

T.C.
ONDOKUZ MAYIS ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



ÇOK DEĞİŞKENLİ FONKSİYONEL VERİLERİN KÜMELEME
YÖNTEMLERİ İLE İNCELENMESİ: TÜRKİYE İKLİM ÖRNEĞİ

ÇAĞLAR SÖZEN

Doktora TEZİ

T.C.
ONDOKUZ MAYIS ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

DOKTORA TEZİ

ÇOK DEĞİŞKENLİ FONKSİYONEL VERİLERİN KÜMELEME
YÖNTEMLERİ İLE İNCELENMESİ: TÜRKİYE İKLİM ÖRNEĞİ

ÇAĞLAR SÖZEN

İSTATİSTİK ANABİLİM DALI

SAMSUN
2019

Her hakkı saklıdır.

TEZ ONAYI

Çağlar Sözen tarafından hazırlanan “Çok Değişkenli Fonksiyonel Verilerin Kümeleme Yöntemleri ile İncelenmesi: Türkiye İklim Örneği” adlı tez çalışması 19/07/2019 tarihinde aşağıdaki jüri tarafından Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı’nda **Doktora Tezi** olarak kabul edilmiştir.

Danışman Doç. Dr. Yüksel ÖNER
Ondokuz Mayıs Üniversitesi
İstatistik Anabilim Dalı

Jüri Üyeleri

Başkan	Prof. Dr. Vedat SAĞLAM Ondokuz Mayıs Üniversitesi İstatistik Anabilim Dalı
Üye	Doç. Dr. Mehmet GÜRCAN Fırat Üniversitesi İstatistik Anabilim Dalı
Üye	Doç. Dr. Yüksel ÖNER Ondokuz Mayıs Üniversitesi İstatistik Anabilim Dalı
Üye	Doç. Dr. Necati Alp ERİLLİ Sivas Cumhuriyet Üniversitesi Ekonometri Anabilim Dalı
Üye	Dr. Öğr. Üyesi Tolga ZAMAN Çankırı Karatekin Üniversitesi İstatistik Anabilim Dalı

Yukarıdaki sonucu onaylıyorum. .../.../20...

Prof. Dr. Bahtiyar ÖZTÜRK
Enstitü Müdürü

ETİK BEYAN

Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım bu tez içindeki bütün bilgilerin doğru ve tam olduğunu, bilgilerin üretilmesi aşamasında bilimsel etiğe uygun davrandığımı, yararlandığım bütün kaynakları atıf yaparak belirttiğimi beyan ederim.

19.07.2019

Çağlar SÖZEN

ÖZET

Doktora Tezi

ÇOK DEĞİŞKENLİ FONKSİYONEL VERİLERİN KÜMELEME YÖNTEMLERİ İLE İNCELENMESİ: TÜRKİYE İKLİM ÖRNEĞİ

Çağlar SÖZEN

Ondokuz Mayıs Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Doç. Dr. Yüksel ÖNER

İstatistiksel analizlerde, analiz edilmek istenen gözlemlerin sayısı arttıkça, bu gözlemler reel değerli bir fonksiyondan örneklenmekte olduğu kabul edilmektedir. Bu durumda türetilen veriler bir fonksiyonel yapının örnekleri olmaktadır. Araştırmada yer alan birimlere ait özelliklerin belirli bir zaman içerisinde incelenmesi durumunda, her birime ait veri yapısı bir fonksiyonel yapı gösterecektir. Bu tür verilerin analiz edilmesinde klasik tek değişkenli ve çok değişkenli yöntemler yetersiz kalmaktadır. Bu yüzden bu türde verilerin analiz edilmesinde fonksiyonel veri analizinin kullanılması daha uygun olacaktır. Fonksiyonel veri analizinde ilk önce ayrı noktalarda gözlenen verileri reel fonksiyonlardan oluşan bir veri örneğine dönüştürmek gerekir. Fonksiyonların oluşturulması sürecinde baz fonksiyon ve pürüzlü ceza yaklaşımı kullanılmaktadır. Oluşturulan bireysel fonksiyonların analizi için fonksiyonel temel bileşenler analizi, kümeleme analizi ve fonksiyonel ANOVA kullanmak mümkündür.

Bu çalışmada Türkiye’de bulunan 81 ilin bazı iklimsel değişkenlerine ait (günlük ortalama sıcaklık, günlük toplam yağış, günlük sıcaklık değişimi, günlük güneşlenme süresi ve günlük buharlaşma süresi) veriler 365 ayrı noktada incelendi. Veriler periyodik bir yapı izlediklerinden baz fonksiyon yaklaşımı olarak Fourier baz fonksiyonu kullanıldı. Veriler Fourier baz ve pürüzlü ceza yaklaşımları kullanılarak sürekli bir fonksiyon haline dönüştürüldü. Değişkenlere ait düzgünleştirme parametresinin bulunmasında genelleştirilmiş çapraz geçerlilik yöntemi kullanıldı. Değişkenlere ait değişimlerin temel nedenini belirleyebilmek için fonksiyonel temel bileşenler analizinden yararlandı. Oluşturulan bireysel fonksiyonların sınıflandırılmasında model tabanlı kümeleme yöntemi kullanıldı. Kümelerin, ortalama fonksiyonları bakımından aralarında istatistiksel açıdan fark olup olmadığının belirlenmesi işleminde FANOVA uygulandı.

