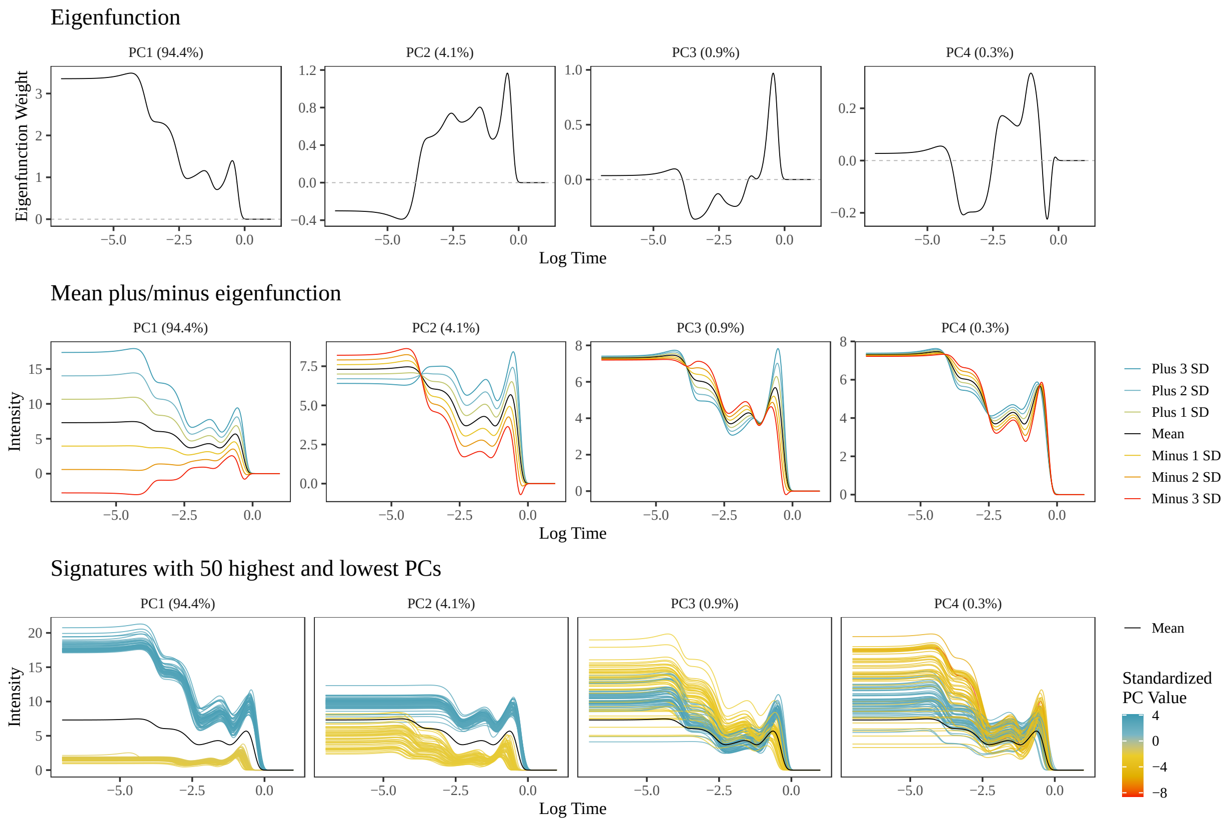

Figure 6: Visualizations of the first four fPCs from the simulated optical-spectral temporal explosion signatures.

- *fPC 2*: The eigenfunction for fPC 2 makes it clear that fPC 2 explains a contrast between time points before and after -3.75. The visualizations of the point-wise mean function plus/minus the eigenfunctions and the signatures corresponding to the 50 highest and lowest fPC 2 values indicate that fPC 2 captures a contrast between signatures with high starting values and 3 peaks that occur after the mean function and signatures with lower starting values and four peaks before the mean function.
- *fPC 3*: The eigenfunction of fPC 3 indicates that the fPC explains a contrast in variability between the times of (-3.75, -1.75) and (-1.75, 0). The mean plus/minus eigenfunctions and signatures with extreme fPC 3 values suggest that fPC 3 captures the variability between signatures with lower values between the first time interval and a large fourth peak during the second time interval and signatures with higher values during the first time interval and a small fourth peak during the second time interval.
- *fPC 4*: The eigenfunction for fPC 4 depicts that fPC 4 explains a contrast between the two time intervals of (-4,-2.5) and (-0.5,0) and the time interval of (-2.5,-0.5). The other two visualizations for fPC 4 indicate that fPC 4 captures the variability between signatures with a steep decrease during the time interval of (-4,-2.5) and dramatic third and fourth peaks during the time intervals of (-2.5,-0.5) and (-0.5,0), respectively and signatures with less intense peaks throughout the entire time interval of (-4,0).

