T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

FONKSİYONEL TEMEL BİLEŞEN ANALİZİ

Sümeyye İNAL

YÜKSEK LİSANS TEZİ

İstatistik Anabilim Dalı

İstatistik Programı

Danışman

Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK

Haziran, 2020

T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

FONKSİYONEL TEMEL BİLEŞEN ANALİZİ

Sümeyye İNAL tarafından hazırlanan tez çalışması 26.06.2020 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, İstatistik Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK, Danışman
Yıldız Teknik Üniversitesi _____

Prof. Dr. Fatma NOYAN TEKELİ, Üye
Yıldız Teknik Üniversitesi _____

Doç. Dr. Seher ARIKAN TEZERGİL, Üye
Marmara Üniversitesi _____

Danışmanım Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK sorumluluğunda tarafımca hazırlanan Fonksiyonel Temel Bileşen Analizi başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Sümeyye İNAL

İmza

Aileme, biricik eşime

ve

oğlum Mahir'e

TEŞEKKÜR

Tez çalışmamda bana her türlü kolaylığı sağlayan, bilgi ve tecrübeleriyle beni destekleyen, bu çalışmanın planlanmasında, araştırılmasında ve yürütülmesinde her anlamda ilgi ve desteğini esirgemeyen çok değerli hocam Sayın Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK ' e en içten dileklerimle teşekkür eder saygılarımı sunarım.

Ayrıca çalışmalarım sırasında her daim yanımda olan ve beni destekleyen çok sevgili eşim Osman ve oğlum Mahir'e sonsuz teşekkürlerimi sunarım.

Sümeyye İNAL

İÇİNDEKİLER

SİMGE LİSTESİ vi

KISALTMA LİSTESİ viii

ŞEKİL LİSTESİ ix

TABLO LİSTESİ xi

ÖZET xii

ABSTRACT xiv

1 GİRİŞ 1

1.1 Literatür Özeti ... 1

1.2 Tezin Amacı ... 7

1.3 Hipotez ... 7

1.4 Temel Bileşenler Analizi ... 8

2 FONKSİYONEL VERİ ANALİZİ 13

2.1 Fonksiyonel Veriler İçin Tanımlayıcı İstatistikler ... 15

2.2 Fonksiyonların Baz Fonksiyonlar ile Gösterimi ... 16

2.3 Fonksiyonel Verileri Düzgünleştirme ... 21

3 FONKSİYONEL TEMEL BİLEŞEN ANALİZİ 29

3.1 Fonksiyonel Temel Bileşen Analizi (FTBA) ... 30

3.2 Düzgünleştirilmiş Fonksiyonel Temel Bileşen Analizi ... 36

4 UYGULAMA 44

4.1 Uygulama 1: Elektrik Tüketimi ... 44

4.2 Uygulama 2: Covid-19 Pandemisi ... 68

5 SONUÇ VE ÖNERİLER 83

KAYNAKÇA 85

TEZDEN ÜRETİLMİŞ YAYINLAR 90

SİMGE LİSTESİ

α	Alfa
B	Beta
β	Beta
I	Birim matris
X	Chi
δ	Delta
D	Diferansiyel
ε	Epsilon
H	Eta
$\mathcal{H}$	Hilbert uzayı
c	Katsayılar vektörü
K	Kappa
κ	Kappa
V	Kovaryans operatörü
ξ	Xi(ksi)
ψ	Ksi
λ	Lambda
$\mathcal{L}^p$	Lebesque integrallenebilir uzay
μ	Mü
Ω	Omega
ω	Omega
u	Özvektör
φ	Phi
ϕ	Phi
Ψ	Psi
π	Pi
P	Ro
ρ	Ro
Σ	Sigma
σ	Sigma
S	Simetrik Şapka matrisi

T	Tau
τ	Tau
θ	Teta
Z	Zeta
ζ	Zeta

KISALTMA LİSTESİ

CV	Çarpraz Geçerlilik Ölçütü (Cross Validation)
COVID-19	Yeni Koronavirüs Hastalığı (SARS-CoV-2 Enfeksiyonu)
ÇFTBA	Çok Değişkenli Fonksiyonel Temel Bileşenler Analizi
EEG	Elektroensefalografi (Beyin Çizgesi Yöntemi)
EPDK	Enerji Piyasası Düzenleme Kurumu
FPCA	Functional Principal Components Analysis
FTBA	Fonksiyonel Temel Bileşenler Analizi
FVA	Fonksiyonel Veri Analizi
GCV	Genelleştirilmiş Çarpraz Geçerlilik Ölçütü (Generalization CV)
MSE	Ortalama Hata Kareler
PCA	Principal Components Analysis
PENSSE	Cezalı Artık Kareler Toplamı
SSE	Ortalama Hata Kareler Toplamı
TBA	Temel Bileşenler Analizi
TBAPSV	Cezalı Temel Bileşen Örneklem Varyansı
TEDAŞ	Türkiye Elektrik Dağıtım A.Ş.
TEİAŞ	Türkiye Elektrik İletim A.Ş.
TÜİK	Türkiye İstatistik Kurumu
OWID	Our World In Data

ŞEKİL LİSTESİ

Şekil 4.1 2017 yılı elektrik tüketiminin tüketici türüne göre dağılımı(EPDK) 44
Şekil 4.2 Türkiye'nin yıllara göre kurulucu gücündeki değişim 45
Şekil 4.3 Türkiyedeki elektrik dağıtım bölgeleri haritası 45
Şekil 4.4 2013-2019 yıllarına ait illerin aylık ortalama elektrik tüketimi 49
Şekil 4.5 B-Spline baz fonksiyonlar 50
Şekil 4.6 Bireysel fonksiyonlar 51
Şekil 4.7 Farklı lambda değerlerine karşılık gelen GCV kriteri 52
Şekil 4.8 Farklı lambda değerlerine karşılık bireysel fonksiyonlar 53
Şekil 4.9 Birinci türev fonksiyonları 54
Şekil 4.10 Yıllara göre birinci türev fonksiyonları 54
Şekil 4.11 Birinci temel bileşen fonksiyonu 55
Şekil 4.12 İkinci temel bileşen fonksiyonu 56
Şekil 4.13 Üçüncü temel bileşen fonksiyonu 57
Şekil 4.14 Dördüncü temel bileşen fonksiyonu 58
Şekil 4.15 Temel bileşen skorlarının dağılımı 59
Şekil 4.16 Gümüşhane, Çanakkale, Balıkesir ve Yalova illerinin türev fonksiyonu 60
Şekil 4.17 İllerin aylık ortalama sıcaklığı 62
Şekil 4.18 Fourier baz fonksiyonları 63
Şekil 4.19 Sıcaklık verisi için bireysel fonksiyonlar 64
Şekil 4.20 Türev fonksiyonları 65
Şekil 4.21 Birinci temel bileşen fonksiyon 66
Şekil 4.22 Temel bileşen fonksiyonları skor dağılımları (sıcaklık verisi) 67
Şekil 4.23 Ülkelere göre covid-19 vaka sayılarının dağılımı 72
Şekil 4.24 Covid-19 vakaları görülen ülkelerin bireysel fonksiyonları 73
Şekil 4.25 Covid-19 türev fonksiyonları 74
Şekil 4.26 Covid-19 birinci temel bileşen fonksiyonu 75
Şekil 4.27 Covid-19 ikinci temel bileşen fonksiyonu 76
Şekil 4.28 Temel bileşen fonsiyonları skorlarının dağılımı (covid-19) 77
Şekil 4.29 Butan Krallığı bireysel fonksiyonu 77
Şekil 4.30 İran bireysel fonksiyonu 78

Şekil 4.31 Türkiye bireysel fonksiyonu 78
Şekil 4.32 Güney Kore bireysel fonksiyonu 79
Şekil 4.33 Almanya bireysel fonksiyonu 79
Şekil 4.34 Brezilya bireysel fonksiyonu 80
Şekil 4.35 İtalya bireysel fonksiyonu 80
Şekil 4.36 Amerika bireysel fonksiyonu 81

TABLO LİSTESİ

Tablo 4.1 Elektrik tüketim verileri için tanımlayıcı istatistikler 48
Tablo 4.2 Farklı lambda değerlerine karşılık GCV değerleri..... 52
Tablo 4.3 Temel bileşen fonksiyonlarının skor değerleri 58
Tablo 4.4 İllerin aylık ortalama sıcaklık verisi için tanımlayıcı istatistikler..... 61
Tablo 4.5 Temel bileşen fonksiyonu skor değerleri (sıcaklık verisi) 66
Tablo 4.6 Covid-19 koronavirüs salgının yayıldığı ülkeler..... 68
Tablo 4.7 Covid-19 vakası görülen ülkeler için tanımlayıcı istatistikler..... 69

ÖZET

Fonksiyonel Temel Bileşen Analizi

Sümeyye İNAL

İstatistik Anabilim Dalı

Yüksek Lisans Tezi

Danışman: Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK

Veri elde etme ve depolamada meydana gelen gelişmeler özellikle fonksiyonel verilerin varlığını büyük ölçüde artırmakta ve klasik istatiksel yöntemler bu verileri analiz etmede yetersiz kalmaktadır. Bu verileri analiz etmek için kullanılan yöntem ve teknikler Fonksiyonel Veri Analizi (FVA) ile adlandırılmaktadır. Fonksiyonel veri analizinde, ayrık veri noktaları Baz Fonksiyonlar ve Pürüzlü Ceza Yaklaşımıyla sürekli birer fonksiyona dönüştürüldüğünden, tek tek veri noktalarını incelemek yerine fonksiyonlar incelenir. Bu da daha hızlı ve daha zengin bir analiz imkanı sunduğundan FVA her geçen gün modern veri analizinde daha önemli bir yer tutmaktadır.

Bu araştırmanın amacı bir fonksiyonel veri analizi yöntemi olan Düzgünleştirilmiş Fonksiyonel Temel Bileşen Analizini (FTBA) ayrıntılı bir şekilde ele alarak iki farklı gerçek veri seti üzerinde farklı baz fonksiyonlarını kullanarak uygulamasını yapmaktır.

Bu çalışmada yapılan ilk uygulamada Türkiye deki illerin sıcaklık ve hane halkı elektrik tüketimi verileri incelenmiş ve Düzgünleştirilmiş FTBA ile iller arasındaki elektrik tüketimi ve sıcaklıklardaki değişkenlik yapısı ortaya konmuştur. Fonksiyonel temel bileşen skorları ve dağılım grafikleri, iller arasındaki temel farklılıkları görsel olarak sunmuştur ve sonuç olarak hava sıcaklığının elektrik

tüketimini artırdığı görülmüştür. İkinci uygulamada Covid-19 toplam vaka sayılarının ülkelere göre değişimi incelenmiş ve bireysel fonksiyonlar, türev fonksiyonları ve temel bileşen fonksiyonlarının incelenmesiyle vaka sayısındaki artışın ilk 15 günde maksimum seviyelere ulaştığı sonuçlarına varılmıştır.

Anahtar Kelimeler: Fonksiyonel veri analizi, pürüzlü ceza yöntemi, düzgünleştirilmiş fonksiyonel temel bileşen analizi, elektrik tüketimi, Covid-19

ABSTRACT

Functional Principal Component Analysis

Sümeyye İNAL

Department of Statistics

Master of Science Thesis

Advisor: Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK

The development of data collection and storage methods have greatly increased the presence of functional data and classical statistical methods used to analyze these data have become insufficient. The methods and techniques used to analyze these data are called Functional Data Analysis (FDA). FDA involves transforming data points to continuous functions, which is basically done using Basis Functions and Roughness Penalty Approach. In FDA analysis is done in terms of functions instead of single data points. Since it provides a faster and richer analysis opportunity, FDA has a more important place in modern data analysis day by day.

The aim of this study is to investigate Functional Principal Component Analysis (FPCA), which is a FDA method, and make applications by using different basis functions on two different real datasets.

In the first application in this study, the temperature and household electricity consumption data of the provinces in Turkey is examined, and Regularized Functional Principal Components Analysis is used with the aim of analyze the variation structure in the data. Functional principal component analysis scores and scatterplots visually presented the main differences between provinces. As a result, it is concluded that the electricity consumption increased as the temperature increased. In the second application, the variation of Covid-19 total case numbers

by country is examined. Individual functions, principle component functions and derivative functions are interpreted and it is concluded that , the increase in the number of cases reached maximum levels in the first 15 days.

Keywords: Functional data analysis, roughness penalty approach, regularized functional principal component analysis, electricity consumption, Covid-19

1
GİRİŞ

1.1 Literatür Özeti

Fonksiyonel veriler, her biri zaman, uzay, dalga boyu ve olasılık gibi bir süreklilik boyunca değişen eğri veya yüzey örneklerinden oluşur [1]. Bu tür veriler, tekrarlanan ölçümlerin kısa bir aralıkta yoğun bir şekilde toplanmasına izin veren gelişmiş ölçüm araçlarının ortaya çıkması nedeniyle birçok alanda yaygınlaşmıştır. Fonksiyonel veriler mühendislikten jeoloji, biyoloji, tıp ve kimyaya kadar her bilim dalında ortaya çıkmaktadır. Örneğin Viviani, Grön ve Spitzer, 2005 [2]; Hwang, Jung, Takane ve Woodward, 2012 [3]; nin üzerinde çalıştığı beyin aktivasyonlarının zamanın bir fonksiyonu olarak tarandığı nörogörüntüleme verileri, Jackson ve Sirois, 2009 [4]; un üzerinde çalıştığı göz izleme verileri ve Vines, Wanderley, Krumhansl, Nuzzo ve Levitin, 2005 [5]; in üzerinde çalıştığı melodik algıdaki değişikliklerin zaman içinde ölçüldüğü müzik algısı verileri fonksiyonel verilere bir örnek olarak gösterilebilir.

Temel Bileşen Analizi (TBA), yüksek boyutlu veri kümelerinin boyutsallığını azaltmayı amaçlayan çok değişkenli bir istatistiksel tekniktir. Bunu, orijinal veri kümesini en iyi temsil eden yeni, çok daha küçük ilişkisiz değişkenler (Ana Bileşenler) hesaplayarak başarır. Her yeni değişken, orijinal değişkenlerin doğrusal bir kombinasyonudur. İlk temel bileşen, maksimum varyans miktarını açıklayan orijinal değişkenlerin doğrusal birleşimidir. İkinci temel bileşen birincisine diktir ve verilerdeki kalan varyansın maksimum miktarını açıklar. Tüm temel bileşenler birbirine diktir, bu nedenle gereksiz bilgi yoktur. Geriye kalan tüm ana bileşenler benzer şekilde tanımlanır, böylece en düşük dereceli bileşenler çok az sapma oluşturur ve genellikle yok sayılabilir[6]. Çok değişkenli istatiksel analizde sıklıkla kullanılan TBA nın fonksiyonel veri analizindeki karşılığı fonksiyonel temel bileşen analizidir (FTBA). FTBA, fonksiyonel verilerin kolayca anlaşılmayan kovaryans yapısına bakmanın bilgilendirici bir yolunu sunar ve fonksiyonel verilerdeki ana varyasyon modunu temsil eder.

Temel bileşen fonksiyonları görselleştirme, aykırı değer tespiti (Jones and Rice 1992; Hyndman and Shang 2009), diskriminant analizi (Hall, Poskitt, and Presnell 2001) ve temel bileşen regresyonu (Yao, Muller, and Wang 2005 ; Reiss and Ogden 2007) için kullanılmaktadır. Temel bileşen fonksiyonları aynı zamanda eğrilerin zaman serilerinin dinamik modellemesi için zaman serileri analizine girdi olarak da kullanılmaktadır (Hyndman ve Ullah 2007 ; Shen ve Huang 2008 ; Hyndman ve Shang 2009 ; Shen 2009) [7].

FTBA için temel fikirler, 1940 ların başlarında Karhunen (1946) ve Loève'nin (1946) nin birbirinden bağımsız olarak olarak geliştirdiği, sürekli bir stokastik sürecin optimal seri genişlemesi ve öz analizin bir simetrik matristen integral operatörlere kadar genişletildiği teorisine dayanmaktadır. Daha sonra Rao (1958) ve Tucker (1958), gözlemlenen fonksiyon değerlerine çok değişkenli TBA uygulayarak Karhunen-Loève (KL) genişlemesinin fonksiyonel verilere uygulanmasını sağlamıştır. Sonsuz boyutlu durumlarda FTBA tahmincilerinin önemli asimtotik özellikleri Dauxois, Pousse ve Romain(1982) [8] tarafından rasgele fonksiyonların bir vektörü için incelenmiş ve böylece FTBA için daha kapsamlı bir çerçeve oluşturulmuştur. Bunun ardından Besse ve Ramsay (1986) [9], Castro (1986) [10], Rice and Silverman (1991) [11] ve Silverman (1996) [12] daki çalışmaları da FTBA ile ilgili ilk çalışmalardandır.

Fonksiyonel temel bileşen analizi ile ilgili literatürdeki çalışmalar aşağıda grublandırılarak verilmiştir:

1. Fonksiyonel TBA: Choi, Hwang ve diğerleri (2017) [13], fonksiyonel manyetik rezonans görüntüleme verilerini kullanarak bireyler arasında yaygın olarak etkinleştirilen sinirsel aktivite ağlarını tanımlamak için fonksiyonel temel bileşen analizi ve fonksiyonel çoklu-set kanonik korelasyon analizini eş zamanlı olarak birlikte uygulayarak daha yorumlanabilir sonuçlar elde etmişlerdir. Feng ve Qian (2018) [14] 2002 - 2016 aylık getiri verilerini kullanarak Çin faiz oranları yapısını FTBA ile analiz etmişlerdir ve ilk iki bileşenin getiri eğrisindeki toplam varyasyonun yüzde 98,68 ini oluşturduğu sonucunu elde etmişlerdir. Wang, Sun, ve Li (2014) [15], Şanghay borsası 50 endeksinin aylık getirilerdeki değişim FTBA ile incelenmiş. Harrison ve diğerleri (2009) [6], rekabete dayalı yarış

yürüyüşçülerinde diz ekleminin kinematiği ve kinetiğini değerlendirmek ve sınıflandırmak için FTBA tekniğini kullanmış ve bunun sonucunda yüksek ve düşük vasıflı sporcular arasında potansiyel olarak önemli teknik farklılıklar ortaya konulmuştur. Han Lin Shang (2014) [16], 1816-2009 yılları arasındaki Fransız erkeklerinin yaşa bağlı ölüm oranını incelemiştir. Viviani, Groan, ve Spitzer (2005) [2], fonksiyonel manyetik rezonans görüntüleme (fMRI) verileri için FTBA uygulanmış ve FTBA nın kümeleme, sıradan TBA ve tasarım matrislerine dayanan yöntemlerden farklılıkları ele alınmıştır. Ingrassia ve Costanzo (2005) [17], 3 Ocak 2000'den 30 Aralık 2002'ye kadar MIB30 sepetindeki hisselerin günlük işlem hacmi FTBA ile incelenmiştir. Kneip ve Utikal(2011) [18], Kernel düzgünleştirme metodu ile İngiliz hane halkı gelir verisi üzerinde FTBA uygulamış ve 1968-1988 yılları arasında hane halkı gelir ve yaş dağılımlarının evrimi ve etkileşimi üzerinde çalışmıştır. Hall ve Hosseini-Nasab (2006) [19], Fonksiyonel TBA'da özdeğer ve özfonksiyon tahmincilerinin stokastik 'Taylor serisi açılımlarının' nasıl geliştirileceği ve bu açılımların doğruluğunu teorik olarak ele alınmıştır. Ramsay ve Dalzell (1991) [20], L-Spline teorisinin doğrusal modelleme ve FTBA da kullanımı, Kanada hava istasyonları arasında sıcaklık yağış modellerindeki farklılıklar üzerine bir çalışma ile açıklamaktadır. Yao ve Lee (2006) [21], Fonksiyonel TBA için cezalı splayn modeller üzerine bir çalışma yapmışlar ve fonksiyonel veya boylamsal verilerde yaygın olarak bulunan örneklem içi korelasyonu azaltmak için, ortalama fonksiyonun kestirimini iyileştirmek için tekrarlı bir tahmin prosedürü önermişlerdir. Grith, Wagner, Hardle ve Kneip (2018) [22], çok değişkenli eğrilerin türevleri için fonksiyonel temel bileşen analizi üzerinde çalışmışlar. Shang (2014) [16], fonksiyonel temel bileşen analizinin, fonksiyonel verilerin açıklayıcı analiz, modelleme, tahmin ve sınıflandırılmasında kullanımı üzerinde bir çalışma sunmaktadır.

2. Düzgünleştirilmiş FTBA: Literatürde düzgünleştirilmiş fonksiyonel temel bileşen analizi iki farklı şekilde ele alınmıştır. Birincisi FTBA uygulamadan önce fonksiyonel verileri düzgünleştirmektir. Ramsay ve Dalzell 1991 [20]; Foutz and Jank 2010 [23] daki çalışmaları bu şekilde yapılan çalışmalara örnek olarak gösterilebilir. İkincisi örneklem kovaryansına pürüzlülük cezası

terimi ekleyerek örneklem kovaryansını maksimize etmektir (Pezzulli ve Silverman 1993 [24]; Silverman 1996 [12]). Ramsay ve Silverman 2002 [25]de krimonoloji verileri üzerine bir düzgünleştirilmiş FTBA uygulaması yapmıştır. Yao, Muller ve diğerleri (2003) [26], California-Davis Üniversitesi'nde yapılan bir deneyde, 13 sağlıklı yetişkin gönüllüden 200 günlük alınan toplam 50 numunede (20 si ilk gün) plazmadaki folat yoğunluğunu incelemiş ve plazmadaki 14C-folata özgü aktivitenin dinamik davranışını karakterize etmek için düzgünleştirilmiş fonksiyonel temel bileşen analizi uygulamıştır. Ertaş ve Keser (2008) [27], İMKB 100 endeksinde yer alan 13 tane şirketin haftalık hisse senedi kapanış fiyatları verisi üzerinde düzgünleştirilmiş FTBA uygulamıştır. Li Luo, Yun Zhu and Momiao Xiong (2013) [28], genetik varyasyonun tüm alelik spektrumunun ilişkisini test etmek için düzgünleştirilmiş FTBA uygulamış ve sonuçları düzgünleştirilmemiş FTBA ile kıyaslamışlardır. Sonuç olarak düzgünleştirilmiş FTBA tarafından tanımlanan anlamlı derecede ilişkili gen sayısının, düzgünleştirilmemiş FTBA tarafından tanımlanandan çok daha fazla olduğunu ve düzgünleştirilmiş FTBA tarafından elde edilen P değerlerinin, düzgünleştirilmemiş FTBA tarafından elde edilen P değerlerinden çok daha az olduğu sonuçlarına varmışlardır.

3. Çok değişkenli fonksiyonel TBA: Berrendero, Justel ve Svarc (2011) [29] de çok değişkenli FTBA üzerinde çeşitli uygulamalar yapmıştır. Zhou ve Pan (2014) [7] Teksas hava istasyonları tarafından kaydedilen 100 yıllık sıcaklık verileri kullanılarak Teksas'taki sıcaklık değişimini iki değişkenli FTBA ve TBA için karma modeli ayrı ayrı uygulayarak sonuçları kıyaslamışlardır. İki değişkenli fonksiyonel TBA bileşenleri aynı birim için ölçüldüğünde ve benzer bir varyasyona sahip olduğunda çok iyi sonuçlar verir. Fakat değişken sayısı arttıkça ve temel bileşenlerden birisi nispeten büyük bir değişkenliğe sahip olduğunda çok değişkenli FTBA iyi sonuçlar veremeyebilir. Bu nedenle Chiou, Chen ve Yang (2014) [30], klasik Karhunen-Loève genişlemesinin genişletilmiş bir versiyonunu kullanarak normalleştirilmiş çok değişkenli FTBA yaklaşımını sunmuşlardır.
4. Mekansal ilişkili (spatially correlated) veriler için FTBA: Liu, Ray ve Hooker (2017) [31], mekânsal olarak ilişkili fonksiyonel verilerin analizi üzerinde

çalışmış ve seyrek gözlem çerçevesinde, mekansal korelasyonları açıkça tahmin etmek ve bireysel eğrileri yeniden yapılandırmak için koşullu beklenti ile mekansal bileşen analizinde yeni bir yaklaşım önermektedir. Li ve Guan (2014) [32] 1992'den 2009'a kadar teşhis edilen 8230 pankreas kanseri insidansının CTR(Amerikanın en eski kanser kayıt defteri) verilerini kullanarak pankreas kanseri vakalarında mekânsal- zamansal örüntü incelenmiştir.

5. Dayanıklı (Robust) FTBA: Temel bileşenlerin tahmin edicilerine yönelik önemli bir dezavantaj, aykırı değerlere olan duyarlılıklarıdır. Locantore ve diğ. (1999) [33] sınırlı etki (bounded influence) yaklaşımını benimseyerek temel bileşenler için dayanıklı tahminciler önermişlerdir. Gervini (2008) [34], çok değişkenli medyan ve küresel (spherical) ana bileşenlerin fonksiyonel versiyonları olan tamamen fonksiyonel dayanıklı tahminciler önermiştir. Hyndman ve Ullah (2007), zamana göre gözlenen yaşa bağlı mortalite ve doğurganlık oranlarını tahmin etmek için dayanıklı bir tahminci önermiştir. Bu yaklaşımının asimptotik özellikleri Bali ve diğerleri (2011) ele almıştır. Ortalama fonksiyon ve temel bileşenlerin robust tahmincilerine alternatif olarak, aykırı değerleri belirleyebilir ve kaldırabiliriz. Bu fikri takiben, Fraiman ve Muniz (2001), kesikli araçları fonksiyonel derinlik kavramını kullanarak önerdiler. Daha sonra, Cuevas ve diğerleri (2007), López-Pintado ve Romo (2007, 2009), Cuevas ve Fraiman (2009), Cuesta-Albertos ve Nieto-Reyes (2010) ve Gervini (2012) de bir dizi derinlik tabanlı tahminci önermişlerdir [16].
6. Yapısal FTBA: Shou, Zipunnikov, Crainiceanu ve Greven (2015) [35], belirli bir doğrusal yapıya sahip herhangi bir fonksiyonel model için TBA yoluyla değişkenliğin ayrıştırılması için bir yöntem olarak yapılandırılmış fonksiyonel temel bileşen analizini önermişlerdir. Bu çalışmalarıyla hem iç içe hem de çapraz tasarımlar için geçerli olan, ancak çok daha geniş bir model alanı için genelleştirilebilen veri odaklı bir yaklaşım sunmuşlardır. Yapısal FTBA, yüksek frekanslı ivmeölçer verileri, dilbilimsel veriler ve EEG verileri de dahil olmak üzere çeşitli yapısal verilere uygulanabilir.
7. Seyrek(Sparse) FTBA: Veri noktaları seyrek ve çok düzensiz olduğunda FTBA tahmini düzenlenmelidir [10]. Rice ve Wu (2001) fonksiyonel temel

bileşenleri tahmin etmek için karma etkiler modelinin kullanılmasını önermiştir. Bu modeller, X_i tarafından belirtilen her fonksiyonel verinin sadece i. nci fonsiyonla değil, tüm fonksiyonlarla tahmin edilebildiği seyreklik sorununu ele alır. Bununla birlikte, rasgele vektörlerin yüksek boyutlu varyans-kovaryans matrisi nümerik olarak stabil olmayabilir, bu da iyi tanımlanmamış problemlere yol açabilir. James ve diğerleri (2000) ve James ve Sugar (2003), karma etkiler modelinden kaynaklanan iyi tanımlanmamış problemlerden kaçınan indirgenmiş rank (reduced rank) modelini önermişlerdir. Zhou ve diğerleri (2008), indirgenmiş rank modelini, cezalandırılmış bir tahmin ve splaynlar yoluyla iki boyutlu seyrek ana bileşen modeline genişletmiştir. Kayano ve Konishi (2010) indirgenmiş rank modelini çok boyutlu seyrek fonksiyonel verilere genişletmiştir. Yao ve diğerleri (2005) [36], deneklerin sayısının artması ve tüm numuneden toplanan zaman noktalarının veri alanında yoğun hale gelmesi şartıyla, seyrek ve düzensiz fonksiyonel veriler konusunu çözmek için koşullu beklenti yoluyla bir FTBA önermiştir[16]. Selene Yue Xu ve diğerleri (2019) [37] da 578 aşırı kilolu kadında (% 60 kanserden kurtulan) zamana göre aktivite desenlerini belirlemek için fonksiyonel temel bileşen karma model yaklaşımını uygulamış ve temel bileşen skorlarıyla temsil edilen aktivite modelleri üzerinde sağlıkla ilişkili olup olmadığını test etmiştir.

8. Ortak (Common) FTBA: Bir grup fonksiyonel veri örneği için Benko ve arkadaşları (2009) [38] ortak fonksiyonel temel bileşen tahmini önermiş ve iki fonksiyonel veri örneğinin özdeğerlerinin, özfonksiyonlarının ve ortalama fonksiyonlarının eşitliğini incelemek için bir bootstrap testi sunmuştur. Boente ve diğerleri (2010) [39] ortak fonksiyonel temel bileşenler altında istatistiksel çıkarımda bulunmuştur. Coffey ve diğerleri (2011) [40] ortak fonksiyonel temel bileşenler fikrini insan hareketi çalışmalarına uyguladılar, Fengler ve diğerleri (2003) [41] ve Benko and Härdle (2005) [42] örtük volatilite yüzey dinamiklerini modellemek için ortak fonksiyonel temel bileşen uyguladılar[16].
9. Çok seviyeli (Multilevel) FTBA: FTBA genellikle bir stokastik süreçten üretilen fonksiyonel verilere uygulanır. Bununla birlikte, bazen en az iki stokastik süreçten üretilen fonksiyonel verileri gözlemleriz, örneğin, bir

hastanın sağlık durumu, bir süre boyunca gün içi kan basıncı ve kalp atışı ile izlenir. Di ve diğerleri (2009) [43], çok düzeyli fonksiyonel verilerin geometrik bileşenlerini çıkarmak için tasarlanmış çok düzeyli FTBA'yı tanıttı. Crainiceanu (2009) [44], çok düzeyli FTBA fikrini fonksiyonel regresyona genişletirken, Zipunnikov (2011) [45], manyetik rezonans görüntüleme kullanılarak taranan yüzlerce beyin görüntüsünü analiz etmek için hızlı ve ölçeklenebilir çok düzeyli FTBA önerdi [16]. Argiento ve diğerleri (2015) [46] , aynı binada farklı laboratuvarlar tarafından gerçekleştirilen bağımsız tekrarlanan ölçümlerden oluşan cephe ses yalıtımı üzerine yapılan deneysel bir çalışmanın verilerini analiz etmektedir ve fonksiyonel varyansı hem ölçüm düzeyinde hem de grup düzeyinde ayrıştırmak için çok düzeyli FTBA uygulamıştır. Di ve diğerleri (2014) [47] çok düzeyli FTBA yı seyrek örneklenmiş verilere uyguladı.

1.2 Tezin Amacı

Bu çalışmanın amacı, son yıllarda modern veri analizinde gittikçe önemi artan Fonksiyonel Veri Analizinin en yaygın boyut indirgeme tekniği olan Fonksiyonel Temel Bileşenler Analizini teorik ve uygulamalı olarak tüm ayrıntılarıyla ortaya koymak ve farklı baz fonksiyonlarını kullanarak iki farklı gerçek veri seti üzerinde uygulamasını yapmaktır.

1.3 Hipotez

Türkiye' de ve dünyada her sene ortalama sıcaklıklar ve yıllık elektrik tüketimi hızla artmaktadır. Buna bağlı olarak yapılan araştırmalar enerji kaynaklarının 30 yıl içerisinde tükeneceğini göstermektedir. Bu durumda tükenmeyen ve yenilenebilir bir enerji kaynağı olan güneş enerjisi ön plana çıkmaktadır. Özellikle Türkiye güneş enerjisi potansiyeli yüksek bir ülke olmasına rağmen yeterince güneş enerjisi sistemi kurulmadığından elektrik üretiminin sadece %0.009 unu güneş enerjisinden karşılamaktadır. Bu çalışmada artan elektrik tüketiminin karşılanması amacıyla elektrik tüketiminin en fazla olduğu illerin tespit edilmesi, güneş enerjisi kullanımının desteklenmesi ve yatırım yapılması için teşvik edilmesi başlıca hipotez konularındandır.

2019 yılında ilk kez Çin'de ortaya çıkan Covid19 koronavirüs salgını hızla yayılarak tüm dünyayı etkisi altına almış ve bir çok ülke bu salgının yayılmasını önlemede geç kalmıştır. Bu nedenle bu tez çalışmasında Covid19 toplam vaka sayısı artışının kontrol alınabilmesi için; ilk vakanın ortaya çıkmasıyla birlikte özellikle ilk 15 gün ilgili kurumlar tarafından gerekli önlemlerin alınması ve tedbirlerin arttırılması, normalleşme sürecine girmek için acele edilmemesi diğer hipotez konularındandır.

1.4 Temel Bileşenler Analizi

Çok değişkenli dağılım gösteren herhangi bir veri setinde p adet değişken arasındaki ilişki, eğer değişkenler bağımlı bağımsız değişken olarak ayrılabiliyorsa regresyon teknikleri ile incelenir. Bu ayrımın yapılamaması ve değişkenler arasındaki birlikte değişimin yorumlanması gereksinimi çeşitli boyut indirgeme yöntemlerinin kullanılmasına neden olmuştur ki bu yöntemlerden biri de temel bileşenler analizidir [48].

Tarım, biyoloji, kimya, tıp, sağlık bilimleri, iklimbilimi, nüfus bilgisi, ekoloji, ekonomi, gıda, araştırma, genetik, jeoloji, meteoroloji, okyanus bilimi, psikoloji ve sosyoloji gibi birçok bilim dalında sıkça yararlanılan temel bileşenler analizi, değişkenler arasındaki bağımlılık yapısını yok etmek, boyut indirgemek ve çeşitli ölçütler bakımından sıralamalar yapmakta kullanılmaktadır.

Temel bileşenler analizi, başlı başına bir analiz tekniği olduğu gibi, başka analiz teknikleri için veri hazırlama tekniği olarak da kullanılmaktadır. Temel bileşenler analizi, çok değişkenli analizler olan Faktör Analizi, Kanonik Korelasyon analizi, Diskriminant (Ayırma) Analizi, Lojistik Regresyon Analizi ve Çok Değişkenli Regresyon Analizi için de ilk adımı oluşturmaktadır [49].

Fonksiyonel Temel Bileşen Analizi, çok değişkenli istatistikte en çok kullanılan boyut indirgeme tekniği olan temel bileşenler analizinin fonksiyonel halidir. Bu nedenle bu çalışmanın birinci bölümünde temel bileşenler analizinin tanımı, özellikleri, gerekliliği ve sağladığı faydalara yer verilmektedir.

