ARL-TR-9323 ● Sep 2021

Functional Data Analysis of the RF Tactical Data

by Vinod K Mishra and Bhikhari Tharu

NOTICES

Disclaimers

ARL-TR-9323 ● SEP 2021

Functional Data Analysis of the RF Tactical Data

Vinod K Mishra
Computational and Information Sciences Directorate,
DEVCOM Army Research Laboratory

Bhikhari Tharu
Department of Mathematics, Spelman College

REPORT DOCUMENTATION PAGE			*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED (From - To)**
September 2021	Technical Report	June 2020–May 2021

4. TITLE AND SUBTITLE Functional Data Analysis of the RF Tactical Data	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Vinod K Mishra and Bhikhari Tharu	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) DEVCOM Army Research Laboratory ATTN: FCDD-RLC-NC Aberdeen Proving Ground, MD 21005	**8. PERFORMING ORGANIZATION REPORT NUMBER** ARL-TR-9323
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES

ORCID IDs: Vinod K Mishra 0000-0001-9432-9082, Bhikhari Tharu 0000-0002-6497-6366

14. ABSTRACT

The extraction of useful information from large, disparate, and heterogeneous data sets requires a careful statistical analysis and identification of hidden patterns and metrics. In this work, we apply the methods and techniques of Functional Data Analysis to the signal data in the tactical arena for this purpose.

15. SUBJECT TERMS

Functional Data Analysis, FDA, PCA, ARIMA, RF signals

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Vinod K Mishra
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	UU	17	**19b. TELEPHONE NUMBER (Include area code)**
Unclassified	Unclassified	Unclassified			(401) 278-0114

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

1. Introduction

The tactical edge, with its complicated electromagnetic environment is a very important part of the defense operations. In general, it contains a mix of friendly and adversarial radio frequency signal sources. A method for distinguishing the signals in the tactical arena will be very useful for telling blue and red teams apart. The Function Data Analysis (FDA) methods offer a promising approach to find their underlying signatures. The FDA contains techniques for understanding and analyzing large and complex data sets with hidden underlying properties. It is particularly useful in situations in which one records the data continuously during a time interval or intermittently at several discrete time points. It can also uncover nonlinear functional dependence hidden in such data.

In current work, we use FDA techniques to uncover the hidden continuous functions in the noisy field data. The measured data is a result of the combination of the signal and noise introduced by solar, atmospheric, and other electromagnetic signals present in the surrounding. The report consists of general theory behind FDA (Section 2), steps in the analysis of the field data (Section 3), and numerical results (Section 4). Finally, in Section 5 we summarize the results and point out the next steps.

2. Functional Data Analysis

Functional data is a realization of a random object taking values in a function space. They usually arise in a time series but can be also in the space of frequency, location, wavelength, and more. The measurements can be exact or error-contaminated, sparse or dense. Usually, functional data has very high or possibly infinite dimensions.

Let $\{x_i, y_t(x_i)\}$ represent the measured time-series data, where

> $y_t(x_i)$ (t= $1, 2, 3, .., n$) = measured signal at time t as a function of x_i ($n \times 1$ vector)
>
> x_i ($i = 1, 2, 3, .., m$) = time-dependent group of parameters

The FDA needs the following steps.

Step 1: *Characterization of the measured signal*

The functional time-series model assumes that the measured signal is expressible as a sum of smooth underlying function and an error term due to noise.

$$y_t(x_i) = s_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i} \qquad (1)$$

where

- $y_t(x_i)$ = the tactical data of each type xi in year t
- $s_t(x_i)$ = $n \times 1$ vector for fitted smooth functional signal at *t* as a function of x_i
- $\varepsilon_{t,i}$ = $n \times 1$ vector for errors as independent and identically distributed standard normal variates
- $\sigma_t(x_i)$ = $n \times 1$ vector for variable noise

We have implemented a nonparametric smoothing technique[1] to the data $y_t(x_i)$to prepare a smooth curve $s_t(x_i)$ as functional data object.

