4-1-2019

Functional Singular Spectrum Analysis and the Clustering of Time-Dependent Data

Jordan Trinka
Marquette University

Degree Name:
Master of Science (MS)
Department:
Mathematics, Statistics, and Computer Science
Advisor:
Mehdi Maadooliat

Functional Singular Spectrum Analysis and the Clustering of Time-Dependent Data

by

Jordan Trinka

An essay submitted to the Faculty of the Mathematics, Statistics, and Computer Science Department,
Marquette University,
in Partial Fulfillment of the Requirements for the Degree of
Master of Science

The Klinger College of Arts and Sciences

Milwaukee, Wisconsin

April 2019

ABSTRACT

Functional Singular Spectrum Analysis and the Clustering of Time-Dependent Data

Jordan Trinka

Marquette University, 2019

The present work extends the application of the recently submitted functional singular spectrum analysis (FSSA) into the realm of structure level subsequence clustering. We begin with a comprehensive review of principal component analysis (PCA), functional principal component analysis (FPCA), singular spectrum analysis (SSA), and the recently submitted FSSA. We computationally show that the novel FSSA-FPCA hybrid clustering technique can be employed as an effective structure-based subsequence clustering method for call-center functional time series data where the method behaves as a dimension reduction technique for time-dependent data. Metrics, such as the F-ratio from k-means clustering, the w-correlation between reconstructed functional time series, and the Rand index are offered to determine the quality of clustering results of labeled functional data. We find that these outcomes are dependent on the grouping stage of FSSA for the call-center data. We also find that our measurements are not significantly sensitive to changes in groupings. Our investigations show that FSSA behaves as a type of temporal to frequency domain transformation similar to that of a Fourier analysis. The results shown in the present essay can be used to extend FSSA in its maturation and offer insight into how the hybrid method should be used and the challenges one faces with it.

ACKNOWLEDGEMENTS
Jordan Trinka

I would like to thank God for the inspiration to pursue this type of work and my fiance Sidney for her love, support, and reassurance. I am also thankful for the love and support of my family. I would also like to thank my advisers Dr. Mehdi Maadooliat, and Dr. Hossein Haghbin for their guidance as well as other members of the MSCS department that have helped me in this project. I would also like to thank the graduate school at Marquette for helping me achieve my dreams.

Contents

List of Tables

List of Figures

1 Introduction

Functional data analysis can be seen as an area of study that extends modern statistical techniques from the finite dimensional realm to the infinite dimensional realm. This allows for a functional approach to data that has a strong theoretical background in infinite dimensional vector spaces. One such technique that has been extended is principal component analysis (PCA) which reduces the dimension of the data that one is working with so that we describe the data with fewer, non-parametric variables. The functional extension of PCA is functional principal component analysis (FPCA) which takes a collection of functional objects and works to find the key functional components that best describe the vector space spanned by those objects. FPCA has been often used in clustering of subjects in order to determine the number of groups present in an unlabeled data set. Another approach just recently extended to the functional realm is singular spectrum analysis (SSA) which acts similar to PCA in the fact that it finds new variables that describe the key components of a time series. SSA has a corresponding functional counterpart known as functional singular spectrum analysis (FSSA) [1] which finds the key components of a functional time series such as trend and seasonality.

Preliminary work shows that the coupling of FSSA and FPCA in a hybrid clustering technique leads to better results for time series data than just applying FPCA directly to the functional time series [1]. The present work aims to give the novel hybrid clustering technique better traction by showing the relationship between FSSA groupings and clustering of time series data. The rest of the paper is organized to first introduce the reader to the background of the problem, give the current state of the problem, describe the methodologies used to achieve our goals, results, a discussion of those outcomes, and a conclusion.

2 Background

In order to fully develop an idea of how the new hybrid technique works, one needs to understand the ideas behind principal component analysis and how this work leads into FPCA and FSSA.

2.1 PCA

PCA is a non-parametric dimension reduction technique that works to find the components that explain most of the data in an n by p data matrix, X, where n is the number of subjects and p is the number of covariates. For this particular problem, we consider each observation to have

had the mean of observations subtracted from it such that $E[X_i] = 0$ for all $i = 1, ..., n$. One particular technique to achieve the goal of finding the principal component that explains most of the variability, and thus information, in the data is to find the unit vector, ξ, that points most in the direction of all n of the 1 by p subjects [2]. This problem can be stated as

$$\frac{1}{n} \max_{\|\xi\|_{\mathbb{R}^p}=1} \sum_{i=1}^{n} (X_i \cdot \xi)^2 \tag{1}$$

where $X_i \cdot \xi$ is the dot product between the i^{th} subject and ξ and $\| \cdot \|_{\mathbb{R}^p}$ is the norm of $\mathbb{R}^p$ induced by the dot product. As according to [2], we rewrite (1) as

$$\frac{1}{n} \max_{\|\xi\|_{\mathbb{R}^p}=1} \xi^T X^T X \xi = \frac{1}{n} \max_{\|\xi\|_{\mathbb{R}^p}=1} \xi^T C \xi = \frac{1}{n} \max_{\|\xi\|_{\mathbb{R}^p}=1} \langle \xi, C\xi \rangle_{\mathbb{R}^p} \tag{2}$$

where C is the p by p variance/covariance matrix and $\langle \cdot, \cdot \rangle_{\mathbb{R}^p}$ is the dot product. From [3], the problem presented in equation (2) is equivalent to finding the numerical radius of C defined as $\rho(C) = \max_{\|\xi\|_{\mathbb{R}^p}=1} |\langle \xi, C\xi \rangle_{\mathbb{R}^p}|$. Proposition 10.2.5 in [3] tells us that since C is symmetric (self-adjoint), $\rho(C) = |\lambda|$ where λ is the largest eigenvalue of C. Notice that if ξ is the eigenvector that corresponds to λ, then we get

$$\langle \xi, C\xi \rangle_{\mathbb{R}^p} = |\lambda| \langle \xi, \xi \rangle_{\mathbb{R}^p} = |\lambda| \|\xi\|^2_{\mathbb{R}^p} = |\lambda| = \rho(C). \tag{3}$$

Thus, our problem becomes that of finding the eigenvector ξ of C that corresponds to λ. To get the remaining components we solve the same maximization problem presented in (2) but now subject to a new constraint that $\langle \xi_{k_1}, \xi_{k_2} \rangle_{\mathbb{R}^p} = 0 \; \forall \; k_1 < k_2$ [2]. This process repeated to completion where $k_2 = p$ will form an orthonormal basis that spans $\mathbb{R}^p$.

Using the basis of principal components $\{\xi_i\}_{i=1}^p$ we find the set of principal component scores, $\{c_{i,j}\}_{j=1}^p = \{\langle X_i, \xi_j \rangle_{\mathbb{R}^p}\}_{j=1}^p$, for each observation $i = 1, ..., n$. These principal component scores are important in describing how much of each data point $X_i \in \mathbb{R}^p$ are in the direction of each $\xi_j \in \mathbb{R}^p$ [2]. Using these scores and their principal components we have the following finite Karhunen-Loeve representation of each data point [3] given by

$$X_i = \sum_{j=1}^{p} \langle X_i, \xi_j \rangle \xi_j, \; i = 1, ..., n. \tag{4}$$

These principal component scores (PCS) are useful because we determine from them which subjects behave similarly. As such, these scores can be used in an unsupervised fashion for

clustering to determine the number of groupings of unlabeled subjects and also the reasons for these groupings [4]. We extend these ideas to the functional realm to obtain FPCA which is used extensively for clustering of functional data.

2.2 FPCA

Let $\mathbb{H}$ be a Hilbert space such that it fulfills all of the requirements of a complete inner product space of functions where the domain of those elements is τ. We can think of FPCA as the decomposition of the space spanned by a finite collection of functions $\{f_i\}_{i=1}^n \in \mathbb{H}$ into functional principal components. Each f_i should be seen as a subject that lives in our Hilbert space. By [2], the maximization problem given in (2) becomes

$$\frac{1}{n} \max_{\|\xi\|_{\mathbb{H}}^2=1} \langle \xi, \Gamma\xi \rangle_{\mathbb{H}} \tag{5}$$

where Γ is the compact, linear, and self-adjoint variance/covariance integral operator over $\mathbb{H}$ and $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ is the inner product over $\mathbb{H}$. We define the image of Γ as

$$\text{Im}\,(\Gamma) = \{\Gamma g(u) \in \mathbb{H} : \Gamma g(u) = \int_\tau \nu(s,u) g(s) ds\} \tag{6}$$

where $g \in \mathbb{H}$ [2]. We also have that the kernel of Γ, denoted as ν, is seperable such that $\nu(s,u) = \frac{1}{n}\sum_{i=1}^n f_i(s) f_i(u)$ [2]. By the Hilbert-Schmidt theorem [5], there exists an orthonormal basis $\{\xi_i\}_{i=1}^\infty$ and corresponding family of scalars $\{\lambda_i\}_{i=1}^\infty$ where $(\lambda_i)_{i=1}^\infty \to 0$ monotonically such that

$$(\Gamma g)(u) = \sum_{i=1}^\infty \lambda_i \langle g, \xi_i \rangle_{\mathbb{H}} \xi_i(u) \tag{7}$$