İkinci bölümde, Fonksiyonel Veri Analizinin önemi, avantajları ve diğer istatiksel yöntemlerle ortak yönlerine yer verilecektir. Fonksiyonel Veri Analizinin ilk adımı olan kesikli verileri, sürekli fonksiyonel verilere dönüştürmek için izlenecek yollar ayrıntılarıyla verilecektir. Bu çalışmada fonksiyonel verileri elde etmek için Baz

Fonksiyon yaklaşımı ve Pürüzlü Ceza yöntemleri kullanılmıştır. Bu nedenle bu yöntemler ve en çok tercih edilen baz fonksiyonlar detaylı bir şekilde ele alınmıştır. Bu tez çalışmasında da özellikle Fourier ve B-Splayn Baz Fonksiyonları kullanılmıştır.

Üçüncü Bölümde Fonksiyonel Temel Bileşen Analizi, Düzgünleştirilmiş FTBA ve Çok Değişkenli FTBA ele alınmıştır ve uygulama sırasında izlenmesi gereken yollar sırasıyla verilmiştir.

Dördüncü Bölümde FTBA nın iki farklı gerçek veri seti üzerinde bir uygulamasına yer verilmektedir. İlk olarak; İzmir, Manisa, Bolu, Kocaeli, Sakarya, Düzce, Balıkesir, Bursa, Çanakkale, Yalova, Artvin, Giresun, Gümüşhane, Rize ve Trabzon illerinin Ocak-2013 ve Mart-2019 yılları arasındaki aylık ortalama sıcaklık ve elektrik tüketimi üzerine bir uygulama yapılmıştır. İkinci bir uygulama olarak ülkelere göre ilk Covid19 vakasının ortaya çıkmasıyla 98 ülkede ilk 83 günde toplam vaka sayılarındaki değişim FTBA ile incelenmiştir.

1.4.1 Temel Bileşenler Analizi Tanımı

Çok değişkenli istatistiksel analizde n tane bireye (nesne) ilişkin p tane değişken (özellik) incelenmektedir. Bu özelliklerden birçoğunun birbirleriyle ilişkili (bağımlı) ve p sayısının çok büyük olması analizde sorun yaratmaktadır. Örneğin insanın anatomik özellikleri değişkenleri ifade ediyor olsun. Bu durumda karın çevresi, ağırlık, göğüs çevresi, boy uzunluğu, kol uzunluğu, omuz genişliği, bacak uzunluğu vs. çok sayıda değişken bulunmaktadır. Bu değişkenlerin bazıları birbirleri ile ilişkilidir. Oysaki bu durum değişkenlerin (yaklaşık da olsa) bağımsızlığı kuralını zedeler. Ayrıca, çok sayıda değişkenle çalışmak, işlem yükünü artıracağı ve elde edilecek sonuçların yorumunda bazı güçlüklere neden olacağı için arzulanan bir durum değildir. Bilgisayar olanaklarının çok geliştiği günümüzde işlem yükü bir sorun olarak görülmese de, çok sayıda değişkene ilişkin analiz sonuçlarının yorumlanması ve özetlenmesi gerçekten zor olabilmektedir. Böyle durumlarda başvurulan tekniklerden en önemlisi Temel Bileşenler Analizi (Principal Component Analysis)'dir [50].

Temel bileşenler analizi, orijinal p değişkenlerinden oluşan veri setini, daha az sayıda ve bu değişkenlerin doğrusal bileşenleri olan yeni değişkenlerle ifade etme yöntemidir. Yani, aralarında korelâsyon bulunan p sayıda değişkenin açıkladığı

yapıyı, aralarında korelâsyon bulunmayan ve sayıca orijinal değişken sayısından daha az sayıda (k<p) orijinal değişkenlerin doğrusal bileşenleri olan k tane değişkenle ifade etme yöntemine temel bileşenler analizi denir [51].

Temel bileşenler analizi, veri indirgemek, şekil sıkıştırıp özellik çıkarmak için esaslı ve önemli bir tekniktir. Temel bileşenler analizi, bir veri setinin varyans-kovaryans yapısını, bu değişkenlerin doğrusal birleşimleri yardımıyla açıklayarak, boyut indirgenmesi ve yorumlanmasını sağlayan çok değişkenli bir istatistik yöntemdir [52].

Temel bileşenler analizi birbiri ile ilgili çok sayıda değişkenden, en az bilgi kaybı ile bağımsız, kavramsal açıdan anlamlı az sayıda yeni değişkenler (boyutlar) bulmayı, ortaya çıkarmayı amaçlar. TBA'nin genel amacı veri indirgeme ve yorumlamadır. Yukarıdaki açıklamalardan da anlaşılacağı gibi temel bileşenler analizinde ele alınan değişkenler arasında, bağımlı ve bağımsız değişkenler olarak adlandırılacak bir yapı yoktur. Veri setindeki tüm değişkenler, eşanlı olarak bir yapıyı oluşturan birbirleriyle ilişkili değişkenlerdir [53].

1.4.2 Temel Bileşenler Analizinin Özellikleri

Geometrik olarak, birbirinden bağımsız değişkenler ve eksenler üreten temel bileşenler analizinde önemli olan hususlardan bazıları şunlardır [54].

- Birimlerin her bir eksene göre koordinatları yeni değişkenlerin birim değerlerini vermektedir. Yeni eksenlere veya değişkenlere temel bileşenler ve yeni değişkenlerin birim değerlerine temel bileşen değerleri (puanları) adı verilmektedir.
- Her bir yeni değişken, orijinal değişkenlerin doğrusal bir birleşimidir.
- Birinci temel bileşen, verilerdeki maksimum varyansı açıklayacak şekilde türetilmektedir.
- İkinci temel bileşen geriye kalan toplam varyansa maksimum katkıda bulunmaktadır. Yani birinci temel bileşen en çok, diğer bileşenler ise gittikçe azalan miktarlarda toplam varyansa katkıda bulunurlar. Bu nedenle, az sayıda bileşenle toplam varyansın büyük bir kısmı açıklanabilmektedir.
- Maksimum olarak türetilebilecek temel bileşen sayısı orijinal değişken sayısına eşittir.

- Yeni bileşenler birbirinden bağımsız olarak türetilmektedir.

1.4.3 Temel Bileşenler Analizinin Gerekliliği ve Sağladığı Faydalar

Temel bileşenler analizi, bağımlılık yapısını yok etme ve boyut indirgeme amaçlarıyla kullanılmaktadır. **Z** matrisinin (X matrisinin standartlaştırılmış hali) değişkenleri arasında tam ya da tama yakın bağımsızlık olması durumunda; bağımsızlaştırma bir amaç olamayacağı gibi boyut indirgemeden de önemli bir kazanç sağlanamaz. Çünkü;

$$R = \frac{1}{n-1}\sum z_j z_j^t = ZZ^t = \mathrm{I} \tag{1.1}$$

özelliği nedeniyle z_j'lerin yerine y_i'lere dönüştürülmesinden de yine birim ilişki matrisine ulaşılacaktır. Yukarıda belirtilen deneysel noktaların ilişki matrisi eşitliğindeki gibi ise bu noktaların p boyutlu uzayda bir küre oluşturdukları söylenebilir. Bu kürenin eksenleri z_j'ler yerine y_i'ler de olsa durumda önemli bir farklılık olmayacaktır. Bu nedenle böyle bir dönüştürme ile y_i'lerin bulunmasına gerek yoktur. Ayrıca tüm eksenlerde eşit varyans olduğundan, dönüştürülmüş y_i'lerden m tanesinin seçilerek boyut indirgenmesi de uygun değildir. Deney ya da gözlemlerden elde edilen veriler için $|\mathrm{R}| = \mathrm{I}$ olma olasılığı sıfırdır. Hangi durumlarda ilişki matrisinin birim matris olarak kabul edileceği konusunda Barlett, tek değişkenli analizlerdeki varyans karşılığı olan genelleştirilmiş varyans kavramını ve aşağıdaki testleri önermiştir [50].

$$-\left[(n-1) - \frac{1}{6}(2p+5)\log|R| \sim \chi^2_{\frac{1}{2}p(p-1)}\right] \tag{1.2}$$

$$-\left[n - \frac{1}{6}(2p+11)\log|R| \sim \chi^2_{\frac{1}{2}p(p-1)}\right] \tag{1.3}$$

Burada |R| , R matrisinin determinantı yani genelleştirilmiş varyansıdır. Küresellik testi olarak bilinen test için,

$$H_0: R = \mathrm{I}$$
$$H_1: R \neq \mathrm{I} \tag{1.4}$$

hipotezleri kullanılmakta ve H_0 hipotezinin red edilmesi durumunda temel bileşenler analizinin kullanılması öngörülmektedir.

Temel bileşenler analizinin kullanımının yararlarından birincisi temel bileşenlerin aralarında korelâsyon yapısı bulunmamasıdır. Temel bileşenler analizinin avantajı, temel bileşenlerin ortogonal olmasından kaynaklanır. Bu avantaj iki şekilde gerçekleşir. İlki, nesneler arasında temel bileşenler analizi skorlarının alınması sonucu yapılacak karşılaştırmalarda farklı ana bileşenler birbirleriyle ilişkisiz olacaktır ve takip eden analizlerde temel bileşen skorları uygun bir şekilde kullanılabilecektir. İkincisi, temel bileşenler analizinin gizli değişkeni varyans yüzdesi olarak hiyerarşik bir yapıda sıralayabilmesidir.

Temel bileşenler analizinin kullanım yararlarından ikincisi, temel bileşenler analizi sonucunda elde edilen doğrusal bileşenlerin, orijinal değişkenlerin diğer doğrusal bileşenlerinden farklı olarak, mümkün olabilecek az sayıda olarak değişkenler arasındaki varyasyonun yüzdesinin maksimum yapacak şekilde oluşturmasıdır [55].

2
FONKSİYONEL VERİ ANALİZİ

Klasik istatiksel yöntemlerde bir bireyin boy, kilo, yaş vb. ölçümleri yapılıp bunların analizi yapılabilirken, veri toplama yöntemlerinde kaydedilen ilerlemeler ve teknolojik gelişmelerle birlikte bir bireyin 3 boyutlu beyin dalgaları da veri olarak kullanılabilmektedir. Belirli bir aralık üzerinde tanımlı ve sürekli olan bu tür veriler fonksiyonel veri olarak adlandırılır. Fonksiyonel veriler genellikle fonksiyonlardan ve düzgün eğrilerden meydana gelirler. Süreç genellikle zamana göredir fakat aynı zamanda uzaysal konum ya da dalga boyu da olabilir. Bu tür verileri analiz edebilmek için geliştirilen yöntem ve teknikler Ramsay ve Dalzell (1991) [20] tarafından "Fonksiyonel Veri Analizi" (Functional Data Analysis) terimi ile adlandırılmıştır. Fakat bu konuyla ilgili çalışmalar daha eski tarihlere dayanmakta ve Grenander (1950) [56] ve Rao (1958) [57] 'e kadar uzanmaktadır.

Fonksiyonel Veri Analizi (FVA) nin son yıllarda kullanımı ve uygulama alanı hızla genişlemekte. Bu tür veriler mühendislikten jeoloji, biyoloji, tıp ve kimyaya kadar her bilim dalında ortaya çıkmaktadır. Arkeologlar tarafından bulunan kemiklerin şekillerinden, uzun yıllar toplanan ekonomik verilere ve bir hokkabazın parmağının izlediği yola kadar değişen bir çok alanda ortaya çıkan fonksiyonel veriler mevcuttur. Dolayısıyla fonksiyonel veriler bazen çok sayıda bağımsız değişkenin tahmini olasılık yoğunluğuyken bazen de bir yüzeyde veya uzayda izlenen eğriler olabilmektir [28].

Fonksiyonel veri analizinin amaçları istatistiğin diğer branşlarıyla benzerdir ve aşağıdaki gibi özetlenebilir.

- Veriyi ileriki analizlere yardımcı olacak şekilde ifade etmek,
- Veriler arasındaki değişkenliği açıklamak,
- Verinin önemli kısımlarını açığa çıkaracak şekilde görselleştirmek,
- Bağımsız değişkenlerle bağımlı değişkendeki varyasyonu açıklamak [1].

Veri analizini fonksiyonel açıdan ele almanın pratik nedenleri şu şekilde sıralanabilir:

- Otomatik veri toplama imkanları daha fazla araştırmacı tarafından erişilebilir hale geldiğinden fonksiyonel gözlemler uygulama çalışmalarında daha sık ortaya çıkmaktadır. Ayrıca, pürüzsüzleştirme ve interpolasyon prosedürleri sonlu gözlem kümelerinin fonksiyonel gösterimlerini sağlayabilir.
- Sınırlı sayıda gözlem olmasına rağmen bazı modelleme problemlerini fonksiyonel terimlerle düşünmek daha doğaldır.
- Bir analizin amaçları, bir fonksiyonun tamamını, türevlerini veya diğer fonksiyonlara ilişkin değerleri tahmin etmek için sonlu verilerin kullanılması durumunda olduğu gibi, doğada fonksiyonel olabilir.
- Fonksiyonel analizin bir adımı olan düzgünleştirme çok değişkenli veri analizinde kullanılması durumunda analiz için önemli sonuçlar verebilir [20].

Fonksiyonel veri analizi genellikle zamansal verilerle ilgilenmesine rağmen, kapsamı ve hedefleri zaman serisi analizinden oldukça farklıdır. Zaman serisi analizi temel olarak verilerin modellenmesine veya gelecekteki gözlemlerin öngörülmesine odaklanırken, fonksiyonel veri analizinde vurgu yörüngeler ve şekiller üzerinedir. Ayrıca fonksiyonel veri analizinde tanım aralığının eşit aralıklı olmadığı yani her bir fonksiyonun farklı sayıda gözlem değerine sahip olduğu durumlarda ve kayıp verilerin olduğu durumlarla da çalışılabilir [17].

Fonksiyonel veri analizi kesikli veri noktalarını, sürekli türevlenebilir fonksiyonlara dönüştürür. Böylece tek tek veri noktalarını incelemek yerine fonksiyonlar açısından analiz yapılır. Bu, zamanla daha az varsayım içereceğinden ayrık verilerin analizinde avantaj sağlar. Bir diğer avantajı da fonksiyonların trendlerini, maksimum ve minimum noktalarını, ivmelerini ve değişim oranlarını görmeyi mümkün kıldığından daha zengin bir analiz imkanı sağlar.

Fonksiyonel veriler tipik olarak $\mathrm{I} = [0, \mathrm{T}]$ kompakt aralığı üzerinde tanımlı bağımsız, reel değerli $X_1(t), X_2(t), \dots, X_n(t)$fonksiyonlarının rasgele bir örneğinden oluşur.

Klasik çok değişkenli analizde gözlemler T boyutlu rassal X vektörü ile modellenir. X (Ω, A, P) olasılık uzayından $(\mathbb{R}^T, \mathcal{B}^T)$ ölçülebilir reel uzayına tanımlı ölçülebilir bir rassal vektördür. $\mathcal{B}^T$ $\mathbb{R}^T$ üzerinde tanımlı Borel kümeleridir ve bu veri setleri

$$X: (\Omega, A, P) \rightarrow (\mathbb{R}^T, \mathcal{B}^T) \quad \{\chi : x_{i1}, x_{i2}, \dots, x_{iT} \ , \ i=1,2,\dots.N \} \tag{2.1}$$

ile temsil edilir.

Fonksiyonel veri analizinde gözlemler, t∈ *J ve J* $\mathbb{R}$ üzerinde tanımlı sınırlı bir aralık olmak üzere bir X(t) stokastik süreci ile modellenir. X (Ω, A, P) olasılık uzayından $(\mathcal{H}, \mathcal{B}_{\mathcal{H}})$ Hilbert uzayına tanımlı rassal bir fonksiyon olmak üzere, fonksiyonel veri setleri,

$$X: (\Omega, A, P) \rightarrow (\mathcal{H}, \mathcal{B}_{\mathcal{H}}) \quad \{\chi : x_i(t), \ , \ i=1,2,\dots.N \ , t \in J\} \tag{2.2}$$

ile temsil edilir [58].

L^p uzayları fonksiyonel analiz uygulamalarında önemli bir yer tutmaktadır. L^p, bir ölçü uzayında p ninci mertebeden Lebesque integrallenebilir fonksiyonların normlu bir vektör uzayını gösterir. Bu vektör uzayı p normlu bir tam uzaydır bir diğer deyişle p normlu bir Banach uzayıdır.(Tüm tam uzaylar bir Banach uzayıdır aksi doğru değildir). p=2 özel durumunda ise bu bir Hilbert uzayıdır ve bu çalışmada çoğunlukla L^2 fonksiyonel Hilbert uzayları üzerinde çalışılmıştır.

L^2 içindeki fonksiyonlara karesi integrallenebilir fonksiyonlar denir ve L^2 üzerinde iç çarpım şu şekilde tanımlıdır.

$$< x, y > = \int_a^b x(t)\, y(t) dt \qquad t \in [a, b] = J \qquad x, y \ \in \ L^2(J) \tag{2.3}$$

Hilbert uzayı üzerinde tanımlı bir x fonksiyonun normu ise

$$\| x \| = < x, x >^{1/2} = \left(\int_a^b x(t)^2 dt\right)^{1/2} \tag{2.4}$$

şeklindedir.

2.1 Fonksiyonel Veriler İçin Tanımlayıcı İstatistikler

Fonksiyonel veriler için tanımlayıcı istatistikler klasik istatistikteki gibidir. n farklı noktada gözlemlenen N tane fonksiyon eğrisi için ortalama fonksiyonu

$$\bar{x}\,(t) = \frac{1}{N}\sum_{i}^{N} x_i(t_j) \qquad i = 1,2,\dots,N \qquad j = 1,2,\dots,n \tag{2.5}$$

şeklindedir. Benzer şekilde varyans fonsiyonu

$$var_X(t) = \frac{1}{N-1}\sum_{i}^{N}(x_i(t) - \bar{x}(t))^2 \tag{2.6}$$

ile gösterilir ve standart sapma fonksiyonu varyans fonksiyonunun kareköküdür.

Kovaryans fonksiyonu t_1 ve t_2 gibi farklı argüman değerleri için gözlemlerin bağımlılığını özetler ve şu şekilde gösterilir:

$$cov_X(t_1,t_2) = \frac{1}{N-1}\sum_{i=1}^{N}\{(x_i(t_1) - \bar{x}(t_1))(x_i(t_2) - \bar{x}(t_2))\} \tag{2.7}$$

Korelasyon fonksiyonu ise

$$corr_X(t_1,t_2) = \frac{cov_X(t_1,t_2)}{\sqrt{var_X(t_1)var_X(t_2)}} \tag{2.8}$$

şeklindedir.

Çarpraz kovaryans ve korelasyon fonksiyonu da aşağıdaki gibidir:

$$cov_{X,Y}(t_1,t_2) = \frac{1}{N-1}\sum_{i=1}^{N}\{(x_i(t_1) - \bar{x}(t_1))(y_i(t_2) - \bar{y}(t_2))\} \tag{2.9}$$

$$corr_{X,Y}(t_1,t_2) = \frac{cov_{X,Y}(t_1,t_2)}{\sqrt{var_X(t_1)var_Y(t_2)}} \tag{2.10}$$

2.2 Fonksiyonların Baz Fonksiyonlar ile Gösterimi

K boyutlu bir fonksiyon uzayındaki her fonksiyon, K adet lineer bağımsız fonksiyon cinsinden yazılabilir. Bu fonksiyonlara baz fonksiyon denir. Bir x fonksiyonunun φ_k baz fonksiyonlar ile gösterimi

$$x(t) = \sum_{k=1}^{K} c_k\varphi_k(t) \tag{2.11}$$

şeklindedir ve c_k bu baz fonksiyonlara karşılık gelen katsayılardır.

Düzgünleştirme derecesini belirleyen K parametresi verinin özelliklerine göre seçilir. Nispeten az sayıda baz fonksiyon kullanarak daha tatmin edici sonuçlar elde edilir. K ne kadar küçükse baz fonksiyonlar o kadar verinin özelliklerini yansıtır [1].

Çoğu fonksiyonel veri analizinde periyodik durumlar için Fourier baz fonsiyonlar, periyodik olmayan durumlar için ise B- Splayn baz fonksiyon yaklaşımı tercih edilmektedir. Türevlerin gerekli olmadığı durumlarda da dalgacık baz fonksiyonları (**wevelet basis**) sıklıkla kullanılmaktadır. Bununla birlikte her zaman için kullanılabilecek, en iyi diyebileceğimiz bir baz fonksiyonun varlığından söz edemeyiz.

En çok tercih edilen baz fonksiyonlar;

- Fourier
- B-splayn
- Üstel ve Güçler
- Dalgacık

fonksiyonlarıdır.

Fourier Baz fonksiyonları: Periyodik fonksiyonlar için en çok tercih edilen baz Fourier serileridir. Bir f(x) fonksiyonunun Fourier seri açılımı

$$f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \sin(nx) + b_n \cos(nx) \qquad (2.12)$$

şeklindedir.

Bir Fourier serisi özellikle değişimin kararlı olduğu yani farklı bir yapının olmadığı, eğriliğin hemen hemen her yerde aynı olma eğiliminde olduğu fonksiyonlar için kullanışlıdır [1].

J aralığı üzerinde tanımlı fonksiyonlar için Fourier baz fonksiyonu

$$\varphi_k(t) = \begin{cases} 1, & k = 0 \\ \sin(rwt), & k = 2r - 1 \\ \cos(rwt), & k = 2r \end{cases} \qquad (2.13)$$

ile tanımlanır. w parametresi periyodu temsil eder ve J aralığının uzunluğu $\frac{2\pi}{w}$ dir.

Eğer t_j değerleri J aralığı üzerinde eşit aralıklı ve periyod J aralığının uzunluğuna eşitse, $\varphi'\varphi$ çarpraz iç çarpım matrisi köşegen olduğundan baz ortogonal olur. Baz fonksiyonlar uygun bir sabite bölünerek bu matris birim matrise dönüştürülebilir. Böylece yukarıda tanımlanan Fourier bazı ortonarmal bir baza dönüştürülür.

Fourier serilerinin istenilen dereceden türevi alınabilir. Özellikle türev fonksiyonunun incelenebilmesi eldeki veriden daha fazla bilgi elde edilmesini sağlar. Fourier baz fonksiyonunun m.inci mertebeden türevi:

$$\begin{aligned} D^m\varphi_0(t) &= 0, \qquad m \geq 1 \\ D^m\varphi_{2r-1}(t) &= (rw)^m \sin\left(rwt + \frac{m\pi}{2w}\right), m \geq 0 \\ D^m\varphi_{2r}(t) &= (rw)^m \cos\left(rwt + \frac{m\pi}{2w}\right), m \geq 0 \quad (\text{r=1,...}\frac{K}{2}) \end{aligned} \tag{2.14}$$

B Splayn Baz Fonksiyonları: Periyodik olmayan fonksiyonel veriler için en çok tercih edilen baz fonksiyon B Splayn baz fonksiyonudur.

J aralığı üzerinde tanımlı parçalı bir polinom eğrisi için baz fonksiyon

$$\varphi_j(t) = B_{j,K}(t) \quad j = 1,2,\dots,m+p-2 \tag{2.15}$$

ile tanımlıdır ve $B_{j,K}(t)$ K.nci dereceden , j.inci B splaynı temsil eder.T $= (t_1, \dots, t_m)$ düğüm vektörü olmak üzere, t_i ler azalmayandır yani her $i = 1,2,\dots,m-1$ için $t_i \leq t_{i+1}$ dir.

p dereceli i. nci B splayn baz fonksiyonu i=0,1,..,m+p ve p=0,1,...,K olmak üzere aşağıdaki gibi tanımlanır:

$$\begin{aligned} B_{i,0}(t) &= \begin{cases} 1, & t_i \leq t \leq t_{i+1} \\ 0, & diğer\ durumlarda \end{cases} \\ B_{i,p}(t) &= \frac{t - t_i}{t_{i+p} - t_i} B_{i,p-1}(t) + \frac{t_{i+p+1} - t}{t_{i+p+1} - t_{i+1}} B_{i+1,p-1}(t) \end{aligned} \tag{2.16}$$

2.15 teki eşitlikte verildiği gibi baz fonksiyon sayısı m+p-2 dir yani bir başka deyişle

Baz fonksiyon sayısı =Düğüm sayısı+mertebe-2 (2.17)

dir. Bir B splayn oluşturmak için gerekli olan düğüm sayısı en az p+2 tanedir.

Örneğin; ***T***=(0,0,0,1,1,1) düğüm vektörü olmak üzere 0.ıncı ve 1.inci dereceden B Splayn baz fonksiyonları şu şekildedir.

$$B_{0,0} = B_{1,0} = 0$$

$$B_{2,0} = \begin{cases} 1, & 0 \le t < 1 \\ 0, & dd \end{cases}$$

$$B_{3,0} = B_{4,0} = 0$$

$$B_{0,1} = \frac{t-0}{0-0} B_{0,0} + \frac{0-t}{0-0} B_{1,0} = 0$$

$$B_{1,1} = \frac{t-0}{0-0} B_{1,0} + \frac{1-t}{1-0} B_{2,0} = \begin{cases} 1-t, & 0 \le t < 1 \\ 0, & dd \end{cases} \qquad (2.18)$$

$$B_{2,1} = \frac{t-0}{1-0} B_{2,0} + \frac{1-t}{1-1} B_{3,0} = \begin{cases} t, & 0 \le t < 1 \\ 0, & dd \end{cases}$$

$$B_{3,1} = \frac{t-1}{1-1} B_{3,0} + \frac{1-t}{1-1} B_{4,0} = 0.$$

B spline baz fonksiyonları aşağıda verilen şu özelliklere sahiptir:

- $B_{i,p}(t)$ her zaman pozitiftir, $t \in (t_i, t_{i+p+1})$ için $B_{i,p}(t) > 0$
- $B_{i,p}(t) = 0$ ise $t \notin (t_i, t_{i+p+1})$ dir.
- $B_{i,p}(t)$ p dereceli parçalı bir polinomdur.
- $t \in [t_r, t_{r+1})$ olmak üzere $B_{i,p}(t)$ baz fonksiyonların toplamı 1 e eşittir. $\sum_{i=r-p}^{r} B_{i,p}(t) = 1$.
- Bir düğüm vektöründe bir düğüm tekrarlanabilir. Bu düğümün tekrar etme sayısı o düğümün katı olarak adlandırılır. **T** düğüm vektöründe k katlı bir düğüm olduğu varsayılırsa $B_{i,p}(t)$ baz fonksiyonu C^{p-k} sürekli olur. Bu nedenle katlılık arttıkça sürekliliğin seviyesi azalacaktır [59].

p.inci dereceden B splayn baz fonksiyonun m.inci mertebeden türevi m<K olmak üzere,

$$D^m B_{i,p}(t) = (p-1)\left\{\frac{D^m B_{i,p-1}(t)}{t_{i+p} - t_i} + \frac{D^m B_{i+1,p-1}(t)}{t_{i+p+1} - t_{i+1}}\right\} \qquad (2.19)$$

şeklindedir.

Birden fazla değişken olması durumunda, sıradan tek boyutlu B-splaynların çok boyutlu B-splaynlara genelleştirilmesi gerekmektedir. Bunun için de Tansör iç çarpım yaklaşımı kullanılarak sıradan bir boyutlu B-spline'ların basit cebirsel özelliklerinin çoğu taşınabilir. Bu konuyla ilgili ayrıntılı bilgi için Schumaker'a (1981) başvurulabilir [59].

Güçler ve Üstel baz fonksiyonları: Her iki baz fonksiyonu da periyodik olmayan durumlar için uygundur.

Üstel baz fonksiyonları

$$e^{\lambda_1 t}, e^{\lambda_2 t}, \dots, e^{\lambda_k t}, \dots \tag{2.20}$$

λ_k oran parametrelerinin hepsinin birbirinden farklı ve genellikle $\lambda_1 = 0$ olduğu üstel fonksiyonlardan oluşmaktadır. Sabit katsayılı lineer differansiyel denklemler için uygundur.

Güçler baz fonksiyonları

$$t^{\lambda_1}, t^{\lambda_2}, \dots, t^{\lambda_k}, \dots \tag{2.21}$$

şeklindeki polinomlardan meydana gelmektedir. t genellikle pozitiftir fakat negatif olduğu Güçler de mevcuttur.

Dalgacık baz fonksiyonları: $(-\infty, \infty)$ aralığı üzerinde karesi integrallenebilir tüm fonksiyonlar için uygun bir φ ana dalgacık fonksiyonu seçilerek j ve k tamsayıları için

$$\varphi_{j,k}(t) = 2^{j/2}\varphi(2^j t - k) \tag{2.22}$$

eşitliğinin tüm genişlemeleri göz önüne alınarak bir baz oluşturulabilir. φ $L^2(R)$ üzerinde bir tek fonksiyon ve $\varphi_{j,k}$ bu fonksiyondan türetilmiş ortonormal bir bazdır.

Dalgacık(Wavelet) serilerinde $L^2(R)$ üzerinde tanımlı bir f fonksiyonu

$$f(x) = \sum_{j,k=-\infty}^{\infty} c_{j,k}\varphi_{j,k}(x) \tag{2.23}$$

ile verilir. $\varphi_{j,k}$ ortonormal dalgacık fonksiyonlarını ve $c_{j,k}$ ler ise dalgacık fonksiyonunun katsayılarını temsil etmektedir.

Dalgacık baz fonksiyonları Fourier baz fonksiyonlarına benzemektedir. Genellikle aralarındaki fark izin verdikleri analizin türünde meydana gelmektedir.

Dalgacık Dönüşümünün çok çözünürlüklü (multi-resolition) bir dönüşüm olması, yani bir zaman-frekans analizine izin vermesi aralarındaki en temel farktır. Örneğin sinyallerin analizinde Fourier Dönüşümü kullanıldığında sonuç, sinyalde bulunan frekansların çok hassas bir analizidir, ancak bu frekansların ne zaman meydana geldiği hakkında bilgi yoktur. Dalgacık dönüşümünde, belirli özelliklerin ne zaman oluştuğu ve sinyalin ölçek özellikleri hakkında bilgi alınabilir [60].

Son yıllarda, bir çok sinyal türünden elde edilen ampirik sürekli verileri analiz etmek için dalgacık analizinde birçok gelişme yaşanmıştır. Dalgacık dönüşümünün şimdiye kadarki en yaygın uygulaması veri sıkıştırması için olmuştur. Bu, ayrık Fourier dönüşümünün alt bant ayrışımı ile yakından ilişkilidir. Sinyal işleme, bilgisayar görme, sismoloji, türbülans, bilgisayar grafikleri, görüntü işleme, dijital iletişim, örüntü tanıma, yaklaşım teorisi, nümerik analiz ve istatistik önemli uygulama alanlarıdır.

2.3 Fonksiyonel Verileri Düzgünleştirme

Fonksiyonel veriler tipik olarak $\mathrm{I} = [0, \mathrm{T}]$ kompakt aralığı üzerinde tanımlı bağımsız, reel değerli $\mathrm{x}_1(\mathrm{t}), \ldots, \mathrm{x}_\mathrm{n}(\mathrm{t})$ fonksiyonlarının rasgele bir örneğinden oluşur. Spesifik olarak, xi fonksiyonu $\mathrm{j} = 1, \ldots, n_i$ olmak üzere (t_{ij}, y_{ij}) çiftlerinden oluşur. Verilerin toplandığı T aralığı ve t_{ij} değerleri gözlemden gözleme değişebilir. Normal olarak, x_i fonksiyonel gözlemlerin y_{ij} ayrık verileri kullanarak oluşturulması, her kayıt için ayrı ayrı veya bağımsız olarak gerçekleştirilir.

Fonksiyonel veri analizinin ilk adımı y_{ij} ayrık verilerini herhangi bir t değeri için hesaplanabilen x_i sürekli fonksiyonuna dönüştürmektir. Eğer bu kesikli verilerin hatasız olduğu varsayılırsa süreç interpolasyondur. Fakat verilerde ortadan kaldırılması gereken bazı gözlemsel hatalar varsa, ayrık verilerden fonksiyonel verilere dönüşüm düzgünleştirmeyi içerebilir [61].

İnterpolasyon eldeki veriden bilinmeyen gözlem değerlerini tahmin etmek için kullanılan bir yöntemdir. İstatistik gibi deneylere ve ölçümlere dayalı bilim

dallarında, genelde toplanan verilerin bir fonksiyon eğrisine uydurulması amacıyla kullanılmaktadır ve verinin aşırı heterojen olduğu durumlarda önem kazanmaktadır. Fakat türev bilgisi kullanılmak istendiğinde bu yaklaşım yetersiz kalmaktadır.

Düzgünleştirme ise verilerden daha fazla bilgi elde etmek ve daha esnek ve güçlü analizler yapabilmeye yardımcı olduğundan istatiksel analizde sıklıkla kullanılan bir yöntemdir.

Popüler veri düzgünleştime yöntemleri arasında baz fonksiyonlar, regresyon splaynları, pürüzlülük cezası(ruughness penalty), çekirdek (kernel) düzgünleştiricileri ve splayn düzgünleştiricileri bulunur. Bunlar arasında, regresyon spline veri düzgünleştirme metodu en yaygın kullanılan yöntemdir. Regresyon splayn yöntemi, klasik regresyonda olduğu gibi bir "tasarım matrisi" oluşturmak için splayn adı verilen baz fonksiyonları kullanarak fonksiyonel veri uydurma yaklaşımını ifade eder. Bu eğri sıradan en küçük kareler yöntemi ile elde edilir [62].

Genelleştirilmiş en küçük kareler tahmincisi, baz fonksiyonun düzgünlüğü üzerindeki tüm kontrolü baz fonksiyonlara bırakır. Bu nedenle, düzgünlük kontrolü yalnızca yeterli sayıda ve uygun baz fonksiyon belirtilerek ayrı bir şekilde yapılabilir.

Farklı fonksiyonel veriler için seçilen baz sistemlerinin çeşitli fonksiyonel gözlemlerin özelliklerini ayrı ayrı yansıtması gerektiği unutulmamalıdır. Şu anda, doğru veri uyumu için geniş bir eğri koleksiyonuna otomatik olarak uygulanabilen böyle bir birleşik baz sistem yoktur [1]. Bunun üstesinden gelmek için de genellikle pürüzlü ceza yaklaşımı kullanılmaktadır.

Pürüzlülük cezasına dayalı düzgünleştirme metotları, eğrilerin pürüzlüğünün bir ölçüsünden oluşan bir ceza terimi ile bir baz sistem kullanan yöntemlerdir.

Fonksiyonel veri analizinde sıklıkla kullanılan Fourier, B-Splayn, Üstel, Güçler ve Dalgacık baz sistemleri bir önceki kısımda verilmektedir. Aşağıda ise düzgünleştirme yöntemi olarak baz sistemleri kullanan pürüzlü ceza yöntemi ayrıntılarıyla verilmektedir.

