

**T.C.
MİMAR SİNAN GÜZEL SANATLAR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**UZAKTAN ALGILAMA VERİLERİNE
FONKSİYONEL VERİ ANALİZİ YAKLAŞIMININ UYGULANMASI**

DOKTORA TEZİ

Nihan ACAR DENİZLİ

İstatistik Anabilim Dalı

İstatistik Programı

Tez Danışmanı: Prof. Dr. Gülay BAŞARIR

Tez Eş Danışmanı: Doç. Dr. Pedro DELICADO

KASIM 2016

**T.C.
MİMAR SİNAN GÜZEL SANATLAR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**UZAKTAN ALGILAMA VERİLERİNE
FONKSİYONEL VERİ ANALİZİ YAKLAŞIMININ UYGULANMASI**

DOKTORA TEZİ

**Nihan ACAR DENİZLİ
(20117229)**

İstatistik Anabilim Dalı

İstatistik Programı

Tez Danışmanı: Prof. Dr. Gülay BAŞARIR

Tez Eş Danışmanı: Doç. Dr. Pedro DELICADO

KASIM 2016

Nihan ACAR DENİZLİ tarafından hazırlanan “UZAKTAN ALGILAMA VERİLERİNE FONKSİYONEL VERİ ANALİZİ YAKLAŞIMIN UYGULANMASI” adlı bu tezin Doktora tezi olarak uygun olduğunu onaylarım.

Prof. Dr. Gülay BAŞARIR
Tez Yöneticisi

Bu çalışma, jürimiz tarafından İstatistik Anabilim Dalında Doktora tezi olarak kabul edilmiştir.

Başkan : Prof. Dr. Gülay BAŞARIR
Mimar Sinan Güzel Sanatlar Üniversitesi

Eş Danışman : Doç. Dr. Pedro DELICADO
Universitat Politecnica de Catalunya

Jüri Üyeleri : Prof. Dr. Aydın ERAR
Mimar Sinan Güzel Sanatlar Üniversitesi

Prof. Dr. Müjgan TEZ
Marmara Üniversitesi

Prof. Dr. Ahmet Mete ÇİLİNGİRTÜRK
Marmara Üniversitesi

Doç. Dr. Eylem DENİZ HOWE
Mimar Sinan Güzel Sanatlar Üniversitesi

Bu tez, Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygundur.

ÖNSÖZ

Bu konuyu çalışmam konusunda beni teşvik eden, bu süreç boyunca bana elinden gelen desteği veren ve fikirleriyle gelişmeye katkıda bulunan tez danışmanım Prof. Dr. Gülay Başarır'a, bitmek tükenmez sorularına sabırla cevap veren, bu alanda çalışma yapan pek çok kişi ile tanışmama vesile olan kendisinden çok şey öğrendiğim bu projenin fikir babası eş danışmanım Pedro Delicado'ya, bu projede verilerini ve bilgisini cömertlikle benimle paylaşan ve iş arkadaşlığını güzel bir dostluğa dönüştüren sevgili Isabel Caballero'ya bana olan inançları ve destekleri için teşekkür ederim.

Kafama takılan soruları samimiyetle cevaplayan ve değerli yorumlarıyla bu çalışmaya şekil veren sevgili Hocalarım Prof. Dr. Aydın Erar'a ve Prof. Dr. Müjgan Tez'e, konunun teorik altyapısı için bana kaynaklar öneren ve ek kısmını düzenlememe yardımcı olan Yrd. Doç. Dr. Gülay İlon Telsiz Kayaoğlu'na, İngilizce tercüme konusunda benden yardımını esirgemeyen dostum Elif Ertürk'e, Ebru Payaslığlu'na ve çalışmam süresince yaptığım yolculukları anlayışla karşılayan, akademik anlamda bana duyduğu güveni ve manevi desteğini hep hissettiren değerli Hocam Prof. Dr. Nalan Cinemre'ye bu çalışmaya katkılarından dolayı teşekkürü bir borç bilirim.

Çalışmamı İspanya'da sürdürdüğüm dönemde bana destek olan tüm UPC çalışanları, Hocaları ve doktora öğrencisi arkadaşlarıma, en yoğun dönemimde bana içtenlikle yardım eli uzatan başta Liliana Tolomei ve Daris Hernandez olmak üzere tüm Arjantinli, Meksikalı, Katalan ve diğer milletlerden dostlarıma, bu süreçte bana ve eşime manevi anlamda desteği büyük olan sevgili Manuel Ramos-Casals ve Antoni Sisó-Almirall'a, bana gösterdikleri anlayış ve destek için tüm Bölüm Hocalarıma ve arkadaşlarıma ve uzakta olsak da bir telefonla zor zamanlarımda yanımda olduğunu hissettiğim tüm dostlarıma;

Bu zorlu süreçte beni görmemeye sabırla katlanan ve tezimi bitirmemi dört gözle bekleyen anneme, benimle daha fazla zaman geçireceği günleri bir kenarda sabırsızlıkla bekleyen sevgili yeğenim Ela'ya, akademik hayat konusunda beni teşvik eden ve beni hala bir yerlerden gülümseyerek izlediğine inandığım babama, her zaman desteğini yanımda hissettiğim ağabeyime, minik yeğenim Alp'e ve tüm aileme bana duydukları güven ve verdikleri manevi destek için teşekkür ederim.

En önemlisi de bu proje sayesinde yollarımızın kesiştiği, gece gündüz demeden çalışmama sabırla katlanan, bu süreçte yanımda olmak için çok uğraşan ve her anlamda bana yardımcı olan, yol arkadaşım, hayattaki en büyük desteğim, eşim Belçin Denizli'ye çok ama çok teşekkür ederim.

Kasım 2016

Nihan ACAR DENİZLİ

İÇİNDEKİLER

	<u>Sayfa</u>
ÖNSÖZ	iii
İÇİNDEKİLER	iv
KISALTMALAR	viii
ÇİZELGE LİSTESİ	ix
ŞEKİL LİSTESİ	x
SEMBOL LİSTESİ	xi
ÖZET	xiii
SUMMARY	xv
1. GİRİŞ	1
2. FONKSİYONEL VERİ ANALİZİNİN GENEL ÇERÇEVESİ	6
2.1 Fonksiyonel Verinin Yapısı.....	6
2.2 Fonksiyonel Keşifsel Veri Analizi	7
2.2.1 Ortalama, Varyans ve Kovaryans Fonksiyonları	8
2.2.2 Derinlik Ölçüleri.....	10
2.3 Fonksiyonel Verinin Temsili.....	13
2.3.1 Taban Açılımı	13
2.3.1.1 Fourier Tabanı.....	14
2.3.1.2 B-splayn tabanı	16
2.3.1.3 Yanlılık - Varyans Dengesi	17
2.3.2 Kernel (Çekirdek) Düzleştirme	18
2.3.3 Fonksiyonel Temel Bileşenler Tabanı	19
2.3.4 Fonksiyonel Kısmi En Küçük Kareler (FKEKK) Tabanı.....	20
2.4 Fonksiyonel Veriyi Düzleştirme	20
2.4.1 En Küçük Kareler Yöntemi	21
2.4.1.1 Sıradan (Ağırlıksız) En Küçük Kareler Yöntemi	21
2.4.1.2 Ağırlıklı En Küçük Kareler (AEKK) Yöntemi.....	22
2.4.2 Yerel En Küçük Kareler (YEKK) Yöntemi.....	23
2.4.2.1 Yerel Taban Fonksiyonları ile Düzleştirme	24
2.4.2.2 Yerel Çoklu Terimli Düzleştirme.....	24
2.4.3 Pürüzlü Ceza Yaklaşımı.....	25
2.4.3.1 Düzleştirme Parametresinin Seçimi.....	27
3. FONKSİYONEL TEMEL BİLEŞENLER ANALİZİ	28
3.1 Temel Bileşenler Analizi Yaklaşımı	29
3.2 Fonksiyonel Veri İçin Temel Bileşenler Analizi.....	32

4. FONKSİYONEL DOĞRUSAL MODELLER.....	39
4.1 Skaler Yanıt Değişkeni için FDRM.....	40
4.1.1 Taban Temsili ile FDRM	41
4.1.2 Pürüzlü Ceza Yaklaşımı.....	43
4.1.3 Fonksiyonel Temel Bileşenler Regresyonu (FTBR).....	46
4.1.4 Fonksiyonel Kısmi En Küçük Kareler Regresyonu (FKEKKR).....	51
5. UYGULAMA	58
5.1 Gereç ve Yöntemler.....	59
5.1.1 Uzaktan Algılanan Veri	59
5.1.2 Denizden Toplanan Veri (In-situ Verisi).....	61
5.1.3 Geçerlilik Aşaması	61
5.1.4 İstatistiksel Yöntemler.....	62
5.1.4.1 Genelleştirilmiş Toplamsal Modeller (GTM).....	64
5.1.4.2 En Küçük Mutlak Daraltma ve Seçim Operatörü (Least Absolute Shrinkage and Selection Operator, LASSO)	65
5.1.5 Sonuçlar.....	65
5.1.5.1 Üstel Regresyon Modelleri	65
5.1.6 GTM and LASSO Modelleri	69
5.1.7 Fonksiyonel Doğrusal Regresyon Modelleri.....	71
5.1.8 Uygulama Sonuçları	77
6. SİMÜLASYON ÇALIŞMASI.....	79
6.1 Fonksiyonel Açıklayıcı Değişkenin Seçimi	79
6.2 Yanıt Değişkeninin Üretilmesi	80
6.3 Tahmin Modellerinin Karşılaştırılması.....	82
6.4 Simülasyon Sonuçları	84
7. SONUÇ VE ÖNERİLER	87
KAYNAKLAR.....	90
EKLER	99
Fonksiyonel Uzaylar Üzerinde Tanımlar.....	101
1.1 Metrik Uzaylar ve Özellikleri.....	101
1.2 İç Çarpım Uzayları ve Özellikleri	102
1.3 Normlu Uzaylar ve Özellikleri	102
1.4 Hilbert Uzayları	103
Operatörler ve Fonksiyonel Veri İçin Bazı Yararlı Teoremler.....	104
2.1 Doğrusal Operatör, İntegral Operatörü ve Hilbert-Schmidt Operatörü.....	104
2.2 Kovaryans Operatörü ve Özellikleri	104
2.3 Bir Doğrusal Operatörün Tekil Değer Ayrışımı (Singular Value Decomposition)	105
2.4 Bir Hilbert-Schmidt Operatörünün Spektral Ayrışımı.....	105
2.5 Parseval Eşitliği	106
2.6 Riesz Temsil Teoremi	106
2.7 Mercer Teoremi	106
2.8 Karhunen-Loève Açılımı.....	106
R Programında Yazılan Simülasyon Fonksiyonu.....	107

ÖZGEÇMİŞ **112**





KISALTMALAR

FVA	: Fonksiyonel Veri Analizi
MRI	: Manyetik Rezonans Görüntüleme (Magnetic Resonance Imaging)
HSD	: Yarım-uzay Derinliği (Half-Space Depth)
FMD	: Fraiman-Muniz Derinliği
FD	: Fonksiyonel Derinlik
SD	: Basit Derinlik (Simplicial Depth)
RPD	: Rasgele İzdüşüm Derinliği (Random Projection Depth)
RTD	: Rasgele Tukey Derinliği (Random Tukey Depth)
MSE	: Hata Kareler Ortalaması (Mean Square Error)
KNN	: K en yakın komşuluk (K Nearest Neighbourhood) kestiricisi
FTB	: Fonksiyonel Temel Bileşenler
EKK	: En Küçük Kareler
KEKK	: Kısmi En Küçük Kareler
FKEKK	: Fonksiyonel Kısmi En Küçük Kareler
SSE	: Hata Kareler Toplamı (Sum of Square Errors)
PENSSE	: Cezalı Hata Kareler Toplamı (Penalized Sum of Square Errors)
SEKK	: Sıradan En Küçük Kareler
AEKK	: Ağırlıklı En Küçük Kareler
LLS	: Yerel En Küçük Kareler (Localized Least Squares)
SSE_{LLS}	: Yerel En Küçük Kareler Yönelimi için Hata Kareler Toplamı
LPS	: Yerel Çoklu Terimli Düzleştirme (Localized Polynomial Smoothing)
SSE_{LPS}	: Yerel Çoklu Terimli En Küçük Kareler için Hata Kareler Toplamı
PCY	: Pürüzlü Ceza Yaklaşımı
CEKK	: Cezalı En Küçük Kareler
PEN	: Ceza ölçüsü (Penalization)
CV	: Çapraz Geçerlilik (Cross Validation)
GCV	: Genelleştirilmiş (Generalized Cross Validation)
Df	: Serbestlik derecesi (Degrees of freedom)
FANOVA	: Fonksiyonel Varyans Analizi (Functional Analysis of Variance)
FDM	: Fonksiyonel Doğrusal Modeller
FDRM	: Fonksiyonel Doğrusal Regresyon Modelleri
TBR	: Temel Bileşenler Regresyonu
KEKKR	: Kısmi En Küçük Kareler Regresyonu
FTBR	: Fonksiyonel Temel Bileşenler Regresyonu
FKEKKR	: Fonksiyonel Kısmi En Küçük Kareler Regresyonu
FKEKKB	: Fonksiyonel Kısmi En Küçük Kareler Bileşeni
MSC	: Model Seçim Kriterleri (Model Selection Criteria)
SIC	: Schwarz Bilgi Kriteri (Schwarz Information Criterion)
SICc	: Düzeltilmiş Schwarz Bilgi Kriteri (Corrected Schwarz Information Criterion)
AIC	: Akaike Bilgi Kriteri (Akaike Information Criterion)
AICc	: Düzeltilmiş Akaike Bilgi Kriteri (Corrected Akaike Information Criterion)
TSS	: Toplam Katı Maddeler (Total Suspended Solids)
Rrs	: Uzaktan Algılamanın Yansımaları (Reflectance of Remote Sensing)

ÇİZELGE LİSTESİ

1.1	Çok Değişkenli ve Fonksiyonel Veri İçin Doğrusal ve Parametrik Olmayan Modeller	4
3.1	TBA ve FTBA yöntemlerinin karşılaştırılması	28
4.1	Fonksiyonel Doğrusal Regresyon Modelleri	39
5.1	L2 maddeleri	60
5.2	Üstel Regresyon Modelleri için AMEP Değerleri	68
5.3	TBR ve KEKKR Modellerine Ait AMEP Değerleri	69
5.4	GTM için AMEP Değerleri	70
5.5	FDRM için AMEP değerleri.....	73
5.6	FTBR CV modelinin Parametre Kestirimleri	75
5.7	FKEKKR CV modeline ait parametre kestirimleri.....	76
6.1	Örneklem Büyüklüğü 25 için Karşılaştırma Tablosu (medMEP değerlerinin ortalaması)	85
6.2	Örneklem Büyüklüğü 50 için Karşılaştırma Tablosu (medMEP değerlerinin ortalaması)	86
6.3	Örneklem Büyüklüğü 100 için Karşılaştırma Tablosu (medMEP değerlerinin ortalaması)	86

ŞEKİL LİSTESİ

5.1	Çalışma alanı ve ilgilenilen alanı gösteren harita.	62
5.2	Değişkenler arasındaki korelasyon çizimleri	66
5.3	In-situ TSS değerleri ile 665 nm ve 681 nm bandlarındaki Rrs değerleri arasındaki saçılım grafiği.....	67
5.4	Üstel regresyon modellerine ait artık grafikleri	69
5.5	Bir dizi lambda (λ) değeri için hata kareler ortalaması.....	71
5.6	LASSO modeli.....	71
5.7	Düzleştirilmiş eğriler	72
5.8	log TSS and Rrs arasındaki FTBR CV modeline ait özet grafikler. ...	74
5.9	log TSS and Rrs arasındaki FKEKKR CV modeline ait özet grafikler.	75
5.10	FTBR CV ve FKEKKR CV Modellerince Seçilen Fonksiyonel Bileşenler.....	77
5.11	FTBR CV ve FKEKKR CV modellerinden kestirilen parametre fonksiyonları.	78
6.1	Farklı dalga boyları için uydudan elde edilen görüntüler	80
6.2	Simülasyonda Kullanılan Fonksiyonel Gözlemler	80
6.3	Simülasyon Tasarımı.....	83

SEMBOL LİSTESİ

X	: Vektör uzayında tanımlı bağımsız değişken
χ	: Fonksiyonel uzayda tanımlı bağımsız değişken
\mathbb{R}	: Vektör uzayı
F	: Fonksiyon uzayı
$L^2(T)$: Karesi integrallenebilir fonksiyonların uzayı
t_j	: Fonksiyonel değişkenin argüman değerleri
n	: Fonksiyonel değişkenin argüman sayısı
T	: Fonksiyonel verinin tanımlandığı aralık
N	: Fonksiyonel gözlem sayısı
\mathbb{E}	: Beklenen değer
$\mu_\chi(t)$: Kitle ortalama fonksiyonu
$\sigma_\chi^2(t)$: Kitle varyans fonksiyonu
$c_\chi(t, s)$: Kitle kovaryans fonksiyonu
$r_\chi(t, s)$: Kitle korelasyon fonksiyonu
Γ_χ	: Kitle kovaryans operatörü
$\hat{\mu}_\chi(t)$: Örneklem ortalama fonksiyonu
$\hat{\sigma}_\chi^2(t)$: Örneklem varyans fonksiyonu
$\hat{c}_\chi(t, s)$: Örneklem kovaryans fonksiyonu
$\hat{r}_\chi(t, s)$: Örneklem korelasyon fonksiyonu
$\hat{\Gamma}_\chi$: Örneklem kovaryans operatörü
$\hat{c}_{\chi, \gamma}(t, s)$: Çapraz kovaryans fonksiyonu
$\hat{r}_{\chi, \gamma}(t, s)$: Çapraz korelasyon fonksiyonu
F_N	: Kümülatif dağılım fonksiyonu
y_i	: y rassal değişkeninin kesikli gözlem değerleri
ϕ_k	: Taban fonksiyonu
c_k	: Katsayı vektörü
K	: Taban fonksiyonu sayısı
D	: Türev operatörü
τ_l	: Splayn kesim noktaları
m	: Polinomun sırası
$K(\cdot)$: Çekirdek fonksiyonu
h	: Çekirdek (kernel) düzleştirmede bant genişliği
$w_i(t)$: Ağırlık fonksiyonu
$S_j(t)$: Doğrusal düzleştirici
f_{ij}	: Temel Bileşen Skorları
ξ_j	: Öz Fonksiyonlar
φ_l	: FKEKK bileşenleri
v_{il}	: FKEKK skor değerleri
S	: Düzleştirme matrisi
R	: Ceza matrisi

λ	: Splayn düzleştirmede düzleştirme parametresi
ω_j	: t_j noktasındaki ağırlık değeri
$\Xi(v)$: Cezalandırma fonksiyonu
x_{ij}	: Merkezleştirilmiş gözlem vektörü
\mathbf{u}_j	: j 'inci bileşene ait ağırlık vektörü
f_{ik}	: k 'ncü bileşenin i 'inci gözlemine ait ağırlık vektörü
$\chi_i^c(t)$: Merkezleştirilmiş gözlem eğrileri
$\hat{\xi}_j(t)$: j 'inci temel bileşene ait ampirik ağırlık fonksiyonu
Y_i	: Skaler yanıt değişkeni
$\gamma_i(t)$: Fonksiyonel yanıt değişkeni
$\varepsilon_i(t)$: Fonksiyonel hata terimi
$\boldsymbol{\varepsilon}_i$: Vektörel hata terimi
\mathbf{Z}	: $N \times p$ boyutlu dizayn matrisi
$\boldsymbol{\beta}$: $p \times 1$ boyunda parametre vektörü
$\beta(s)$: Parametre fonksiyonu
$\langle \cdot, \cdot \rangle$: İç çarpım
\hat{Y}_i	: Yanıt değişkeninin kestirimi
$\hat{\beta}(s)$: Parametre fonksiyonu kestirimi
$\hat{\beta}_{FTB}$: FTBR modelinden elde edilen parametre fonksiyonu kestirimi
$\hat{\beta}_{FKEKK}$: FKEKKR modelinden elde edilen parametre fonksiyonu kestirimi
$\beta^*(s)$: Parametre fonksiyonu kestiriminin taban açılımı
θ_l	: Parametre fonksiyonu kestirimi için kullanılan taban fonksiyonu
b_l	: Parametre fonksiyonu için kullanılan taban açılımı katsayıları
\hat{b}_l	: $\hat{\beta}$ fonksiyonuna ait taban katsayı vektörü
\mathbf{C}	: $\chi_i(s)$ fonksiyonuna ait taban katsayı vektörü
$\mathbf{J}_{\phi\theta}$: ϕ ve θ taban fonksiyonunun iç çarpımından oluşan $K \times L$ boyutlu matris
$\boldsymbol{\zeta}$: $L + 1$ uzunluğunda parametre vektörü
$\hat{\boldsymbol{\zeta}}$: $L + 1$ uzunluğunda parametre vektörünün kestirimi
\mathbf{H}	: Şapka matrisi
h_{ii}	: Şapka matrisinin köşegen elemanları
p_N	: Bilgi kriterlerinin genelleştirilmesinde kullanılan bir parametre
\tilde{Y}_{exp}	: Üstel regresyon modelinden türetilen yanıt vektörü
\tilde{Y}_{taban}	: B-splayn açılımı ile FDRM kullanılarak türetilen yanıt vektörü
\tilde{Y}_{FTB}	: FTBR yöntemi ile tüm bileşenler kullanılarak türetilen yanıt vektörü
\tilde{Y}_{FTBcv}	: FTBR CV yöntemi ile seçilen bileşenler kullanılarak türetilen yanıt vektörü

UZAKTAN ALGILAMA VERİLERİNE FONKSİYONEL VERİ ANALİZİ YAKLAŞIMININ UYGULANMASI

ÖZET

Son zamanlarda modern bilimin gelişmesi ve özellikle de verilerin sürekli zamanda ölçülme eğiliminin artmasıyla Fonksiyonel Veri Analizi (FVA) istatistikte önem kazanmıştır. Klasik istatistikte vektör uzayları üzerinde tanımlanan diskriminant analizi, temel bileşenler analizi, regresyon analizi gibi sınıflama, boyut indirgeme ve modelleme amaçlarıyla kullanılan pek çok yöntem fonksiyonel duruma uyarlanmıştır. FVA, bir süreklilik üzerinde tanımlanan ya da sürekli yapıya sahip olan değişkenlerin analizi ile ilgilenir. Bu nedenle, FVA kemometri, tıp ve çevrebilim gibi alanlarda bir spektrum üzerinde tanımlanan ya da görüntü olarak kaydedilen verilerin analizinde önemli yere sahiptir. Özellikle çevrebilimde, uydulardan elde edilen görüntü verilerinin analiz edilmesi karada toprak kullanımı, sudaki kirlilik ya da mineral oranı gibi pek çok konuda daha ucuz ve hızlı bir şekilde bilgi sahibi olmamızı sağlar.

Bu çalışmanın amacı, FVA yaklaşımını önermek, Fonksiyonel Doğrusal Regresyon Modellerini (FDRM) uzaktan algılama verileri üzerinde İspanya'nın Cadiz bölgesinde bulunan Guadalquivir nehir ağzında biriken katı madde oranının tahmin edilmesi için uygulamak ve çeşitli FDRM ile klasik istatistiksel modellerden elde edilen sonuçları uygulamalı olarak karşılaştırmaktır. Ayrıca, uygulama sonuçlarının desteklenmesi amacıyla bir simülasyon çalışması tasarlamak ve bu şekilde en iyi tahmin performansını gösteren modelleri belirlemek amaçlanmaktadır.

Bu amaç doğrultusunda çalışmanın izleyen bölümünde, fonksiyonel veri analizi konusunda literatür taraması yapılarak bu alandaki çalışmalardan bahsedilmiştir.

İkinci bölümde, fonksiyonel veri analizi genel çerçevesi ile açıklanmış, fonksiyonel değişkenler için betimsel istatistiklerin hesabı ve keşifsel veri analizi konularına değinilerek kesikli bir değişkeni fonksiyonel değişkene dönüştürmek için kullanılacak yöntemler tanıtılmıştır.

Üçüncü bölümde, önemli bir boyut indirgeme yöntemi olan temel bileşenler analizinin çok değişkenli analizden fonksiyonel veri analizine uyarlanması ve teorik alt yapısı ele alınmıştır.

Dördüncü bölümde, yanıt değişkenin skaler olduğu durumda kullanılacak FDRM kapsamlı olarak açıklanmıştır.

Beşinci bölümde, uygulamada kullanılan uydu verilerinin ve skaler yanıt vektörünü oluşturan katı madde oranı değerlerinin nasıl elde edildiği anlatılmış, katı madde oranının uydu verilerinden tahmin edilmesi için bazı klasik istatistiksel yöntemler ve fonksiyonel yöntemler kullanılarak elde edilen sonuçlar karşılaştırılmıştır.

Altıncı bölümde, uygulamadan elde edilen sonuçların desteklenmesi ve kullanılan fonksiyonel modellerin tahmin performansının karşılaştırılabilmesi için tasarlanan simülasyon çalışması anlatılmış ve sonuçları verilmiştir.

Çalışmanın son bölümünde ise elde edilen sonuçlar genel olarak yorumlanmış, literatürde yapılan çalışmalarla karşılaştırılmış ve ileriye yönelik önerilerde bulunulmuştur.

Anahtar kelimeler: Fonksiyonel veri analizi, fonksiyonel doğrusal regresyon modelleri, fonksiyonel temel bileşenler regresyon, fonksiyonel kısmi en küçük kareler regresyon, uzaktan algılama verileri.



A FUNCTIONAL DATA ANALYSIS APPROACH TO REMOTE SENSING DATA

SUMMARY

Functional Data Analysis (FDA) is a statistical field which has gained importance due to the progress in modern science, mainly in the ability to measure in continuous time results of an experiment and the possibility to record them. Many methods such as discriminant analysis, principal components analysis and regression analysis that are used on vector spaces for classification, dimension reduction and modelling have been adapted to the functional case. FDA is concerned on variables that are defined on a continuum or that have continuous structure. Therefore, FDA has an important role in the analysis of spectral data sets and images that are mostly recorded in the fields of chemometry, medicine and ecology. Especially in ecology, the analysis of images that are recorded in satellite sensors inform us in a fast and economical way about the use of land, the crop production in land, the water pollution and the amount of minerals include the water.

The aim of this study is to propose the use of FDA approach and to predict the amount of Total Suspended Solids (TSS) in the estuary of Guadalquivir river in Cadiz on remote sensing data by using different Functional Linear Regression Models (FLRM). Besides, it is purposed to compare the results obtained from various FLRMs and classical statistical methods practically, to design a simulation study in order to support findings and to determine the best prediction model.

In accordance with this purpose, the following chapter reviews the studies on this area and the literature on FDA.

In the second chapter, the general framework of FDA is explained, descriptive statistics and exploratory data analysis for functional variables are handled and the methods which can be used to convert a discrete variable to a functional variable are presented.

In the third chapter, the theoretical background of the extension of principal component analysis from multivariate case to the functional case is explained.

In the fourth chapter, FLRM for the case of scalar response is explained comprehensively.

In the fifth chapter, the acquisition of the satellite images and in-situ data are explained in detail. In order to estimate the amount of TSS on satellite data, some classical statistical models and functional models are used and their results are compared.

In the sixth chapter, a simulation study is designed to support findings and to measure the performance of the FLRMs and later its results are presented.

As to final chapter, the results obtained are discussed and compared to the results of the studies in the literature and prospective suggestions are made.

Key Words: Functional data analysis, functional linear regression models, functional principal components regression, functional partial least squares regression, remote sensing data.



1 GİRİŞ

Modern bilimin gelişmesiyle, son zamanlarda tıp, ekonomi ve doğa bilimleri gibi pek çok alanda büyük veri tabanları ile çalışmak daha yaygınlık kazanmıştır. Geniş veri setlerinin analizi, var olan istatistiksel yöntemlerin geliştirilmesini gerektirmektedir. Fonksiyonel veri analizi (FVA), çok değişkenli istatistiksel analiz yöntemlerinin vektörler yerine eğrilerin analizi ile ilgilenen bir uzantısıdır. FVA zaman, uzay ya da belli bir süreklilik üzerinde tanımlanan verilerin analiz edilmesinde önemli bir yaklaşımdır. Bundan dolayı, biomedikal bilim ve biyoinformatik alanlarında önem kazanmıştır (Wu ve Müller, 2010; Escabias ve diğ., 2012). Ullah ve Finch (2013)'e göre, 2005-2010 yılları arasında FVA yaklaşımı kullanan yayınlanmış makalelerin yüzde 21'i biyomedikal bilim ile ilgilidir. Literatürde biyomekanik, manyetik rezonans görüntüleme (Magnetic Resonance Imaging, MRI) ve gen profilleri çalışmalarını içeren pek çok FVA uygulaması mevcuttur (Ullah ve Finch, 2013; Ramsay, 2000; Viviani ve diğ., 2005; Müller ve diğ., 2008).

Çok değişkenli veri analizinden farklı olarak FVA, veriyi üreten sürecin düzgün olduğu varsayımı altında bir dizi gözlem noktalarını tek bir gözlem olarak ele alır. Bir fonksiyonun düzgün olması, belli bir dereceye kadar türevlenebilir olduğu anlamına gelir. FVA, fonksiyonlarda ya da onların türevlerinde saklı olan ve geleneksel istatistiksel yöntemler aracılığıyla elde edilmesi mümkün olmayan ek bilgiyi ortaya çıkarmaya yarar (Levitin ve diğ., 2007). Fonksiyonel veri analizinin temel amaçları şunlardır: Veriyi ileri analizlere uygun şekilde tanıtmak, veriye ait önemli karakteristikler ortaya çıkacak şekilde verinin görselleştirilmesi, verideki örüntü ve değişimin çalışılması, bir yanıt ya da bağımlı değişkendeki değişimin bağımsız değişkenlere ait bilgiler kullanılarak açıklanması, aynı ya da farklı fonksiyonlar içeren iki ya da daha fazla sayıda veri setinin ortak tekrar kümeleri açısından karşılaştırılması. Bu bağlamda, çok değişkenli istatistiksel analizden fonksiyonel veri analizine uyarlanan bazı yöntemler fonksiyonel betimsel istatistik, fonksiyonel temel

bileşenler analizi, fonksiyonel doğrusal modeller ve fonksiyonel kanonik korelasyon analizidir (Ramsay ve Silverman, 2005).

Zamana bağlı tekrarlı ölçümlerden oluşan gözlem yapısından dolayı fonksiyonel veri analizi çok değişkenli bir yöntem olan boylamsal (longitudinal) analiz ile karıştırılabilmektedir. Her iki yöntemin de ilgilendiği veri yapısı benzer gibi görülse de, gerçekte birbirinden oldukça farklıdır.

Boylamsal veri analizi her bir birey için farklı zamanlarda toplanan az sayıda tekrarlı ölçümlerle ilgilenir. Fonksiyonel veri analizi ise değişik zaman noktalarından olabileceği gibi, farklı frekanslardan ya da herhangi bir süreklilik üzerinden toplanan çok sayıda tekrarlı ölçümlerin analizi ile ilgilenir. Bunun yanı sıra, uzlamsal veri parametrik bir modelde rassal vektör olarak ele alınırken, fonksiyonel veri az sayıda yapısal varsayım gerektirdiğinden nonparametrik ya da semiparametrik yöntemler gerektiren, kesikli zaman noktalarında gözlemlenen düzgün bir süreç olarak ele alınır (Davidian ve diğ., 2004). FVA ve boylamsal veri analizi yöntemlerinin detaylı olarak karşılaştırılması Rice (2004)'ın ve Hall ve diğ. (2006)'ın çalışmalarında bulunabilir.

Fonksiyonların önemi ve klasik ve modern matematiksel analiz yöntemleri arasındaki farklar Dieudonné (1960) tarafından şu şekilde açıklanmıştır:

"Bir f fonksiyonelinin tek başına değişen bir obje olması geniş bir fonksiyonel uzayda tek bir nokta olarak düşünülüp değişkenlik göstermesi demektir. Aslında analizin klasik ve modern içeriği arasındaki fark şudur: Klasik matematikte $f(x)$ yazıldığında, f sabit ve x de değişken olarak görüntülenir. Ancak bugünlerde hem f hem de x değişken olarak ele alınmaktadır. (hatta bazen sabit olan x , değişkenlik gösteren nesne ise f olmaktadır)."

Fonksiyonel veri analizinin önemi 60'lı yıllarda belirtilmiş olmasına rağmen, fonksiyonların rassal gözlemler olarak ele alındığı ilk makaleler 80'li yıllarda basılmıştır (Dauxois ve diğ., 1982; Ramsay, 1982; Besse ve Ramsay, 1986). Ramsay (1982) istatistiksel içerikleri fonksiyonel veriye uyarlayarak verinin fonksiyonel analitik görünümünü açıklamış, Besse ve Ramsay (1986) veri seti fonksiyonlardan oluştuğunda temel bileşenler analizi için yeni bir yöntem önermiştir. Stokastik bir süreçten gelen rassal bir fonksiyonun bağımsız gözlemleri olduğu varsayılan eğrilerden oluşan bir örneklem ortalamasının kestirimi için bir yaklaşım 1991 yılında

Rice ve Silverman (1991) tarafından tanımlanmıştır. Bunu takiben, fonksiyonel veri analizi kavramı ilk kez Ramsay ve Dalzell (1991)'in doğrusal modelleme ve temel bileşenler analizi gibi klasik istatistiksel yöntemleri sonsuz boyutlu uzaya uyguladığı çalışmasında önerilmiştir. Daha sonra Silverman (1996) fonksiyonel durumda temel bileşenler analizi için düzgünleştirme parametresine bağlı bir yaklaşım önermiştir. Manteiga ve Vieu (2007) ve Shang (2014)'de ifade edildiği üzere temel bileşenler analizi ve doğrusal regresyon modelleri klasik istatistikten fonksiyonel veriye uyarlanan ilk yöntemlerdir. Fonksiyonel temel bileşenler analizi (FTBA) fMRI verilerinin analizinden uçak yörüngelerinin hesaplanmasına kadar pek çok veri seti üzerinde geniş bir uygulama alanı bulmuştur (Viviani ve diğ., 2005; Nicol, 2013). Shang (2014)'in çalışmasında FTBA üzerine geniş bir literatür taraması bulunmaktadır.

Değişkenlerden en az birinin fonksiyonel olduğu durumda iki değişken arasındaki ilişkinin açıklanabilmesi için fonksiyonel doğrusal regresyon modelleri (FDRM) önerilmiştir. Bu modeller ilk olarak hem yanıt hem açıklayıcı değişkenin fonksiyonel yapıda olması durumu için geliştirilmiş (Ramsay ve Dalzell, 1991), daha sonraları ise açıklayıcı değişkenin fonksiyonel yanıt değişkeninin skaler olduğu durum ve yanıt değişkeninin fonksiyonel, açıklayıcı değişkenlerin skaler olduğu durumlara uyarlanmıştır (Cardot ve diğ., 1999; James, 2002). İlerleyen süreçte ikili yanıt değişkenini fonksiyonel açıklayıcı değişkenlerden tahmin etmek için kullanılan fonksiyonel lojistik regresyon modelleri geliştirilmiştir (Escabias ve diğ., 2004; Müller ve Stadtmüller, 2005). Çoklu bağlantıdan ve büyük boyutluluktan kaynaklanan problemleri engellemek için fonksiyonel lojistik regresyonun temel bileşenler analizi ile birlikte kullanımı önerilmiştir Escabias ve diğ. (2005). Lojistik regresyonun fonksiyonel veri üzerinde değişik uygulamaları Ratcliffe ve diğ. (2002); Müller ve Stadtmüller (2005); Aguilera (2008)'in çalışmalarında bulunabilir. Fonksiyonel lojistik regresyon modelleri daha sonra Genelleştirilmiş Doğrusal Regresyon Modellerine (GDRM) genişletilmiştir (James, 2002; Müller ve Stadtmüller, 2005). Cuevas ve diğ. (2004) fonksiyonel veriler için ANOVA testi önermiştir. Fonksiyonel veri analizi yaklaşımı değişik alanlarda, insan parmaklarının gücünün analizi (Ramsay ve diğ., 1995), dudak hareketlerinin analizi (Ramsay ve diğ., 1996), el yazılarının analizi (Ramsay, 2000), kişisel büyüme eğrilerinin analizi, iklim değişimi, kriminoloji,

Çizelge 1.1: Çok Değişkenli ve Fonksiyonel Veri İçin Doğrusal ve Parametrik Olmayan Modeller

Veri Türü	Doğrusal Modeller	Parametrik Olmayan Modeller
Çok Değişkenli	$X \in \mathbb{R}^p$ $Y = a_0 + \sum_{j=1}^p a_j X_j + \varepsilon$ $C = \{r \text{ doğrusal}\}$	$X \in \mathbb{R}^p$ $Y = r(X_1, \dots, X_p) + \varepsilon$ $C = \{r \text{ sürekli}\}$
Fonksiyonel	$\chi \in F = L^2(T)$ $Y = \int_T \rho(t) \chi(t) dt + \varepsilon$ $C = \{\chi \mapsto \int_T \rho(t) \chi(t) dt \in \mathbb{R}\}$	$\chi \in F = L^2(T)$ $Y = r(\chi) + \varepsilon$ $C = \{r \text{ sürekli}\}$

tüketim malları indeksi ve kalça ve diz eklemlerinin analizi (Ramsay ve Silverman, 2002, 2005), sanatçıların müzik performanslarının analizi (Almansa ve Delicado, 2009) ve uçuş trafiğinin analizi (Nicol, 2013) gibi pek çok farklı amaçla uygulanmıştır.

Fonksiyonel modeller klasik istatistikte olduğu gibi parametrik ve parametrik olmayan modeller olmak üzere ikiye ayrılır. Fakat tanımlı oldukları uzay bakımından fonksiyonel modellerin yapısı klasik istatistiksel modellerin yapısından oldukça farklıdır. Bir eğri ile kategorik yanıt değişkeni arasındaki ilişkinin parametrik olmayan çekirdek yöntemler aracılığıyla araştırılması problemi, fonksiyonel veri ile nonparametrik analizi birleştirerek Ferraty ve Vieu (2003) tarafından önerilen yeni yaklaşımın temelini oluşturmuştur. Parametrik olmayan bu yeni yaklaşım eğrilerin denetimli ve denetimsiz olarak sınıflandırılmasında kullanılmıştır. Parametrik olmayan yöntemlerin kemometri, konuşma tanımlama ve elektrik tüketimi gibi pek çok alanda farklı uygulamaları bulunmaktadır (Ferraty ve Vieu, 2006).

Regresyon modellerinin genel biçimini göz önüne alalım:

$$Y = r(X) + \varepsilon.$$

X , \mathbb{R} vektör uzayı üzerinde tanımlı rassal değişkenleri, $\chi = \chi(t)$, $t \in T = (t_1, t_n)$ ise sınırlı bir fonksiyonel uzay olan $L^2(T)$ üzerinde tanımlı fonksiyonel değişkenleri ifade etmek üzere çok değişkenli ve fonksiyonel veri için, parametrik ve parametrik olmayan modeller arasındaki ayrım Tablo (1.1)'deki gibi özetlenebilir.

Fonksiyonel veri analizinde bir başka önemli konu da aykırı değerlerin belirlenmesidir. Febrero ve diğ. (2007, 2008) tarafından derinlik ölçülerine dayanan farklı aykırı değer belirleme yöntemleri geliştirilmiştir. Bunun yanısıra, sağlam aykırı değer belirleme yöntemleri de önerilmiştir (Sawant ve diğ., 2012).