Again, the problem statement in (5) boils down to finding the numerical radius of Γ and since Γ is a compact and self-adjoint, we have $\rho(\Gamma) = |\lambda_1| = \|\Gamma\|_{\text{op}}$ [3] where λ_1 is the leading eigenvalue of Γ, and $\|\cdot\|_{\text{op}}$ is the operator norm for bounded operators in $\mathbb{H}$. As a result, we use the eigenfunction, $\xi_1 \in \mathbb{H}$ that corresponds to λ_1 to find our solution to the maximization problem. Thus, the Hilbert-Schmidt theorem provides us with our orthonormal basis that is our functional principal component basis. In application, we use the technique presented in [6] to solve the homogeneous Fredholm integral equation of the second kind given by

$$(\Gamma\xi_i)(u) = \lambda_i\xi_i = \int_\tau \nu(s,u)\xi_i(s)ds \tag{8}$$

to find $\{\xi_i\}_{i=1}^K$ and $\{\lambda_i\}_{i=1}^K$ where we truncate at the K^{th} component. Our functional principal component scores, $c_{i,j}$, can be expressed as $\{c_{i,j}\}_{j=1}^K = \{\langle f_i, \xi_j\rangle_{\mathbb{H}}\}_{j=1}^K$ for each $i = 1, ..., n$. Without truncation, we obtain the following Karhunen-Loeve representation of each $f_i \in \mathbb{H}$ [3] as

$$f_i = \sum_{j=1}^{\infty}\langle f_i, \xi_j\rangle_{\mathbb{H}}\xi_j,\ i = 1, ..., n. \tag{9}$$

This is one technique that can be used to derive the functional principal components and corresponding scores. There has been another technique that was developed by [7] which uses penalized low rank approximations in order to try to minimize the difference between the estimations and the original collection of functions. This can be seen as the functional version of minimizing the reconstruction error between a matrix $A \in \mathbb{R}^{n\times p}$ and its low rank approximation $A^* \in \mathbb{R}^{n\times p}$. Regardless of how the functional principal component scores are obtained, they can be used to cluster subjects in our Hilbert space. One particular realm that multivariate FPCA (MFPCA), has seen some success is in clustering daily traffic flow patterns [8]. The issue with this approach is that FPCA and MFPCA assume independence of data from time where traffic flow data very much is dependent on the time of day. SSA satisfies this need for a dimension reduction technique that is applicable to time series.

2.3 SSA

A time series is a collection of data points observed over some measure of time that often have characteristics such as a mean function, seasonality, noise, etc. The goal of SSA is to separate out these characteristics from one another for reconstruction, forecasting, dimension reduction, and other purposes [9]. SSA can be broken up into the four steps of embedding, decomposition, grouping, and reconstruction.

1. Embedding: Let $y_N = \{y_i\}_{i=1}^N$ be a time series with N entries. The first step of SSA is to use a window length parameter, $0 < L < \frac{N}{2}$, to construct an L by $K = N - L + 1$ trajectory matrix X given by

$$X = \begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_{N-L+1} \\ y_2 & y_3 & y_4 & \cdots & y_{N-L+2} \\ y_3 & y_4 & y_5 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & y_{L+2} & \cdots & y_N \end{bmatrix}. \tag{10}$$

We call the columns of X, L-lagged vectors. Notice that X is Hankel such that the antidiagonal elements are equal [9]. As such, the embedding step is a linear bijection $T : \mathbb{R}^N \to \mathbb{R}_H^{L\times K}$ where $\mathbb{R}_H^{L\times K}$ is the collection of Hankel matrices that operate over $\mathbb{R}^K$. In this sense, we see that $X = Ty_N$.

2. Decomposition: Let X^* be the transpose of X and $C = X^*X$ be the variance/covariance matrix of L-lagged vectors. In the second step of SSA we use PCA to decompose C to find the principal component eigenbasis $\{\xi_i\}_{i=1}^K$ and corresponding eigenvalues $\{\lambda_i\}_{i=1}^K$ used for grouping. This decomposition offers us components and corresponding eigenvalues that allow us to determine how we can best separate the characteristics of the time series y_N.

3. Grouping: In this step, we use our decomposition from step 2 to group principal components together for reconstruction. Let $\{X_i\}_{i=1}^K = \{X\xi_i\xi_i^*\}_{i=1}^K$ be the collection of rank one matrices that approximate X such that

$$X = \sum_{i=1}^{K} X_i \tag{11}$$

We find $M \leq K$ groupings of these rank one matrices such that $\hat{X}_m = \sum_j X_j$ where $j \in 1, ..., K$ for $m = 1, ..., M$. As such, equation (11) becomes

$$X = \sum_{m=1}^{M} \hat{X}_m \tag{12}$$

We decide on which X_j's for $j = 1, ..., K$ to include in our sum to build each $\hat{X}_m$ based on eigenvalues that have similar intensity, geometry in pair-plots of principal components, and so that each $\hat{X}_m$ is approximately Hankel [9]. In doing this we form separated $\hat{X}_m$ trajectory matrices that each describe a portion of the original trajectory matrix X.

4. Reconstruction: We reconstruct these additive $\hat{X}_m$ trajectory matrices into additive time series, $\tilde{y}_N^{(m)}$, for each $m = 1, ..., M$ such that

$$y_N = \sum_{m=1}^{M} \tilde{y}_N^{(m)}. \tag{13}$$

We Hankelize each $\hat{X}_m$ using the processes described in [9] such that each entry of a Hankelized trajectory reconstruction matrix $\tilde{X}_m$, is given by

$$\tilde{x}_{i,j}^{(m)} = \begin{cases} \frac{1}{s-1} \sum_{l=1}^{s-1} \hat{x}_{l,s-l} \text{ for } 2 \leq s \leq L-1 \\ \frac{1}{L} \sum_{l=1}^{L} \hat{x}_{l,s-l} \text{ for } L \leq s \leq K+1 \\ \frac{1}{K+L-s+1} \sum_{l=s-K}^{L} \hat{x}_{l,s-l} \text{ for } K+2 \leq s \leq K+L \end{cases} \tag{14}$$

where $\hat{x}_{i,j}$ is the i, j^{th} element of $\hat{X}_m$. One can view this diagonal averaging as a Hankelization of $\hat{X}_m$ where we apply a linear operation $\mathbf{H} : \mathbb{R}^{L \times K} \rightarrow \mathbb{R}_H^{L \times K}$ and this sense, we obtain $\tilde{X}_m = \mathbf{H}\hat{X}_m$. As a result, we get the separated, additive components of y_N given by $\tilde{y}_N^{(m)} = T^{-1}\tilde{X}_m$ for $m = 1, ..., M$.

One measure of how well the components are separated is given by the w-correlation (weighted orthogonality) measure, given by

$$\rho^{(w)}\left(\tilde{y}_N^{(m_1)}, \tilde{y}_N^{(m_2)}\right) = \frac{\left(\tilde{y}_N^{(m_1)} \cdot \tilde{y}_N^{(m_2)}\right)_w}{\sqrt{\left(\tilde{y}_N^{(m_1)} \cdot \tilde{y}_N^{(m_1)}\right)_w}\sqrt{\left(\tilde{y}_N^{(m_2)} \cdot \tilde{y}_N^{(m_2)}\right)_w}} \tag{15}$$

for $m_1, m_2 \in 1, ..., M$ where

$$\left(\tilde{y}_N^{(m_1)} \cdot \tilde{y}_N^{(m_2)}\right)_w = \sum_{i=1}^{N} w_i \tilde{y}_i^{m_1} \tilde{y}_i^{m_2}. \tag{16}$$

The weights are given by

$$w_i = \begin{cases} i+1 \text{ for } 0 \leq i \leq L^* - 1 \\ L^* \text{ for } L^* \leq i \leq K^* \\ N - i \text{ for } K^* \leq i \leq N-1 \end{cases} \tag{17}$$

where $L^* = \min(\{L, K\})$ and $K^* = \max(\{L, K\})$ [9]. A w-correlation close to 0 is desired since we want to separate the original time series out into the different components that

make it up.

SSA, as a technique, has been used on data where noise is present and showed that the results are highly dependent on the type of time series and corresponding choice in L [10]. An application that multivariate SSA (MSSA) has seen success is in phase synchronization in a large system of coupled oscillators with high observational noise where the algorithm was able to detect many different oscillatory patterns [11]. The algorithm was also able to detect which patterns are “shared by clusters of phase-and-frequency locked oscillators."[11]. SSA has seen success in its use but there has been a need to extend this method to account for functional time series data.

2.4 FSSA

The extension of SSA into the functional realm is similar to the process of extending PCA to FPCA. FSSA is an algorithm that decomposes a functional time series, $\mathrm{F}_N = \{f_1, ..., f_N\}$ where $f_i \in \mathcal{L}^2([0,1])$ [1], into the additive components that make it up. As was the case in SSA, the process is broken down into four steps.

1. Embedding: We define $\mathbb{H}^L$ to be the Hilbert space built from the direct product of L copies of $\mathbb{H} = \mathcal{L}^2([0,1])$ such that the inner product of some $\mathbf{g}, \mathbf{h} \in H^L$ is given by

$$\langle \mathbf{g}, \mathbf{h} \rangle_{\mathbb{H}^L} = \sum_{i=1}^{L} \langle g_i, h_i \rangle_{\mathbb{H}}. \tag{18}$$

   For an integer $0 < L < \frac{N}{2}$ define an L-lagged functional time series vector as $\mathbf{x}_j(s) = \{f_j, f_{j+1}, \ldots, f_{j+L-1}\} \in \mathbb{H}^L$ for $j = 1, ..., K$ where $s \in [0,1]$.