2.3.1 Pürüzlü Ceza Yaklaşımı

Fonksiyonel veriler ile klasik çok değişkenli veriler arasındaki en önemli fark fonksiyonel verilerin düzgünleştirme sürecidir. Düzgünleştirme ile fonksiyonların türevlerindeki bilgiye kolaylıkla erişilebilir.

Pürüzlülük bir x(t) fonksiyonun t deki bir lineer fonksiyondan sapmasının ölçüsüdür [59]. Leurgans, Moyeed & Silverman'a (1993) [63] göre pürüzlülük, x(t) fonksiyonunun değişkenliğinin hızının bir ölçüsüdür.

(a,b) aralığı üzerinde tanımlanmış bir x(t) fonksiyonun pürüzlülüğünü ya da kıvrımlılığını ölçmenin bir çok yolu vardır. Bir x(t) fonksiyonun ikinci mertebeden türevinin karesi $(D^2x(t))^2$o fonksiyonun eğriliğini verir. Toplam eğrilik ise yani bu fonksiyonun ikinci mertebeden türevinin karesinin integrali pürüzlülüğü verir. Bu dünya çapında kabul gören pürüzlülük ölçütüdür [64].

$$PEN_2(x) = \int (D^2x(t))^2 \mathrm{dt} \tag{2.24}$$

Sabit ve doğrusal bir fonksiyonun ikinci mertebeden türevi sıfır olduğundan özellikle regresyon bağlamında, sabit ve doğrusal bir fonksiyon eklenmesi pürüzlülük ölçüsünü değiştirmez [65].

Pürüzlülük cezası yaklaşımı, genelleştirilmiş en küçük karelerin bir uzantısıdır. Ana fikir ortalama hata kare (MSE) yi minimize etmektir. Yanlılık, hata karelerinin toplamı(SSE) olmak üzere ;

$$MSE = Yanlılık^2 + Varyans \tag{2.25}$$

dir [17]. $\boldsymbol{y} = x(\boldsymbol{t}) + \boldsymbol{\varepsilon}$ modeli ele alınsın cezalı artık kareler toplamı şu şekilde tanımlanır.

$$PENSSE_\lambda(x \mid \mathbf{y}) = \{\boldsymbol{y} - x(\boldsymbol{t})\}^T \boldsymbol{W}\{\boldsymbol{y} - x(\boldsymbol{t})\} + \lambda PEN_2(x) \tag{2.26}$$

$PENSSE_\lambda(x)$ yi minimum yapan x değeri bulunarak $PEN_2(x)$ değeri elde edilir. λ düzgünleştirme parametresidir ve artık kareler toplamı ile ölçülen verilere uyum iyiliği ile $PEN_2(x)$ ile ölçülen x fonksiyonunun değişkenliği arasındaki değişim oranını ölçer. **W** simetrik bir ağırlık matrisidir ve ceza terimi olarak $PEN_2(x)$ pürüzlülüğünü kullanır.

Gözlem sayısının çok olduğu durumlarda pürüzsüzlük ve veriye neredeyse mükemmel uyum aynı anda sağlanamaz. λ düzgünleştirme parametresi pürüzsüzlük ve verilere uyum arasındaki kontrolü sağlar. λ büyüdükçe pürüzsüzlük daha çok ön plana çıkar ve daha büyük bir $PEN_2(x)$ pürüzlülük cezası uygulanır. Sonuç olarak x fonksiyonu düzgünleştikçe verilere uyum iyiliği daha az önem kazanır. Diğer yandan λ sıfıra yaklaştıkça pürüzsüzlük daha az önem kazanır ve nerdeyse verilere tam olarak uyan bir fonksiyon elde edilir. Fakat her iki durum da kullanışlı değildir. Bu nedenle asıl zorluk uygun bir lamda değeri bulmaktadır.

De Boor (2002) ve düzgünleştirme üzerine olan bir çok metinde bulunan dikkate değer bir teorem, $PENSSE_\lambda(x \mid \mathbf{y})$ minimize eden x eğrisinin, düğümleri t_j veri noktalarında yer alan kübik bir splayn olduğunu belirtir. Düğümleri veri noktalarına yerleştirmek regresyon splayn kullanımındaki düğümlerin nereye yerleştirileceği sorununu ortadan kaldırır. Düzgünleştirme splaynları doğal olarak örnekleme noktalarının eşit olmayan aralıklarına uyum sağlar. Böylece veri yoğunluğunun yüksek olduğu ve aynı zamanda özellikle az sayıda gözlemin olduğu bölgelerde pürüzsüz olan bölgelerden otomatik olarak yararlanır. Splayn düzgünleştirme için kullanılan en yaygın hesaplama tekniği dördüncü dereceden B-Splayn baz fonksiyon kullanımı ve bu baz fonksiyonunun katsayılarına ilişkin 2.20 deki minimizasyon kriterini kullanmaktır [1].

Bir $x(t)$ fonksiyonunun baz fonksiyon gösterimi

$$x(t) = \sum_{k=1}^{K} c_k \varphi_k(t) = \boldsymbol{c}'\varphi(t) = \varphi'(t)\boldsymbol{c} \tag{2.27}$$

şeklindedir. $\boldsymbol{c}$ K boyutlu katsayılar vektörü ve φ K boyutlu baz fonksiyon vektörüdür. Bu durumda artık kareler toplamını minimize eden katsayılar vektörü

$$\hat{\boldsymbol{c}} = (\boldsymbol{\Phi}'\boldsymbol{W}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\boldsymbol{W}'\boldsymbol{y} \tag{2.28}$$

ile gösterilir. $\boldsymbol{\Phi}$ n tane örneklem noktasındaki K tane baz fonksiyonu içeren nxK boyutlu bir matrisdir. **W** artıklar (örneklem hataları) arasındaki kovaryans yapısına izin veren bir ağırlık matrisidir ve **y** düzgünleştirilecek kesikli verilerin vektörüdür.

$\hat{\boldsymbol{y}} = x(\boldsymbol{t})$ veri uydurma vektörü

$$\widehat{\boldsymbol{y}} = \boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{W}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\boldsymbol{W}\boldsymbol{y} = \boldsymbol{S}_{\varphi}\boldsymbol{y} \tag{2.29}$$

şekildeki gibidir ve $\boldsymbol{S}_{\varphi}$ φ baz fonksiyon yapısına karşılık gelen izdüşüm(projection) operatörüdür ve aynı zamanda düzgünleştirme matrisidir.

$$\boldsymbol{S}_{\varphi} = \boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{W}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\boldsymbol{W} \tag{2.30}$$

$PEN_m(x)$ pürüzlülük cezası matris cinsinden şu şekilde ifade edilir.

$$\begin{aligned} PEN_m(x) &= \int [D^m x(s))]^2 ds \\ &= \int [D^m \boldsymbol{c}'\varphi(s)]^2 ds \\ &= \int \boldsymbol{c}' D^m\varphi(s) D^m\varphi'(s)\boldsymbol{c} ds \\ &= \boldsymbol{c}'\left[\int D^m\varphi(s) D^m\varphi'(s)\, ds\right]\boldsymbol{c} \\ &= \boldsymbol{c}'\boldsymbol{R}\boldsymbol{c} \end{aligned} \tag{2.31}$$

$\boldsymbol{R} = \int D^m\varphi(s) D^m\varphi'(s)\, ds$ ile gösterilen ifade fonksiyonların dış çarpım vektörünün integralidir. Bu durumda ağırlıklandırılmış cezalı artık kareler toplamını şu şekilde ifade edebiliriz:

$$PENSSE_m(\boldsymbol{y} \mid \mathbf{c}) = (\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{c})'W(\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{c}) + \lambda \boldsymbol{c}'\boldsymbol{R}\boldsymbol{c} \tag{2.32}$$

Bu ifadenin her iki tarafının türevi alındığında tahmini katsayılar vektörü $\hat{\boldsymbol{c}}$ elde edilir:

$$\hat{\boldsymbol{c}} = (\boldsymbol{\Phi}'\mathbf{W}\boldsymbol{\Phi} + \lambda\mathbf{R})^{-1}\boldsymbol{\Phi}'\boldsymbol{W}\boldsymbol{y} \tag{2.33}$$

$\widehat{\boldsymbol{y}} = x(\boldsymbol{t})$ veri uydurma vektörü (data fitting vector)

$$\widehat{\boldsymbol{y}} = \boldsymbol{\Phi}(\boldsymbol{\Phi}'\mathbf{W}\boldsymbol{\Phi} + \lambda\mathbf{R})^{-1}\boldsymbol{\Phi}'\boldsymbol{W}\boldsymbol{y} = \boldsymbol{S}_{\boldsymbol{\varphi},\lambda}\boldsymbol{y} \tag{2.34}$$

şeklindedir. $\boldsymbol{S}_{\boldsymbol{\varphi},\boldsymbol{\lambda}}$ ifadesi n boyutlu simetrik şapka matrisini (hat matrix) belirtir.

$$\boldsymbol{S}_{\boldsymbol{\varphi},\boldsymbol{\lambda}} = \boldsymbol{\Phi}(\boldsymbol{\Phi}'\mathbf{W}\boldsymbol{\Phi} + \lambda\mathbf{R})^{-1}\boldsymbol{\Phi}'\boldsymbol{W} \tag{2.35}$$

(2.30) ile(2.35) ifadeleri arasındaki tek fark $\boldsymbol{\Phi}'\mathbf{W}\boldsymbol{\Phi}$ dış çarpım matrisinin tersinin alınmadan önce $\lambda\mathbf{R}$ ifadesinin eklenmesidir. $\lambda = 0$ olduğunda bu iki operatörü birbirine eşit olur.

$$\boldsymbol{S}_{\boldsymbol{\varphi},\boldsymbol{\lambda}} = \boldsymbol{S}_{\boldsymbol{\varphi}}, \qquad \lambda = 0 \tag{2.36}$$

$\boldsymbol{S}_{\boldsymbol{\varphi}}$ düzgünleştirme matrisi simetrik ve idempotent bir izdüşüm operatörü iken $\boldsymbol{S}_{\boldsymbol{\varphi},\boldsymbol{\lambda}}$ idempotent olma özelliği taşımaz. Bu nedenle alt izdüşüm operatörü olarak adlandırılır.

$$\begin{aligned} \boldsymbol{S}_{\boldsymbol{\varphi}}\boldsymbol{S}_{\boldsymbol{\varphi}} &= \boldsymbol{S}_{\boldsymbol{\varphi}} \\ \boldsymbol{S}_{\boldsymbol{\varphi},\boldsymbol{\lambda}}\boldsymbol{S}_{\boldsymbol{\varphi},\boldsymbol{\lambda}} &\neq \boldsymbol{S}_{\boldsymbol{\varphi},\boldsymbol{\lambda}} \end{aligned} \tag{2.37}$$

Pürüzlülük ceza tahmincisi yukarıda da görüldüğü gibi Genelleştirilmiş en küçük kareler tahmincisine benzer bir yapıya sahiptir. Özellikle fonksiyonel verileri sonlu baz fonksiyonların bir açılımı ile temsil ederken eğri tahminine doğal ve esnek bir yaklaşım sağlar. Pürüzlülük cezası yaklaşımına başlamadan önce iki parametre tahmin edilmelidir: λ düzgünleştirme parametresi ve gerekirse cezalandırma için kullanılan doğrusal diferansiyel operatör [59].

2.3.1.1 Düzgünleştirme Operatörünün Belirlenmesi

Düzgünleştirme parametresi λ yı seçme sorununa karşı Green & Silverman (1994) [65], iki farklı felsefi yaklaşımı öne sürmüşlerdir. İlk yaklaşım, düzgünleştirme parametresini öznel olarak seçmektir. Farklı ölçeklerde ortaya çıkan verilerin lambda ya göre değişimi incelenir ve içlerinden en iyi görüneni düzgünleştirme parametresi olarak seçilir.

İkinci yaklaşım, düzgünleştirme parametresi değerinin veriler tarafından seçildiği otomatik bir yöntemdir. Her istatistiksel prosedürde olduğu gibi, yöntemin kendisinin seçiminde keyfi kararlar olduğundan böyle bir yöntem için objektif yerine otomatik kelimesini kullanmak daha doğrudur. Bununla birlikte, bu yaklaşım kullanılan otomatik yönteme bağlı olarak, düzgünleştirme parametresi seçiminin gerçekten objektif olduğu durumdur. Özellikle fonksiyonel veri bağlamında büyük veri setlerinin kullanıldığı durumlarda karşılaştırılabilir sonuçlar elde etmek için otomatik bir yöntem şarttır.

Düzgünleştirme parametresi seçimi için bir çok yöntem vardır. Bunlar arasında en çok tercih edilen yöntemler Çarpraz Geçerlilik (Cross-Validation, CV) ve Genelleştirilmiş Çarpraz Geçerlilik (Generalized Cross-Validation, GCV)

yöntemleridir. Düzgünleştirme parametresi seçimine rehberlik etmek için çapraz geçerlilik kullanımına ilişkin önemli bir referans Craven & Wahba'dır (1979) [66].

Çarpraz geçerlilik yöntemi verilerin bir kısmını doğrulama örneği olarak ayırıp veriyi geriye kalan kısma uydurmayı içerir. Bu şekilde, modelin modeli tahmin etmek için kullanılmayan verilere ne kadar iyi uyduğu tespit edilir. Doğrulama örneği olarak sadece bir gözlem değeri ayrıldığında veri geri kalan gözlem değerlerine uydurulur ve başta ayırılan gözlem için uygun bir değer tahmin edilir. Bu prosedür bütün gözlem değerleri için tekrarlandığında ortaya çıkan hataların kareleri toplamı alınırsa çarpraz geçerlilik ölçütü elde edilir. Bu kriter bir dizi lambda değeri için hesaplanır ve minumum değeri veren lambda değeri seçilir.

m_λ^{-i} , $X_i(t)$ hariç λ düzgünleştirme parametresi tarafından hesaplanan düzgünleştirilmiş örneklem ortalaması olsun . m_λ^{-i} nin $X_i(t)$ yi ne kadar iyi tahmin ettiğini görmek için global uyum iyiliği ölçüsü olan entegre hata kareleri (integrated squared error (ISE)) hesaplanır.

$$ISE\{m_\lambda^{-i}\} = \int \{m_\lambda^{-i}(t) - X_i(t)\}^2 dt \tag{2.38}$$

Çarpraz geçerlilik ölçütü $i = 1,2, \dots, n$ olmak üzere her bir i için hesaplanan entegre hata kareleri toplamı ile elde edilir ve şu şekilde gösterilir.

$$CV(\lambda) \sum_{\text{İ}=1}^{n} \int \{m_\lambda^{-i}(t) - X_i(t)\}^2 dt \tag{2.39}$$

Çapraz geçerlilik yöntemi, gözlemlerin birbirinden nispeten bağımsız olduğu varsayımına dayanır ve çok çeşitli durumlarda kullanılabilir. Ancak, yöntemin iki sorunu vardır. Birincisi, genellikle hesaplama yoğunluğu vardır ve binlerce örneklem büyüklüğü için uygun olabilecek bir şey değildir. İkinci sorun, CV'nin en aza indirilmesinin verilerin yetersiz düzgünleşmesine sebep olabilmesidir [1].

Çarpraz geçerlilik yöntemi çok zaman alıcı olduğundan ve verilerin yetersiz düzgünleşmesine sebep olduğundan Eubank (1988) [64], Genelleştirilmiş Çarpraz Geçerlilik Yöntemini ortaya koymaktadır.

Craven and Wahba (1979) [66] tarafından geliştirilen GCV, başlangıçta çarpraz geçerlilik yönteminin daha basit bir versiyonu olarak geliştirildi ve n kez yeniden

düzgünleştirme ihtiyacını ortadan kaldırdı. GCV yöntemi yetersiz düzgünleşmeye sebebiyet verme açısından CV yöntemiyle kıyaslandığında daha güvenilir bulunmuştur ve genellikle şu şekilde ifade edilir [1].

$$GCV(\lambda) = \frac{nSSE(\lambda)}{(n - df_{\lambda})^2} \tag{2.40}$$

df_{λ} , $\widehat{X}_i(t) = \Phi_i(t)\widehat{\boldsymbol{c}}_i$ nin $X_i(t)$ yi tahmin etmede kullandığı etkili parametre sayısı veya serbestlik derecesidir.

CV (λ), GCV (λ) ve diğer λ düzgünleştirme parametresi seçim yöntemleri ile ilgili daha ayrıntılı ve teorik bilgi için Gu(2003) [67] ü inceleyiniz.

3
FONKSİYONEL TEMEL BİLEŞEN ANALİZİ

Temel bileşen analizi (Jolliffe 2002, [68]), fonksiyonel verilere genişletilen ve fonksiyonel temel bileşen analizi (FTBA) olarak adlandırılan çok değişkenli veriler için önemli bir boyut küçültme aracıdır. Temel fikirler Grenander (1950) [56], Karhunen (1946) [69], Loeve (1946) [70] ve Rao (1958) [57] e dayandığı düşünülmektedir. Fakat FTBA için daha kapsamlı bir çerçeve ilk olarak Toulouse Üniversitesi'nde yayımlanan Dauxois ve Pousse (1976) tezinde oluşturuldu. O zamandan beri, bu yaklaşım fonksiyonel veri analizindeki en yaygın araç haline geldi. Bunun nedeni FTBA'nin doğal olarak sonsuz boyutlu fonksiyonel verilerini rastgele skorların sonlu boyutlu bir vektörüne dönüştürülmesini kolaylaştırmasıdır [58].

Ramsay ve Dalzell'in (1991) [20] vurguladığı gibi "sadece sınırlı sayıda gözlem olmasına rağmen bazı modelleme problemlerini fonksiyonel olarak düşünmek daha doğaldır". Öte yandan, temel bileşenler gibi klasik boyut küçültme teknikleri, bilgiyi fazla kayıp olmadan sıkıştırmamıza izin vermeleri bakımından önemlidir.

Fonksiyonel veriler için temel bileşenler analizi bir çok nedenden ötürü dikkate alınması gereken kilit bir tekniktir. Fonksiyonel verilerin kaydedilmesi görüntülenmesi gibi ön adımlardan sonra bu fonksiyonları karakterize eden özellikleri görmek için verilerin incelenmesi gerekir. Bu amaçla, fonksiyonel verilerin varyans yapısını gözlemlemek için başvurulan teknik fonksiyonel temel bileşen analizi (FTBA) dir.

Klasik çok değişkenli durumda olduğu gibi, varyans-kovaryans ve korelasyon fonksiyonlarının yorumlanması zor olabilir ve bu fonksiyonlar gözlemlenen verilerdeki değişkenlik yapısının her zaman tam olarak anlaşılabilir bir gösterimini veremeyebilirler. Temel bileşenler analizi, varyans-kovaryans fonksiyonunun doğrudan incelenmesini ve çok daha bilgilendirici olabilen kovaryans yapısına bakmanın bir yolunu sunar [1].

Temel bileşen analizi çok değişkenli regresyon durumunda eğer değişkenler birbirleriyle ilişkili ise eş-doğrusallıktan kaynaklanan sorunlardan kaçınmak için kullanılan bir tekniktir [51]. Bu durum fonksiyonel veriler için de geçerlidir.

Çok değişkenli temel bileşen analizine giriş bölümünde değinilmiştir. Bu bölümde fonksiyonel temel bileşen analiziyle ilgili teorik bilgiler verilecektir.

3.1 Fonksiyonel Temel Bileşen Analizi (FTBA)

Çok değişkenli verilerde olduğu gibi fonksiyonel verilerde de, veride güçlü bir şekilde temsil edilen varyasyon türlerini belirtmek veya görüntülemek için ağırlıklar seçilir. Her bir temel bileşen, fonksiyonel verilerle aynı t aralığında tanımlanan bir temel bileşen ağırlık fonksiyonu $\xi(t)$ ile belirtilir. f_i temel bileşen skorlarının lineer(doğrusal) bir kombinasyonudur ve şu şekilde gösterilir:

$$f_i = \sum_{j=1}^{p} \xi_j x_{ij} \ , \ \text{klasik TBA} \tag{3.1}$$

$$f_i = \int \xi(t) x_i(t)\, dt \ , \quad \text{fonksiyonel TBA} \tag{3.2}$$

Fonksiyonel temel bileşenler analizinde x_{ij} değişken değerleri $x_i(t)$ fonksiyonlarıyla, j kesikli indeksi t sürekli indeksle, ξ_j temel bileşen ağırlıkları $\xi(t)$ fonksiyonuyla ve toplam notasyonu integral notasyonuyla yer değiştirir.

FTBA nın ilk adımı,

$$\int \xi_1(t)^2 dt = \| \xi_1 \|^2 = 1 \tag{3.3}$$

kısıtı altında f_i temel bileşen skorlarının varyansını maksimum yapacak $\xi_1(t)$ ağırlık fonksiyonunu bulmaktır. Bir diğer deyişle $\xi_1(t)$ ortalama kareyi (mean square)

$$N^{-1} \sum_i f_{i1}^2 = N^{-1} \sum_i (\int \xi_1 x_i)^2 \tag{3.4}$$

en üst düzeye çıkaran ağırlık fonksiyonudur [25].

FTBA uygulanmadan önce genellikle her bir fonksiyonun ortalaması karşılık gelen fonksiyon değerlerinden çıkarılır. Böylece, temel bileşen skorlarının ortalama karesinin maksimize edilmesi, örneklem varyanslarının maksimize edilmesine karşılık gelir. Buradan 3.2 de verilen temel bileşen skorlarının doğrusal bir birleşimi olan f_i ve f_i nin varyansı şu şekilde verilir.

$$f_i = \langle \xi_i, x - Ex \rangle = \int \xi_i(t) \{x(t) - Ex(t)\} dt \quad (3.5)$$

$$Var\, f_i = Var\langle \xi_i, x - Ex \rangle = \iint \xi_i(s) Cov(s,t) \xi_i(t) ds dt \quad (3.6)$$

FTBA daki bu ilk adımın amacı ortalama kareyi en üst düzeye çıkararak değişkenlerdeki en güçlü ve en önemli varyasyon modunu tanımlamaktır[1].

İkinci temel bileşen de aynı şekilde temel bileşen skorlarını maksimum yapacak şekilde ve (3.3) kısıtına ek olarak

$$\int \xi_2(t) \xi_1(t) dt = 0 \quad (3.7)$$

kısıtı altında tanımlanır. Bu kısıt j.ninci temel bileşen için şu şekilde genelleştirilir:

$$\int \xi_j(t) \xi_1(t) dt = \int \xi_j(t) \xi_2(t) dt = \cdots = \int \xi_j(t) \xi_{j-1}(t) dt = 0 \quad (3.8)$$

Bu kısıt ile tahmini temel bileşenlerin karşılıklı olarak birbirlerine dik(ortogonal) olması sağlanır.

Çok değişkenli veri analizinde olduğu gibi fonksiyonel veriler için de temel bileşen analizi, kovaryans veya korelasyon matrisinin özdeğerlerini ve özvektörlerini bulma görevi olarak tanımlanabilir. Bu durumda özvektör yerine özfonksiyon ifadesi kullanılır ve özdenklem şu şekilde gösterilir:

$$\int v(s,t) \xi(t) dt = \rho \xi(s) \quad (3.9)$$

$v(s,t)$ ile gösterilen ifade kovaryans fonksiyonunu temsil eder ve aşağıdaki gibi tanımlanır.

$$v(s,t) = N^{-1} \sum_{i=1}^{N} x_i(s) x_i(t) \quad (3.10)$$

$\xi_1, \xi_2, \ldots$ öz fonksiyonlarına $\rho_1, \rho_2, \ldots$ özdeğerleri karşılık gelir ve bu özdeğerler arasında $\rho_1 \geq \rho_2 \geq \cdots$ olacak şekilde bir ilişki vardır. Böylece 3.6 da verilen temel bileşen skorlarının varyansını maksimum yapan $\xi(t)$ ağırlık fonksiyonunu bulma problemi, artık 3.9 da verilen özdenkleme karşılık gelen en büyük ρ özdeğeri bulma problemine dönüşür. Bu durumda özdeğerler ve öz fonksiyonlar arasında şu şekilde bir bağlantı gerçekleşir [1].

$$\rho_i = Var(f_i) = Var\left[\int \xi_i(t)\{x(t) - Ex(t)\}dt\right] = \iint \xi_i(s)Cov(s,t)\xi_i(t)dsdt \quad (3.11)$$

3.9 da verilen özdenklem V kovaryans operatörü olarak adlandırılan integral dönüşümüyle kısaca şu şekilde gösterilir.

$$V\xi = \int v(.,t)\xi(t)dt \quad (3.12)$$

$$V\xi = \rho\xi \quad (3.13)$$

Çok değişkenli ve fonksiyonel özanaliz problemleri arasında, özdeğer-özfonksiyon çiftlerinin maksimum sayısı açısından önemli bir fark vardır. Çok değişkenli durumda p değişken sayısına karşılık , fonksiyonel durumda fonksiyon değerleri yani sonsuzluk karşılık gelir. Bununla birlikte, xi fonksiyonları lineer bağımlı değilse, V operatörünün rankı N - 1 olur ve sadece N - 1 tane sıfır olmayan özdeğere sahip olur [1].

3.1.1 FTBA İçin Baz Fonksiyon Yaklaşımı

3.9 da verilen $\int v(s,t)\xi(t)dt = \rho\xi(s) = \int Cov(s,t)\xi(t)dt = \rho\xi(s)$ özdenklemini matris formuna dönüştürmenin bir yolu, her bir x_i fonksiyonunu φ_k baz fonksiyonlarının lineer bir birleşimi olarak ifade etmektir. Bu durumda her bir x_i fonksiyonunun baz fonksiyon genişlemesine sahip olduğu varsayımı altında şu şekilde ifade edilebilir.

$$x_i = \sum_{k=1}^{K} c_{ik}\varphi_k \text{ i=1,2,...,N} \quad (3.14)$$

K baz fonksiyonlarının sayısı bir çok hususa bağlıdır. Örneğin, orijinal verideki kaç tane kesikli örneklem noktası(n) olduğuna, K<n olmak üzere düzgünleştirme kullanıp kullanılmayacağına, baz fonksiyonları orijinal fonksiyonların davranışlarını yansıtmada ne kadar etkili ve güçlü olduğu gibi hususlara bağlıdır. Örneğin aylık sıcaklık verileri üzerinde çalışıldığında, [0,12] aralığı üzerinde tanımlı ortonormal Fourier serilerini kullanmak oldukça mantıklıdır. Eğri başına sadece 12 örnekleme noktası olduğundan, K=12 aylık sıcaklık verileri için baz fonksiyonların mümkün olan maksimum boyutudur [1].

(3.14) de verilen ifade $\boldsymbol{x} = (x_1, x_2, \dots, x_N)$ ve $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \dots, \varphi_K)$ olacak şekilde vektör değerli $\boldsymbol{x}$ ve $\boldsymbol{\varphi}$ fonksiyonları tanımlandığında kısaca şu şekilde ifade edilebilir.

$$\boldsymbol{x} = \boldsymbol{C\varphi} \tag{3.15}$$

Bu ifadede **C**, (NxK) boyutlu katsayılar matrisidir. Kovaryans fonksiyonu da aynı şekilde matris formunda daha kompakt bir şekilde ifade edilebilir:

$$Cov(s,t) = N^{-1}\boldsymbol{\varphi}(s)'\boldsymbol{C}'\boldsymbol{C\varphi}(t) = \boldsymbol{\varphi}(s)'Cov\boldsymbol{C\varphi}(t) \tag{3.16}$$

Burada $\boldsymbol{\varphi}(s)'$ ifadesi $\boldsymbol{\varphi}(s)$ baz fonksiyonlarının transpozunu temsil etmektedir.

(3.9) da verilen özdenklemde ξ öz fonksiyonları da baz fonksiyon yaklaşımıyla

$$\xi(t) = \sum_{k=1}^{K} b_k \varphi_k(t) \tag{3.17}$$

ve $\xi(t) = \boldsymbol{\phi}(t)'\boldsymbol{b}$ matris gösterimiyle ifade edilebilir. Matris notasyonuyla ifade edilen öz fonksiyon ve kovaryans fonksiyonu özdenklemde yerine yazıldığında

$$\begin{aligned} \int Cov(s,t)\xi(t)dt &= \int \boldsymbol{\varphi}(s)'Cov\boldsymbol{C\varphi}(t)\boldsymbol{\phi}(t)'\boldsymbol{b}dt \\ &= \boldsymbol{\varphi}(s)'\boldsymbol{CovCWb} \\ \boldsymbol{\varphi}(s)'Cov\boldsymbol{CWb} &= \boldsymbol{\rho\varphi}(s)'\boldsymbol{b} \end{aligned} \tag{3.18}$$

ifadesi elde edilir ve $\boldsymbol{W} = \int \varphi\phi'$ olacak şekilde K boyutlu simetrik bir matris olmak üzere bu durumda öz denklem artık kısaca şu şekilde ifade edilebilir.

$$\boldsymbol{CovCWb} = \boldsymbol{\rho b} \tag{3.19}$$

$\| \xi \| = 1$ olduğundan $\boldsymbol{b'Wb} = 1$ olur. Benzer şekilde ξ_1 ve ξ_2 fonksiyonları yalnızca katsayılar vektörlerine karşılık gelen $\boldsymbol{b_1'Wb_2} = 0$ olması ile ortogonal olurlar. Temel bileşenleri elde etmek için $\boldsymbol{u} = \boldsymbol{W}^{1/2}\boldsymbol{b}$ ifadesi tanımlanır, bu ifadeye karşılık gelen

$$\boldsymbol{W}^{1/2}Cov\boldsymbol{C}\boldsymbol{W}^{1/2}\boldsymbol{u} = \boldsymbol{\rho u} \tag{3.20}$$

simetrik özdeğer problemi çözülür ve her bir özvektör için $\boldsymbol{b} = \boldsymbol{W}^{-1/2}\boldsymbol{u}$ ifadesi hesaplanır. Özetle algoritma şu şekildedir [1]:

- Birinci adım olarak $\boldsymbol{C}$ ve $\boldsymbol{W}$ hesaplanır.

- Cholesky ayrıştırma yöntemi kullanılarak $\boldsymbol{W}^{1/2}$ hesaplanır.
- $\boldsymbol{W}^{1/2}Cov\boldsymbol{C}\boldsymbol{W}^{1/2}$ simetrik özdeğer matrisi kullanılarak özdeğer ve özvektörler($\mathbf{u}$) elde edilir.
- $\boldsymbol{b} = \boldsymbol{W}^{-1/2}\boldsymbol{u}$ ifadesi hesaplanır ve $\xi(t) = \boldsymbol{\phi}(t)'\boldsymbol{b}$ denkleminde yerine yazılarak temel bileşen ağırlık fonksiyonları elde edilir

Baz fonksiyonlar ortonormal ise, yani $\boldsymbol{W} = \boldsymbol{I}$ durumunda fonksiyonel TBA problemi $\mathbf{C}$ katsayılar dizisinin standart çok değişkenli TBA'sına indirgenir ve sadece K mertebeli simerik Cov$\mathbf{C}$ dizisinin özanalizi hesaplanır. Her durumda, prensip olarak baz fonksiyon yaklaşımı ile hesaplanabilecek maksimum özfonksiyon sayısının, bazın boyutu olan K olduğu unutulmamalıdır [1].

3.1.2 Çok Değişkenli Fonksiyonel Temel Bileşen Analizi

Bir çok uygulama alanındaki teknik ilerlemelerle birlikte, birim gözlem başına düşen fonksiyonel veri sayısının birden fazla olduğu verilerle karşılaşılmaktadır. Bu durum çok değişkenli fonksiyonel verilerle çalışmayı motive eden bir durumdur.

Birden fazla fonksiyonu ayrı ayrı incelemek yerine, bir bütün olarak ele alıp, bu fonksiyonların eş zamanlı varyasyonunu incelemenin bir yolu çok değişkenli fonksiyonel temel bileşen analizidir(ÇFTBA). Burada dikkat edilmesi gereken iki husus vardır. İlk husus, bütün fonksiyonlar aynı argümana göre örneğin zamana göre ele alınmış olmalıdır. İkinci husus ise genellikle bütün fonksiyonlar aynı birim cinsinden örneğin dereceye göre ölçülmüş olmalıdır [1]. Farklı birimlerde ölçülen veriler için Jacques ve Preda (2014) [71] ve Chiou, Yang ve Chen (2014) [30], normalize edilmiş bir kovaryans operatörüne dayalı ÇFTBA nın normalize edilmiş bir versiyonunu tartışmaktadır.

ÇFTBA için mevcut yaklaşımlar, sonlu, tek boyutlu aralıkta gözlemlenen fonksiyonlarla sınırlıdır (Ramsay ve Silverman 2005 [1]; Berrendero, Justel ve Svarc 2011 [29]; Jacques ve Preda 2014 [71]; Chiou, Yang ve Chen 2014 [30]). Berrendero, Justel ve Svarc (2011) [29] hariç hepsi, verilerin çok değişkenli fonksiyonel Karhunen-Loève temsiline yönelmektedir [72].

Çok değişkenli fonksiyonel TBA da veriler elemanları fonksiyonlar olan bir matris formunda düzenlenir ve her birim de bu fonksiyonların bir vektörü şeklindedir. Tüm fonksiyonların $\mathbb{R}$ de değer aldığı ve aynı reel kompakt bir J aralığı üzerinde

tanımlandığı varsayılır. Çok değişkenli fonksiyonel veriler için i. inci bireysel fonksiyona karşılık gelen j.ninci vektör değerli fonksiyonlar nxp boyutlu bir matrisi oluşturur ve şu şekilde temsil edilir:

$$\chi : x_{ij}(t), \quad , \quad \text{i=1,2,....N , j=1,2,...,p} \quad t \in J \tag{3.21}$$

Veri matrisinin her bir satırı (Ω, F, P) olasılık uzayı üzerinde tanımlı p boyutlu $X \coloneqq (X_1, X_2, \dots, X_p)'$ stokastik süreci ile modellenir. Genelliği kaybetmeden her $t \in J$ için $X(t) \coloneqq (X_1(t), X_2(t), \dots, X_p(t))'$ rassal vektörün ortalaması sıfıra eşittir $(\mu(t) = 0)$. $\mu(t) \neq 0$ durumunda $X(t)$ yerine $\tilde{X}(t) \coloneqq X(t) - \mu(t)$ ifadesi kullanılır ve $X(t)$ nin pozitif tanımlı

$$\sum(t) = X(t)X(t)' \tag{3.22}$$

kovaryans matrisine sahip olduğu varsayılır. Buradan $t \in J$ olmak üzere $\zeta : J \to \mathbb{R}^P$ herhangi bir ζ fonksiyonu için

$$Var[\zeta(t)'X(t)] = \zeta(t)'VarX(t)\zeta(t) = \zeta(t)' \sum(t)\, \zeta(t) \tag{3.23}$$

eşitliği sağlanır. Bu durumda $\| \zeta(t) \| = 1$ kısıtı altında ζ_1 ağırlık fonksiyonu

$$\int \zeta'(t) \sum(t)\, \zeta(t)\, \mathrm{dt} \tag{3.24}$$

temel bileşen fonksiyonlarının varyansını maksimum yapan fonksiyon olarak tanımlanır. X in birinci temel bileşen fonksiyonu $Z_1 = \zeta_1(t)'X(t)$ şeklindedir ve $r \geq 2$ olmak üzere r.ninci temel bileşen fonksiyonu

$$\| \zeta(t) \| = 1 \quad \zeta(t)\zeta_l(t) = 0 \ , l = 1,2, \dots, r-1 \tag{3.25}$$

kısıtları altında şu şekilde gösterilir:

$$Z_r = \zeta_r(t)'X(t) \quad r = 2, \dots, p \tag{3.26}$$

Her $t \in J$ için $\lambda_1(t) > \cdots > \lambda_p(t) > 0$ olmak üzere $\sum(t)$ kovaryans matrisinin özdeğerleri olsun. Bu durumda

- Her $t \in J$ ve $r = 1,2, \dots, p$ için $\zeta_r(t)$, $\lambda_r(t)$ özdeğerine karşılık gelen birim normlu bir özvektörüdür.