Step 2: *Decomposition of the functional data*

The functional signal $s_t(x)$ decomposes into statistical components as

$$s_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + e_t(x), t = 1, 2, 3, \dots, n \qquad (2)$$

Here x stands x_i $(i = 1, 2, 3, .., m)$ and

- $\mu(x) = \frac{1}{n}\sum_{t=1}^{n} s_t(x)$ = the mean curve of $s_t(x)$
- $\{\phi_k(x)\}$ = a set of orthonormal basis functions for signal expansion ($k = 1, 2, 3, .., $K). They can belong to the well-known basis-function sets like B-splines, Fourier functions, and the like, or one can estimate them by applying the Functional Principal Component (FPC) method. Application of these methods to the smooth curves $\{s_t(x_i)\}$ provides the minimum number of basis-functions and enables informative interpretation. The basis-functions are modeled as a weight assigned to the variable taking the maximum and the minimum values at the highest and the lowest peak of the curve.
- $\beta_{t,k}$ = uncorrelated coefficients[1,2] obtained by using a univariate time series method known as Auto-Regressive Integrated Moving Average (ARIMA) model
- $e_t(x)$ = the uncorrelated error of the model

The ARIMA model method identifies a small number of basis-functions, and therefore simplifies interpretations and finds the uncorrelated coefficients. In our analysis, two basis-functions provide a reasonable fit and explain most of the variability in the data.

Step 3: *Finding the optimum K*

- The expression for the mean integrated square error (MISE) is

$$\text{MISE} = \frac{1}{n}\sum_{t=1}^{n}\int\{e_t(x)\}^2 dx = \frac{1}{n}\sum_{t=1}^{n}\int\left[s_t(x) - \left\{\mu(x) + \sum_{k=1}^{K}\beta_{t,k}\phi_k(x)\right\}\right]^2 dx \tag{3}$$

- Minimization of MISE with respect to K gives the optimum value of K leading to the fitted *n*-vector.

$$x(t_i) = \sum_{k=1}^{K}\phi_k(t_i)\beta_k = \Phi\beta \tag{4}$$

Step 4: *Finding the smoothest fit*

This study adapts the penalized regression splines smoothing[3] method for this task. Since the choice of the smoothing parameter λ is crucial, the generalized cross-validation (GCV) criteria has been adapted for its determination.[4] The GCV determines the smoothness of fit through the parameter λ and is defined as

$$GCV(\lambda) = \left(\frac{n}{n-df(\lambda)}\right)\left(\frac{SSE}{n-df(\lambda)}\right) \tag{5}$$

where

- $SSE = \sum_{i=1}^{n}[y_i - x(t_i)]^2$ is the sum of squared errors (SSE),
- n = the total number of time instants, and
- $df(\lambda)$ = the degrees of freedom associated with λ parameter.
- GCV's right factor $\left(\frac{SSE}{n-df(\lambda)}\right)$ is the unbiased estimate of the error variance σ^2 similar to regression analysis, and represents some discounting of $df(\lambda)$ by n.
- Multiplication by GCV's left factor $\left(\frac{n}{n-df(\lambda)}\right)$ further discounts this estimate.[5]

Step 5: *Forecasting*

The univariate time series model are fitted to each coefficients $\beta_{t,k}$ (t = 1, · · ·, n), and these estimates are used to find the coefficients $\beta_{t,k}$ (t = n+1, · · ·, n + h). We obtain the basis-functions $\{\phi_k(x)\}$ and the coefficients by FPC and univariate time-series method, respectively. The coefficients $\beta_{t,k}$ and $\beta_{t,l}$ are assumed to be uncorrelated for k ≠ l; therefore, univariate method is applicable for forecasting each time series $\hat{\beta}_{t,k}$. We use forecast coefficients with Eq. 2 to obtain $s_t(x)$,

$t = T + 1, \cdots, T + h$. From Eq. 2, forecasts of $s_t(x)$ are also the forecasts of $y_t(x)$.[6] Combining Eqs. 1 and 2, we can write

$$y_t(x_i) = \mu(x_i) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x_i) + e_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i} \tag{6}$$

$$t = 1, 2, 3, \dots, n \tag{2}$$

This implies

$$\eta_{T,h} = \mathrm{Var}\left[y_{T+h}|y_t(x_i)\right]$$

$$\hat{\sigma}_\mu^2(x_i) + \sum_{k=1}^{K} \mathrm{Var}(\beta_{T+h}|\beta_{1,k},,,\beta_{T,k})\hat{\phi}_k^2(x_i) + \mathrm{Var}(x_i) + \sigma_{T+h}^2(x_i) \tag{7}$$

The various variances are obtained as follows:

- $\mathrm{Var}(\beta_{T+h}|\beta_{1,k},,,\beta_{T,k})$ from the time series model,
- $\hat{\sigma}_\mu^2(x_i)$ or variance of smooth estimate $\hat{\mu}(x_i)$ from the smoothing method,
- $\mathrm{Var}(x_i)$ or variance of observational error by assuming binomial distribution of mortality rates, and
- $\sigma_{T+h}^2(x_i)$ or variance of the model error by averaging $\hat{e}_t^2(x_i)$ for each x_i.