   Defining $\mathbb{H}^N$ to be the Hilbert space created by the direct product of N copies of $\mathbb{H}$ and $\mathbb{H}_H^{L\times K}$ to be the collection of bounded, Hankel operators, we define the linear bijection $\mathcal{T} : \mathbb{H}^N \to \mathbb{H}_H^{L\times K}$. The embedding process then forms the trajectory operator $\mathcal{X} = \mathcal{T}\mathrm{F}_N$ where $\mathcal{X} : \mathbb{R}^K \to H^L$ such that for every $a \in \mathbb{R}^K$, we have

$$\mathcal{X}a = \sum_{j=1}^{K} a_j \mathbf{x}_j = \begin{bmatrix} \sum_{j=1}^{K} a_j f_j \\ \sum_{j=1}^{K} a_j f_{j+1} \\ \vdots \\ \sum_{j=1}^{K} a_j f_{j+L-1} \end{bmatrix}. \tag{19}$$

Notice that if we evaluate $\mathcal{X}a$ at some $s \in [0,1]$, we obtain a matrix product [1] between X which is an L by K matrix and a where

$$X(s) = \left[\mathbf{x}_1(s), \dots \mathbf{x}_K(s)\right]. \tag{20}$$

By proposition 3.1 [1] that an adjoint of $\mathcal{X}$ is a linear operator $\mathcal{X}^* : \mathbb{H}^L \to \mathbb{R}^K$ where for all $\mathbf{z} \in \mathbb{H}^L$, we have

$$\mathcal{X}^*\mathbf{z} = \begin{bmatrix} \sum_{i=1}^{L} \langle f_i, z_i \rangle \\ \sum_{i=1}^{L} \langle f_{i+1}, z_i \rangle \\ \vdots \\ \sum_{i=1}^{L} \langle f_{i+k-1}, z_i \rangle \end{bmatrix}. \tag{21}$$

2. Decomposition: Define $\mathcal{S} = \mathcal{X}\mathcal{X}^*$ to be a linear variance/covariance integral operator where $\mathcal{S} : \mathbb{H}^L \to \mathbb{H}^L$ such that

$$\mathcal{S}\mathbf{z} = \sum_{i=1}^{K}\sum_{i=1}^{L} \langle f_{i+j-1}, z_i \rangle \mathbf{x}_j = \sum_{j=1}^{K} \langle \mathbf{x}_j, \mathbf{z} \rangle_{\mathbb{H}^L} \mathbf{x}_j = \sum_{j=1}^{K} (\mathbf{x}_j \otimes \mathbf{x}_j)\, \mathbf{z}. \tag{22}$$

As stated in [1], $\mathcal{S}$ can be seen as an L by L matrix with operator entries $\mathcal{S}_{i.j} : \mathbb{H} \to \mathbb{H}$ where for a component z_j of $\mathbf{z}$, we have

$$\mathcal{S}_{i,j} z_j(u) = \int_0^1 c_{i,j}(s,u) z_j(s) ds. \tag{23}$$

We also have that $c_{i,j}(s,u)$ is a separable kernel [1] such that

$$c_{i,j}(s,u) = \sum_{k=1}^{K} f_{i+k-1}(s) f_{j+k-1}(u), \text{ for } s,u \in [0,1]. \tag{24}$$

For $u \in [0,1]$ we write

$$\mathcal{S}\mathbf{z}(u) = \int_0^1 \mathbf{C}(s,u)(z)(s) ds = \begin{bmatrix} \sum_{i=1}^{L} \int_0^1 c_{i,1}(s,u) z_i(s) ds \\ \vdots \\ \sum_{i=1}^{L} \int_0^1 c_{i,L}(s,u) z_i(s) ds \end{bmatrix} \tag{25}$$

where $\mathbf{C}(s,u) = X(s)X(u)^*$ [1]. According to proposition 3.2 [1], $\mathcal{S}$ is self-adjoint,

positive, bounded, continuous, and compact. Using the Hilbert-Schmidt theorem [5], we have that there exists an orthonormal eigenbasis $\{\xi_i\}_{i=1}^{\infty} \in \mathbb{H}^L$ and corresponding eigenvalues $\{\lambda_i\}_{i=1}^{\infty}$ so that

$$\mathcal{S}\mathbf{z}(u) = \sum_{i=1}^{\infty} \lambda_i \langle \mathbf{z}, \xi_i \rangle_{\mathbb{H}^L} \xi_i. \tag{26}$$

The eigenbasis is our functional principal component basis that also gives us our functional principal component scores of $\{c_{i,j}\}_{i=1}^{\infty} = \{\langle \xi_i, \mathbf{x}_j \rangle_{\mathbb{H}^L}\}_{i=1}^{\infty}$ for $j = 1, ..., K$. By [1], we define the rank one, elementary operator as $\mathcal{X}_i : \mathbb{R}^k \to \mathbb{H}^L$ where

$$\mathcal{X}_i = (\xi_i \otimes \xi_i) \sum_{j=1}^{K} a_j \mathbf{x}_j. \tag{27}$$

Proposition 3.3 [1] guarantees that each $\mathcal{X}_i$ is linear, bounded, of rank one, and that the following equality holds

$$\mathcal{X} = \sum_{i=1}^{\infty} \mathcal{X}_i \tag{28}$$

As such, in the grouping stage we group these rank one operators.

3. Grouping: We follow similar grouping techniques as those in the grouping step of subsection 2.3 to obtain $m = 1, ..., M$ separated trajectory operators $\hat{\mathcal{X}}_m : \mathbb{R}^k \to \mathbb{H}^L$ such that

$$\mathcal{X} = \sum_{m=1}^{M} \hat{\mathcal{X}}_m. \tag{29}$$

where

$$\hat{\mathcal{X}}_m = \sum_{j} \mathcal{X}_j, \text{ for some } j \in \mathbb{N}. \tag{30}$$

4. Reconstruction: To finish, we reconstruct each $\hat{X}_m$ into corresponding additive functional time series components, $\tilde{\mathrm{F}}_N^{(m)}$ for $m = 1, ..., M$. We define the Hankelization operator $\mathbf{H} : \mathbb{H}^{L \times K} \to \mathbb{H}_H^{L \times K}$. Then $\tilde{\mathcal{X}}_m = \mathbf{H}\hat{\mathcal{X}}_m$ for $m = 1, ..., M$ are our reconstructed additive trajectory operators. This linear transformation is given by

$$\tilde{\mathbf{x}}_{i,j}^{(m)} = \frac{1}{n_s} \sum_{(k,l):k+l=s} \hat{\mathbf{x}}_{k,l}^{(m)} \tag{31}$$

where $s = i + j$ and n_s is the number of (l, k) pairs such that $l + k = s$ [1]. Applying $\tilde{\mathrm{F}}_N^{(m)} = \mathcal{T}^{-1}\tilde{\mathcal{X}}_m$ for $m = 1, ..., M$ gives us our additive functional time series components that sum to F_N.

As was the case in SSA, we measure how well we did in separating out the additive components in the grouping stage, so we have a measure of separability given by

$$\rho^{(w)}\left(\tilde{\mathrm{F}}_N^{(m_1)}, \tilde{\mathrm{F}}_N^{(m_2)}\right) = \frac{\langle \tilde{\mathrm{F}}_N^{(m_1)}, \tilde{\mathrm{F}}_N^{(m_2)} \rangle_w}{\sqrt{\langle \tilde{\mathrm{F}}_N^{(m_1)}, \tilde{\mathrm{F}}_N^{(m_1)} \rangle_w} \sqrt{\langle \tilde{\mathrm{F}}_N^{(m_2)}, \tilde{\mathrm{F}}_N^{(m_2)} \rangle_w}} \tag{32}$$

for $m_1, m_2 \in 1, ..., M$ where

$$\langle \tilde{\mathrm{F}}_N^{(m_1)}, \tilde{\mathrm{F}}_N^{(m_2)} \rangle_w = \sum_{i=1}^{N} w_i \langle \tilde{\mathrm{F}}_N^{(m_1)}, \tilde{\mathrm{F}}_N^{(m_2)} \rangle_{\mathbb{H}}. \tag{33}$$

The weights are given by $w_i = \min\{i, L, N - i + 1\}$ [1]. If $\rho^{(w)}\left(\tilde{\mathrm{F}}_N^{(m_1)}, \tilde{\mathrm{F}}_N^{(m_2)}\right) = 0$, then we call $\tilde{\mathrm{F}}_N^{(m_1)}$ and $\tilde{\mathrm{F}}_N^{(m_2)}$ w-orthogonal which is desirable.

2.5 Clustering Time Series

Clustering of time series data can be viewed in two forms. The first involves whole clustering where if we have a set of K time series $\{y_N^{(k)}\}_{k=1}^K$ from possibly different processes the goal is to cluster these time series together based on likeness over their time periods [12]. The second form is known as subsequence clustering where one uses a sliding window to try to find similar or different behavior over time periods for one time series [12]. The focus of our work is on the latter with the goal. The way we measure similarity of time series is either through the shape level or the structure level where shape level clustering is used for short timer series. Shape-based methods tend to perform poorly on long time series and so we turn to structure-based methods which extract underlying features of the time series such as trend and seasonality.

