i = w_i \circ \gamma_i$  is given by  $\hat{x}_i^{[M]} = \hat{w}_i^{[M]} \circ \Psi^{-1}(\hat{v}_i^{[M]})$ . For a given  $M$ , an optimal value of  $C$  can be found as

$$\hat{C} = \arg \min_{C>0} \frac{1}{N} \sum_{i=1}^N \|x_i - \hat{x}_i^{[M]}\|_2^2.$$

This choice makes sure that the PCA also yields optimal reconstructions with respect to the originally observed data.

The eigenvalue  $v_m$  associated with the principal component  $\phi_m$  can be interpreted as the amount of variability in  $z_i$  that is explained by this principal component. Typically, an optimal choice for  $M$  is obtained by

$$\hat{M} = \min \left\{ M \in \mathbb{N} : \frac{\sum_{m=1}^M \hat{v}_m}{\sum_{m=1}^{\infty} \hat{v}_m} > \tau \right\}, \quad (4)$$

where  $\tau$  is a threshold for the minimum proportion of variability in  $z_1, \dots, z_N$ , that is to be explained by the first  $\hat{M}$  components (typically  $\tau = 0.95$  or  $\tau = 0.99$ ). In practice, the infinite sum in the denominator is replaced by the sum over all calculated eigenvalues.

As seen in expression (4), the value  $\hat{M}$  relates to the variability in  $z_i = (w_i, v_i)$  and not in the decomposition of the original data,  $\zeta_i = (w_i, \gamma_i)$ . In the following, we show that the ratio in (4) can also be interpreted as a proportion of Fréchet variances for  $\zeta_i$  by defining an appropriate metric. The Fréchet mean and variance generalize the concept of mean and variance to random variables on general metric spaces and have been studied in the context of random densities in  $B^2(\mathcal{T})$  by van den Boogaart, Egozcue, and Pawlowsky-Glahn (2014) and Petersen and Müller (2016), among others. Our findings represent new results and have not been considered in Lee and Jung (2016) or Hadjipantelis et al. (2015).

**Definition 1.** For  $\zeta_1, \zeta_2 \in \mathcal{G} := L^2(\mathcal{T}) \times \Gamma(\mathcal{T})$  with  $\zeta_i = (w_i, \gamma_i)$  and  $z_i = (w_i, \Psi(\gamma_i))$ ,  $i = 1, 2$ , define

$$d(\zeta_1, \zeta_2) = \|z_1 - z_2\|_w = \left[ \|w_1 - w_2\|_2^2 + C^2 \|\Psi(\gamma_1) - \Psi(\gamma_2)\|_2^2 \right]^{1/2} \quad \text{with } C > 0.$$

As  $d$  is induced by the norm  $\|\cdot\|_w$  on  $\mathcal{H}$ , it is immediately clear that  $d$  defines a valid metric on  $\mathcal{G}$ . On the basis of this metric, we obtain the following result.

**Theorem 1.** Consider a random element  $\zeta \in \mathcal{G}$  with  $\zeta = (w, \gamma)$ . Then the Fréchet mean  $\mathbb{E}_d(\zeta)$  and Fréchet variance  $\text{Var}_d(\zeta)$  of  $\zeta$  based on the metric  $d$  are given by

$$\mathbb{E}_d(\zeta) = (\mathbb{E}(w), \Psi^{-1}(\mathbb{E}(\Psi(\gamma)))) , \quad \text{Var}_d(\zeta) = \int_{\mathcal{T}} \text{Var}(w(t)) + C^2 \text{Var}(\Psi(\gamma(t))) dt,$$

where  $\mathbb{E}(f)(t) = \mathbb{E}[f(t)]$  denotes the usual mean function for random functions  $f \in L^2(\mathcal{T})$ .

*Proof.* Fréchet mean:

$$\begin{aligned} \mathbb{E}_d(\zeta) &= \arg \inf_{\theta \in \mathcal{G}} \mathbb{E} [d(\zeta, \theta)^2] = \arg \inf_{(w_\theta, \gamma_\theta) \in L^2(\mathcal{T}) \times \Gamma(\mathcal{T})} \mathbb{E} \left[ \|w - w_\theta\|_2^2 + C^2 \|\Psi(\gamma) - \Psi(\gamma_\theta)\|_2^2 \right] \\ &= \left( \arg \inf_{w_\theta \in L^2(\mathcal{T})} \mathbb{E} \left[ \|w - w_\theta\|_2^2 \right], \arg \inf_{\gamma_\theta \in \Gamma(\mathcal{T})} \mathbb{E} \left[ \|\Psi(\gamma) - \Psi(\gamma_\theta)\|_2^2 \right] \right) = (\mathbb{E}(w), \Psi^{-1}(\mathbb{E}(\Psi(\gamma)))) , \end{aligned}$$

where the last step can be seen, for example, by completing the squares.