- Temel bileşen fonksiyonlarının iç çarpımı

$$\langle Z_r, Z_s \rangle \coloneqq E[\int Z_r(t) Z_r(t)\, dt] \tag{3.27}$$

şeklinde tanımlıdır ve $r \neq s$ için $\langle Z_r, Z_s \rangle = 0$ dır.

- $\| Z_r \| = (E[\int Z_r(t)^2 dt])^{1/2}$ olmak üzere $\| Z_r \|^2 = \int \lambda_r(t) dt$ dir [1].

Çok değişkenli FTBA da izlenilen yol FTBA dakiyle aynıdır. Bu nedenle Bölüm 3.1 de verilen ifadelere bu bölümde kısaca değinilmiştir Daha ayrıntılı bilgi için Berrendero , Justel ve Svarc(2011) [29] ve Happ ve Greven(2018) [72] nin çalışmalarını inceleyiniz.

3.2 Düzgünleştirilmiş Fonksiyonel Temel Bileşen Analizi

Fonksiyonel temel bileşen analizi ile elde edilen özfonksiyonlar genellikle çok pürüzlüdür. Bu fonksiyonları düzgünleştirmek daha kararlı ve daha iyi yorumlanabilir sonuçlara yol açabilir. Bu fonksiyonları düzgünleştirmek için genellikle popüler bir düzgünleştirme metodu olan pürüzlülük cezası(roughness penalty) yaklaşımı uygulanır. Bu tekniğin dezavantajı L^2 anlamında ortogonalliğin kaybedilmesidir [73].

Standart eğri düzgünleştirmesinde olduğu gibi ξ bir temel bileşen eğrisi olduğu varsayılışa ξnin pürüzlülüğü ikinci türevin karesinin integrali $PEN_2(\xi) = \| D^2 \xi \|^2$ ile cezalandırılır.

FTBA nın ilk adımı $\| \xi \|^2 = 1$ kısıtı altında $\int \xi x_i$ temel bileşen skorlarının varyansını maksimum yapacak birinci temel bileşen ağırlık fonksiyonunu bulmaktır ve bu maksimum varyans problemi Bölüm 3.1 de belirtildiği gibi $V\xi = \rho\xi$ özdenklemi ile çözülür.

Düzgünleştirilmiş temel bileşen fonksiyonları elde etmenin bir yolu, ξ bir temel bileşen fonksiyonu olmak üzere , $\| \xi \|^2 = 1$ kısıtı altında $\int \xi x_i$ temel bileşen örneklem varyansının $\{1 + \lambda \times PEN_2(\xi)\}$ ile bölünmesidir. Bu durumda cezalı temel bileşen örneklem varyansı elde edilir ve şu şekilde gösterilir:

$$TBAPSV(\xi) = \frac{Var \int \xi x_i}{\| \xi \|^2 + \lambda \times PEN_2(\xi)} = \frac{\iint \xi_i(s) Cov(s,t) \xi_i(t) ds dt}{\int \xi_i(t)^2 \, dt + \lambda \int (D^2 \xi_i(t))^2 dt} \quad (3.28)$$

Burada amaç sadece varyansı maksimum yapmak değil aynı zamanda ξ temel bileşen ağırlık fonksiyonunun $PEN_2(\xi) = \int \xi''(t)^2 \, dt$ pürüzlülüğünün de çok büyük olmasını önlemektir. Pürüzlülük cezası yaklaşımı ile bu olası çatışma durumu ortaya çıkar ve bu durum $\lambda \geq 0$ düzgünleştirme parametresi ile kontrol edilir.

λ düzgünleştirme parametresini sabit tutup ξ temel bileşen fonksiyonunun pürüzlülüğünün arttırılması $PEN_2(\xi)$ ifadesini arttıracağından, TBAPSV(ξ) cezalı temel bileşen örneklem varyansını azaltır. Hatta $\lambda \to 0$ durumunda (3.28) deki eşitlik

$$TBAPSV(\xi) = Var \int \xi x_i \quad (3.29)$$

şeklinde olacağından cezalı temel bileşen varyansı temel bileşenin varyansına eşit olur. Diğer yandan λ ne kadar büyük olursa cezalı temel bileşen varyansı ξ nin pürüzlülüğünden o kadar etkilenir. $\lambda \to \infty$ durumunda $a$ve b sabitler olmak üzere ξ bileşeni periyodik durumda $\xi = a$ periyodik olmayan durumda $\xi = a + bt$ formuna dönüşür [1].

ξ_m J aralığı üzerinde tanımlı, dördüncü mertebeye kadar türevi olan karesi integrallenebilir bir fonksiyon olduğu varsayıldığında $D^2\xi$ ve $D^3\xi$ aşağıdaki koşullardan birini sağlar [73].

1. J aralığının uç noktalarında $D^2\xi$ ve $D^3\xi$ sıfıra eşittir.
2. ξ , $D\xi$, $D^2\xi$ ve $D^3\xi$ periyodik sınır koşullarını sağlar.(Periyodik sınır koşulları, genellikle birim hücre adı verilen küçük bir parça kullanılarak büyük (sonsuz) bir sisteme yaklaşmak için seçilen bir dizi sınır koşuludur.

Bu durumda a ve b J aralığının sınırları olmak üzere pürüzlü ceza şu şekilde yazılabilir[73]:

$$\| D^2\xi \|^2 = \int D^2\xi(t)D^2\xi(t)dt \tag{3.30}$$

$$= D\xi(b)D^2\xi(b) - D\xi(a)D^2\xi(a) - \int D\xi(t)D^3\xi(t)dt \tag{3.31}$$

$$= D\xi(b)D^3\xi(b) - D\xi(a)D^3\xi(a) - \int \xi(t)D^4\xi(t)dt \tag{3.32}$$

$$= \langle \xi, D^4\xi \rangle \tag{3.33}$$

3.31 ve 3.32 da verilen ifadelerin ilk iki terimi yukarıda belirtilen her iki koşula göre de sıfırdır. Bu durumda pürüzlü ceza $PEN_2(\xi) = \langle \xi, D^4\xi \rangle$ ifadesi ile de gösterilebilir. Böylece

$$\begin{aligned} 1 + \lambda \times PEN_2(\xi) &= 1 + \lambda\langle \xi, D^4\xi \rangle \\ &= \int \xi(t)(\mathrm{I} + \lambda D^4)\xi(t)dt \end{aligned} \tag{3.34}$$

eşitliği sağlandığından 3.28 de ifade edilen cezalı temel bileşen örneklem varyansı şu şekilde gösterilir:

$$TBAPSV(\xi) = \frac{\iint \xi_i(s)Cov(s,t)\xi_i(t)dsdt}{\int \xi(t)(\mathrm{I} + \lambda D^4)\xi(t)dt} \tag{3.35}$$

Cezalı temel bileşen örneklem varyansı TBAPSV nin maksimumu,

$$\int Cov(s,t)\xi(t)dt = \rho(\mathrm{I} + \lambda D^4)\xi(s) \tag{3.36}$$

fonksiyonel özdenkleminin en büyük özdeğerine karşılık gelen bir özfonksiyondur. Bu özdenklemin çözülmesi ile tüm özfonksiyonlar diğer bir deyişle tüm temel bileşen ağırlık fonksiyonları elde edilebilir.

Yukarıda belirtildiği gibi, sonuçta ortaya çıkan ağırlık fonksiyonları artık L^2 anlamında ortonormal değildir. Bu nedenle ağırlık fonksiyonları, temel bileşen fonksiyonlarının düzgünleştirilmiş tahmin edicileri olarak kullanıldığından, bu fonksiyonlar $\| \xi \|^2 = 1$ olacak şekilde yeniden ölçeklendirilmelidir. Ağırlık fonksiyonları ξ_j için Sobolev tipi modifiye skaler iç çarpımı

$$(f, g) \stackrel{\text{def}}{=} \langle f, g \rangle + \lambda\langle D^2 f, D^2 g \rangle \tag{3.37}$$

uygulanarak ortogonallik şartı sağlanır [73].

İkinci ve daha yüksek dereceden temel bileşen fonksiyonlarının $\| \xi \|^2 = 1$ kısıtına ek olarak ikinci bir kısıt olan ortogonallik şartını sağlaması için modifiye edilmiş ortogonallik formu uygulanır:

$$\int \xi_i(t)\xi_j(t)dt + \lambda \int D^2\xi_i(t)D^2\xi_j(t)dt, \quad \text{j} = 1,2,\dots,\text{i} - 1 \tag{3.38}$$

Modifiye edilmiş ortogonallik koşulunun kullanılması, tek bir özdeğer problemini çözerek gerekli tüm temel bileşenlerin tahminlerinin bulunabileceği anlamına gelmektedir [1].

Düzgünleştirme parametresinin belirlenmesi: Düzgünleştirilmiş FTBA nde temel sorunlardan biri de λ düzgünleştirme parametresinin belirlenmesidir. Bunun için otomatik bir yöntem olan çarpraz geçerlilik yöntemi kullanılabilir. Bu yaklaşım $\xi_1, \xi_2, \dots, \xi_m$ temel bileşen fonksiyonlarının bir baz fonksiyon olarak görülmesine dayanmaktadır.

Bunun için aşağıdaki adımlar takip edilmelidir [1]:

1. İlk olarak x_i gözlemlerinden genel ortalama çıkartılır.
2. Belirli bir λ düzgünleştirme parametresi için, $\xi_j^{[i]}(\lambda)\ j = 1,2,\dots,m$ olmak üzere x_i hariç tüm veriden elde edilen ξ_j nin tahmini olarak ifade edilir
3. $\zeta_m^{[i]}(\lambda)$, $\xi_j^{[i]}(\lambda)$ nın alt uzayına ortogonal x_i nin bileşeni olarak tanımlanır.
4. Çarpraz geçerlilik skorlarını elde etmek için $\zeta_m^{[i]}(\lambda)$

$$CV_m(\lambda) = \sum_{i=1}^{n} \| \zeta_m^{[i]}(\lambda) \| \tag{3.39}$$

şeklinde yazılır böylece çarpraz geçerlilik ölçütü aşağıdaki gibi elde edilir.

$$CV(\lambda) = \sum_{m=1}^{\infty} CV_m(\lambda) \tag{3.40}$$

n tane eğri için en fazla n-1 tane temel bileşen tahmin edildiğinden bu toplamın m=n-1 de kesilmesi gerekir.

5. Son olarak $CV(\lambda)$ ifadesini minumum yapan λ değeri düzgünleştirme parametresi olarak belirlenir.

3.2.1 Düzgünleştirilmiş FTBA İçin Baz Fonksiyon Yaklaşımı

Düzgünleştirilmiş temel bileşenler, uygun baz fonksiyonlarıyla çalışılarak kolaylıkla elde edilebilir. Bölüm 2 de baz fonksiyonlar ayrıntılarıyla ele alınmıştı. Baz fonksiyonlar periyodik durumlar ve periyodik olmayan durumlar olarak iki ayrı şekilde ele alınabilir. Bu nedenle Düzgünleştirilmiş FTBA için baz fonksiyon yaklaşımı iki ayrı şekilde ele alınacaktır.

Periyodik Durumlarda Baz Fonksiyon Yaklaşımı: Periyodik durumlarda algoritma kurmanın en kolay olduğu baz fonksiyonlar Fourier serileridir. $\boldsymbol{T}$ [0,1] aralığı olmak üzere, periyodik sınır koşulları tüm fonksiyonlar için geçerli olduğu varsayıldığında, bu $x_i(s)$ verilerinin periyodik olduğu anlamına gelmektedir. $\{\varphi_m\}$,

$$c_0 + c_1 \sin(wt) + c_2 \cos(wt) + c_3 \sin(2wt) + c_4 \cos(2wt) + \cdots \tag{3.41}$$

$w_{2j-1} = w_{2j} = 2\pi j$ şeklinde tanımlanan Fourier fonksiyonlarının bir serisi olsun. Herhangi bir periyodik fonksiyon x verildiğinde $c_m = \int x\varphi_m$ Fouirer katsayıları olmak üzere ,x

$$x(s) = \sum_m c_m \varphi_m(s) = \boldsymbol{c'\varphi(s)} \tag{3.42}$$

şeklinde yazılabilir. Her m için;

$$D^2\varphi_m = -w_m^2 \varphi_m \tag{3.43}$$

yazılabildiğinden 3.42 de verilen ifadenin 2. Mertebeden türevi aşağıdaki gibidir:

$$D^2 x(s) = -\sum_m w_m^2 c_m \varphi_m(s) \tag{3.44}$$

Trigonometrik fonksiyonların ortogonallik özelliğinden, φ_m baz fonksiyonları ortonormaldir ve sonuç olarak $\| D^2 x \|^2$ pürüzlülük cezası c_m katsayılarının ağırlıklı kareler toplamı olarak yazılabilir:

$$\| D^2 x \|^2 = \int \left(-\sum_m w_m^2 c_m \varphi_m \right)\left(-\sum_m w_m^2 c_m \varphi_m \right) = \sum_m w_m^4 c_m^2 \quad) \tag{3.45}$$

$\boldsymbol{y}$ herhangi bir ξ olası temel bileşen eğrisinin katsayılar vektörü olarak tanımlandığında

$$\xi(s) = \sum_m y_m \varphi_m(s) = \boldsymbol{y}'\boldsymbol{\varphi}(\boldsymbol{s}) \tag{3.46}$$

eşitliği sağlanır. Böylece Fourier baz fonksiyonlar yaklaşımı ile cezalı temel bileşen örneklem varyansı şu şekilde ifade edilir:

$$TBAPSV = \frac{\boldsymbol{y}'\boldsymbol{V}\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{S}^{-2}\boldsymbol{y}} \tag{3.47}$$

3.47 de verilen ifadede yer alan $\boldsymbol{V}$, c_i katsayılar vektörünün kovaryans matrisidir ve $\boldsymbol{S}$, diagonal matrisi $S_{mm} = (1 + \lambda w_m^4)^{-1/2}$ olmak üzere düzgünleştirme operatörüne karşılık gelir.

$y_{(i)}$ ξ_j eğrilerinin Fourier katsayılarının vektörü olarak gösterildiğinde 3.38 de verilen ortogonallik kısıtı şu şekildedir.

$$\boldsymbol{y}'_{(i)}\boldsymbol{S}^{-2}\boldsymbol{y}_{(j)} = 0 \qquad j = 1,2,\dots,i-1 \tag{3.48}$$

Bu durumda özvektör denklemi aşağıdaki gibi yazılır:

$$\boldsymbol{V}\boldsymbol{y} = \rho\boldsymbol{S}^{-2}\boldsymbol{y} \tag{3.49}$$

$\boldsymbol{S}\boldsymbol{c}_i$ düzgünleştirme operatörü S nin uygulanmasıyla elde edilen düzgünleştirilmiş Fourier katsayılar vektörüdür. Bu vektörlerin kovaryans matrisi $\boldsymbol{SVS}$ ile gösterilmek üzere 3.49 daki özdenklem

$$(\boldsymbol{SVS})(\boldsymbol{S}^{-1}\boldsymbol{y}) = \rho(\boldsymbol{S}^{-1}\boldsymbol{y})\ ) \tag{3.50}$$

şeklinde yazılabilir. Bu özdenklemin çözülebilmesi için $\boldsymbol{SVS}$ nin özvektörü $\boldsymbol{u}$ olarak tanımlanır ve ardından düzgünleştirilmiş katsayılar vektörü $\boldsymbol{S}\boldsymbol{c}_i$ için standart temel bileşen analizi uygulanır. $\| \boldsymbol{y} \|^2 = 1$ olması gerektiğinden $\boldsymbol{y} = \boldsymbol{S}\boldsymbol{u}/\| \boldsymbol{S}\boldsymbol{u} \|$ olacak şekilde yeniden normalleştirilir ve ardından $\boldsymbol{y}$ ye karşılık gelen temel bileşen ξ 3.46 da verilen eşitlikten hesaplanır.

Düzgünleştirilmiş fonksiyonel temel bileşen analizi için aşağıdaki adımlar takip edilmelidir [1]:

1. Her bir x_i fonksiyonu için φ baz fonksiyon gösterimi ile ortaya çıkan $\boldsymbol{c}_i$ katsayıları hesaplanır.

2. Düzgünleştirme operatörü S bu katsayılara uygulanarak düzgünleştirilmiş katsayılar vektörü $\boldsymbol{Sc}_i$ elde edilir.
3. Bu düzgünleştirilmiş katsayılar vektörüne standart temel bileşen analizi uygulanır.
4. Elde edilen $\boldsymbol{u}$ özvektörlerine S düzgünleştirme operatörü uygulanır ve $\boldsymbol{Su}$ elde edilir.
5. $\| \boldsymbol{y} \|^2 = 1$ olacak şekilde y yeniden normalleştirilir.
6. ξ temel bileşen fonksiyonu 3.46 da verilen eşitlikten hesaplanır.

Periyodik Olmayan Durumlarda Baz Fonksiyon Yaklaşımı: Periyodik olmayan durumlarda en sık kullanılan baz fonksiyonlar B-Splayn baz fonksiyonlarıdır. $\boldsymbol{T}$ [0,1] aralığı olmak üzere, bu aralık üzerinde tanımlı periyodik olmayan x fonksiyonu

$$x(s) = \sum_m c_m \varphi_m(s) = \boldsymbol{c}'\boldsymbol{\varphi}(\boldsymbol{s}) \tag{3.51}$$

şeklinde yazılabilir. Periyodik durumda olduğu gibi $\boldsymbol{y}$ herhangi bir ξ olası temel bileşen eğrisinin katsayılar vektörü olarak tanımlansın ($\xi(s) = \sum_m y_m \varphi_m(s) = \boldsymbol{y}'\boldsymbol{\varphi}(\boldsymbol{s})$) ve $\boldsymbol{c_i}$ katsayılar vektörünün kovaryans matrisi **V** ile gösterilsin. $\boldsymbol{J}$ matrisi elemanları $\int \varphi_j \varphi_k$ olan $\int \varphi\varphi'$ olarak ve $\boldsymbol{K}$ matrisi elemanları $\int D^2\varphi_j D^2\varphi_k$ olan bir matris olarak tanımlandığında cezalı temel bileşen örneklem varyansı şu şekilde ifade edilir:

$$TBAPSV = \frac{\boldsymbol{y}'\boldsymbol{JVJy}}{\boldsymbol{y}'\boldsymbol{Jy} + \boldsymbol{\lambda y}'\boldsymbol{Ky}} \tag{3.52}$$

Böylece 3.49 da verilen özdenklem periyodik olmayan durumlar için şu şekildedir.

$$\boldsymbol{JVJy} = \rho(\boldsymbol{J} + \boldsymbol{\lambda K})\boldsymbol{y} \tag{3.53}$$

Bu özdenklemin çözülebilmesi için $\boldsymbol{S} = \boldsymbol{L}^{-1}$ ve $\boldsymbol{LL}' = \boldsymbol{J} + \boldsymbol{\lambda K}$ olacak şekilde tanımlanır. $\boldsymbol{L}$ matrisinin bulunabilmesi için matrisleri ayrıştırmak için kullanılan Tekil Değer Ayrışımı (Singular Value Decomposition, (SVD)) metodu veya Cholesky ayrıştırma metodu kullanılabilir. Burada $\boldsymbol{L}$ üst üçgen matristir. Bu durumda 3.53 teki öz denklem şu şekilde ifade edilir.

$$(\boldsymbol{SJVJS}')(\boldsymbol{L}'\boldsymbol{y}) = \rho\boldsymbol{L}'\boldsymbol{y} \tag{3.54}$$

Özetle periyodik olmayan durumlarda fonksiyonel temel bileşen analizi için baz fonksiyon yaklaşımını kullanmak için şu adımlar takip edilmelidir[1].

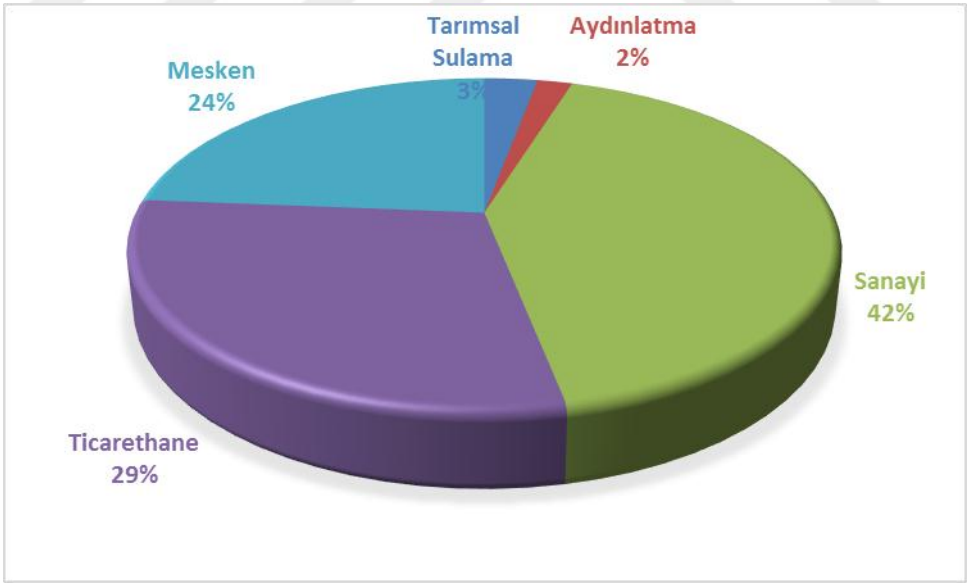
1. x_i gözlem değerleri uygun bir baz fonksiyon ile genişletilir ve $\boldsymbol{c}_i$ katsayı vekörleri elde edilir.
2. $\boldsymbol{J}$ ve $\boldsymbol{K}$ matrisleri hesaplanır.
3. $\boldsymbol{S} = \boldsymbol{L}^{-1}$ ve $\boldsymbol{L}\boldsymbol{L}' = \boldsymbol{J} + \boldsymbol{\lambda}\boldsymbol{K}$ olacak şekilde tanımlanır ve ardından SVD veya Cholesky ayrışımı kullanılarak $\boldsymbol{L}$ ve $\boldsymbol{S}$ matrisleri hesaplanır.
4. $\boldsymbol{L}d_i = \boldsymbol{S}\boldsymbol{c}_i$ eşitliğini çözmek için her bir i için $d_i = \boldsymbol{S}\boldsymbol{J}\boldsymbol{c}_i$ vektörleri bulunur.
5. d_i katsayı vektörleri üzerinde standart temel bileşen analizi uygulanır.
6. $\boldsymbol{L}'\boldsymbol{y} = u$ ile elde edilen $\boldsymbol{u}$ özvektörlerine düzgünleştirme operatörü $\boldsymbol{S}'$ uygulanır ve $\boldsymbol{y}$ vektörleri $\boldsymbol{y}'\boldsymbol{J}\boldsymbol{y} = 1$ olacak şekilde yeniden normalleştirilir.
7. Son olarak ξ temel bileşen fonksiyonları 3.46 da verilen ifadeden elde edilir.

4
UYGULAMA

Bu bölümde Fonksiyonel Veri Analizinde sıklıkla kullanılan Fonksiyonel Temel Bileşenler Analizi metodunun Türkiye'deki illerin hane halkı elektrik tüketimi ve bütün dünyayı etkisi altına alan Covid19 verileri üzerine bir uygulaması yer almaktadır.

4.1 Uygulama 1: Elektrik Tüketimi

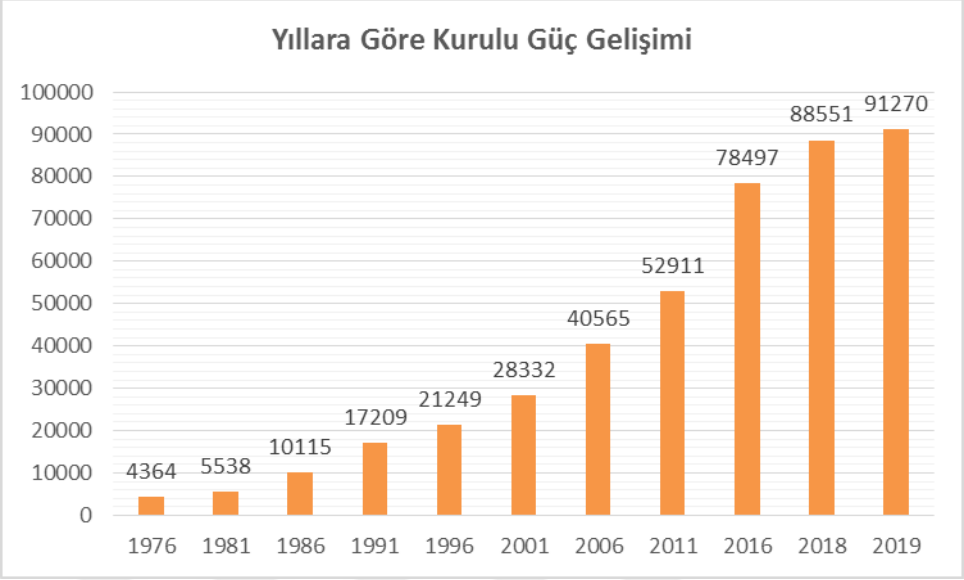
Veri : Türkiye'de elektrik enerjisi tüketimi nüfusun artmasına bağlı olarak her yıl artmaktadır. En son 2018 yılında EPDK tarafından yayımlanan yıllık rapora göre elektrik tüketimi bir önceki yıla göre %2,2 artarak 304,2 milyar kWh olmuştur. Toplam tüketimin yaklaşık %25 ini mesken (hane halkı) elektrik tüketimi oluşturmaktadır. EPDK nın 2017 yılındaki raporunda tüketimin tüketici türüne göre dağılımı aşağıdaki grafikteki gibi verilmektedir.



Şekil 4.1 2017 yılı elektrik tüketiminin tüketici türüne göre dağılımı(EPDK)

Buna bağlı olarak 2018 yılındaki rapora göre elektrik üretimi ise aynı oranda %2,2 artarak 304,8 milyar kWh e ulaşmıştır. Bunun yanı sıra 2019 yılı Eylül ayı sonu itibariyle Türkiye'nin kurulu gücü 90.720 MW'a ulaşmıştır ve bu gücün yüzde 31,4'ü hidrolik enerji, yüzde 28,6'sı doğal gaz, yüzde 22,4'ü kömür, yüzde 8,1'i rüzgâr,

yüzde 6,2'si güneş, yüzde 1,6'sı jeotermal ve yüzde 1,7'si ise diğer kaynaklar şeklindedir. Türkiye'nin kurulu gücünün yıllara göre değişimi şu şekilde özetlenebilir.



Şekil 4.2 Türkiye'nin yıllara göre kurulucu gücündeki değişim

Türkiye'de elektrik dağıtımı, enerji talebi, coğrafi yakınlık, yönetimsel yapı gibi bazı değişkenler göz önüne alınarak 21 dağıtım bölgesi üzerinde gerçekleşmektedir. Bu çalışmada kullanılmak üzere 81 ilin hane halkı elektrik tüketimi aşağıda belirtilen dağıtım şirketlerinden talep edilmiştir.



Şekil 4.3 Türkiyedeki elektrik dağıtım bölgeleri haritası

1. Dicle Elektrik Dağıtım A.Ş.
2. Vangölü Elektrik Dağıtım A.Ş.
3. Aras Elektrik Dağıtım A.Ş.
4. Çoruh Elektrik Dağıtım A.Ş.
5. Fırat Elektrik Dağıtım A.Ş.
6. Çamlıbel Elektrik Dağıtım A.Ş.
7. Toroslar Elektrik Dağıtım A.Ş.
8. Meram Elektrik Dağıtım A.Ş.
9. Başkent Elektrik Dağıtım A.Ş.
10. Akdeniz Elektrik Dağıtım A.Ş.
11. Gediz Elektrik Dağıtım A.Ş.
12. Uludağ Elektrik Dağıtım A.Ş.
13. Trakya Elektrik Dağıtım A.Ş.
14. Anadolu Yakası Elektrik Dağıtım A.Ş.
15. Sakarya Elektrik Dağıtım A.Ş.
16. Osmangazi Elektrik Dağıtım A.Ş.
17. Boğaziçi Elektrik Dağıtım A.Ş.
18. Kayseri ve Civarı Elektrik Türk A.Ş.
19. Menderes Elektrik Dağıtım A.Ş.
20. Göksu Elektrik Dağıtım A.Ş.
21. Yeşilırmak Elektrik Dağıtım A.Ş.

Özel bilgilerin korunması ve bu konuda çok fazla veri talebi olması gerekçesiyle bu dağıtım şirketlerinin birçoğundan olumlu yanıt alınamamıştır. Bunun sonucunda bu çalışmada İzmir, Manisa, Bolu, Kocaeli, Sakarya, Düzce, Balıkesir, Bursa, Çanakkale, Yalova, Artvin, Giresun, Gümüşhane, Rize ve Trabzon illerinin Ocak-2013 ve Mart-2019 yılları arasındaki hane halkı elektrik tüketimi verisi üzerine bir uygulamaya yer verilmektedir.

Elektrik tüketimi popülasyonun yoğun olduğu bölgelerde daha yoğundur. Özellikle Türkiye'de nüfusu en yoğun olan 10 il arasında sırasıyla İzmir ve Bursa illeri de bulunmaktadır. Nüfusun yoğun olduğu illerdeki tüketim daha fazla olacağından, tüketim verisini nüfus etkisinden arındırmak gerekmektedir. Bu amaçla Türkiye

İstatistik Kurumu (TÜİK) den 2013-2019 yıllarına ait toplam nüfus verisi elde edilmiştir. Bunun sonucunda her ilin elektrik tüketimi, o ilin toplam nüfusuna bölünerek tüketim verisi nüfus etkisinden arındırılmıştır.

Elektrik tüketimini etkileyen en önemli faktörlerden biri de sıcaklıktır. Özellikle sıcaklığın arttığı yaz aylarında elektrik tüketimi de artmaktadır. Sıcaklığa bağlı olarak elektrik tüketimindeki değişimi incelemek amacıyla bu çalışmada kullanılmak üzere Orman ve Su İşleri Bakanlığı Meteoroloji Genel Müdürlüğünden İzmir, Manisa, Bolu, Kocaeli, Sakarya, Düzce, Balıkesir, Bursa, Çanakkale, Yalova, Artvin, Giresun, Gümüşhane, Rize ve Trabzon illerinin Ocak-2013 ve Mart-2019 yılları arasındaki aylık ortalama sıcaklık verisi temin edilmiştir.

Yöntem: n tane bireye ait p tane değişkenin incelendiği çok değişkenli analizde temel bileşenler analizi yöntemi ile bir veri setinin varyans-kovaryans yapısını, bu değişkenlerin doğrusal birleşimleri yardımıyla açıklanması, boyut indirgenmesi ve yorumlanması mümkündür. Fakat veri seti büyüdükçe ve klasik temel bileşen analizini uygulanmasının mümkün olmadığı durumlarda (n<p) verileri fonksiyonel olarak ele almak daha avantajlıdır. Bu çalışmada fonksiyonel temel bileşen analizinin tercih edilmesini şu şekilde sıralayabiliriz.

- ✓ Örnek hacmi n (=15) değişken sayısı p(=75) den küçük olması
- ✓ Kayıp veriler (missing value) ile çalışılmasının mümkün olması
- ✓ Temel bileşen fonksiyonları ve bireysel fonksiyonların türev fonksiyonlarının incelebilmesi
- ✓ Görselliğin çok daha iyi olması.

Bu uygulamada yapılan ilk adım 75 ayrık noktada 15 ile ait aylık ortalama hane halkı elektrik tüketimi verileri sürekli birer fonksiyona dönüştürmek için Baz fonksiyonlar ve Pürüzlü Ceza Yaklaşımı metotları kullanılmıştır. Tüketim verilerinde periyodik bir yapı gözlenmediğinden B-Splayn baz fonksiyonları, sıcaklık verilerinde ise tüketim verilerinin aksine periyodik bir yapı olduğu gözlemlendiğinden Fouirer Baz fonksiyonları kullanılmıştır. Ardından lambda düzgünleştirme parametresinin belirlenmesinde en çok tercih edilen yöntem olan Genelleştirilmiş Çapraz Geçerlilik Yöntemi kullanılmıştır. Farklı lambda değerlerine karşılık gelen GCV değerleri kıyaslanmış ve minumum GCV değerine karşılık gelen lambda düzgünleştirme parametresi olarak seçilmiştir. Ardından ortalama

fonksiyonu ve fonksiyonların maksimum ve minumum noktalarının daha rahat gözlemlendiği türev fonksiyonları incelenmiştir. Ardından iller arasındaki aylık ortalama tüketim miktarları arasındaki değişimi incelemek amacıyla Düzgünleştirilmiş Fonksiyonel Temel Bileşenler Analizi(FTBA) uygulanmış ve en yüksek değişkenliği açıklama oranına sahip olan birinci temel bileşen fonksiyonu üzerinde en etkili olan iller tespit edilmiştir. Benzer şekilde iller arasındaki aylık ortalama sıcaklıklar arasındaki değişimi incelemek amacıyla sıcaklık verisine de Düzgünleştirilmiş FTBA uygulanmış ve birinci temel bileşen fonksiyonu üzerinde en etkili iller tespit edilmiştir. Son olarak sıcaklığın ve elektrik tüketiminin en yüksek ve en düşük olduğu iller karşılaştırılarak aralarında bir ilişki olup olmadığı incelenmiştir. Bu çalışmada açık kaynak kodlu R programından yararlanılmıştır.