Work has been done in structure level clustering of long time series by using self-organizing map (SOM) which reduces the dimension of a high dimensional time series [12]. Another approach that has been used in structure-based clustering of time series data is that of the discrete Fourier transform which transforms the series from the temporal realm to the frequency realm and

reduces the dimension in the process [13]. We have also seen the comparison of other clustering techniques such as k-means and shape-based clustering methods applied to time series data to try to determine the periods of time-varying operations in complex energy systems [14]. Another approach involves estimating the spectral density functions for a collection of stationary time series that share features and clustering on those features [15]. The structure based methods all revolve around some sort of feature selection or feature extraction where clustering is performed using those features and as such, PCA seems applicable. The issue is that PCA is not an effective structure level clustering technique for time series because it assumes independence of data from time [16].

FSSA is the answer to the need for a subsequence structure-based dimension reduction technique for time-dependent functional data that can be coupled with FPCA for clustering purposes. Our goal is to show that the performance in the FSSA-FPCA hybrid clustering method depends on how the grouping step in FSSA is performed. We execute the hybrid method by first separating a functional time series into its additive features by using FSSA. We then perform FPCA on those additive elements and cluster the resulting functional principal component scores (PCS) within each additive component individually. Preliminary work has been done in an earlier version of [1] which uses this hybrid clustering method for time-dependent daily call-center data. They observed better separation of daily call-center behavior with the FSSA-FPCA hybrid method than with straight application of FPCA to the same time series as shown in the following figure.

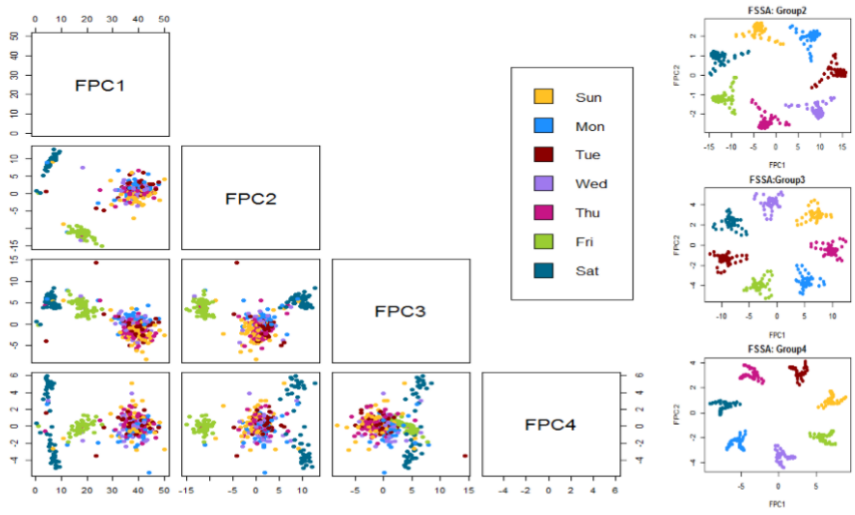

Figure 1: Direct FPCA vs. Hybrid FSSA-FPCA Clustering [Haghbin; submitted 2019]

We graphically see better separation between daily call-center behavior and there is a need for metrics that can be used to measure the quality of clustering using the FSSA-FPCA hybrid method as we change how we perform step three of FSSA.

3 Research Problem

In the grouping step of FSSA, we choose how to group functional principal components together for reconstruction purposes. Metrics used to measure quality of separation of reconstructions and clustering will depend on how this grouping is done and so we break our goal into the following pieces:

1. Recreate the results from the earlier version of [1] using the call-center data
2. Study the connections between the quality of groupings and the w-correlation as well as the clustering results
3. Show the connection between the w-correlation score between two reconstructions from FSSA and clustering results of FPCA scores from those same groupings using the call-center example

4 Methodology

The idea here is to treat the grouping of functional principal components in the third step of FSSA as our independent variable and our metrics of clustering quality and w-correlation as our dependent variables.

4.1 FSSA-FPCA on Call-Center Data Background

The steps to recreate the results shown in Figure 1 are as follows:

1. We partition the daily call-center data into 24 hour time periods which results in 365 days of call-center daily behavior. To this end, we cast the original time series, $y_{87,600}$, into a data matrix, X, which is 240 by 365 where each column contains 240 discrete samples of a continuous function $f_i \in \mathcal{L}^2([0,1])$. We have 240 discrete samples since each record in our time series is a the number of calls received in a 6-minute time interval. For scaling purposes, we scale down by taking the square root of each observation.

2. We use the “fda" package [17] , to estimate each daily function to obtain our functional time series $\mathrm{F}_N = \{f_i\}_{i=1}^N$ where $N = 365$. The following figure shows a plot of these functions.

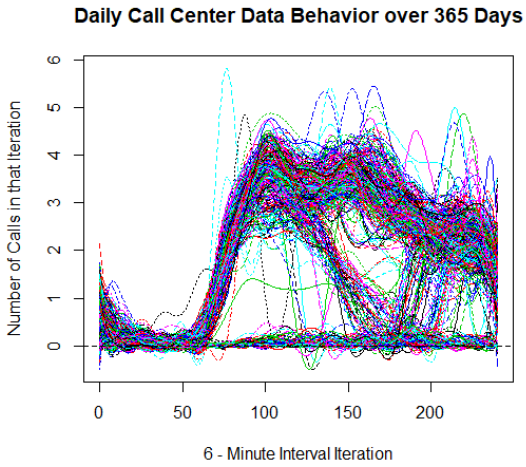

Figure 2: Elements of F_N

3. Next, we apply FSSA to F_N with $L = 28$ to obtain our truncated set of orthonormal functional principal components $\{\xi_i\}_{i=1}^{50}$ and corresponding eigenvalues $\{\lambda_i\}_{i=1}^{50}$ of $\mathcal{S}$ which will help us decide on groupings.

4. We achieve our groupings by using the following respective plots of $\{\sqrt{\lambda_i}\}_{i=1}^{50}$ and paired plots of the eigenfunctions.

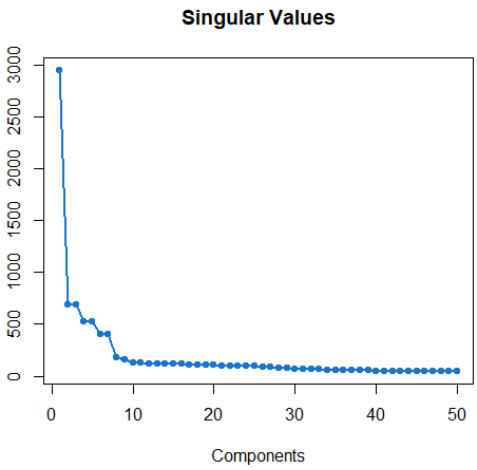

Figure 3: The Singular Values of $\mathcal{X}$

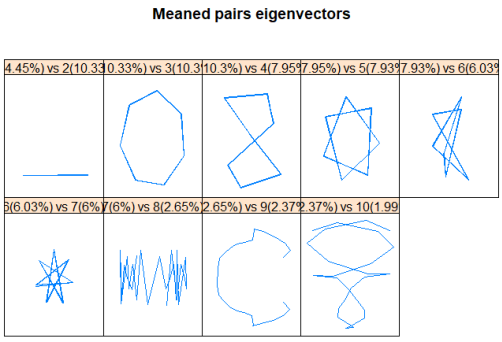

Figure 4: Mean Paired Plot of Eigenfunctions

Due to the alike intensities shown in the singular values and the geometry in the pair plots, we group the first functional principal component for reconstruction by itself, the second with the third, the fourth with the fifth, the sixth with the seventh, and all other $\{\xi_i\}_{i=8}^{50}$ components together as noise.

5. Using these groupings and the "FSSA" package [18], we reconstruct a set of five functional time series, $\tilde{\mathrm{F}}_N^{(i)},\ i = 1, ..., 5$ given in the following table.

Table 1: FSSA Reconstructions and Components Used to Create Reconstruction

	FSSA Components
$\tilde{\mathrm{F}}_N^{(1)}$	ξ_1
$\tilde{\mathrm{F}}_N^{(2)}$	$\{\xi_2, \xi_3\}$
$\tilde{\mathrm{F}}_N^{(3)}$	$\{\xi_4, \xi_5\}$
$\tilde{\mathrm{F}}_N^{(4)}$	$\{\xi_6, \xi_7\}$
$\tilde{\mathrm{F}}_N^{(5)}$	$\{\xi_i\}_{i=8}^{50}$

The following plots show the elements of each functional time series reconstruction.