Yakın zamanda FVA yaklaşımı ile ilgili farklı literatür çalışmaları yayınlanmıştır (Cuevas, 2014; Wang ve diğ., 2016; Morris, 2014; Reiss ve diğ., 2016). Cuevas (2014) fonksiyonel veri için istatistiksel çıkarımların teorisi, regresyon, sınıflama ve boyut indirgeme konularında kısmi bir literatür incelemesi yapmıştır. Bu çalışmalardan Wang ve diğ. (2016) FVA'nin zamanla gelişimini ele alırken, Morris (2014) fonksiyonel regresyon üzerine, Reiss ve diğ. (2016) ise özel olarak skaler yanıt değişkenli fonksiyonel modeller üzerine kapsamlı literatür çalışmaları yapmışlardır. Konunun popülerliğinin artmasıyla birlikte *Statistica Sinica*, issue 14, 3 (2004), *Computational Statistics*, 22, 3 (2007), *Computational Statistics & Data Analysis*, 51, 10 (2007) ve *Journal of Multivariate Analysis*, 101, 2 (2010) gibi pek çok dergi bu alanda çalışmalar yapan Davidian ve diğ. (2004); Manteiga ve Vieu (2007); Valderrama (2007) ve Ferraty (2010)'nin önsözleri ile fonksiyonel veri analizi konusunda özel sayılar çıkarmışlardır. Son zamanlarda, fonksiyonel veri analizi kavramı ulusal literatürümüzde de önem kazanmaya başlamıştır (Keser, 2010, 2007; Gündüz, 2012; Sözen, 2014; Özçomak ve Gündüz, 2014). Keser (2010) ve Sözen (2014) Ege ve Karadeniz Bölgelerine ait yağış verilerini fonksiyonel temel bileşenler analizini kullanarak analiz etmiştir. Gündüz (2012)'ün çalışmasında fonksiyonel kanonik korelasyon analizi Borsa Değişim İndeksinin hesaplanması için kullanılmıştır. Ulusal çalışmalar daha çok Fonksiyonel Temel Bileşenler Analizi ve Fonksiyonel Kanonik Korelasyon Analizi üzerinde yoğunlaşmıştır. Henüz, fonksiyonel veri analizinin farklı dalları üzerine yoğunlaşan çalışmalar ulusal literatürde bulunmamaktadır.

2 FONKSİYONEL VERİ ANALİZİNİN GENEL ÇERÇEVESİ

Klasik istatistiksel analiz nokta olarak tanımlanan gözlemlerle ilgilenir. Ancak, fonksiyonel veri analizi gerçel fonksiyonlar şeklinde tanımlanan gözlemlerle ilgilenir. Fonksiyonel veri örnekleme, bir T kümesi üzerinde tanımlı $\chi_1, \chi_2, \dots, \chi_N$ ile gösterilen N adet fonksiyondan oluşur. Fonksiyonel veri analizinde örneklem uzayı vektör uzayı değil, sonsuz boyutlu bir fonksiyonel uzaydır. Örneklem uzayının sonsuz boyutlu olmasından dolayı FVA istatistiksel teorisinin yeni bir branşı olarak görülmektedir (Cuevas, 2014).

2.1 Fonksiyonel Verinin Yapısı

Çizelge 1.1’de de görüldüğü üzere çok değişkenli ve fonksiyonel modeller arasındaki en önemli ayırım değişkenin tanımlı olduğu uzaydır. Bir χ fonksiyonel değişkeni metrik ya da semi-metrik olan bir F fonksiyon uzayında değerler alan bir rassal değişkendir. Fonksiyonel bir veri seti χ ile aynı dağılımdan gelen $t_j, (j = 1, \dots, n)$ argüman değerlerine sahip $\chi_1, \chi_2, \dots, \chi_N$ ya da $\chi_1(t), \chi_2(t), \dots, \chi_N(t)$ olarak gösterilen N adet fonksiyonel değişkenden oluşur (Febrero-Bande ve Oviedo de la Fuente, 2012; Ferraty ve Vieu, 2006). t_j argüman değerleri tüm ölçüm noktaları için aynı olabileceği gibi kayıttan kayıda değişiklik de gösterebilmektedir (Ramsay ve Silverman, 2005).

Fonksiyonel veri setinin tanımını göz önünde bulundurularak t_1, t_2, \dots, t_n kesikli noktalarında değerler alan herhangi bir $\{\chi(t) : t \geq 0\}$ süreci, fonksiyonel veri olarak adlandırılabilir. Ancak bu şekilde tanımlanan her kesikli süreç her zaman fonksiyonel yapıda değildir. Cuevas (2014)’a göre böyle bir süreci fonksiyonel olarak ele almanın iki nedeni vardır:

"Birincisi argümanların çok sık aralıklarla gözlenme olasılığı (en azından teorik olarak) ve limitte sabit bir t değeri için $x(t)$ 'nin gözlenmesidir. İkincisi onu yaklaşık olarak temsil edecek fonksiyonel modelin seçimidir."

Fonksiyonel veri setinin bazı özellikleri şu şekilde sıralanabilir (Ramsay ve Silverman, 2002):

- Fonksiyonel veri kesikli noktalarda gözlenmiş olsa bile, içerik olarak sürekli tanımlanmıştır.
- Fonksiyonel veri setine ait bir gözlem nokta olarak değil, tüm bir fonksiyon olarak gözlenir. Fonksiyonel veri setini oluşturan gözlemlerin birbirinden bağımsız olduğu varsayımı olmasına rağmen, aynı fonksiyonel veriyi oluşturan gözlemlerin bağımsızlığına yönelik bir varsayım söz konusu değildir.
- Fonksiyonel veri zamanın fonksiyonlarından oluşur. Zaman çok değişkenli analizdeki gibi ayrı bir değişken olarak ele alınmaz.
- Fonksiyonel veri düzgün olarak gözlenmemiş olsa bile, fonksiyonel veriyi modellemek için düzgünlük gereklidir.

Klasik veri analizinde olduğu gibi, ortalamanın ve standart sapmanın kestirilmesi ve aykırı değer analizi fonksiyonel veri analizi için de önemli adımlardır. Ancak bu durumda verinin fonksiyonel yapısından ötürü ortalamanın ve standart sapmanın hesabı çok değişkenli duruma göre biraz daha zordur.

2.2 Fonksiyonel Keşifsel Veri Analizi

Fonksiyonel bir örnekleme ait ortalama, standart sapma, varyans ve derinlik gibi diğer özet istatistikler çok değişkenli duruma benzer şekilde hesaplanır. Aralarındaki en önemli fark fonksiyonel durumda bu istatistikler tekrarlar üzerinden nokta bazlı olarak hesaplanır ve sonuç artık skaler bir değer değil bir fonksiyondur.

2.2.1 Ortalama, Varyans ve Kovaryans Fonksiyonları

$L^2(T)$ uzayında tanımlı karesi integrallenebilir rassal bir $\chi = \chi(t)$, $t \in T = [t_1, t_n]$ fonksiyonunu ele alalım. Bir fonksiyonun karesinin integrallenebilir olması (2.1) ile verilen özelliğin gerçekleşmesini gerektirir.

$$\mathbb{E}\|\chi\|^2 = \mathbb{E} \int_T \chi^2(t) dt < \infty. \quad (2.1)$$

Karesi integrallenebilen fonksiyonlardan oluşan bir kitleye ait iki önemli kitle parametresi ortalama fonksiyonu ve varyans fonksiyonudur. Bu fonksiyonlar sırasıyla (2.2) ve (2.3) eşitliklerindeki gibi tanımlanır.

$$\mu_\chi(t) = \mathbb{E}[\chi(t)], \quad (2.2)$$

$$\text{Var}_\chi(t) = \sigma_\chi^2(t) = \mathbb{E}[(\chi(t) - \mu(t))^2]. \quad (2.3)$$

Fonksiyonel bir veri seti için kovaryans fonksiyonu, farklı argüman değerleri arasındaki kayıtların bağımsızlığı hakkında bilgi verir. Fonksiyonel bir kitleye ait $t, s \in T$ argüman değerleri için kovaryans fonksiyonu (2.4) eşitliğinden bulunur.

$$\text{Cov}_\chi(t, s) = c_\chi(t, s) = \mathbb{E}[(\chi(t) - \mu_\chi(t)) (\chi(s) - \mu_\chi(s))]. \quad (2.4)$$

Kitleye ait korelasyon fonksiyonu ise (2.3) ve (2.4) eşitlikleri aracılığıyla şu şekilde hesaplanır:

$$\text{Cor}_\chi(t, s) = r_\chi(t, s) = \frac{\text{Cov}_\chi(t, s)}{\sqrt{\text{Var}_\chi(t)\text{Var}_\chi(s)}}. \quad (2.5)$$

Fonksiyonel durum için bir başka önemli parametre de (2.6) eşitliği ile tanımlanan kovaryans operatörüdür.

$$\Gamma_\chi(\xi) = \mathbb{E}[\langle (\chi - \mu_\chi), \xi \rangle (\chi - \mu_\chi)]. \quad (2.6)$$

Kovaryans operatörü fonksiyonel veri analizinde varyans-kovaryans matrisinin rolünü üstlenir. Bundan ötürü, özellikle Bölüm 3'te detaylı olarak açıklanan fonksiyonel temel bileşenlerin hesabında büyük öneme sahiptir.

$\chi_1(t), \chi_2(t), \dots, \chi_N(t)$, $L^2(T)$ uzayında tanımlı, karesi integrallenebilen, $\chi(t)$ ile aynı dağılıma sahip ve birbirinden bağımsız N adet fonksiyonel gözlem olsun. Bu fonksiyonel örnekleme ait ortalama, varyans ve kovaryans fonksiyonları (2.2), (2.3) ve (2.4) eşitlikleri ile verilen denklemlerin örneklem eşleniklerinden kestirilir. Örneklem ortalama fonksiyonu (2.7) eşitliğinden hesaplanır:

$$\hat{\mu}_\chi(t) = \bar{\chi}(t) = \frac{1}{N} \sum_{i=1}^N \chi_i(t). \quad (2.7)$$

Benzer şekilde, örneklem varyans fonksiyonu da (2.8) eşitliğinden bulunur:

$$\hat{\sigma}_\chi^2(t) = \frac{1}{N-1} \sum_{i=1}^N [\chi_i(t) - \bar{\chi}(t)]^2. \quad (2.8)$$

Örnekleme ait standart hata fonksiyonu, varyans fonksiyonunun nokta bazlı karekökü alınarak hesaplanır ve $\hat{\sigma}_\chi(t)$ ile gösterilir.

Kovaryans fonksiyonu, korelasyon fonksiyonu ve kovaryans operatörünün ampirik versiyonları sırasıyla şu eşitliklerden hesaplanır:

$$\hat{c}_\chi(t, s) = \frac{1}{N-1} \sum_{i=1}^N [\chi_i(t) - \bar{\chi}(t)][\chi_i(s) - \bar{\chi}(s)], \quad (2.9)$$

$$\hat{r}_\chi(t, s) = \frac{\hat{c}_\chi(t, s)}{\sqrt{\hat{\sigma}_\chi^2(t)\hat{\sigma}_\chi^2(s)}}, \quad (2.10)$$

$$\hat{\Gamma}_\chi(\xi) = (N-1)^{-1} \sum_{i=1}^N [(\chi_i - \bar{\chi}), \xi(t)] (\chi_i - \bar{\chi}), \quad \xi(t) \in L^2(T). \quad (2.11)$$

Fonksiyonel bir örneklemin ortalamasının hesabı fonksiyonel bir dağılımın merkezinin kestirilmesine eş değerdir. Fonksiyonel örneklemin kovaryans fonksiyonunun hesabı ise fonksiyonel dağılımın yayılım ve korelasyon yapısının kestirimidir (Febrero ve diğ., 2007).

Gözlenen iki fonksiyonelin birbiri ile ilişkisi sırasıyla (2.12) ve (2.13) eşitlikleri ile verilen çapraz-kovaryans ve çapraz-korelasyon fonksiyonları aracılığıyla araştırılır.

$$\text{Cov}_{\chi,\gamma}(t,s) = \hat{c}_{\chi,\gamma}(t,s) = \frac{1}{N-1} \sum_{i=1}^n [\chi_i(t) - \bar{\chi}(t)][\gamma_i(s) - \bar{\gamma}(s)], \quad (2.12)$$

$$\text{Cor}_{\chi,\gamma}(t,s) = \hat{r}_{\chi,\gamma}(t,s) = \text{Cov}_{\chi,\gamma}(t,s) / \text{Var}_{\chi}(t) \text{Var}_{\gamma}(s). \quad (2.13)$$

Bir veri setine ait merkezi ve yayılım ölçülerini kestirmek için ortalama ve standart sapma fonksiyonları dışında başka kestiriciler de kullanılabilir. Medyan ve kesilmiş ortalama sağlam merkezi ölçüler iken, ortanca mutlak sapma (median absolute deviation) ve kesilmiş standart sapma sağlam yayılım ölçüleridir. $0 \leq \alpha \leq (n-1)/n$ olmak üzere, fonksiyonel α -kesilmiş ortalama en merkezi $n - [\alpha n]$ eğrinin ortalamasını verir. En merkezi eğrinin bulunmasında derinlik ölçüleri de kullanılabilir.

2.2.2 Derinlik Ölçüleri

Derinlik, verilen bir veri bulutu içinde gözlemin merkeziliğini ölçer. Skaler değişkenler için en sık kullanılan derinlik ölçüleri yarı-uzay derinliği (Half-Space Depth, HSD), basit derinlik (Simplicial Depth, SD) ve Fraiman-Muniz derinliğidir (FMD).

Tek değişkenli durumda, bir noktanın derinliği X_1, X_2, \dots, X_N gözlemlerinden oluşan bir örneklemin kümülatif dağılım fonksiyonu F_N 'ye bağlı olarak hesaplanır.

$$F_N(x) = \frac{1}{N} \sum_{k=1}^N \mathbf{I}\{X_k \leq x\}. \quad (2.14)$$

X_i gözlemine ait HSD değeri, (2.14) kullanılarak şu şekilde hesaplanır:

$$\text{HSD}(X_i) = \min(F(X_i), 1 - F(X_i)). \quad (2.15)$$

X_i 'nin ortanca olması durumunda, $F(X_i)$ değeri $1/2$ 'ye eşittir. Bundan dolayı, $\text{HSD}(X_i)$ de $1/2$ değeri ile olası en büyük derinliği verir. Eğer X_i veri setindeki en büyük nokta ise, $F(X_i) = 0$ ve $\text{HSD}(X_i) = 0$ olur ve bu da olası en az derinliği verir (Horváth ve Kokoszka, 2012).

SD ise, olası en büyük ve en küçük derinlikler sırasıyla 1/2 ve 0 değerlerini alacak şekilde, (2.16) eşitliğinden hesaplanır.

$$SD(X_i) = 2F(X_i)(1 - F(X_i)). \quad (2.16)$$

Fraiman ve Muniz (2001) tarafından önerilen derinlik ölçüsü FMD 1/2 ve 1 arasında değerler alır ve

$$FMD(X_i) = 1 - \left| \frac{1}{2} - F(X_i) \right|, \quad (2.17)$$

eşliğinden hesaplanır.

Fonksiyonlardan oluşan $\{\chi_i(t), t \in T, i = 1, 2, \dots, N\}$ gibi bir örneklem için $t \in [t_{\min}, t_{\max}]$ noktasındaki ampirik kümülatif dağılım fonksiyonu, $\mathbf{I}(\cdot)$ gösterge fonksiyonu olmak üzere,

$$F_{N,t}(\chi_i(t)) = \frac{1}{N} \sum_{k=1}^N \mathbf{I}\{\chi_k(t) \leq \chi_i(t)\}, \quad (2.18)$$

ile ifade edilir (Febrero ve diğ., 2007).

X_i 'nin t noktasındaki derinliği \mathbb{R} uzayında D ile gösterilirse, $\chi_1, \chi_2, \dots, \chi_N$ eğrilerinin fonksiyonel derinliği (FD) tek değişkenli derinliklerin integralinden bulunur:

$$FD(\chi_i) = \int_{t_{\min}}^{t_{\max}} D(\chi_i(t)) dt, \quad i = 1, \dots, N. \quad (2.19)$$

Örneğin, $\chi_i(t)$ eğrisinin fonksiyonel Fraiman-Muniz derinliği (FMD),

$$FMD(\chi_i) = \int_{t_{\min}}^{t_{\max}} \left[1 - \left| \frac{1}{2} - F_{N,t}(\chi_i(t)) \right| \right] dt, \quad (2.20)$$

ile tanımlanır.

Tek değişkenli duruma benzer şekilde, maksimum ve minimum değer alan eğriler sırasıyla en derin ve en az derin eğrilerdir (Febrero ve diğ., 2007).

Bunlar dışında, modal derinlik, Rasgele İzdüşüm Derinliği (Random Projection Depth, RPD) ve bant-derinliği gibi başka derinlik ölçüleri de vardır. Bu teknikler hakkında

ayrıntılı bilgi Febrero ve diğ. (2008) ve López-Pintado ve Romo (2009) kaynaklarında bulunabilir.

Derinlik, aykırı değerliliğin önemli bir ölçüsüdür. Çünkü, anlamlı derecede küçük derinliğe sahip eğriler aykırı değer olmaya daha yatkındır. Fonksiyonel durumda, aynı dağılımdan geldiği varsayılan eğrilerin oluşturduğu bir fonksiyonel uzayda, diğer eğrilerinkinden farklı bir dağılıma sahip olan eğri fonksiyonel aykırı değer olarak nitelendirilir.

Derinlik ölçüleri kullanılarak aykırı değer belirlenmesi için aşamalar şu şekilde özetlenebilir (Horváth ve Kokoszka, 2012):

1. $F(\chi_1), F(\chi_2), \dots, F(\chi_N)$ değerleri hesaplanır.
2. C gibi bir kesim değeri belirlenir ve C değerinden daha küçük derinliğe sahip eğriler veri setinden çıkarılarak aykırı değer olarak sınıflandırılır. Eğer böyle bir eğri bulunamıyorsa, sürece bu aşamada son verilir.
3. Birinci adıma geri dönülür ve ikinci adımda aykırı değer olarak belirlenen eğriler çıkarıldıktan sonra geriye kalan örnekleme aynı adımlar uygulanır.

Febrero-Bande ve Oviedo de la Fuente (2012) bootstrap örneklemlerine dayalı iki farklı aykırı değer belirleme yöntemi önermişlerdir. Bu yöntemlerden biri veri setinin iki uçtan kesilmesine (trimming) dayalı iken diğeri verinin ağırlıklandırılmasına dayalıdır. Bu yöntemler sırasıyla "outliers.depth.trim()" ve "outliers.depth.pond()" fonksiyonları ile "fda.usc" R paketine eklenmiştir (Febrero-Bande ve Oviedo de la Fuente, 2012; Oviedo de la Fuente, 2011). Birinci yöntemde, bootstrap örneklemleri veri eğrilerinin $\% \alpha$ kadarı çıkarıldıktan sonra seçilir. İkinci yöntemde ise, her gözlem (her eğri) derinlik değerine bağlı olarak ağırlıklandırılır. Her iki yöntemin de son aşamaları benzerdir. "outliers.depth.pond()" fonksiyonunun algoritması şu şekilde özetlenebilir:

1. SD, FMD veya RPD gibi bir fonksiyonel derinlik ölçüsü seçilir ve $\chi_1(t), \dots, \chi_n(t)$ fonksiyonel veri seti için $D(\chi_1), \dots, D(\chi_n)$ fonksiyonel derinlikleri hesaplanır.
2. Her eğri kendi derinliğine orantılı olarak ağırlıklandırılacak şekilde fonksiyonel veri setinden B sayıda bootstrap örnekleme χ_i^b alınır.

3. $Z_i^b(t_1), \dots, Z_i^b(t_N)$ 'nin sıfır ortalamaya ve Σ_χ , $\chi(t_1), \dots, \chi(t_N)$ 'ye ait kovaryans matrisini ve γ düzgünleştirme parametresini ifade etmek üzere $\gamma\Sigma_\chi$ kovaryans matrisine sahip normal dağılımdan geldiği göz önüne alınarak bootstrap örneklemi $Y_i^b = \chi_i^b + Z_i^b$ şeklinde düzleştirilir.

4. Her bootstrap örnekleme için ($b = 1, \dots, B$), C^b kesim değeri $D(Y_i^b)$ derinliklerinin dağılımının c yüzdeliği olarak $P(D_n(\chi_i) \leq C^b) = c$, $i = 1, \dots, N$ eşitliğinden seçilir. Genelde $c = 0.01$ olarak alınır. 5. C^b değerlerinin ortancası alınarak kesim değeri C belirlenir (Oviedo de la Fuente, 2011).

Daha sonra Horváth ve Kokoszka (2012)'nin önerdiği bir önceki algoritmada verilen ikinci ve üçüncü adımlar uygulanarak aykırı gözlemler bulunur.

2.3 Fonksiyonel Verinin Temsili

Ham verinin fonksiyonel objelere dönüştürülmesi fonksiyonel veri analizinin ilk adımını oluşturur (Levitin ve diğ., 2007). Bu, iki şekilde yapılabilir. Eğer kesikli verilerin hatasız olduğu varsayılıyorsa, interpolasyon kullanılır. Eğer hatanın gözlemlendiği varsayılıyorsa, bu dönüşüm düzleştirme ile yapılır (Ramsay ve Silverman, 2005). Bundan dolayı, interpolasyon ve düzgünleştirme FVA'nın en önemli kısmıdır. Bu işlem fonksiyonel bir verinin kesikli noktalarını düzgün bir fonksiyona dönüştürmeye yarar.

Düzensizlik, bir χ fonksiyonunu oluşturan iki komşu noktanın bağlantılı olduğunu ve birbirinden çok da farklı olmadığını ifade eder (Ramsay ve Silverman, 2005). Eğer veri bir L^2 veya Hilbert uzayında tanımlanmışsa, düzgünleştirme işlemi taban fonksiyonları aracılığıyla yapılabilir. Aksi halde, düzgünleştirmede çekirdek yöntemler kullanılabilir (Febrero-Bande ve Oviedo de la Fuente, 2012).

2.3.1 Taban Açılımı

Taban sistemi, birbirinden bağımsız olan fonksiyonların oluşturduğu bir sistemdir. Taban fonksiyonu yaklaşımı bir χ fonksiyonunu K adet bilinen ϕ_k taban fonksiyonlarının doğrusal birleşimi olarak tanımlar:

$$\chi(t) = \sum_{k=1}^K c_k \phi_k(t). \quad (2.21)$$

\mathbf{c} , K uzunluğunda c_k katsayılarından oluşan bir vektör ve $\boldsymbol{\phi}$, $n \times K$ boyutunda, $\phi_k(t_j)$ taban fonksiyonlarını içeren bir matris olmak üzere (2.21) ile verilen ifade matris gösteriminde de aşağıdaki gibi yazılabilir (Ramsay ve Silverman, 2005):

$$\chi(t) = \mathbf{c}' \boldsymbol{\phi} = \boldsymbol{\phi}' \mathbf{c}. \quad (2.22)$$

c_k katsayıları, eğri oluşturulurken her bir ϕ_k taban fonksiyonuna verilen ağırlığı ifade eder.

Açılımın boyutu, taban fonksiyonlarının sayısı olan K 'ya bağlıdır. Taban açılımı yaklaşımı sonsuz boyutlu fonksiyonları, sonlu boyutlu uzayda resmederek onlarla çalışılmasını sağlar. Düzgünleştirmede en önemli adımlar kullanılacak baz fonksiyonun ve K fonksiyon sayısının seçimidir. K sayısı ne kadar küçük ise hesaplamada o kadar kolaylık sağlar.

Hangi taban açılımının kullanılacağına verinin yapısına göre karar verilir. Sıklıkla kullanılan taban fonksiyonları Fourier ve B-splayn fonksiyonlarıdır. Fourier tabanı daha çok periyodik veriler için tercih edilirken, periyodik olmayan veriler için genelde B-splayn tabanı kullanılır.

Belirtilen bu iki yöntem dışında tek terimli (monomial) tabanlar, dalgacık (wavelet) tabanları, çok terimli (polynomial) tabanlar, üssel ve kuvvet tabanları, poligonal tabanlar ve sabit taban gibi taban fonksiyonları da bulunmaktadır. Dalgacık tabanları daha çok türevin gerekli olmadığı durumlarda tercih edilir. Sabit taban ise bu taban fonksiyonları arasında en basit olanıdır. Daha çok skaler bir gözlemin fonksiyonel veri olarak ifade edilmesinde kullanılır. Bundan dolayı, özellikle fonksiyonel doğrusal modellerde sabit terimin (intercept) belirlenmesinde önemlidir. Çok terimli taban fonksiyonları ise daha basit fonksiyonel problemler için kullanılır.

2.3.1.1 Fourier Tabanı

En iyi bilinen taban açılımı (2.23) ile verilen ve katsayıları sırasıyla $\phi_0(t) = 1$, $\phi_{2r-1}(t) = \sin r\omega t$, $\phi_{2r}(t) = \cos r\omega t$ eşitliklerini gerçekleyen,

$$\hat{\chi}(t) = c_0 + c_1 \sin \omega t + c_2 \cos \omega t + c_3 \sin 2\omega t + c_4 \cos 2\omega t + \dots, \quad (2.23)$$

Fourier serisi açılımıdır. Bu seri $2\pi/\omega$ periyodu ile periyodiktir.

Bir serinin türevinin baz yaklaşımı,

$$D\hat{\chi}(t) = \sum_{i=1}^K \hat{c}_i D\phi_i(t) = \hat{c}' D\phi(t), \quad (2.24)$$

şeklinde yazılabilir.

Fonksiyonu iyi temsil eden bir taban fonksiyonu her zaman türev kestirimi için uygun olmayabilir. Bundan dolayı, bir ya da daha fazla türevin yaklaşımının iyi olup olmadığının kontrol edilmesi önerilir (Ramsay ve Silverman, 2005).

Bir fonksiyonunun türevleri için Fourier serisi açılımı katsayıları şu şekilde hesaplanır:

$$D\sin r\omega t = r\omega \cos r\omega t$$

$$D\cos r\omega t = -r\omega \sin r\omega t.$$

Yüksek mertebeden türevler için Fourier açılımı, katsayıların $r\omega$ 'nin uygun kuvvetleri ile çarpılmasından kolaylıkla bulunabilir. Örneğin, bir fonksiyonun ikinci dereceden türevi $D^2\chi$ 'ye ait katsayı vektörü,

$$(0, -\omega^2 c_1, -\omega^2 c_2, -4\omega^2 c_3, -4\omega^2 c_4, \dots),$$

şeklinde bulunabilir.

Fourier tabanının eşit aralıklı ve periyodik veriler için hesap kolaylığı sağlayan özellikleri vardır. Özellikle eğriliğin her yerde aynı dereceden olmaya meyilli olduğu durağan fonksiyonlar için uygundur. Fourier serisi açılımları, matematik, istatistik ve mühendislik alanındaki uygulamalarda çok popülerdir. Fakat, periyodik olmayan ve eşit aralıkla ölçülmemiş veriler için kullanımı önerilmez. Bu durumda, B-splayn gibi başka türde taban fonksiyonlarının kullanılması daha iyi olur.

Ramsay ve Silverman (2005)'in kitabında da belirtildiği gibi:

"Fourier serileri margarin gibidir: Ucuzdur ve pratik olarak herşeyin üzerine sürülebilir ancak ortaya çıkan sonucun her zaman heyecanla yiyilebileceğini ümit etme."

2.3.1.2 B-splayn tabanı

Splayn fonksiyonları periyodik olmayan fonksiyonel veri ya da parametreler için en sık kullanılan yaklaşım sistemi seçeneğidir (Ramsay ve Silverman, 2005). Splayn fonksiyonlarının avantajı hızlı hesap edilebilmeleri ve az sayıda fonksiyon ile veriye iyi bir yaklaşım sergileyebilmeleridir (Horváth ve Kokoszka, 2012).

Bir splayn oluşturmak için ilk adım verinin tanımlı olduğu aralığı, L adet alt aralığa bölmektir. Bir aralığı diğerinden ayıran noktalara kesim noktaları adı verilir ve $l = 1, \dots, L - 1$ olmak üzere τ_l ile gösterilirler. Kesim noktası sayısı arttıkça splaynlar esneklik kazanırlar. Bir splayn tabanının kesim noktalarındaki değerleri dizisi düğüm noktaları olarak adlandırılır. Bir kesim noktası ile ilişkili düğüm noktalarının sayısı birden fazla olabilmektedir. Düğüm noktaları ayrık ise, düğüm noktaları ve kesim noktaları aynı şeyi ifade eder.

Bir splayn fonksiyonu her aralıkta tanımlanan sabit dereceli bir polinomdur. Bir polinomun derecesi, polinomun en yüksek kuvvetine eşittir. Polinomun sırası ise, polinomu tanımlamak için gerekli olan sabit sayısını ifade eder. Polinomun sırası, polinomun derecesinin bir fazlasına eşittir ve m ile gösterilir (Ramsay ve diğ., 2009). $m = 1$ ile sıfıncı dereceden bir adım fonksiyonu, $m = 2$ ile parçalı doğrusal bir fonksiyon, $m = 3$ ve $m = 4$ ile ise sırasıyla kuadratik ve kübik bir polinom ifade edilir. Uygulamalarda genellikle dördüncü sıradan splaynlar kullanılır. Çünkü oldukça düzgün olmalarının yanısıra birinci ve ikinci dereceden türevleri mevcuttur. Eğer düzgün türev fonksiyonlarına ihtiyaç varsa, Ramsay ve diğ. (2009) en yüksek dereceden türevin en az iki fazlası sıraya sahip polinomun kullanılmasını önerirler.

Düğüm noktalarının kesim noktasına eşit olması durumunda splayn sisteminin baz fonksiyonlarının sayısı K , polinomun sıra sayısı m ile iç düğüm nokta sayısının toplamına eşittir (Ramsay ve Silverman, 2005). Bu toplam şu şekilde ifade edilir:

$$K = m + L - 1. \quad (2.25)$$

Splaynların bazı temel özellikleri şu şekilde özetlenebilir:

1. Her taban fonksiyonu $\phi_k(t)$, τ düğüm serisine sahip, m sıralı bir splayn fonksiyonudur.

2. Splaynların toplamı, farkı, çarpımı ve doğrusal kombinasyonu yine bir splayn fonksiyonudur.
3. m sıralı herhangi bir splayn ve τ düğüm serisi bu taban fonksiyonlarının doğrusal bir kombinasyonu olarak ifade edilebilir.

2.3.1.3 Yanlılık - Varyans Dengesi

Taban fonksiyonu sayısına karar verilmesi düzgünleştirmenin önemli bir adımıdır. K taban sayısını belirlerken en iyi yaklaşımı bulmak için kestirimin yanlılığına ve varyansına dikkat edilmelidir. Yanlılık ve varyans sırasıyla (2.26) ve (2.27) eşitlikleri ile tanımlanır.

$$\text{Yan}[\hat{\chi}(t)] = \chi(t) - \mathbb{E}[\hat{\chi}(t)], \quad (2.26)$$

$$\text{Var}[\hat{\chi}(t)] = \mathbb{E}[\hat{\chi}(t) - \mathbb{E}[\hat{\chi}(t)]^2]. \quad (2.27)$$

Fazla sayıda taban fonksiyonu kullanmak yanlılığı azaltarak veriyi iyi temsil edebilir. Ancak bu $\hat{\chi}(t)$ kestiriminin varyansını artırır. Benzer şekilde, az sayıda taban fonksiyonu kullanmak varyansı düşürürken kestirimin yanlılığını artırır (Levitin ve diğ., 2007). Taban sayısı ve düzgünlük arasındaki çelişkiden kaynaklanan bu duruma yanlılık-varyans dengesi adı verilir.

Bu nedenle, taban sayısını belirlerken Hata Kareler Ortalaması (Mean Square Error, MSE) adı verilen bir kriter kullanılır. MSE hem kestirimin yanlılığını hem de varyansını göz önüne alır ve eşitlik (2.28)'daki gibi hesaplanır.

$$\text{MSE}[\hat{\chi}(t)] = \mathbb{E}[\hat{\chi}(t) - \chi(t)]^2. \quad (2.28)$$

Bu kriter, yanlılık ve varyans tanımları kullanılarak (2.29)'deki gibi de yazılabilir.

$$\text{MSE}[\hat{\chi}(t)] = \text{Yan}^2[\hat{\chi}(t)] + \text{Var}[\hat{\chi}(t)]. \quad (2.29)$$

MSE örneklem varyansında büyük bir azalış olduğunda az derece yanlılığın göz ardı edilmesine izin verir (Ramsay ve Silverman, 2005).

Taban sayısını belirlemek için önerilen bir başka yöntem de adımsal değişken seçimidir. Bu yöntem, taban fonksiyonlarının adımsal olarak eklenmesine dayanır. Her adımda uyumun gelişip gelişmediği test edilerek taban fonksiyonu sayısına karar verilir (Ramsay ve Silverman, 2005).

2.3.2 Kernel (Çekirdek) Düzleştirme

Fonksiyonel veri, ağırlıklandırma prensibine dayalı parametrik olmayan yöntemler kullanılarak da temsil edilebilir. Kernel yöntemi popüler parametrik olmayan yöntemlerden biridir. Kernel düzleştirmede ağırlık değerleri (2.30)'daki gibi tanımlanan bir kernel fonksiyonuna bağlıdır.

$$w_i(t) = K\left(\frac{t_i - t}{h}\right). \quad (2.30)$$

$K(\cdot)$ ile tanımlanan çekirdek fonksiyonu Normal (Gaussian), Tekdüze (Uniform), Kuadratik (Quadratic), Epanechnikow, "Triweight" ya da "Cosine Kernel" çekirdek fonksiyonlarından biri olarak seçilebilir. Burada, h bant genişliği parametresi yoğunluğun bir ölçüsüdür. $w_i(t)$ ise bir t noktasından uzaklığa bağlı olarak elde edilen ağırlık değerlerini tanımlar. h değerinin küçük olması sadece t noktasına yakın gözlemlerin ağırlıklandırılacağını gösterir. Ters durumda ise, t noktasından uzakta bulunan gözlemler de ağırlıklandırılır.

Kernel düzleştirmede belli bir noktadaki kestirim değeri, doğrusal düzleştirici adı verilen yerel ağırlıklara bağlı bir ağırlık ölçüsü $S_j(t)$ ve y_j gözlemlerinin doğrusal bir kombinasyonu olarak yazılabilir:

$$\hat{\chi}(t_j) = \sum_{j=1}^n S_j(t) y_j. \quad (2.31)$$

En popüler çekirdek kestiricisi (2.32) eşitliği ile verilen Nadaraya-Watson kestiricisidir.

$$S_j(t) = \frac{K\left[\frac{t_j - t}{h}\right]}{\sum_{l=1}^n K\left[\frac{t_l - t}{h}\right]}. \quad (2.32)$$

Çekirdek düzleştirmede kullanılabilen diğer olası kestiriciler k en yakın komşuluk (k nearest neighbors, KNN) kestiricisi, yerel doğrusal regresyon kestiricisi ve yerel polinomial kestiricidir (Wasserman, 2006; Febrero-Bande ve Oviedo de la Fuente, 2012).

Çekirdek düzleştirmede türevler kestiricinin türevi alınarak kolayca hesaplanamaz. Çünkü, tekdüze ve kuadratik çekirdekler gibi bazı çekirdek fonksiyonları türevlenebilir değildir. Türevin kestirilebilmesi için, bant genişliği parametresinin fonksiyonu kestirirken kullanılan değerden daha büyük olması gerekmektedir (Ramsay ve Silverman, 2005).

2.3.3 Fonksiyonel Temel Bileşenler Tabanı

Temel bileşenler analizi, FVA'da da çok değişkenli durumda olduğu gibi bir boyut indirgeme yöntemi olarak kullanılır. Aralarındaki en önemli fark fonksiyonel durumda bileşenler vektör yerine fonksiyonlardır (Levitin ve diğ., 2007).

Fonksiyonel Temel Bileşenler (FTB) tabanı, fonksiyonel verinin bileşen skorlarının varyansı $N^{-1} \sum_{j=1}^K f_{ij}^2$ 'yi en büyükleyen ortonormal değişkenler cinsinden temsil edilmesini sağlar. Verideki en önemli özelliklerin ortaya çıkarılabilmesi için bu maksimizasyon gereklidir (Ramsay ve Silverman, 2005).

FTB'nin temel algoritması şu şekildedir:

1. Hata kareler toplamını minimize etmek için $\sum f_{i1}^2$ 'nin $\|\xi_1\|^2 = 1$ şartına göre $\sum f_{i1}^2$ 'yi en büyükleyen normu bire eşit olan $f_{i1} = \int_T \xi_1(t) \chi_i(t) dt$ skorlu birinci temel bileşen ağırlık fonksiyonunun ($\xi_1(t)$) belirlenmesi.
2. İki bileşenin ortogonalite şartı $\int \xi_1(t) \xi_2(t) dt = 0$ göz önünde bulundurularak $\sum f_{i2}^2$ 'nin $\|\xi_2\|^2 = 1$ 'e göre en büyüklenmesinden bir sonraki ağırlık fonksiyonu $\xi_2(t)$ 'nin ve temel bileşen skorları $f_{i2} = \int \xi_2(t) \chi_i(t) dt$ 'nin hesaplanması.
3. Bu adımların temel bileşen sayısı kadar tekrarı.

Fonksiyonel değişkenler, f_{ij} temel bileşen skorları ve ξ_j öz fonksiyonları kullanılarak sonlu ortonormal tabanda aşağıdaki gibi yazılabilirler:

$$\hat{\chi}_i(t) = \sum_{j=1}^K f_{ij} \xi_j(t), \quad i = 1, \dots, N. \quad (2.33)$$

Bu yöntem Bölüm 3'te daha detaylı olarak anlatılmıştır.

2.3.4 Fonksiyonel Kısmi En Küçük Kareler (FKEKK) Tabanı

Kısmi En Küçük Kareler (KEKK), Temel Bileşenler Analizine (TBA) alternatif bir boyut indirgeme yöntemidir. KEKK yaklaşımı ve regresyon analizi genellikle birlikte kullanılırlar. Bundan dolayı, bu yöntem Kısmi En Küçük Kareler Regresyonu (KEKKR) olarak da bilinmektedir. KEKKR, açıklayıcı değişkenlerin stokastik bir süreçten oluştuğu duruma Preda ve Saporta (2005) tarafından uyarlanmıştır. Fonksiyonel bağlamda bu yöntem, skaler yanıt değişkeni için fonksiyonel doğrusal regresyon modelleri ile kullanılır ve Fonksiyonel Kısmi En Küçük Kareler Regresyonu (FKEKKR) adını alır. Bu yöntem Bölüm 4'te ayrıntılı olarak açıklanmıştır. Genel anlamda bu yöntem, skaler yanıt değişkeni ve fonksiyonel açıklayıcı değişkenler arasındaki korelasyon değeri dikkate alınarak KEKK bileşenlerinin belirlenmesini hedefler. FKEKKR yönteminin başta kemometri alanında olmak üzere literatürde pek çok uygulaması mevcuttur (Saeys ve diğ., 2008; Aguilera ve diğ., 2010; Preda ve Schiltz, 2011).

FKEKK yönteminin ana fikri $\chi_i(t)$ fonksiyonel değişkeninin φ_l ; FKEKK bileşenleri ve v_{il} ; FKEKK skor değerleri kullanılarak temsil edilmesidir:

$$\hat{\chi}_i(t) = \sum_{l=1}^K v_{il} \varphi_l(t), \quad i = 1, \dots, N. \quad (2.34)$$

Bu bileşenlerin elde edilmesi için gerekli algoritma izleyen bölümlerde açıklanmıştır. Bu yöntemle ilgili detaylı bilgi Preda ve Saporta (2005) ve Aguilera ve diğ. (2010) kaynaklarında bulunabilir. FKEKK algoritması R programında fda.usc paketi içinde de bulunmaktadır (Febrero-Bande ve Oviedo de la Fuente, 2012).

2.4 Fonksiyonel Veriyi Düzleştirme

Fonksiyonel verinin düzleştirilmesindeki en önemli nokta ham veriye en uygun olan modelin belirlenmesidir. Bu aşamada, kullanılan düzleştirme yöntemine bağlı olarak çeşitli kriterler kullanılabilir. Fonksiyonel verinin düzleştirilmesinde kullanılan en temel kriterler: Sıradan (Ağırlıksız) En Küçük Kareler (SEKK) yöntemi, Ağırlıklı

En Küçük Kareler (AEKK) yöntemi, Yerel En Küçük Kareler (YEKK) ve Pürüzlü Ceza Yöntemidir (PCY).

EKK genellikler taban fonksiyonu ile düzleştirmede, taban fonksiyonu katsayılarının elde edilmesinde kullanılır. Otokorelasyon durumunda EKK yerine Ağırlıklı En Küçük Kareler (AEKK) yöntemi tercih edilir. YEKK kriteri genellikle ağırlık fonksiyonlarının çekirdek bir fonksiyon tarafından belirlendiği çekirdek düzleştirme yönteminde kullanılır. PCY ise veriyi düzleştirirken fazladan bir ceza parametresinin yanısıra Cezalı En Küçük Kareler (CEKK) adı verilen bir kriter kullanır.