Elektrik Tüketim Verilerinin Düzgünleştirilmiş Fonksiyonel Temel Bileşen Analizi

Bu çalışmada 2013-2019 yıllarına ait İzmir, Manisa, Bolu, Kocaeli, Sakarya, Düzce, Balıkesir, Bursa, Çanakkale, Yalova, Artvin, Giresun, Gümüşhane, Rize ve Trabzon illerinin aylık tüketim verileriyle çalışılmaktadır. Bu verilerin dağılımı logaritmik normal dağılıma uygundur ve betimleyici istatistikler şu şekildedir:

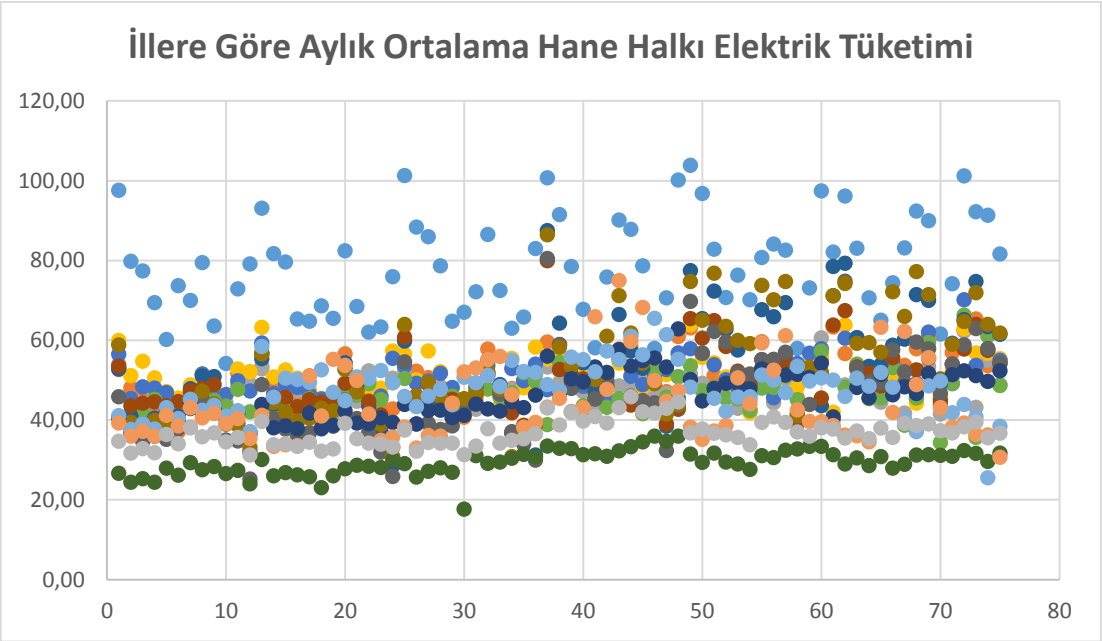
Tablo 4.1 Elektrik tüketim verileri için tanımlayıcı istatistikler

	Min.	Q1(1.Kartil)	Medyan	Ortalama	Q3(3.Kartil)	Mak.
İzmir	54.25	68.60	78.57	78.05	86.28	**103.87**
Manisa	37.11	45.79	49.05	50.64	55.03	65.38
Bolu	37.99	44.46	47.38	47.61	50.20	62.36
Kocaeli	41.84	47.44	51.44	51.63	55.38	63.93
Sakarya	40.17	47.21	49.85	51.15	54.34	70.16
Düzce	34.4	43.62	45.77	47.17	49.59	66.38
Balıkesir	28.53	41.88	50.15	52.22	60.47	87.52
Bursa	33.18	44.55	47.98	49.45	53.47	80.02
Çanakkale	25.24	37.87	45.12	46.77	54.47	80.54
Yalova	30.85	41.77	50.39	53.15	62.70	86.52
Artvin	36.15	40.76	44.86	45.94	51.24	62.86
Giresun	**17.71**	27.62	29.53	29.52	31.50	35.98
Gümüşhane	25.54	43.55	48.47	47.74	51.30	65.55
Rize	30.68	37.27	41.21	44.21	49.94	75.03
Trabzon	31.27	34.25	36.57	36.80	38.93	45.85

Min.	Q1	Medyan	Ortalama	Q3	Mak.
17.71	40.85	47.28	48.80	54.15	103.87

Tablo 4.1 incelendiğinde iller bazında en düşük tüketimin Giresun ve en yüksek tüketiminde İzmir illerine ait olduğu rahatlıkla gözlenmektedir.

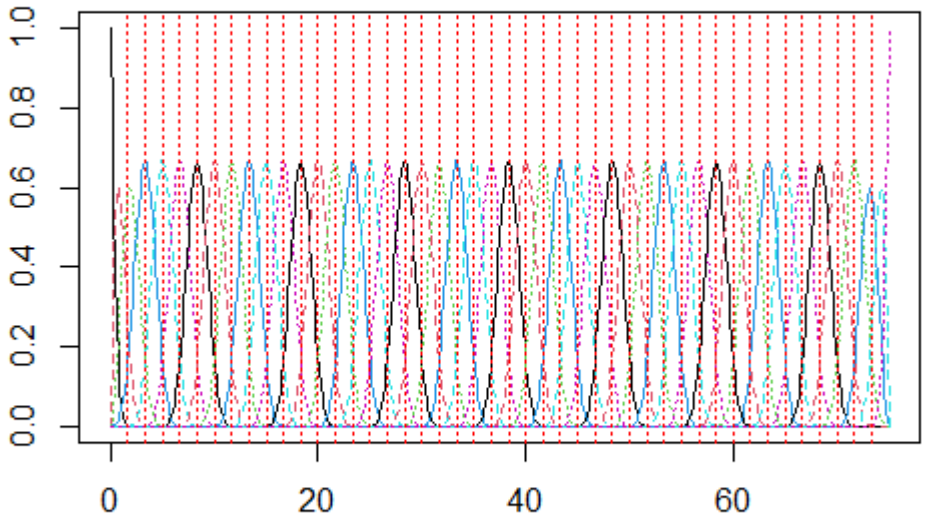
Aşağıdaki grafikte görüldüğü gibi 15x75 ayrık veri noktası bulunmaktadır.



Şekil 4.4 2013-2019 yıllarına ait illerin aylık ortalama elektrik tüketimi

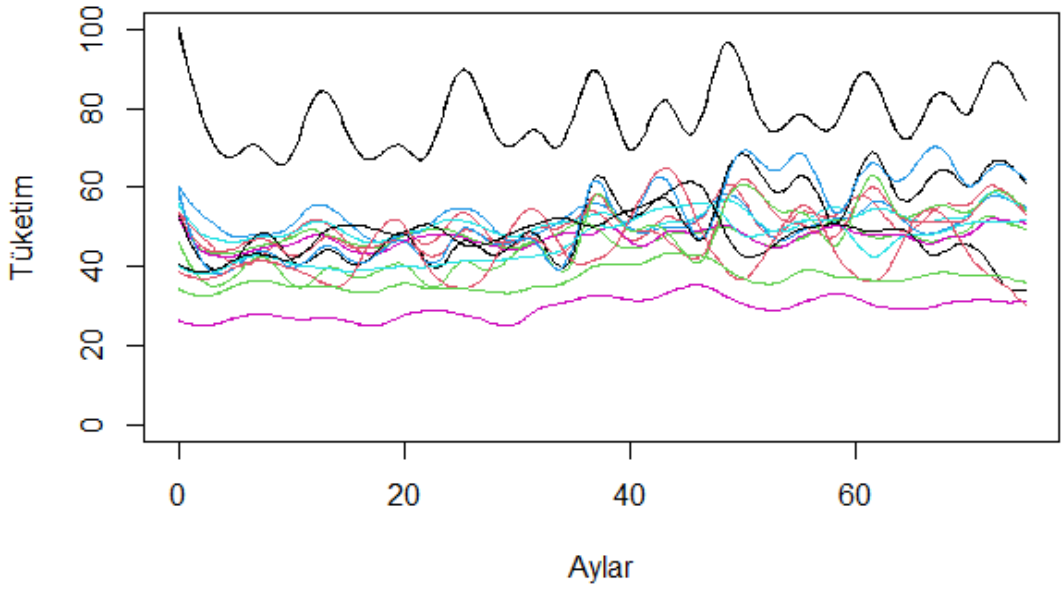
Şekil 4.4 incelendiğinde illerin bireysel olarak aylık ortalama elektrik tüketimlerini incelemek bazı noktalarda oldukça zorlaşmaktadır. Daha büyük bir veriyle çalışıldığında bu verileri bireysel olarak incelemek çok daha zor olacağından veri setini fonksiyonel olarak ele almak daha iyi bir analiz yapma imkanı sunmaktadır.

Fonksiyonel verilerle çalışmak için yapılması gereken ilk şey ayrık veri noktalarını sürekli birer fonksiyona dönüştürmektir. Bu çalışmada fonksiyonların türev bilgisine de başvurulacağından interpolasyon yerine veri düzgünleştirme metotları kullanılmıştır. Bu amaçla öncelikle elektrik tüketimi verileri için B-Splayn Baz fonksiyonları oluşturulmuştur.



Şekil 4.5 B-Spline baz fonksiyonlar

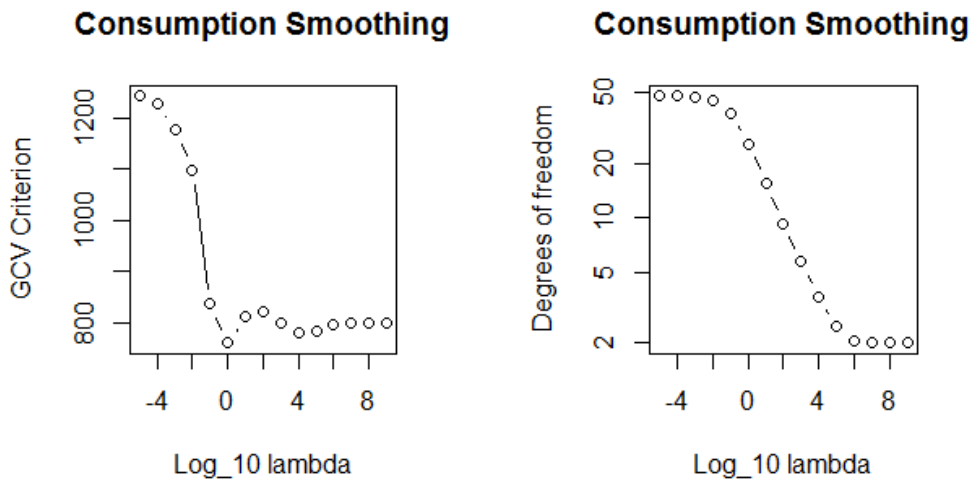
Baz fonksiyonlar elde edildikten sonra pürüzlü ceza yöntemiyle bireysel fonksiyonlar elde edilmiştir.



Şekil 4.6 Bireysel fonksiyonlar

Şekil 4.6 da Pürüzlü Ceza yöntemiyle oluşturulan bireysel fonksiyonlar incelendiğinde en üstteki eğrinin maksimum tüketime sahip olan İzmir ili ve en alttaki eğrinin de minumum tüketime sahip Giresun ili olduğu rahatlıkla gözlenmektedir. İzmir ve Giresun dışındaki illere ait eğriler birbirlerine yakın değerlere sahip olduğundan belli bir bölgede yoğunlaşmıştır. Bu nedenle bu eğrilerin bireysel davranışları hakkında çok fazla yorum yapılamamaktadır. Fakat genel olarak incelediğimizde 36. noktaya denk gelen Aralık- 2015 ten itibaren tüketimde genel bir artış olduğu ve yine 36. ve 60. noktaya denk gelen Aralık-2015, Aralık-2017 yıllarında tüketimin maksimuma ulaştığı gözlenmektedir.

Pürüzlü ceza yönteminde kullanılan lambda düzgünleştirme parametresini belirlemek için Genelleştirilmiş Çarpraz Geçerlilik ölçütü kullanılmıştır. Farklı lambda değerlerine karşılık gelen GCV değerleri kıyaslanarak minumum GCV değerine karşılık gelen lambda değeri düzgünleştirme parametresi olarak seçilmiştir.



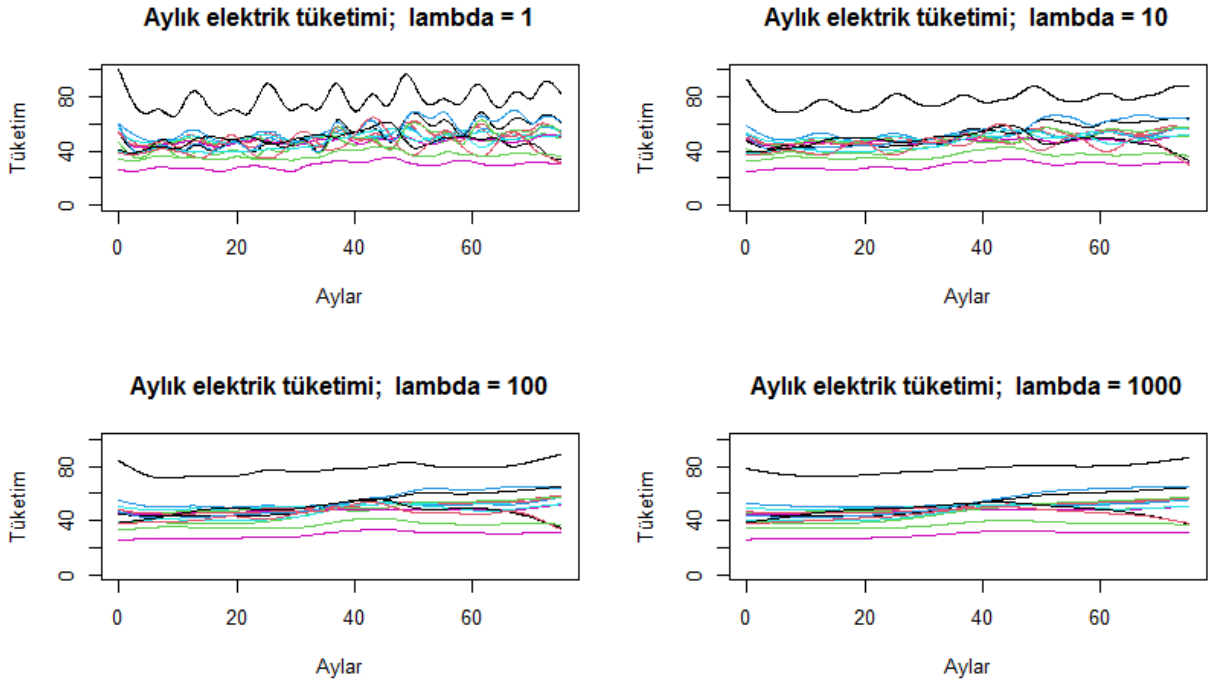
Şekil 4.7 Farklı lambda değerlerine karşılık gelen GCV kriteri

Tablo 4.2 Farklı lambda değerlerine karşılık GCV değerleri

	loglam	df	gcv
-5	-5	47.977245	1241.9517
-4	-4	47.791362	1226.2726
-3	-3	46.830924	1175.8528
-2	-2	44.969205	1098.7039
-1	-1	38.485858	836.8559
0	0	26.022004	759.8884
1	1	15.716711	810.9650
2	2	9.358860	822.8270
3	3	5.710876	798.1579
4	4	3.651062	780.7963
5	5	2.501334	785.1402
6	6	2.071456	796.5741
7	7	2.007488	799.0395
8	8	2.000752	799.3116
9	9	2.000076	799.3391

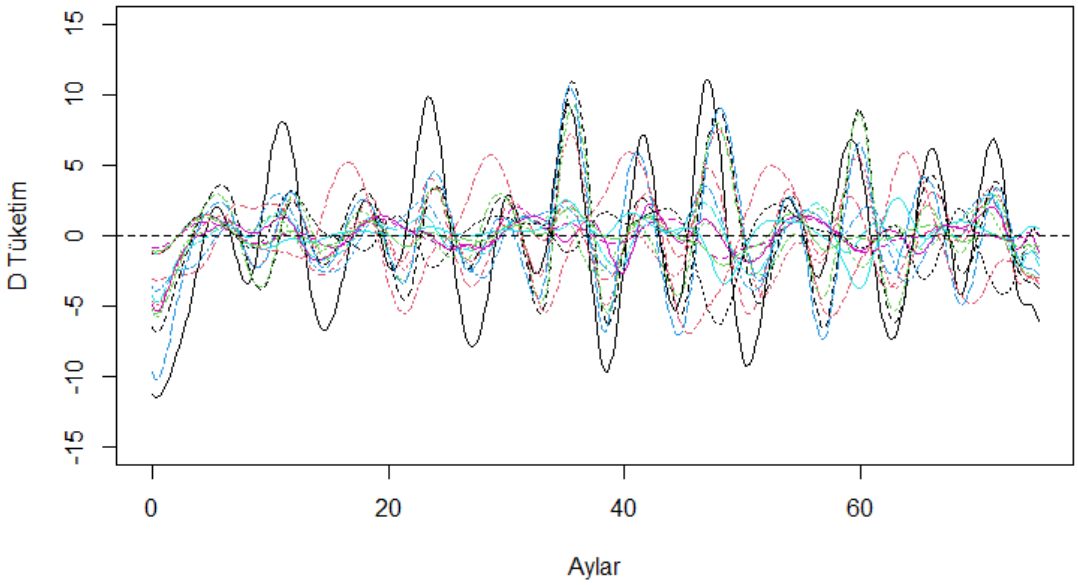
Tablo 4.2 ve Şekil 4.7 incelendiğinde 10^{-5} ten 10^{9} a kadar 15 farklı lambda değerine karşılık gcv değerleri görülmektedir. Minumum gcv değerine karşılık gelen lambda değeri 10^{0} yani 1 olduğundan lambda düzgünleştirme parametresi 1 olarak belirlenmiştir.

Aşağıdaki farklı lambda değerlerine karşı bireysel fonksiyonlarda meydana gelen değişim gösterilmektedir. Lambda değeri büyüdükçe orijinal veriden uzaklaşılmakta bu da veri kaybına sebep olmaktadır.

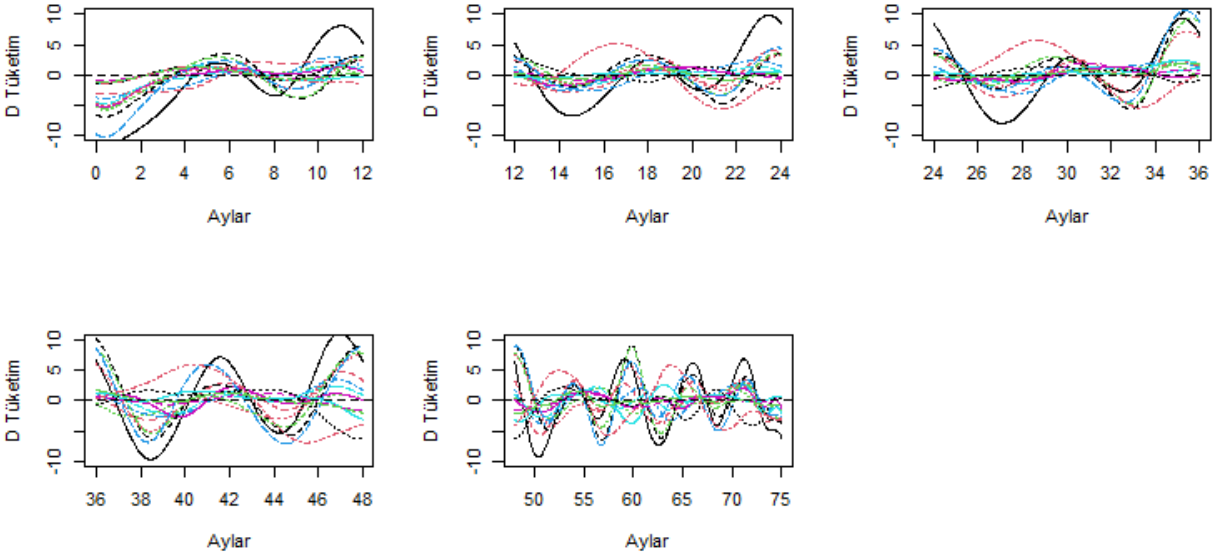


Şekil 4.8 Farklı lambda değerlerine karşılık bireysel fonksiyonlar

Şekil 4.6 da gösterilen bireysel fonksiyonların maksimum ve minumum noktalarını, fonksiyonların iniş- çıkış yaptığı noktaları tespit etmek ve bu fonksiyonları daha yakından inceleyebilmek için aşağıdaki şekilde gösterilen türev fonksiyonları elde edilmiştir.



Şekil 4.9 Birinci türev fonksiyonları



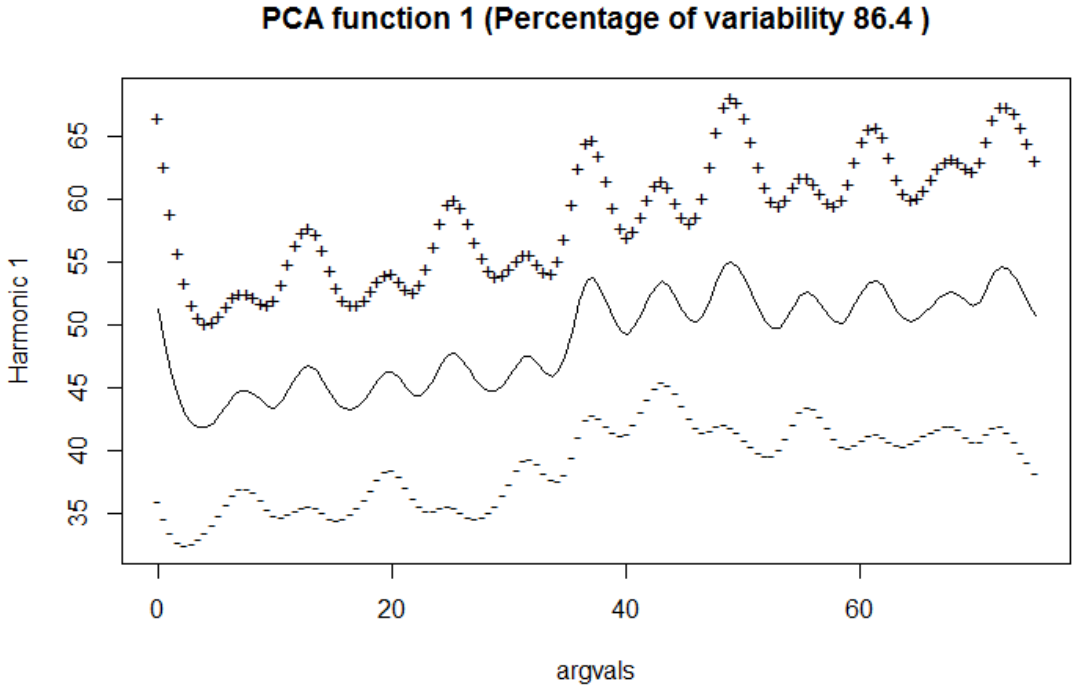
Şekil 4.10 Yıllara göre birinci türev fonksiyonları

Türev fonksiyonları incelendiğinde illerin tüketim konusunda birbirinden ne kadar çok farklılık gösterdiği görülmektedir. Bazı illerin elektrik tüketiminin maksimum olduğu aylarda bazı iller minumum tüketime sahip olabilmektedir. Bu nedenle bazı

noktaları daha yakından inceleyebilmek için Şekil 4.10 da yıllara göre türev fonksiyonları verilmiştir.

2013 yılından itibaren her yılın 6. ve 11. Ayına denk gelen Haziran ve Kasım aylarında genel olarak tüketim miktarlarının maksimuma ulaştığı ve önceki yıllara göre kıyaslandığında en fazla tüketimin 2016 yılında yapıldığı söylenebilir. Bununla birlikte her yılın Mart ve Eylül aylarında elektrik tüketiminin genellikle azaldığı ve bu azalmanın en fazla 2016 yılında gerçekleştiği rahatlıkla görülmektedir.

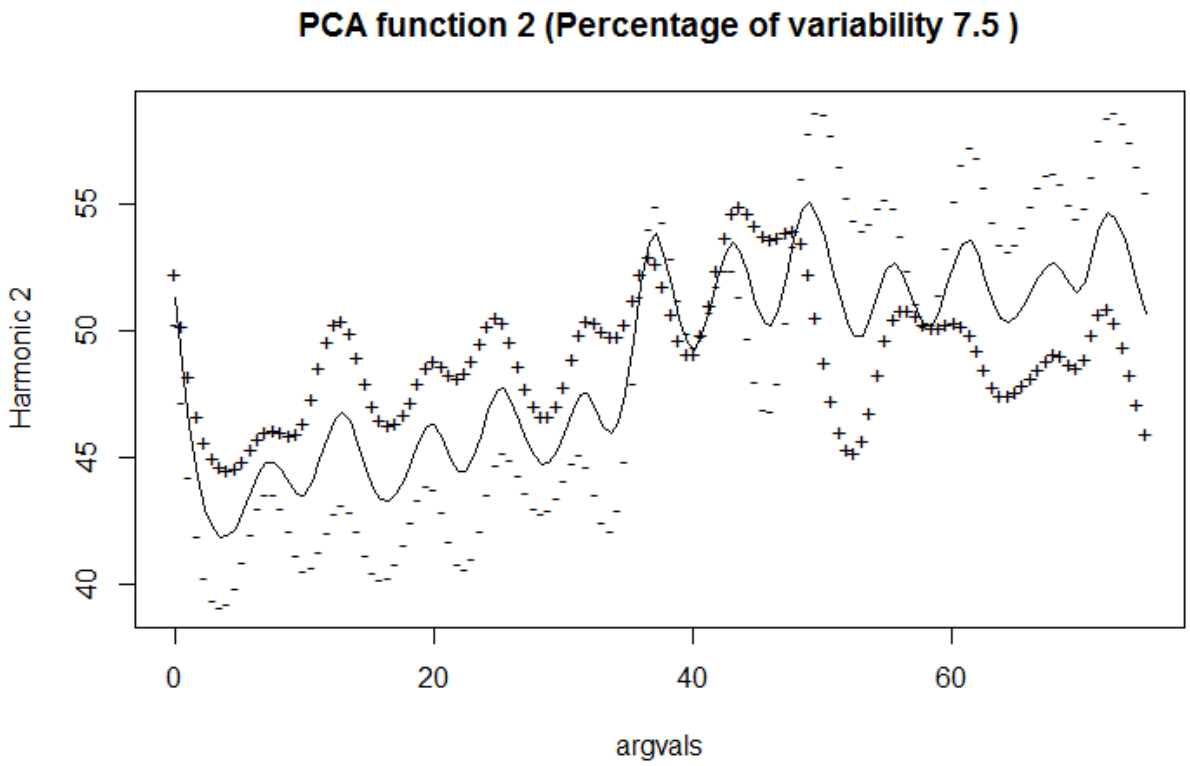
Ayrık veri noktaları sürekli birer fonksiyona dönüştürülüp, bireysel fonksiyonlar ve türevleri incelendikten sonra bu fonksiyonlar arasındaki değişimi incelemek amacıyla Fonksiyonel Temel Bileşenler Analizi uygulanmıştır.



Şekil 4.11 Birinci temel bileşen fonksiyonu

Şekil 4.11 da görülen eğri ortalama fonksiyonunu, + ve – ler ise ortalama fonksiyonuna birinci temel bileşen fonksiyonunun bir sabitle çarpımının eklenmesi ve çıkarılmasının etkilerini göstermektedir. Temel bileşen fonksiyonlarını tek başına yorumlamak zor olduğundan yukarıdaki gibi ortalama fonksiyonuyla kıyaslamak çok daha yararlıdır.

Birinci temel bileşen fonksiyonu incelendiğinde toplam değişkenliğin %86.4 üne karşılık geldiği ve hiçbir şekilde işaret değiştirmediği, yani her daim pozitif olduğu gözlenmektedir. Şekil 4.11 e bakıldığında + ve – ler ortalamadan uzak olduğundan ortalamadan sapmaların yüksek olduğu, özellikle bu sapmaların Kasım aylarında daha da arttığı görülmektedir. Buradan %86.4 değişkenliği açıklama yüzdesine sahip birinci temel bileşen fonksiyonu ile iller arasındaki elektrik tüketimindeki temel değişkenliğin özellikle kış aylarından kaynaklandığı söylenebilir.



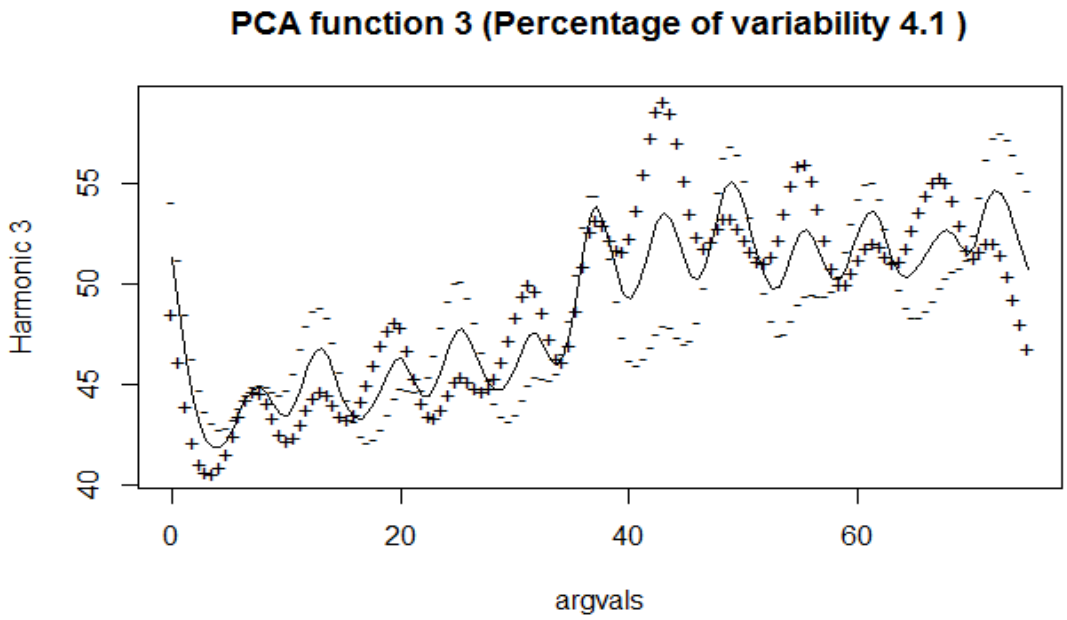
Şekil 4.12 İkinci temel bileşen fonksiyonu

İkinci temel bileşen fonksiyonu toplam değişkenliğin %7.5 ine karşılık gelmektedir ve birinci temel bileşen fonksiyonuna kıyasla ortalamadan sapmalar daha azdır. Fakat Şekil 4.12 incelendiğinde ikinci temel bileşen fonksiyonunun 36, 42 ve 48. inci noktalarda yani 2016 yılının Ocak, Haziran ve Aralık aylarında işaret değiştirdiği ve 2016 yılının ilk yarısında ortalamadan sapmaların minumuma ulaştığı görülmektedir. İşaret değiştirdiği noktalarda temel bileşen ağırlık fonksiyonunun mutlak değerce ortalamadan daha büyük olduğu ve bu noktalardan önce artan(azalan) tüketimin bu noktalardan sonra azalışa(artışa) geçtiği yorumu

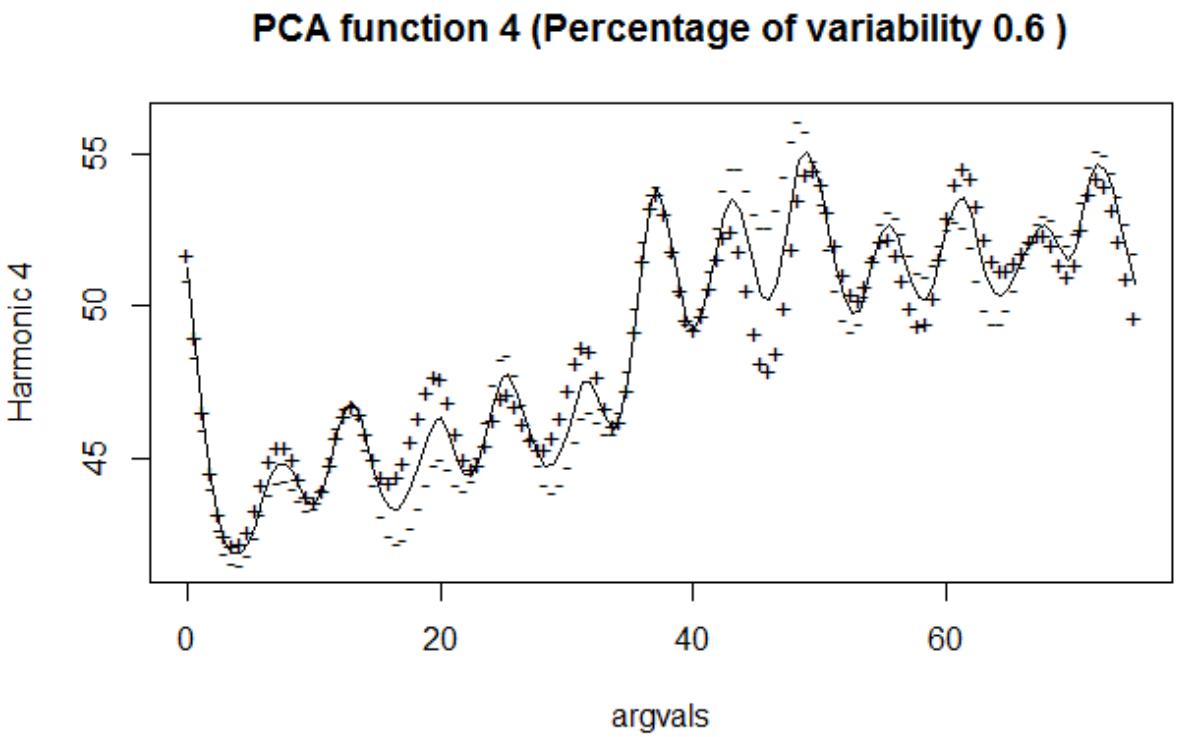
yapılabilir. Bu nedenle ikinci temel bileşen fonksiyonunu şok bileşeni (shock component) olarak adlandırabilir.

Türkiye Elektrik İletim AŞ (TEİAŞ) verilerinden de derlenen bilgiye göre tüm zamanların en çok elektrik tüketimi 2016 yılında gerçekleşmiştir. Bu çalışmada kullanılan verilerde bu bilgiyi desteklemektedir. Bu nedenle %7.5 değişkenlik açıklama yüzdesine sahip ikinci temel bileşen fonksiyonu ile iller arasındaki elektrik tüketimindeki değişimin 2016 yılındaki tüketimdeki patlamadan kaynaklandığı söylenebilir.