Assuming the errors are normally distributed, a $100(1-\alpha)\%$ prediction interval is $\hat{y}_{T,h}(x_i) \pm z_{\alpha/2}\sqrt{\eta_{T,h}}$. We also use the exponential smoothing state-space model[7–9] for forecasting the data and prediction intervals[10–12] and the mean integrated squared forecasting error for evaluating the accuracy of the predictions for the estimates. The statistical package "*ftsa*"[8] facilitated the work for obtaining these results.

3. Steps of the Data Analysis

We applied FDA techniques to the field measured signals provided by Aberdeen Testing Center at Aberdeen Proving Ground. The data resulted from a continuous recording of RF signals for a defined time interval or intermittently at several discrete time points. We chose a few files containing almost 100K rows of IQ data for analysis by following the steps:

Step 1: Run the statistical analysis language package R and load data (in *.mat* file)

Step 2: Run the *ftsa* package to obtain smooth functional curves as functional data object

Step 3: Run the *ftsa* module to calculate the MISE for a given K and selected basis functions $\phi_k(x)$.

We found that two basis functions were optimal since they explain more than 95% of the variations in the data. A model is chosen based on the minimum MISE for the basis function with a given number of K. The exponential smoothing state-space model was then used to forecast the transferred rates and prediction intervals.

4. Numerical Results

The following plots present the results of the FDA analysis of the field signal data.

Figure 1 displays the observed signals for both raw signals (top chart) and smoothed signals (bottom chart) for 100 sampled units. Since the observed signal does not have much noise, smoothing seems similar to observed data. We have incorporated the penalized regression splines[3] for smoothing the observed data. The colors in the figures represent a chronicle order of the rainbow starting at the first as red, and the most recent as purple. These figures are somewhat periodic in nature.

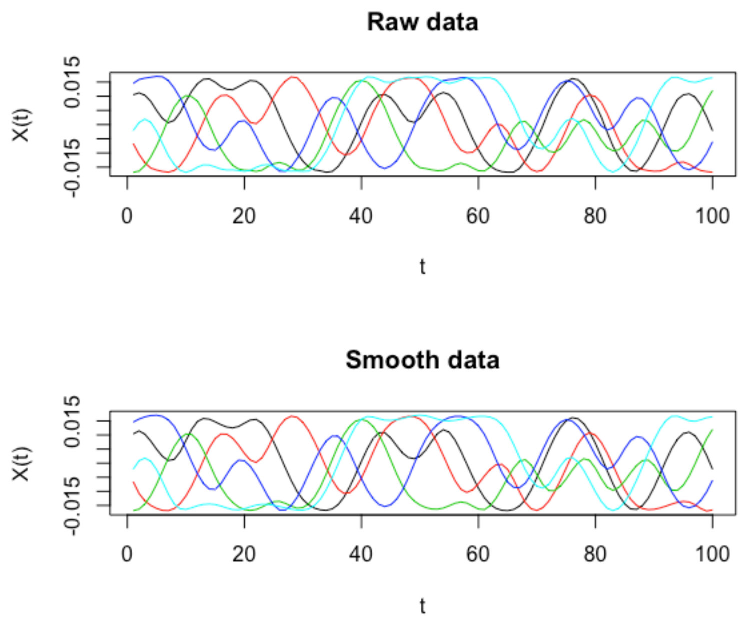

Fig. 1 Raw and smooth tactical signal

Figure 2 shows the first two functional principal components and their associated scores (in black color) with 80% prediction intervals (in yellow color) using an exponential smoothing state-space model for the data. From left to right, the first graph represents the estimated mean curve, the second and third graphs represent basis functions, the fourth and fifth represent coefficient corresponding to Basis Function 1 and Basis Function 2, respectively. In the bottom panel, the black curves represent estimated coefficients by the functional principal component method, and the blue curves are estimated through the ARIMA time-series model. First two basis functions explain most of the variability (more than 90%) present in the data. The first basis function $\phi_1(x)$ models around the second time unit, but the second basis function $\phi_2(x)$ models around the fourth time unit of the data observed. Similarly, the coefficients are primarily periodic throughout the unit time we observed.