2.4.1 En Küçük Kareler Yöntemi

Fonksiyonel ham veriyi bir taban fonksiyonu ile temsil etme problemi, $y_j = (y_1, y_2, \dots, y_n)$ gözlenen kesikli veri vektörü, $\chi(t_j)$ (2.21) formunda bir açılım, \mathbf{t} , t_j , $j = 1, \dots, n$ kesikli gözlem noktalarından oluşan bir vektör ve ε_j hata terimlerini içeren n uzunluğunda bir vektör olmak üzere,

$$y_j = \chi(t_j) + \varepsilon_j, \quad (2.35)$$

şeklinde ifade edilen bir regresyon problemidir.

2.4.1.1 Sıradan (Ağırlıksız) En Küçük Kareler Yöntemi

(2.21) eşitliğindeki c_k katsayıları veriye en iyi uyumu sağlayacak şekilde belirlenmelidir. Bu nedenle, bu katsayılar en küçük kareler kestirimi kullanılarak hata kareler toplamının (Sum of Square Errors, SSE),

$$SSE = \sum_{j=1}^n \left[y_j - \sum_{k=1}^K c_k \phi_k(t_j) \right]^2 = \sum_{j=1}^n [y_j - \boldsymbol{\phi}(t_j)' \mathbf{c}]^2, \quad (2.36)$$

en küçüklenmesinden elde edilir. Bu ifade matris formunda şu şekilde yazılabilir:

$$(\mathbf{y} - \boldsymbol{\phi} \mathbf{c})' (\mathbf{y} - \boldsymbol{\phi} \mathbf{c}). \quad (2.37)$$

(Ramsay ve diğ., 2009; Hooker, 2010). Fonksiyonel notasyonda bu ifade $\| \mathbf{y} - \boldsymbol{\phi} \mathbf{c} \|^2$ normuna denktir.

(2.37) ifadesinin \mathbf{c} 'ye göre türevi alınıp bu problem \mathbf{c} için çözümlerse,

$$2\boldsymbol{\phi}\boldsymbol{\phi}'\mathbf{c} - 2\boldsymbol{\phi}'\mathbf{y} = 0, \quad (2.38)$$

$\hat{\mathbf{c}}$ olarak ifade edilen sıradan en küçük kareler kestirimi elde edilir.

$$\hat{\mathbf{c}} = (\boldsymbol{\phi}'\boldsymbol{\phi})^{-1}\boldsymbol{\phi}'\mathbf{y}. \quad (2.39)$$

Kestirilen eğri değerleri $\hat{\chi}(t)$, (2.39) denklemi kullanılarak (2.40)'deki gibi hesaplanır.

$$\hat{\chi}(t) = \boldsymbol{\phi}(\boldsymbol{\phi}'\boldsymbol{\phi})^{-1}\boldsymbol{\phi}'\mathbf{y}. \quad (2.40)$$

Düzleştirilmiş $\hat{\chi}(t)$ değerleri, düzleştirme matrisi $S_j(t)$ matrisi aracılığıyla, gözlenen kesikli y_j değerlerinin doğrusal dönüşümü olarak yazılabilir. Bu işleme doğrusal düzleştirme adı verilir.

$$\hat{\chi}(t) = \sum_{j=1}^n S_j(t) y_j. \quad (2.41)$$

Burada $S_j(t)$ yerel ağırlıklara bağlı olarak hesaplanan bir ağırlık ölçüsüdür ve j 'sinci kesikli veri değerini ağırlıklandırır.

$\hat{\chi}(\mathbf{t})$, $\chi(\mathbf{t})$ fonksiyonunun t_j noktalarındaki kestirim değerlerinden oluşan bir sütun vektörü ve \mathbf{S} bağımlı değişken vektörünü uyumlu değerlere atayan düzleştirme matrisi olmak üzere, (2.41) ifadesi matris formunda (2.42)'deki gibi yeniden yazılabilir.

$$\hat{\chi}(\mathbf{t}) = \mathbf{S}\mathbf{y}. \quad (2.42)$$

SEKK yöntemine ait düzleştirme matrisi (2.43)'teki gibi tanımlanır:

$$\mathbf{S} = \boldsymbol{\phi}(\boldsymbol{\phi}'\boldsymbol{\phi})^{-1}\boldsymbol{\phi}'. \quad (2.43)$$

2.4.1.2 Ağırlıklı En Küçük Kareler (AEKK) Yöntemi

SEKK yaklaşımının, ε_j hatalarının sıfır ortalama ve σ^2 sabit varyansı ile bağımsız ve eş dağıldığı durumda kullanılması uygundur. Eğer hatalar arasında otokorelasyon

mevcut ise veya gözlem zamanına göre varyans değişiyorsa verinin düzleştirilmesinde Ağırlıklı En Küçük Kareler (AEKK) yönteminin kullanılması önerilir (Ramsay ve Silverman, 2005).

AEKK kriteri simetrik ve pozitif tanımlı olan bir \mathbf{W} matrisinin eklenmesiyle,

$$(\mathbf{y} - \boldsymbol{\phi}\mathbf{c})'\mathbf{W}(\mathbf{y} - \boldsymbol{\phi}\mathbf{c}), \quad (2.44)$$

olarak tanımlanır.

Bu durumda, AEKK kestirimi $\hat{\mathbf{c}}$,

$$\hat{\mathbf{c}} = (\boldsymbol{\phi}'\mathbf{W}\boldsymbol{\phi})^{-1}\boldsymbol{\phi}'\mathbf{W}\mathbf{y}, \quad (2.45)$$

eşitliğinden bulunur.

Buradan, AEKK kestirimi için düzleştirme matrisi

$$\mathbf{S} = \boldsymbol{\phi}(\boldsymbol{\phi}'\mathbf{W}\boldsymbol{\phi})^{-1}\boldsymbol{\phi}'\mathbf{W}, \quad (2.46)$$

şeklinde yazılır.

Klasik istatistikteki gibi \mathbf{W} ağırlık matrisi en basit şekilde hataların varyans-kovaryans matrisinden elde edilebilir:

$$\mathbf{W} = \boldsymbol{\Sigma}_e^{-1}. \quad (2.47)$$

2.4.2 Yerel En Küçük Kareler (YEKK) Yöntemi

Bir fonksiyonun bir t_j noktasındaki kestirim değerleri bu noktaya yakın gözlemlerden etkilenir. Bu durum göz önüne alınarak, t_j noktasındaki fonksiyon kestirimi yerel ağırlık fonksiyonu w_j kullanılarak,

$$\chi(t_j) = \sum_j^n w_j y_j, \quad (2.48)$$

şeklinde yazılabilir. Burada, w_j ile gösterilen yerel ağırlıklar (2.30) eşitliğindeki gibi ifade edilen bir çekirdek fonksiyonundan elde edilir.

2.4.2.1 Yerel Taban Fonksiyonları ile Düzleştirme

EKK kriteri hataların yerel bir ölçüsüne dayanarak genişletilebilir. Yerel En Küçük Kareler (Localized Least Squares, LLS) adı verilen bu yöntemin temeli taban fonksiyonu kestirimi ve çekirdek fonksiyonu kestirimi yöntemlerinin birleştirilmesi fikrine dayalıdır.

Bu yönteme ait yerel taban fonksiyonları kestiricisi, $w_j(t)$ (2.30)'daki gibi bir çekirdek fonksiyonundan elde edilen ağırlık değerleri olmak üzere,

$$SSE_{LLS} = \sum_{j=1}^n w_j(t) \left[y_j - \sum_{k=1}^K c_k \phi_k(t_j) \right]^2, \quad (2.49)$$

ile verilen hata kareler ortalamasının en küçüklenmesinden elde edilir.

Bu ifade, matris formunda şu şekilde yazılabilir:

$$SSE_{LLS} = (\mathbf{y} - \boldsymbol{\phi}\mathbf{c})' \mathbf{W} (\mathbf{y} - \boldsymbol{\phi}\mathbf{c}). \quad (2.50)$$

YEKK kestirimleri,

$$\hat{\mathbf{c}} = (\boldsymbol{\phi}' \mathbf{W} \boldsymbol{\phi})^{-1} \boldsymbol{\phi}' \mathbf{W} \mathbf{y}, \quad (2.51)$$

denkleminde bulunur.

2.4.2.2 Yerel Çoklu Terimli Düzleştirme

Düzleştirme için önerilen bir başka yöntem de yerel çoklu terimli düzgünleştirme (Localized Polynomial Smoothing, LPS) yöntemidir. Bu yöntemde, kestirilen $\hat{\chi}_i(t)$ eğri değerleri

$$SSE_{LPS} = \sum_{j=1}^n \text{Kern}_h(t_j, t) \left[y_j - \sum_{l=0}^L c_l (t - t_j)^l \right]^2, \quad (2.52)$$

formülünün minimizasyonundan elde edilir. Genelde L değeri en az bire eşit olarak alınır. Türevlerin kullanılması durumunda, l parametresi en yüksek dereceli türevin derecesinin iki fazlası olarak alınmalıdır (Ramsay ve Silverman, 2005).

2.4.3 Pürüzlü Ceza Yaklaşımı

Düzleştirmede başka güncel bir yöntem de Pürüzlü Ceza Yaklaşımı (PCY) yöntemidir. Bu yöntem, fonksiyon kompleksliğinin bir ölçüsünü cezalandırırken fazla sayıda taban fonksiyonunun kullanımını mümkün kılar (Ramsay ve diğ., 2009). Bir χ fonksiyonun pürüzlülüğünün popüler bir ölçüsü eğrilik olarak da adlandırılan fonksiyonun ikinci türevinin karesidir. Bu ceza ölçüsü $\text{PEN}_2(\chi)$ ile gösterilir ve

$$\text{PEN}_2(\chi) = \int [D^2\chi(t)]^2 dt, \quad (2.53)$$

olarak tanımlanır.

Pürüzlü cezanın seçimi veriye ve düzleştirmenin amacına bağlıdır. Eğer amaç ikinci türevlerin düzleştirilmesi ise, ikinci türevin eğikliğinin cezalandırılması iyi sonuç verir.

$$\text{PEN}_4(\chi) = \int [D^4\chi(t)]^2 dt. \quad (2.54)$$

Periyodik bir veri söz konusu ise, (2.55) ile ifade edilen

$$L\chi = D^3\chi + \omega^2 d\chi, \quad (2.55)$$

harmonik ivme operatörünün cezalandırılması pürüzlüğün bir ölçüsü olarak kullanılabilir.

$$\text{PEN}_L(\chi) = \int [(L\chi)^2](t) dt = \|L\chi\|^2. \quad (2.56)$$

Pürüzlü cezaların genel yapısı, \mathbf{R} ceza matrisi,

$$\mathbf{R} = \int D^m \phi(\mathbf{t}) D^m \phi(\mathbf{t})' dt, \quad (2.57)$$

olmak üzere (2.58)'deki gibi matris biçiminde verilebilir:

$$\begin{aligned}
\text{PEN}_m(\chi) &= \int [D^m \chi(t)]^2 dt = \int [D^m \mathbf{c} \boldsymbol{\phi}(t)]^2 dt \\
&= \int \mathbf{c}' D^m \boldsymbol{\phi}(t) D^m \boldsymbol{\phi}(t)' \mathbf{c} dt = \mathbf{c}' \left[\int D^m \boldsymbol{\phi}(t) D^m \boldsymbol{\phi}'(t) dt \right] \mathbf{c} \\
&= \mathbf{c}' \mathbf{R} \mathbf{c}
\end{aligned} \tag{2.58}$$

PCY'ye ait hata kareler toplamı, $\text{PEN}_m(\chi)$ 'nin düzgünleştirme parametresi λ ile çarpımının, ağırlıklı hata kareler toplamına eklenmesinden elde edilir.

$$\text{PENSSE} = (\mathbf{y} - \boldsymbol{\phi} \mathbf{c})' \mathbf{W} (\mathbf{y} - \boldsymbol{\phi} \mathbf{c}) + \lambda \mathbf{c}' \mathbf{R} \mathbf{c}. \tag{2.59}$$

(2.59) eşitliğinin türevi alınırsa,

$$-2 \boldsymbol{\phi}' \mathbf{W} \mathbf{y} + \boldsymbol{\phi}' \mathbf{W} \boldsymbol{\phi} \mathbf{c} + \lambda \mathbf{R} \mathbf{c} = 0, \tag{2.60}$$

olduğu görülür.

Katsayı vektörünün kestirimi bu denklemden,

$$\hat{\mathbf{c}} = (\boldsymbol{\phi}' \mathbf{W} \boldsymbol{\phi} + \lambda \mathbf{R})^{-1} \boldsymbol{\phi}' \mathbf{W} \mathbf{y}, \tag{2.61}$$

olarak elde edilir.

Buradan düzgünleştirme matrisi,

$$\mathbf{S}_{\boldsymbol{\phi}, \lambda} = \boldsymbol{\phi} (\boldsymbol{\phi}' \mathbf{W} \boldsymbol{\phi} + \lambda \mathbf{R})^{-1} \boldsymbol{\phi}' \mathbf{W}, \tag{2.62}$$

biçimini alır. $\lambda = 0$ olması durumunda, düzgünleştirme matrisi (2.46) ile verilen AEKK yöntemindeki düzgünleştirme matrisine eşittir.

EKK kestiriminin serbestlik derecesi (Degrees of freedom, Df) düzgünleştirme matrisinin izine eşittir. Bu da, katsayı vektörünün içerdiği parametre sayısı K 'ya eşittir (Ramsay ve Silverman, 2005).

$$Df = \text{iz}(\mathbf{S}) = K. \tag{2.63}$$

2.4.3.1 Düzleştirme Parametresinin Seçimi

Düzleştirme parametresinin değerine karar verilmesinde en sık kullanılan yöntemler Çapraz-Geçerlilik (Cross-Validation, CV) ve Genelleştirilmiş Çapraz Geçerlilik (Generalized Cross Validation, GCV) kriterleridir.

CV kriteri, $\hat{y}_{j(-j)}$ j 'inci gözlem çifti (t_j, y_j) dışarıda bırakıldığında elde edilen kestirim değerini ve ω_j, t_j noktasındaki ağırlık değerini ifade etmek üzere,

$$CV(\mathbf{v}) = \frac{1}{n} \sum_j^n (y_j - \hat{y}_{j(-j)})^2 \omega_j, \quad (2.64)$$

biçiminde tanımlanır. Sıklıkla kullanılmasına rağmen bu yöntemin yoğun hesap gerektirmesi ve veriyi az düzeltmesi gibi bazı problemleri vardır (Ramsay ve Silverman, 2005). Bu nedenle veriyi daha fazla düzeltme eğilimi olan GCV yönteminin kullanılması önerilir. GCV kriteri, (2.65) eşitliğinden hesaplanır:

$$GCV(\mathbf{v}) = \frac{1}{n} \sum_j^n (y_j - \hat{y}_{j(-j)})^2 \omega_j \Xi(\mathbf{v}). \quad (2.65)$$

Burada, $\Xi(\mathbf{v})$ cezalandırma fonksiyonunu, \mathbf{v} taban sayısı K 'ya bağlı bir parametreyi ve λ ceza parametresini temsil eder.

GCV kriteri için $\Xi(\mathbf{v})$,

$$\Xi(\mathbf{v}) = (1 - Tr(\mathbf{S}) n^{-1})^{-2}, \quad (2.66)$$

olarak verilir.

Bu fonksiyondaki değişime bağlı olarak, düzgünleştirme parametresinin seçiminde Akaike Bilgi Kriteri (Akaike Information Criterion, AIC), Sonlu Hata Tahmini (Finite Prediction Error, FPE), Rice bant genişliği seçici (Rice Bandwidth Selector) ve Shibata'nın model seçimi (Shibata's Model Selector) kriterleri gibi farklı yöntemler de kullanılabilir. Bu kriterler için $\Xi(\mathbf{v})$ fonksiyonun aldığı değerler Febrero-Bande ve Oviedo de la Fuente (2012)'nin çalışmasında ayrıntılı olarak açıklanmıştır.

3 FONKSİYONEL TEMEL BİLEŞENLER ANALİZİ

Temel bileşenler analizi, değişkenler arasındaki korelasyon ve kovaryans değerlerini dikkate alarak, rassal değişkenlerin doğrusal birleşiminden oluşan bağımsız (ortogonal) yeni değişkenlerin ortaya çıkarılmasına ve boyut indirgemeye yarayan önemli bir çok değişkenli tekniktir.

Fonksiyonel Temel Bileşenler Analizi (FTBA), Temel Bileşenler Analizinin (TBA) bir uzanımıdır. Bu iki yöntem arasındaki en temel fark FTBA'nın vektörler yerine eğrilerle uğraşmasıdır. Bundan dolayı klasik TBA'da kullanılan kovaryans matrisi ve özvektörler yerini sırasıyla doğrusal operatör ve öz fonksiyonlara bırakır. Temel bileşenler artık \mathbb{R}^p uzayında rassal değişkenler değil, L^2 uzayının elemanlarıdır. FTBA yönteminin temeli, stokastik bir süreçten gelen rassal değişkenleri ifade etmek için kullanılan Karhunen-Loève ayrışımına dayanır. Bu açılım hakkında ayrıntılı bilgi Ek 2.8'te mevcuttur. FTBA ve TBA arasındaki temel farklar Çizelge 3.1'deki gibi özetlenebilir.

Çizelge 3.1: TBA ve FTBA yöntemlerinin karşılaştırılması

	TBA	FTBA
Veri	\mathbb{R}^p uzayında vektör	$L^2(T)$, $T = [t_1, t_n]$ uzayında fonksiyon
Değişkenler	$X = [X_1, X_2, \dots, X_p]$ $X_j = [x_{1j}, \dots, x_{nj}]$, $j = 1, \dots, p$	$\chi(t) = [\chi_1(t), \dots, \chi_N(t)]$ $t \in T = [t_1, t_n]$
Kovaryans	V matrisi $\mathbf{V} = \text{Cov}(X) \in \mathbb{R}$	t_1 ve t_n arasında sınırlı Γ_χ operatörü $\Gamma_\chi : L^2(T) \rightarrow L^2(T)$
Özyapı	$u_j \in \mathbb{R}^p$ vektörü $\mathbf{V}u_j = \lambda_j u_j$	$\xi_j \in L^2(T)$, $t \in T$ fonksiyonu $\int_T c(s, t) \xi_j(s) ds = \lambda_j \xi_j(t)$
Bileşenler	\mathbb{R}^p 'de rassal değişkenler	$L^2(T)$ 'de rassal fonksiyonlar

Burada ifade edilen tüm bu farklılıklara rağmen, iki yöntem de verideki değişkenliği en iyi açıklayan anlamlı faktörlerin elde edilmesini amacını taşır.

3.1 Temel Bileşenler Analizi Yaklaşımı

Çok değişkenli durumda, temel bileşenler verideki en önemli değişimleri açığa çıkaracak şekilde seçilen $\mathbf{u}_j = (u_{1j}, u_{2j}, \dots, u_{pj})'$ yükleri ile j 'inci değişkene ait merkezleştirilmiş gözlemlerin (x_{ij}) doğrusal birleşimi olarak ifade edilirler. Bu doğrusal birleşim, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$ bir vektör ve \mathbf{u}_j , j 'inci bileşene ait ağırlık vektörü olmak üzere,

$$f_{ik} = \sum_{j=1}^p \mathbf{u}_{kj} x_{ij} = \mathbf{u}'_j \mathbf{x}_i, \quad i = 1, \dots, N, \quad j = 1, \dots, p, \quad k = 1, \dots, p, \quad k \leq p, \quad (3.1)$$

şeklinde yazılır.

Birinci bileşene ait ağırlık vektörü $\mathbf{u}_1 = (u_{11}, \dots, u_{p1})'$,

$$\|\mathbf{u}_1\|^2 = \mathbf{u}'_1 \mathbf{u}_1 = 1, \quad (3.2)$$

koşulu altında birinci bileşenin varyansı f_{i1} 'nin,

$$f_{i1} = \sum_{j=1}^p \mathbf{u}_{j1} x_{ij} = \mathbf{u}'_1 \mathbf{x}_i, \quad (3.3)$$

en büyüklenmesinden elde edilir.

İkinci bileşene ait ağırlık vektörü \mathbf{u}_2 , ikinci bileşenin varyansı f_{i2} 'nin $\|\mathbf{u}_2\|^2 = \mathbf{u}'_2 \mathbf{u}_2 = 1$ koşuluna ek olarak bileşenlerin dikliğini, bir başka deyişle bağımsızlığını ifade eden,

$$\mathbf{u}'_2 \mathbf{u}_1 = 0, \quad (3.4)$$

koşulu altında en büyüklenmesinden elde edilir.

Genel anlamda m 'inci bileşene ait ağırlık vektörü, f_{im} 'nin varyansı

$$\|\mathbf{u}_m\|^2 = \mathbf{u}'_m \mathbf{u}_m = 1 \quad (3.5)$$

ve

$$\mathbf{u}'_l \mathbf{u}_m = 0, \quad l < m \quad (3.6)$$

koşulları altında en büyüklenecek şekilde bu adımların tekrarından bulunur.

Bu şekilde, bileşenlerin varyansı her adımda azalacak şekilde verideki en önemli değişim noktaları belirlenir.

\mathbf{X} ; merkezleştirilmiş x_{ij} gözlemlerinden oluşan $N \times p$ boyutlu bir matris ve \mathbf{u} ; p uzunluğunda bir ağırlık vektörü olmak üzere f_{ik} temel bileşen skor değerleri,

$$\max \quad N^{-1} \mathbf{u}' \mathbf{X}' \mathbf{X} \mathbf{u}, \quad \mathbf{u}' \mathbf{u} = 1, \quad (3.7)$$

şeklinde matris formunda verilen maksimizasyon kriterinden elde edilir.

(3.8)'deki gibi tanımlanan varyans-kovaryans matrisi,

$$\mathbf{V} = N^{-1} \mathbf{X}' \mathbf{X}, \quad (3.8)$$

göz önüne alınarak (3.7) eşitliği yeniden yazılabilir:

$$\max \quad \mathbf{u}' \mathbf{V} \mathbf{u}, \quad \mathbf{u}' \mathbf{u} = 1. \quad (3.9)$$

Bu maksimizasyon probleminin çözümü,

$$\mathbf{V} \mathbf{u}_j = \lambda_j \mathbf{u}_j, \quad (3.10)$$

şeklinde tanımlanan öz denklem probleminin çözümünden elde edilir. Burada, λ_j özdeğerleri ve \mathbf{u}_j varyans-kovaryans matrisinin özvektörlerini ifade eder. Her j değeri için, özvektörler sırasıyla (3.5) ve (3.6) ile verilen koşulları sağlar.

Merkezileştirmeden dolayı, merkezleştirilmiş veri matrisinin rankı $N - 1$ 'e eşittir ve $p \times p$ boyutlu simetrik \mathbf{V} matrisinin en çok $\min(p, N - 1)$ kadar sıfıra eşit olmayan λ_j özdeğeri vardır. (3.10) koşulunu sağlayan farklı özdeğer-özvektör çiftleri $(\lambda_j, \mathbf{u}_j)$ bulunabilir.

\mathbf{V} 'nin $p \times p$ boyutlu simetrik bir matris olduğu düşünülecek olursa,

$$\max \quad \mathbf{x}' \mathbf{V} \mathbf{x}, \quad \mathbf{x}' \mathbf{x} = 1. \quad (3.11)$$

biçiminde verilen herhangi bir maksimizasyon problemi kuadratik biçim kullanılarak çözülebilir.

$\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$, \mathbf{V} matrisinin azalan sırada verilen pozitif özdeğerleri ve \mathbf{u}_j \mathbf{V} matrisinin ortonormallik koşulunu sağlayan ilgili özvektörleri olsun. \mathbf{D} köşegen elemanları λ_j , ($j = 1, \dots, p$) olan diyagonal bir matris ve \mathbf{U} , $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_p]$ ile verilen, sütunlarını \mathbf{V} 'nin özvektörlerinin oluşturduğu ortonormal bir matris olsun. \mathbf{U} ortonormal matrisinin $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p$ elemanları \mathbb{R}^p uzayında ortonormal bir taban oluşturur.

Tekil Değer Ayrışımı (Singular Value Decomposition, SVD) teoremine göre, \mathbf{V} ,

$$\mathbf{V} = \mathbf{U}\mathbf{D}\mathbf{U}', \quad (3.12)$$

olarak yazılabilir. Ortonormallikten ötürü, $\mathbf{U}' = \mathbf{U}^{-1}$ ve

$$\mathbf{U}\mathbf{U}' = \mathbf{U}'\mathbf{U} = \mathbf{I}, \quad (3.13)$$

eşitlikleri sağlanır.

Birim uzunlukta bir \mathbf{x} vektörünü elde etmek yerine, $\mathbf{y} = \mathbf{U}'\mathbf{x}$ şeklinde tanımlanan ve $\mathbf{x} = \mathbf{U}\mathbf{y}$ 'yi sağlayan \mathbf{y} vektörünü göz önüne alalım. $\|\mathbf{u}_j\| = 1$ olduğundan, $\mathbf{x}'\mathbf{x}$ ve $\mathbf{y}'\mathbf{y}$ ifadeleri denktir (Horváth ve Kokoszka, 2012).

$$\mathbf{x}'\mathbf{x} = \mathbf{y}'\mathbf{U}'\mathbf{U}\mathbf{y} = \mathbf{y}'\mathbf{y}. \quad (3.14)$$

Varyans kovaryans matrisi \mathbf{V} 'nin tekil değer ayrışımı (3.12) eşitliğinde kullanılır ve $\mathbf{x} = \mathbf{U}\mathbf{y}$ eşitliği (3.11)'ye yerleştirilirse, maksimizasyon problemi,

$$\max \mathbf{y}'\mathbf{D}\mathbf{y}, \quad \mathbf{y}'\mathbf{y} = 1, \quad (3.15)$$

biçimini alır. Bu problem, \mathbf{y} birim vektöre eşitlenerek, $\mathbf{y} = [1, 0, \dots, 0]'$, çözülebilir. Böylelikle, \mathbf{U} matrisinin ilk sütunu \mathbf{x} 'e eşit olarak bulunur, yani λ_1 en büyük özdeğeri olmak üzere $\mathbf{x} = \mathbf{u}_1$ 'dir.

Bu adımlar,

$$\mathbf{x}'\mathbf{u}_1 = \mathbf{x}'\mathbf{u}_2 = \dots = \mathbf{x}'\mathbf{u}_{m-1} = 0,$$

ek koşulları altında, \mathbf{V} matrisine ait tüm özvektörlerin bulunması için uzatılabilir.

m 'inci adımda, $\mathbf{x}'\mathbf{V}\mathbf{x} = \lambda_m \mathbf{u}_m' \mathbf{u}_m = \lambda_m$ m 'inci özdeğeri ifade etmek üzere, $\mathbf{x} = \mathbf{u}_m$ olarak bulunur. Böylelikle, (3.9) denklemindeki \mathbf{u}_j ağırlık vektörlerinin varyans kovaryans matrisinin ilgili özdeğerine ilişkin özvektörlerine eşit olduğu kanıtlanmış olur.

3.2 Fonksiyonel Veri İçin Temel Bileşenler Analizi

FTBA'da veri sürekli fonksiyonlardan oluşur. Çok değişkenli durumdaki \mathbf{u}_j ağırlık vektörü ve \mathbf{x}_i merkezleştirilmiş gözlem vektörü artık fonksiyondurlar ve sırasıyla $\xi_j(t)$ ve $\chi^c(t)$ ile gösterilirler. Bu nedenle, vektörlerin doğrusal birleşimini ifade etmek için kullanılan toplam yerine artık integral kullanılır. Bu durumda, FTBA skorları, q boyutlu fonksiyonel uzayda $\hat{\xi}_j(t)$ ampirik ağırlık fonksiyonları ile $\chi^c(t)$ merkezleştirilmiş gözlem eğrilerinin iç çarpımı olarak hesaplanır.

Fonksiyonel temel bileşenlerin skorları (f_{ij}), her bir gözlem eğrisinden ortalama fonksiyonunun çıkartılması ile elde edilen $\chi_i^c(t)$ merkezleştirilmiş gözlem eğrileri kullanılarak,

$$f_{ij} = \int_T \hat{\xi}_j(t) \chi_i^c(t) dt, \quad i = 1, \dots, N, \quad j = 1, \dots, q, \quad (3.16)$$

olarak hesaplanır.

FTBA'nın adımları işlemlerin fonksiyonlar üzerinde olması farkı dışında, TBA'nın adımlarına benzerlik gösterir. İlk adımda,

$$\int_T \hat{\xi}_1^2(t) dt = \|\hat{\xi}_1\|^2 = 1, \quad (3.17)$$

koşulu altında temel bileşen skorlarının varyansı,

$$f_{i1} = \int_T \hat{\xi}_1(t) \chi_i^c(t) dt, \quad (3.18)$$

en büyüklükte birinci fonksiyonel temel bileşene ait ağırlık fonksiyonu $\hat{\xi}_1(t)$ bulunur.

İleriki adımlarda, $\|\hat{\xi}_m\|^2 = 1$ ve $m - 1$ adet ek diklik kısıtları altında,

$$\int_T \hat{\xi}_l(t) \hat{\xi}_m(t) dt = 0, \quad l < m, \quad (3.19)$$

fonksiyonel temel bileşenlerin varyansını en büyükleyen ağırlık fonksiyonları hesaplanır.

Fonksiyonel içerikte \mathbf{V} ile ifade edilen varyans kovaryans matrisinin yerini ampirik kovaryans operatörü $\hat{\Gamma}_\chi$ alır. $\chi_i^c(t)$ merkezleştirilmiş eğrilerinin, $L^2[T]$ Hilbert uzayının elemanları olduğunu ve her $t \in T = [t_1, t_n]$ için $\mathbb{E}\|\chi\|^2 = \mathbb{E}[\int \chi(t)^2 dt] < \infty$ ifadesinin gerçekleştiğini varsayalım.

Bu durumda, $\hat{\xi} \in H$ olmak üzere fonksiyonel veri için maksimizasyon probleminin karesel biçimi şu şekilde tanımlanır:

$$\max \langle \hat{\Gamma}_\chi \hat{\xi}, \hat{\xi} \rangle, \quad \|\hat{\xi}\| = 1. \quad (3.20)$$

Ek 2.6'de açıklanan Riesz temsili kullanılarak bu maksimizasyon problemi şu şekilde yeniden yazılabilir:

$$\langle \hat{\Gamma}_\chi \hat{\xi}, \hat{\xi} \rangle = \mathbb{E}[\langle \chi_i^c(t), \hat{\xi}(s) \chi_i^c(s) \rangle, \hat{\xi}(t)] = \mathbb{E}[\langle \chi_i^c, \hat{\xi} \rangle^2], \quad \hat{\xi} \in L^2[T]. \quad (3.21)$$

Öte yandan, fonksiyonel temel bileşenler analizinin amacı boyut indirgemek ve veriyi hesap kolaylığı sağlaması açısından q boyutlu uzayda temsil etmektir. $\hat{\xi}_1, \hat{\xi}_2, \dots, \hat{\xi}_q$ ortonormal fonksiyonların oluşturduğu ortonormal bir taban olsun.

$$\text{SSE} = \sum_{i=1}^N \|\chi_i^c - \hat{\chi}_i^c\|^2, \quad (3.22)$$

ile ifade edilen hata kareler toplamını (SSE) minimize edecek böyle bir taban bulunuyorsa, $q < N$ için herbir $\hat{\chi}_i^c$ eğrisi,

$$\hat{\chi}_i^c = \sum_{j=1}^q \langle \chi_i^c, \hat{\xi}_j \rangle \hat{\xi}_j, \quad (3.23)$$

olarak temsil edilebilir. $\langle \chi_i^c, \hat{\xi}_j \rangle$ iç çarpımı, (3.16) denklemi ile belirtilen χ_i 'nin j 'inci skoruna eşittir.

$\|\hat{\xi}_j\|^2 = 1$ ortonormallik koşulu altında (3.22) eşitliği tekrar şu şekilde ifade edilebilir:

$$\begin{aligned}
\sum_{i=1}^N \|\chi_i^c - \langle \chi_i^c, \hat{\xi}_j \rangle \hat{\xi}_j\|^2 &= \sum_{i=1}^N \|\chi_i^c\|^2 - 2 \sum_{i=1}^N \sum_{j=1}^q \langle \chi_i^c, \hat{\xi}_j \rangle^2 + \sum_{i=1}^N \sum_{j=1}^q \langle \chi_i^c, \hat{\xi}_j \rangle^2 \|\hat{\xi}_j\|^2 \\
&= \sum_{i=1}^N \|\chi_i^c\|^2 - \sum_{i=1}^N \sum_{j=1}^q \langle \chi_i^c, \hat{\xi}_j \rangle^2.
\end{aligned} \tag{3.24}$$

Buradan da görüldüğü gibi, SSE'yi minimize etmek $\sum_{i=1}^N \sum_{j=1}^q \langle \chi_i^c, \hat{\xi}_j \rangle^2$ ifadesini, yani $\langle \hat{\Gamma}_\chi \hat{\xi}, \hat{\xi} \rangle$ iç çarpımını, maksimize etmeye denktir. Bu maksimizasyon probleminin sonuçları aşağıda verilen teoremden elde edilir.

Teorem (Horváth ve Kokoszka, 2012): Ψ , v_j özfonksiyonları ve $\lambda_1 > \lambda_2 > \dots > \lambda_p > \lambda_{p+1}$ özdeğerlerine sahip simetrik, pozitif tanımlı bir Hilbert-Schmidt operatörü olsun. O zaman,

$$\sup \{ \langle \Psi \xi, \xi \rangle : \|\xi\| = 1, \langle \xi, v_j \rangle = 0, \quad 1 \leq j \leq i-1, i < p. \} = \lambda_i \tag{3.25}$$

gerçeklenir ve $\xi = v_i$ noktasında supremuma ulaşılır.

\hat{v}_j 'ler örneklem kovaryans operatörü $\hat{\Gamma}_\chi$ 'nin ortonormal özfonksiyonları olsun. O halde, Ek 2.4'de verilen Spektral Ayrışım Teoremi'nden, maksimizasyon problemi, $\hat{\lambda}_1 > \hat{\lambda}_2 > \dots$ azalan sırasında verilen $\hat{\lambda}_j$ ampirik özdeğerleri göstermek üzere şu şekilde açılabilir:

$$\begin{aligned}
\sum_{i=1}^N \sum_{j=1}^q \langle \chi_i^c, \hat{\xi}_j \rangle^2 &= \sum_{j=1}^q \langle \hat{\Gamma}_\chi \hat{\xi}_j, \hat{\xi}_j \rangle \\
&= \sum_{j=1}^{\infty} \hat{\lambda}_j \langle \hat{\xi}_1, \hat{v}_j \rangle^2 + \sum_{j=1}^{\infty} \hat{\lambda}_j \langle \hat{\xi}_2, \hat{v}_j \rangle^2 + \dots + \sum_{j=1}^{\infty} \hat{\lambda}_j \langle \hat{\xi}_q, \hat{v}_j \rangle^2.
\end{aligned} \tag{3.26}$$

Yukarıda verilen teoremden $\hat{\xi}_1 = \hat{v}_1, \hat{\xi}_2 = \hat{v}_2, \dots, \hat{\xi}_q = \hat{v}_q$ olduğu açıkça görülmektedir. Sonuç olarak, $\hat{\xi}_j$ ağırlık fonksiyonları aslında örneklem kovaryans operatörünün ortonormal özfonksiyonlarıdır ve ampirik fonksiyonel temel bileşenler (Empirical Functional Principal Component, EFPC) olarak adlandırılırlar. χ_i^c eğrisinin j 'inci skor değeri, $\hat{\xi}_j$ fonksiyonel temel bileşeninin χ_i^c eğrisine katkısının ağırlığı olarak yorumlanır.

Kovaryans operatörünün ampirik özfonksiyonları $\hat{\xi}_j$, (3.27) öz denkleminin çözümleridir:

$$\hat{\Gamma}_\chi \hat{\xi}_j = \hat{\lambda}_j \hat{\xi}_j, \quad j \geq 1. \quad (3.27)$$

Ampirik kovaryans operatörü, $\hat{c}(s,t)$ çekirdeği ile integral dönüşüm olarak ifade edilebileceğinden bu denklem,

$$\int_T \hat{c}(t,s) \hat{\xi}_j(s) ds = \hat{\lambda}_j \hat{\xi}_j(t), \quad j \geq 1, \quad (3.28)$$

biçiminde de yazılabilir.

Fonksiyonel temel bileşenleri ifade etmek için farklı yöntemler vardır. En sık kullanılan yöntem taban temsili yöntemidir.

Her gözlenen fonksiyonunun, geçen bölümde bahsedildiği gibi (2.21) ya da (2.22) şeklinde bir taban açılımı olduğu göz önüne alınırsa, özfonksiyonlar taban fonksiyonları cinsinden,

$$\hat{\xi}(t) = \sum_{k=1}^K b_k \phi_k(t), \quad (3.29)$$

ya da denk olarak matris biçiminde,

$$\hat{\xi}(t) = \boldsymbol{\phi}(t)' \mathbf{b}, \quad (3.30)$$

olarak ifade edilebilir.

Fonksiyonel verinin matris formundaki taban açılımı kullanılarak,

$$\hat{\chi} = \mathbf{C} \boldsymbol{\phi}, \quad (3.31)$$

ampirik varyans kovaryans fonksiyonu (3.32)'deki gibi matris biçiminde yazılabilir:

$$\hat{c}(t,s) = \frac{1}{N-1} \boldsymbol{\phi}(t)' \mathbf{C}' \mathbf{C} \boldsymbol{\phi}(s). \quad (3.32)$$

K sıralı simetrik bir \mathbf{W} matrisi tanımlanırsa,

$$\mathbf{W} = \int \boldsymbol{\phi}(s) \boldsymbol{\phi}(s)' ds, \quad (3.33)$$

özdenklem,

$$\begin{aligned}\int_T \hat{c}(t,s)\hat{\xi}(s)ds &= \int \frac{1}{N-1} \boldsymbol{\phi}(t)' \mathbf{C}' \mathbf{C} \boldsymbol{\phi}(s) \boldsymbol{\phi}(s)' \mathbf{b} ds \\ &= \frac{1}{N-1} \boldsymbol{\phi}(t)' \mathbf{C}' \mathbf{C} \mathbf{W} \mathbf{b}.\end{aligned}\quad (3.34)$$

ifadesine denktir.

Ağırlık fonksiyonları (3.30)'da verildiği gibi taban fonksiyonları cinsinden yazılabildiğinden özdenklem problemi,

$$\frac{1}{N-1} \boldsymbol{\phi}(t)' \mathbf{C}' \mathbf{C} \mathbf{W} \mathbf{b} = \hat{\lambda} \boldsymbol{\phi}(t)' \mathbf{b}, \quad (3.35)$$

şeklini alır. Bu eşitlik tüm $t \in T$ değerleri için geçerlidir. Bu durumda, $\boldsymbol{\phi}(t)'$ denklemden çıkarılarak şu şekilde yazılabilir:

$$\frac{1}{N-1} \mathbf{C}' \mathbf{C} \mathbf{W} \mathbf{b} = \hat{\lambda} \mathbf{b}. \quad (3.36)$$

Burada, $\|\hat{\xi}\| = 1$ kısıtı $\mathbf{b}' \mathbf{W} \mathbf{b} = 1$ olma durumunu ifade eder. İki fonksiyonun ortogonalliğini ifade eden $\hat{\xi}_1$ and $\hat{\xi}_2$ ek kısıtı ise, $\mathbf{b}_1' \mathbf{W} \mathbf{b}_2 = 0$ olarak ifade edilebilir.

Temel bileşenleri elde etmek için bir $\mathbf{u} = \mathbf{W}^{1/2} \mathbf{b}$ özvektörü tanımlanır ve (3.36) denkleminde yerleştirilirse,

$$\frac{1}{N-1} \mathbf{W}^{1/2} \mathbf{C}' \mathbf{C} \mathbf{W}^{1/2} \mathbf{u} = \hat{\lambda} \mathbf{u}, \quad (3.37)$$

olur ve her vektör için $\mathbf{b} = \mathbf{W}^{-1/2} \mathbf{u}$ hesaplanır.