Diğer temel bileşen fonksiyonları da şu şekildedir:



Şekil 4.13 Üçüncü temel bileşen fonksiyonu



Şekil 4.14 Dördüncü temel bileşen fonksiyonu

Şekil 4.13 ve Şekil 4.14 ten de görüldüğü üzere üçüncü temel bileşen fonksiyonu toplam değişkenliğin yüzde 4.1 ine karşılık gelirken dördüncü temel bileşen fonksiyonu %0.6 sına karşılık gelmektedir. Üçüncü temel bileşen fonksiyonu incelendiğinde kış aylarında ortalamadan sapmaların arttığı görülürken dördüncü temel bileşen fonksiyonunda ortalamadan sapmalar yok denecek kadar azdır.

Yukarıda temel bileşen fonksiyonları ortalamayla kıyaslanarak grafikleri incelenmiş ve yorumlanmıştır. Aşağıdaki tabloda bu temel bileşen fonksiyonlarının skorları verilmektedir:

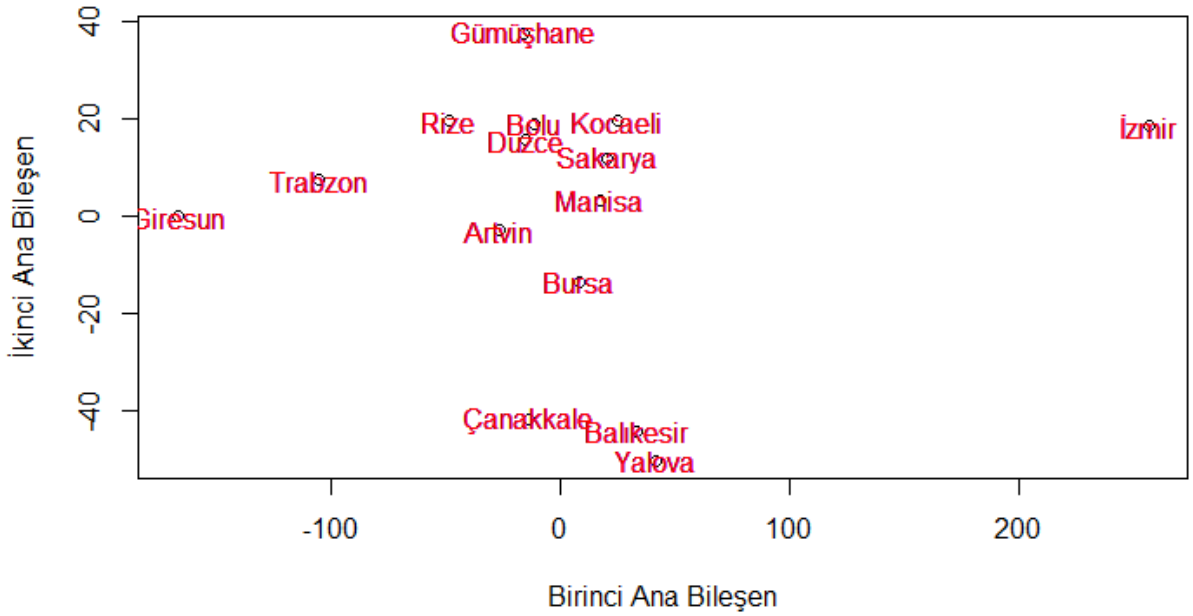
Tablo 4.3 Temel bileşen fonksiyonlarının skor değerleri

	PCA.1	PCA.2	PCA.3	PCA.4
İzmir	256.578896	18.73786911	-1.682748	-5.7488270
Manisa	17.788862	3.28480196	-11.374738	-5.6029158
Bolu	-11.318618	19.10334051	-12.757847	4.9371067
Kocaeli	25.378938	19.60625179	-21.135082	3.5889326
Sakarya	20.255047	11.82037011	-10.415299	2.2142540
Düzce	-14.952706	15.70725530	-11.030341	2.8056279
Balıkesir	33.297645	-44.13745720	5.784369	1.4182472
Bursa	8.267043	-13.48175435	-14.860260	14.8826174
Çanakkale	-13.771162	-41.87848657	-1.482847	5.1085914

Tablo 4.3 Temel bileşen fonksiyonlarının skor değerleri(devamı)

Yalova	41.830167	-50.22831260	16.098494	-3.1564539
Artvin	-26.634312	-2.94760606	11.117178	-17.6607867
Giresun	-166.826999	-0.07622067	-18.725204	-6.9240513
Gümüşhane	-15.694876	37.52468047	23.001324	0.8204062
Rize	-48.775253	19.57564513	53.301587	7.4062456
Trabzon	-105.422672	7.38962306	-5.838585	-4.0889944

En yüksek değişkenliği açıklama yüzdelerine sahip birinci ve ikinci temel bileşen fonksiyonu skorlarının dağılımı şu şekildedir:

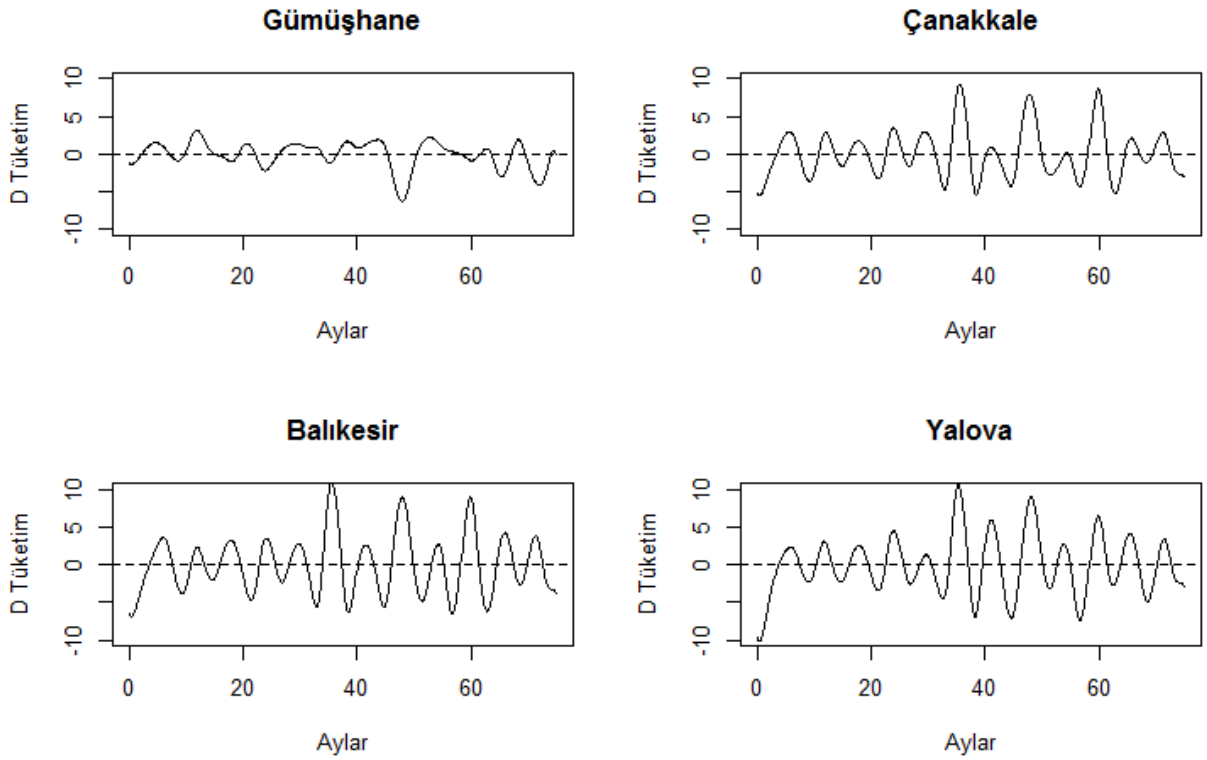


Şekil 4.15 Temel bileşen skorlarının dağılımı

Şekil 4.15 %86.4 ile en yüksek değişkenliği açıklama yüzdesine sahip birinci temel bileşen fonksiyonu açısından değerlendirildiğinde en yüksek temel bileşen skoruyla İzmir ilinin en etkili il olduğu, Giresun ve Trabzon illerinin de en yüksek negatif skor değerlerine sahip iller olduğu görülmektedir. Veri seti incelendiğinde yıl boyunca özellikle kış aylarında elektrik tüketiminin en yüksek olduğu il İzmir, en düşük olduğu iller ise Giresun ve Trabzon'dur.

%7.5 değişkenliği açıklama yüzdesine sahip ikinci temel bileşen fonksiyonu açısından bir değerlendirilme yapıldığında en yüksek pozitif skor değerine sahip Gümüşhane ilinin en etkili il olduğu ve Çanakkale, Balıkesir ve Yalova illerinin en yüksek negatif skor değerlerine sahip iller olduğu görülmektedir.

İkinci temel bileşen fonksiyonundaki değişkenliği daha iyi açıklamak için mutlak değerce en yüksek skor değerlerine sahip illerin türev fonksiyonu aşağıdaki şekilde verilmiştir.



Şekil 4.16 Gümüşhane, Çanakkale, Balıkesir ve Yalova illerinin türev fonksiyonu

İkinci temel bileşen fonksiyonundaki değişkenliğin 2016 yılındaki aşırı elektrik tüketiminden kaynaklandığı ve önceden tüketimin artışta(azalışta) olduğu illerde yıl içerisinde 3 kere seyir değiştirerek azalışa(artışa) geçmesinden kaynaklandığı belirtilmişti. Gerçekten de Şekil 4.16 dan da görüleceği üzere 2016 yılına denk gelen 36 ve 48.inci noktalar arasında diğer illerde tüketimin arttığı noktalarda Gümüşhane ilinde azalma meydana geldiği görülmektedir. Veri seti incelendiğinde bu illerin minumum tüketimleriyle maksimum tüketimlerinin arasında diğer illere göre çok büyük bir fark olduğu, elektrik tüketiminin yıllara göre değişkenlik gösterdiği görülmektedir.

Sıcaklık Verilerinin Düzgünleştirilmiş Fonksiyonel Temel Bileşen Analizi

İllerin aylık ortalama elektrik tüketimi üzerine yukarıda yapılan uygulamada genel olarak İzmir ilinde ortalamanın çok üstünde, Giresun ve Trabzon illerinde ise ortalamanın oldukça altında bir tüketim olduğu, Gümüşhane, Çanakkale, Balıkesir

ve Yalova illerinde de yıllara göre elektrik tüketiminin oldukça değişken olduğu göze çarpmaktadır. Bu veriler analiz edilmeden önce nüfus etkisinden arındırıldığı için bu iller arasındaki farklılığın sıcaklıkla bir etkisi olup olmadığı merak edilmektedir. Bu nedenle bu uygulamada İzmir, Manisa, Bolu, Kocaeli, Sakarya, Düzce, Balıkesir, Bursa, Çanakkale, Yalova, Artvin, Giresun, Gümüşhane, Rize ve Trabzon illerinin Ocak-2013 Mart-2019 dönemleri arasındaki aylık ortalama sıcaklık verisi üzerine fonksiyonel temel bileşenler analizi yapılarak bu illerin aylık ortalama sıcaklıklardaki değişkenliği incelenecektir.

İllerin 2013- 2019 dönemlerinde aylık ortalama sıcaklığı logaritmik normal dağılıma sahiptir ve betimleyici istatistikler şu şekildedir:

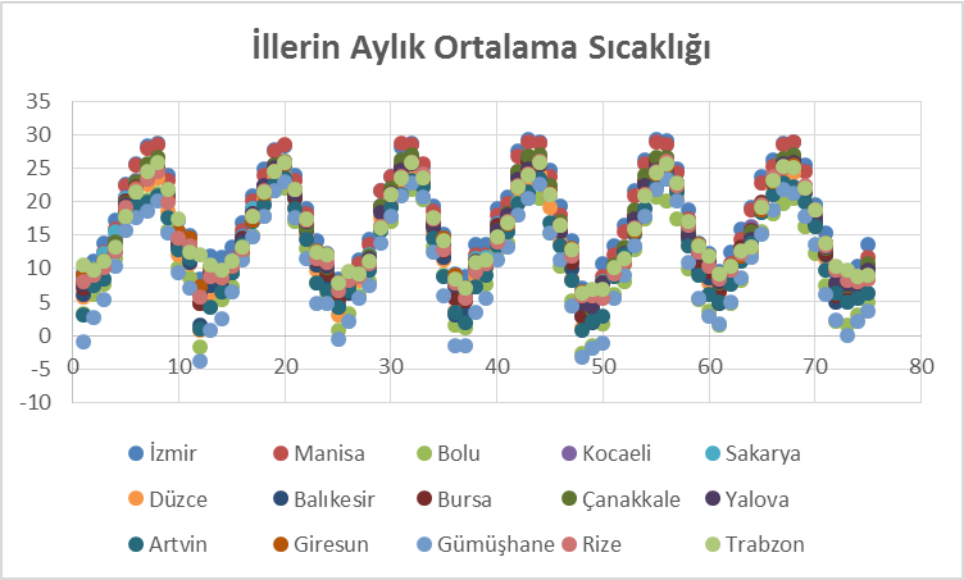
Tablo 4.4 İllerin aylık ortalama sıcaklık verisi için tanımlayıcı istatistikler

	Min.	Q1(1.Kartil)	Medyan	Ortalama	Q3(3.Kartil)	Mak.
İzmir	6.7	12.35	17.3	18.4	25	**29.4**
Manisa	4.2	10.7	15.9	17.1	24.2	29
Bolu	-2.6	5.35	10.4	11.05	17.4	22.1
Kocaeli	4.3	10.35	15	15.84	22.1	26.2
Sakarya	3.9	10.15	15.2	15.61	21.6	25.7
Düzce	0.7	8.45	13.5	13.96	20.3	24.4
Balıkesir	1.6	8.8	13.6	14.74	21.6	26.3
Bursa	2.9	9.55	14.5	15.44	22.05	26.4
Çanakkale	3.9	10.3	14.8	15.97	22.25	27.1
Yalova	4.2	9.65	14.1	15.5	21.9	25.9
Artvin	0.8	7.4	13.4	12.99	19.25	23.7
Giresun	6.6	10.1	14.6	15.65	21.55	25.9
Gümüşhane	**-3.8**	4.1	11	10.41	17.8	23.3
Rize	5.5	9.65	13.5	15.35	21.8	26.1
Trabzon	6.4	10.75	14.1	15.9	21.65	25.9

```
   Min.      Q1      Medyan   Ortalama Q3      Mak.
 -3.80    9.30    14.00    14.93    21.30    29.40
```

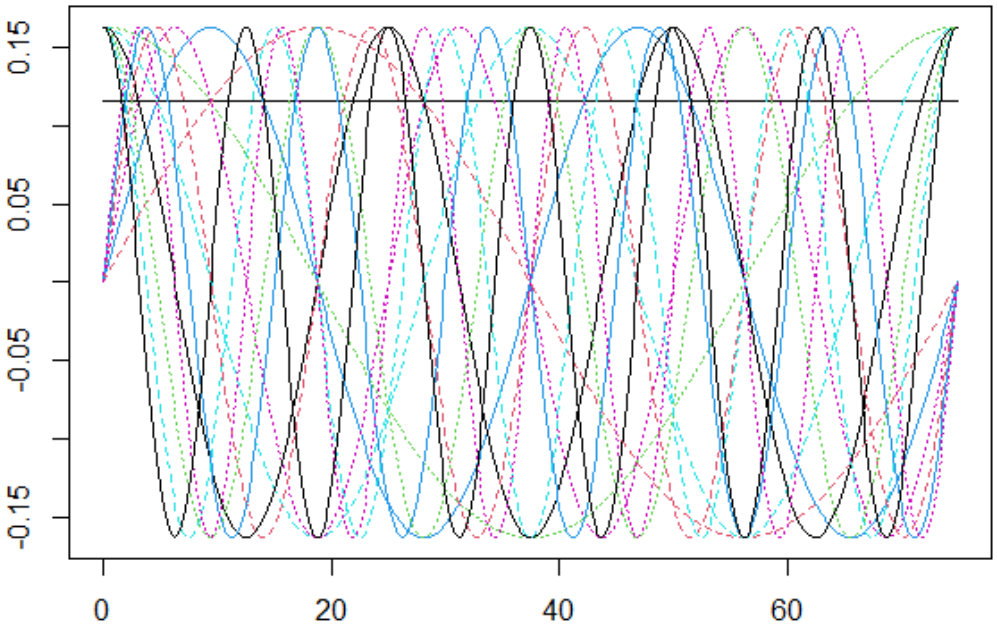
Tablo 4.4 incelendiğinde aylık ortalama en düşük sıcaklığa sahip ilin Gümüşhane ve en yüksek sıcaklığa sahip ilin İzmir olduğu görülmektedir.

Aşağıdaki grafikte illerin aylık ortalama sıcaklığı verisine ait 15x75 ayrık veri noktası bulunmaktadır.



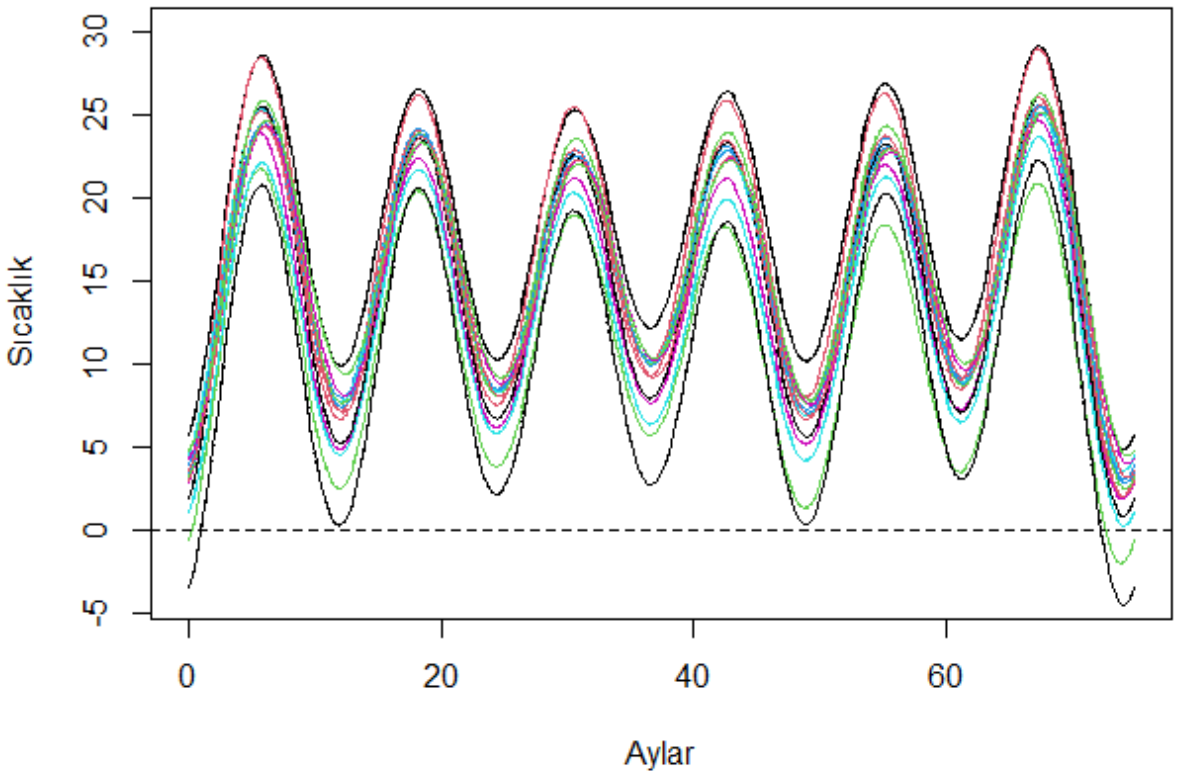
Şekil 4.17 İllerin aylık ortalama sıcaklığı

Şekil 4.17 incelendiğinde illerin aylık ortalama sıcaklığını bireysel olarak incelemenin mümkün olmadığı görülmektedir. Fonksiyonel veri analizi, analiz açısından daha avantajlı olduğundan sıcaklık verisi, elektrik verisinde olduğu gibi fonksiyonel olarak ele alınmıştır. Şekil 4.17 den da görüleceği üzere sıcaklık verilerinde periyodik bir yapı gözlenmektedir. Bu nedenle ayrık veri noktalarını sürekli birer fonksiyona dönüştürmek için ilk adım olarak Fourier Baz fonksiyonları oluşturulmuştur.



Şekil 4.18 Fourier baz fonksiyonları

Baz fonksiyonlar elde edildikten sonra pürüzlü ceza yöntemiyle bireysel fonksiyonlar elde edilmiştir.

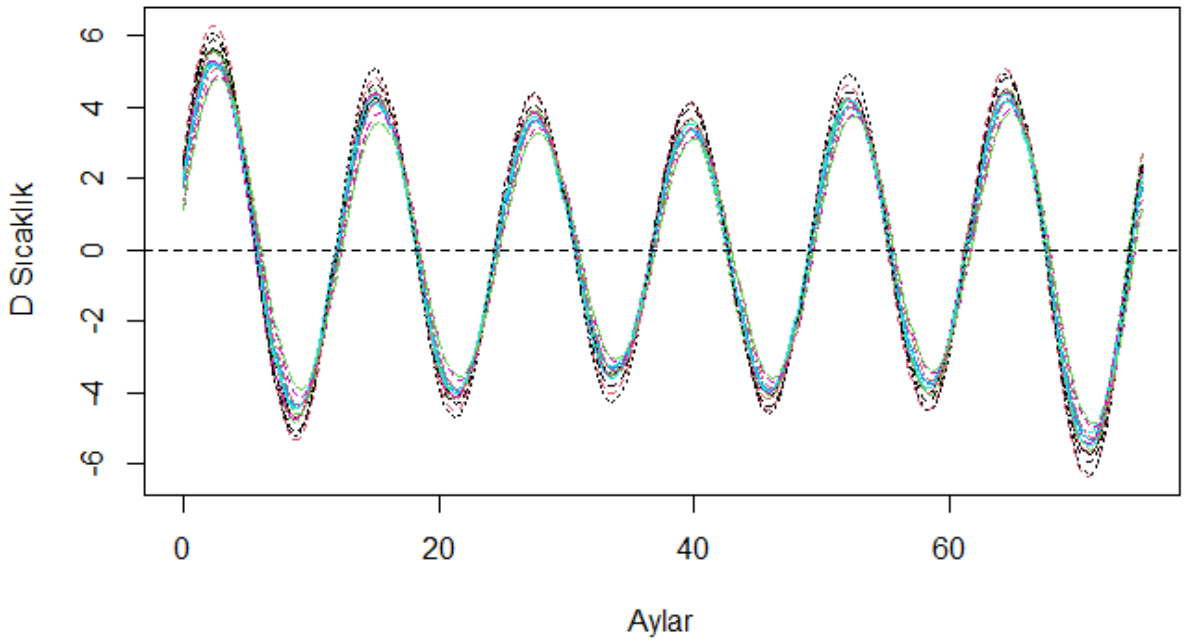


Şekil 4.19 Sıcaklık verisi için bireysel fonksiyonlar

Şekil 4.19 da Pürüzlü Ceza yöntemiyle belirlenen bireysel fonksiyonlar incelendiğinde her yılın Temmuz ayında sıcaklıkların maksimum ve Aralık ayında ise minumum olduğu görülmektedir. 2013 ve 2018 yılları arasında bu illerde sıfırın altında sıcaklığa rastlanmazken Aralık- 2018 de bütün illerde sıcaklıklarda genel bir düşüş olduğu ve özellikle Bolu ve Gümüşhane illerinde sıcaklıkların sıfırın altına indiği görülmektedir. İller genelinde en yüksek sıcaklıklara ise Temmuz-2018 de ve daha sonra Temmuz-2013 de rastlanmaktadır.

Pürüzlü ceza yönteminde kullanılan lambda düzgünleştirme parametresini belirlemek için Genelleştirilmiş Çarpraz Geçerlilik ölçütü kullanılmıştır. Farklı lambda değerlerine karşılık gelen GCV değerleri kıyaslanarak minumum GCV değerine karşılık gelen 1 değeri düzgünleştirme parametresi olarak seçilmiştir.

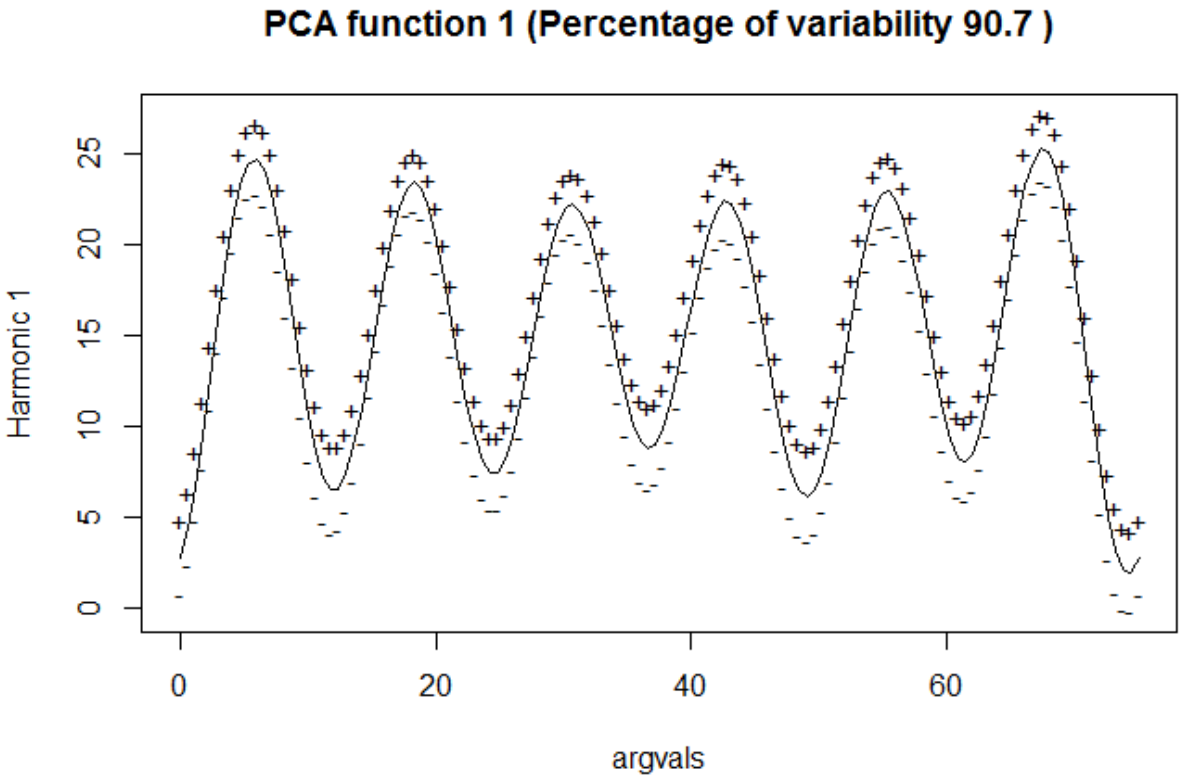
Sıcaklık verisini daha ayrıntılı incelemek için aşağıda türev fonksiyonu verilmektedir.



Şekil 4.20 Türev fonksiyonları

Şekil 4.20 incelendiğinde her yıl Temmuz ayına kadar sıcaklıkların artıp maksimuma ulaştığı ve daha sonra Aralık ayına kadar minumuma düştüğü görülmektedir. Bunun yanı sıra sıcaklıkların düşme hızında en çok 2013 ve 2018 yıllarında bir artış olduğu görülmektedir.

Yukarıda sıcaklık verisi için öncelikle ayrık veri noktaları sürekli birer fonksiyona dönüştürülüp, bireysel fonksiyonlar ve türev fonksiyonları incelenmiştir. Bu fonksiyonlar arasındaki değişimi incelemek amacıyla bir sonraki adımda Fonksiyonel Temel Bileşenler Analizi uygulanmıştır.



Şekil 4.21 Birinci temel bileşen fonksiyon

Şekil 4.21 de Birinci Temel bileşen fonksiyonu incelendiğinde sıcaklıkların genellikle ortalama etrafında olduğu ancak sıcaklıkların en yüksek ve en düşük olduğu noktalarda yani Temmuz ve Aralık aylarında ortalamadan sapmaların arttığı görülmektedir. Buradan %90.7 değişkenliği açıklama yüzdesine sahip birinci temel bileşen fonksiyonu ile iller arasındaki aylık ortalama sıcaklıklardaki değişkenliğin özellikle Temmuz ve Aralık aylarından kaynaklandığı söylenebilir.

Temel bileşen fonksiyonlarını incelemenin bir diğer yolu temel bileşen skor değerlerini incelemektir. Aşağıdaki tabloda 4 tane temel bileşen fonksiyonunun skor değerleri verilmektedir:

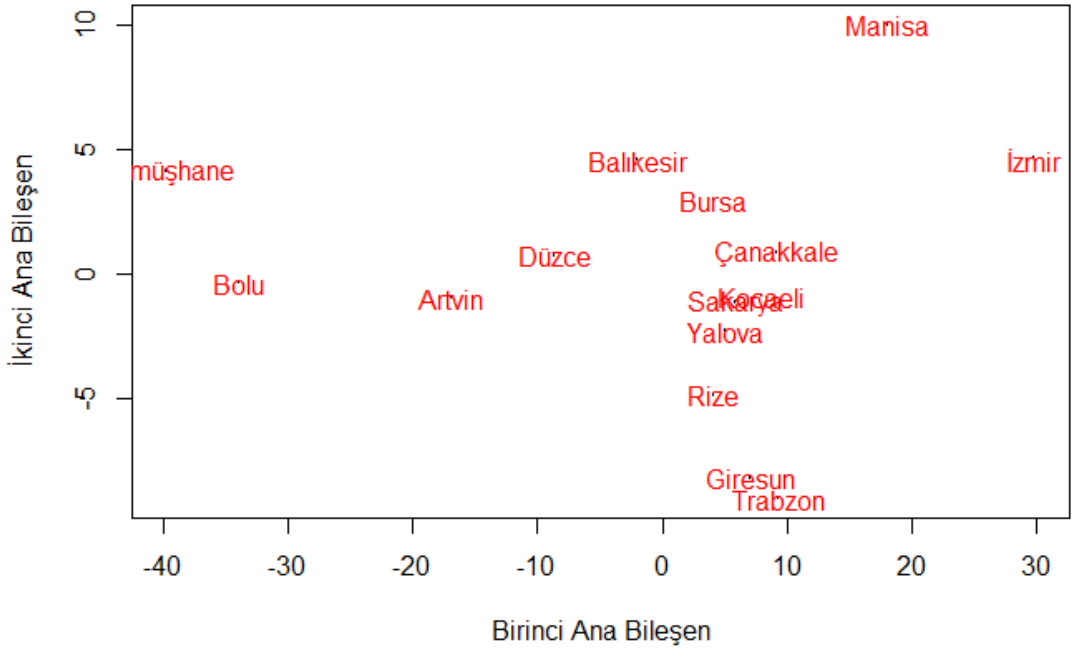
Tablo 4.5 Temel bileşen fonksiyonu skor değerleri (sıcaklık verisi)

	PCA.1	PCA.2	PCA.3	PCA.4
	%90.7	%6.8	%1.5	%0.5
İzmir	29.883743	4.6224038	0.46669795	0.8719430
Manisa	18.083431	10.0208075	-0.92611928	-0.8958439
Bolu	-33.913176	-0.3728106	4.80326861	-2.2173281
Kocaeli	7.972744	-0.8922036	1.46014408	-0.4181718
Sakarya	5.959075	-1.1646716	3.44408664	0.3620429
Düzce	-8.581876	0.7172225	1.72707891	2.9024107
Balıkesir	-1.943734	4.5817839	-0.85586896	-0.7373504

Tablo 4.5 Temel bileşen fonksiyonu skor değerleri (sıcaklık verisi)(devamı)

Bursa	4.158476	2.9772322	0.75877992	0.5659524
Çanakkale	9.184399	0.8397706	-0.99878940	-1.7154366
Yalova	5.154495	-2.3077406	0.09505256	-0.4071792
Artvin	-16.857597	-0.9765281	-0.14799342	2.5991835
Giresun	7.089664	-8.2302517	-1.72099353	-0.7809858
Gümüşhane	-39.710219	4.1281376	-4.53536151	-0.1456212
Rize	4.141053	-4.8930838	-2.91033494	0.2598965
Trabzon	9.379522	-9.0500681	-0.65964764	-0.2435122

En yüksek değişkenliği açıklama yüzdesine sahip birinci ve ikinci temel bileşen fonksiyonları skorlarının dağılımı şu şekildedir:



Şekil 4.22 Temel bileşen fonksiyonları skor dağılımları (sıcaklık verisi)

Temel bileşen fonksiyonlarının dağılımını %90.7 ile en yüksek değişkenliği açıklama yüzdesine sahip birinci temel bileşen fonksiyonu açısından değerlendirildiğinde, en yüksek pozitif skor değerlerine sahip İzmir ve Manisa illeri ve en yüksek negatif skor değerlerine sahip Gümüşhane ve Bolu illerinin birinci temel bileşen fonksiyonu üzerinde en etkili iller olduğu görülmektedir. Birinci temel bileşen fonksiyonu ile iller arasındaki aylık ortalama sıcaklıklardaki değişkenliğin özellikle Temmuz ve Aralık aylarından kaynaklandığı bilgisi yukarıdaki sonuçları desteklemektedir. Gerçekten de en yüksek ortalama sıcaklıklara sahip İzmir ve Manisa illeri ile en

düşük ortalama sıcaklıklara sahip Gümüşhane ve Bolu illeri birinci temel bileşen fonksiyonu üzerinde en etkili illerdir.

İzmir ili en yüksek sıcaklıklara sahip il olması yanında elektrik tüketiminin de en çok yapıldığı ilken en düşük sıcaklıklara sahip Gümüşhane ve Bolu illerinde elektrik tüketiminin de düşük olduğu görülmektedir. Buradan hava sıcaklığının yüksek olmasının elektrik tüketimini tetikleyen bir durum olduğu sonucu çıkartılabilir.

Son yıllarda Türkiye'deki klima satışlarında %200 e varan artışlar da göz önüne alındığında hava sıcaklığının artmasıyla klima kullanımının özelikle İzmir gibi yüksek sıcaklıklara sahip illerde yaz aylarında elektrik tüketimini arttırdığı görülmektedir. Fakat kış aylarında elektrikli ısıtıcılar ya da klima kullanmak doğalgaza oranla daha maliyetli olduğundan, havaların soğumasının elektrik tüketimini arttırdığı fakat çok büyük bir değişikliğe sebep olmadığı görülmektedir.

4.2 Uygulama 2: Covid-19 Pandemisi

İlk kez 2019 yılında Çin'in Wuhan şehrinde ortaya çıkan koronavirüs salgını hızla yayılarak bütün dünyayı etkisi altına almıştır. Dünya Sağlık Örgütünün Mart ayında yaptığı açıklama ile bu salgın pandemi olarak ilan edilmiş ve salgınının daha fazla yayılmasını önlemek için tüm dünyada gerekli önlemler alınmıştır. Ancak toplam vaka sayısı 7 milyona, toplam ölüm ise 400 bine ulaşmıştır ve bu sayılar artmaya devam etmektedir.[74]

Bu çalışmada aşağıdaki ülkelerde ilk vakanın ortaya çıkmasıyla 83 günde toplam vaka sayılarındaki değişkenliği incelemek amacıyla Fonksiyonel Temel Bileşen Analizinin bir uygulamasına yer verilmektedir. Veriler Our World In Data (OWID)[75] internet sitesinden temin edilmiştir.