Fréchet variance:

$$\begin{aligned} \text{Var}_d(\zeta) &= \mathbb{E}(d(\zeta, \mathbb{E}_d(\zeta))^2) = \mathbb{E} \left[ \|w - \mathbb{E}(w)\|_2^2 + C^2 \|\Psi(\gamma) - \Psi(\Psi^{-1}(\mathbb{E}(\Psi(\gamma))))\|_2^2 \right] \\ &= \mathbb{E} \left[ \int_{\mathcal{T}} (w(t) - \mathbb{E}(w(t)))^2 dt + C^2 \int_{\mathcal{T}} (\Psi(\gamma(t)) - \mathbb{E}(\Psi(\gamma(t))))^2 dt \right] \\ &= \int_{\mathcal{T}} \text{Var}(w(t)) dt + C^2 \int_{\mathcal{T}} \text{Var}(\Psi(\gamma(t))) dt. \end{aligned}$$

□

The Fréchet mean of  $\zeta$  based on  $d$  is hence given by the mean function of the transformed data  $\mathbf{z}$ , with the warping part back transformed to  $\Gamma(\mathcal{T})$  via  $\Psi^{-1}$ . Similarly, the Fréchet variance can be obtained as a weighted sum of the integrated pointwise variances of the elements of  $\mathbf{z}$ .

With the use of the infinite bivariate Karhunen–Loève expansion  $\mathbf{z} = \boldsymbol{\mu}_z + \sum_{m=1}^{\infty} \rho_m \boldsymbol{\phi}_m$  of a random element  $\mathbf{z} \in \mathcal{H}$ , the Fréchet variance of the associated  $\zeta$  can be rewritten as

$$\begin{aligned} \text{Var}_d(\zeta) &= \int_{\mathcal{T}} \text{Var}(w(t)) dt + C^2 \int_{\mathcal{T}} \text{Var}(\Psi(\gamma(t))) dt \\ &= \int_{\mathcal{T}} \text{Var} \left( \sum_{m=1}^{\infty} \rho_m \boldsymbol{\phi}_m^{(w)}(t) \right) + C^2 \text{Var} \left( \sum_{m=1}^{\infty} \rho_m \boldsymbol{\phi}_m^{(v)}(t) \right) dt \\ &= \sum_{m=1}^{\infty} v_m \int_{\mathcal{T}} \left[ \boldsymbol{\phi}_m^{(w)}(t)^2 + C^2 \boldsymbol{\phi}_m^{(v)}(t)^2 \right] dt = \sum_{m=1}^{\infty} v_m \|\boldsymbol{\phi}_m\|_w^2 = \sum_{m=1}^{\infty} v_m, \end{aligned}$$

because the scores  $\rho_m$  are uncorrelated with variance  $\text{Var}(\rho_m) = v_m$  and the principal components  $\boldsymbol{\phi}_m$  have weighted norm 1. Thus, the Fréchet variance of  $\zeta$  equals the total variance in  $\mathbf{z}$ .

For a given  $M$ , we obtain the fundamental variance decomposition

$$\text{Var}_d(\zeta) = \sum_{m=1}^M v_m + \sum_{m=M+1}^{\infty} v_m = \text{Var}_d(\zeta^{[M]}) + \text{Var}_d(\zeta - \zeta^{[M]}),$$