Ortonormal taban açılımı için ağırlık matrisi birim matrise denktir. Yani, $\mathbf{W} = \mathbf{I}$ 'dir. Bu durumda FTBA,

$$\frac{1}{N-1} \mathbf{C}' \mathbf{C} \mathbf{u} = \hat{\lambda} \mathbf{u}, \quad (3.38)$$

özdenklemini ile klasik TBA yöntemine indirgenmiş olur.

Fonksiyonel temel bileşenleri hesaplamak için ikinci bir yol da veri güdümlü taban kullanılmasıdır (Ramsay ve Silverman, 2005). Bu durumda öz denklem probleminin sol yanında yer alan integral, kesikli değerlerin toplamı olarak ifade edilmelidir. Bir f

fonksiyonu için bu yaklaşım, t_j argüman değerlerini, n argüman değerlerinin sayısını ve w_j toplamdaki her fonksiyon değerine ait ağırlıkları göstermek üzere şu şekilde yazılabilir:

$$\int f(t)dt \approx \sum_{j=1}^n w_j f(t_j). \quad (3.39)$$

Eğer (3.39) denklemindeki gibi bir yaklaşım uygulanırsa kovaryans operatörü, \mathbf{V} , t_j ve t_k argüman noktalarındaki $\hat{c}(t_j, t_k)$ kovaryans değerlerinden oluşan bir matrisi, $\tilde{\xi}$, $\tilde{\xi}(t_j)$ değerlerini içeren n 'inci dereceden bir vektörü ve \mathbf{W} köşegen elemanları w_j ağırlık değerlerinden oluşan köşegen bir matrisi göstermek üzere,

$$\hat{\Gamma}_x \hat{\xi} \approx \mathbf{V} \mathbf{W} \tilde{\xi}, \quad (3.40)$$

biçimini alır.

Öz analiz probleminin matris biçimi,

$$\mathbf{V} \mathbf{W} \tilde{\xi} = \lambda \tilde{\xi}, \quad (3.41)$$

olarak yazılabilir.

Bu durumda ortonormallik ve diklik kısıtları ağırlık matrisi kullanılarak (3.42) denklemindeki gibi belirtilir:

$$\tilde{\xi}_m' \mathbf{W} \tilde{\xi}_m = 1 \quad \text{ve} \quad \tilde{\xi}_m' \mathbf{W} \tilde{\xi}_k = 0, \quad k < m. \quad (3.42)$$

(3.37) eşitliğine benzer şekilde (3.41) eşitliği $\mathbf{u} = \mathbf{W}^{1/2} \tilde{\xi}$ kullanılarak ortonormal vektör \mathbf{u} cinsinden (3.43)'teki gibi tekrar yazılabilir.

$$\mathbf{W}^{1/2} \mathbf{V} \mathbf{W}^{1/2} \mathbf{u} = \hat{\lambda} \mathbf{u}. \quad (3.43)$$

Bu problemin çözümü nümerik iterasyonlar gerektirir ve dört adımdan oluşur:

1. n, w_j ve t_j değerlerinin belirlenmesi,
2. $\hat{\lambda}_j$ özdeğerlerinin hesabı ve $\mathbf{W}^{1/2} \mathbf{V} \mathbf{W}^{1/2}$ matrisinin \mathbf{u}_j özvektörlerinin hesabı,

3. $\tilde{\xi}_m = \mathbf{W}^{1/2} \mathbf{u}_m$ vektörünün hesabı,
4. $\tilde{\xi}_m$ vektörünün kesikli değerlerinin, $\hat{\xi}_m$ gibi bir fonksiyona dönüştürülmesi için interpolasyon yönteminin kullanılması.

Bu sürecin sonunda, n argüman sayısı kadar özfonksiyon elde edilir.

Fonksiyonel temel bileşen sayısı q 'nün belirlenmesi önemli bir adımdır ve bu sayının belirlenmesi için çeşitli yöntemler önerilmiştir. Bu yöntemlerin en popüler olanı "scree plot" adı verilen $\hat{\lambda}_j$ özdeğerlerinin j değerlerine karşı grafik çizimidir. Grafikte özdeğerlerin ani düşüşe geçtiği noktadaki q bileşen sayısı olarak belirlenir (Horváth ve Kokoszka, 2012).

Bileşen sayısına karar vermekte kullanılan bir başka yöntem ise toplam varyansın birikimli yüzdesidir (Cumulative Percentage of total Variance, CPV). CPV kriteri şu şekilde tanımlanır:

$$\text{CPV}(q) = \frac{\sum_{k=1}^q \hat{\lambda}_k}{\sum_{k=1}^N \hat{\lambda}_k}. \quad (3.44)$$

Horváth ve Kokoszka (2012), CPV kriterinin 85% değerini geçtiği noktadaki kadar temel bileşen sayısının alınması önerir. Bu yöntemlere alternatif olarak yapay-AIC (pseudo-AIC) veya çapraz geçerlilik yöntemleri de temel bileşen sayısına karar verilmesi amacıyla kullanılabilir.

FTBA'da bir başka önemli husus da sonuçların görselleştirilmesidir. Bu adım elde edilen bileşenlerin yorumlanabilmesi için gerekli ve önemli bir adımdır. Sonuçları görselleştirmek için fonksiyonel veriye ait ortalama fonksiyonu hesaplanır ve her bileşen eğrisinin uygun bir çarpanı sırayla hesaplanan fonksiyonel ortalamaya eklenir ve ortalamadan çıkarılır. Her bir bileşen için ortalama fonksiyonu ile bu fonksiyona ilgili bileşenin belirli bir katının eklenmesi ve çıkarılmasından elde edilen etkiler grafik olarak çizilir. Burada, temel bileşen fonksiyonunun çarpanının seçimi önemlidir. Bu çarpan sabit bir değerdir ve C ile gösterilir. Genelde, öznel olarak seçilir. C sabitinin değeri fonksiyondan fonksiyona değişebilirse de hesap kolaylığı açısından tüm temel bileşen fonksiyonları için aynı alınması önerilir (Ramsay ve Silverman, 2005). Ramsay ve Silverman (2005)'in çalışmasında kolay yorumlanabilir sonuçlar elde edilmesi açısından $C = 0.2$ olarak alınmıştır.

4 FONKSİYONEL DOĞRUSAL MODELLER

Değişkenler arası doğrusal ilişkiyi ortaya çıkarmak ve ileriye yönelik tahminlerde bulunmak amacıyla kullanılan doğrusal modellerin klasik istatistikte önemli bir yeri vardır. Temel bileşenler analizi gibi doğrusal modeller de Fonksiyonel Doğrusal Modeller (FDM) adı altında fonksiyonel veriye uyarlanabilir.

Doğrusal modeli oluşturan değişkenlerden en az birinin fonksiyonel yapıda olması durumunda model fonksiyonel doğrusal model adını alır. FDM yanıt değişkeni ve açıklayıcı değişkenlerin yapısına göre farklı gruplara ayrılır. Fonksiyonel yanıt değişkeni ve kategorik açıklayıcı değişkenlerden oluşan bir doğrusal model fonksiyonel varyans analizi (Functional Analysis of Variance, FANOVA) modeli adını alırken fonksiyonel yanıt değişkeninden ya da skaler yanıt değişkeni ve fonksiyonel açıklayıcı değişkenlerden oluşan doğrusal modeller ise fonksiyonel çok değişkenli regresyon modeli adını alır.

Fonksiyonel Doğrusal Regresyon Modelleri (FDRM) temel olarak Çizelge 4.1'deki gibi özetlenebilir.

Çizelge 4.1: Fonksiyonel Doğrusal Regresyon Modelleri

Model Adı	Yanıt Değişkeni	Açıklayıcı Değişkenler	Fonksiyonel Doğrusal Model
Fonksiyonel Yanıt Modeli	Fonksiyonel	Skaler	$\gamma_i(t) = \beta(t)X_i + \varepsilon_i(t)$
Tam Fonksiyonel Model	Fonksiyonel	Fonksiyonel	$\gamma_i(t) = \int \beta(t,s)\chi_i(s)ds + \varepsilon_i(t)$
Skaler Yanıt Modeli	Skaler	Fonksiyonel	$Y_i = \int \beta(s)\chi_i(s)ds + \varepsilon_i$

Bu bölümde, gerçel bir rassal değişken ile bir T aralığında tanımlı fonksiyonel rassal değişkenler arasındaki ilişkiyi modellemek için kullanılan skaler yanıt değişkeni için

FDRM üzerinde durulacaktır. Literatürde, skaler bir yanıt değişkenini fonksiyonel değişkenlerle açıklamak amacıyla kullanılan pek çok çalışma vardır (Ramsay ve Silverman, 2005; Cardot ve diğ., 1999; Aguilera ve diğ., 2010; Preda ve Schiltz, 2011). Bu modeller özellikle Ferraty ve Vieu (2006)'nın ve Aguilera ve diğ. (2010)'nin çalışmalarındaki gibi açıklayıcı değişkenlerin dalgaboylarının fonksiyonları şeklinde tanımlandığı kemometri alanında büyük öneme sahiptir.

4.1 Skaler Yanıt Değişkeni için FDRM

Skaler yanıtı olan bir klasik doğrusal regresyon modeli,

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (4.1)$$

\mathbf{Y} ; $N \times 1$ 'lik bir yanıt vektörü, \mathbf{Z} ; $N \times p$ boyutlu dizayn matrisi, $\boldsymbol{\beta}$; $p \times 1$ boyunda parametre vektörü ve $\boldsymbol{\varepsilon}$; $N \times 1$ boyutunda hata terimini göstermek üzere (4.1) eşitliğindeki gibi tanımlanır.

Doğrusal bir regresyon modeli ile FDRM arasındaki en temel fark fonksiyonel durumda modeldeki bazı elemanların vektörler yerine karesi integrallenebilen fonksiyonlardan oluşmasıdır. Bu bağlamda, fonksiyonel skaler yanıt modeli,

$$\mathbf{Y} = \langle \boldsymbol{\chi}, \boldsymbol{\beta} \rangle + \boldsymbol{\varepsilon} = \int_T \boldsymbol{\chi}(s)\boldsymbol{\beta}(s)ds + \boldsymbol{\varepsilon}, \quad (4.2)$$

biçimini alır.

Burada, tahmin edici $t \in T$ noktalarında değerler alan ve $E[\|\boldsymbol{\chi}\|^2] \leq \infty$ ile $\mu_{\boldsymbol{\chi}}(t) = E[\boldsymbol{\chi}(t)]$ koşullarını sağlayan karesi integrallenebilen bir fonksiyondur. Fonksiyonel modelin eğimi $\boldsymbol{\beta}(s)$, $L^2(T)$ uzayında karesi integrallenebilen bir fonksiyon, hata terimi $\boldsymbol{\varepsilon}$ ise 0 ortalama ve $\sigma_{\boldsymbol{\varepsilon}}^2$ varyanslı gerçek bir rassal değişkendir (Febrero-Bande ve diğ., 2015).

(4.2) eşitliği ile verilen FDRM'nin, yanıt değişkenini mükemmel derecede tahmin eden sonsuz sayıda çözümü vardır (Ramsay ve Silverman, 2005). Bu nedenle, yorumlanabilir bir $\boldsymbol{\beta}(s)$ parametre fonksiyonu kestirimi bulmak için taban fonksiyonlarına ve parametrik olmayan düzgülendirmeye dayalı farklı yöntemler önerilir (Ramsay ve Silverman, 2005; Ferraty ve Vieu, 2006; Aguilera ve diğ., 2010).

Bu tezde en sık kullanılan üç yöntem anlatılacaktır. İlk yöntem klasik taban açılımı temsilidir. Diğerleri ise, boyut indirgeme amacıyla kullanılan fonksiyonel temel bileşenler analizi ve fonksiyonel kısmi en küçük kareler analizi yöntemlerine bağlıdır. Bu nedenle sırasıyla Fonksiyonel Temel Bileşenler Regresyonu (Functional Principal Components Regression, FTBR) ve Fonksiyonel Kısmi En Küçük Kareler Regresyon (Functional Partial Least Squares Regression, FKEKKR) olarak adlandırılırlar. FTBR ve FKEKKR modelleri iki farklı yaklaşım temel alınarak kurulabilir. Bunlardan ilki, taban açılımı yaklaşımıdır. Diğeri ise, Febrero-Bande ve Oviedo de la Fuente (2012) tarafından önerilen ve verinin öz yapısından elde edilen fonksiyonel bileşenlere dayalı olan yaklaşımıdır. Bu çalışmada ikinci yaklaşım üzerinde durulacaktır. Ancak, taban fonksiyonları aracılığıyla FTBR ve FKEKKR modellerinin elde edilmesine ilişkin ayrıntılı bilgi Ocaña ve diğ. (2007); Aguilera Morillo (2013); Aguilera ve diğ. (2010) çalışmalarında bulunabilir.

4.1.1 Taban Temsili ile FDRM

(4.2) ile verilen fonksiyonel doğrusal model aşağıdaki eşitlikten kestirilir:

$$\hat{Y}_i = \int_T \chi_i(s) \hat{\beta}(s) ds. \quad (4.3)$$

Bu ifade $\langle \chi_i, \hat{\beta} \rangle$ iç çarpımını tanımlar ve taban fonksiyonu yaklaşımı kullanılarak açılabilir.

$\beta(s)$ parametre fonksiyonunun bir kestirimi θ_l 's taban fonksiyonlarından oluşan taban açılımıdır:

$$\beta^*(s) = \sum_{l=1}^L b_l \theta_l(s). \quad (4.4)$$

Burada, b_1, b_2, \dots, b_L katsayıları,

$$\sum_{i=1}^N \left[Y_i - \int_T \chi_i(s) \beta^*(s) ds \right]^2 = \sum_{i=1}^N \left[Y_i - \sum_{l=1}^L \langle \chi_i, \theta_l \rangle b_l \right]^2. \quad (4.5)$$

olarak verilen SSE kriterinin minimize edilmesi ile kestirilir. Böylelikle parametre fonksiyonunun kestirimi,

$$\hat{\beta}(s) = \sum_{k=1}^L \hat{b}_k \theta_k(s) = \boldsymbol{\theta}(s)' \hat{\mathbf{b}}, \quad (4.6)$$

olarak elde edilir.

Öte yandan, $\chi_i(s)$ fonksiyonel tahmin edicileri Bölüm 2'de belirtildiği gibi $\phi_1(s), \phi_2(s), \dots, \phi_K(s)$ taban fonksiyonları cinsinden yazılabilir,

$$\chi_i(s) = \sum_{k=1}^K c_k \phi_k(s) = \mathbf{C} \boldsymbol{\phi}(s). \quad (4.7)$$

(4.6) ve (4.7) eşitlikleri ile verilen açılımlar (4.3) modeline yerleştirilirse, \mathbf{C} ve $\hat{\mathbf{b}}$ sırasıyla $\chi_i(s)$ ve $\hat{\beta}(s)$ fonksiyonlarına ait taban katsayı vektörlerini ifade etmek üzere fonksiyonel doğrusal model,

$$\hat{Y}_i = \int \mathbf{C} \boldsymbol{\phi}(s) \boldsymbol{\theta}(s)' \hat{\mathbf{b}} ds, \quad (4.8)$$

biçimini alır.

$$\mathbf{J}_{\phi\theta} = \int \boldsymbol{\phi}(s) \boldsymbol{\theta}'(s) ds, \quad (4.9)$$

(4.9) ile tanımlanan $\mathbf{J}_{\phi\theta}$, taban fonksiyonlarının iç çarpımlarından oluşan $K \times L$ boyutlu bir matris olsun. O zaman, (4.8) eşitliği matris biçiminde,

$$\hat{Y}_i = \mathbf{C} \mathbf{J}_{\phi\theta} \hat{\mathbf{b}}, \quad (4.10)$$

yazılabilir.

b_0 sabit terimi ifade etmek üzere, $\boldsymbol{\zeta} = (b_0, b_1, \dots, b_K)$, $L + 1$ uzunluğunda bir parametre vektörü ve $\mathbf{Z} = [1 \ \mathbf{C} \mathbf{J}_{\phi\theta}]$, $N \times K$ boyutlu bir tasarım matrisi olarak tanımlanabilir. (4.10) ifadesi ile verilen FDRM, $\hat{\boldsymbol{\zeta}}$ kestirilen parametre vektörü olmak üzere,

$$\hat{\mathbf{Y}} = \mathbf{Z} \hat{\boldsymbol{\zeta}}, \quad (4.11)$$

şeklinde sıradan bir doğrusal regresyon problemine indirgenir.

Fonksiyonel modele ait,

$$SSE = \sum_{i=1}^N (Y_i - \hat{Y}_i) = \sum_{i=1}^N (Y_i - \langle \chi_i, \hat{\beta} \rangle), \quad (4.12)$$

olarak tanımlanan hata kareler toplamı (4.11) ve (4.12) eşitlikleri kullanılarak matris formunda yazılabilir:

$$SSE = \|\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\zeta}}\|^2. \quad (4.13)$$

Bu problem klasik bir doğrusal regresyon problemidir ve $\boldsymbol{\zeta}$ parametresinin en küçük kareler kestirimi (4.13) eşitliğinin minimize edilmesinden hesaplanır. Bu da,

$$\mathbf{Z}'\mathbf{Z}\hat{\boldsymbol{\zeta}} - \mathbf{Z}'\mathbf{Y} = 0, \quad (4.14)$$

eşitliğinin gerçekleşmesine yol açar. Buradan, $\hat{\boldsymbol{\zeta}}$,

$$\hat{\boldsymbol{\zeta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}, \quad (4.15)$$

olarak elde edilir. (4.15) değerinin, (4.11) denkleminde yerleştirilmesiyle, yanıtın tahmin değerlerinden oluşan $\hat{\mathbf{Y}}$ vektörü elde edilir.

$$\hat{\mathbf{Y}} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}. \quad (4.16)$$

Bu eşitlikte, $\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ ifadesi klasik çok değişkenli doğrusal regresyon problemindeki \mathbf{H} ile gösterilen, $D_f = \text{iz}(\mathbf{H})$ serbestlik dereceli, şapka (izdüşüm) matrisine eşittir.

4.1.2 Pürüzlü Ceza Yaklaşımı

(4.3) biçiminde verilen fonksiyonel doğrusal regresyon modeline ait parametre fonksiyonu Pürüzlü Ceza Yaklaşımı (PCY) yöntemi kullanılarak daha düzgün bir şekilde kestirilebilir. Bu durumda, λ düzleştirme parametresi ve D^m , $\beta(s)$ parametre fonksiyonunun m 'inci türevini gösteren doğrusal diferansiyel operatör olmak üzere, fonksiyonel modele ait cezalandırılmış SSE ,

$$\sum_{i=1}^N [Y_i - \hat{Y}_i]^2 + \lambda \int [D^m \beta(s)]^2 ds, \quad (4.17)$$

olarak tanımlanır. (4.17)'deki son terim ceza terimidir ve $\beta(s)$ parametre fonksiyonunun yapısına bağlı olarak seçilir. Ceza parametresi olarak genelde parametre fonksiyonunun ikinci türevi kullanılır (Ramsay ve Silverman, 2005).

\mathbf{R} , eşitlik (2.57)'deki gibi tanımlanan bir ceza matrisi olsun. Böylelikle pürüzlü ceza terimi matris biçiminde yazılabilir. (4.10) eşitliği göz önüne alınacak olursa, cezalı hata kareler toplamı,

$$\text{PENSSE}_\lambda(\hat{b}_0, \hat{\mathbf{b}}) = \|\mathbf{Y} - \hat{b}_0 - \mathbf{C}\mathbf{J}_{\phi, \theta} \hat{\mathbf{b}}\|^2 + \lambda \hat{\mathbf{b}}' \mathbf{R} \hat{\mathbf{b}}, \quad (4.18)$$

olarak yazılır.

\mathbf{R}_0 , sıfırlardan oluşan K boyutlu bir vektörün ceza matrisinin satır ve bir sütunlarına eklenmesiyle oluşan yeni matris olsun. Bu durumda, (4.18) ifadesi $\hat{\boldsymbol{\zeta}} = (\hat{b}_0, \hat{\mathbf{b}})'$ parametre vektörü ve \mathbf{Z} tasarım matrisi cinsinden daha basit biçimde matris formunda yeniden yazılabilir.

$$\text{PENSSE}_\lambda(\hat{\boldsymbol{\zeta}}) = \|\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\zeta}}\|^2 + \lambda \hat{\boldsymbol{\zeta}}' \mathbf{R}_0 \hat{\boldsymbol{\zeta}}. \quad (4.19)$$

Bu problemin çözümü,

$$(\mathbf{Z}'\mathbf{Z} + \lambda \mathbf{R}_0) \hat{\boldsymbol{\zeta}} = \mathbf{Z}'\mathbf{Y}, \quad (4.20)$$

eşitliğinin çözümünden elde edilir.

Buradan, $\boldsymbol{\zeta}$ parametre vektörü,

$$\text{Var}[\hat{\boldsymbol{\zeta}}] = \sigma_\varepsilon^2 (\mathbf{Z}'\mathbf{Z} + \lambda \mathbf{R}_0)^{-1} \mathbf{Z}'\mathbf{Z} (\mathbf{Z}'\mathbf{Z} + \lambda \mathbf{R}_0)^{-1}, \quad (4.21)$$

varyansı ile

$$\hat{\boldsymbol{\zeta}} = (\mathbf{Z}'\mathbf{Z} + \lambda \mathbf{R}_0)^{-1} \mathbf{Z}'\mathbf{Y}, \quad (4.22)$$

olarak kestirilir.

λ düzleştirme parametresi ise öznel olarak ya da Bölüm 2’de bahsedilen yöntemlerden biri kullanılarak seçilebilir.

Örneğin çapraz geçerlilik algoritması bu yaklaşıma (4.23) eşitliğindeki gibi uyarlanabilir:

$$CV_\lambda = \sum_{i=1}^N [Y_i - \hat{b}_0^{(-i)} - \int z_i(s) \hat{\beta}_\lambda^{(-i)}(s)]^2 ds. \quad (4.23)$$

$\hat{b}_0^{(-i)}$ ve $\hat{\beta}_\lambda^{(-i)}(s)$ sırasıyla sabit terim ve β parametresinin veri setinden i 'nci gözlem çifti (z_i, y_i) çıkarıldıktan sonra kalan tüm gözlemler üzerinden hesaplanan cezalı hata kareler ortalamasının en küçüklenmesiyle elde edilen kestirim değerlerini ifade eder. CV kriterini en küçükleyen λ değeri, optimum değer kabul edilir (Ramsay ve Silverman, 2005).

CV_λ 'nın başka bir tanımı da izdüşüm matrisi \mathbf{H} 'ye dayalı olarak verilir. Cezalı en küçük kareler regresyonda \mathbf{H} şapka matrisi,

$$\mathbf{H} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{R})^{-1}\mathbf{Z}', \quad (4.24)$$

olarak tanımlanır. h_{ii} , \mathbf{H} matrisinin köşegen elemanlarını göstermek üzere CV_λ ,

$$CV_\lambda = \sum_{i=1}^N \left(\frac{Y_i - \hat{Y}_i^2}{1 - h_{ii}} \right)^2, \quad (4.25)$$

eşitliğinden hesaplanabilir.

Hesap kolaylığından ötürü, FDM'nin çözümünde genelde taban açılımı yaklaşımı kullanılır. Ancak çoklu bağlantı durumunda, tek başına bu yöntem yeterli değildir. Böyle bir durumda çoklu bağlantıyı engellemek ve boyut indirmek için Fonksiyonel Temel Bileşenler Regresyonu (FTBR) ya da Fonksiyonel Kısmi En Küçük Kareler Regresyonu (FKEKKR) adı verilen yöntemlerin kullanılması önerilir. Bu yöntemlerden ikincisi, skaler bir yanıt değişkeni ile fonksiyonel yapıdaki açıklayıcı değişkenler arasındaki ilişkiye dayanır ve daha çok fonksiyonel skaler yanıt modellerinde kullanılır.

4.1.3 Fonksiyonel Temel Bileşenler Regresyonu (FTBR)

FPCR'nin ana fikri skaler bir \mathbf{Y} yanıt vektörünü fonksiyonel temel bileşenler skorları üzerinden tahmin etmektir. Γ_χ kovaryans operatörüne ait ξ_1, ξ_2, \dots özfonksiyonlarının $L^2(T)$ uzayında ortonormal bir taban oluşturduğu göz önüne alınır, $\mathbb{E}[\chi] = \mu_\chi$ ortalamalı $\chi(t)$, $t \in T$ stokastik süreci, Ek (B.12)'de verilen Karhunen-Loève Teoremi kullanılarak fonksiyonel temel bileşenler cinsinden şu şekilde temsil edilebilir:

$$\hat{\chi}(t) = \mu_\chi + \sum_{j=1}^{\infty} f_j \xi_j(t), \quad i = 1, \dots, N. \quad (4.26)$$

Burada f_j , $j \geq 1$, temel bileşen skorlarının oluşturduğu sıfır ortalamalı ve ilişkisiz rassal değişkenleri ifade etmektedir.

\mathbf{Y} yanıt vektörünün ortalamasının $\mathbb{E}[\mathbf{Y}] = \mu_{\mathbf{Y}}$ ve varyansının $\sigma_{\mathbf{Y}}^2 = \mathbb{E}[(\mathbf{Y} - \mu_{\mathbf{Y}})^2]$ olduğunu varsayalım. \mathbf{Y} skaler yanıt değişkenine sahip FDM,

$$\mathbf{Y} = \mu_{\mathbf{Y}} + \langle \chi - \mu_\chi, \beta \rangle + \varepsilon = \mu_{\mathbf{Y}} + \int_T (\chi(t) - \mu_\chi(t)) \beta(t) dt + \varepsilon, \quad (4.27)$$

şeklinde ifade edilir.

$\chi(t)$ ve FTB skorları f_j 'lerin kümesi aynı uzayı gerer. Bu yüzden, \mathbf{Y} skaler yanıt vektörünün $\chi(t)$ üzerindeki fonksiyonel doğrusal regresyonu, aynı yanıt vektörünün fonksiyonel temel bileşen skorları f_j 'ler üzerinden yapılan doğrusal regresyonuna denktir (Preda ve Saporta, 2005). Bu denklik, (4.27) eşitliğindeki χ fonksiyonel değişkeni üzerinde (4.26) açılımının yapılmasıyla kolaylıkla görülebilir.

$$\begin{aligned} \mathbf{Y} &= \mu_{\mathbf{Y}} + \int_T \chi(t) \beta(t) dt - \int_T \mu_\chi(t) \beta(t) dt + \varepsilon \\ &= \mu_{\mathbf{Y}} + \int_T \mu_\chi(t) \beta(t) dt + \int_T \sum_{j=1}^{\infty} f_j \xi_j(t) \beta(t) dt - \int_T \mu_\chi(t) \beta(t) dt + \varepsilon \\ &= \mu_{\mathbf{Y}} + \int_T \sum_{j=1}^{\infty} f_j \xi_j(t) \beta(t) dt + \varepsilon. \end{aligned} \quad (4.28)$$

(4.28) denklemi aslında $f_j = \langle \chi^c, \xi \rangle$ açıklayıcı değişkenler, $b_j = \int_T \xi_j(t) \beta(t) dt = \langle \beta, \xi_j \rangle$ de katsayı vektörü olan,

$$\mathbf{Y} = \mu_{\mathbf{Y}} + \sum_{j=1}^{\infty} f_j b_j + \varepsilon, \quad (4.29)$$

biçiminde klasik bir doğrusal regresyon problemine denktir.

$\beta(t)$ fonksiyonel eğimi, ortonormal fonksiyonel temel bileşenler cinsinden yazılabilir:

$$\beta(t) = \sum_{j=1}^{\infty} b_j \xi_j(t). \quad (4.30)$$

(4.29) modelindeki f_j terimi de açılırsa, fonksiyonel doğrusal regresyon modeline ait sabit terim,

$$b_0 = \mu_{\mathbf{Y}} - \int_T \mu_{\chi}(t) \beta(t) dt, \quad (4.31)$$

olarak bulunur. $\beta(t)$ parametre fonksiyonu karesi integrallenebilir bir fonksiyon olduğundan, $\sum_{j=1}^{\infty} b_j^2 < \infty$ eşitliğini sağlar. Böylelikle, (4.29) modelinin b_j regresyon katsayıları klasik basit doğrusal regresyon problemine benzer şekilde (4.32) eşitliğinden hesaplanır.

$$b_j = \frac{\text{Cov}(\mathbf{Y}, f_j)}{\text{Var}(f_j)}. \quad (4.32)$$

Temel bileşen skorlarının varyansının, $\Gamma_{\chi} \xi_j = \lambda_j \xi_j$ denkleminde elde edilen ilgili özdeğere eşit olduğu düşünülürse $\beta(t)$ fonksiyonel eğimi,

$$\beta(t) = \sum_{j=1}^{\infty} \frac{\text{Cov}(\mathbf{Y}, f_j)}{\text{Var}(f_j)} \xi_j(t) = \sum_{j=1}^{\infty} \frac{c_{\mathbf{Y}, f_j}}{\lambda_j} \xi_j(t), \quad (4.33)$$

eşitliğindeki gibi yeniden yazılabilir.

Bu sonuç Horváth ve Kokoszka (2012) tarafından aşağıda verilen Lemma ile ifade edilmiştir.

Lemma (Horváth ve Kokoszka, 2012): $\chi(t)$ merkezleştirilmiş bir süreç ve Y ile ε sıfır ortalamalı rassal değişkenler olsun. β 'nin $\int \int \beta^2(t) dt < \infty$ koşulu ile karesi integrallenebilir bir fonksiyon olması koşulu ile,

$$Y = \int_T \beta(t) \chi(t) dt + \varepsilon, \quad (4.34)$$

doğrusal modeli elde edilir.

$\xi_j(t)$ fonksiyonel temel bileşenleri ve $f_j = \langle \chi, \xi_j \rangle$ iç çarpımı temel bileşen skorlarını belirtsin. O zaman,

$$\beta(t) = \sum_{j=1}^{\infty} \frac{\mathbb{E}[f_j Y]}{\mathbb{E}[f_j^2]} \xi_j(t), \quad (4.35)$$

ifadesi gerçektir.

FDM'nin uyum iyiliği R^2 ile gösterilen belirtme katsayısı aracılığıyla ölçülür. R^2 skaler yanıt değişkenindeki değişimin χ fonksiyonel tahmin edicisi tarafından açıklanan varyans oranıdır (Febrero-Bande ve diğ., 2015). Fonksiyonel durumda bu değer,

$$R^2 = \frac{\text{Var}(\mathbb{E}[\mathbf{Y}|\chi])}{\text{Var}[\mathbf{Y}]}, \quad (4.36)$$

eşitliğinden hesaplanır.

(4.36) formülünün payı,

$$\text{Var}(\mathbb{E}[\mathbf{Y}|\chi]) = \sum_{j=1}^{\infty} \lambda_j b_j^2 = \sum_{j=1}^{\infty} \frac{c_{\mathbf{Y},f_j}^2}{\lambda_j}, \quad (4.37)$$

değerine eşittir. O halde skaler yanıt değişkeni ve FTB skorları arasındaki korelasyon $r_{\mathbf{Y},f_j}^2$ ile göstermek üzere, belirtme katsayısı R^2 ,

$$R^2 = \frac{1}{\text{Var}[\mathbf{Y}]} \sum_{j=1}^{\infty} \frac{c_{\mathbf{Y},f_j}^2}{\lambda_j} = \sum_{j=1}^{\infty} r_{\mathbf{Y},f_j}^2, \quad (4.38)$$

şeklinde tekrar yazılabilir. Buradan, FTBR modelinin uyum iyiliğinin FTB skorları ile yanıt değişkeni arasındaki ilişkiye bağlı olduğu söylenebilir.

$\chi(t)$ ile aynı dağılıma sahip, bağımsız ve aynı dağılan $\chi_1, \chi_2, \dots, \chi_N$ gözlemlerinden oluşan bir örnekleme ele alalım. (χ, \mathbf{Y}) 'den seçilen bir örnekleme ait fonksiyonel doğrusal regresyon modeli,

$$\hat{Y}_i = \bar{Y} + \int_T (\chi_i(t) - \bar{\chi}(t)) \hat{\beta}_j(t) dt, \quad i = 1, 2, \dots, N, \quad j = 1, \dots, K, \quad (4.39)$$

biçimindedir.

Örneklem kovaryans operatörü $\hat{\Gamma}_\chi$ 'nin özfonksiyonları $\hat{\xi}_j$, özdeğerleri $\hat{\lambda}_j$ ile verilsin. $i = 1, \dots, N$ ve $j = 1, \dots, K$ için $\hat{f}_{ij} = \langle \chi - \bar{\chi}, \hat{\xi}_j \rangle$ iç çarpımından elde edilen örneklem FTB skorları $\hat{f}_j = (\hat{f}_{1j}, \dots, \hat{f}_{Nj})^T$ ile gösterilsin. Bu halde fonksiyonel doğrusal regresyon modelinin eğimi, açıklayıcı değişkenlerin en büyük K adet $\hat{\lambda}_j$ özdeğeri ile ilişkili $\hat{\xi}_j, j = 1, \dots, K$ ampirik fonksiyonel bileşeni üzerindeki yansımından kestirilir.

$$\hat{\beta}_j(t) = \sum_{j=1}^K \hat{b}_j \hat{\xi}_j(t). \quad (4.40)$$

$\hat{\beta}_j(t)$ 'yi hesaplamak için öncelikle (4.41) denklemini en küçükleyen \hat{b}_j değerleri kestirilmelidir.

$$\sum_{i=1}^N [Y_i - \sum_{j=1}^K \langle \chi_i^c(t), \hat{\xi}_j \rangle \hat{b}_j]^2. \quad (4.41)$$

Yukarıda bahsi geçen Lemma'ya göre (Horváth ve Kokoszka, 2012), \hat{b}_j regresyon katsayıları skaler yanıt değişkeni ile örneklem temel bileşen skorları arasındaki kovaryansın, f_{ij} FTB skorlarının varyansına, yani $\hat{\lambda}_j$ ilgili özdeğerine, oranından elde edilir.

$$\hat{b}_j = \frac{\text{Cov}(\mathbf{Y}, f_j)}{\text{Var}(f_j)} = \frac{\hat{c}_{\mathbf{Y}, f_j}}{\hat{\lambda}_j}, \quad j = 1, \dots, K. \quad (4.42)$$

Böylece, $\beta(t)$ parametre fonksiyonunun kestirimi kovaryans fonksiyonunun, özfonksiyonların ve özdeğerlerin örneklem eşlenikleri cinsinden,

$$\hat{\beta}_{FTB}(t) = \sum_{j=1}^K \frac{\hat{c}_{\mathbf{Y}, f_j}}{\hat{\lambda}_j} \hat{\xi}_j(t), \quad (4.43)$$

şeklinde yeniden yazılabilir.

\hat{b}_j katsayı vektörü kestirimi ilk K temel bileşen cinsinden

$$\hat{b}_j = \left(\frac{f_{.1}^T \mathbf{Y}}{N \hat{\lambda}_1}, \frac{f_{.2}^T \mathbf{Y}}{N \hat{\lambda}_2}, \dots, \frac{f_{.K}^T \mathbf{Y}}{N \hat{\lambda}_K} \right). \quad (4.44)$$

olarak ifade edilir.

H matrisi,

$$\mathbf{H} = \left(\frac{f_{.1}f_{.1}^T \mathbf{Y}}{N\hat{\lambda}_1}, \frac{f_{.2}f_{.2}^T \mathbf{Y}}{N\hat{\lambda}_2}, \dots, \frac{f_{.K}f_{.K}^T \mathbf{Y}}{N\hat{\lambda}_K} \right). \quad (4.45)$$

elemanlarından oluşmak üzere yanıt vektörünün tahmin değerleri $\hat{\mathbf{Y}} = \mathbf{H}\mathbf{Y}$ eşitliğinden elde edilir. Burada **H**, $\text{iz}(\mathbf{H}) = K$ serbestlik derecesine sahip bir şapka matrisidir.

Buradan hata vektörü,

$$\hat{\sigma}_\varepsilon^2 = \frac{\varepsilon_i \varepsilon_i^T}{N - K - 1}, \quad (4.46)$$

varyansı ile,

$$\hat{\varepsilon}_i = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{H})\mathbf{Y}, \quad (4.47)$$

vektörüne eşittir.

Kestirilen bir fonksiyonel doğrusal regresyon modelinin uyum iyiliğini hesaplamak için,

$$\hat{R}^2 = \sum_{j=1}^K \hat{\lambda}_j \hat{b}_j^2 = \sum_{j=1}^K \hat{r}_{\mathbf{Y}, f_j}^2, \quad (4.48)$$

şeklinde tanımlanan örneklem belirtme katsayısı \hat{R}^2 kullanılır.

FTBR'deki en önemli problem, optimum sayıda temel bileşenin seçilmesidir. Yanıtı en iyi kestiren K bileşen sayısına karar verilmesinde önceki bölümlerde anlatılan CV kriteri ya da benzer başka bir model seçimi kriteri kullanılabilir (Febrero-Bande ve Oviedo de la Fuente, 2012).

CV kriteri,

$$\text{CV}(K) = \frac{1}{N} \sum_{i=1}^N (Y_i - \langle \mathbf{x}_i, \hat{\boldsymbol{\beta}}_{(-i, K)} \rangle)^2, \quad (4.49)$$

formülü ile verilir.

Diğer Model Seçim Kriterleri (Model Selection Criteria, MSC) için genel formül,

$$MSC(K) = \log \left[\frac{1}{N} \sum_{i=1}^N \left(Y_i - \langle \chi_i, \hat{\beta}_{(i,K)} \rangle \right)^2 \right] + p_N \frac{K}{N}, \quad (4.50)$$

şeklinde yazılabilir. Bu formülde, eşitliğin sağ yanındaki son terim, p_N , kullanılan bilgi kriterine göre farklılık gösterir. En çok kullanılan kriterler Schwarz Bilgi Kriteri (Schwarz Information Criterion, SIC), Düzeltilmiş Schwarz Bilgi Kriteri (Corrected SIC, SICc), Akaike Bilgi Kriteri (Akaike Information Criterion, AIC) ve Düzeltilmiş Akaike Bilgi Kriteri (Corrected AIC, AICc)'dir. MSC'yi en küçükleyen K değeri, yansımaların sayısı kadar alınır.

SIC, SICc, AIC and AICc kriterleri için p_N değerleri sırasıyla şu şekildedir (Febrero-Bande ve Oviedo de la Fuente, 2012):

$$SIC(p_N) = \frac{\log(N)}{N}, \quad (4.51)$$

$$SICc(p_N) = \frac{\log(N)}{N - K - 2}, \quad (4.52)$$

$$AIC(p_N) = 2, \quad (4.53)$$

$$AICc(p_N) = \frac{2N}{N - K - 2}. \quad (4.54)$$

4.1.4 Fonksiyonel Kısmi En Küçük Kareler Regresyonu (FKEKKR)

Kısmi En Küçük Kareler Regresyon (KEKKR), yanıt değişkeni ve tahmin ediciler arasındaki ilişkiyi dikkate alan, Temel Bileşenler Regresyona (TBR) alternatif popüler bir kestirim yöntemidir (Delaigle ve diğ., 2012). Bu yöntem daha çok kemometri alanında tercih edilir (Aguilera ve diğ., 2010). KEKK yaklaşımı fonksiyonel duruma Preda ve Saporta (2005) tarafından uyarlanmıştır. Daha sonrasında Wang ve diğ. (2009) fonksiyonel veride karşılaşılan çoklu bağlantı probleminin çözümü için yeni bir KEKK yöntemi de önermiştir. Yanıt değişkeninin skaler olduğu durum için literatürde pek çok FKEKKR uygulaması bulunmaktadır (Preda ve Saporta, 2005; Reiss ve Ogden, 2007; Aguilera ve diğ., 2010; Febrero-Bande ve diğ., 2015). Yakın zamanda

Preda ve Schiltz (2011) bu yaklaşımı hem yanıt hem de açıklayıcı değişkenlerin fonksiyonel durumda olduğu tam fonksiyonel modele genişletmiştir.