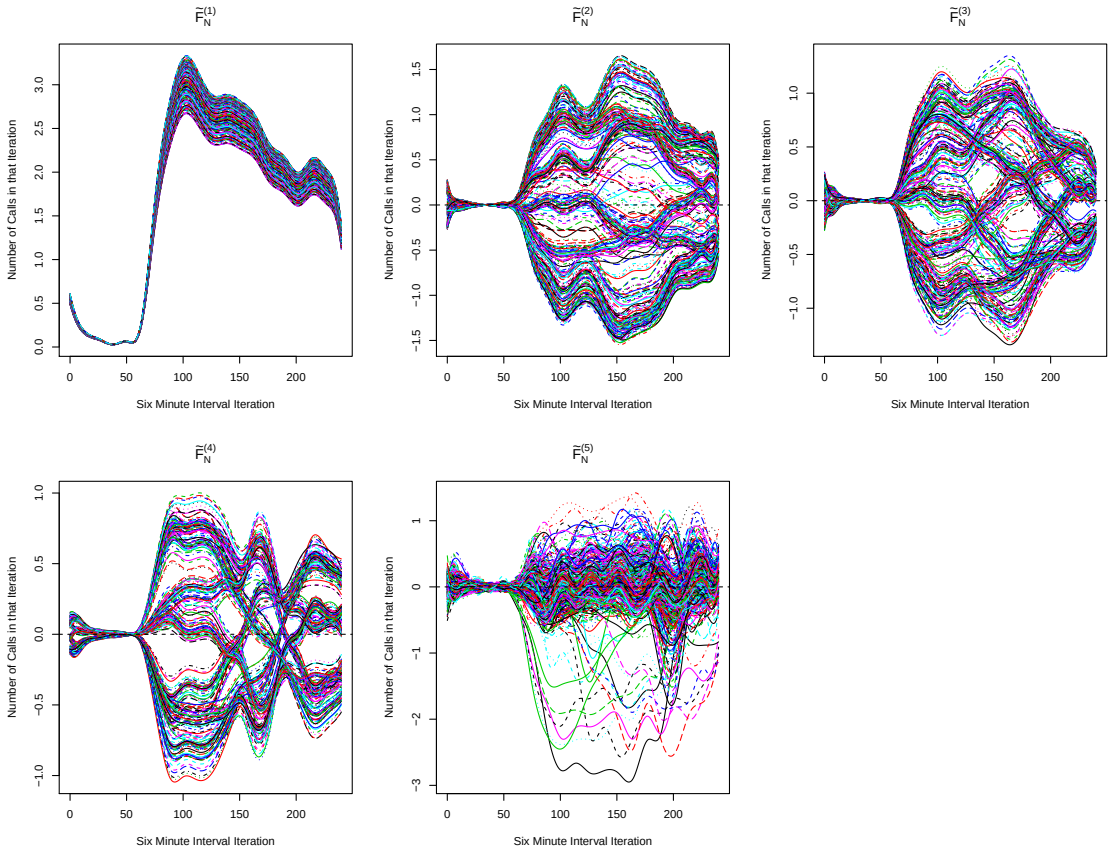

Figure 5: Plots of Additive Elements of F_N

We then apply FPCA to $\tilde{F}_N^{(i)}$ for $i = 2, 3, 4$ since these groupings are the only ones that capture daily behavior in a week as shown in Figure 4.

6. We cluster the subjects based on their resulting first and second FPC scores for each FSSA reconstruction as according to results shown in 1. To achieve this, we apply k-means clustering which searches for each cluster's centroid given a preset number of clusters to search for [4]. We apply this particular algorithm here since we expect to see 7 clusters representative of different call-center behavior on each of the 7 days in a week. This two-step approach of FPCA then clustering has also been used in clustering functional principal component scores as according to [19]. K-means also provides us a measure of how well the clustering was done defined by

$$F = \frac{\text{SS}_{\text{between}}}{\text{TSS}_{\text{within}}} \tag{34}$$

where $\text{SS}_{\text{between}}$ is the sum of square error between clusters and $\text{TSS}_{\text{within}}$ is the total sum

of square error within clusters. As is the case with F-statistics, the goal is to maximize (34). We also employ another measurement known as the Rand index as a measure of how well clustering was performed by comparing to the known true groupings [20]. We let $X = \{X_1, ..., X_7\}$ be seven groupings we obtain from using k-means and we let $Y = \{Y_1, ..., Y_7\}$ be the seven true groupings. Let a be the number of pairs in the same X_i that have corresponding matching pairs in some Y_i and let b be the number of pairs different in some X_i that have corresponding non-matching pairs in some Y_i. The Rand index that compares the clustering of k-means and the true groupings is given by

$$\text{RI} = \frac{a+b}{\binom{n}{2}} \tag{35}$$

where a Rand index close to one is desired. We compare the clustering of subjects using k-means on the functional principal component scores for an FSSA reconstruction and the true groupings and we compare the performance of k-means on the functional principal component scores of F_N versus the true groupings in the following figure.

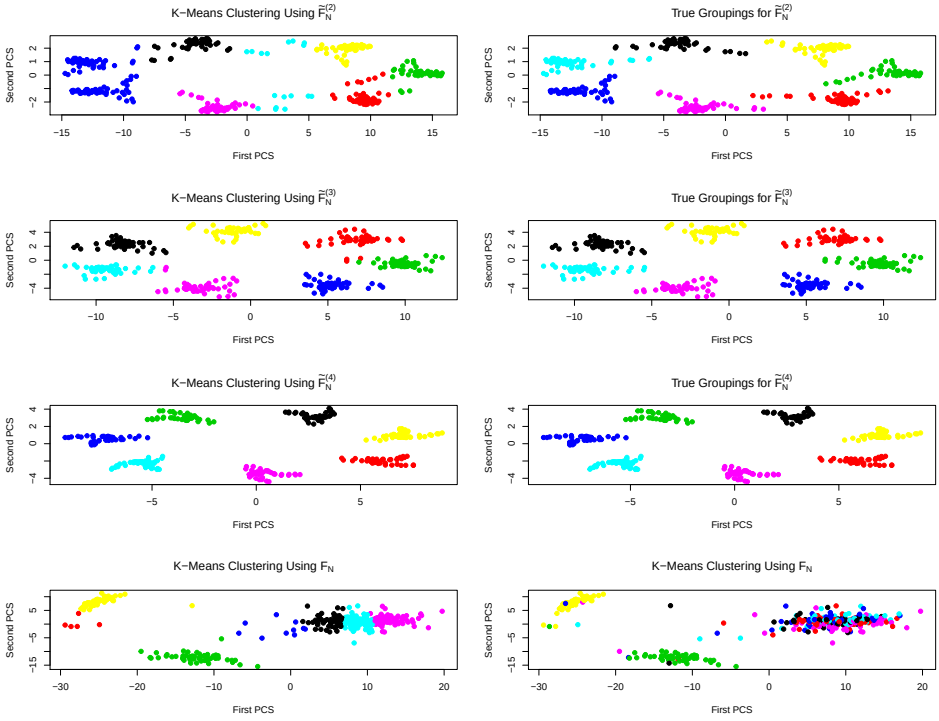

Figure 6: K-Means Clustering using FPCA Scores vs. True Groupings

The following table shows the resulting F-ratio value and Rand index for k-means clustering of FPC scores using reconstructions $\tilde{\text{F}}_N^{(2)}$, $\tilde{\text{F}}_N^{(3)}$, $\tilde{\text{F}}_N^{(4)}$, and FPCA directly applied to F_N.

Table 2: Clustering Results For K-Means

	$\tilde{\mathrm{F}}_N^{(2)}$	$\tilde{\mathrm{F}}_N^{(3)}$	$\tilde{\mathrm{F}}_N^{(4)}$	F_N
F-Ratio	42.10	34.43	59.30	33.88
Rand Index	0.93	0.992	1	0.798

As reflected in these results, k-means clustering performed best on the fourth reconstruction due to the fact that it is a locally based method that performs best when there is large separation between clusters [4].

4.2 Noise Groupings vs. F-Ratio, w-Correlation, and the Rand Index

We use new reconstructions that include the FSSA components of $\tilde{\mathrm{F}}_N^{(4)}$ to computationally build a relationship between the grouping done in the third step of FSSA and the F-ratio, w-correlation, and Rand index. We choose $\tilde{\mathrm{F}}_N^{(4)}$ as the reconstruction to apply our computational techniques to since k-means performed best on its FPCA scores.

Recall we used $\{\xi_6, \xi_7\}$ to build $\tilde{\mathrm{F}}_N^{(4)}$ and $\{\xi_i\}_{i=8}^{50}$ to build $\tilde{\mathrm{F}}_N^{(5)}$. We expect that if we use functional principal component terms associated with noise and principal component terms $\{\xi_6, \xi_7\}$ to reconstruct a new functional time series, then clustering results will suffer due to the creation of a more noisy reconstruction. To this end we build a new reconstruction $\tilde{\mathrm{F}}_N^{(p)}$ made from FSSA components $\{\xi_i\}_{i=6}^{p_1}$ with $p_1 = 7, ..., 50$. We expect that the F-ratio will decrease, the w-correlation between $\tilde{\mathrm{F}}_N^{(p_1)}$ and $\tilde{\mathrm{F}}_N^{(5)}$ will increase, and the Rand index will decrease. Our proposed technique to estimate a model for these relationships is as follows:

1. Use FSSA principal components $\{\xi_i\}_{i=6}^{p}$ to reconstruct $\tilde{\mathrm{F}}_N^{(p)}$
2. For the F-ratio, we apply FPCA to $\tilde{\mathrm{F}}_N^{(p)}$ and cluster resulting FPCA scores using k-means
3. For the w-correlation score, we just directly measure $\rho^{(w)}\left(\tilde{\mathrm{F}}_N^{(p)}, \tilde{\mathrm{F}}_N^{(5)}\right)$
4. We measure the Rand index between the k-means clustering of FPCA scores from the decomposition of $\tilde{\mathrm{F}}_N^{(p_1)}$ and the true groupings
5. Repeat steps 1-3 for $p_1 = 7, ..., 49$

4.3 Non-Noise Groupings vs. the Rand Index

We investigate how the Rand index changes as we include non-noise components in our reconstruction of the original time series. We expect that clustering results will be worse because as we add more components in, we will be approximating F_N and we will simply recreate the results shown in the direct FPCA method of Figure 1. We do not consider the F-ratio or w-correlation here because the k-means algorithm is local algorithm that will find false groupings which inflates F and the w-correlation will oscillate due to true FSSA non-noise groupings being orthogonal to one another. To this end, we employ the following steps.