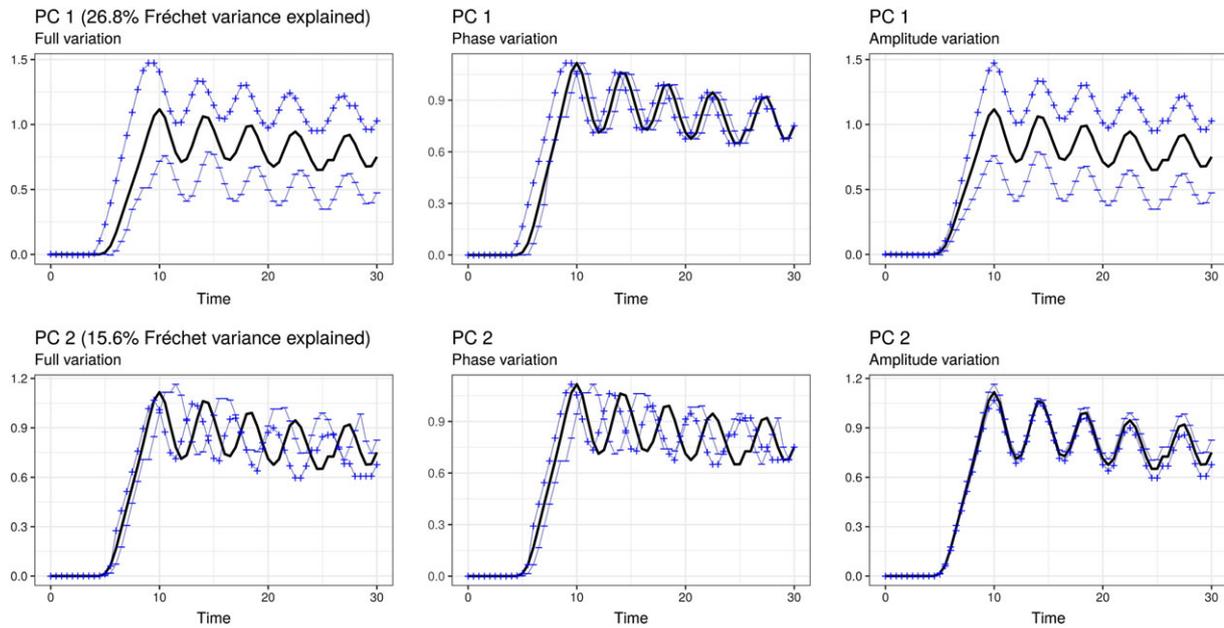
that is, the Fréchet variance explained by the first  $M$  principal components is  $V_{[M]} := \text{Var}_d(\zeta^{[M]})$  with  $\zeta^{[M]} = (w^{[M]}, \Psi^{-1}(v^{[M]}))$  corresponding to a truncated version of  $\zeta$  including the first  $M$  principal components. The fraction arising in the definition of  $\hat{M}$  in (4) is thus the proportion that the Fréchet variance in the first  $M$  components,  $V_{[M]}$ , explains with respect to the total Fréchet variance in  $\zeta$ . A similar result has been obtained for density functions in Petersen and Müller (2016). With the use of the metric  $d$  as introduced in Definition 1, our results show that the variance decomposition is valid in the more complex space  $\mathcal{G}$ , which combines warping functions  $\gamma_i$  and registered functions  $w_i$ .

## 4 | APPLICATION TO EARTHQUAKE SIMULATIONS

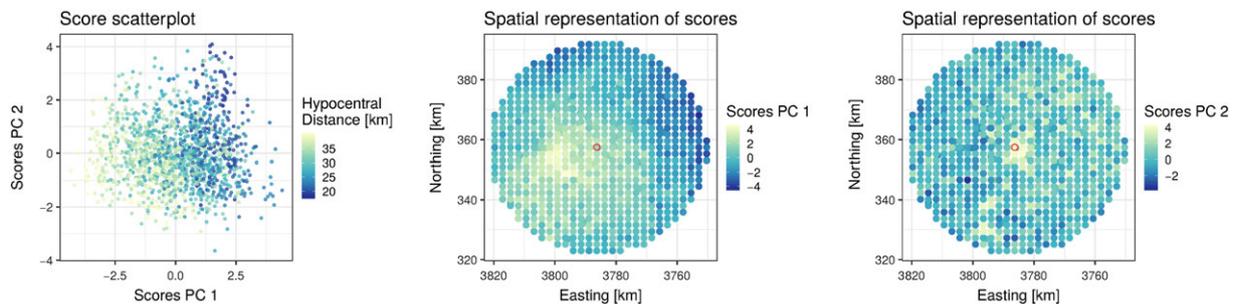
Using the proposed Bayes space transformation  $\Psi_B$  and multivariate functional principal component approach, we apply our method to a subset of data from a seismological in silico experiment based on the Mw 6.7 1994 Northridge earthquake, a blind thrust event that was felt over 200,000 km<sup>2</sup>. The induced ground shaking exceeded engineering building codes and resulted in 60 fatalities, >7,000 injured, 40,000 damaged buildings, and \$44 billion in economic losses. The hypocentre was located at about 19 km depth on a fault dipping southward at about 35° below the San Fernando Valley in the Los Angeles metropolitan area (Hauksson, Jones, & Hutton, 1995). The strongly pronounced topographic relief in the vicinity of the fault is associated with dextral transpression at the Pacific–North American plate boundary (Montgomery, 1993). The patterns of damage that occurred during the Northridge event showed irregular distributions. Generally, the region closest to the earthquake was shaken most severely. However, there were also isolated pockets of damage at distant locations.