Tablo 4.6 Covid-19 koronavirüs salgının yayıldığı ülkeler

Afghanistan	Belgium	Canada	Ecuador	Kenya
Albania	Bhutan	Chile	Egypt	Kuweit
Algeria	Bolivia	China	Ethiopia	Lebanon
Andorra	BosniaH	C.Rica	France	Luxembourg
Argentina	Brazil	C. d'Ivoire	Gabon	Macedonia
Armenia	Brunei	Cuba	Georgia	Malaysia
Australia	Bulgaria	Czech R.	Germany	Maldives
Austria	B.Faso	Congo	Ghana	Malta

Tablo 4.6 Covid-19 koronavirüs salgının yayıldığı ülkeler (devamı)

Azerbaijan	Cambodia	Denmark	Honduras	Moldova
Bahrain	Cameroon	Dominican	Iceland	Monaco
Netherlands	Portugal	S.Africa	Trinidad	
Nigeria	Qatar	S.Korea	Tunusia	
Norway	Romania	Spain	Turkey	
Oman	Russia	Sri Lanka	Ukraine	
Pakistan	S. Marino	Sudan	U. Arab	
Palestine	S. Arabia	Sweden	U.K	
Panama	Serbia	Switchzerland	USA	
Peru	Singapore	Taiwan	Vietnam	
Philippines	Slovakia	Thailand		
Poland	Slovenia	Togo		

Bu uygulama ilk olarak 98 ülkenin 83 ayrık noktada gözlemlenen Covid-19 toplam vaka verileri, B-Splayn Baz Fonksiyonlar ve Pürüzlü Ceza Yöntemi ile sürekli birer fonksiyona dönüştürülmüştür. Böylelikle bireysel fonksiyonlar ve türev fonksiyonları ayrıntılı bir şekilde incelenerek ülkelerin toplam vaka sayılarındaki artışın zamana göre değişimi gözlemlenmiştir. Ardından ülkeler arasındaki günlük vaka sayılarındaki değişkenliği incelemek amacıyla Düzgünleştirilmiş Fonksiyonel Temel Bileşenler Analizi(FTBA) uygulanmış ve en yüksek değişkenliği açıklama oranına sahip olan birinci temel bileşen fonksiyonu üzerinde etkili olan ülkeler tespit edilmiştir.

Bu çalışmada Covid-19 pandemisinden en çok etkilenen ülkelere ait 98x83 ayrık veri noktası bulunmaktadır ve tanımlayıcı istatistikler şu şekildedir.

Tablo 4.7 Covid-19 vakası görülen ülkeler için tanımlayıcı istatistikler

	Min.	Q1(1.Kartil)	Medyan	Ortalama	Q3(3.Kartil)	Mak.
Afghanistan	1	75	784	2225	3308	11831
Albania	2	204.5	548	536	853	1099
Algeria	1	170	1825	2487	4224	7918
Andorra	1	433.5	724	577.1	760.5	851
Argentina	1	639.5	2560	3579.1	5276.5	13920
Armenia	1	453	1339	2214	3244	8927
Australia	1	15	59	1542	2982	6458
Austria	2	1174	12297	9294	15222	16154
Azerbaijan	1	143.5	1283	1424.2	2165.5	4568
Bahrain	1	217.5	700	1590	2618	6747
Belgium	1	1	886	10196	19061	45325
Bhutan	1	5	7	11.28	17.5	47

Tablo 4.7 Covid-19 vakası görülen ülkeler için tanımlayıcı istatistikler(devamı)

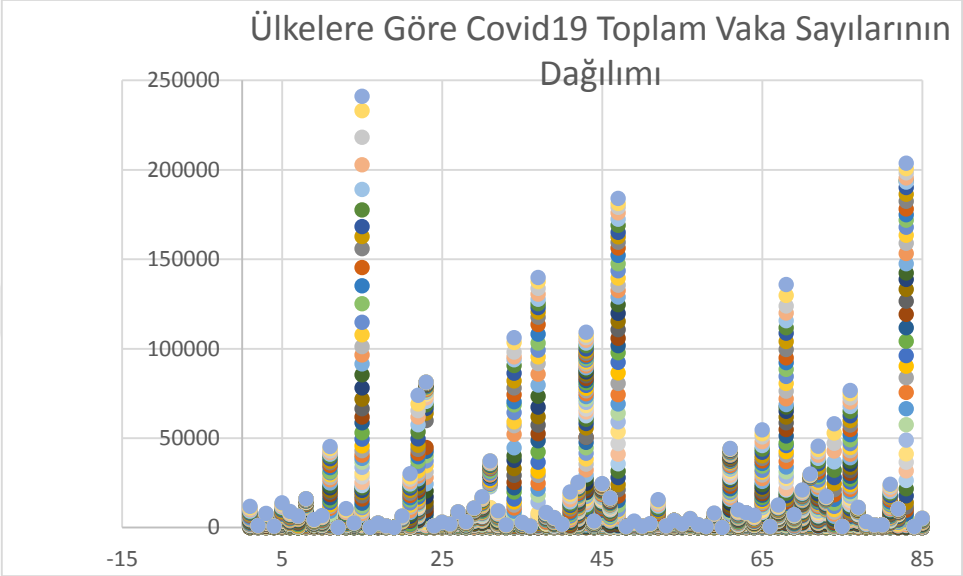
Bolivia	2	119	609	1980	2898	10531
BosniaH	2	498	1367	1322	2170	2534
Brazil	1	262.5	12056	45076.9	64194.5	241080
Brunei	1	128	138	122.2	141	141
Bulgaria	2	352.5	915	1121.7	1960	2513
B.Faso	2	274.5	609	515.1	769.5	883
Cambodia	1	1	24	53.92	116	122
Cameroon	1	252	1163	1859	2744	6585
Canada	1	8	51	4452	4346	30081
Chile	1	1032	7917	14999	21330	73997
China	27	310.5	40206	41120.1	80031.5	81346
C.Rica	1	279	649	536.5	769	1000
C. d'Ivoire	1	206	1004	1135	1942	3024
Cuba	2	251	1235	1081	1820	2107
Czech R.	3	1080	5902	4726	7746	8813
Congo	1	123	377	796.1	1205.5	3325
Denmark	1	1070	5071	4984	8774	10968
Dominican	1	1005	4964	5997	10490	17285
Ecuador	1	1307	7858	13845	29465	37355
Egypt	1	57	609	1969	3238	9400
Ethiopia	1	32	116	236.4	267.5	1486
France	3	11	423	19709	27194	106206
Gabon	1	19.5	167	633.3	1054	2902
Georgia	1	46	234	295.9	560.5	721
Germany	1	15	902	30785	54923	139897
Ghana	2	204.5	1154	2591.2	5469	8548
Honduras	2	243	519	1425	2286	5690
Iceland	1	369.5	1648	1174.8	1797	1803
India	1	3	73	2472	1681	19984
Indonesia	2	1220	6284	8080	13378	25216
Iran	2	8521	44606	47005	84154	109286
Iraq	1	159	1031	1183	1847	3554
Ireland	1	845.5	8928	10820.2	21004.5	24506
Israel	1	106	6252	7310	14737	16548
Italy	3	3	12462	52628	108183	183957
Jamaica	2	45.5	252	274.9	508	590
Japan	1	22.5	144	543.2	819	3654
Jordan	1	288.5	428	413.7	579	755
Kenya	1	116	320	526.8	747.5	2216
Kuweit	3	136	743	2611	3364	15691
Lebanon	1	120	541	456.5	708.5	911
Luxembourg	1	1718	3480	2725	3865	4008
Macedonia	1	210	1081	943.5	1555.5	2040
Malaysia	1	18.5	50	1075	1913.5	5072

Tablo 4.7 Covid-19 vakası görülen ülkeler için tanımlayıcı istatistikler (devamı)

Maldives	2	17	52	404	812.5	1672
Malta	1	144	413	331.6	485	615
Moldova	1	280.5	2472	2899.2	4897	8098
Monaco	1	53.5	94	74.07	96	99
Netherlands	1	2256	20549	21079	38609	44249
Nigeria	1	114	627	2302	4275	9855
Norway	1	1366	5863	4697	7569	8249
Oman	2	53.5	546	1486.7	2465	6794
Pakistan	2	835.5	5374	12575.9	19644.5	54601
Palestine	7	94	313	323.5	547	625
Panama	1	945	4273	4766	8176	12531
Peru	1	607.5	12491	32034.8	56671.5	135905
Philippines	1	3	111	1547	3056	7192
Poland	1	691.5	6674	7703.2	13534	20931
Portugal	1	1150	15472	13980	25089	29912
Qatar	1	531.5	3428	10352.8	16666.5	45465
Romania	1	292.5	5202	6536.1	12109	17387
Russia	2	2	30	5641	3162	57999
S. Marino	1	138.5	333	338.7	566	656
S. Arabia	1	956	5862	18374	31095	76726
Serbia	1	420.5	5318	5246	9819.5	11275
Singapore	3	54	112	456.9	563	3252
Slovakia	1	295	1049	879.6	1450	1520
Slovenia	1	552.5	1248	1003	1448.5	1469
S.Africa	1	818	2506	5414	7690	24264
S.Korea	1	26	3526	4413	8848	10450
Spain	1	2	5958	56565	122899	202649
Sri Lanka	1	1	42	103.6	181.5	588
Sudan	1	7	162	1120	1891	5310
Sweden	1	1	620	3260	5206	16004
Switchzerland	1	2425	21574	17112	29030	30504
Taiwan	1	18	40	108	180	388
Thailand	1	19	35	447.1	482.5	2518
Togo	1	39.5	88	149	228.5	452
Trinidad	1	93	115	95.45	116	117
Tunusia	1	252.5	866	665.7	1031	1071
Turkey	1	19528	101790	86462	143932	166422
Ukraine	1	895.5	7647	9191.9	16636	25411
U. Arab	1	8.5	85	1257	1024	8756
U.K	2	9	456	21720	27312	129044
USA	1	12.5	89	62687.2	30976.5	**529951**
Vietnam	1	14	16	78.67	145	268

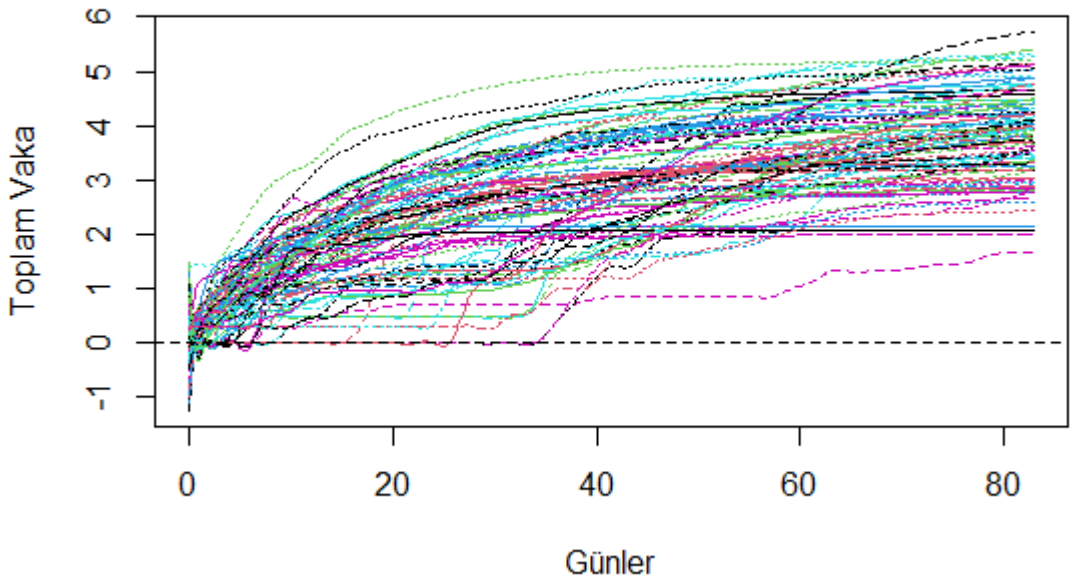
Min.	Q1	Medyan	Ortalama	Q3	Mak.
1	71	663	8382	4254	**529951**

Tablo 4.7 deki minumum değerler ilk günkü vaka sayısını, maksimum değerler ise toplam 83 gündeki vaka sayısını göstermektedir. Tanımlayıcı istatistik değerleri incelendiğinde Çin ilk günkü vaka sayısındaki fazlalık ile ve Amerika maksimum vaka sayısındaki artış ve fazlalık ile dikkat çekmektedir. Bu verilerin dağılımı logaritmik normal dağılıma uygundur ve ilk vakanın ortaya çıkmasıyla bu ülkelerdeki toplam vaka sayısındaki değişim aşağıdaki grafikte verilmektedir.



Şekil 4.23 Ülkelere göre covid-19 vaka sayılarının dağılımı

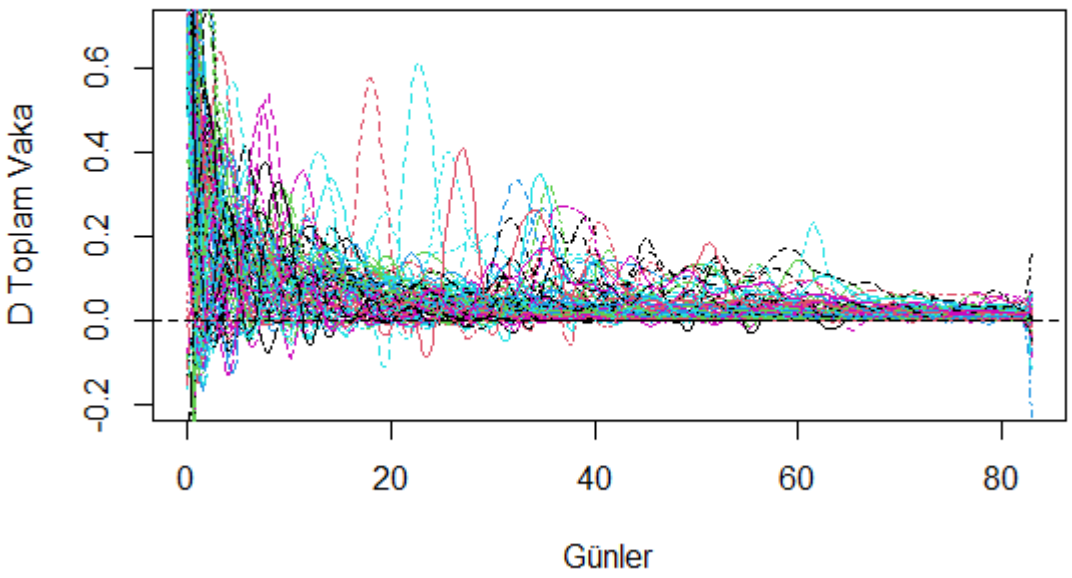
Şekil 4.23 ten toplam vaka sayılarının bazı günlerde diğer günlerden çok daha fazla olduğu özellikle 15. 47. ve 80. Günlerde dikkate değer bir artış olduğu gözlemlenmektedir. Fakat her bir ülkenin bireysel olarak günlük vaka sayılarını bu grafikten incelemek mümkün olmadığından B- Splayn Baz fonksiyonları ve Pürüzlü Ceza Yöntemi kullanılarak bu veri seti fonksiyonel bir forma dönüştürülmüştür.



Şekil 4.24 Covid-19 vakaları görülen ülkelerin bireysel fonksiyonları

Şekil 4.24 te verilen bireysel fonksiyonlar incelendiğinde hemen hemen her ülkede ilk 15 günde toplam vaka sayısının hızla arttığı görülmektedir. Veri seti incelendiğinde yukarıdaki grafikte en üstteki iki eğrinin yani ilk 15 günde vaka sayısındaki artışın en yüksek olduğu ülkelerin sırasıyla Türkiye ve İran olduğu görülmektedir. Vaka sayısında artışın ilk 15 günde en az olduğu ülke Belçika iken 83 günde en az vakaya sahip ülke Butan Krallığıdır. Bireysel fonksiyonların incelenmesinde dikkat çeken bir diğer eğri ise 70. Günden itibaren aniden vaka sayısındaki artış ile Türkiye ve İran'ı da geçen Amerika'dır.

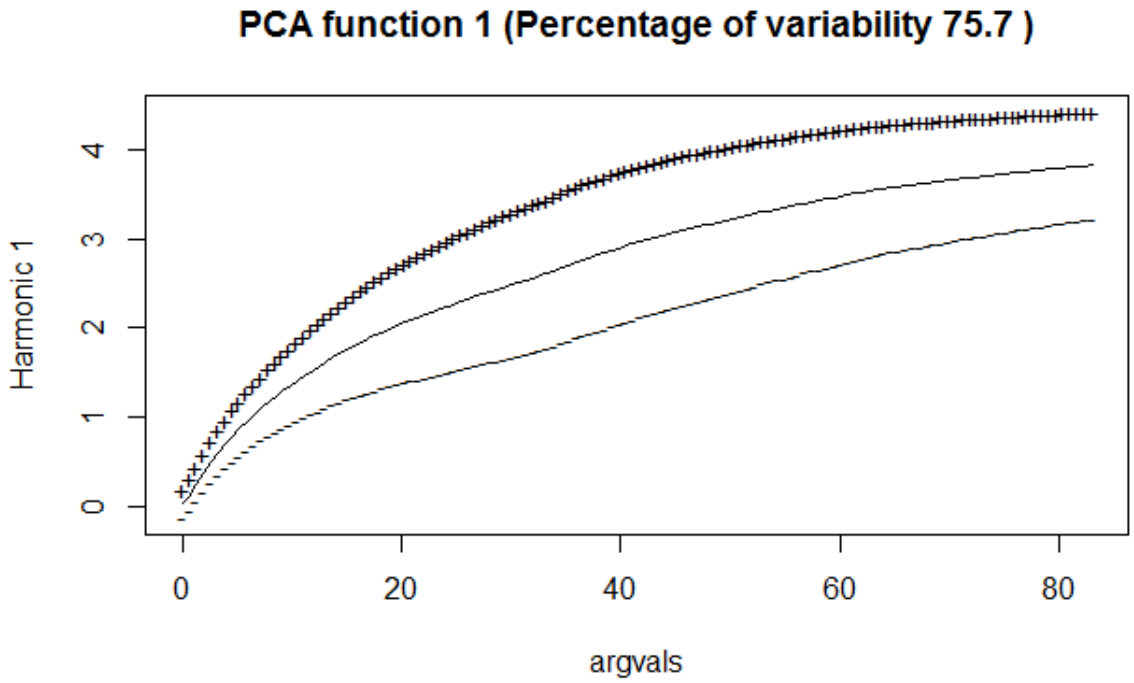
Şekil 4.24 te gösterilen bireysel fonksiyonların maksimum ve minumum noktalarını, fonksiyonların iniş- çıkış yaptığı noktaları tespit etmek ve bu fonksiyonları daha yakından inceleyebilmek için aşağıdaki şekilde gösterilen türev fonksiyonları elde edilmiştir.



Şekil 4.25 Covid-19 türev fonksiyonları

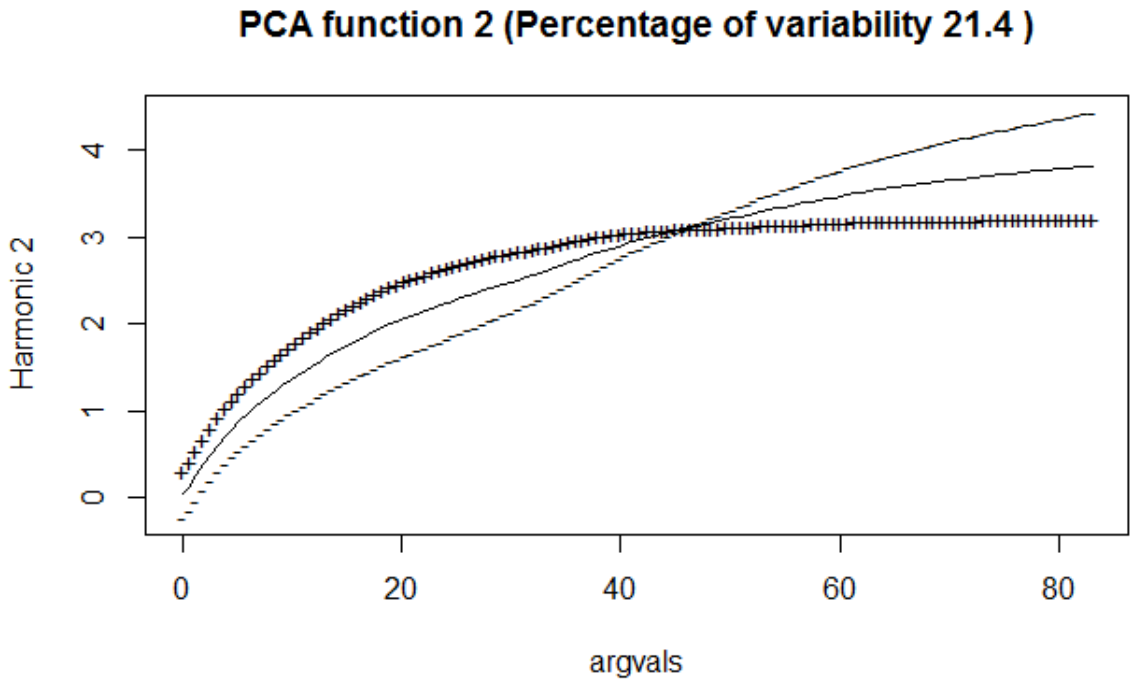
Şekil 4.25 te verilen birinci türev fonksiyonları incelendiğinde vaka sayısının sürekli arttığı ve genel olarak vaka artış hızının ilk 15 günde maksimum seviyelere ulaştığı gözlemlenmektedir. Buna rağmen bazı eğrilerde 65. güne kadar vaka artışlarında zaman zaman ani artış ve azalışların olduğu fakat bu noktadan itibaren tüm eğrilerde vaka artışlarında bir normalleşme olduğu söylenebilir.

Bireysel fonksiyonlar ve türev fonksiyonları incelendiğinde vaka artış hızının ülkeden ülkeye değiştiği görülmektedir. Ülkeler arasındaki vaka artış hızındaki değişkenliği açıklamak için bir sonraki adım olarak Fonksiyonel Temel Bileşenler Analizi uygulanmıştır.



Şekil 4.26 Covid-19 birinci temel bileşen fonksiyonu

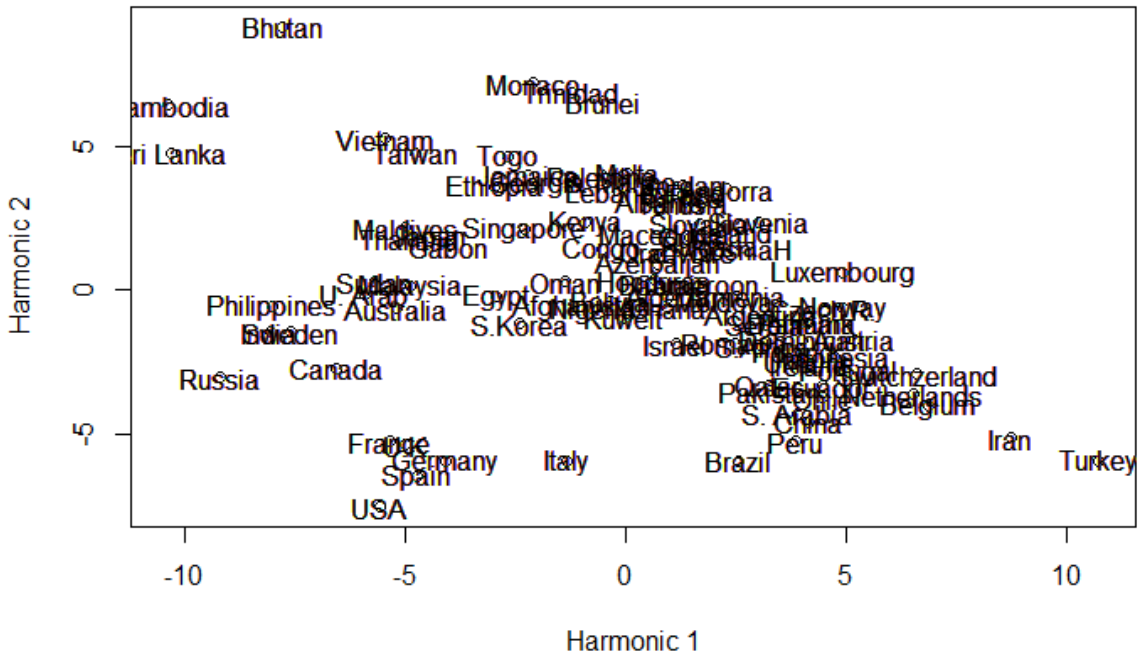
Şekil 4.26 da verilen birinci temel bileşen fonksiyonu incelendiğinde birinci temel bileşen fonksiyonunun toplam değişkenliğin %75.7 sine karşılık geldiği ve hiçbir noktada işaret değiştirmediği gözlemlenmektedir. Ortalamadan sapmaların genel anlamda yüksek olduğu fakat özellikle 15. ve 65. günler arasında arttığı söylenebilir. Türev fonksiyonu incelendiğinde vaka artış hızının 15. günde maksimuma ulaşmasına rağmen bazı ülkelerde vaka sayılarında zaman zaman ani artışların olduğu ve 65. güne kadar bu şekilde devam ettiği gözlenmişti. Buradan %75.7 değişkenliği açıklama yüzdesine sahip birinci temel bileşen fonksiyonu ile ülkeler arasındaki vaka artış hızındaki temel değişkenliğin özellikle 15. Ve 65. Günler arasında kaynaklandığı söylenebilir.



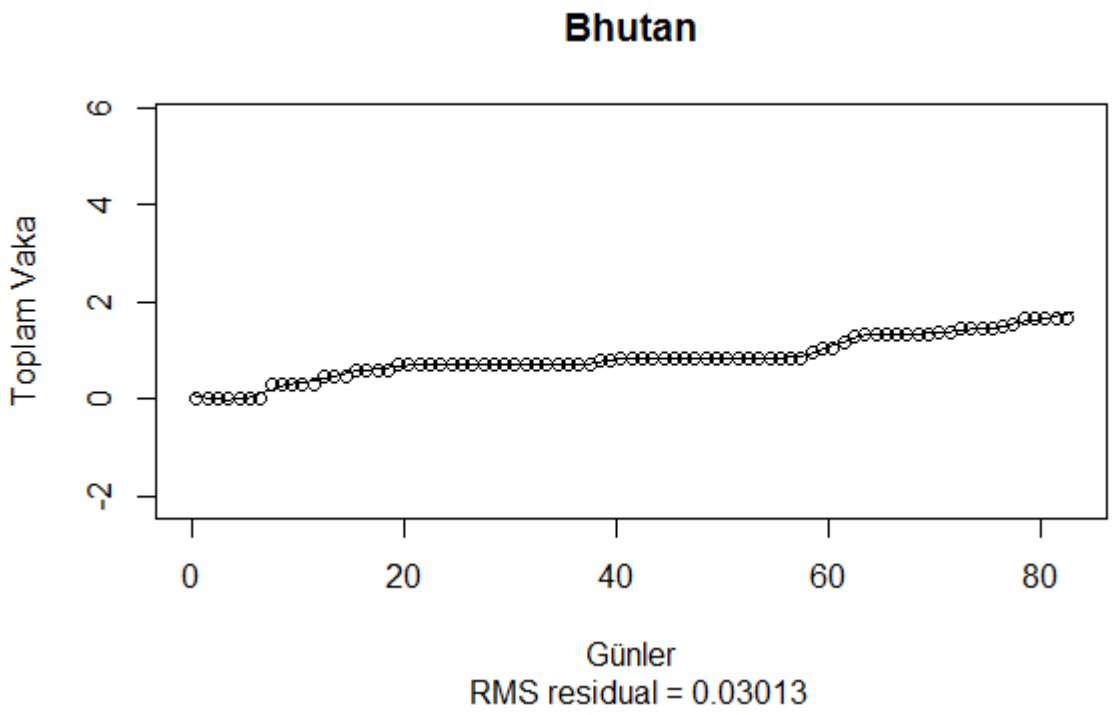
Şekil 4.27 Covid-19 ikinci temel bileşen fonksiyonu

Şekil 4.27 incelendiğinde ikinci temel bileşen fonksiyonunun toplam değişkenliğin %21.4 üne karşılık geldiği ve 47. inci noktada yani ilk vakanın ortaya çıkmasının 47. gününde işaret değiştirdiği görülmektedir. Buradan vaka sayısındaki artışta 47.günün bir dönüm noktası olduğu anlaşılmaktadır. Bu nedenle %21.4 değişkenlik açıklama yüzdesine sahip ikinci temel bileşen fonksiyonu ile ülkeler arasındaki vaka sayısı artışındaki temel değişkenliğin 47. günden kaynaklandığı söylenebilir.

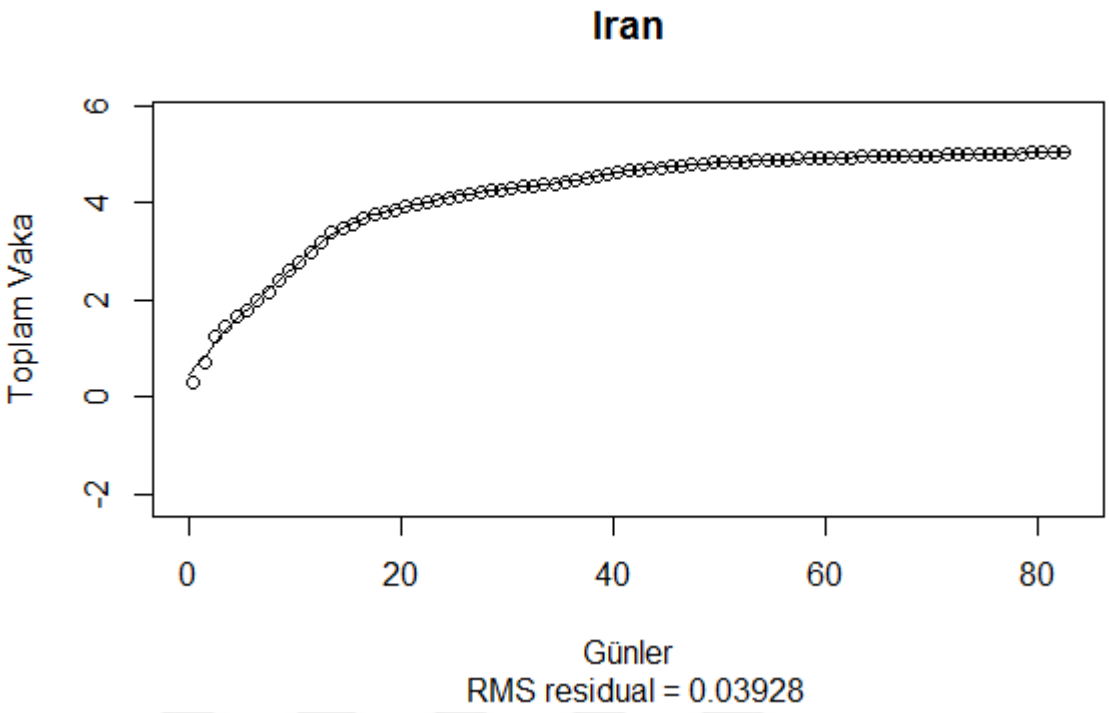
Yukarıda temel bileşen fonksiyonları ortalamayla kıyaslanarak grafikleri incelenmiş ve yorumlanmıştır. Daha ayrıntılı bir analiz için aşağıda bu temel bileşen fonksiyonlarının skorlarının dağılımı ve bazı ülkelerin bireysel fonksiyonları verilmiştir.