FKEKK yöntemi $\chi(t)$ fonksiyonel bağımsız değişkeni ile \mathbf{Y} skaler yanıt değişkeni arasındaki kovaryansın maksimize edilmesine dayanır. Bundan dolayı, FKEKK bileşenleri yanıtta ki değişimle daha çok ilgilidir ve sonucu tahmin etmek için daha kullanışlıdır (Oviedo de la Fuente, 2011; Reiss ve Ogden, 2007). Bu yaklaşımın çözümünde kullanılan maksimizasyon kriteri literatürde Tucker kriteri olarak da bilinir ve $\varphi(t)$, $L^2(T)$ uzayının $\|\varphi(t)\|_{L^2(T)} = 1$ normuna sahip fonksiyonel bir elemanı olmak üzere, şu şekilde tanımlanır (Aguilera ve diğ., 2010):

$$\max \text{Cov}^2\left(\int_T \chi(t)\varphi(t)dt, \mathbf{Y}\right). \quad (4.55)$$

FKEKK bileşenleri iteratif bir alitmadan hesaplanır. Bu algoritmanın adımları şu şekilde özetlenebilir (Febrero-Bande ve diğ., 2015):

1. Yanıt vektörü ve açıklayıcı değişkenler merkezileştirilerek sırasıyla $Y^c = Y - \mu_Y$ ve $\chi^c = \chi - \mu_\chi$ olarak tanımlanır. İlk adım için $l = 0$ kabul edilir. Y_l yanıt vektörü ve χ_l tahmin edicisinin başlangıç değerleri $Y_0 = Y^c$ ve $\chi_0 = \chi^c$ olarak alınır.
2. $\varphi_{l+1} \in L^2(T)$ ilgili ağırlık fonksiyonu (4.56) eşitliğindeki gibi tanımlanır:

$$\varphi_{l+1} = \frac{\text{Cov}(Y_l, \chi_l)}{\|\text{Cov}(Y_l, \chi_l)\|} = \frac{E[Y_l \chi_l(t)]}{\int_T E[Y_l \chi_l(t)] dt}, \quad t \in [0, T]. \quad (4.56)$$

Tucker kriterinde verildiği gibi $\text{Cov}^2[Y_l, v_{l+1}] = c_{Y_l, v_{l+1}}^2$ ifadesini en büyükleyen FKEKK bileşen skorları $v_{l+1} = \langle \chi_l, \varphi_{l+1} \rangle$ ile gösterilir.

3. Y_l ve χ_l değişkenlerinin (4.57) ve (4.58) modellerinde verildiği gibi FKEKK bileşenleri üzerinde regresyon analizi yapılır.

$$Y_l = \psi_{l+1} v_{l+1} + \varepsilon_l(t), \quad (4.57)$$

$$\chi_l = \delta_{l+1} v_{l+1} + \eta_l(t), \quad (4.58)$$

Bu denklemlerden elde edilen regresyon katsayıları $\psi_{l+1} \in \mathbb{R}$ ve $\delta_{l+1} \in L^2[T]$, (4.59) ve (4.60) eşitliklerindeki gibi tanımlanır:

$$\psi_{l+1} = \frac{\text{Cov}[Y_l, v_{l+1}]}{\text{Var}[v_{l+1}]}, \quad (4.59)$$

$$\delta_{l+1} = \frac{\text{Cov}[\chi_l, v_{l+1}]}{\text{Var}[v_{l+1}]}. \quad (4.60)$$

4. (4.57) ve (4.58) modellerine ait hata terimleri Y_{l+1} ve χ_{l+1} ile gösterilir ve sırasıyla (4.61), (4.62) eşitliklerindeki gibi hesaplanır:

$$Y_{l+1} = Y_l - \psi_{l+1} v_{l+1}, \quad (4.61)$$

$$\chi_{l+1} = \chi_l - \delta_{l+1} v_{l+1}. \quad (4.62)$$

5. $l = l + 1$ alınır ve bu algoritma ikinci adımdan itibaren tekrar uygulanır.

FKEKK bileşenleri v_l 'ler, $\{\chi(t) : t \in [0, T]\}$ fonksiyonel değişkeni ve $\varphi_l(t)$ ağırlık fonksiyonlarının iç çarpımı ile (4.63) eşitliğindeki gibi tanımlanır:

$$v_l = \langle \chi_{l-1}, \varphi_l \rangle = \int_T \chi_{l-1}(t) \varphi_l(t) dt. \quad (4.63)$$

FTBR'dekine benzer şekilde, FKEKKR de yanıt vektörünün FKEKK bileşenleri üzerinde regresyonunun yapılmasına dayanır. (4.29) fonksiyonel doğrusal modeli FKEKK bileşenlerinin doğrusal bir kombinasyonu olarak yazılabilir (Oviedo de la Fuente, 2011).

$\beta(t)$ parametre fonksiyonu,

$$\mathbf{Y} = \mu_{\mathbf{Y}} + \sum_{l=1}^{\infty} c_l v_l + \varepsilon, \quad (4.64)$$

denkleminin açılımından bulunur.

Başlangıç olarak tek bir FKEKK bileşeni kullanıldığını varsayalım. $l = 1$ için, (4.64) denklemi,

$$\mathbf{Y} = \mu_{\mathbf{Y}} + c_1 v_1 + \varepsilon, \quad (4.65)$$

şeklinde yazılır.

FKEKK bileşenlerinin (4.63) ile verilen tanımından, v_1 ,

$$v_1 = \int_T \chi_0(t) \phi_1(t) dt, \quad (4.66)$$

olur.

$\chi_0 = \chi^c$ ilk değeri kullanılıp $\phi_1(t)$ 'nin $\phi_1(t)$ 'ye eşit olduğu göz önüne alınırsa denklem,

$$v_1 = \int_T \chi^c(t) \phi_1(t) dt, \quad (4.67)$$

biçimini alır.

Böylelikle (4.65) eşitliği,

$$\mathbf{Y} = \mu_{\mathbf{Y}} + c_1 \int_T \chi^c(t) \phi_1(t) dt + \varepsilon, \quad (4.68)$$

olarak tekrar yazılabilir.

Buradan kolaylıkla görüldüğü gibi, $l = 1$ için $\beta(t)$ fonksiyonel parametresi,

$$\beta(t) = c_1 \phi_1(t), \quad (4.69)$$

değerine eşittir.

Şimdi bir yerine iki adet FKEKK bileşeni alalım. Bu durumda $l = 2$ için (4.64) denkleminin açılımı,

$$\begin{aligned} \mathbf{Y} &= \mu_{\mathbf{Y}} + \sum_{l=1}^2 c_l v_l + \varepsilon \\ &= \mu_{\mathbf{Y}} + c_1 v_1 + c_2 v_2 + \varepsilon, \end{aligned} \quad (4.70)$$

olarak yazılabilir.

v_1 'in (4.67) ile verilen açılımı ve (4.62) ifadesi kullanılırsa, $\chi_1(t)$ 'nin,

$$\chi_1(t) = \chi_0(t) - \delta_1 v_1 = \chi^c(t) - \delta_1 v_1, \quad (4.71)$$

eşitliğini sağladığı görülür.

v_2 terimi ise,

$$\begin{aligned} v_2 &= \langle \chi_1, \varphi_2 \rangle = \int_T (\chi^c - \delta_1 v_1) \varphi_2(t) dt \\ &= \int_T (\chi^c \varphi_2(t) - \langle \delta_1, \varphi_2 \rangle v_1) dt \\ &= \int_T \chi^c \varphi_2(t) dt - \int_T \chi^c(t) \langle \delta_1, \varphi_2 \rangle \varphi_1 \\ &= \int_T \chi^c(t) [\varphi_2(t) - \langle \delta_1, \varphi_2 \rangle \varphi_1(t)] dt, \end{aligned} \quad (4.72)$$

şeklinde yazılabilir. (4.72) ile verilen bu eşitlik (4.70) denkleminde yerleştirilirse,

$$Y = \mu_Y + c_1 \int_T \chi^c(t) \varphi_1(t) dt + \int_T \chi^c(t) [\varphi_2(t) - \langle \delta_1, \varphi_2 \rangle \varphi_1(t)] dt, \quad (4.73)$$

modeli elde edilir.

Buradan, $\beta(t)$ parametresinin iki bileşenle yaklaşımı,

$$\beta(t) = c_1 \varphi_1 + c_2 [\varphi_2(t) - \langle \delta_1, \varphi_2 \rangle \varphi_1(t)] = c_1 \varphi_1(t) + c_2 \varphi_2(t), \quad (4.74)$$

olarak bulunur.

Buna dayanarak $\phi_2(t) = \varphi_2(t) - \langle \delta_1, \varphi_2 \rangle \varphi_1(t)$ olduğu görülür.

Üç bileşen kullanılır ve $l = 3$ alınır, v_1 ve v_2 terimlerinin yanısıra v_3 teriminin açılımının da

$$Y = \mu_Y + \sum_{i=1}^3 c_i v_i = c_1 v_1 + c_2 v_2 + c_3 v_3, \quad (4.75)$$

denkleminde yerine konulması gerekir.

(4.63) tanımını ve (4.62) eşitliği kullanılarak, üçüncü FKEKK bileşenine ait açılım,

$$\begin{aligned}
v_3 &= \langle \chi_2, \varphi_3 \rangle = \int_T (\chi^c - \delta_1 v_1 - \delta_2 v_2) \varphi_3(t) dt \\
&= \int_T (\chi^c \varphi_3(t) - \langle \delta_1, \varphi_3 \rangle v_1) dt - \langle \delta_2, \varphi_3 \rangle v_2 dt \\
&= \int_T \chi^c \varphi_3(t) dt - \int_T \chi^c(t) \langle \delta_1, \varphi_3 \rangle \phi_1 - \int_T \chi^c(t) \langle \delta_2, \varphi_3 \rangle \phi_2 \\
&= \int_T \chi^c(t) [\varphi_3(t) - \langle \delta_1, \varphi_3 \rangle \phi_1 - \langle \delta_2, \varphi_3 \rangle \phi_2] dt
\end{aligned} \tag{4.76}$$

olarak bulunur. Buradan, $\phi_3(t) = \varphi_3(t) - \langle \delta_1, \varphi_3 \rangle \phi_1 - \langle \delta_2, \varphi_3 \rangle \phi_2$ olur ve $\beta(t)$ parametre fonksiyonunun üç bileşenle yaklaşımı,

$$\begin{aligned}
\beta(t) &= c_1 \phi_1(t) + c_2 \phi_2(t) + c_3 [\varphi_3(t) - \langle \delta_1, \varphi_3 \rangle \phi_1 - \langle \delta_2, \varphi_3 \rangle \phi_2] \\
&= c_1 \phi_1(t) + c_2 \phi_2(t) + c_3 \phi_3(t).
\end{aligned} \tag{4.77}$$

ile verilir.

İleriki adımlar hesaplanırsa, $\phi_l(t)$ fonksiyonlarının aynı yapıya sahip olduğu görülür. Bu yüzden, $\phi_l(t)$ fonksiyonları şu şekilde genellenebilir:

$$\phi_1 = \varphi_1, \quad \text{for } l = 1, \tag{4.78}$$

$$\phi_l = \varphi_l - \langle \delta_1, \varphi_l \rangle \phi_1 - \dots - \langle \delta_{l-1}, \varphi_l \rangle \phi_{l-1}, \quad \text{for } l \geq 2. \tag{4.79}$$

ϕ_l sırasıyla (4.78) ve (4.79) koşullarını sağlamak üzere, fonksiyonel doğrusal modelin sonlu durumda regresyon katsayı kestirim fonksiyonu,

$$\beta = \sum_{l=1}^{\infty} c_l \phi_l, \tag{4.80}$$

(4.80) eşitliğinden hesaplanır.

q sayıda FKEKK bileşen kullanılması durumunda, fonksiyonel doğrusal modelin eğimi

$$\hat{\beta}_{FKEKK} = \sum_{l=1}^q \hat{c}_l \hat{\phi}_l, \tag{4.81}$$

şeklinde kestirilir.

FKEKK bileşenlerinin sayısına, FTB sayısının belirlenmesinde kullanılan yöntemlerden biri ile karar verilebilir (Oviedo de la Fuente, 2011).

FKEKKR modelinin uyum iyiliği, skaler yanıt değişkeni \mathbf{Y} ve FKEKK bileşenleri arasındaki korelasyonun karesini ifade eden belirtme katsayısı,

$$R^2 = \sum_l r_{\mathbf{Y},v_l}^2, \quad (4.82)$$

kullanılarak hesaplanır.



5 UYGULAMA

FVA, fonksiyonel verinin yoğun bir zaman aralığı veya farklı frekans kanallarından ya da dalga boylarından oluşan bir spektrum gibi sürekli bir aralık üzerinde ölçüldüğünü varsayar. FVA teknikleri, özellikle bir spektrum üzerinde ölçülmüş spektral veri setlerini analiz etmekte büyük öneme sahiptir. Spektral veri setleri üzerinde FVA tekniklerinin, başta kemometri alanında olmak üzere pek çok uygulaması mevcuttur (Saeys ve diğ., 2008; Aguilera ve diğ., 2013). Elektromanyetik enerjinin uzaktan algılama sensörleri aracılığıyla iletilmesinden elde edilen uydu verileri spektral veriye bir örnektir. Pidwirny (2006)'nin de belirttiği gibi, uzaktan algılamanın; toprak kullanımının haritalanması, tarım, ormancılık ve okyanus bilimi gibi farklı alanlarda pek çok uygulaması mevcuttur (Caballero ve diğ., 2014a; Faivre ve Fischer, 1997; Caballero ve diğ., 2014b; Nezhlin ve DiGiacomo, 2005). Okyanus biliminde uzaktan algılama verileri, deniz yüzey sıcaklığı (Sea Surface Temperature, SST), Chlorophyll-a (Chl-a) içeriği ve toplam katı madde miktarı (Total Suspended Solids, TSS) gibi karakteristik okyanus parametrelerini tahmin etmekte kullanılır (Caballero ve diğ., 2014a,b; Clarke ve diğ., 2006; Schwarz ve diğ., 2008). FVA, uzaktan algılama sensör veri setlerini analiz etmekte son zamanlarda önem kazanmaya başlamıştır. Cardot ve diğ. (2003) ve Besse ve diğ. (2005) Spot 4 uydusunun Vegetation sensöründen elde edilen uzaktan algılama verileri aracılığıyla toprak kullanımını tahmin etmek için FVA yaklaşımını kullanmıştır. Liu ve diğ. (2012) periyodik uzaktan algılama verileri üzerinde bir uygulamayla fonksiyonel faktör analizi için yeni bir dönüşüm yöntemi önermiştir. Gong ve diğ. (2015) FTBA'ni Viktoria Gölü'ne ait yüksek boyutlu sıcaklık eğrileri ve sıcaklık yüzeylerini modellemekte kullanmıştır. Son olarak Ferraty ve diğ. (2016), 400-2500 nm dalgaboyları arasında 2101 noktada ölçülmüş hiperspektral veriler üzerinden klorofil içeriğini tahmin etmek için parametrik olmayan fonksiyonel modelleri kullanmaktadır. Çok değişkenli analiz tekniklerinin okyanus biliminde uzaktan algılama verileri üzerine uygulamasına dayanan pek çok çalışma olmasına

rağmen (Caballero ve diğ., 2014a,b; Clarke ve diğ., 2006; Nezlin ve DiGiacomo, 2005), FVA yaklaşımını içeren az sayıda çalışma vardır (Gong ve diğ., 2015; Ferraty ve diğ., 2016).

Bu çalışmada, Guadalquivir nehir ağzına komşu sahil kesimindeki TSS oranını kestirmek amacıyla uydu verilerine FVA tekniklerinin uygulanması ve FVA yönteminin diğer alternatif klasik istatistiksel yöntemler ile uygulamalı olarak karşılaştırılması ve sonuçların bir simülasyon çalışması ile desteklenmesi amaçlanmaktadır.

5.1 Gereç ve Yöntemler

Çalışmada kullanılan veri seti uydudan kaydedilen uzaktan algılama verileri ile denizde yerinde ("in-situ") toplanan örneklemeler olmak üzere iki kısımdan oluşmaktadır. Uydu verisi, "Environmental Satellite" (ENVISAT) uydusu platformunda yer alan 300 m tam uzamsal çözünürlüğe (Full Spatial Resolution, FRS) sahip "MEDium Resolution Imaging Spectrometer" (MERIS) sensöründen elde edilmiştir. "In-situ" verisi okyanustan değişik zamanlarda farklı firmalara ait gemiler tarafından toplanan örneklemelerden oluşmaktadır.

5.1.1 Uzaktan Algılanan Veri

Çalışma alanı, İberya Yarımadasının güneybatı sahilinde bulunan (35.5° – 37.5° N enlem ve 1° – 10° W boylam) Cadiz körfezinin sahil kesimidir. Özel olarak, Batı Avrupa'nın en büyük ve en üretken nehir ağzı sistemlerinden biri olan Guadalquivir nehir ağzı incelenecektir. İlgili alana (Region Of Interest, ROI) ait uydu verileri hdf biçiminde "Ocean Colour" internet sayfasından (<http://oceancolor.gsfc.nasa.gov>) indirilmiştir. Veri seti 2002-2011 yılları arasında sekiz farklı dalga boyunda (413 nm, 443 nm, 490 nm, 510 nm, 560 nm, 620 nm, 665 nm, 681 nm) 300 m tam uzamsal çözünürlükle kaydedilmiş 2. düzey (Level-2) yani okyanus yüzeyinden kaydedilmiş uzaktan algılama yansıma (Reflectance of remote sensing, Rrs) (sr^{-1}) değerlerinden oluşmaktadır. 2. düzey verileri sensor kalibrasyonu ve atmosferik düzeltmenin sonucudur, standart NASA süreçleri kullanılarak ilgili radyometrik düzeltme yapılmış atmosferin üstünden kaydedilen Level-1A ürünlerinden türetilmiş jeofiziksel değişkenleri içerir. MERIS uydusunun Merkez Avrupa için geçiş zamanı her üç günde bir küresel kaplama ile 9:30 and 11:00 UTC saatleri arasındadır.

Çizelge 5.1: L2 maddeleri

L2 maddeleri	Tanım
LAND	Pikselin karada olması
CLOUD	Bulut kirliliği
ATMFAIL	Atmosferik düzeltme hatası
HIGLINT	Yüksek güneş ışığı
HILT	Diz seviyesinden yukarıdaki toplam radyans
HISATZEN	Yüksek uydu zirvesi
CLDICE	Bulutlar ve/veya buz
COCCOLITH	Kokolitoforid belirlenmesi
HISOLZEN	Yüksek güneş zirvesi
LOWLW	Çok düşük sudan ayrılma radyansı
CHLFAIL	Klorofil algoritması hatası
NAVWARN	Şüpheli navigasyon
MAXAERITER	Maksimum iterasyonlar
CHLWARN	Aralık dışı klorofil
ATMWARN	Şüpheli atmosferik düzeltme

Veriyi hdf biçiminden ascii biçimine dönüştürmek için SeaDAS görüntü analizi yazılımı (SeaDAS image analysis software) (SeaWifs Data Analysis System, version 6, <http://seadas.gsfc.nasa.gov/>) ve VMware Workstation 12 Player (<https://www.vmware.com/>) arayüzü kullanılmıştır.

Şüpheli ve düşük kaliteli verileri temizlemek için veri seti Çizelge 5.1 ile verilen ve L2 maddeleri (L2 flags) olarak adlandırılan özelliklere ilişkin bir kalite kontrol sürecinden geçirilmiştir. Bu temizleme süreci MATLAB 7.12.0-R2011a yazılımı kullanılarak yapılmış ve temizlenen veri tüm dalga boyları ve yıllar için ".mat" biçiminde kaydedilmiştir.

İlgilenilen çalışma alanının enlem ve boylamları arasındaki fark sırasıyla 2° ve 9°'dir. Her derece arasında 111.12 km uzaklık olduğu göz önüne alınırsa, bölgenin enlem ve boylamları arasındaki uzaklıklar sırasıyla yaklaşık olarak 222 km ve 999 km'dir. Görüntülerin çözünürlüğü 300 m olduğundan enlem için 740, boylam için 3330 piksel vardır. Yani, veri seti 740 × 3330 piksel görüntülerden oluşmaktadır. İndirme sürecinde her dalga boyu için kaydedilen gözlemler 2464200 pikselden oluşan vektörlerdir. Üretilen bu veri dosyalarının MATLAB 7.12.0 yazılımı kullanılarak düzenlenip analiz edilmesinden sonra ROI 36.01 – 37° N enlem ve 7 – 6.01° W boylamları ile belirtilen

alana indirgenerek MATLAB içindeki "m_map" araç kutusu kullanılarak bu alan görüntülenmiştir.

5.1.2 Denizden Toplanan Veri (In-situ Verisi)

Denizden toplanan veri, bir başka deyişle "in-situ" verisi Endülüs Hükümeti'ne ait "Junta de Andalucía" istasyonu ile "Reserva" ve "Fluctuaciones" istasyonları tarafından Guadalquivir nehir ağzından alınan örneklemelerden kaydedilen TSS değerlerinden oluşmaktadır. Analize alınan yüzey örneklemeleri "rosette örneklemcisi" ile su yüzeyinin 5 metre altından, sahilden 1 km'den 25 km'ye kadar değişen açıklıktan toplanmıştır.

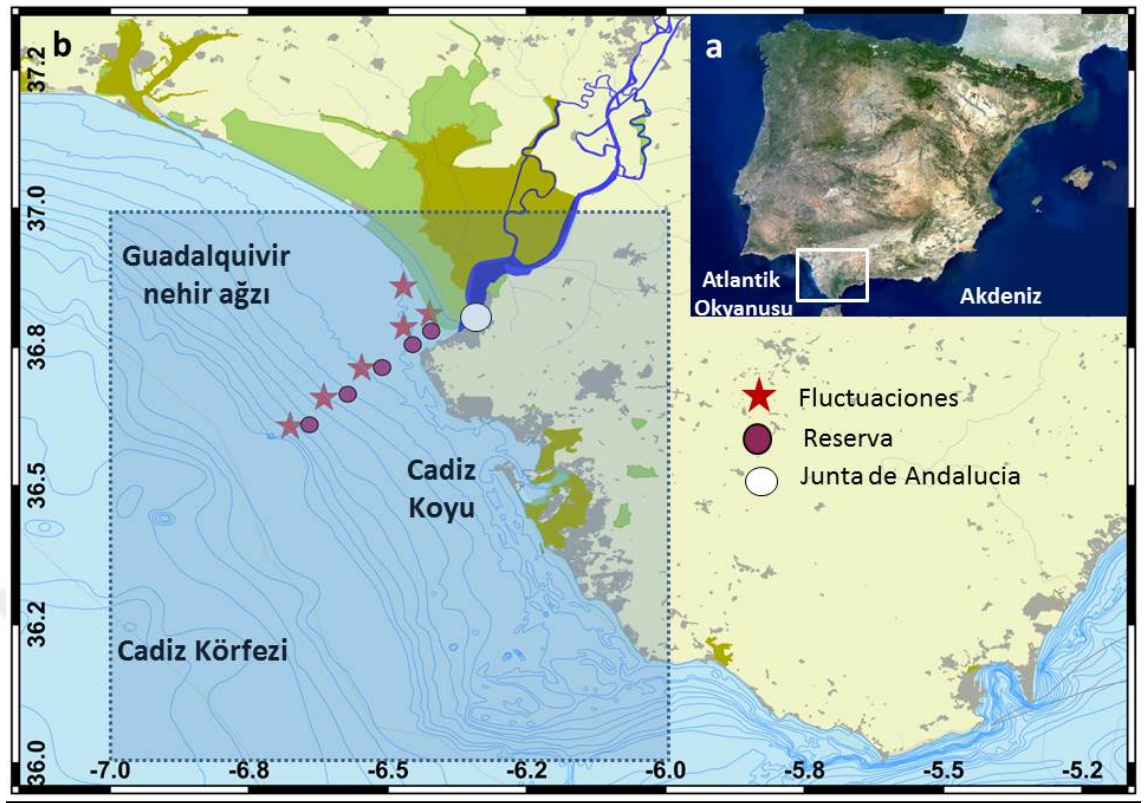
Örneklemeler değişik zaman periyotlarında toplanmıştır. "Junta de Andalucía" tarafından toplanan örneklemeler Nisan 2008 - Mayıs 2011 periyodunu kapsarken, "Reserva" ve "Fluctuaciones" gemileri tarafından toplanan örneklemeler sırasıyla Temmuz 2002 - Eylül 2004 ve Mayıs 2005 - Mayıs 2007 periyotlarını kapsamaktadır. Her bir örneklem sefer sırasında belirlenen bir koordinattan alınmıştır. Endülüs Hükümeti'ne ait istasyonun koordinatı 36.78° N enlem ve 6.37° W boylam ile sabittir. "Reserva" ve "Fluctuaciones" gemileri için ise koordinatlar sefer planlamasına göre seçilmiştir. Örneklem toplanan koordinatlar Şekil 5.1'de görüldüğü gibidir.

Toplanan örneklemelerdeki TSS yoğunluk miktarları Caballero ve diğ. (2014b)'te belirtilen protokoller kullanılarak ölçülmüştür.

5.1.3 Geçerlilik Aşaması

Geçerlilik aşamasında, uydudan elde edilen filtrelenmiş veri, "Junta de Andalucía", "Reserva" ve "Fluctuaciones" istasyonlarından toplanan veriler ile örneklemin toplanma zamanı dikkate alınarak eşleştirilmiştir. Bu aşamada, "in-situ" verilerinin ve uydu verilerinin koordinatları 2×2 piksellerin oluşturduğu bir kutu alan içinde yer alacak şekilde verinin toplandığı koordinata en yakın olan nokta seçilerek eşleştirilmiştir. Tüm eşleştirmeler MATLAB 7.12.0-R2011a programında yapılmıştır.

Eşleştirme sürecunda toplam 71 gözlem elde edilmiştir. Nehir ağzı sularının mekan-zamansal değişkenliğini düşürmek amacıyla yerinde veriler ile uydu verileri arasındaki zaman farkının en fazla 1,5 saat olduğu koşulu konulmuştur. Bu nedenle



Şekil 5.1: Çalışma alanı ve ilgilenilen alanı gösteren harita. (a) Çalışma alanı. (b) Guadalquivir nehir ağzı haritası ve ROI'nin yer aldığı Cadiz körfezi sahil kesimi. Pembe yıldızlar ve daireler sırasıyla "Fluctuaciones" and "Reserva" istasyonlarını ifade etmektedir. Beyaz daire ise "Junta de Andalucía" istasyonu konumudur.

gözlem sayısı 71'den 31'e düşmüştür. Dört gözlem su yüzeyinden toplanmadığı gerekçesiyle, bir gözlem filtreleme süreci sırasındaki ölçüm hatasından dolayı, bir gözlem de 13 nm ve 665 nm dalga boylarındaki eksik Rrs değerleri nedeniyle analizden çıkarılmıştır. Analizler kalan 25 gözlem üzerinden yapılmıştır.

5.1.4 İstatistiksel Yöntemler

TSS ve Rrs değerleri arasındaki ilişkiyi incelemek için pek çok klasik ve fonksiyonel istatistiksel yöntem kullanılmış ve bu yöntemlerin sonuçları karşılaştırılmıştır. FDRM'nin uydudan elde edilen tüm veriyi kullanma kapasitesine sahip olması ve spektral verinin yapı itibarıyla gözlemlerin noktalar yerine eğriler olarak alınmasına izin vermesi nedenleriyle diğer yöntemlerden daha iyi sonuç vereceği öngörülmektedir.

Verinin spektral yapısından ötürü, TSS ve Rrs arasındaki ilişkiyi analiz etmek için skaler yanıt değişkeni için FDRM kullanılmıştır. Burada, skaler yanıt değişkeni TSS

değerlerinden oluşan yanıt vektörünü, fonksiyonel açıklayıcı değişken ise sekiz farklı dalga boyunda kaydedilmiş Rrs değerlerini ifade etmektedir.

Genelde, TSS yoğunluğunu tahmin etmede Rrs 665 dalgaboyundan elde edilen veriler kullanılmaktadır (Binding ve diğ., 2003, 2005; Caballero ve diğ., 2014b). Nechad ve diğ. (2010) MERIS verisi ile TSS'i modellemek için 665 nm ve 681 nm dalgaboylarını kullanmayı önermiştir. Caballero ve diğ. (2014b)'nin çalışmasında TSS ve Rrs 665 arasındaki ilişkinin analizinde basit üstel regresyon modeli anlamlı bulunmuştur.

Bu çalışmada, en yüksek korelasyon 665 nm ve 681 nm dalgaboylarında görüldüğü için TSS değerleri ile 665 nm ve 681 nm dalga boylarındaki Rrs değerleri arasında ayrı ayrı basit üstel regresyon modeli kurulmuştur. Bunu takiben, yanıt değişkenini açıklayan optimum sayıda dalga boyunun seçilebilmesi için adımsal üstel regresyon modeli uygulanmıştır.

TSS'i modellemek için FDRM'ye ve üstel regresyon modellerine alternatif olarak, En Küçük Mutlak Daraltma ve Seçim Operatörü (Least Absolute Shrinkage and Selection Operator, LASSO) modeli ve farklı bağlantı fonksiyonları ve parametreler kullanılarak Genelleştirilmiş Toplamsal Modeller (GTM) uygulanmıştır. Modeller her seferinde bir gözlem dışarıda bırakılarak çapraz geçerlilik (LOOCV) olarak da bilinen Jackknife yöntemiyle hesaplanan düzeltilmiş Hata Tahmin Ortalaması (Adjusted Mean Error of Prediction, AMEP) kullanılarak karşılaştırılmıştır.

2-katlı çapraz geçerlilik için MEP şu şekilde ifade edilir:

$$MEP = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / n}{\text{Var}(y)} = \frac{\text{SSE}}{\text{Var}(y)}. \quad (5.1)$$

Burada, y tahmin edilecek yanıt değişkenini göstermektedir (Febrero-Bande ve Oviedo de la Fuente, 2012).

LOOCV durumunda, her adımda sadece tek bir gözlem dışarıda bırakıldığı için bir gözlemlilik veri setinde varyans hesaplanamaz. Bundan dolayı, modelleri karşılaştırmada MEP'in düzeltilmiş bir ölçüsü olan AMEP kriteri kullanılabilir. AMEP, (5.2) eşitliğindeki gibi tanımlanır:

$$AMEP = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / n}{\sum_{i=1}^n (y_i - \bar{y}_{-i})^2} = \frac{\text{SSE}}{\sum_{i=1}^n (y_i - \bar{y}_{-i})^2}. \quad (5.2)$$

Bu eşitliğin paydasında bulunan $\sum_{i=1}^n (y_i - \bar{y}_{-i})^2$ ifadesi sabit bir değerdir ve farklı modellere ait MEP değerlerini ölçeklemek için kullanılmıştır.

Tüm istatistiksel analizler Windows için R v3.2.3 istatistiksel yazılım programı kullanılarak yapılmıştır.

5.1.4.1 Genelleştirilmiş Toplamsal Modeller (GTM)

GTM ilk olarak Hastie ve Tibshirani (1986) tarafından tanıtılmıştır. GTM, Genelleştirilmiş Doğrusal Modellerin (GDM) parametrik olmayan bir uzantısıdır ve yanıtın beklenen değerinin bir fonksiyonunu açıklayıcı değişkenlerin parametrik ya da parametrik olmayan fonksiyonlarının toplamları cinsinden modellemeye yarar (Hothorn ve Everitt, 2014). Bu model, yanıt değişkeni üstel dağılım ailesi üyesi bir dağılımdan gelmek üzere, eşitlik (5.3) ile ifade edilir (Wood, 2006):

$$g(\mathbb{E}[Y|X_1, X_2, \dots, X_p]) = \alpha + f(X_1) + f(X_2) + \dots + f(X_p). \quad (5.3)$$

Bağlantı fonksiyonu g , yanıtın dağılımına bağlı olarak birim, lojit, probit ya da log bağlantı fonksiyonu olabilir. Birim bağlantı fonksiyonu genellikle normal dağılımdan gelen yanıt değişkenleri için kullanılırken log bağlantı fonksiyonu Poisson dağılımlı sayma verileri için kullanılır. İki terimli olasılıkları modellemek için ise lojit veya probit bağlantı fonksiyonları tercih edilir (Friedman ve diğ., 2001; Hastie ve diğ., 2011).

GTM, genelleştirilmiş parametrik olmayan çoklu regresyon modeli ile GDM arasında yer alır. Çok sayıda açıklayıcı değişken olması durumunda, parametrik olmayan regresyon modellerinde "boyut laneti (curse of dimensionality)" adı verilen ve bir t noktasının komşuluğunun hiç bir gözlem içermemesini ifade eden bir problem ile karşılaşılabilir (Delicado, 2015). Verinin büyük olması durumunda, GTM parametrik olmayan çoklu regresyon modellerine bir alternatiftir.

Eğer eşitlik (5.3)'in sağ yanındaki fonksiyonlar doğrusal ise GTM, GDM'ye denktir. Ayrıca modelin sağ yanındaki fonksiyonların bazılarının yanıt üzerinde doğrusal bir etkisi varken diğerlerinin yoksa o zaman bu modeller "yarı-parametrik model" olarak adlandırılır. Yarı-parametrik modeller parametrik olmayan modellere göre daha esneklerdir.

5.1.4.2 En Küçük Mutlak Daraltma ve Seçim Operatörü (Least Absolute Shrinkage and Selection Operator, LASSO)

Bir çoklu doğrusal regresyon probleminde açıklayıcı değişkenlerin sayısı p , gözlem sayısı n 'den daha büyükse, büyük boyutluluktan kaynaklanabilecek problemleri engellemek için kestirimin farklı yöntemler kullanılarak cezalandırılması önerilir. Değişken seçimi amacıyla kullanılan bu yöntemlerden biri LASSO'dur. LASSO kestirimi, cezalı HKT'nin L_1 normu ile cezalandırılıp, $\sum_{j=1}^p |\beta_j| \leq t$ koşulu altında en küçüklenmesiyle elde edilir (Hastie ve diğ., 2011).

$$\sum_{i=1}^N \left(Y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p |\beta_j|, \quad (5.4)$$

Burada t "tunning" parametresi olarak adlandırılır. Yorum kolaylığı açısından, değişkenler standardize edilmiş "tunning" parametrelerine, $s = \frac{t}{\sum_{j=1}^p |\beta_j|}$, karşı çizdirilir. Optimum s değeri çapraz geçerlilik kullanılarak belirlenir. Optimum s değeri için, değer alan değişkenler analize alınır.

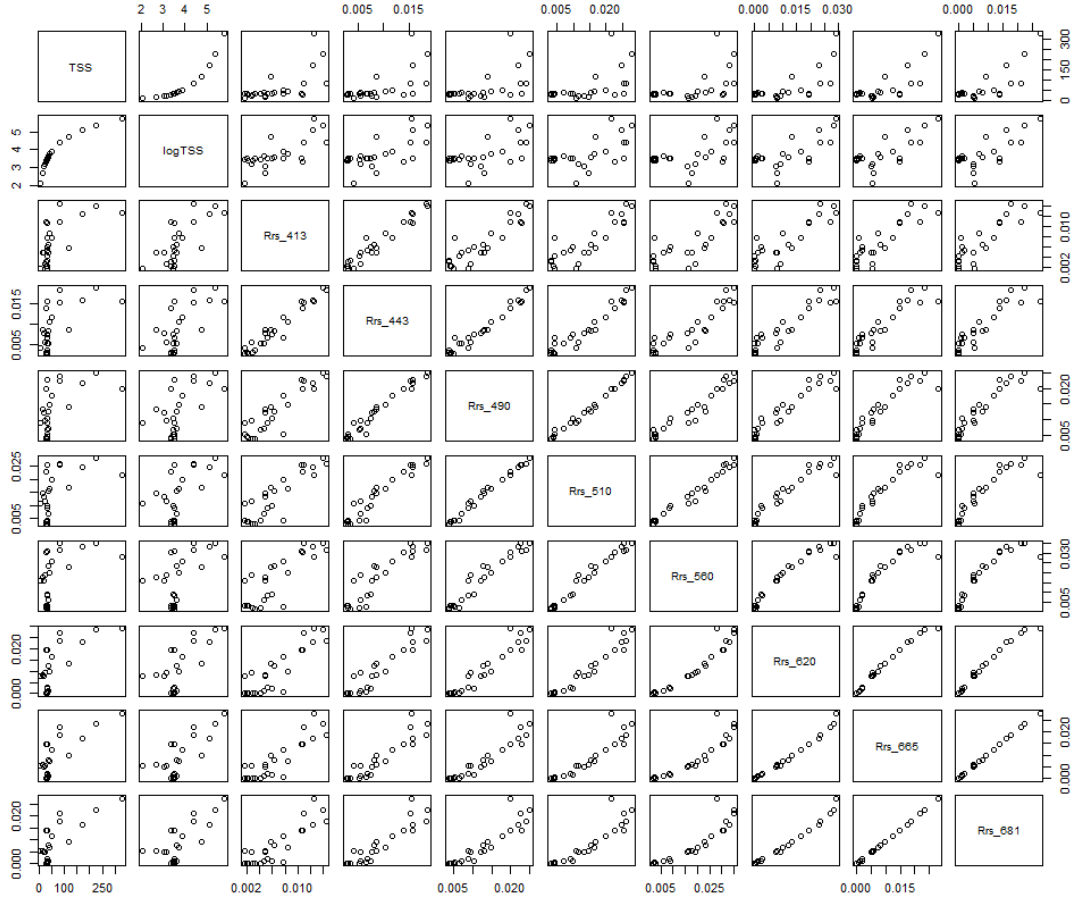
5.1.5 Sonuçlar

Veri seti üç farklı istasyondan toplanmış 25 gözlemden oluşmaktadır. 14 gözlem "Junta de Andalucía" istasyonundan, 7 gözlem "Reserva" gemisi seferlerinden, 4 gözlem ise "Fluctuaciones" gemisi seferlerinden alınmıştır.

Şekil 5.2'deki korelasyon çizimleri dalga boyları arasında yüksek derecede ilişki olduğunu ve yerinde TSS değerleri ile her dalga boyundaki Rrs değerleri arasında üstel bir ilişki olduğunu göstermektedir.

5.1.5.1 Üstel Regresyon Modelleri

Üstel regresyon modelinde hangi dalga boyunun kullanılacağına karar vermek için TSS değişkeninin logaritması ile dalgaboylarında kaydedilen Rrs değerleri arasındaki korelasyon değerleri dikkate alınmıştır. TSS'nin logaritması ile 413 nm, 443 nm, 490 nm, 510 nm, 560 nm, 620 nm, 665 nm, 681 nm dalga boyları arasındaki korelasyonlar sırasıyla 0.723 ($p < 0.001$), 0.671 ($p < 0.001$), 0.596 ($p = 0.002$), 0.567 ($p = 0.003$), 0.533 ($p = 0.005$), 0.681 ($p < 0.001$), 0.729 ($p < 0.001$) ve 0.734 ($p < 0.001$) olarak



Şekil 5.2: Değişkenler arasındaki korelasyon çizimleri

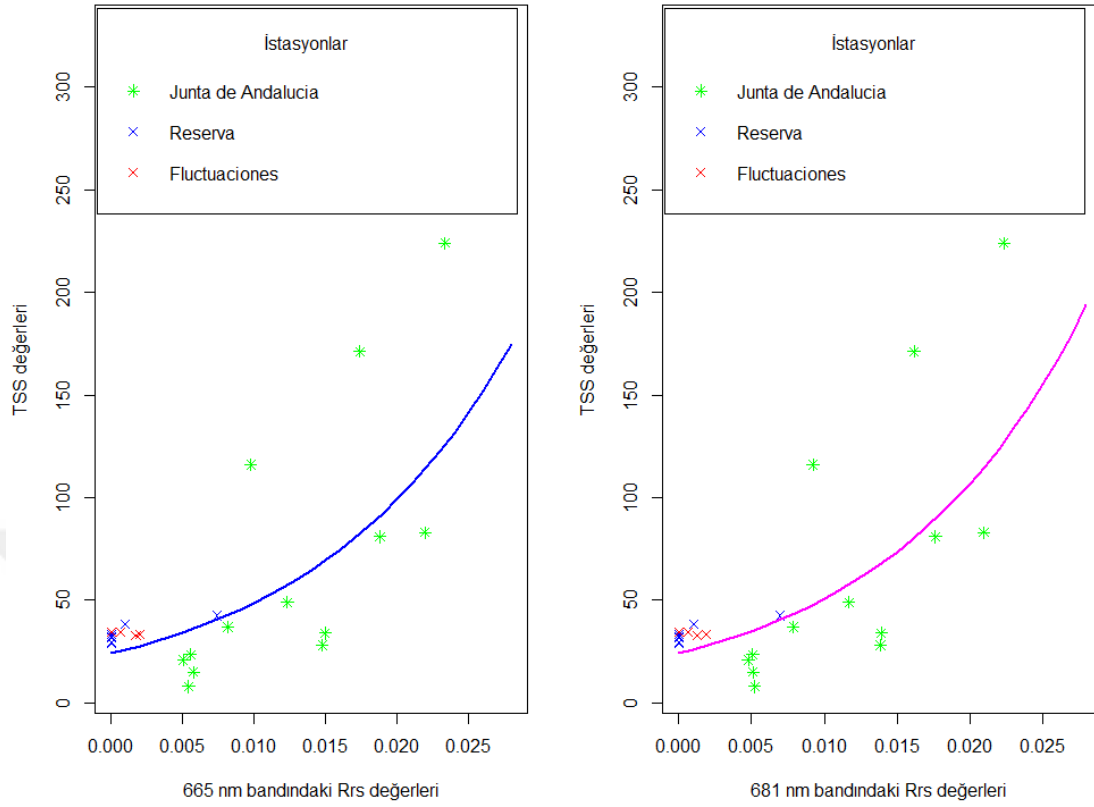
bulunmuştur. En yüksek korelasyonlar 665 nm ve 681 nm dalga boylarında görüldüğü için TSS ve bu iki dalga boyundan elde edilen Rrs değerleri arasında basit üstel regresyon modeli kurulmuştur.