Estimation of realistic ground motions for complex surface topography is a long-standing challenge in computational seismology (e.g., Chaljub et al., 2010). The influence of topography on local site responses to seismicity is challenging to consider in current seismic hazard assessment typically based on empirical ground motion prediction equations, even though surface topography can have a significant impact on seismic wave propagation and earthquake ground motions (Boore, 1973; Bouchon, 1973). For the data analysed here, physics-based earthquake scenarios were modelled using three-dimensional unstructured meshes (Rettenberger, Meister, Bader, & Gabriel, 2016) constructed from geological constraints such as high-resolution topography data and the Southern California Earthquake Center Community Fault Model combined with a one-dimensional subsurface structure (Wald, Heaton, & Hudnut, 1996). With the use of SeisSol, an open-source earthquake simulation package that couples three-dimensional seismic wave propagation to the simulation of dynamic rupture propagation across earthquake fault zones (Heinecke et al., 2014; Pelties, Gabriel, & Ampuero, 2014; Uphoff et al., 2017), multiple simulations varying the initial fault stress and strength conditions were performed. Time series of absolute ground velocity were recorded at a dense network of virtual seismometers distributed across Southern California. A more detailed description of the data and an analysis on the full set of original, unregistered data can be found, for example, in Bauer, Scheipl, Küchenhoff, and Gabriel (2018).

For our example, we focus on seismometer locations with a maximum distance of 40 km from the hypocentre from two simulations, which showed the strongest ground velocity movements on average. This results in a total of 1,558 observation units, each of which is recorded at 2 Hz over 30 s for a total of 61 timepoints. We presmoothed the ground velocity curves using a Tweedie distribution with log-link for the response and 40 cubic regression splines with the penalized second derivative using the R-package `mgcv` (Wood, Pya, & Säfken, 2016) to model the evolution over time before registration. The corresponding smoothed curves representing  $\log(1 + V_i(t))$  for ground velocity  $V$  at locations  $s_i$  and time  $t$  are shown in Figure 4. The figure also shows the warping functions and the aligned functions after SRVF-based warping (Tucker, 2014). This is the same warping approach as used in Lee and Jung (2016), but other approaches are possible instead. The results of the warping are the inputs of the analysis. The warping functions  $\gamma_i$  are transformed to  $L^2(\mathcal{T})$  via the Bayes space transform  $\Psi_B$ . Together with the registered functions  $w_i$ , the transformed warping functions are then fed into the MFPCA, which is calculated using the R-package `MFPCA` (Happ, 2018). We choose



**FIGURE 5** Results for the first two principal components of the joint principal component (PC) analysis for the earthquake dataset. The thick black curves represent the mean  $\bar{w} \circ \Psi_B^{-1}(\bar{v})$ , and the blue lines correspond to  $(\bar{w} \pm \alpha_w \hat{v}_m^{1/2} \hat{\phi}_m^{(w)}) \circ \Psi_B^{-1}(\bar{v} \pm \alpha_v \hat{v}_m^{1/2} \hat{\phi}_m^{(v)})$ . First or second row: Effect of the first ( $m = 1$ ) or second ( $m = 2$ ) principal component. Left: Joint effect ( $\alpha_w = 1, \alpha_v = 1$ ). Middle: Phase effect ( $\alpha_w = 0, \alpha_v = 1$ ). Right: Amplitude effect ( $\alpha_w = 1, \alpha_v = 0$ ). Note that these visualizations show the absolute time of the aligned functions on the horizontal axis



**FIGURE 6** Scores for the first two principal components (PCs) of the combined principal component analysis. Left: Scatterplot of the scores  $\rho_{11}, \rho_{12}$  depending on the hypocentral distance. Middle or right: Spatial distribution of the scores for the first or second principal component. Spatial locations of the seismometers are given in Universal Transverse Mercator coordinates. The red circle corresponds to the seismometer with minimal hypocentral distance (red curve in Figure 4)

$M = 10$  principal components, for which the optimal weight is found to be  $\hat{C} = 1.03$ . The weight of the aligned functions  $w_i$  in the MFPCA thus approximately equals that of the transformed warping functions.