1. Build reconstruction $\tilde{\mathrm{F}}_N^{(p_2)}$ using $\{\xi_6, \xi_7\} \cup \{\xi_i\}_{i=1}^{p_2}$.
2. Measure the quality of the k-means clustering of the FPCA functional principal component scores from the FSSA reconstruction $\tilde{\mathrm{F}}_N^{(p_2)}$ as compared to the true groupings using the Rand index
3. Repeat steps 1-2 for $p_2 = 1, ...5$

5 Results

In general, we expect to see worse clustering performance, reflected in our metrics, as we use improper groupings in our FSSA reconstructions.

5.1 Results for Groupings vs. the F-Ratio, w-Correlation, and Rand Index

Using the technique covered in subsection 4.2, we obtain the following plot of the number of noise components included in the grouping versus the F-ratio.

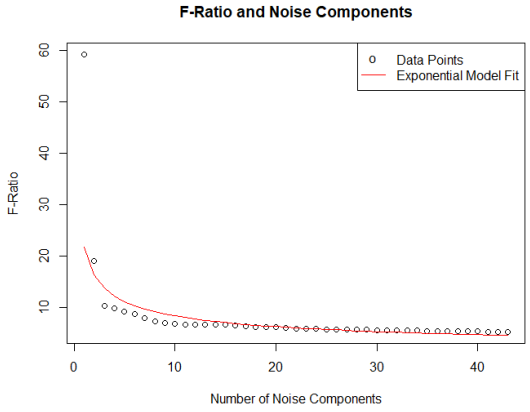

Figure 7: Plot of F-Ratio as More Noise is Added

As expected, increasing the number of noise components in the reconstruction of $\tilde{\mathrm{F}}_N^{(p_1)}$ decreases the F-ratio. We fit an exponential model to the data points in Figure 7 to estimate

$$F(k) = e^{\beta_0 + \beta_1 k} \text{ for } k \in \mathbb{N} \tag{36}$$

where F is the F-ratio as a function of k noise components included in the reconstruction. We also obtain the following summary table which gives the results of this exponential fit.

Table 3: Exponential Model Results for F-ratio vs. Noise

	Estimate	Standard Error	t-value	p-value
β_0	3.081	0.105	29.46	2×10^{-16}
β_1	-0.415	0.035	-11.72	1.13×10^{-14}

As we see from Table 5.1, we fail to reject that null hypothesis that there is no exponential relationship between F and k and as such we believe there is an exponential relationship present.

Using the same technique, we obtain the following plot of number of noise components included in the reconstruction vs. the w-correlation between $\tilde{\mathrm{F}}_N^{(p_1)}$ and $\tilde{\mathrm{F}}_N^{(5)}$.

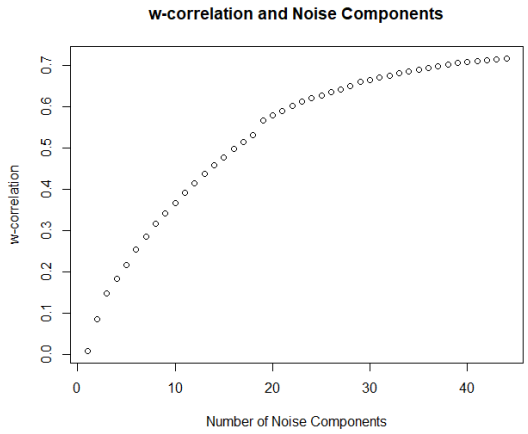

Figure 8: Plot of w-Correlation between $\tilde{\mathrm{F}}_N^{(p_1)}$ and $\tilde{\mathrm{F}}_N^{(5)}$

As we expected, we get worse separation between $\tilde{\mathrm{F}}_N^{(p_1)}$ and $\tilde{\mathrm{F}}_N^{(5)}$ as p_1 gets larger.

The following is a visual of how the Rand index changes as we increase the number of noise components included in the reconstruction of $\mathrm{F}_N^{(p_1)}$.

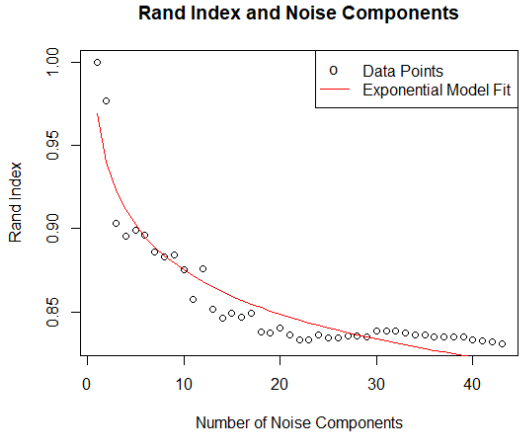

Figure 9: Plot of the Rand Index as More Noise is Added

Similar to the F-ratio, there appears to be an exponential relationship between the Rand index and the number of noise components included in the reconstruction of $\tilde{\mathrm{F}}_N^{(p_1)}$ and so we fit the model given in the right-hand side of equation (36) and obtain the following table summarizing the fit.

Table 4: Exponential Model Results for Rand Index vs. Noise

	Estimate	Standard Error	t-value	p-value
β_0	-0.031	0.007	-4.487	5.75×10^{-5}
β_1	-0.044	0.002	-18.69	2×10^{-16}

We see that there does seem to be an exponential relationship between the Rand index and the number of noise components included in the reconstruction.

Notice that our hypothesis that as the w-correlation increases, the F-ratio decreases is computationally confirmed as illustrated in the following figure.

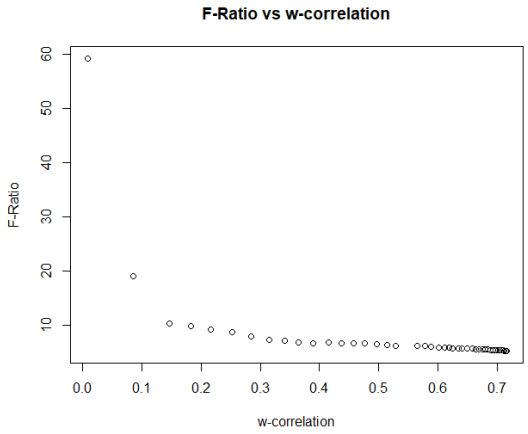

Figure 10: F-Ratio vs. w-Correlation

5.2 Results for Rand Index vs. Non-Noise Components

We expect that as the number of non-noise components we include in our reconstruction of F_N that the clustering results will suffer simply because as more components are added, we will be regenerating the results seen in Figure 1. Clustering using different FPC scores of a reconstruction will also change the F-ratio and Rand index since we may capture seasonal behaviors in different pairwise components. For the purpose of deciding which FPCA functional principal components should be used in clustering for each reconstruction, we calculate the percentage of variability explained by each component for each reconstruction in the following table.

Table 5: Percent of Variability Explained by Each FPC for each Reconstruction $\tilde{\mathrm{F}}_N^{(p_2)}$

	Reconstruction $\{\xi_1, \xi_6, \xi_7\}$	Reconstruction $\{\xi_1, \xi_2, \xi_6, \xi_7\}$	Reconstruction $\{\xi_1, \xi_2, \xi_3, \xi_6, \xi_7\}$	Reconstruction $\{\xi_1, \xi_2, \xi_3, \xi_4, \xi_6, \xi_7\}$	Reconstruction $\{\xi\}_{i=1}^{7}$
FPC 1	79.00%	73.28%	83.62%	83.54%	83.55%
FPC 2	19.73%	25.27%	15.60%	15.69%	15.73%
FPC 3	1.26%	1.35%	0.70%	0.67%	0.59%
FPC 4	9×10^{-4}%	0.07%	0.05%	0.05%	0.05%

As according to Table 5, most of the variability in the data is explained in the first three components with a little present in the fourth. For graphical investigative purposes, we plot k-means clustering results using the first two FPCA functional principal components that arise from varying FSSA reconstructions of $\tilde{\mathrm{F}}_N^{(p_2)}$ and we plot the true groupings.