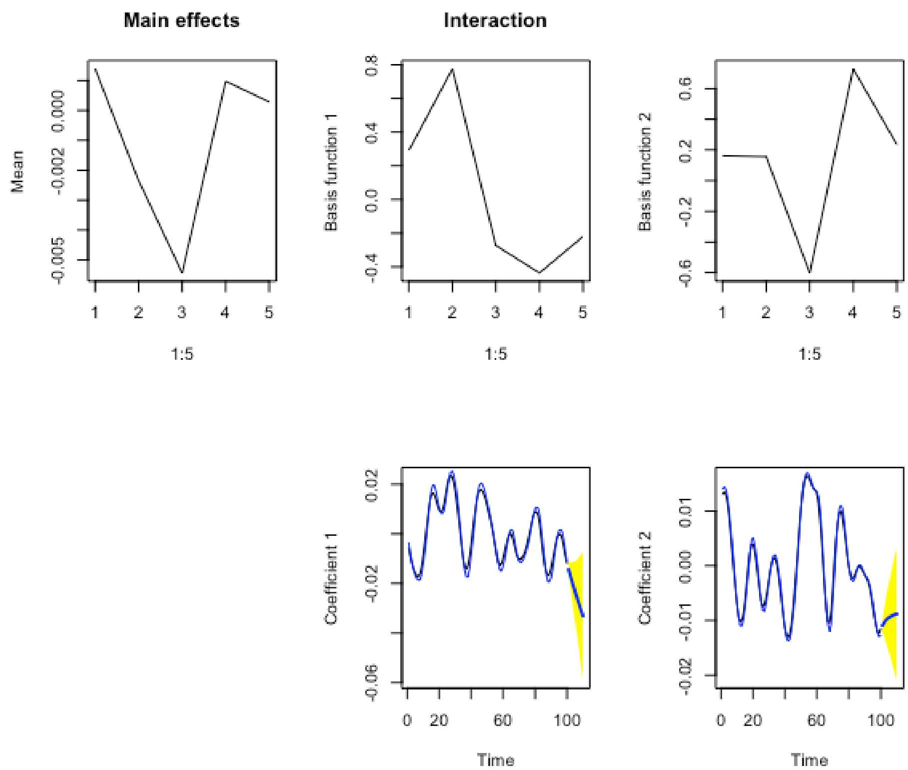

Fig. 2 Basis functions and principal components

The rainbow plot in Fig. 3 displays the forecast of the signal for next 10 units' time using the exponential smoothing state-space model.[7] This rainbow plot contains curves that are ordered chronologically within the rainbow: the first unit is red and the most recent unit (10th) is purple. To enhance the forecast accuracy of the model, sufficiently large number of principal components, K = 2 has been chosen in the study.

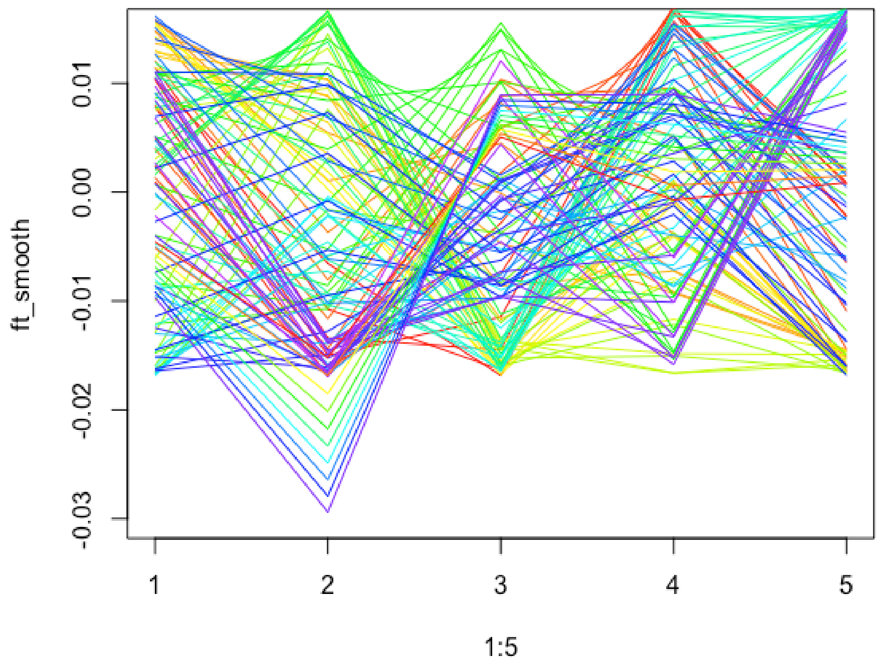

Fig. 3 Functional data including predicted curves

Figure 4 depicts raw observed data with 100 units with 10 step predictions (top panel) and smoothed data with 10 step predictions (bottom panel). Our result does not seem very consistent with the data observed, which indicates that more work needs to be done to capture the underline nature of the data. One possible issue to look at is that the FDA seems to predict linearly. Another possible point is to look at the modeling with functional time-series analysis to forecast the data points. Future analysis will help us in understanding the implication of this observation of RF modulation classification.