The first principal component, which explains 26.8% of the (Fréchet) variability in the data, is illustrated in Figure 5. Positive scores in this component are associated with a larger amplitude than the mean with five clearly defined peaks that occur earlier than in the mean. For negative scores, one observes a lower amplitude and slightly delayed peaks. Visualizing phase and amplitude variation for this component separately (middle and right panels, respectively) shows that the phase variation in this first principal component is a fairly constant shift in time up to the last peak, whereas its main characteristic is driven by amplitude variation. In other words, for positive scores in the first principal component, ground movement is stronger and begins slightly earlier, whereas for negative scores, ground movement is somewhat delayed and markedly less severe. Therefore, one would naturally expect rather positive scores for the first principal component for seismometers closer to the hypocentre and more negative values for those further away.

The second principal component (representing 15.6% of Fréchet variability) shows a different pattern. Here, the main component is phase variation, whereas amplitude variation plays only a minor role affecting local maxima and minima after 20 s; see Figure 5, bottom row. For functions with positive scores, we observe that peaks occur earlier than in the mean function and that this temporal shift increases over time. Moreover, the amplitude tends to decrease faster than in the mean. Conversely, for functions with negative scores, we observe five increasingly delayed peaks with more slowly decreased amplitude after 20 s.

Figure 6 contains a scatter plot of the first two principal component scores coloured by hypocentral distance (left panel) and the spatial distribution of the first two PC score vectors (middle and right panels). As expected, seismometers that are closer to the hypocentre are characterized by positive scores for the first principal component (earlier and larger movement), that is, having scores in the right part of the first

scatterplot, whereas seismometers with a higher hypocentral distance show the opposite behaviour. We observe pronounced directivity effects caused by the unilateral earthquake rupture and dynamically enforced by the fault morphology. The dipping orientation of the fault adds further spatial complexity to the ground shaking. The middle panel of Figure 6 shows larger and earlier ground motions in lighter colours. Besides source effects, topographic effects such as amplification, deamplification, scattering, and channelling of seismic waves strongly affect the seismic data analysed. Concerning the second principal component, the image is more heterogeneous with some local spots having especially high values; that is, here, the ground movements arrive earlier but also decrease more rapidly.

The seismological *in silico* experiment presented in this section is exemplary for data in which both amplitude and phase variation are of interest. Using the Bayes space transformation for the warping functions, we preserve as much of the structure in the data as possible, as detailed in Section 2. The bivariate functional principal components of the transformed warping functions and the registered functions are interpretable and provide interesting insights into the joint variation of amplitude and phase. In addition, the results found in Section 3 allow us to quantify the proportion of variability in the data explained by each principal component in terms of the Fréchet variance using the metric defined in Definition 1. It is expected that more in-depth analyses on the complete data will shed light on long-standing questions of source and path effects on ground motions.

## DATA AVAILABILITY STATEMENT

The Supporting Information contains R code and data to fully reproduce the toy example in Section 2 and the application in Section 4.

## ACKNOWLEDGEMENTS

We would like to thank Almond Stöcker for drawing our attention to the theory of Bayes spaces for probability densities and David Rügamer for his valuable comments on the text. Regarding the seismological application, we are grateful to Helmut Küchenhoff and Alexander Bauer for their collaboration. Fabian Scheipl's work has been supported by the German Federal Ministry of Education and Research (BMBF) under Grant No. 01IS18036A.

Alice-Agnes Gabriel's work was supported by the German Research Foundation (DFG) (projects no. KA 2281/4-1, GA 2465/2-1, GA 2465/3-1), by BaCaTec (project no. A4), by KONWIHR - the Bavarian Competence Network for Technical and Scientific High Performance Computing (project NewWave), by the Volkswagen Foundation (ASCETE, grant no. 88479), by KAUST-CRG (GAST, grant no. ORS-2016-CRG5-3027 and FRAGEN, grant no. ORS-2017-CRG6 3389.02), by the European Union's Horizon 2020 research and innovation program (ExaHyPE, grant no. 671698 and ChEESE, grant no. 823844). Computing resources were provided by the Institute of Geophysics of LMU Munich (Oeser et al., 2006), the Leibniz Supercomputing Centre (LRZ, projects no. h019z, pr63qo, and pr45fi on SuperMUC).