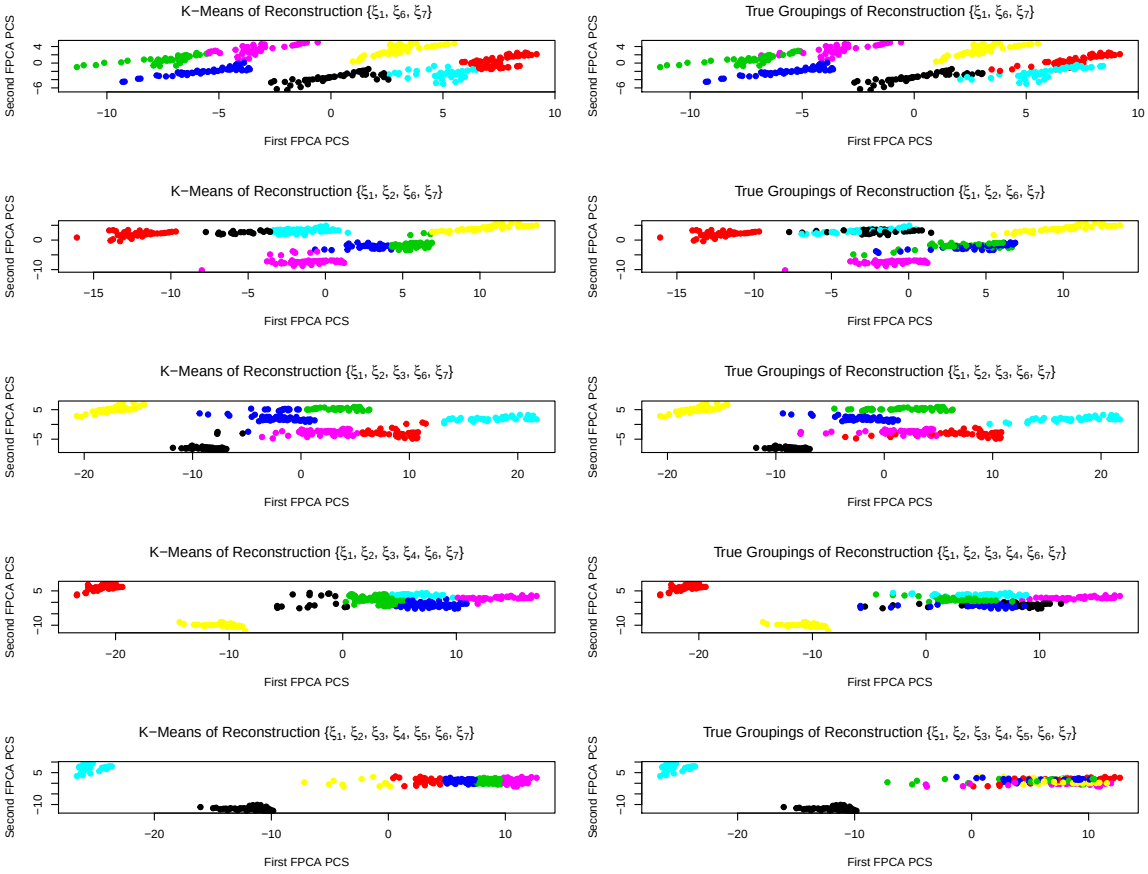

Figure 11: Clustering FPCA Scores with Various FSSA Reconstructions of $\tilde{\mathrm{F}}_N^{(p_2)}$

As we graphically see, k-means clustering results are suffering as we include more non-noise components in our reconstruction. We use the Rand index to measure the quality of clustering for each reconstruction using combinations of the first four FPC scores and we obtain the following table.

Table 6: Rand Index from Clustering Based on Varying Reconstructions and FPC's

	Reconstruction $\{\xi_1, \xi_6, \xi_7\}$	Reconstruction $\{\xi_1, \xi_2, \xi_6, \xi_7\}$	Reconstruction $\{\xi_1, \xi_2, \xi_3, \xi_6, \xi_7\}$	Reconstruction $\{\xi_1, \xi_2, \xi_3, \xi_4, \xi_6, \xi_7\}$	Reconstruction $\{\xi\}_{i=1}^{7}$
FPC $\{1, 2\}$	0.9493	0.9107	0.9606	0.8985	0.8276
FPC $\{1, 3\}$	0.9490	0.9137	0.9618	0.9055	0.8484
FPC $\{1, 4\}$	0.9490	0.9140	0.9643	0.9055	0.8484
FPC $\{2, 3\}$	0.8299	0.9312	0.9471	0.9573	0.8874
FPC $\{2, 4\}$	0.8299	0.9312	0.9487	0.9591	0.9111
FPC $\{3, 4\}$	0.7508	0.7779	0.8662	0.8860	0.8551

We see that none of the reconstructions we used have a Rand index of one and therefore, the best clustering results appear when we reconstruct $\tilde{\mathrm{F}}_N^{(p_2)}$ with components ξ_6 and ξ_7. Recall that the first FSSA principal component captures mean behavior and not weekly behavior as shown in Figure 5. Including this component in our reconstructions for this example weakens our clustering results by decreasing the locality of clusters as seen in the following figure.

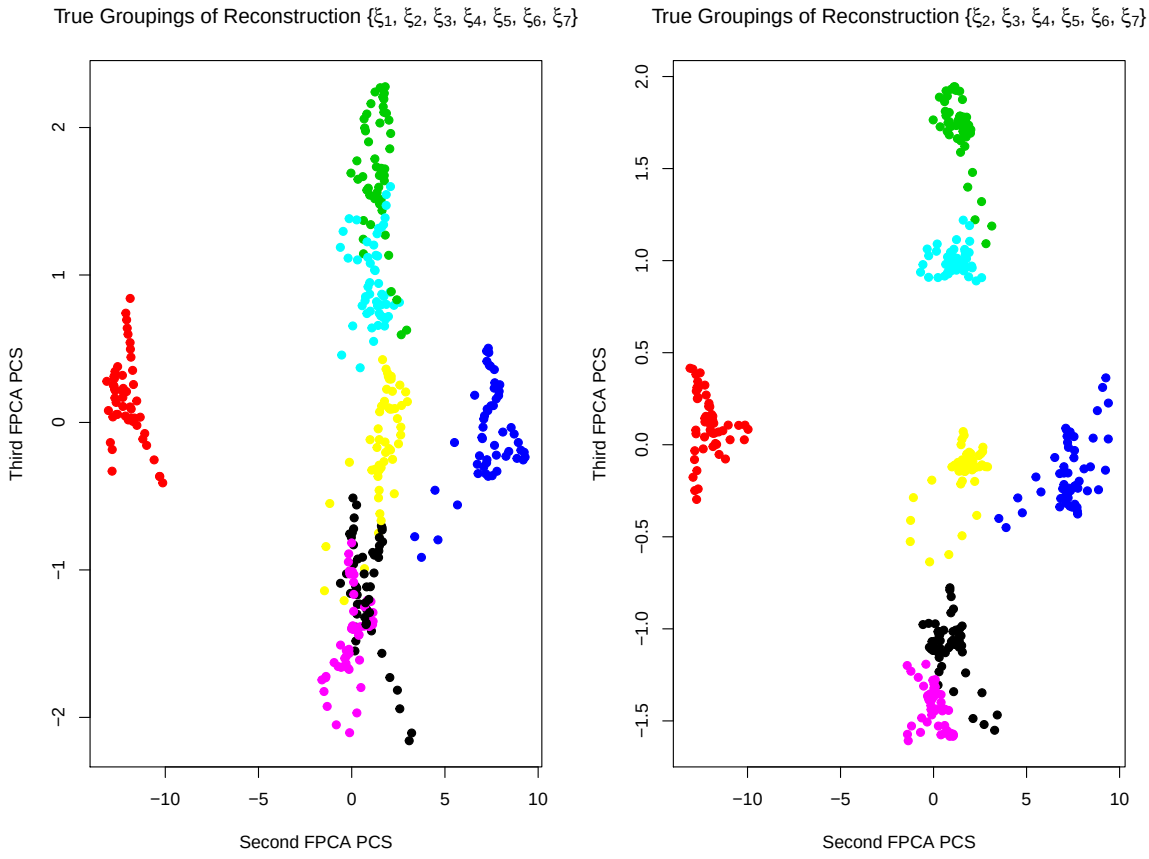

Figure 12: Comparing Reconstructions

The reason for this poor performance in clustering is due to the fact that more energy is introduced into the reconstruction causing a wider spread in the data. We take the first component completely out of reconstructions to try to improve on locality issues however, we find that

k-means performs even worse simply because k-means searches for spherical clusters as shown in the following figure.