Yerinde TSS ölçümleri 3 - 327 mg/L aralığında değerler alırken 681 nm dalga boyundaki Rrs değerleri 0.000 - 0.0275 sr^{-1} aralığında 665 nm dalga boyundaki Rrs değerleri 0.000 - 0.028 sr^{-1} aralığında değerler almaktadır. Şekil 5.3'de verilen dağılım grafiğinden de görüldüğü gibi, Rrs değerlerinin yayılımı her iki bantta da benzerdir.

681 nm bandı için kurulan basit üstel regresyon modeli,

$$\text{TSS} = 24.14 * \exp(74.38 * \text{Rrs } 681), \quad (5.5)$$

olarak elde edilmiştir. Açıklanan varyans oranı 0.554, standart hata 0.57'dir ve model anlamlıdır ($p < 0.001$).



Şekil 5.3: In-situ TSS değerleri ile 665 nm ve 681 nm bandlarındaki Rrs değerleri arasındaki saçılım grafiği. Mavi eğri TSS ile 665 nm dalga boyundaki Rrs değerleri arasındaki ilişkiyi, pembe eğri ise TSS ile 681 nm dalgaboyundaki Rrs değerleri arasındaki ilişkiyi gösterir.

TSS değerleri ile ikinci en yüksek korelasyona sahip 665 nm dalgaboyu için kurulan basit üstel regresyon modeli,

$$TSS = 24.04 * \exp(70.83 * Rrs\ 665), \quad (5.6)$$

olarak elde edilmiştir. Bu modelin R^2 değeri 0.53, standart hatası ise 0.57'dir.

Bu modelleri takiben TSS değerlerinin logaritması ile Rrs değerleri arasında adımsal doğrusal regresyon modeli uygulanmıştır. Sonuç olarak, eşitlik (5.7)'deki AIC kriteri ile 413 nm ($p = 0.121$), 560 nm ($p = 0.003$) ve 620 nm ($p = 0.003$) dalgaboylarının analize alındığı model anlamlı bulunmuştur ($p < 0.001$). Bu modele ait belirtme katsayısı 0.69, düzeltilmiş belirtme katsayısı 0.65 ve model standart hatası 0.49'dur.

$$\log \text{TSS} = 3.42 + 82.90 * \text{Rrs } 413 - 100.46 * \text{Rrs } 560 + 142.92 * \text{Rrs } 620. \quad (5.7)$$

Bu modellerin yanısıra iki parametrelili üstel doğrusal regresyon modelleri de denenmiştir. 620 nm ve 681 nm dalgaboylarının birlikte kullanıldığı üstel regresyon modeli ile 620 nm ve 665 nm dalgaboylarının birlikte kullanıldığı üstel regresyon modelleri anlamlı bulunmuştur. Ancak bu modellerin düzeltilmiş belirtme katsayısı adimsal regresyon modeline göre daha düşük bulunmuştur.

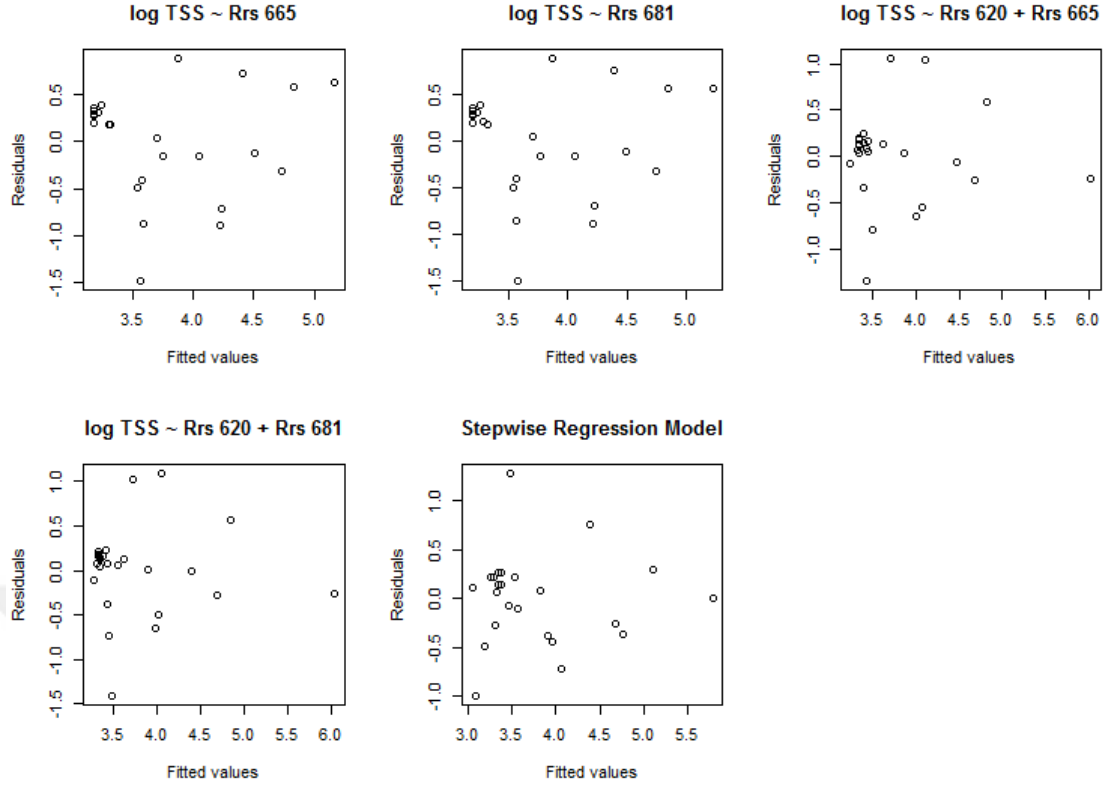
Bahsedilen tüm modellere ait standart hata, R^2 , Düzeltilmiş R^2 (Adj. R^2) ve AMEP değerleri Çizelge 5.2’de görüldüğü gibidir.

Çizelge 5.2: Üstel Regresyon Modelleri için AMEP Değerleri

Model	Std Hata	Adj R^2	AMEP
$\log \text{TSS} = b_0 + b_1 \text{ Rrs } 681$	0.57	0.52	0.46
$\log \text{TSS} = b_0 + b_1 \text{ Rrs } 665$	0.58	0.51	0.48
$\log \text{TSS} = b_0 + b_1 \text{ Rrs } 620 + b_2 \text{ Rrs } 681$	0.53	0.58	5.07
$\log \text{TSS} = b_0 + b_1 \text{ Rrs } 620 + b_2 \text{ Rrs } 665$	0.53	0.59	2.98
$\log \text{TSS} = b_0 + b_1 \text{ Rrs } 413 + b_2 \text{ Rrs } 560 + b_3 \text{ Rrs } 620$	0.50	0.65	1.37

AMEP değerlerine göre, tek açıklayıcı değişkenli (sadece 665 nm ya da sadece 681 nm bandı kullanılan) basit üstel regresyon modelleri daha uygun görünmektedir. Bu modellerin standart hatası ve açıklanan varyans oranı daha düşük olmasına rağmen TSS değişkenini tahmin etmek için daha uygundur. Öncelikle çoklu bağlantı probleminden etkilenmezler. İkinci olarak, Şekil 5.4 ile verilen artık grafiklerinden de görüldüğü gibi diğer üstel regresyon modelleri varyans homojenliği varsayımını sağlamamaktadır.

Adimsal regresyon modeline ek olarak değişkenler arası korelasyonu engellemek ve boyut indirgemek amacıyla veri setine sırasıyla Temel Bileşenler Regresyon (TBR) ve Kısmi En Küçük Kareler Regresyon (KEKKR) modelleri de uygulanmıştır. Her iki model için de optimum bileşen sayısı çapraz geçerlilik yöntemiyle belirlenmiştir. Tüm bileşenleri içeren sıradan TBR ve KEKKR modellerine ait AMEP değerleri ile optimum bileşen sayısının modellemede kullanıldığı çapraz geçerlilik (Cross Validated, CV) uygulanan TBR ve KEKKR modellerine ait AMEP değerleri Çizelge 5.3’te verilmektedir.



Şekil 5.4: Üstel regresyon modellerine ait artık grafikleri

Çizelge 5.3: TBR ve KEKKR Modellerine Ait AMEP Değerleri

Model	MSE	R ²	AMEP
TBR	0.19	0.70	3.91
KEKKR	0.19	0.70	3.40
CV yöntemiyle seçilen 2 bileşenle TBR	0.29	0.55	0.86
CV yöntemiyle seçilen 2 bileşenle KEKKR	0.23	0.64	0.90

Buna göre CV kriteri ile belirlenen bileşenlerle kurulan regresyon modelleri yanıt tüm bileşenlerle kurulan regresyon modelinden daha iyi tahmin etmektedir. Ancak bu durumda klasik TBR, klasik KEKKR yöntemine göre daha iyi sonuç vermektedir.

5.1.6 GTM and LASSO Modelleri

Çalışmada, TSS değişkenini modellemek için temelde üç farklı GTM kullanılmıştır. İlk olarak, Normal dağılım ve birim bağlantı fonksiyonu kullanılarak TSS değerlerinin logaritması 665 nm ve 681 nm dalga boylarındaki Rrs değerleri cinsinden modellenmiştir. Bu şekilde kullanılan basit üstel regresyon modellerinden daha esnek

bir model elde edilmesi amaçlanmıştır. İkinci olarak, bağımlı değişken TSS olarak alınmış ve modellemede ters bağlantı fonksiyonu ile Gamma dağılımı kullanılmıştır. Bunun yanısıra, tek parametrelili modellere alternatif olarak TSS değerlerindeki değişimi en iyi açıklayan modelin bulunması için iki parametrelili GTM modelleri de araştırılmıştır.

İlgili GTM'ye ait düzeltilmiş R^2 , açıklanan sapma (Deviance Explained) yüzdesi ve AMEP değerleri Çizelge 5.4'deki gibi özetlenebilir:

Çizelge 5.4: GTM için AMEP Değerleri

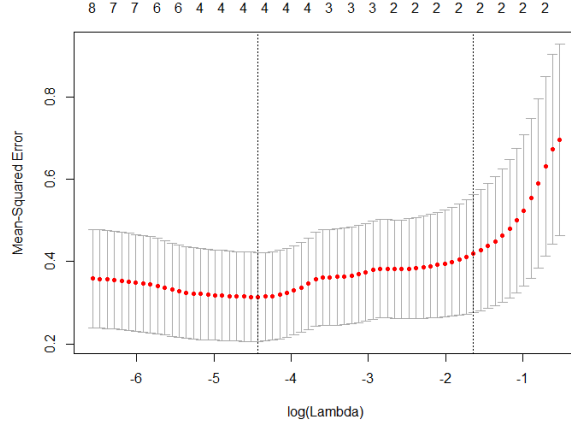
Model	Dağılım Ailesi	Bağlantı Fonksiyonu	Adj R^2	Açıklanan Sapma Yüzdesi	AMEP
$\log TSS \sim s(\text{Rrs } 665)$	Normal	Birim	0.89	93.2	966.01
$\log TSS \sim s(\text{Rrs } 681)$	Normal	Birim	0.87	92.0	1866.72
$TSS \sim s(\text{Rrs } 620) + s(\text{Rrs } 665)$	Normal	log	0.88	89.0	1.53
$TSS \sim s(\text{Rrs } 620) + s(\text{Rrs } 681)$	Normal	log	0.85	90.6	1.53
$TSS \sim s(\text{Rrs } 665)$	Gamma	ters	0.73	69.0	1.63
$TSS \sim s(\text{Rrs } 681)$	Gamma	ters	0.74	69.0	1.63

TSS'nin logaritması bağımlı değişken olarak alınıp LASSO modeli kurulmuştur. Analize alınması gereken değişkenleri belirlemek için 10-katlı CV yönteminden elde edilen optimum λ değeri kullanılmıştır. Bu değer iki şekilde bulunabilir: ya ortalama CV hatasını minimum yapan değer seçilerek ya da 1 standart hata içinde kalan hatanın maksimumu alınarak.

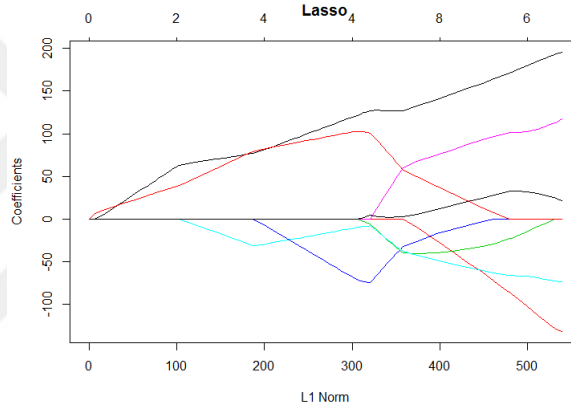
Şekil 5.5'te gösterildiği gibi ortalama CV hatasına bağlı olarak seçilen λ değeri 1.01 olarak bulunmuş ve buna göre analize sokulması gereken değişken sayısı 4 olarak belirlenmiştir. Standart hataya bağlı olarak belirlenen λ değeri ise 1.21 olarak bulunmuş ve bu değer için değişken sayısı 2 olarak belirlenmiştir. Söz konusu şeklin üst kısmında λ 'nın değişik değerleri için analize sokulması gereken değişken sayısı verilmiştir.

L_1 normuna göre aşama aşama analize sokulan değişkenler Şekil 5.6'dan görülebilir.

LASSO modeline ait optimum parametre sayısı CV kriteri kullanılarak minimum 2, maksimum 4 olarak belirlenmiş ve AMEP değeri 33.71 olarak hesaplanmıştır. Bu değer, kullanılan diğer modellere (TBR, KEKKR, GTM) göre daha yüksek



Şekil 5.5: Bir dizi λ değeri için hata kareler ortalaması. Şeklin sol tarafındaki kesikli çizgi CV hatasına bağlı λ değerlerini gösterirken şeklin sağ yanındaki kesikli çizgi standart hataya bağlı λ değerlerini göstermektedir.



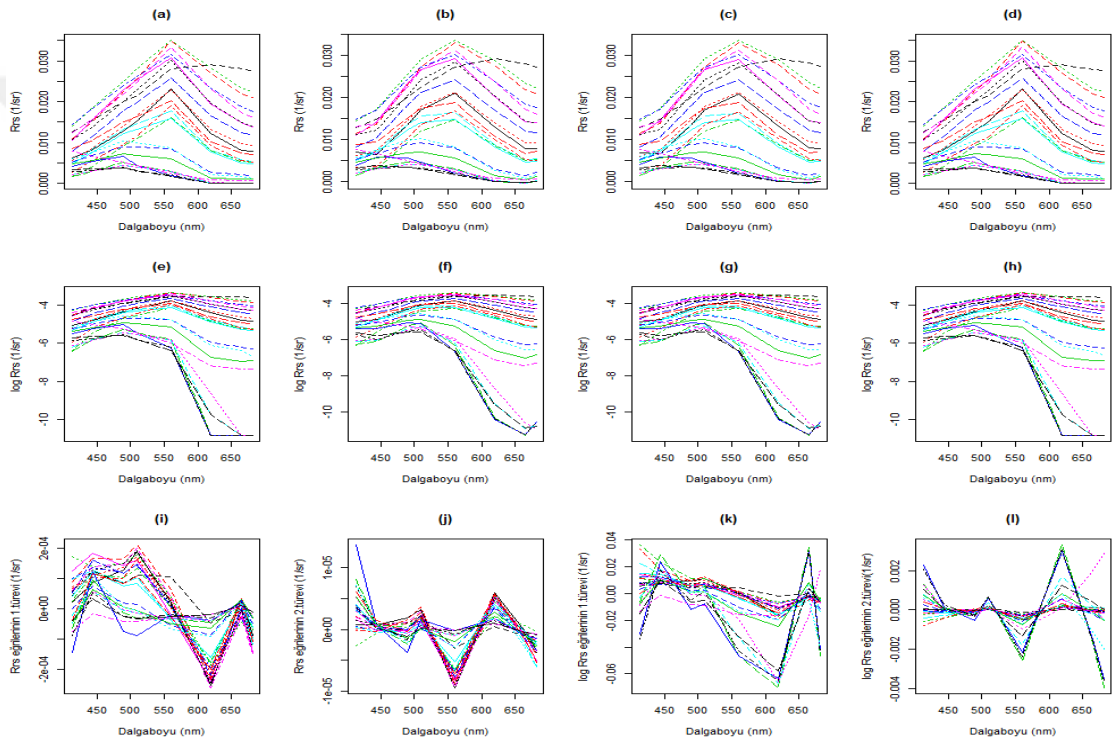
Şekil 5.6: LASSO modeli. Bir dizi L_1 norm değeri için analizdeki değişken sayısı. x-ekseninde norm değeri görülmektedir. Her bir eğri bir değişkeni temsil etmektedir.

olduğundan LASSO modelinin TSS değerlerinin tahmin edilmesinde iyi bir seçenek değildir.

5.1.7 Fonksiyonel Doğrusal Regresyon Modelleri

Skaler yanıt değişkenini modellemek için kullanılan modeller B-spline taban fonksiyonları aracılığıyla FDRM, tüm bileşenler ile FTBR, çapraz geçerlilik yöntemi ile FTBR (FTBR CV), tüm bileşenleri içeren FKEKKR, çapraz geçerlilik yöntemi ile FKEKKR (FKEKKR CV) ve Rrs eğrilerinin birinci ve ikinci dereceden türevlerinin açıklayıcı değişken olarak kullanıldığı FDRM'dir. Tüm modeller (5.2) eşitliğinden elde edilen AMEP değerleri kullanılarak karşılaştırılmıştır. İlk olarak, skaler yanıt

değişkeni olarak ham TSS değerleri alınmış ve TSS ile Rrs eğrileri arasında FDRM kurulmuştur. Daha sonra TSS değerlerinin logaritması yanıt değişkeni olarak ele alınmış ve sırasıyla Rrs eğrileri, Rrs eğrilerinin logaritması ve bu eğrilerin birinci ve ikinci türevleri üzerine fonksiyonel regresyon analizi yapılmıştır. Fonksiyonel modellemelerin gerçekleştirilmesi için öncelikle eğriler düzleştirilmiştir. Düzleştirme için sırasıyla optimum taban sayısı CV kriteri ile belirlenen B-splayn taban yaklaşımı, optimum taban sayısı ve cezalandırma parametresi CV kriteri ile belirlenen cezalı B-splayn taban yaklaşımı ve Nadaraya-Watson çekirdek yöntemleri kullanılmıştır. Bu yöntemlerle düzleştirilmiş eğriler Şekil 5.7 'deki gibidir.



Şekil 5.7: Düzleştirilmiş eğriler. (a) Ham Rrs eğrileri. (b) Optimum taban sayısı CV yöntemiyle belirlenen B-splayn yöntemi ile düzleştirilmiş Rrs eğrileri. (c) Optimum taban sayısı ve ceza parametresi CV yöntemiyle belirlenen cezalandırılmış B-splayn yöntemi ile düzleştirilmiş Rrs eğrileri. (d) Nadaraya-Watson yöntemi ile düzleştirilmiş Rrs eğrileri. (e) Rrs eğrilerinin logaritması (log Rrs eğrileri). (f) log Rrs eğrilerinin optimum taban sayısı CV yöntemiyle belirlenen B-splayn yöntemi ile düzleştirilmiş hali. (g) log Rrs eğrilerinin optimum taban sayısı ve ceza parametresi CV yöntemiyle belirlenen cezalandırılmış B-splayn yöntemi ile düzleştirilmiş hali. (h) Nadaraya-Watson yöntemi ile düzleştirilmiş log Rrs eğrileri. (i) Rrs eğrilerinin 1.türevinin 8 tabanlı B-splayn kullanılarak düzleştirilmiş hali. (j) Rrs eğrilerinin 2.türevinin 8 tabanlı B-splayn kullanılarak düzleştirilmiş hali. (k) log Rrs eğrilerinin 1.türevinin 8 tabanlı B-splayn kullanılarak düzleştirilmiş hali. (l) log Rrs eğrilerinin 1.türevinin 8 tabanlı B-splayn kullanılarak düzleştirilmiş hali.

Buna göre farklı yöntemlerle düzleştirilmiş eğriler benzerlik göstermektedir. Bu nedenle, hesaplama kolaylığından dolayı Rrs eğrilerinin düzleştirilmesinde 8 tabanlı B-splayn düzleştirme yöntemi tercih edilmiştir. Rrs ve log Rrs eğrilerinin düzleştirilmesinde de taban sayısı 8 olan B-splayn yöntemi kullanılmıştır.

Veri setindeki fonksiyonel aykırı değerlerin belirlenebilmesi için Febrero-Bande ve Oviedo de la Fuente (2012) tarafından önerilen ağırlıklandırma tekniğine ve bootstrap yöntemlerine dayalı bir yöntem uygulanmıştır. Bu yöntem 4 farklı derinlik ölçüsü üzerinden uygulanmıştır: Fraiman-Muniz derinliği (FMD), Modal Derinliği (MD), Rastgele Tukey Derinliği (RTD), Rasgele Yansıma Derinliği (Random Projection, RPD). Bootstrap örneklemeden elde edilen kesim değerini belirlemek için Febrero-Bande ve Oviedo de la Fuente (2012) tarafından önerildiği gibi bootstrap örneklem sayısı 200, kantil değeri ise 0.5 olarak alınmıştır. Veri setinde aykırı değer olmasından kuşkulanan gözlemler olmasına rağmen, hiçbir gözlem aykırı değer olarak belirlenmemiştir. Analizler 25 gözlem üzerinden yapılmıştır.

İlgili FDRM'ye ait AMEP değerleri Çizelge 5.5'te verildiği gibidir.

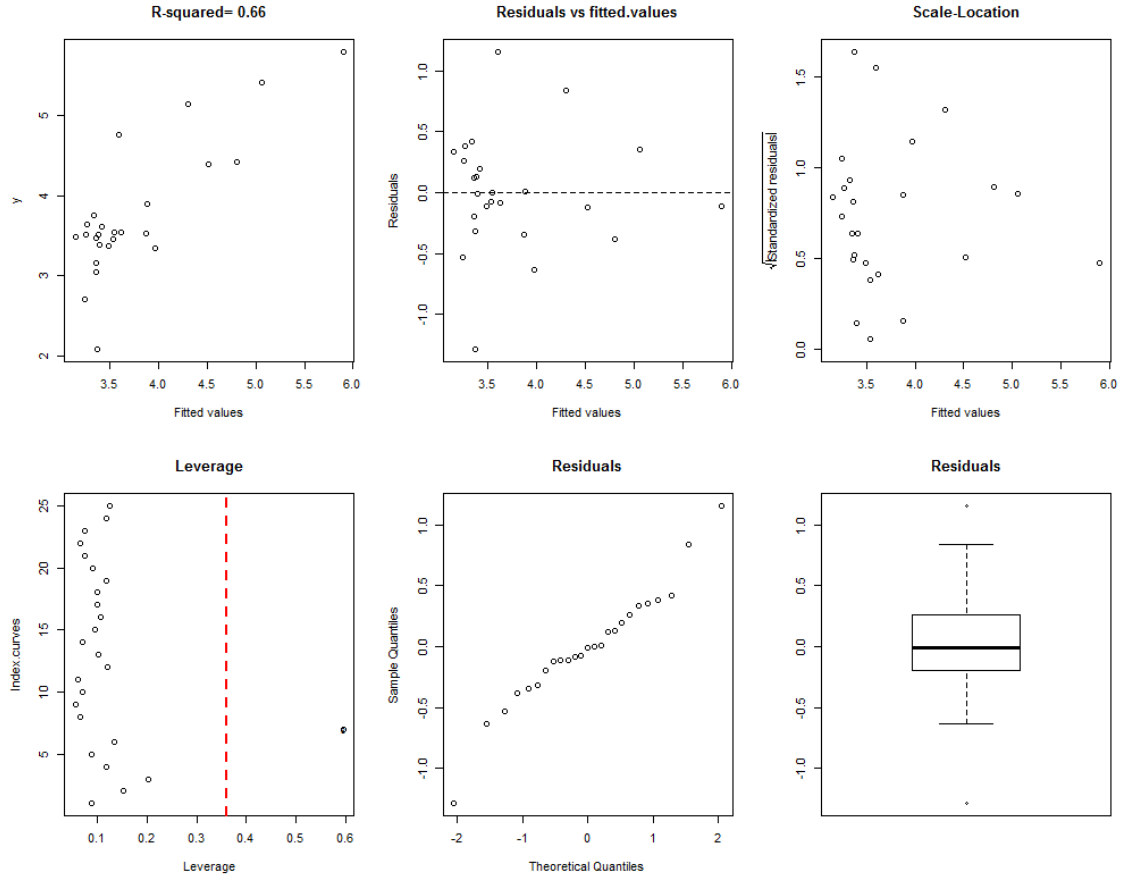
Çizelge 5.5: FDRM için AMEP değerleri

MODEL	TSS ~ Rrs	log TSS ~ Rrs	log TSS ~ log Rrs
B-Splayn Tabanı	0.34	0.43	0.77
FTBR	0.64	0.36	0.85
FTBR CV	0.34	0.22	0.75
FKEKKR	0.64	0.36	0.85
FKEKKR CV	0.33	0.24	0.79
FDRM 1.türev	0.57	0.64	0.96
Cezalandırılmış FDRM 2.Türev	0.54	0.63	0.97
FDRM 2.türev	0.57	0.39	1.02
Cezalandırılmış FDRM 2.türev	0.51	0.33	0.92

Çizelge 5.5'te verilen tüm modeller anlamlı bulunmuştur ($p < 0.05$). En yüksek AMEP değerleri genelde log TSS ve log Rrs arasında kurulan FDRM'den elde edilmiştir. Bu nedenle, TSS değerini Rrs eğrileri üzerinden tahmin etmenin uygun olmadığı söylenebilir. Öte yandan, en düşük AMEP değerleri B-splayn taban yöntemi ve birinci dereceden türevin açıklayıcı değişken olarak ele alındığı modeller dışında TSS değerlerinin logaritması ile Rrs eğrileri arasında kurulan FDRM'den elde edilmiştir. Bu yöntemler, TSS değerlerinin Rrs eğrileri üzerinden modellenmesinde daha iyi

sonuç vermiştir. Genel olarak bakıldığında, çapraz geçerlik yöntemine dayalı FTBR ve FKEKKR daha düşük AMEP değerleri vermiştir. En iyi tahminler ise, "log TSS \sim Rrs" arasında kurulan ve sırasıyla 0.22 ve 0.24 AMEP değerlerine sahip FTBR CV ile FKEKKR CV yöntemlerinden elde edilmiştir. CV kriteri tarafından her iki yöntem için de iki bileşen seçilmiştir. FTBR için beşinci ve ikinci temel bileşenler alınırken, FKEKKR için ilk iki temel bileşen analize alınmıştır.

FTBR CV modeline ait belirtme katsayısı (R^2) ve standart sapma değerleri sırasıyla 0.66 ve 0.50 olarak bulunmuştur. Şekil 5.8'in üst orta grafiğinden de görüldüğü gibi heterojen varyanslılık söz konusudur. Bunun temel nedeni az sayıda gözlemle çalışılmasıdır. Şekil 5.8'in alt ortasında yer alan uç değerleri gösteren "Leverage" grafiğine bakılırsa, 8 numaralı eğrinin olası bir etkin gözlem olduğu görülür. Artık dağılımını gösteren Q-Q plot ve kutu çizimlerinden de görüleceği gibi artıklar hafif çarpık olmakla birlikte yaklaşık olarak normal dağılmaktadır ve iki uç değer gözlenmektedir.



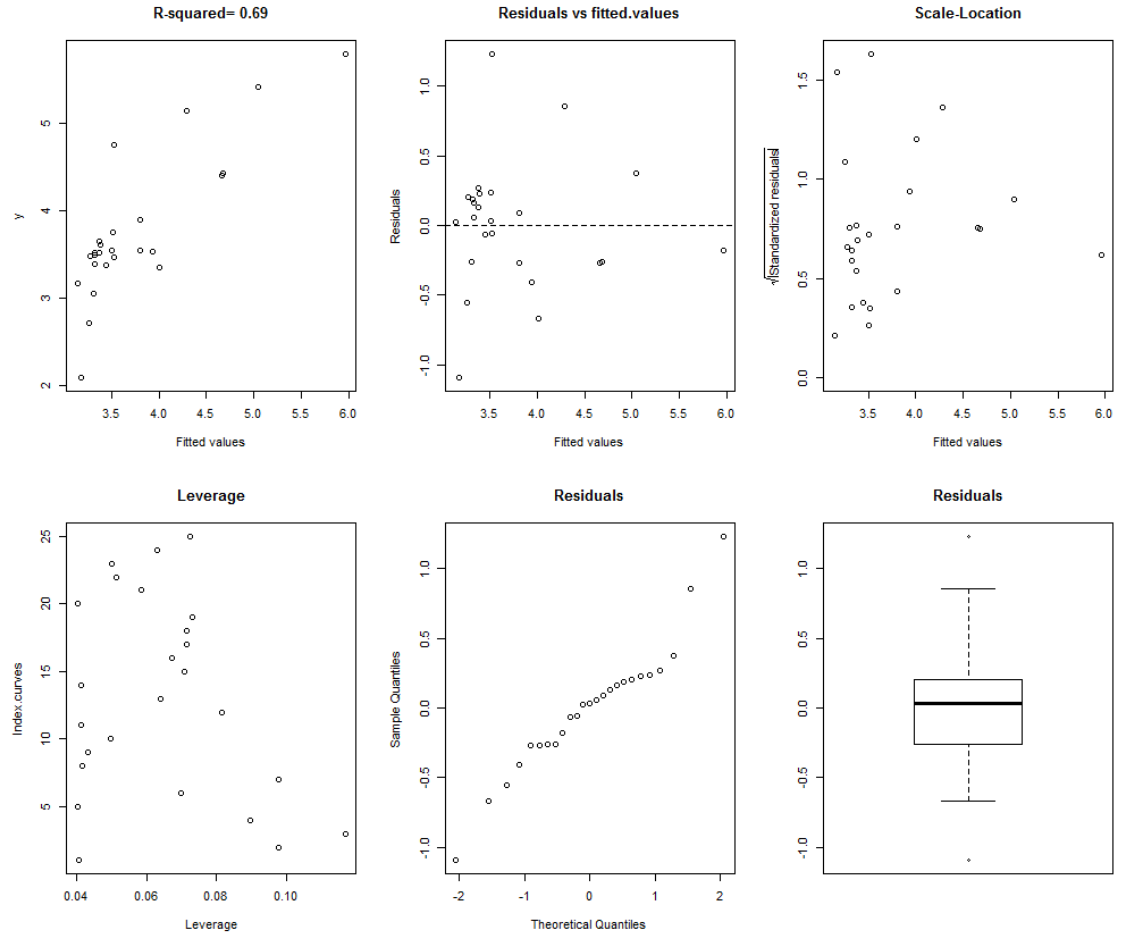
Şekil 5.8: log TSS and Rrs arasındaki FTBR CV modeline ait özet grafikler.

CV yöntemi ile seçilen beşinci ve ikinci fonksiyonel temel bileşenler sırasıyla toplam varyansın yüzde 4.3'ünü ve yüzde 6.8'ini açıklamaktadır. Bu bileşenlere ait parametre değerleri, standart hata, t istatistikleri ve p değerleri Çizelge 5.6'da verilmiştir.

Çizelge 5.6: FTBR CV modelinin Parametre Kestirimleri

Parametre	Kestirim	Standart Hata	t istatistik değ.	p değeri
Sabit terim	3.76	0.10	37.4	< 0.001
FTB 5	-24.05	3.92	-6.13	< 0.001
FTB 2	-8.35	3.12	-2.68	0.013

FKEKKR CV modeline ait belirtme katsayısı değeri ve model standart sapması ise sırasıyla 0.69 ve 0.23 olarak bulunmuştur. Bu modele ait R^2 değeri FTBR CV modeline ait R^2 değerinden biraz daha yüksek, standart sapma değeri ise daha düşüktür. FKEKKR CV modeline ait özet grafikleri Şekil 5.9 ile verilmiştir.



Şekil 5.9: log TSS and Rrs arasındaki FKEKKR CV modeline ait özet grafikler.

Şekil 5.9'un alt solunda verilen "Leverage" grafiğinden görüldüğü gibi hiçbir aykırı ya da uç değer bulunmamıştır. Artık değerlerine ait Q-Q plot ve kutu grafikleri FTBR CV modeline benzer şekildedir.

FKEKKR CV modelince seçilen birinci ve ikinci temel fonksiyonel kısmi en küçük kareler bileşenlerine ait parametre kestirimleri ve standart hatalar Çizelge 5.7'de verilmiştir.

Çizelge 5.7: FKEKKR CV modeline ait parametre kestirimleri

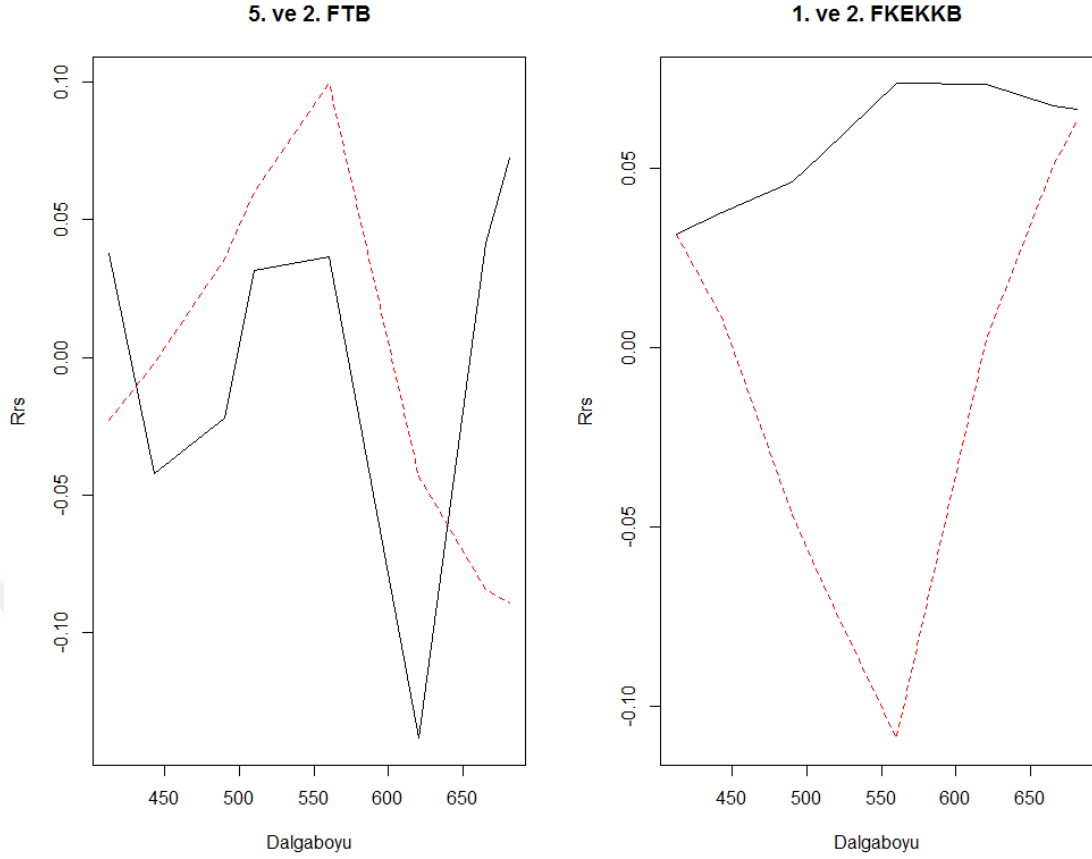
Parametre	Kestirim	Standart Hata	t istatistik değ.	p değeri
Sabit Terim	3.76	0.09	38.89	< 0.001
FKEKKB 1	0.12	0.005	23.89	< 0.001
FKEKKB 2	0.20	0.001	178.77	< 0.001

Çizelge 5.6 ve Çizelge 5.7'de görüldüğü gibi hem FTBR modeli için hem de FKEKKR modeli için CV kriterince seçilen fonksiyonel bileşenler anlamlı bulunmuştur. Seçilen FTBR ve FKEKKR bileşenleri Şekil 5.10'da çizilmiştir.

Şekil 5.10'un sol panelindeki grafik seçilen fonksiyonel temel bileşenleri göstermektedir. Kırmızı eğri ile belirtilen ikinci fonksiyonel temel bileşen 550 nm dalgaboyuna daha fazla ağırlık verirken en düşük ve en yüksek dalgaboylarına daha düşük ağırlık vermektedir. Siyah eğri ile gösterilen beşinci fonksiyonel temel bileşen ise dalgalı bir yapıya sahip olduğundan yorumlanması zordur. Sağ panelde gösterilen FKEKKR bileşenlerinden siyah eğri ile gösterilen birinci FKEKKR bileşeni yüksek band değerlerine yüksek ağırlık veren bir fonksiyondur. Kırmızı eğri ile ifade edilen ikinci FKEKKR bileşeni ise orta derece dalgaboylarına az, alçak ve yüksek dalgaboylarına daha fazla ağırlık vermektedir. Bu bileşen, ikinci fonksiyonel temel bileşenin tersi olarak yorumlanabilir.

Çapraz geçerlilik kriteri ile uygulanan FTBR ve FKEKKR yöntemlerinden kestirilen parametre fonksiyonları Şekil 5.11'de gösterilmiştir.