## ORCID

Clara Happ  <https://orcid.org/0000-0003-4737-3835>

## REFERENCES

- Bauer, A., Scheipl, F., Küchenhoff, H., & Gabriel, A. A. (2018). An introduction to semiparametric function-on-scalar regression. *Statistical Modelling*, 18(3-4), 346–364.
- Boore, D. M. (1973). The effect of simple topography on seismic waves: Implications for the accelerations recorded at Pacoima Dam, San Fernando Valley, California. *Bulletin of the Seismological Society of America*, 63(5), 1603–1609.
- Bouchon, M. (1973). Effect of topography on surface motion. *Bulletin of the Seismological Society of America*, 63(2), 615–632.
- Chakraborty, A., & Panaretos, V. M. (2017). Functional registration and local variations. arXiv: 1702.03556.
- Chaljub, E., Moczo, P., Tsuno, S., Bard, P.-Y., Kristek, J., Käser, M., ..., & Kristekova, M. (2010). Quantitative comparison of four numerical predictions of 3D ground motion in the Grenoble Valley, France. *Bulletin of the Seismological Society of America*, 100(4), 1427–1455.
- Chiou, J.-M., Yang, Y.-F., & Chen, Y.-T. (2014). Multivariate functional principal component analysis: A normalization approach. *Statistica Sinica*, 24, 1571–1596.
- Egozcue, J. J., Díaz-Barrero, J. L., & Pawlowsky-Glahn, V. (2006). Hilbert space of probability density functions, V based on Aitchison geometry. *Acta Mathematica Sinica, English Series*, 22(4), 1175–1182.
- Hadjipantelis, P. Z., Aston, J. A. D., Müller, H. G., & Evans, J. P. (2015). Unifying amplitude and phase analysis: A compositional data approach to functional multivariate mixed-effects modeling of Mandarin Chinese. *Journal of the American Statistical Association*, 110(510), 545–559.
- Happ, C. (2018). MFPCA: Multivariate functional principal component analysis for data observed on different dimensional domains [Computer software manual]. Retrieved from <https://github.com/ClaraHapp/MFPCA> R package version 1.3-1.
- Happ, C., & Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association*, 113(522), 649–659.
- Hauksson, E., Jones, L. M., & Hutton, K. (1995). The 1994 Northridge earthquake sequence in California: Seismological and tectonic aspects. *Journal of Geophysical Research: Solid Earth*, 100(B7), 12335–12355.
- Heinecke, A., Breuer, A., Rettenberger, S., Bader, M., Gabriel, A.-A., Pelties, C., ..., & Dubey, P. (2014). Petascale high order dynamic rupture earthquake simulations on heterogeneous supercomputers. In *International Conference for High Performance Computing, Networking, Storage and Analysis (SC14)*. New Orleans, USA: IEEE, pp. 3–14.