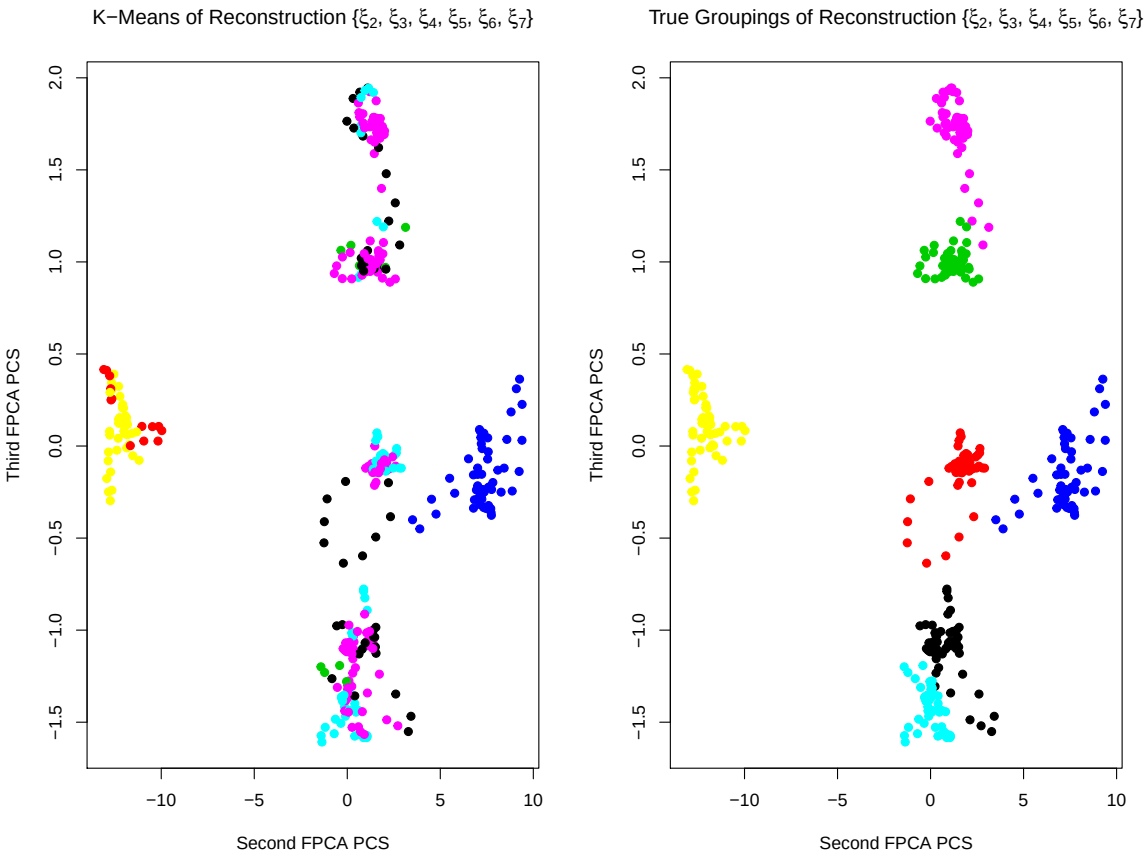

Figure 13: K-Means vs. True Groupings First Component Left Out

6 Discussion

As we see, from section 5, if we execute the grouping step of FSSA poorly, we obtain worse clustering results.

6.1 Sensitivity Analysis

In order to estimate by how much the F-ratio will change as we add another noise functional principal component into our reconstruction of $\tilde{F}_N^{(p)}$, we employ

$$S_i = \frac{|\Delta F_i|}{|\Delta k|} = |\Delta F_i|. \tag{37}$$

We consider measuring the discrete sensitivity of the results shown in Figure 7 and we obtain

the following Figure

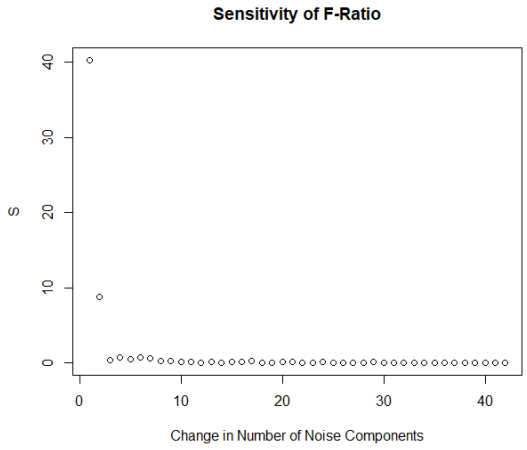

Figure 14: Sensitivity of F-Ratio with Added Components

Recall that the first data point corresponds to the F-ratio when no noise terms are added into the reconstruction of $\tilde{\mathrm{F}}_N^{(p_1)}$. As we see, our measure of the F-ratio is not very sensitive to changes in the number of noise components past the addition of the first noise component. We expect similar sensitivity behavior in the w-correlation between $\tilde{\mathrm{F}}_N^{(p_1)}$ and $\tilde{F}_N^{(5)}$ and the Rand index since these measurements do not change much for larger amounts of noise components as shown in Figures 8 and 9.

For the sensitivity of the Rand index as compared to the grouping and FPCA components used for the k-means clustering, we employ a method where we measure the change in the Rand index across the rows of Table 6. Labeling the first column of Table 6 reconstruction 1, the second as reconstruction 2, the third as reconstruction 3, the fourth as reconstruction 4, and the fifth as reconstruction 5, we obtain the following table of sensitivities and average sensitivities of the Rand index as we change the reconstruction of $\tilde{F}_N^{(p_2)}$.

Table 7: Sensitivity of Rand Index

	Reconstruction 1 to 2	Reconstruction 2 to 3	Reconstruction 3 to 4	Reconstruction 4 to 5	Average Sensitivity
FPC $\{1,2\}$	0.0386	0.0499	0.0621	0.0709	0.0554
FPC $\{1,3\}$	0.0353	0.0481	0.0563	0.0571	0.0492
FPC $\{1,4\}$	0.0350	0.0503	0.0588	0.0580	0.5053
FPC $\{2,3\}$	0.1014	0.0158	0.0102	0.0688	0.0491
FPC $\{2,4\}$	0.1014	0.0174	0.0104	0.0480	0.0443
FPC $\{3,4\}$	0.0271	0.0884	0.0200	0.0301	0.0413

We see all of our average sensitivities are below 0.10 thus our Rand index measurement is not very sensitive to changes in the groupings.

6.2 Future Work

There are a lot of avenues to explore further with this work. The first being that of using a different clustering algorithm to try to avoid the issue of locality in k-means. One of these other methods we can apply is hierarchical clustering which will allow us the flexibility of not selecting a set number of clusters to search for. This can be very helpful in clustering unlabeled data. This leads into another avenue of work where we apply this same analysis to other functional time series data where we do not know the proper labels. We can also experiment with clustering based off of the FSSA principal component scores themselves. Even though our reconstructions are additive functional time series components, we see that FPCA performed on these reconstructions gives us good clustering results. As such, our reconstructions seem to not be dependent on time and FSSA appears to be behaving as a temporal to frequency domain transformation similar to that of a Fourier analysis which needs to be expanded upon.

7 Conclusion

In this work we first covered the literature review necessary to do research using the new FSSA algorithm presented in [1]. One particular area FSSA has seen some use is in the realm of subsequence clustering on a structural level where it was shown in [1] that the novel FSSA-FPCA algorithm has done a better job in clustering FPCA scores of reconstructed time series rather than FPCA applied to the original functional time series. The present work extended this example further and shows that the quality of clustering in the new hybrid algorithm depends on how well the grouping of functional principal components for reconstruction is done in the third step of FSSA. We also showed that there's a connection between the w-correlation score between reconstructions of the original time series and the clustering results using k-means. Using this work, we have been able to obtain the clues to pursue theoretical work in this field to further mature the idea of FSSA.

8 Bibliography

References

[1] Haghbin, H., Najibi, S.M., Mahmoudvand, R., Maadooliat, M. (submitted 2019). Functional Singular Spectrum Analysis. *The Annals of Statistics*.

[2] Ramsay, J.O., Silverman, B.W. (1997). Functional Data Analysis. *Springer Series in Statistics*

[3] Shalit, O. (2017). A First Course in Functional Analysis. *CRC Press Taylor & Francis Group*.

[4] James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R. *Springer Texts in Statistics*, 10(3): 385-386.

[5] Reed, S., Simon, B. (1972). Methods of Modern Mathematical Physics, Volume I: Functional Analysis. *Academic Press, Inc.*.

[6] Zemyan, S. (2012). The Classical Theory of Integral Equation. *Birkhäuser*.

[7] Huang, J.Z., Shen, H., Buja, A. (2008). Functional principal components analysis via penalized rank one approximation. *Election Journal of Statistics Vol.2*.

[8] Chiou, J.M., Chen, Y.T., Yang, Y.F., (2014). Multivariate Functional Principal Component Analysis: A Normalization Approach. *Statistica Sinica 24*, on pages 1571-1596.

[9] Golyandina, N., Nekrutkin, V., Zhigljavsky, A.A. (2001). Analysis of Time Series Structure SSA and Related Techniques. *Chapman & Hall/CRC*.

[10] Golyandina, N., Stepanov, D. (2005). SSA-based approaches to analysis and forecast of multidimensional time series. In: *Proceedings of the 5th St. Petersburg Workshop on Simulation*, June 26-July 2, 2005, St. Petersburg State University, St. Petersburg, on pages 293-298.

[11] Groth, A., Ghil, M. (2011). Multivariate singular spectrum analysis and the road to phase synchronization. *Physical Review E, Statistical, nonlinear, and soft matter physics*.

[12] Wang, X., Smith, K., Hyndman, R. (2006). Characteristic-Based Clustering for Time Series Data. *Data Mining and Knowledge Discovery*, on pages 335-364.

[13] Roelofsen, P. (2018). Time series clustering. *Master's Essay*. Vrije Universiteit, Amsterdam, Netherlands

[14] Teichgraeber, H., Brandt, A. (2019). Clustering methods to find representative periods for the optimization of energy systems: an initial framework and comparison. *Applied Energy*, Volume 239, on pages 1283-1293.

[15] Maadooliat, M., Sun, Y., Chen, T. (2018). Nonparametric collective spectral density estimation with an application to clustering the brain signals. *Statistics in Medicine*, Volume 37, Issue 30.

[16] Li, L., Prakash, B. (2011). Time Series Clustering: Complex is Simpler!. In: *Proceedings of the 28th International Conference on Machine Learning*, June 28- July 2, 2011, Bellevue, Washington, on pages 185-192.

[17] Ramsay, J., Wickham, H., Graves, S., Hooker, G. (2018). fda: Functional Data Analysis. *R package version 2.4.8*. https://CRAN.R-project.org/package=fda

[18] Hossein, H., Najibi, S. (2019). Rfssa: Functional Singular Spectrum Analysis. *R package version 0.0.1*. https://CRAN.R-project.org/package=Rfssa

[19] Jacques, J., Preda, C. (2014). Functional data clustering: a survey. Advances in Data Analysis and Classification, *Springer Verlag*

[20] Rand, W. (2018). Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association*, Volume 66, Issue 336, on pages 846-850.