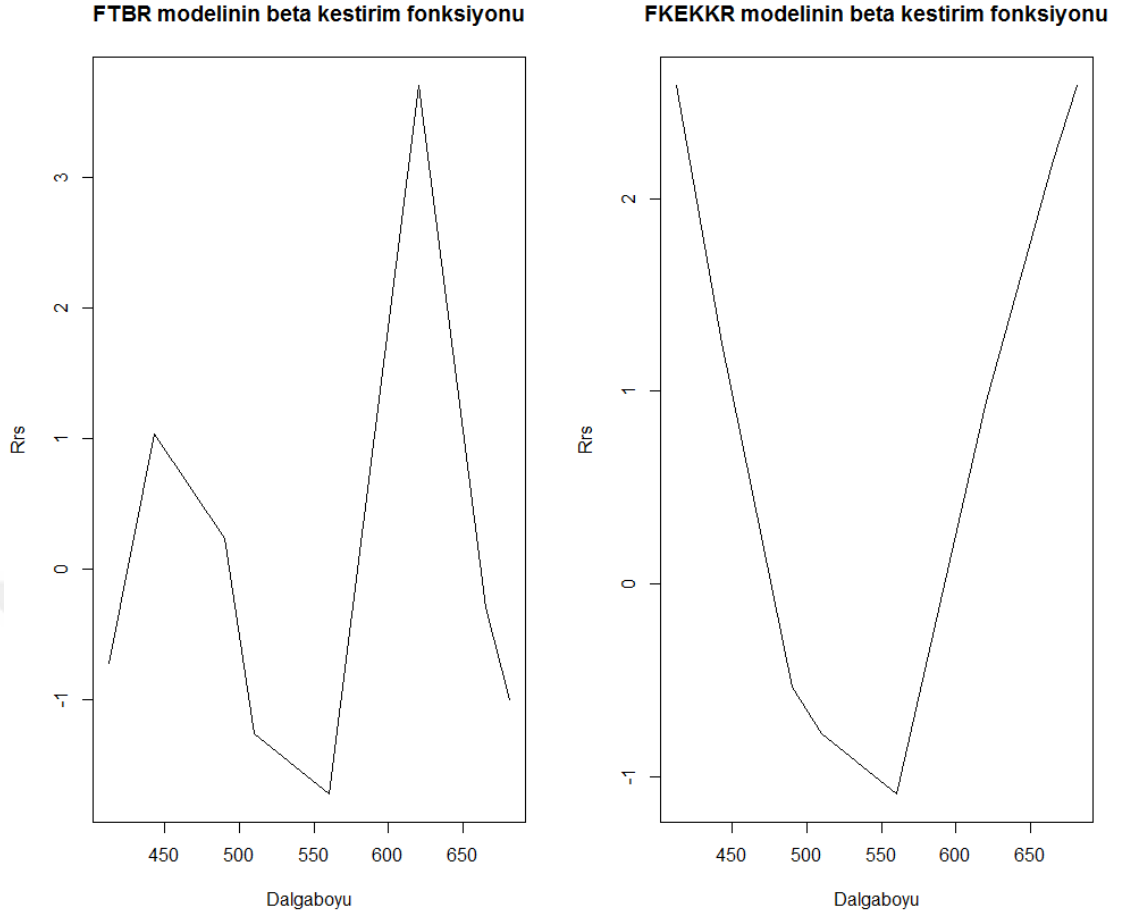
Şeklin sol panelinde görüldüğü gibi, her iki durumda da yüksek band değerleri parametre fonksiyonun kestiriminde daha etkilidir. İki parametre fonksiyonu arasındaki en temel fark düşük band değerlerinin FKEKKR'da FTBR'ye göre daha önemli olmasıdır.



Şekil 5.10: FTBR CV ve FKEKKR CV Modellerince Seçilen Fonksiyonel Bileşenler.

5.1.8 Uygulama Sonuçları

Diğer istatistiksel yöntemlere kıyasla FDRM genellikle daha düşük AMEP değerleri vermiştir. Klasik istatistiksel modeller arasında (Üstel regresyon, GAM ve LASSO), 681 nm ve 665 nm bandlarından elde edilen Rrs değerleri kullanılarak yapılan basit üstel regresyon modelleri TSS değerlerini daha iyi tahmin etmiştir. Bu modeller aynı zamanda literatürde de TSS'i tahmin etmek için sıkça kullanılan ve önerilen modellerdir. Skaler yanıt değişkeni TSS olarak alındığında, B-splayn taban yaklaşımı ile FDRM, FTBR CV ve FKEKKR CV yöntemleri daha iyidir. log TSS ve Rrs değişkenleri arasında kurulan, açıklayıcı değişkenin RSS eğrilerinin birinci türevi olarak alındığı fonksiyonel modelleri hariç, tüm fonksiyonel doğrusal regresyon modelleri yanıtı, üstel regresyon modellerinden ve diğer istatistiksel yöntemlerden daha iyi tahmin etmiştir. Tüm modeller arasında en iyi tahmini FTBR CV ve FKEKKR CV modelleri vermiştir.



Şekil 5.11: FTBR CV ve FKEKCR CV modellerinden kestirilen parametre fonksiyonları.

Sonuç olarak, TSS değişkeninin Rrs değerleri kullanılarak tahmin edilmesinde FDRM, klasik istatistiksel yöntemlere iyi bir alternatif oluşturmaktadır. Özellikle, daha fazla dalgaboyunun kullanılmasına imkan veren gelecek çalışmalarda FDRM uzaktan algılama verilerinin modellenmesinde daha büyük öneme sahip olacaktır.

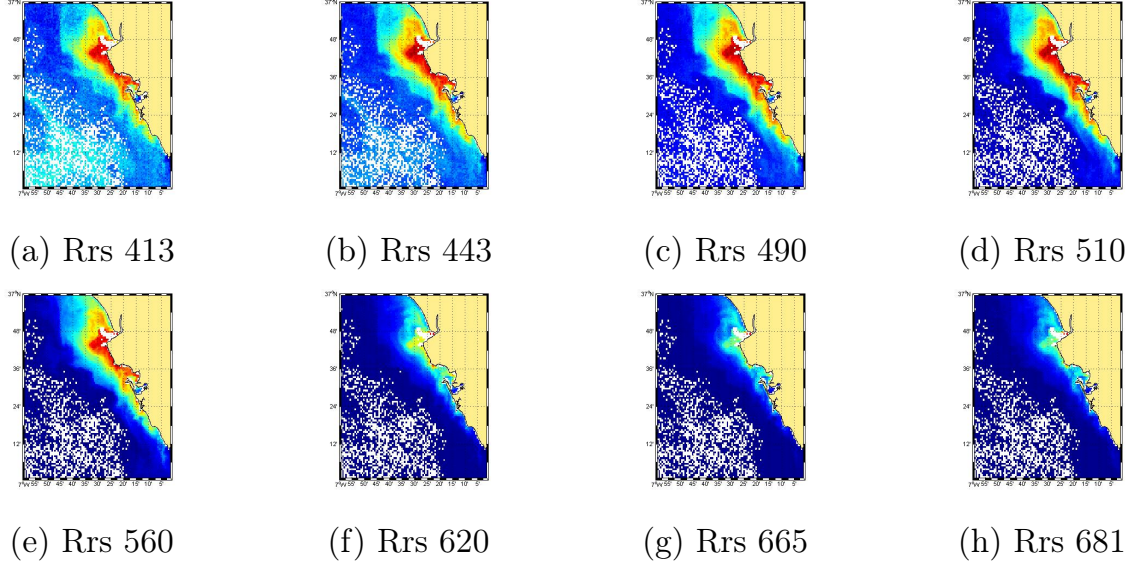
6 SİMÜLASYON ÇALIŞMASI

Uygulamada kullanılan yöntemlerin tahmin performanslarının karşılaştırılması ve sonuçların desteklenmesi amacıyla bir simülasyon çalışması tasarlanmıştır. Simülasyon çalışması üç temel adımdan oluşmaktadır. Birinci adımda, skaler yanıt vektörünün oluşturulması için gerekli olan fonksiyonel açıklayıcı değişken kaydedilen uydu görüntülerinden seçilmiştir. İkinci adımda, bir önceki adımda seçilen uydu görüntüsü kullanılarak \tilde{Y} yanıt değişkeni oluşturulmuştur. Bu aşamada, yanıt değişkenini üretmek için dört farklı model ve üç farklı hata varyansı kullanılmıştır. Üçüncü adımda ise üretilen yanıt değişkeni belirlenen pikseller için 7 farklı model üzerinden tahmin edilmiştir. Bu bağlamda, üç farklı örneklem büyüklüğü seçilerek iki katlı Çapraz Geçerlilik (CV) yöntemine bağlı MEP değerleri hesaplanarak modeller karşılaştırılmıştır. Gelecek bölümlerde bu adımlar ayrıntılı olarak açıklanmıştır.

6.1 Fonksiyonel Açıklayıcı Değişkenin Seçimi

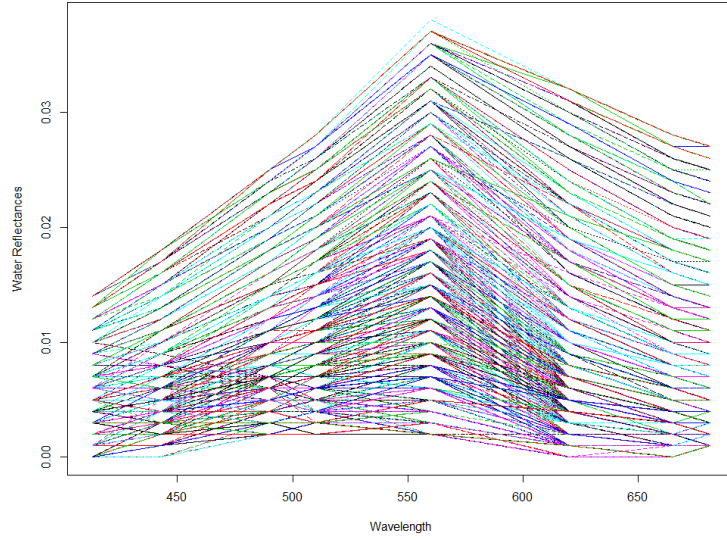
Fonksiyonel açıklayıcı değişkenin oluşturulmasında 413 nm, 443 nm, 490 nm, 510 nm, 560 nm, 620 nm, 665 nm, 681 nm dalga boylarında kaydedilmiş uydu görüntüleri kullanılmıştır. Görüntüler 36.01 – 37° N enlem, 7 – 6.01° W boylamları ile sınırlandırılan alanı gösterecek şekilde düzenlenmiştir. Bu nedenle, her görüntü toplam 136900 pikselden oluşmaktadır. Her piksel sudan yansıma değerleri ile ilgili bilgi içermektedir. Olabildiğince fazla bilgi içeren fonksiyonel değişkenin belirlenebilmesi için karaya ait pikseller çıkarılarak Rrs değeri 0'dan büyük olan pikseller ele alındı. Dalgaboyu değeri arttıkça, piksellerdeki bilgi değeri düştüğü için (bkz. Şekil 6.1) bu işlem son dalgaboyu 681 nm referans alınarak yapılmıştır.

En çok bilgi içeren, başka bir deyişle 0'dan büyük değer içeren en fazla piksele sahip olan görüntünün 9843 piksel ile 2009 yılına ait olduğu görülmüştür. Fonksiyonel açıklayıcı değişken $\tilde{\chi}(t)$ olarak kullanılmak üzere bu güne ait sekiz dalga boyundaki Rrs değerleri birleştirilerek fonksiyonel obje oluşturulmuştur. Bazı dalga boylarında



Şekil 6.1: Farklı dalga boyları için uydudan elde edilen görüntüler

bulunan eksik değerler sonucu toplam 66 gözlem analizden çıkarılmıştır. Sonuç olarak, 9777 ($N=9777$) adet fonksiyonel gözlem kalmıştır. Her bir piksele ait eğriler Şekil 6.2’te görüldüğü gibidir.



Şekil 6.2: Simülasyonda Kullanılan Fonksiyonel Gözlemler

6.2 Yanıt Değişkeninin Üretilmesi

\tilde{Y} skaler yanıt vektörü, uygulamada TSS değerinin logaritması ve Rrs değerleri arasında kurulan modellerden elde edilen parametre kestirim fonksiyonlarına dayanarak üretilmiştir. TSS değerinin logaritması üretilirken dört model ($M = 4$)

kullanılmıştır: 665 nm dalgaboyu açıklayıcı değişkeni ile basit üstel regresyon, B-splayn taban fonksiyonlarına dayalı FDRM, tüm bileşenler ile FTBR, CV yöntemi ile seçilen bileşenlere dayalı FTBR. FKEKKR, yanıt değişkeni ve açıklayıcı değişkenler arasındaki kovaryansın bilinmesini gerektirdiğinden kullanılamamıştır. Bu aşamada yanıt vektörü bilinmediğinden ve açıklayıcı değişkene bağlı olarak oluşturulmak istendiğinden kovaryans hesaplanamaz.

Başlangıç olarak, yanıt vektörü (\tilde{Y}_{exp}), (5.6) denklemi ile verilen basit klasik üstel regresyon modeli temel alınarak üretilmiştir. Açıklayıcı değişken olarak seçilen gün için 665 nm dalgaboyunda kaydedilmiş Rrs değerleri alınmıştır.

Daha sonra ise yanıt değişkeninin üretilmesinde,

$$\tilde{Y}_i = \int \tilde{\chi}_i^*(t)\beta(t)dt + \varepsilon, \quad (6.1)$$

genel formuyla ifade edilen FDRM kullanılmıştır. Burada, $\tilde{\chi}^*(t)$ seçilen güne ilişkin olarak 8 farklı dalga boyunda kaydedilmiş Rrs değerlerinin oluşturduğu fonksiyonel açıklayıcı değişkeni, $\beta(t)$ sırasıyla uygulamada kullanılan ilgili FDRM'den alınan (sırasıyla B-splayn taban fonksiyonuna dayalı FDRM, FTBR ve FTBR CV modelleri) parametre kestirim fonksiyonunu ifade etmektedir. Bu modellerden üretilen yanıt vektörleri sırasıyla \tilde{Y}_{taban} , \tilde{Y}_{FTB} ve \tilde{Y}_{FTBcv} ile gösterilecektir.

B-splayn yöntemi ile FDRM'e dayalı yanıt vektörü (\tilde{Y}_{taban}), $\mathbf{J}_{\phi\theta}$, sırasıyla $\tilde{\chi}_i^*(t) = \sum_{k=1}^8 c_{ik}^* \phi_k(t)$ ve $\beta(t) = \sum_{l=1}^5 b_l^* \theta_l(t)$ fonksiyonlarının açılımında kullanılan $\phi(t) = [\phi_1(t), \dots, \phi_8(t)]'$ ve $\theta(t) = [\theta_1(t), \dots, \theta_5(t)]'$ şeklindeki taban fonksiyonlarının iç çarpımının oluşturduğu matrisi ifade etmek üzere,

$$\tilde{Y}_{taban} = \int \tilde{\chi}^*(t)\beta(t)dt = \mathbf{c}_i^* \mathbf{J}_{\phi\theta} \mathbf{b}^*, \quad (6.2)$$

modelinden üretilmiştir. Burada, $\mathbf{c}_i^* = [c_{i1}^*, \dots, c_{i8}^*]'$ ve $\mathbf{b}^* = [b_1^*, \dots, b_5^*]'$ fonksiyonel açıklayıcı değişken ve parametre fonksiyonunun açılımında kullanılan katsayı vektörleridir. (6.2) modelindeki \mathbf{b}^* ve $\mathbf{J}_{\phi\theta}$ parameterleri ilgili FDRM'den alınırken, \mathbf{c}_i^* ilk aşamada oluşturulan fonksiyonel açıklayıcı değişken $\tilde{\chi}_i^*(t)$ 'nin uygulamadakin benzer şekilde 8 adet B-splayn taban fonksiyonu üzerinden düzleştirilmesinden elde edilmiştir.

FTBR durumunda, (6.1) denklemleri ile verilen model,

$$\tilde{Y}_{FTB} = \mu_Y + \int \tilde{\chi}^{c*}(t)\beta(t)dt = \mu_Y + \mathbf{F}_{tb}^* \mathbf{b}^*, \quad (6.3)$$

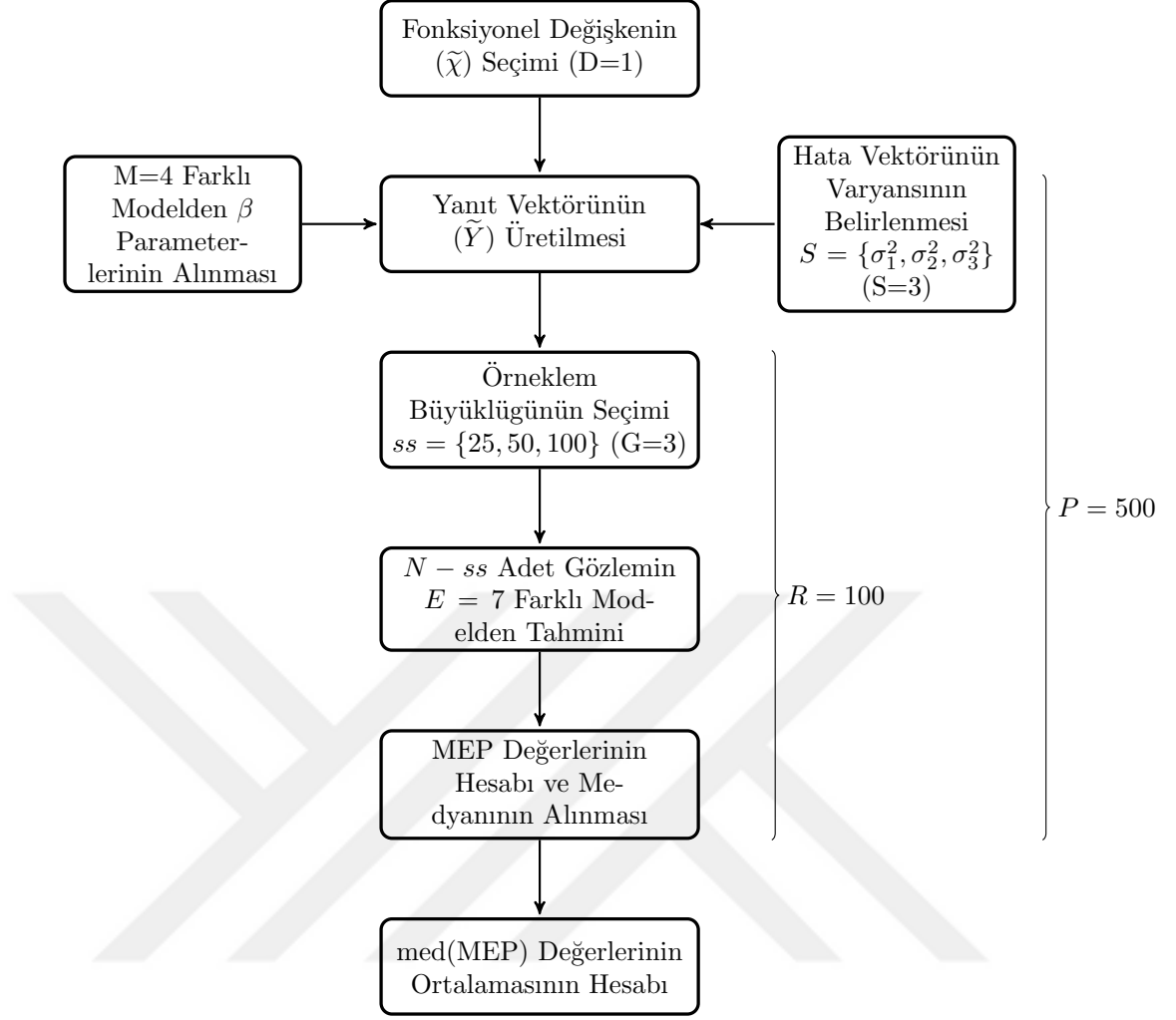
μ_Y gerçek yanıt vektörünün ortalaması, \mathbf{F}_{tb}^* , uygulamadan elde edilen $\hat{\xi}$ öz fonksiyonları aracılığıyla merkezleştirilmiş $\tilde{\chi}^{c*}$ eğrileri kullanılarak $\mathbf{F}_{tb}^* = \langle \tilde{\chi}^{c*}, \hat{\xi} \rangle$ eşitliğinden hesaplanan skor matrisi ve \mathbf{b}^* , uygulamada kullanılan $\beta_{FTB}(t) = \sum_{j=1}^J b_j^* \hat{\xi}_j$ açılımından alınan parametre vektörüdür. Burada, J , verinin düzleştirilmesinde kullanılan bileşen sayısını göstermek üzere (6.3) modeline indirgenir. FTBR modeli için bu sayı 8'e, FTBR CV modeli için ise 2'ye eşittir.

Yanıt vektörünün üretilmesinde kullanılan bir diğer parametre hata terimidir. Hata terimi ε sıfır ortalamalı ve parametre fonksiyonunun alındığı modelin hata varyansına sahip normal dağılımdan gelen bir vektördür. Gözlem sayısının az olmasından ötürü uygulamadaki modellere ait artık varyansları oldukça yüksektir. Bu nedenle, yanıt vektörünün simülasyonunda üç farklı ($S = 3$) varyans değeri göz önüne alınmıştır: uygulamadaki ilgili modelden elde edilen ham artık varyansı ($\sigma_1^2 = \hat{\sigma}^2$), model artık varyansının beşte biri ($\sigma_2^2 = \hat{\sigma}^2/5$) ve model artık varyansının onda biri ($\sigma_3^2 = \hat{\sigma}^2/10$). Her bir modelde, artık varyansının her bir değeri için sıfır ortalamalı normal dağılımdan gelen 500 adet hata vektörü üretilmiştir. Dolayısıyla, her bir modelde her bir artık varyans değeri için 500 ($P = 500$) farklı yanıt vektörü oluşturulmuştur.

6.3 Tahmin Modellerinin Karşılaştırılması

Yanıtın tahmin edilmesinde yedi farklı ($E=7$) model kullanılmıştır: 665 nm dalgaboyu ile üstel regresyon modeli, B-splayn taban açılımına dayalı FDRM, FTBR, FTBR CV, FKEKKR, FKKEKKR CV modelleri ve 2.dereceden türevin açıklayıcı değişken olarak alındığı FDRM (FDRM 2.türev). Modellerin tahmin performansı (5.1) denkleminde hesaplanan 2-katlı CV yöntemine dayalı MEP değerleri aracılığıyla karşılaştırılmıştır.

Bu bağlamda öncelikle veri üç farklı ($G = 3$) örneklem büyüklüğü $ss = \{25, 50, 100\}$ seçilerek ikiye bölünmüştür. Gerçek hayat verilerinde tahminler genellikle sınırlı sayıda gözlem üzerinden yapıldığından en büyük örneklem hacmi 100 olarak belirlenmiştir. Görüntüyü oluşturan 9777 pikselden ss kadarı rassal olarak seçilmiş



Şekil 6.3: Simülasyon Tasarımı

ve geriye kalan $(N - ss)$ pikselleri için yanıt değişkeni seçilen ss gözlem üzerinden tahmin edilmiştir.

Bu süreç, $R = 100$ defa tekrar edilmiş ve her bir yanıt türü ve tahmin modeli için MEP değerleri hesaplanmıştır. Daha sonra, hesaplanan MEP değerleri çarpık bir dağılıma sahip olduğundan MEP değerlerinin 100 tekrar üzerinden medyanı (medMEP) alınmıştır.

Son olarak ise, $P = 500$ tekrar üzerinden medMEP değerlerinin ortalaması alınmış ve modeller ortalama medMEP değerleri kullanılarak karşılaştırılmıştır.

Simülasyon çalışması Şekil 6.3'deki gibi özetlenebilir. Çizelgede görüldüğü gibi tasarlanan simülasyon çalışmasında $M \times S \times G$, yani 36, olası senaryo vardır. Her olası durum için $E = 7$ farklı modelden elde edilen sonuçlar karşılaştırılmıştır.

6.4 Simülasyon Sonuçları

Örneklem büyüklüğüne göre simülasyon sonuçları sırasıyla Çizelge 6.1, Çizelge 6.2 ve Çizelge 6.3'de verilmiştir.

Çizelge 6.1 incelenecek olursa örneklem büyüklüğünün 25 olması durumunda, hata varyansı azaldıkça \tilde{Y}_{taban} ve \tilde{Y}_{FTB} yanıtları için üstel regresyon modeli ve FDRM 2.Türev modelleri hariç tüm modellerin ortalama med(MEP) değerleri de azalmaktadır. Hata varyansının yüksek olması durumunda \tilde{Y}_{exp} ve \tilde{Y}_{taban} yanıtları için sırasıyla FKEKKR CV and FDRM Türev 2 modelleri, \tilde{Y}_{FTBcv} yanıtı için ise FKEKKR CV modelleri iyi tahminler vermişlerdir. Hata varyansı azaldıkça yanıt tipine göre en iyi tahminler genelde yanıtın geldiği modelden elde edilirken \tilde{Y}_{FTBcv} yanıtı için FKEKKR CV yöntemi FTBR CV yöntemine göre daha iyi sonuç vermiştir. Tüm bileşenleri içeren FTBR ve FKEKKR modelleri her zaman eşit sonuçlar vermektedir.

Örneklem genişliği 50 olarak alındığında Çizelge 6.2'de de görüldüğü gibi, yüksek hata varyansı (σ_1^2) durumunda \tilde{Y}_{exp} yanıtı için FKEKKR CV yöntemi, \tilde{Y}_{FTBcv} yanıtı için ise FDRM Taban ve FKEKKR CV yöntemi en iyi tahminleri vermiştir. Bunlar dışında en iyi tahminler genelde yanıtın üretildiği modelden elde edilmiştir. Örneklem büyüklüğü arttırıldığında tüm modellerden hesaplanan medMEP değerleri ortalamasının hata varyansı azaldıkça düştüğü görülmektedir. Ancak bu düşüş Rrs eğrilerinin açıklayıcı değişken olarak alındığı fonksiyonel modellerde diğer modellere kıyasla çok daha yüksektir. Çizelge 6.3'den de görüldüğü gibi örneklem genişliği 100'e çıkarıldığında, düşük hata varyansı durumunda \tilde{Y}_{exp} yanıtını FDRM 2.türev modeli hariç tüm modeller iyi tahmin etmiştir. Fonksiyonel modellerden türetilen yanıtlar içinde \tilde{Y}_{taban} yanıtı için yüksek varyans durumunda en iyi tahmini FDRM Taban yöntemi verirken varyans düştükçe fonksiyonel modellerin tahmin performansı yaklaşmıştır. Boyut indirgemeye dayalı fonksiyonel regresyon modellerinden türetilen \tilde{Y}_{FTB} ve \tilde{Y}_{FTBcv} yanıtlarını ise en iyi türetildikleri modeller tahmin etmiştir.

Genel olarak yorum yapılacak olursa, örneklem büyüklüğü arttıkça ve hata varyansı azaldıkça tüm yanıt türleri için modellerin medMEP değerleri düşmektedir. Özellikle de fonksiyonel regresyon modellerinin medMEP ortalamaları yaklaşmakta ve aralarındaki fark azalmaktadır. Ancak fonksiyonel modeller arasında FDRM 2.Türev modeli diğer modellere kıyasla daha yüksek MEP değerleri ile fark yaratmaktadır.

Yanıt üstel regresyon modelinden geldiğinde, özellikle de örneklem genişliği büyük ise, üstel regresyon modelinden elde edilen MEP değerleri ile diğer FDRM'den elde edilen MEP değerleri arasında fazla fark yoktur. Ancak yanıt değişkeni FDRM kullanılarak türetildiğinde, üstel regresyon modeli ve diğer FDRM'den elde edilen MEP değerleri arasındaki fark artmaktadır.

Tüm yanıt türleri için, FDRM'nin genelde literatürde sıklıkla kullanılan üstel regresyon modeline göre yanıtı daha iyi tahmin ettiği görülmektedir. Özellikle de, boyut indirgeme yöntemine dayalı FDRM (FTBR, FTBR CV, FKEKKR, FKEKKR CV) ile Taban yaklaşımı ile FDRM düşük MEP değerleri ile en iyi tahminleri yapmaktadır.

Çizelge 6.1: Örneklem Büyüklüğü 25 için Karşılaştırma Tablosu (medMEP değerlerinin ortalaması)

Örneklem Büyüklüğü ss=25								
Yanıt Türü	Hata Varyansı	Üstel Reg.	FDRM Taban	FTBR	FTBR CV	FKEKKR	FKEKKR CV	FDRM 2.Türev
\tilde{Y}_{exp}	σ_1^2	0.99	1.26	1.57	1.25	1.57	0.98	1.04
	σ_2^2	0.14	0.25	0.36	0.29	0.36	0.31	0.67
	σ_3^2	0.04	0.07	0.09	0.09	0.09	0.09	0.65
\tilde{Y}_{taban}	σ_1^2	1.10	1.11	1.32	1.15	1.32	1.09	1.10
	σ_2^2	1.19	0.13	0.20	0.17	0.20	0.16	1.03
	σ_3^2	1.20	0.03	0.05	0.05	0.05	0.06	1.02
\tilde{Y}_{FTB}	σ_1^2	1.06	1.22	1.27	1.16	1.27	1.13	1.01
	σ_2^2	1.08	0.76	0.17	0.21	0.17	0.29	0.85
	σ_3^2	1.07	0.75	0.04	0.05	0.04	0.05	0.83
\tilde{Y}_{FTBcv}	σ_1^2	0.88	0.78	0.91	0.82	0.91	0.74	0.78
	σ_2^2	0.52	0.08	0.09	0.08	0.09	0.07	0.15
	σ_3^2	0.49	0.03	0.02	0.02	0.02	0.02	0.11

Sonuç olarak, tasarlanan simülasyon çalışmasından elde edilen sonuçlar uygulamada elde edilen sonuçları desteklemektedir. Böylelikle, sınırlı sayıda dalga boyunda bilgi kaydedilmiş olsa bile uzaktan algılama verilerinin analizinde FDRM'nin bilinen klasik istatistiksel yöntemlere iyi bir alternatif oluşturduğu söylenebilir.

Çizelge 6.2: Örneklem Büyüklüğü 50 için Karşılaştırma Tablosu (medMEP değerlerinin ortalaması)

Örneklem Büyüklüğü ss=50								
Yanıt Türü	Hata Varyansı	Üstel Reg.	FDRM Taban	FTBR	FTBR CV	FKEKKR	FKEKKR CV	FDRM 2.Türev
\tilde{Y}_{exp}	σ_1^2	0.90	1.00	1.08	0.98	1.08	0.89	0.97
	σ_2^2	0.11	0.14	0.16	0.15	0.16	0.20	0.60
	σ_3^2	0.03	0.04	0.04	0.04	0.04	0.04	0.57
\tilde{Y}_{taban}	σ_1^2	1.02	0.84	0.91	0.89	0.91	0.86	1.04
	σ_2^2	0.94	0.08	0.09	0.09	0.09	0.10	0.95
	σ_3^2	0.94	0.02	0.02	0.02	0.02	0.03	0.94
\tilde{Y}_{FTB}	σ_1^2	0.97	0.90	0.85	0.88	0.85	0.88	0.95
	σ_2^2	0.83	0.40	0.08	0.08	0.08	0.09	0.76
	σ_3^2	0.82	0.36	0.02	0.02	0.02	0.02	0.75
\tilde{Y}_{FTBcv}	σ_1^2	0.89	0.71	0.75	0.73	0.75	0.71	0.73
	σ_2^2	0.54	0.07	0.07	0.07	0.07	0.07	0.14
	σ_3^2	0.52	0.02	0.02	0.02	0.02	0.02	0.09

Çizelge 6.3: Örneklem Büyüklüğü 100 için Karşılaştırma Tablosu (medMEP değerlerinin ortalaması)

Örneklem Büyüklüğü ss=100								
Yanıt Türü	Hata Varyansı	Üstel Reg.	FDRM Taban	FTBR	FTBR CV	FKEKKR	FKEKKR CV	FDRM 2.Türev
\tilde{Y}_{exp}	σ_1^2	0.84	0.88	0.91	0.88	0.91	0.86	0.95
	σ_2^2	0.10	0.11	0.12	0.12	0.12	0.12	0.56
	σ_3^2	0.03	0.03	0.03	0.03	0.03	0.03	0.53
\tilde{Y}_{taban}	σ_1^2	0.96	0.72	0.75	0.75	0.75	0.74	1.01
	σ_2^2	0.77	0.07	0.07	0.07	0.07	0.07	0.91
	σ_3^2	0.76	0.02	0.02	0.02	0.02	0.02	0.91
\tilde{Y}_{FTB}	σ_1^2	0.91	0.75	0.69	0.72	0.69	0.72	0.92
	σ_2^2	0.69	0.27	0.06	0.06	0.06	0.06	0.72
	σ_3^2	0.67	0.24	0.01	0.01	0.01	0.01	0.71
\tilde{Y}_{FTBcv}	σ_1^2	0.90	0.69	0.70	0.69	0.70	0.69	0.71
	σ_2^2	0.59	0.07	0.06	0.06	0.06	0.06	0.13
	σ_3^2	0.57	0.02	0.02	0.02	0.02	0.02	0.09

7 SONUÇ VE ÖNERİLER

Bu çalışmada, denizbilim alanında uzaktan algılama verileri üzerinden tahminler yapılmasında bilinen klasik istatistiksel modellere alternatif olarak FVA yaklaşımı önerilmiş ve Cadiz bölgesindeki Guadalquivir nehir ağzındaki TSS miktarının uydu verileri üzerinden tahmin edilmesinde çeşitli istatistiksel modeller uygulanmıştır. Denizbilimde uydu verilerinden TSS değerlerinin tahmin edilmesinde genelde klasik üstel doğrusal regresyon modelleri kullanılırken bu çalışma ile pek çok alternatif yöntem denenmiş ve sonuçları karşılaştırılmıştır. FDRM için farklı yanıtlar ve açıklayıcı değişkenler alınmıştır. Bu modeller arasında, TSS değerlerinin logaritması ve Rrs değerleri arasındaki boyut indirgeme yöntemlerine dayalı FDRM, TSS miktarını GTM, LASSO ve üstel regresyon modellerine kıyasla daha iyi tahmin etmiştir.

Bu çalışmanın önemi, deniz bilim alanında uydu verilerinden TSS parametre tahmini yapılmasında ilk kez fonksiyonel doğrusal regresyon modellerinin uygulanmasıdır. MERIS uydusundan elde edilen uzaktan algılama verilerinin kullanıldığı daha önceki çalışmalarda TSS değerleri tek bantlı üstel regresyon modelleri kullanılarak tahmin edilmiştir. Binding ve diğ. (2003) mineral oranının 665 nm bandındaki yansıma değerleri üzerinden kuadratik regresyon modeli ile tahmin etmiştir. Nechad ve diğ. (2010)'ın çalışmasında TSS değerlerinin modellenmesinde 681 nm bandını kullanılırken Caballero ve diğ. (2016)'ın çalışmasında en yüksek açıklama oranını 665 nm bandında bulunmuştur. MERIS uydusu dışında MODIS, DEIMOS gibi farklı uydulardan elde edilen uzaktan algılama verilerinin kullanıldığı çalışmalar da vardır (Caballero ve diğ., 2014a,b). Ancak MERIS daha yüksek çözünürlüğe sahip olduğundan bu uydulara tercih edilmektedir. Nechad ve diğ. (2010) özellikle düşük yoğunluk değerleri için tek bant yerine çok bantlı modellerin daha uygun olabileceğini belirtmiştir. Bu çalışmada FDRM'nin yanısıra daha önce literatürde kullanılmamış tek

bantlı ve çok bantlı pek çok farklı model uygulanarak karşılaştırılmıştır. En uygun modeller FDRM olarak bulunmuştur.

Bu çalışmanın kısıtları gözlem sayısının ve dalga boyu sayısının azlığıdır. Tasarlanan simülasyon çalışması ile gözlem sayısı artırılmıştır. Ancak ne yazık ki, o dönemde kullanılan sensörlerden dolayı dalgaboyu sayısını arttırmak mümkün değildir. Son zamanlarda, daha fazla dalga boyunda veri kaydeden sensörler kullanımdadır. Ancak, bu dönem ile eşleşen yeterli sayıda toplanmış örneklem bulunmamaktadır. Bu alanda Fransa'da Ferraty ve diğ. (2016) tarafından yürütülen yeni bir projede hiperspektral uzaktan algılama verileri üzerinden nonparametrik FDRM kullanılarak doğa ile ilişkili parametre tahminleri yapılmaktadır. Dalga boyu sayısının artması verinin büyüklüğünü de arttıracaktır.

Uygulama kısmında elde edilen sonuçlar bir simülasyon çalışması ile desteklenmiştir. Simülasyon sonuçlarına göre, örneklem büyüklüğü artıp hata varyansı azaldıkça FDRM'nin tahmin performansının üstel regresyon modellerinden daha iyi olduğu görülmektedir. Uygulamadakine benzer şekilde boyut indirgeme yöntemlerine dayalı FDRM diğer fonksiyonel modellerden genelde daha iyi tahminler vermiştir. Çalışmadaki zaman kısıtı nedeniyle simülasyon belli kısıtlarda yapılmıştır. Ancak ilerleyen çalışmalarda bazı parametreleri değiştirerek simülasyon çalışmasının geliştirilmesi düşünülmektedir. Örneğin, tahmin için kullanılan örneklem büyüklüğünün 100 yerine 200 olması, açıklayıcı değişkeni oluşturan piksellerin rasgele seçilmesi yerine belli rotalarda toplanan örneklem koordinatları gözönünde bulundurularak seçilmesi, açıklayıcı değişken belirlenirken farklı günlerde kaydedilen görüntülerin göz önüne alınıp daha fazla sayıda görüntü kullanılması düşünülen değişikliklerden bazılarıdır. Bunun dışında, uygulamada veri eşleştirmesi sürecinde gözlemin tam olarak kaydedildiği piksel yerine gözlemin 2×2 kutu alanda en yakın olduğu nokta kullanılmıştı. Bu sayede daha fazla gözlem elde edildi. Ancak ileriki çalışmalarda verinin tam koordinatı dikkate alınarak yapılan eşleştirilmenin sonuçları nasıl değiştireceği incelenebilir. Bizim çalışmamızda indirilen 3.5 TB boyutundaki uydu verileri göz önüne alındığında çalışmanın ayrıca bir büyük veri problemi içerdiği görülmektedir. Gelecekte büyük veri setlerinin artmasıyla birlikte sadece bu alanda değil pek çok alanda FDRM'nin diğer istatistiksel yöntemlere göre daha kolaylık sağlayacağı ve kullanımının daha önem kazanacağı öngörülmektedir.

Bu çalışma ülkemizde bu denli ayrıntı içeren ilk detaylı çalışma olduđu için ileride bu alanda araştırma yapacaklara yardımcı olacağı düşünölmektedir.



KAYNAKLAR

- Aguilera, A. M., Escabias, M., Preda, C., ve Saporta, G. (2010). Using basis expansions for estimating functional pls regression: Applications with chemometric data. *Chemometrics and Intelligent Laboratory Systems*, 104(2):289–305.
- Aguilera, A. M., Escabias, M., Valderrama, M. J., ve Aguilera-Morillo, M. C. (2013). Functional analysis of chemometric data. *Open Journal of Statistics*, 3(5):333–343.
- Aguilera, A. M. ve Escabias, M. v. V. M. J. (2008). Discussion of different logistic models with functional data. application to systemic lupus erythematosus. *Computational Statistics & Data Analysis*, 53(1):151–163.
- Aguilera Morillo, M. d. C. (2013). *Penalized Estimation Methods in Functional Data Analysis*. PhD thesis, University of Granada.
- Almansa, J. ve Delicado, P. (2009). Analysing musical performance through functional data analysis: rhythmic structure in schumann’s träumerei. *Connection Science*, 21(2-3):207–225.
- Besse, P. ve Ramsay, J. O. (1986). Principal components analysis of sampled functions. *Psychometrika*, 51(2):285–311.
- Besse, P. C., Cardot, H., Faivre, R., ve Goulard, M. (2005). Statistical modelling of functional data. *Applied Stochastic Models in Business and Industry*, 21(2):165–173.
- Binding, C., Bowers, D., ve Mitchelson-Jacob, E. (2003). An algorithm for the retrieval of suspended sediment concentrations in the irish sea from seawifs ocean colour satellite imagery. *International Journal of Remote Sensing*, 24(19):3791–3806.
- Binding, C., Bowers, D., ve Mitchelson-Jacob, E. (2005). Estimating suspended sediment concentrations from ocean colour measurements in moderately turbid

- waters; the impact of variable particle scattering properties. *Remote sensing of Environment*, 94(3):373–383.
- Caballero, I., Morris, E. P., Ruiz, J., ve Navarro, G. (2014a). Assessment of suspended solids in the Guadalquivir estuary using new Deimos-1 medium spatial resolution imagery. *Remote Sensing of Environment*, 146:148–158.
- Caballero, I., Morris, E. P., Ruiz, J., ve Navarro, G. (2014b). The influence of the Guadalquivir river on the spatio-temporal variability of suspended solids and chlorophyll in the eastern Gulf of Cadiz. *Mediterranean Marine Science*, 15(4):721–738.
- Caballero, I., Ruiz, J., ve Navarro, G. (2016). Dynamics of the turbidity plume in the Guadalquivir estuary (SW Spain): A remote sensing approach. ESA Living Planet Symposium, Prague, Czech Republic.
- Cardot, H., Faivre, R., ve Goulard, M. (2003). Functional approaches for predicting land use with the temporal evolution of coarse resolution remote sensing data. *Journal of Applied Statistics*, 30(10):1185–1199.
- Cardot, H., Ferraty, F., ve Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters*, 45(1):11–22.
- Clarke, E., Speirs, D., Heath, M., Wood, S., Gurney, W., ve Holmes, S. (2006). Calibrating remotely sensed chlorophyll-a data by using penalized regression splines. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55(3):331–353.
- Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference*, 147:1–23.
- Cuevas, A., Febrero, M., ve Fraiman, R. (2004). An ANOVA test for functional data. *Computational Statistics & Data Analysis*, 47(1):111–122.
- Dauxois, J., Pousse, A., ve Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *Journal of Multivariate Analysis*, 12(1):136–154.