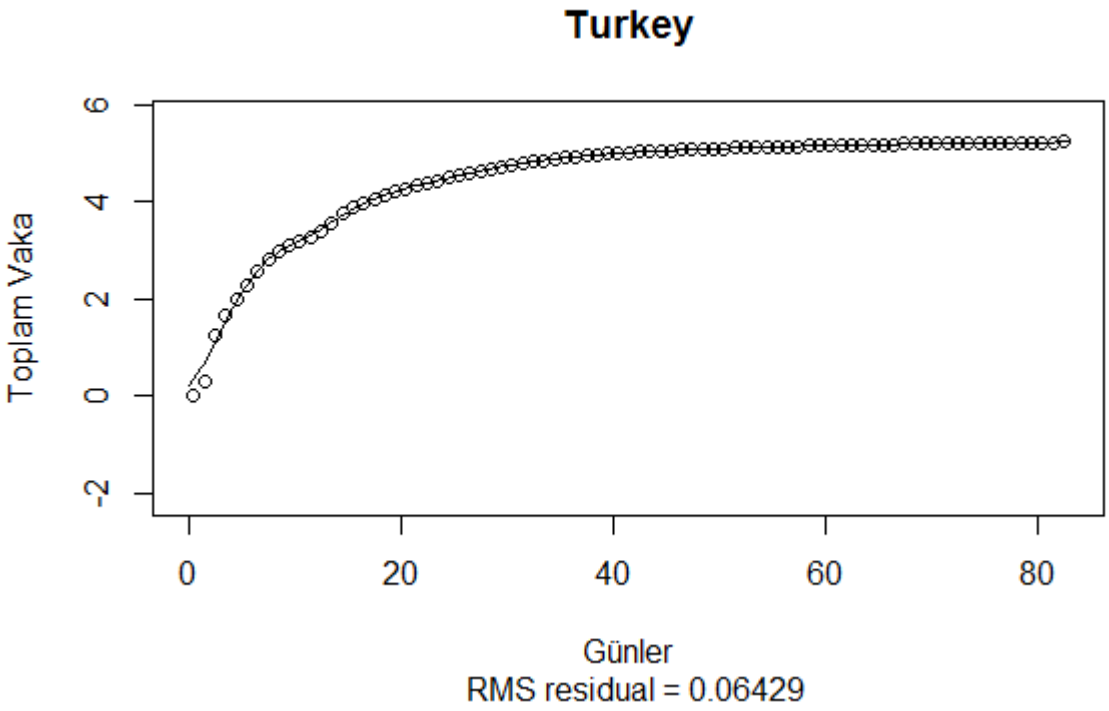
Şekil 4.28 Temel bileşen fonsiyonları skorlarının dağılımı (covid-19)



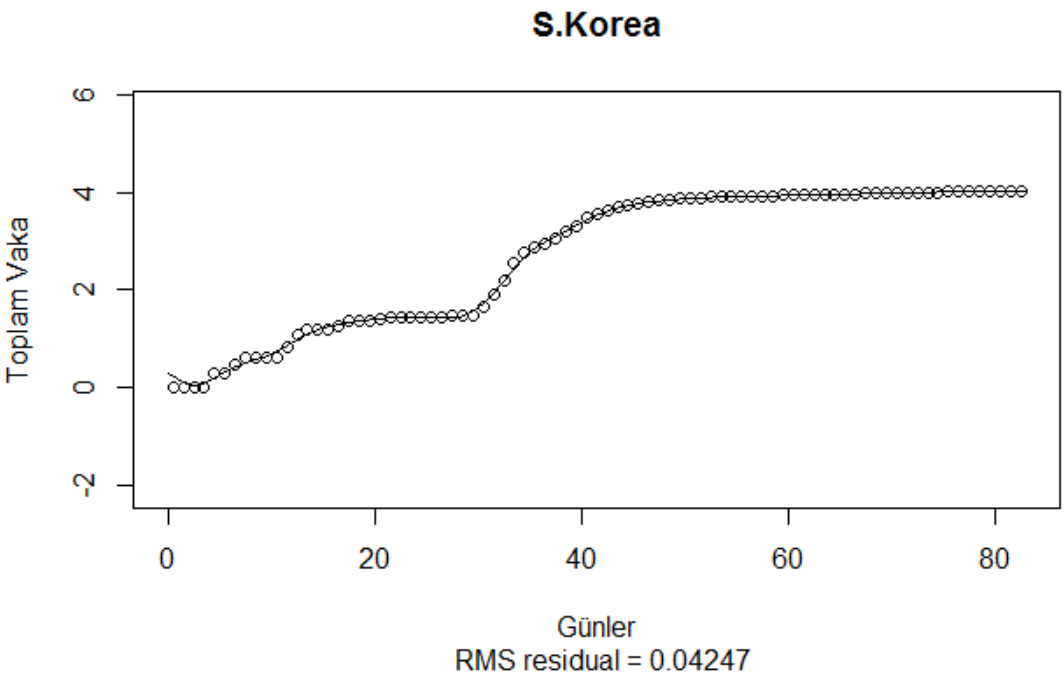
Şekil 4.29 Butan Krallığı bireysel fonksiyonu



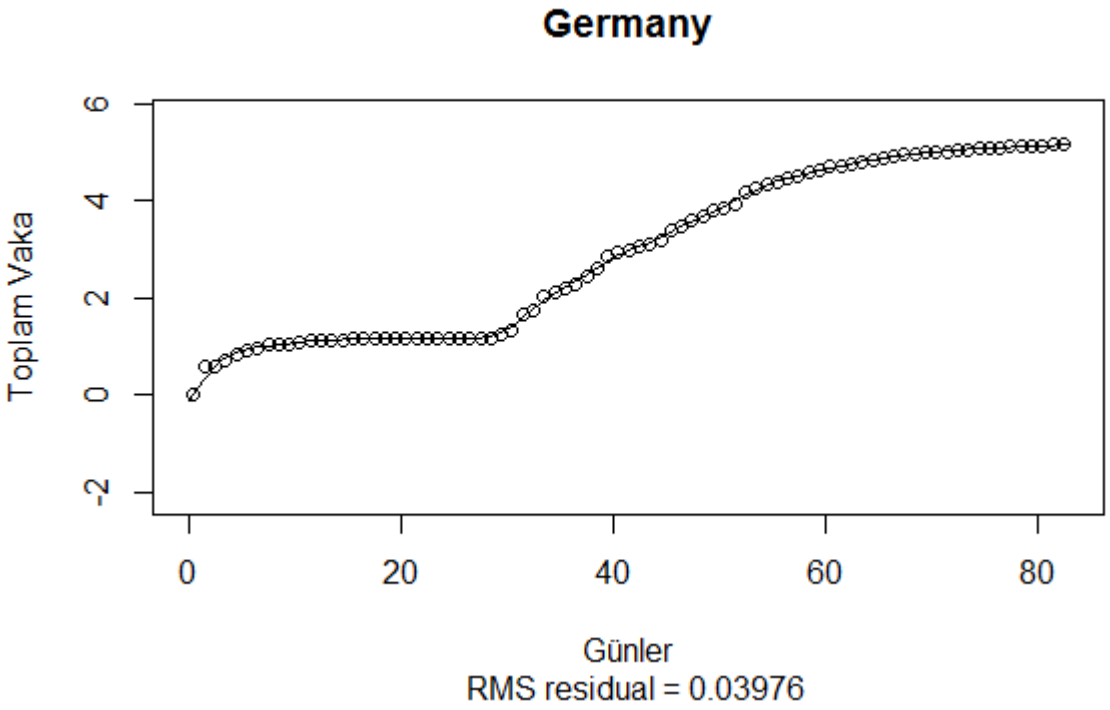
Şekil 4.30 İran bireysel fonksiyonu



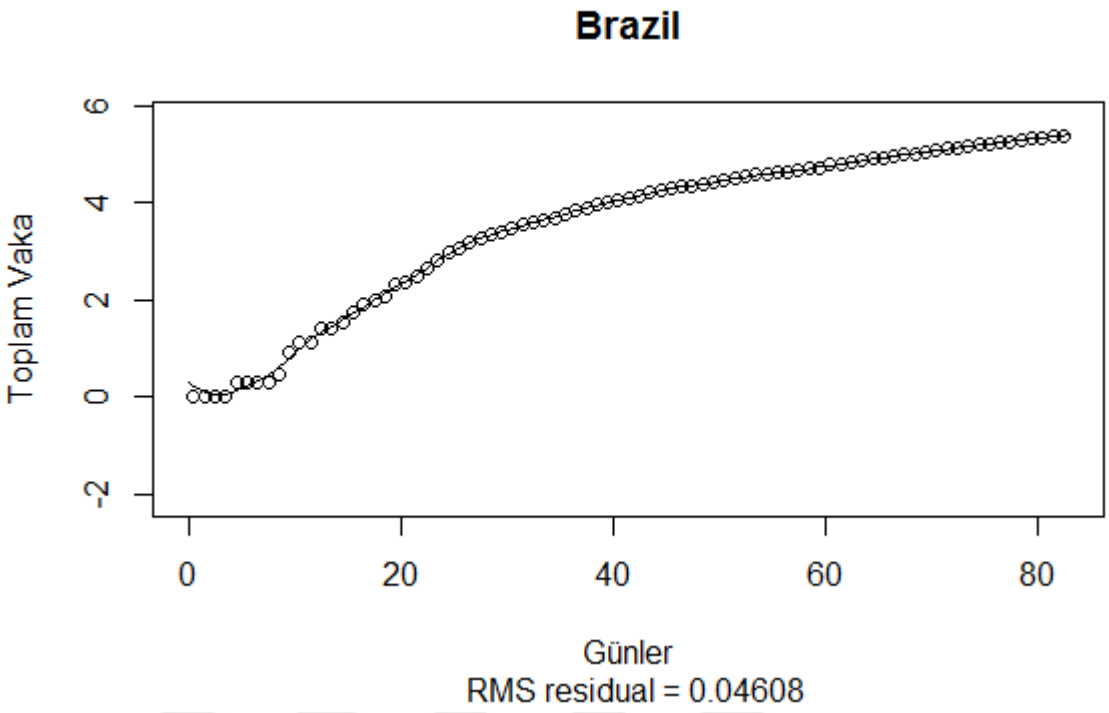
Şekil 4.31 Türkiye bireysel fonksiyonu



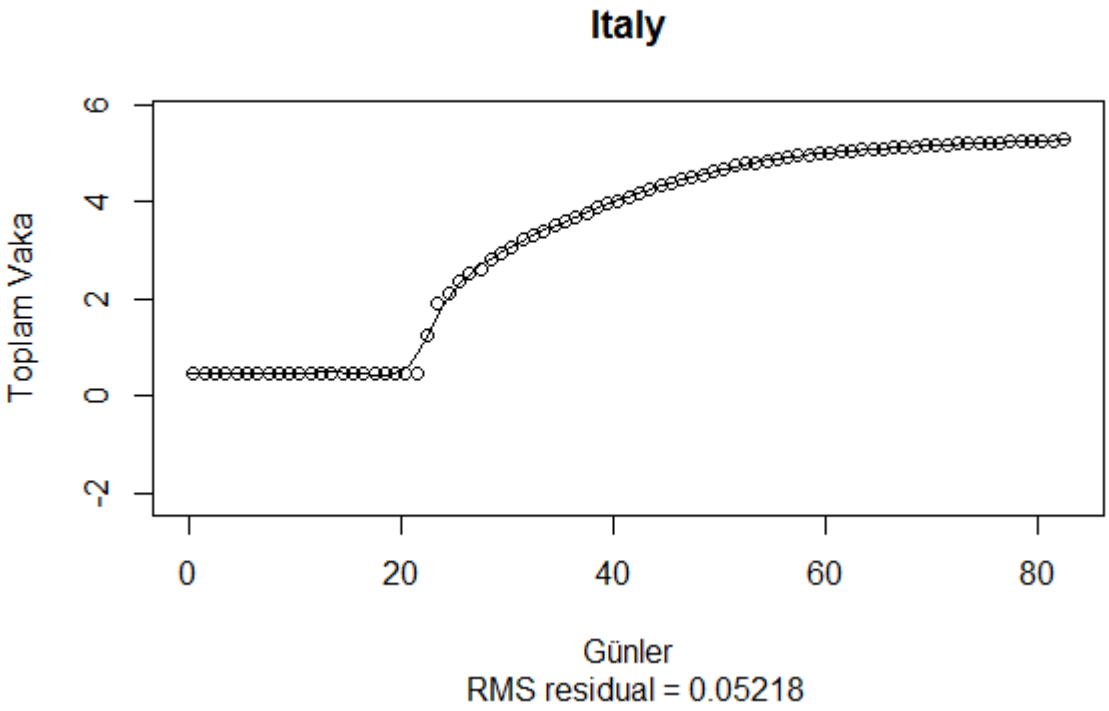
Şekil 4.32 Güney Kore bireysel fonksiyonu



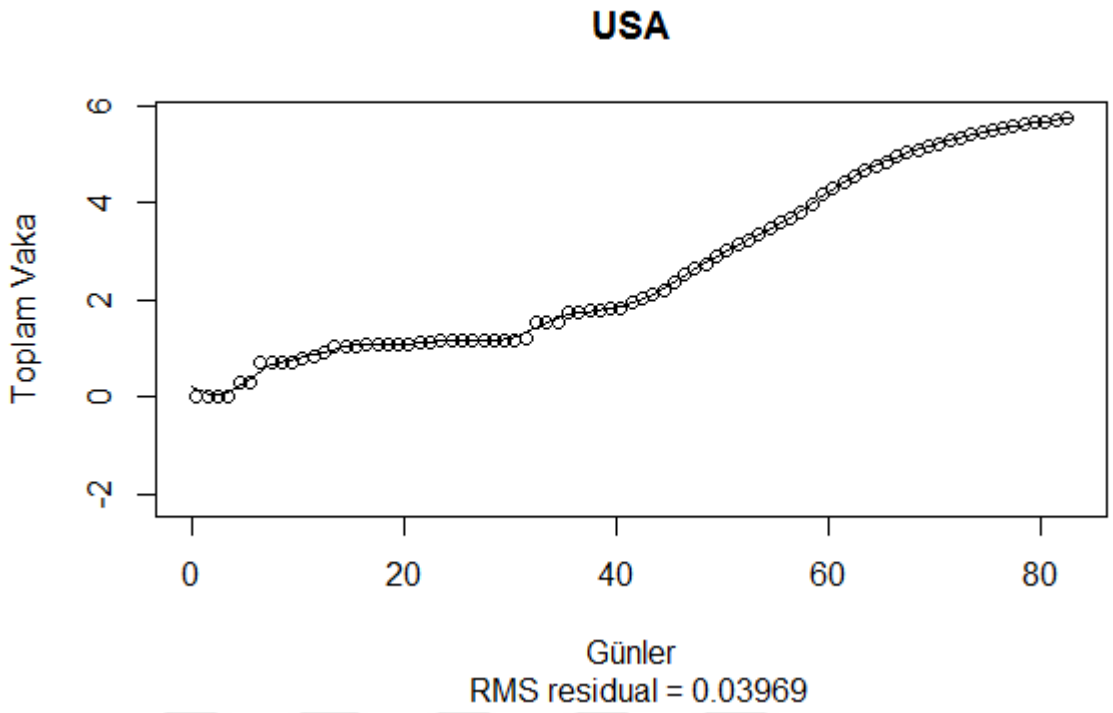
Şekil 4.33 Almanya bireysel fonksiyonu



Şekil 4.34 Brezilya bireysel fonksiyonu



Şekil 4.35 İtalya bireysel fonksiyonu



Şekil 4.36 Amerika bireysel fonksiyonu

Şekil 4.28 incelendiğinde %75. 7 ile en yüksek değişkenliği açıklama yüzdesine sahip birinci temel bileşen fonksiyonu açısından değerlendirildiğinde en yüksek pozitif temel bileşen skoruyla Türkiye ve İran ülkelerinin en etkili ülkeler olduğu gözlenmektedir. Ülkelerin bireysel fonksiyonları da incelendiğinde Türkiye ve İran'ın ilk 15 günde vaka artışı en yüksek olan ülkeler olduğu rahatlıkla gözlenmektedir. Bunun yanı sıra vaka sayılarının ve vaka artış hızının çok olduğu Amerika, İspanya, Fransa, Almanya, İtalya ve Rusya da asıl artışın 15. Günden sonra gerçekleştiği fakat Brezilyada vaka artışının Türkiye ve İran'daki gibi olduğu görülmektedir.

%21.4 değişkenliği açıklama yüzdesine sahip ikinci temel bileşen fonksiyonu açısından bir değerlendirilme yapıldığında en yüksek pozitif skor değerine sahip Butan Krallığı, Monako, Trinidad, Brunei, Kamboçya ve Sri Lanka ülkelerinin en etkili ülkeler olduğu ve en düşük vaka sayılarına sahip olduğu gözlenmektedir. İkinci temel bileşen fonksiyonundaki değişkenliğin 47. Günden kaynaklandığı bilgisiyle bu ülkelerin bireysel fonksiyonları incelendiğinde vaka sayılarında az da olsa 47. günden sonra bir artış olduğu gözlemlenmektedir.

Şekil 4.31 de verilen Türkiye ve Şekil 4.32 de verilen Güney Kore bireysel fonksiyonlarını karşılaştırdığımızda Türkiye'de vaka artış hızının en fazla olduğu günlerin ilk 15 günde olduğu, Güney Kore'de ise 30. Gün ile 42. Gün arasında olduğu gözlenmektedir. Bunun yanı sıra ilk 83 gündeki toplam vaka sayısı Türkiye'de 166.422 iken Güney Kore'de ise 10.450 dir. Türkiye'nin vaka sayısı Güney Kore'nin neredeyse 16 katıdır. Buradan virüsün yayılma hızını yavaşlatmak ve vaka sayılarını minumum seviyede tutmak için ilk 15 gün alınan önlemlerin önemi anlaşılmaktadır. Özellikle iki ülke arasındaki vaka sayılarındaki farkın bu kadar çok olmasının ilk vakanın ortaya çıkmasıyla birlikte Güney Kore'de alınan önlemlerin ve sayıca daha fazla test yapılmasının etkili olduğu düşünülmektedir.

5
SONUÇ VE ÖNERİLER

Veri elde etme ve depolamada meydana gelen gelişmeler özellikle fonksiyonel verilerin varlığını büyük ölçüde artırmaktadır. Fonksiyonel veriler; eğriler, yüzeyler veya zamana göre değişen herhangi bir şey olabilir. Fonksiyonel veri analizinde, ayrık veri noktaları sürekli birer fonksiyona dönüştürüldüğünden, tek tek veri noktalarını incelemek yerine fonksiyonlar incelenir. Bu da daha hızlı ve daha zengin bir analiz imkanı sunmakta ve giderek modern veri analizinde önemli bir yer tutmaktadır. Bu amaçla bu çalışmada en popüler Fonksiyonel Veri Analizi tekniği olan Fonksiyonel Temel Bileşenler Analizi (FTBA) tanıtılmış ve FTBA ile ilgili ayrıntılı bir literatüre yer verilmiştir.

FTBA çok değişkenli istatistiksel analizde iyi bilinen bir boyut küçültme tekniği olan temel bileşen analizinin fonksiyonel halidir. Sıradan temel bileşen analizi ile karşılaştırıldığında, FTBA' nin kayıp verilerle kolaylıkla baş edebilmesi, farklı boyutlarda eşit aralıklı olmayan fonksiyonlarla çalışabilme imkanı vermesi, fonksiyonların türevlenebilir olma özelliğinden istenilen dereceden türev fonksiyonların incelenebilmesi ve bunun sonucunda daha ayrıntılı bir analiz yapma imkanı vermesi ve görsel olarak da bunu desteklemesi diğer boyut indirgeme tekniklerinden üstün kılmaktadır.

Bu çalışmada Fonksiyonel veri analizinde en sık kullanılan boyut indirgeme tekniği olan Fonksiyonel Temel Bileşen Analizinin iki farklı gerçek veri seti üzerine uygulamasına yer verilmiştir. İlk olarak Türkiye'deki 15 ilin aylık ortalama elektrik tüketimi ve sıcaklıkları verisi üzerine ayrı ayrı bir FTBA uygulaması yapılmıştır. Temel bileşen fonksiyonları ile iller arasındaki elektrik tüketimindeki değişkenliğin ana modunun kış aylarından kaynaklandığı, daha sonra en fazla tüketimin olduğu ve tüketimdeki seyrin bazı iller için yön değiştirdiği 2016 yılından kaynaklandığı görülmüştür. Fonksiyonların yıllara ve bazı illere göre türev grafikleri ayrı ayrı incelenerek tüketimdeki değişkenlik hem yıllar hem de iller bazında incelenmiştir. Fonksiyonel temel bileşen analizi skorları ve dağılım grafikleri, iller arasındaki temel farklılıkları görsel olarak sunmuştur. Bu illerin elektrik tüketimleri nüfus

etkisinden arındırıldığından aralarındaki farklılığın sıcaklıkla bir ilişkisi olup olmadığını anlamak için sıcaklık verisine FTBA uygulanmıştır. Elektrik tüketiminin en yüksek olduğu illerin en yüksek ortalama sıcaklık değerlerine sahip olması yüksek sıcaklıkların klima kullanımından kaynaklı olarak elektrik tüketimini tetiklediği sonucuna ulaşılmıştır.

Türkiye bulunduğu konum itibariyle güneş enerjisi potansiyeli yüksek bir ülke olmasına rağmen elektrik üretiminin sadece %0.009 nun güneş enerjisinden sağlamaktadır. Bu çalışma sonucunda elde edilen veriler artan sıcaklığın elektrik tüketimini arttırdığını gösterdiğinden ve birçok enerji kaynağının önümüzdeki yıllarda tükenme riski olduğundan doğal bir enerji kaynağı olan güneş enerjisinden daha fazla faydalanmalı ve yatırım yapılmalıdır.

İkinci bir uygulama olarak tüm dünyayı etkisi altına alan Covid 19 pandemi verileri üzerine bir uygulama yapılmıştır. Genel anlamda toplam vaka sayısındaki artışın ilk 15 günde hızla arttığı ve 60. Günden sonraki günlerde de normalleştiği gözlenmiştir. Temel bileşen fonksiyonları ile ülkeler arasındaki toplam vaka sayılarının artışındaki değişkenliğin ana modunun 15.inci ve 60.ıncı günler arasında gerçekleştiği ve daha sonra 47.inci günün vaka artışının seyrini değiştirdiği sonucuna ulaşılmıştır.

2019 yılında ilk kez Çin'de orta çıkan Covid-19 virüsü hızla yayılırken birçok ülke virüs kendi ülkesinde gözükmeyene dek hiçbir önlem almamış ve daha sonra yapılan ilk şey ise kendi ülkelerindeki vaka artış hızını düşürmek olmuştur. Analiz sonuçlarından elde edilen verilerle ilk 15 günün vaka artışında kritik düzeyde bir öneme sahip olduğu görülmektedir. Buradan ilk vakanın ortaya çıkmasıyla birlikte gerekli önlemler alınırsa vaka artış hızı düşürülebileceği ve 60 gün içerisinde çok daha az vaka ile normalleşmeye gidilebileceği öngörülmektedir.

Bu çalışmanın ana konusu olan Fonksiyonel veri analizi birçok bilimsel alanda popülerlik kazanmaya devam etmekte ve Fonksiyonel Temel Bileşenler Analizi istatistikçiler için yeni bir çalışma alanı sunmaktadır. Özellikle literatürde çok fazla örneği olmayan çok değişkenli verilere FTBA uygulanması üzerine çalışabilir ve bir sonraki adım olarak fonksiyonel regresyon yapılabilir.

KAYNAKÇA

[1] J.O.Ramsay and B.W. Silverman, "Functional data analysis," Second Edition, Springer-Verlag, New York, 2005.

[2] R. Viviani and G. Grön, and M.Spitzer, "Functional Principal Component Analysis of fMRI Data," Human Brain Mapping, vol. 24, pp. 109-129, 2005.

[3] H. Hwang and K.Jung and Y.Takane and T.S. Woodward, "Functional multiple-set canonical correlation analysis," Psychometrica vol.77, pp. 48-64, 2012.

[4] I. Jackson & S. Sirois, "Infant cognition: going full factorial with pupil dilation," Developmental Science, vol.12, pp.670–679, 2009

[5] W. Bradley and W. Vines and R. L. Nuzzo and D. J. Levitin, "Analyzing Temporal Dynamics In Music: Differential Calculus, Physics, And Functional Data Analysis Techniques," An Interdisciplinary Journal, vol. 23, no. 2, pp. 137-152, 2005.

[6] G. Dona and E. Preatoni and C. Cobelli and R. Rodano and & A. J. Harrison, "Application of functional principal component analysis in race walking: An emerging methodology," Sports Biomechanics, vol. 8, no. 4, pp. 284-301, 2009.

[7] L. Zhou & H. Pan, "Principal Component Analysis of Two-Dimensional Functional Data," Journal of Computational and Graphical Statistics, vol. 23, no. 3, pp. 779-801, 2014.

[8] J. Dauxois and A. Pousse and Y. Romain, "Asymptotic Theory for the Principal Component Analysis of a Vector Random Function: Some Applications to Statistical Inference," Journal Of Multivariate Analysis vol.12, pp. 136-154, 1982.

[9] P. Besse and J. O. Ramsay, "Principal Components Analysis Of Sampled Functions," Psychometrika, vol. 51, no. 2, pp. 285-311, 1986.

[10] P. E. Castro and W. H. Lawton and E. A. Sylvestre, "Principal Modes of Variation for Processes With Continuous Sample Curves," Technometrics, vol.28, no.4, pp. 329-337, 1986.

[11] J.A Rice & B.W. Silverman, "Estimating The Mean and Covariance Structure Nonparametrically When The Data Are Curves," Journal of the Royal Statistical Society: Series B, vol.53, no.1, 1991.

[12] B. W. Silverman, "Smoothed Functional Principal Components Analysis By Choice Of Norm," The Annals of Statistics, vol. 24, no. 1, pp. 1-24, 1996.

[13] J.Y. Choi and H. Hwang and M. Yamamoto and K. Jung and T. S. Woodward, "A Unified Approach To Functional Principal Component Analysis And Functional Multiple-Set Canonical Correlation," vol. 82, no. 2, pp. 427-441, 2017.

[14] P.Feng and J.Qian, "Analyzing and Forecasting the Chinese Term Structure Of Interest Rates Using Functional Principal Component Analysis," China Finance Review International, 2018.

[15] Z. Wang and Y. Sun and P. Li, "Functional Principal Components Analysis of Shanghai Stock Exchange 50 Index," Discrete Dynamics in Nature and Society, 2014.

[16] H. L. Shang, "A survey of functional principal component analysis," AStA Advances in Statistical Analysis, vol. 98, no. 2, pp. 121-142, 2014.

[17] G.D Costanzo and S. Ingrassia, "Functional Principal Component Analysis of FinancialTime Series," Dipartimento di Economia e Statistica, 351-358, 2005.

[18] A.Kneip & J.K Utikal, "Inference Density Families Using Functional Principal Component Analysis," Journal of the American Statistical Association, vol.96, no.454, pp.519-542, 2011.

[19] P. Hall & H.M.Nasab, "On Properties Of Functional Principal Components Analysis," Journal of the Royal Statistical Society, vol.68, no.1, Series B, 2006.

[20] J.O. Ramsay & C. Dalzell, " Some Tools For Functional Data Analysis," Journal of the Royal Statistical Society: Series B., vol.53, no.3, 1991.

[21] F. Yao and T. C. M. Lee, "Penalized spline models for functional principal component analysis", Journal of Royal Statistics, vol.68, no.1, pp. 3-25, 2006.

[22] M. Grith and H. Wagner and W.K. Hardle and A. Kneip, "Functional Principal component analysis For Derivatives of multivariate curves," Statistica Sinica, vol.28, pp. 2469-2496, 2018.

[23] N.Z. Foutz and W. Jank, "Prerelease Demand Forecasting for Motion Picturea Using Functional Shape Analysis of Virtual Stock Markets," Marketing Science, vol.29, pp.568-579, 2010.

[24] S. Pezzulli and B.W. Silverman, "Some Properties of Smoothed Principal Components Analysis for Functional Data", Computational Statistics, vol.8, pp. 1–16, 1993.

[25] J.O.Ramsay and B.W. Silverman, "Applied Functional Data Analysis:Methods and Case Studies," Springer-Verlag, New York, 2002.

[26] F. Yao and H.G Müller and A. J. Clifford ve diğerleri, "Shrinkage Estimation for Functional Principal Component Scores with Application to the Population Kinetics of Plasma Folate," Biometrics, vol. 59, pp. 676-685, 2003.

[27] İ.K. Keser, "Düzgünleştirilmiş fonksiyonel ana bileşenler analizi ile İMKB verilerinin incelenmesi," İstanbul Üniversitesi İktisat Fakültesi Ekonometri ve İstatistik Dergisi, vol.8, pp. 1-32, 2008.

[28] L. Luo and Y.Zhu and M. Xiong, "Smoothed functional principal component analysis for testing association of the entire allelic spectrum of genetic variation", European Journal of Human Genetics, vol. 21, pp. 217-224, 2013

[29] J.R. Berrendero and A. Justel and M. Svarc, “Principal components for multivariate functional data,” Computational Statistics & Data Analysis , vol. 55, no. 9, pp. 2619-2634, 2011.

[30] J.M. Chiou and Y.T Chen and Y.F Yang, “Multivariate Functional Principal Component Analysis: A Normalization Approach,” Statistica Sinica, vol. 24, No. 4 , pp. 1571-1596, 2014.

[31] C. Liu and S. Ray and G. Hooker, “Functional principal component analysis of spatially correlated data,” Statistics and Computing vol.27, pp.1639–1654, 2017.

[32] Y.Li & Y.Guan, “Functional Principal Component Analysis of Spatiotemporal Point Processes With Applications in Disease Surveillance, ” Journal of the American Statistical Association, vol. 109, no. 507, pp. 1205-1215, 2014.

[33] N. Locantore et al., “Robust principal component analysis for functional data,” Socicdad de Estadistica e Investigacion Opcrativa, vol. 8, no.1, pp. 1-73, 1999.

[34] D. Gervini, “Robust functional estimation using the median and spherical principal components,” Biometrika, vol. 95, no. 1, pp. 587-600, 2008.

[35] H. Shou, V. Zipunnikov, C.M. Crainiceanu and S.Greven, “Structured functional principal component analysis,” Biometrics, vol.71, no.1, pp. 247-257, 2015.

[36] F. Yao and H.G.Müller and J.L. Wang, “Functional data analysis for sparse longitudinal data,” Journal of American Statistics Association, vol. 100, pp. 577–591, 2005.

[37] S. Y. Xu ve diğerleri, ” Modeling Temporal Variation in Physical Activity Using Functional Principal Components Analysis,” Statistics in Biosciences, vol.11 pp:403-421, 2019.

[38] M. Benko and W. Hardle and A. Kneip, “Common Functional Principal Components,” The Annals of Statistics, vol.37, no.1, pp. 1-34, 2009.

[39] G. Boente and R. Fraiman, “ Kernel-based Functional Principal Components,” Statistics & Probability Letters, vol.48, no. 4, pp:335-345, 2000.

[40] N. Coffey, A.J. Harrison, O.A. Donoghue and K.Hayes, “Common functional principal components analysis: a new approach to analyzing human movement data,” Human Movement Science, vol.30, no.1, pp:44-66, 2011.

[41] M. Fengler and W. Hardle and P. Villa, “ The Dynamics of Implied Volatilities: A common principle components approach,” Review of Derivative Research vol.6 pp: 179-202, 2003.

[42] M. Benko and W. Hardle, “Common Functional Implied Volatility Analysis,” SFB 649 Discussion Papers, Berlin, 2005.

[43] C.Z. Di and C.M. Crainiceanu and B.S Caffo and N.M. Punjabi, “Multilevel Functional Principal Component Analysis,” The Annals of Applied Statistics, vol.3, no.1, pp.458-488, 2009.

[44] C.M. Crainiceanu et al. “ Nonparametric signal extraction and measurementerror in the analysis of electroencephalographic activity dur-

ing sleep," Journal of American Statistical Association, vol. 104, pp:541–555, 2009.

[45] V. Zipunnikov, B. Caffo ve diğerleri, "Multilevel Functional Principal Component Analysis for High-Dimensional Data," Journal of Computational and Graphical Statistics, vol. 20, no. 4, pp. 862-873, 2011.

[46] R. Argiento and P.G. Bissiri and A. Pievatolo and C.Scrosati, "Multilevel Functional Principal Component Analysis of Façade Sound Insulation Data," Special Issue Article, vol. 31, no.7 pp:1239-1253, 2015.

[47] C. Di and C. M. Crainiceanu and W. S. Jank, "Multilevel sparse functional principal component analysis," Stat, vol.3, pp. 126–143, 2014.

[48] F. Bafra, "Çok Değişkenli Veri Matrislerine Uygulanan Bazı Analiz Yöntemleri ve Bunlarla İlgili Bir Paket Program," 1987.

[49] Ö. Alkan, "Temel Bileşenler Analizi ve Bir Uygulama Örneği," Yüksek Lisans Tezi, Sosyal Bilimler Enstitüsü, Erzurum, 2008.

[50] H. Tatlıdil, "Uygulamalı Çok Değişkenli İstatistiksel Analiz," Akademi Matbaası, Ankara, 2002.

[51] R.A Johnson & D.W Wichern, " Applied Multivariate Statistical Analysis," Prentice – Hall: New Jersey, 1998.

[52] A.Ö. Yaycılı, "Temel Bileşenler Analizi İçin Robust Algoritmaları," Gazi Üniversitesi Fen Bilimleri Enstitüsü, 2006.

[53] R. Alpar, "Uygulamalı Çok Değişkenli İstatiksel Yöntemler," Detay Yayıncılık , Ankara, 2013.

[54] A.S. Albayrak, "Uygulamalı Çok Değişkenli İstatistik Teknikleri," 2006.

[55] R.J.Richard, "A Primer of Multivariate Statistics," New York, Academic Pres, 1975.

[56] U. Grenander, "Stochastic Processes and Statistical Inference," Arkiv För Matematik, vol.1, no.17, 1950.

[57] C.R. Rao, "Some Statistical Methods for Comparison of Growth Curves," Biometrics, vol.14, no.1, 1958.

[58] J. L. Wang and J.M. Chiou and H.G. Müller, "Review of functional data analysis," Annual Review of Statistics and Its Application, pp.1-41, 2015.

[59] J. Ulbricht, "Representing Functional Data as Smooth Functions," A Master Thesis, Humboldt University Institute of Statistics and Econometrics, Berlin, 2004.

[60] A. H.A. Al.Tememy, "design Heart Disease Diagnosis Using Discrete Wavelet Transform and Neural Network," 2005.

[61] J.O.Ramsay & B.W Silverman, "Functional Data Analysis," Springer Verlag: New York, 1997.

[62] S. Huang, "Functional Data Smoothing Methods and Their Applications," 2017.

[63] S.E.Leurgans, R.A.Moyeed and B.W. Silverman, "Canonical Correlation Analysis when the Data are Curves," Journal of the Royal Statistical Society, vol.55, no.3, 1993.

[64] R.L. Eubank, "Spline Smoothing and Nonparametric Regression," Marcel Dekker:USA, 1988.

[65] P.J. Green& B.W. Silverman, "Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach," Chapman & Hall:London, 1994.

[66] P.Craven and G. Wahba, "Smoothing Noisy Data with Spline Functions," Numerische Mathematik, vol.31, pp:377-403, 1979.

[67] C. Gu, "Smoothing Spline ANOVA Models," New York:Springer, 2nd ed., 2013.

[68] I.T. Jolliffe, "Principal Component Analysis," Second Edition, Springer- Verlag, New York, 2002.

[69] K. Karhunen, "Zur spektraltheorie stochastischer prozesse," Annales Academiae Scientiarum Fennicae, vol.37, pp: 1–37, 1946.

[70] M. Lo`eve, "Fonctions al´eatoires a decomposition orthogonale exponentielle," La Revue Scientifique, vol.84, pp: 159–162, 1946.

[71] J. Jacques, C.Preda, "Model-based Clustering for Multivariate Functional Data," Computational Statistics and Data Analysis, vol.71, pp:92-106, 2014.

[72] C. Happ S. Greven, "Multivariate Functional Principal Component Analysis For Data Observed On Different (Dimensional) Domains," Journal of the American Statistical Association, vol.113, no.522, pp.649-659, 2018.

[73] M. Benko, "Functional Principal Components analysis, Implementation and Applications", 2004.

[74] T.C. Sağlık Bakanlığı Halk Sağlığı Genel Müdürlüğü , "COVID-19 (SARS-CoV-2 Enfeksiyonu) (Bilim Kurulu Çalışması) Genel Bilgiler, Epidemioloji Ve Tanı", Ankara, 2020, (8Haziran 2020), https://covid19bilgi.saglik.gov.tr/depo/rehberler/covid-19-rehberi/COVID-19_REHBERI_GENEL_BILGILER_EPIDEMIYOLOJI_VE_TANI.pdf).

[75] M. Roser, H. Ritchie," the complete Our World in Data COVID-19 dataset," Statistics and Research, 6 Haziran 2020, https://ourworldindata.org/coronavirus-data

TEZDEN ÜRETİLMİŞ YAYINLAR

İletişim Bilgisi: sumeyye.karatas.inal@gmail.com

Konferans Bildirileri

1. S. Inal ve G.G. Simsek, "Investigation of the Electricity Consumption of Provinces of Turkey using Functional Principal Components Analysis," in ISBIS Young Business and Industrial Statisticians Workshop on Recent Advances in Data Science and Business Analytics, Istanbul, 2019.