- Hron, K., Menafoglio, A., Templ, M., Hrušová, K., & Filzmoser, P. (2016). Simplicial principal component analysis for density functions in Bayes spaces. *Computational Statistics and Data Analysis*, 94, 330–350.
- Kneip, A., & Ramsay, J. O. (2008). Combining registration and fitting for functional models. *Journal of the American Statistical Association*, 103(483), 1155–1165.
- Lee, S., & Jung, S. (2016). Combined analysis of amplitude and phase variations in functional data. arXiv:1603.01775.
- Marron, J. S., Ramsay, J. O., Sangalli, L. M., & Srivastava, A. (2015). Functional data analysis of amplitude and phase variation. *Statistical Science*, 30(4), 468–484.
- Montgomery, D. R. (1993). Compressional uplift in the central California Coast Ranges. *Geology*, 21(6), 543–546.
- Oeser, J., Bunge, H.-P., & Mohr, M. (2006). Cluster Design in the Earth Sciences: TETHYS, in High Performance Computing and Communications – Second International Conference, HPCC 2006, Munich, Germany, Lecture Notes in Computer Science, vol. 4208, edited by Michael Gerndt and Dieter Kranzlmüller, pp. 31–40, Springer. [https://doi.org/10.1007/11847366\\_4](https://doi.org/10.1007/11847366_4)
- Pelties, C., Gabriel, A.-A., & Ampuero, J.-P. (2014). Verification of an ADER-DG method for complex dynamic rupture problems. *Geoscientific Model Development*, 7(3), 847–866.
- Petersen, A., & Müller, H.-G. (2016). Functional data analysis for density functions by transformation to a Hilbert space. *The Annals of Statistics*, 44(1), 183–218.
- Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis* (2nd ed.). New York: Springer.
- Rettenberger, S., Meister, O., Bader, M., & Gabriel, A.-A. (2016). ASAGI: A parallel server for adaptive geoinformation. In *Proceedings of the Exascale Applications and Software Conference 2016 (EASC2016)*. Stockholm, Sweden: ACM, pp. 2.
- Sangalli, L. M., Secchi, P., Vantini, S., & Vitelli, V. (2010). K-mean alignment for curve clustering. *Computational Statistics and Data Analysis*, 54(5), 1219–1233.
- Srivastava, A., Jermyn, I., & Joshi, S. (2007). Riemannian analysis of probability density functions with applications in vision. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 44, IEEE, pp. 1–8.
- Srivastava, A., & Klassen, E. P. (2016). *Functional and shape data analysis*. New York: Springer.
- Tucker, J. D. (2014). *Functional component analysis and regression using elastic methods* (Ph.D. Thesis). Florida State University, Tallahassee, USA.
- Uphoff, C., Rettenberger, S., Bader, M., Madden, E. H., Ulrich, T., Wollherr, S., & Gabriel, A.-A. (2017). Extreme scale multi-physics simulations of the tsunamigenic 2004 Sumatra megathrust earthquake. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC17)*. Denver, USA: ACM, pp. 21:1–21:16.
- van den Boogaart, K. G., Egozcue, J. J., & Pawlowsky-Glahn, V. (2014). Bayes Hilbert spaces. *Australian and New Zealand Journal of Statistics*, 56(2), 171–194.
- Wald, D. J., Heaton, T. H., & Hudnut, K. W. (1996). The slip history of the 1994 Northridge, California, earthquake determined from strong-motion, teleseismic, GPS, and leveling data. *Bulletin of the Seismological Society of America*, 86(1B), S49–S70.
- Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, 111(516), 1548–1563.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Happ C, Scheipl F, Gabriel A-A, Greven S. A general framework for multivariate functional principal component analysis of amplitude and phase variation. *Stat.* 2019;8:e220. <https://doi.org/10.1002/sta4.220>

## APPENDIX A

*Proof of Equation 3.* The first equation is clear by the definition of the mappings  $\tilde{\psi}_{S,\mu}$ ,  $\tilde{\psi}_{S,\mu}^{-1}$  and the spaces  $S_+^\infty(\mathcal{T})$ ,  $S^\infty(\mathcal{T})$ .

For the second equation, choose, for example,  $\mu = q_0$ , the SRVF associated with identity warping. By the Cauchy–Schwarz inequality,  $|\langle q, \mu \rangle_2| \leq \|q\|_2 \|\mu\|_2 = \eta$  for all  $q, \mu \in S_+^\infty(\mathcal{T})$ . Using that  $q(t) \geq 0$  for all  $t \in \mathcal{T}$  further yields  $\langle q, \mu \rangle_2 = \int_{\mathcal{T}} q(t) dt \geq 0$ . Therefore,  $\frac{\langle q, \mu \rangle_2}{\eta} \in [0, 1]$  and  $\theta \in \left[0, \frac{\pi}{2}\right]$ .

Now, let  $v = \tilde{\psi}_{S,\mu}(q)$ . It is easy to show that  $\|v\|_2^2 = \theta^2$  (Srivastava & Klassen, 2016), which implies  $\|v\|_2 \leq \frac{\pi}{2}$ . The boundary cases are given by  $\|v\|_2 \rightarrow 0$  for  $q \rightarrow \mu$ ; hence,  $\langle q, \mu \rangle_2 \rightarrow \eta$  and  $\|v\|_2 \rightarrow \frac{\pi}{2}$  for  $\langle q, \mu \rangle_2 = 0$ , meaning that  $q \in T_\mu(\mathcal{T})$  already. Note that the latter case is only achieved by degenerate functions  $q$  that correspond to stepwise constant warping functions  $\gamma$ .