- Davidian, M., Lin, X., ve Wang, J. (2004). Introduction: Emerging issues in longitudinal and functional data analysis. *Statistica Sinica*, 14(3):613–614.
- Delaigle, A., Hall, P., ve diğ. (2012). Methodology and theory for partial least squares applied to functional data. *The Annals of Statistics*, 40(1):322–352.
- Delicado, P. (2015). Advanced statistical modelling. Lecture Notes.
- Dieudonné, J. (1960). *Foundations of modern analysis*, volume 286. Academic press New York.
- Escabias, M., Aguilera, A. M., ve Valderrama, M. J. (2004). Principal component estimation of functional logistic regression: discussion of two different approaches. *Journal of Nonparametric Statistics*, 16(3-4):365–384.
- Escabias, M., Aguilera, A. M., ve Valderrama, M. J. (2005). Modeling environmental data by functional principal component logistic regression. *Environmetrics*, 16(1):95–107.
- Escabias, M., Valderrama, M. J., ve Aguilera-Morillo, M. C. (2012). Functional data analysis in biometrics and biostatistics. *J Biom Biostat*, 3(8):e120.
- Faivre, R. ve Fischer, A. (1997). Predicting crop reflectances using satellite data observing mixed pixels. *Journal of Agricultural, Biological, and Environmental Statistics*, 2(1):87–107.
- Febrero, M., Galeano, P., ve Manteiga-Gonzales, W. (2007). A functional analysis of nox levels: location and scale estimation and outlier detection. *Computational Statistics*, 22:411–427.
- Febrero, M., Galeano, P., ve Manteiga-Gonzales, W. (2008). Outlier detection in functional data by depth measures, with application to identify abnormal nox levels. *Environmetrics*, 19:331–345.
- Febrero-Bande, M., Galeano, P., ve González-Manteiga, W. (2015). Functional principal component regression and functional partial least-squares regression: An overview and a comparative study. *International Statistical Review*, 0(0):1–23.

- Febrero-Bande, M. ve Oviedo de la Fuente, M. (2012). Statistical computing in functional data analysis: the r package fda.usc. *Journal of Statistical Software*, 51(4):1–28.
- Ferraty, F. (2010). High-dimensional data: a fascinating statistical challenge. *Journal of Multivariate Analysis*, 101(2):305–306.
- Ferraty, F., Fauvel, M., ve Zullo, A. (2016). Nonparametric prediction based on contaminated functional predictor and application to hyperspectral data. 2016 CRoNoS Summer Course on Functional Data Analysis, Oviedo, Spain.
- Ferraty, F. ve Vieu, P. (2003). Curves discrimination: a nonparametric functional approach. *Computational Statistics & Data Analysis*, 44(1):161–173.
- Ferraty, F. ve Vieu, P. (2006). *Nonparametric functional data analysis: theory and practice*. Springer Science & Business Media.
- Fraiman, R. ve Muniz, G. (2001). Trimmed means for functional data. *Test*, 10(2):419–440.
- Friedman, J., Hastie, T., ve Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin.
- Gündüz, M. (2012). *The Survey of The Relationship Between Closing Price of the Companies Trading on Istanbul Stock Exchange and Their Trading Volumes Through Functional Canonical Correlation Analysis*. PhD thesis, Atatürk Üniversitesi.
- Gong, M., Miller, C., ve Scott, E. (2015). Functional pca for remotely sensed lake surface water temperature data. *Procedia Environmental Sciences*, (26):127–130.
- Hall, P., Müller, H.-G., ve Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The annals of statistics*, 34(3):1493–1517.
- Hastie, T. ve Tibshirani, R. (1986). Generalized additive models. *Statistical science*, 1(3):297–310.

- Hastie, T. J., Tibshirani, R. J., ve Friedman, J. H. (2011). *The elements of statistical learning: data mining, inference, and prediction*. Springer.
- Hooker, G. (2010). Functional data analysis: A short course. International Workshop on Statistical Modeling.
- Horváth, L. ve Kokoszka, P. (2012). *Inference for functional data with applications*, volume 200. Springer Science & Business Media.
- Hothorn, T. ve Everitt, B. S. (2014). *A handbook of statistical analyses using R*. CRC press.
- James, G. M. (2002). Generalized linear models with functional predictors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):411–432.
- Keser, I. K. (2007). *A study on regularized functional principal component analysis as a statistical dimension reduction technique*. PhD thesis, Dokuz Eylül Üniversitesi.
- Keser, I. K. (2010). Ege bölgesi yağış verilerinin fonksiyonel veri analizi İle İncelenmesi. *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 25(1):41–67.
- Kreyszig, E. (1978). *Introductory Functional Analysis with Applications*. John Wiley & Sons Inc., New York.
- Levitin, D., Nuzzo, R., Vines, B., ve Ramsay, J. (2007). Introduction to functional data analysis. *Canadian Psychology*, 48:135–155.
- Liu, C., Ray, S., Hooker, G., ve Friedl, M. (2012). Functional factor analysis for periodic remote sensing data. *The Annals of Applied Statistics*, 6(2):601–624.
- López-Pintado, S. ve Romo, J. (2009). On the concept of depth for functional data. *Journal of the American Statistical Association*, 104(486):718–734.
- Manteiga, W. G. ve Vieu, P. (2007). Statistics for functional data. *Computational Statistics & Data Analysis*, 51(10):4788–4792.
- Morris, J. S. (2014). Functional regression. <https://arxiv.org/abs/1406.4068>.

- Müller, H.-G., Chiou, J.-M., ve Leng, X. (2008). Inferring gene expression dynamics via functional regression analysis. *BMC bioinformatics*, 9(1):60.
- Müller, H.-G. ve Stadtmüller, U. (2005). Generalized functional linear models. *Annals of Statistics*, 33(2):774–805.
- Nechad, B., Ruddick, K., ve Park, Y. (2010). Calibration and validation of a generic multisensor algorithm for mapping of total suspended matter in turbid waters. *Remote Sensing of Environment*, 114(4):854–866.
- Nezlin, N. P. ve DiGiacomo, P. M. (2005). Satellite ocean color observations of stormwater runoff plumes along the san pedro shelf (southern california) during 1997–2003. *Continental Shelf Research*, 25(14):1692–1711.
- Nicol, F. (2013). Functional principal component analysis of aircraft trajectories. In *ISIATM 2013, 2nd International Conference on Interdisciplinary Science for Innovative Air Traffic Management*. Toulouse, France.
- Ocaña, F. A., Aguilera, A. M., ve Escabias, M. (2007). Computational considerations in functional principal component analysis. *Computational Statistics*, 22(3):449–465.
- Oviedo de la Fuente, M. (2011). Utilities for statistical computing in functional data analysis: The r package 'fda.usc'. Master's thesis, Universidad de Santiago de Compostela.
- Pidwirny, M. (2006). Introduction to geographic information systems.
- Preda, C. ve Saporta, G. (2005). Pls regression on a stochastic process. *Computational Statistics & Data Analysis*, 48(1):149–158.
- Preda, C. ve Schiltz, J. (2011). Functional pls regression with functional response: the basis expansion approach. In *14th Applied Stochastic Models and Data Analysis Conference*, pages 1126–1133. Università di Roma La Sapienza.
- Ramsay, J. (1982). When the data are functions. *Psychometrika*, 47(4):379–396.
- Ramsay, J. (2000). Functional components of variation in handwriting. *Journal of the American Statistical Association*, 95(449):9–15.

- Ramsay, J. ve Dalzell, C. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(3):539–572.
- Ramsay, J., Munhall, K. G., Gracco, V. L., ve Ostry, D. J. (1996). Functional data analyses of lip motion. *The Journal of the Acoustical Society of America*, 99(6):3718–3727.
- Ramsay, J. ve Silverman, B. (2002). *Applied Functional Data Analysis*. Springer, USA.
- Ramsay, J. ve Silverman, B. (2005). *Functional Data Analysis*. Springer, USA.
- Ramsay, J., Wang, X., ve Flanagan, R. (1995). A functional data analysis of the pinch force of human fingers. *Applied Statistics*, 44(1):17–30.
- Ramsay, J. O., Hooker, G., ve Graves, S. (2009). *Functional Data Analysis with R and MATLAB*. Springer.
- Ratcliffe, S. J., Heller, G. Z., ve Leader, L. R. (2002). Functional data analysis with application to periodically stimulated foetal heart rate data. ii: Functional logistic regression. *Statistics in medicine*, 21(8):1115–1127.
- Reiss, P. T., Goldsmith, J., Shang, H. L., ve Ogden, R. T. (2016). Methods for scalar-on-function regression. *International Statistical Review*, 0(0):1–22.
- Reiss, P. T. ve Ogden, R. T. (2007). Functional principal component regression and functional partial least squares. *Journal of the American Statistical Association*, 102(479):984–996.
- Rice, J. ve Silverman, B. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J.R. Statist. Soc. B*, 53(1):233–243.
- Rice, J. A. (2004). Functional and longitudinal data analysis: perspectives on smoothing. *Statistica Sinica*, 14(3):631–648.
- Saeyns, W., De Ketelaere, B., ve Darius, P. (2008). Potential applications of functional data analysis in chemometrics. *Journal of chemometrics*, 22(5):335–344.
- Sawant, P., Billor, N., ve Shin, H. (2012). Functional outlier detection with robust functional principal component analysis. *Computational Statistics*, 27(1):83–102.

- Schwarz, J., Pinkerton, M., Wood, S., ve Zeldis, J. (2008). Remote sensing of river plumes in the canterbury bight: Stage i report. *NIWA Client Report CHC2008-169*, 68.
- Shang, H. L. (2014). A survey of functional principal component analysis. *AStA Advances in Statistical Analysis*, 98(2):121–142.
- Silverman, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *The Annals of Statistics*, 24(1):1–24.
- Sözen, C. (2014). *Ayrık Noktalarda Gözlenen Verilerin Analizi ve Bir Uygulama*. PhD thesis, On Dokuz Mayıs Üniversitesi.
- Ullah, S. ve Finch, C. F. (2013). Applications of functional data analysis: A systematic review. *BMC medical research methodology*, 13(1):43.
- Valderrama, M. J. (2007). An overview to modelling functional data. *Computational Statistics*, 22(3):331–334.
- Viviani, R., Grön, G., ve Spitzer, M. (2005). Functional principal component analysis of fmri data. *Human brain mapping*, 24(2):109–129.
- Wang, J.-L., Chiou, J.-M., ve Mueller, H.-G. (2016). Functional data analysis. *The Annual Review of Statistics and Its Application*, 3(2):257–295.
- Wang, S., Wang, J., Wang, H., ve Saporta, G. (2009). An Approach for PLS Regression Modeling of Functional Data. In *PLS'09, 6th Int. Conf. on Partial Least Squares and Related Methods, Pekin*, pages 28–33.
- Wasserman, L. (2006). *All of nonparametric statistics*. Springer Science & Business Media.
- Wood, S. (2006). *Generalized additive models: an introduction with R*. CRC press.
- Wu, P.-S. ve Müller, H.-G. (2010). Functional embedding for the classification of gene expression profiles. *Bioinformatics*, 26(4):509–517.
- Özçomak, M. S. ve Gündüz, M. (2014). Borsa İstanbul'da İşlem gören Şirketlerin kapanış fiyatları ile İşlem miktarları arasındaki ilişkinin fonksiyonel kanonik

korelasyon ile analizi. *International Journal of Economic and Administrative Studies*, 7(13):233–253.



EKLER

EK A : Fonksiyonlar Uzaylar Üzerinde Tanımlar

EK B : Operatörler ve Fonksiyonel Veri için Bazı Yararlı Teoremler

EK C : R Programında Yazılan Simülasyon Fonksiyonu





A Fonksiyonel Uzaylar Üzerinde Tanımlar

Fonksiyonel veri analizinde, sonlu boyutlu vektör uzayları yerine fonksiyonel uzaylar üzerinde çalışılır. Fonksiyonel uzaylar genel olarak vektör uzayları ile aynı özelliklere sahiptir. Sadece vektörlerin yerini fonksiyonlar, matrislerin yerini ise doğrusal dönüşümler alır. Fonksiyonel bir veri genelde metrik ya da yarı-metrik bir uzay üzerinde tanımlanır. Fonksiyonel veri analizinin ilk adımı fonksiyonel veri üzerinde çalışılacak uzaya karar vermektir.

Bu bölümde fonksiyonel veri ile çalışılırken bilinmesinde yarar olan bazı tanım ve teoremler verilmiştir. Kullanılan bazı referanslar şunlardır: Kreyszig (1978); Ramsay ve Silverman (2005); Horváth ve Kokoszka (2012) ve Cuevas (2014).

1.1 Metrik Uzaylar ve Özellikleri

X bir küme ve $d : X \times X \rightarrow \mathbb{R}$, X üzerinde tanımlı bir uzaklık fonksiyonu (metrik) olmak üzere her $x, y, z \in X$ için,

1. $d(x, y)$ negatif-olmayan sonlu bir gerçel değer (Pozitif tanımlı),
2. $x = y$ ancak ve ancak $d(x, y) = 0$ için (Birim),
3. $d(x, y) = d(y, x)$ (Simetri),
4. $d(x, y) \leq d(x, z) + d(z, y)$ (Üçgen Eşitsizliği),

özellikleri sağlanıyorsa (X, d) çiftine bir *metrik uzay* denir.

$d(x, y) = |x - y|$ metriği ile gerçel doğru \mathbb{R} ve $d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$ Öklid metriği ile \mathbb{R}^2 düzlemi metrik uzayların örneklerindedir.

Yarı-metrik uzay yukarıda belirtilen ikinci özellik dışında metrik uzaylarla aynı özelliklere sahiptir. İkinci özelliğin sağlanmaması, $x, y \in X$ gibi iki nokta arasındaki $d(x, y)$ uzaklığının yarı-metrik uzaylarda x ve y ayrık noktalar olduğunda da sifıra eşit olabileceği anlamına gelir.

Yarı-metrik uzayın bir örneği,

$$d(\chi, \gamma) = \int_0^1 |\chi(t) - \gamma(t)| dt$$

metriği ile verilen $[0, 1]$ aralığı üzerinde mutlak integrallenebilir fonksiyonların kümesidir.

1.2 İç Çarpım Uzayları ve Özellikleri

İç çarpım pozitiflik, simetri ve doğrusallık özelliklerine sahip bir operatördür. Bir X uzayına ait x ve y elemanlarının iç çarpımı $\langle x, y \rangle$ şeklinde gösterilir ve şu özelliklere sahiptir:

1. Tüm x değerleri için $\langle x, x \rangle \geq 0$. Yalnızca $x = 0$ ise $\langle x, x \rangle = 0$ 'dır. (Pozitiflik)
2. Tüm x ve y değerleri için $\langle x, y \rangle = \langle y, x \rangle$ sağlanır. (Simetri)
3. Tüm $a, b \in \mathbb{R}$ skaler değerleri ve x, y ve z vektörleri için, $\langle ax + by, z \rangle = a\langle x, z \rangle + b\langle y, z \rangle$ sağlanır. (Doğrusallık)

Üzerinde iç çarpım tanımlanmış olan bir vektör uzayı *iç çarpım uzayı* olarak adlandırılır.

Öklit uzayında tanımlı $x = (x_1, x_2, \dots, x_n)$ ve $y = (y_1, y_2, \dots, y_n)$ vektörlerinin iç çarpımı nokta çarpımı olarak tanımlanır:

$$\langle (x_1, x_2, \dots, x_n), (y_1, y_2, \dots, y_n) \rangle = x_1y_1 + x_2y_2 + \dots + x_ny_n = \sum_i x_iy_i.$$

Gerçel fonksiyonların vektör uzayında toplam yerini integrale bırakır. $t \in [0, T]$ olmak üzere $\chi(t)$ ve $\gamma(t)$ fonksiyonları için iç çarpım şu şekilde tanımlanır:

$$\langle \chi, \gamma \rangle = \int_0^T \chi(t)\gamma(t)dt.$$

1.3 Normlu Uzaylar ve Özellikleri

Bir vektör uzayına ait bir x elemanının iç çarpımı onun boyunun bir ölçüsüdür. Bu ölçünün pozitif kare kökü *norm* olarak adlandırılır. Bir x elemanının normu, pozitiflik aksiyomu gereğince $\|x\| \geq 0$ olmak üzere,

$$\|x\|^2 = \langle x, x \rangle,$$

olarak gösterilir.

n -boyutlu bir Öklit uzayına ait bir elemanın normu vektörün uzunluğuna eşittir. Bir χ fonksiyonunun normu \mathcal{L}^2 norm olarak adlandırılır ve

$$\|\chi\| = \left(\int \chi(t)^2 dt \right)^{1/2}$$

olarak gösterilir.

Normun özellikleri iç çarpımın özelliklerine benzerdir ve aşağıdaki gibi özetlenebilir:

1. Tüm x değerleri için $\|x\| \geq 0$. Yalnızca $x = 0$ için $\|x\| = 0$. (Pozitiflik)
2. Tüm x ve y elemanları için $\|x + y\| = \|y + x\|$. (Simetri)

3. Tüm a gerçel sayıları için,
 $\|ax\| = |a|\|x\|$.

X bir vektör uzayı olsun ve $\|\cdot\| : X \rightarrow \mathbb{R}$ fonksiyonu verilsin. Eğer tüm $x, y \in X$ elemanları ve bir $\lambda \in \mathbb{R}$ skaleri için yukarıda bahsedilen özellikler sağlanıyorsa $(X, \|\cdot\|)$ ikilisi bir *normlu uzay* olarak adlandırılır.

Tüm normlu uzaylar metrik uzaydır ancak tüm metrik uzaylar normlu uzay değildir. Benzer şekilde, tüm iç çarpım uzayları normlu uzaydır ancak tersi doğru değildir.

1.4 Hilbert Uzayları

Eğer H tam bir iç çarpım uzayı ise, o zaman H uzayına bir *Hilbert uzayı* denir.

$[0, T]$ üzerinde tanımlı $L^2 = L^2([0, T])$ uzayı, ölçülebilir gerçel değerli fonksiyonların bir kümesidir ve $\int_0^T \chi^2(t)dt < \infty$ özelliğini sağlar. Bu uzaya *karesi integrallenebilir fonksiyonların uzayı* denir. L^2 uzayı, χ ve γ iç çarpımı ile,

$$\langle \chi, \gamma \rangle = \int \chi(t)\gamma(t)dt,$$

ayrılabilir bir Hilbert uzayıdır

B Operatörler ve Fonksiyonel Veri İçin Bazı Yararlı Teoremler

2.1 Doğrusal Operatör, İntegral Operatörü ve Hilbert-Schmidt Operatörü

Bir *doğrusal operatör*, iki vektör uzayı arasında bir eşleştirmedir. Bir $\Psi(x)$ operatörü,

1. $\langle \Psi(x), x \rangle = \langle x, \Psi(x) \rangle$. (Simetri)
2. $\langle \Psi(x), x \rangle \geq 0$. (Pozitif Tanımlılık)

özelliklerini sağlıyorsa simetrik ve pozitif tanımlıdır denir.

$L^2(T)$ uzayı üzerinde $\psi : T \times T \rightarrow \mathbb{R}$ çekirdek fonksiyonu ile tanımlı bir *integral operatörü* $\Psi : \xi \rightarrow \Psi\xi$, $\xi \in L^2(T)$,

$$\Psi(\xi)(t) = \int_T \psi(t, s) \xi(s) ds, \quad \xi \in L_2(T), \quad (\text{B.1})$$

şeklinde gösterilir.

Eğer ψ çekirdek fonksiyonu $\int \int \psi^2(t, s) dt ds < \infty$ özelliğini gerçekleştiriyorsa bu çekirdeğe *Hilbert-Schmidt çekirdeği* denir. Hilbert-Schmidt çekirdeğine sahip doğrusal ve sınırlı bir integral operatörüne *Hilbert-Schmidt operatörü* denir.

Eğer integral operatörünün ilgili çekirdeği simetrik ve pozitif-tanımlı ise integral operatörü de simetrik ve pozitif-tanımlıdır.

2.2 Kovaryans Operatörü ve Özellikleri

$c(t, s) = \mathbb{E}[\chi(t)\chi(s)]$ çekirdekli kovaryans operatörü bir integral operatördür ve

$$\Gamma_\chi \xi(t) = \int c(t, s) \xi(s) ds, \quad \xi \in L^2(T), \quad (\text{B.2})$$

olarak ifade edilir.

$c(s, t)$ kovaryans fonksiyonu simetri

$$c(s, t) = c(t, s), \quad (\text{B.3})$$

ve pozitif tanımlılık

$$\int \int c(t,s)\xi(t)\xi(s)dt ds = \int \int E[\chi^c(t)\chi^c(s)\xi(t)\xi(s)] = E[(\int \chi^c(t)\xi(t)dt)^2] \geq 0, \quad (\text{B.4})$$

özelliklerini sağlar. Bundan dolayı Γ_χ kovaryans operatörü de simetrik ve pozitif tanımlıdır.

$\Gamma_\chi \xi_j = \lambda_j \xi_j$ öz denklemini ele alınacak olursa, kovaryans operatörünün negatif olmayan öz değerleri olduğu gösterilebilir.

$$\lambda_j = \langle \Gamma_\chi \xi_j, \xi_j \rangle = \langle \mathbb{E}[\langle \chi, \xi_j \rangle \chi], \xi_j \rangle = \mathbb{E}[\langle \chi, \xi_j \rangle^2]. \quad (\text{B.5})$$

(B.5) eşitliğinin her iki tarafında da sonlu toplam alınırsa,

$$\sum_{j=1}^{\infty} \lambda_j = \sum_{j=1}^{\infty} \mathbb{E}[\langle \chi, \xi_j \rangle^2] = \mathbb{E}\|\chi\|^2 < \infty, \quad (\text{B.6})$$

(B.6) eşitliğinden de görüldüğü gibi kovaryans operatörünün öz değerleri $\sum_{j=1}^{\infty} \lambda_j < \infty$ koşulunu sağlar.

Bir operatör yalnız ve yalnızca simetrik, pozitif tanımlı ve sınırlı ise bir kovaryans operatörüdür.

2.3 Bir Doğrusal Operatörün Tekil Değer Ayrışımı (Singular Value Decomposition)

$\|\cdot\|$ normunu üreten $\langle \cdot, \cdot \rangle$ iç çarpımı ile ayrılabilir bir Hilbert uzayını göz önüne alalım. \mathcal{L} ,

$$\|\Psi\|_{\mathcal{L}} = \sup\{\|\Psi(x)\| : \|x\| = 1\},$$

normu ile H üzerindeki sınırlı doğrusal operatörlerinin uzayı olsun.

Tekil değer ayrışımı (Singular Value Decomposition, SVD) bir $\Psi \in \mathcal{L}$ kompakt operatörünün η_j and ξ_j ortonormal tabanlar ve λ_j sifira yakınsayan gerçel bir dizi olmak üzere,

$$\Psi(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, \xi_j \rangle \eta_j, \quad (\text{B.7})$$

biçiminde temsil edilebileceğini ifade eder.

2.4 Bir Hilbert-Schmidt Operatörünün Spektral Ayrışımı

Simetrik ve pozitif tanımlı bir Hilbert-Schmidt operatörü, ξ_j 'ler Ψ operatörünün ortonormal taban oluşturan öz fonksiyonları olmak üzere,

$$\Psi(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, \xi_j \rangle \xi_j, \quad x \in H, \quad (\text{B.8})$$

spektral ayrışımını sağlar.

2.5 Parseval Eşitliği

ξ_j , ayırık Hilbert uzayı H 'de ortonormal bir taban olsun.

$$\sum_j \langle \chi, \xi_j \rangle = \|\chi\|^2 = \langle \chi, \chi \rangle, \quad (\text{B.9})$$

biçimindeki eşitlik "Parseval eşitliği" olarak adlandırılır.

2.6 Riesz Temsil Teoremi

Riesz Temsil Teoremi $\Gamma_\chi : L^2(T) \rightarrow L^2(T)$ kovaryans operatörünün,

$$\Gamma_\chi \xi = \mathbb{E}[\langle \chi, \xi \rangle \chi], \quad \xi \in L^2(T), \quad (\text{B.10})$$

şeklinde yazılabileceğini ifade eder.

2.7 Mercer Teoremi

Mercer Teoremi simetrik pozitif-tanımlı bir doğrusal operatörün çekirdeğinin kendi öz değerleri ve özfonksiyonları cinsinden yazılabileceğini ifade eder. Yani, bir Hilbert-Schmidt operatörünün sürekli, simetrik, pozitif tanımlı çekirdeği $\psi(t, s)$, ξ_j 'ler $L^2(T)$ 'de tanımlı ortonormal fonksiyonlar ve λ_j azalan sayıda pozitif sayılardan oluşan özdeğerlerin oluşturduğu bir dizi olmak üzere,

$$\psi(t, s) = \sum_{j=1}^{\infty} \lambda_j \xi_j(t) \xi_j(s) \text{ in } L^2(T \times T), \quad (\text{B.11})$$

biçiminde yazılabilir.

2.8 Karhunen-Loève Açılımı

$\chi(t)$, t 'nin tüm değerleri için $\mathbb{E}[\chi(t)] = 0$ ve $\mathbb{E}[\chi^2(t)] \leq \infty$ ile merkezleştirilmiş bir stokastik süreç olsun. Bu durumda, $\{\xi_j\}_{j \in \mathbb{N}}$, Γ_χ kovaryans operatörünün özfonksiyonlarının oluşturduğu ilgili λ_j özdeğerine ilişkin bir ortonormal tabanı ve f_j , $E[f_j] = 0$, $E[f_i f_j] = 0$ for $i \neq j$ ve $E[f_j^2] = \lambda_j$ eşitliklerini sağlayan, ortogonal rassal değişkenlerinin oluşturduğu bir dizi olmak üzere, $\chi(t)$ fonksiyonu şu şekilde temsil edilebilir (Cuevas, 2014):

$$\hat{\chi}(t) = \sum_{j=1}^{\infty} f_j \xi_j(t). \quad (\text{B.12})$$

C R Programında Yazılan Simülasyon Fonksiyonu

```
library(fda)
library(fda.usc)

#simulation<-function(Xt,rep=100,P=500){
rep=100
P=500
for(p in 1:P){
print(p)
print(date())
mep<-function(newx,newy,mod,exp=FALSE){
pred<-predict(mod,newx)
if(exp) MEP <- ((1 / length(newy)) * sum((newy - exp(pred))^2))
else MEP <- ((1 / length(newy)) * sum((newy - pred)^2))
return(MEP)
}

simexp<-function(data_665s,samp=samp,rep=rep){
amep.665s<-numeric()
for (r in 1:rep){
data_665s.sam<-data_665s[samp[[r]],]
aux<-lm(lY~Xt_665,data=data_665s.sam)
mep.reg665s<-mep(newx=data_665s[-samp[[r]],2,drop=FALSE],newy=exp(data_665s[-
samp[[r]],1]),mod=aux,exp=TRUE)
scalvar<-var(exp(data_665s[-samp[[r]],1]))
amep.665s[r]<-mep.reg665s/scalvar
}
return(median(amep.665s))
}

simfunc<-
function(chi,lgY,mod="bas",samp=samp,rep=rep,bas1=basis1,bas2=basis2,bas2d2=basis2d2){
amep.sim<-numeric()
for (r in 1:rep){
chi.aux<-chi[samp[[r]],]
lgY.aux<-lgY[samp[[r]],]
if(mod=="bas") aux<-fregre.basis(chi.aux,lgY.aux,bas1,bas2)
if(mod=="fpc") aux<-fregre.pc(chi.aux,lgY.aux,l=1:8)
if(mod=="fpcv"){
aux0<-fregre.pc.cv(chi.aux,lgY.aux,8)
```

```

aux<-fregre.pc(chi.aux,lgY.aux,aux0$pc.opt)
}
if(mod=="fpls") aux<-fregre.pls(chi.aux,lgY.aux,l=1:8)
if(mod=="fplscv"){
aux0<-fregre.pls.cv(chi.aux,lgY.aux,8)
aux<-fregre.pls(chi.aux,lgY.aux,aux0$pls.opt)
}
if(mod=="fd2") aux<-fregre.basis(chi.aux,lgY.aux,bas1,bas2d2)

mep.sim<-mep(chi[-samp[[r]],,drop=FALSE],exp(lgY[-samp[[r]]]),mod=aux,exp=TRUE)
scalvar<-var(exp(lgY[-samp[[r]]]))
amep.sim[r]<-mep.sim/scalvar
}
return(median(amep.sim))
}
MEplist_exp<-vector(mode="list",length=P)
MEplist_bas<-vector(mode="list",length=P)
MEplist_pc<-vector(mode="list",length=P)
MEplist_pcv<-vector(mode="list",length=P)
MEplist_d2<-vector(mode="list",length=P)
MEPS<-vector(mode="list",length=P)
lys<-vector(mode="list",length=5)
for(i in 1:5) lys[[i]]<-vector(mode="list",length=3)
###Exponential Regression Model with Rrs 665nm
#a vector composed of rrs at 665 nm
Xt_665<-Xt[,7]
b<-reg.665$coefficients

###generating IY.exp
lys[[1]][[1]]<-IYexp.sd1<-b[1]+Xt_665*b[2]+rnorm(length(Xt_665),sd=sd(reg.665$residuals))
lys[[1]][[2]]<-IYexp.sd2<-b[1]+Xt_665*b[2]+rnorm(length(Xt_665),sd=sd(reg.665$residuals)/5)
lys[[1]][[3]]<-IYexp.sd3<-b[1]+Xt_665*b[2]+rnorm(length(Xt_665),sd=sd(reg.665$residuals)/10)

###generating IY.bas
listXt<-list()
wl<-Xt[,c(1:8)]
listXt$rres<-wl[,]
rres.nm<-as.numeric(substring(names(listXt$rres),5))
Xtfdata<-fdata(listXt$rres,argvals=rres.nm,names=list(main='Water reflectances of box
data',xlab='Wavelength',ylab='Water Reflectances'))
ts<-Xtfdata[["argvals"]]
bsXt<-create.bspline.basis(rangeval=range(ts),nbasis=8)
fdXt <- Data2fd(argvals=ts,y=t(Xtfdata$data),basisobj=bsXt)
Cnew<-fdXt$coefs
Xtfd<-fdata2fd(Xtfdata,nbasis = 8)
J<-flrly$J
B<-flrly$beta.est$coefs
a<-flrly$a.est
lys[[2]][[1]]<-IYbas.sd1<-a+t(Cnew)%*%J%*%B+rnorm(n=nrow(t(Cnew)),sd=sd(flirly$residuals))
lys[[2]][[2]]<-IYbas.sd2<-a+t(Cnew)%*%J%*%B+rnorm(n=nrow(t(Cnew)),sd=sd(flirly$residuals)/5)
lys[[2]][[3]]<-IYbas.sd3<-a+t(Cnew)%*%J%*%B+rnorm(n=nrow(t(Cnew)),sd=sd(flirly$residuals)/10)

```

```

##generating IY.pc
Xtcen<-fdata.cen(Xtfdata) Xtfdatacen<-Xtcen$Xcen
Fpc<-inprod.fdata(Xtfdatacen,fpcly$fdata.comp$rotation)
bpc<-fpcly$coefficients[2:9]
sdpc<-sd(fpcly$residuals)
lys[[3]][[1]]<-IYpc.sd1<-mean(fpcly$y)+Fpc%%as.matrix(bpc)+rnorm(n=9777,0,sd=sdpc)
lys[[3]][[2]]<-IYpc.sd2<-mean(fpcly$y)+Fpc%%as.matrix(bpc)+rnorm(n=9777,0,sd=sdpc/5)
lys[[3]][[3]]<-IYpc.sd3<-mean(fpcly$y)+Fpc%%as.matrix(bpc)+rnorm(n=9777,0,sd=sdpc/10)

##generating IY.pcv
Fpcv<-inprod.fdata(Xtfdatacen,fpclv$fregre.pc$fdata.comp$rotation)
bpcv<-fpclv$fregre.pc$coefficients[2:3]
sdpcv<-sd(fpclv$fregre.pc$residuals)
lys[[4]][[1]]<-IYpcv.sd1<-
mean(fpclv$fregre.pc$y)+Fpcv[,c(2,5)]%%as.matrix(bpcv)+rnorm(n=9777,0,sd=sdpcv)
lys[[4]][[2]]<-IYpcv.sd2<-
mean(fpclv$fregre.pc$y)+Fpcv[,c(2,5)]%%as.matrix(bpcv)+rnorm(n=9777,0,sd=sdpcv/5)
lys[[4]][[3]]<-IYpcv.sd3<-
mean(fpclv$fregre.pc$y)+Fpcv[,c(2,5)]%%as.matrix(bpcv)+rnorm(n=9777,0,sd=sdpcv/10)

##generating IY.der2
der2Xt<-fdata.deriv(Xtfdata, nbasis=8, nderiv=2)
listder2Xt<-list()
wld<-der2Xt[,c(1:8)]
listder2Xt$rres<-wld[,]
rss.nm<-as.numeric(substring(names(listder2Xt$rres),5))
ts<-Xtfdata[["argvals"]]

bsXt<-create.bspline.basis(rangeval=range(ts),nbasis=8)
der2Xtfd <- Data2fd(argvals=ts,y=t(der2Xt$data),basisobj=bsXt)
Cd2<-der2Xtfd$coefs
Jd2<-regd2ly$J
Bd2<-regd2ly$coefficients[2:5]
ad2<-regd2ly$coefficients[1]
lys[[5]][[1]]<-IYd2.sd1<-ad2+t(Cd2)%*%Jd2)%*%Bd2+rnorm(n=nrow(t(Cd2)),sd=sd(regd2ly$residuals))
lys[[5]][[2]]<-IYd2.sd2<-
ad2+t(Cd2)%*%Jd2)%*%Bd2+rnorm(n=nrow(t(Cd2)),sd=sd(regd2ly$residuals)/5)
lys[[5]][[3]]<-IYd2.sd3<-
ad2+t(Cd2)%*%Jd2)%*%Bd2+rnorm(n=nrow(t(Cd2)),sd=sd(regd2ly$residuals)/10)
samp25<-samp50<-samp100<-list()
for(r in 1:rep){
  samp25[[r]]<-sample(1:length(data_665s[,1]),25)
  samp50[[r]]<-sample(1:length(data_665s[,1]),50)
  samp100[[r]]<-sample(1:length(data_665s[,1]),100)
}
cont<-TRUE
for(z in 1:5){
  if(cont){
    MEPaux<-as.data.frame(matrix(nrow=3,ncol=21,NA))
    for(i in 1:3){
      data_665s$Y<-lys[[z]][[i]]
    }
  }
}

```

```

for(j in 1:3){
  if(j==1) sampl<-samp25
  if(j==2) sampl<-samp50
  if(j==3) sampl<-samp100
  if(cont){
    MEPaux[i,((j-1)*7)+1]<-try(simexp(data_665s,samp=sampl,rep=rep),TRUE)
    cont<-is.numeric(MEPaux[i,((j-1)*7)+1])
  }
  if(cont){
    MEPaux[i,((j-1)*7)+2]<-
    try(simfunc(chi=Xtfddata,lgY=as.matrix(lys[[z]][[i]]),samp=sampl,rep=rep,mod="bas"),TRUE)
    cont<-is.numeric(MEPaux[i,((j-1)*7)+2])
  }
  if(cont){
    MEPaux[i,((j-1)*7)+3]<-
    try(simfunc(Xtfddata,as.matrix(lys[[z]][[i]]),samp=sampl,rep=rep,mod="fpc"),TRUE)
    cont<-is.numeric(MEPaux[i,((j-1)*7)+3])
  }
  if(cont){
    MEPaux[i,((j-1)*7)+4]<-
    try(simfunc(Xtfddata,as.matrix(lys[[z]][[i]]),samp=sampl,rep=rep,mod="fpcv"),TRUE)
    cont<-is.numeric(MEPaux[i,((j-1)*7)+4])
  }
  if(cont){
    MEPaux[i,((j-1)*7)+5]<-
    try(simfunc(Xtfddata,as.matrix(lys[[z]][[i]]),samp=sampl,rep=rep,mod="fpls"),TRUE)
    cont<-is.numeric(MEPaux[i,((j-1)*7)+5])
  }
  if(cont){
    MEPaux[i,((j-1)*7)+6]<-
    try(simfunc(Xtfddata,as.matrix(lys[[z]][[i]]),samp=sampl,rep=rep,mod="fplscv"),TRUE)
    cont<-is.numeric(MEPaux[i,((j-1)*7)+6])
  }
  if(cont){
    MEPaux[i,((j-1)*7)+7]<-
    try(simfunc(der2Xt,as.matrix(lys[[z]][[i]]),samp=sampl,rep=rep,mod="fd2"),TRUE)
    cont<-is.numeric(MEPaux[i,((j-1)*7)+7])
  }
}
}
}
colnames(MEPaux)<-
paste(rep(c('exp','bas','fpc','fpcv','fpls','fplscv','der2'),3),'_',rep(c(25,50,100),each=7),sep="")
if(z==1){
  rownames(MEPaux)<-c('IYexp.sd1','IYexp.sd2','IYexp.sd3')
  MEPlist_exp[[p]]<-MEPaux
}
if(z==2){
  rownames(MEPaux)<-c('IYbas.sd1','IYbas.sd2','IYbas.sd3')
  MEPlist_bas[[p]]<-MEPaux
}
if(z==3){
  rownames(MEPaux)<-c('IYpc.sd1','IYpc.sd2','IYpc.sd3')
}
}
}

```



```

MEPlist_pc[[p]]<-MEPaux
}
if(z==4){
rownames(MEPaux)<-c('IYpcv.sd1','IYpcv.sd2','IYpcv.sd3')
MEPlist_pcv[[p]]<-MEPaux
}
if(z==5){
rownames(MEPaux)<-c('IYd2.sd1','IYd2.sd2','IYd2.sd3')
MEPlist_d2[[p]]<-MEPaux
}
if(z==1) MEPSaux<-MEPaux
else MEPSaux<-rbind(MEPSaux,MEPaux)
}
}
MEPS[[p]]<-MEPSaux
if(cont){
arch<-paste("iter_NAD_Y_",p,".csv",sep="")
write.csv2(MEPSaux,arch)
}
}
#}

for (p in 1:P){
if (p==1) sumMEP=as.matrix(MEPS[[1]])
else sumMEP=sumMEP+as.matrix(MEPS[[p]])
}
meanMEP<-sumMEP/P
write.csv2(sumMEP,"sumMEP_500.csv")
write.csv2(meanMEP,"meanMEP_500.csv")

```

ÖZGEÇMİŞ

Ad Soyad: Nihan ACAR DENİZLİ

Doğum Yeri ve Tarihi: 07/03/1985 BURSA

E-Posta: nihan.acar@msgsu.edu.tr

Lisans: Marmara Üniversitesi Matematik Bölümü

Y. Lisans: Mimar Sinan Güzel Sanatlar Üniversitesi İstatistik Programı

TEZDEN TÜRETİLEN BİLDİRİLER

- **Acar-Denizli N.**, Başarır G., Caballero I., Delicado P., 2016: The Application Of Functional Linear Models On Remote Sensing Data In Oceanography. *2016 CRoNoS Summer Course On Functional Data Analysis*, Ağustos 26-28, 2016, Oviedo, İspanya.
- **Acar-Denizli N.**, Delicado P., Başarır G., Caballero I., 2016: A simulation based comparative study of functional linear regression models on remote sensing data. *ERCIM 2016*, Aralık 9-11, 2016, Sevilla, İspanya. (Alicia Nieto-Reyes tarafından organize edilen "EO117: Functional data analysis" oturumuna davet edildi.)
- **Acar-Denizli N.**, Delicado P., Başarır G., Caballero I., 2017: The Evaluation of the Performance of Functional Linear Regression Models on a Remote Sensing Application. *61st World Statistics Congress, ISI 2017*, Temmuz 16-21, 2016, Marakeş, FAS. (Kristina Lurz ve Nihan Acar-Denizli tarafından organize edilen "IPS112: Young researchers session on Functional Data Analysis" oturumuna kabul edildi.)