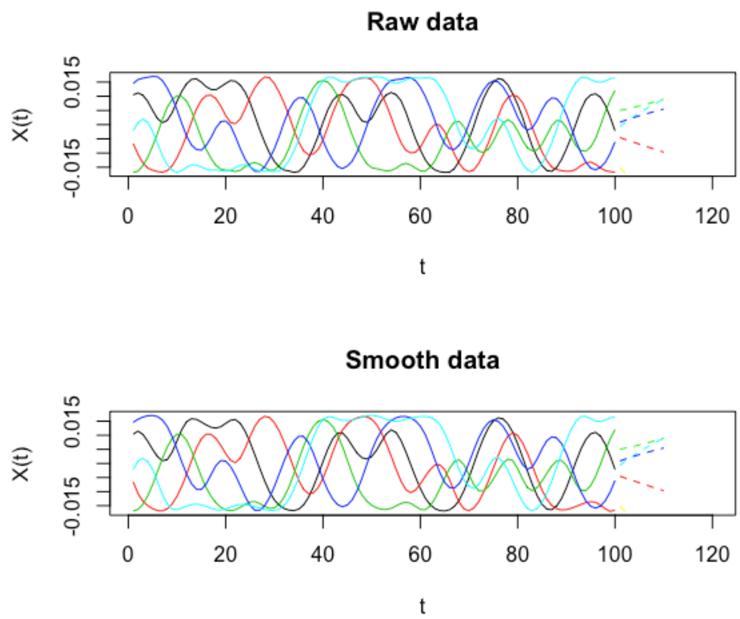

Fig. 4 Actual and smooth data with 10 step prediction estimates

5. Summary, Conclusion, and Next Steps

We have presented the results from application of FDA to signal data. We have observed that the basis functions capture the main features and trends by taking all the data into consideration. The proposed approach captures the noise-free component of the measured data. It has also shown the potential for deeper understanding of their temporal structure.

In future we will expand this framework to a larger collection of field data with known modulations and explore their functional analytic signatures. Our final aim is to identify signals with unknown and complex signatures and develop a real-time machine-learning algorithm for this purpose.

6. References

1. Ferraty F, Vieu P. Nonparametric functional data analysis. Springer-Verlag; 2006.
2. Erbas B, Hyndman RJ, Gertig DM. Forecasting age-specific breast cancer mortality using functional data models. Statistics in Medicine. 2007;26:458–470.
3. Wood SN. Monotonic smoothing splines fitted by cross validation. SIAM Journal on Scientific Computing. 1994;15:1126–1133.
4. Craven P, Wahba G. Smoothing noisy data with spline functions. Numeriche Mathematik. 1979;31:377–403.
5. Chung Seokhyun, Kontar Raed. Functional principal component analysis for extrapolating multi-stream longitudinal data. 2019 Mar 9. arXiv:1903 .03871v1 [stat.ML].
6. Ramsay JO, Silverman BW. Functional data analysis. Springer; 2005.
7. Hyndman RJ, Ullah MS. Robust forecasting of mortality and fertility rates: a functional data approach. Computational Statistics & Data Analysis. 2007;51:4942–4956.
8. Hyndman RJ, Koehler AB, Ord JK, Snyder RD. Forecasting with exponential smoothing: a state space approach. Springer; 2008.
9. Hyndman RJ. FTSA package; 2019. https://cran.r-project/org/web/packages /ftsa/citation.html.
10. Tharu B, Pokhrel K, Aryal G, Kafle RC, Khanal Na. Study of age specific lung cancer mortality trends in the US using functional data analysis. Communications for Statistical Applications and Methods. 2013;20(1):1–16 DOI: http://dx.doi.org/10.5351/CSAM.2013.20.1.001
11. Gu C, Kim Y. Penalized likelihood regression: general formulation and efficient approximation. The Canadian Journal of Statistics. 2008;30:619–628.
12. Ramsay JO, Dalzell CJ. Some tools for Functional Data Analysis. Journal of the Royal Statistical Society: Series B (Methodological). 1991;53:539–561.

List of Symbols, Abbreviations, and Acronyms

ARIMA	Auto-Regressive Integrated Moving Average
FDA	Function Data Analysis
FPC	Functional Principal Component
GCV	generalized cross-validation
MISE	mean integrated square error
RF	radio frequency
SSE	sum of squared errors

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
1 (PDF)	DEVCOM ARL FCDD RLD DCI TECH LIB
1 (PDF)	DEVCOM ARL FCDD RLC NC V MISHRA
1 (PDF)	DEPARTMENT OF MATHEMATICS SPELMAN COLLEGE B THARU