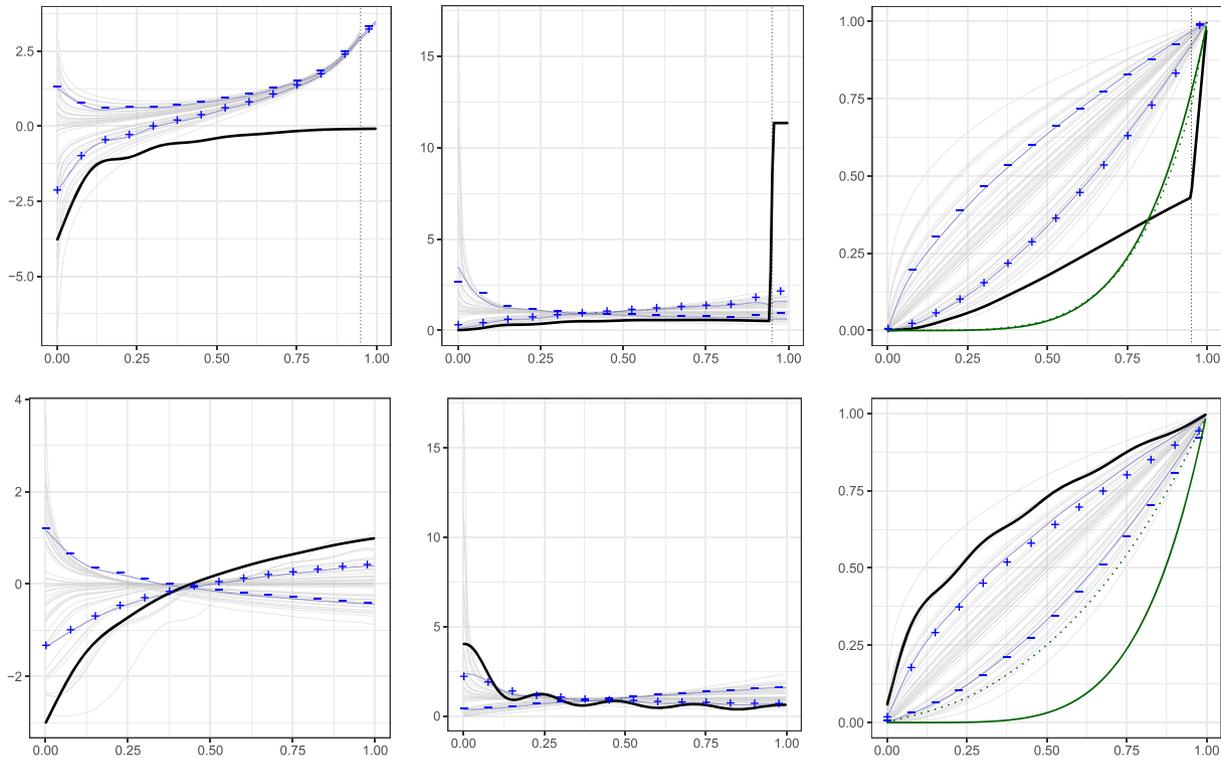
Using  $q(t) \geq 0$  for all  $t \in \mathcal{T}$  further yields

$$v(t) = \frac{\theta}{\eta^{1/2} \sin(\theta)} (q(t) - \cos(\theta)\mu(t)) \geq \frac{q(t) - \cos(\theta)}{\eta^{1/2}} \geq -\eta^{-1/2}.$$

This shows

$$\text{im}(\tilde{\psi}_{S,\mu}) \subseteq \left\{ v \in T_\mu(\mathcal{T}) : \|v\|_2 \leq \frac{\pi}{2}, v(t) \geq -\eta^{-1/2} \forall t \in \mathcal{T} \right\} \subsetneq T_\mu(\mathcal{T}).$$

As  $\|v\|_2 < \pi$  for all  $v$  in  $\text{im}(\tilde{\psi}_{S,\mu})$ , the restriction of  $\tilde{\psi}_{S,\mu}^{-1}$  to  $\text{im}(\tilde{\psi}_{S,\mu})$  is a bijection (Srivastava & Klassen, 2016). For alternative choices of  $\mu$ , analogous results can be obtained.



**FIGURE A1** Principal component analysis of warping functions based on  $\psi_H$  (top row) or  $\psi_Q$  (bottom row) for the same warping functions as in Figure 2 and different stages of transformation. The grey functions represent the (transformed) data, and the black functions the first principal component  $\hat{\phi}_1$ . Blue lines visualize the effect of the first principal component as perturbation from the mean  $(\hat{\mu} \pm \hat{\lambda}_1^{1/2} \hat{\phi}_1)$ . Left:  $L^2(\mathcal{T})$ . Centre: Density space  $B^2(\mathcal{T})$ . Right: Space of warping functions  $\Gamma(\mathcal{T})$ . The dotted vertical line in the top row corresponds to the threshold parameter  $\delta$  for the log-hazard transformation  $\psi_H$ , here chosen as  $\delta = 0.05$ . The green curves correspond to a new observation (solid) and its reconstruction (dotted) based on the first principal component