To connect the interpretations of the first four fPCs to the explosive device characteristics, consider the functional means for the Y_1, Y_2, and Y_3 in Figure 3. PFI identified fPCs 1 and 2 as being important for predicting Y_1. This is reasonable since both fPCs capture a variability in functions with intensity during early times, different timings for peak 1, and the number of peaks (3 or 4). PFI found that fPCs 1 and 3 are important for predicting Y_2, which is reasonable since both fPCs capture a variability between signatures with high and low intensities across the entire time interval. PFI identified fPC 2 as being the most important for predicting Y_3, which is reasonable since fPC 2 captures the variability between signatures with high intensity values and peaks occurring after the mean function and signatures with low intensity values and peaks occurring before the mean function. PFI also identified fPCs 1, 3, and 4 as having some importance for predicting Y_3, and these fPCs pick up on smaller amounts of variability affected by Y_3 such as the intensity in certain regions and the intensity of the fourth peak. By sharing these findings with an SME, we are able to confirm that the fPCs identified by PFI as important capture the type of variability in the signatures that is important to the corresponding explosive device characteristics.

5 Discussion

The prediction of explosive device characteristics using optical spectral-temporal signatures from explosions is an ex-

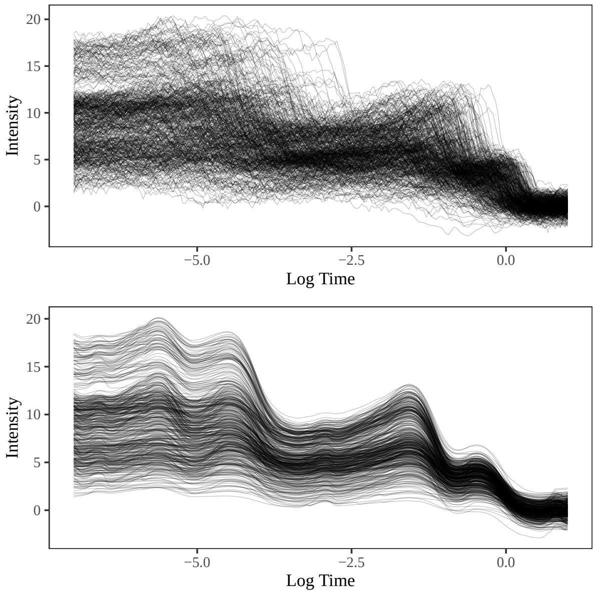

Figure 7: (Top) Examples of simulated optical spectral-temporal signatures from explosions with more variability. (Bottom) The signatures from the plot on the left after applying smoothing and alignment (using box filtering and time warping, respectively) from the fdasrvf R package (version 1.9.3; **? ?**).

ample in national security where machine learning applied to functional data could improve performance in practice. However, this is also an example of a machine learning application where not only is high predictive accuracy important, but it is also imperative that it is possible to explain how the model makes predictions. In this paper, we propose a procedure for training and explaining machine learning models with functional data inputs that accounts for the functional nature of the data. We implement our procedure to provide explainable predictions for neural networks trained to predict explosive device characteristics. In particular, the transformation of the optical spectral-temporal signatures using fPCA permits the identification of fPCs important to prediction in a neural network for an explosive device characteristic using PFI, and visualizations for interpreting the variability captured by the important fPCs allows for the determination of the aspects of the signatures that are important for prediction. The validation from an SME of the meaningfulness of the fPCs identified by PFI allows us to be confident that the neural networks are using trustworthy aspects of the signatures to make predictions.

A limitation of this method is that the ability to explain a prediction made by the neural network is dependent on the ability to interpret the fPCs. In our example, PFI identifies the first four fPCs as important for predicting at least one of the characteristics, and it is possible to determine meaningful variation captured by these fPCs. However, if PFI identifies fPCs that are not able to be interpreted, it would not be possible to explain the aspects of a function that are important to the neural network for prediction. While the earlier fPCs explain larger amounts of the variability in a data set, it is not necessarily true that the earlier fPCs will be the best for discrimination of response characteristics. If PFI identifies a higher numbered fPC as important, it is likely to be more difficult to interpret.

Another aspect not considered in this paper is that fPCA accounts for amplitude variability (vertical variability) but does not account for phase variability (horizontal variability) in the functions. Joint functional principal component analysis (joint fPCA) is a method that can be applied after smoothing and aligning functional data that accounts for both amplitude and phase variability (**?**; **?**). The procedure in this paper could be adjusted by substituting fPCA with smoothing, aligning, and applying joint fPCA to the signatures (Figure 7). With noisier signatures, accounting for phase variability is important to capture the signals in the data.

The focus of this paper is the explanation of the predictions. However, another aspect important to applications connected to high consequence decisions is the ability to report uncertainty quantification to gauge the model's confidence in a prediction. Bayesian neural networks (BNNs) are an example of a machine learning method that returns uncertainty quantification. BNNs produce a distribution for prediction as opposed to a single prediction. If a BNN is used as the machine learning model in our procedure, it would be possible to develop a method that adjusts the computation of PFI to account for the distribution of predictions.

In this paper, we present a method for explaining predictions from a machine learning model trained using functional data. We demonstrate the method on a national security application in which model authentication is crucial. While all statistical methods used within our procedure have been developed previously, we join these techniques in a new way that provides insight to a black-box model. Additionally, our approach accounts for the functional nature of the data, which is an aspect that has been overlooked in the explainable machine learning literature.

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 DE-NA0003525. 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. SAND2020-11247 C

Acknowledgments

The authors would like to thank J. Derek Tucker for his advice on fPCA.

Itaque pariatur quibusdam repellendus labore beatae, corrupti sed quia illum numquam fuga quasi at, accusamus quis cum perferendis quam veniam ullam