Temmuz 2019, 91 sayfa

Anahtar Kelimeler: Fonksiyonel veri, Fonksiyonel veri analizi, Fonksiyonel kümeleme, Fonksiyonel ANOVA, Fonksiyonel model tabanlı kümeleme.

ABSTRACT

Doctoral Dissertation

INVESTIGATION OF MULTIVARIATE FUNCTIONAL DATA WITH CLUSTERING METHODS: THE CASE OF TURKEY CLIMATE

Çağlar SÖZEN

Ondokuz Mayıs University
Graduate School of Sciences
Department of Statistics

Supervisor: Assoc. Prof. Dr. Yüksel ÖNER

Observations to be analyzed are sampled from a real-valued function as they increase in number. The derived data are, therefore, samples of a functional structure. When the properties of units are analyzed within a certain time period, the data structure of each unit corresponds to a functional structure. Classical single variable and multivariate methods fall short of analyzing such data. It would, therefore, be more appropriate to use functional data analysis, in which data observed at discrete time points should first be converted into a data sample consisting of real functions. Base function and roughness penalty approaches are used to create functions. Functional principal component analysis, cluster analysis and functional ANOVA can then be used for the analysis of individual functions.

This study examined the mean daily temperature, daily total precipitation, daily temperature changing, daily sunshine duration and daily evaporation data of 81 cities in Turkey at 365 discrete time points. Data had a periodic structure, and therefore, the Fourier base function was used. Data were transformed into a continuous function using Fourier base and roughness penalty approaches. Generalized cross validity method was used to find the smoothing parameter of variables. Functional principal component analysis was used to determine the main cause of the change in variables. Model-based clustering method was used to classify individual functions. Functional ANOVA was used to determine whether there was a statistical difference in mean functions between clusters.

July 2019, 91 pages

Key Words: Functional data, Functional data analysis, Functional clustering, Functional ANOVA, Functional model-based clustering.

ÖNSÖZ VE TEŞEKKÜR

Akademik eğitim boyunca gerek yüksek lisans gerekse doktora eğitimim boyunca hiçbir konuda desteğini esirgemeyen danışman hocam Sayın Doç. Dr. Yüksel Öner'e teşekkürü bir borç bilirim.

Beni bugünlere getirmek için hiçbir fedakârlıktan kaçınmayan aileme sonsuz teşekkür ederim.

Doktora eğitimim boyunca destekten kaçınmayan arkadaşım Dr. Öğr. Üyesi Hasan BULUT'a teşekkür ederim.

Temmuz 2019, Samsun

Çağlar SÖZEN

İÇİNDEKİLER

ÖZET.....	iii
ABSTRACT.....	iv
ÖNSÖZ VE TEŞEKKÜR.....	v
İÇİNDEKİLER.....	vi
SİMGELER VE KISALTMALAR.....	viii
ŞEKİLLER DİZİNİ.....	ix
ÇİZELGELER DİZİNİ.....	x
1. GİRİŞ.....	1
2. FONKSİYONEL VERİLERİN MODELLENMESİ.....	7
2.1. En Küçük Kareler Yöntemi.....	7
2.2. Baz Fonksiyon Yaklaşımı.....	8
2.2.1. Fourier baz fonksiyon yaklaşımı.....	8
2.2.2. Polinom baz fonksiyon yaklaşımı.....	9
2.2.3. B-splayn baz fonksiyon yaklaşımı.....	9
2.2.4. Dalgacık baz fonksiyon yaklaşımı.....	10
2.3. Düzleştirme Yöntemleri.....	11
2.3.1. Kernel düzleştirme.....	11
2.3.2. Pürüzlülük cezası.....	12
2.4. Düzgünleştirme Parametresinin Belirlenmesi.....	13
3. FONKSİYONEL VERİLERİN KÜMELENMESİNDE KULLANILAN YÖNTEMLER.....	15
3.1. Ham Veri Yöntemleri.....	15
3.2. Filtreleme Yöntemleri.....	16
3.2.1. FTBA puanlarında filtreleme yöntemi.....	16
3.2.2. Baz fonksiyon katsayılarını filtreleme yöntemi.....	18
3.3. Uyarlanabilir Yöntemler.....	18
3.4. Uzaklığa Dayalı Yöntemler.....	19
3.5. Model Tabanlı Kümeleme Yöntemleri.....	20
3.5.1. Karışım modeli.....	21
4. FONKSİYONEL VERİLER İÇİN ÖRNEK PROBLEMLER.....	24
4.1. Stokastik Süreçler.....	24
4.1.1. Gauss süreçleri.....	24
4.1.2. Wishart süreçleri.....	25
4.1.3. χ^2 -tipi karışımlar.....	25
4.1.4. Kümülanlar.....	26
4.1.5. F -tipi karışımlar.....	27
4.2. Fonksiyonel Veriler için Tek Örnek Problemi.....	27
4.2.1. Tek örnek için noktasal testler.....	28
4.2.2. Tek örnek için L_2 -norm tabanlı test.....	29
4.2.3. Tek örnek için F -tipi test.....	31
4.2.4. Tek örnek için bootstrap testi.....	32
4.3. Fonksiyonel Veriler için İki Örnek Problemi.....	32
4.3.1. İki örnek için eksensel test fonksiyonu.....	33
4.3.2. İki örnek için noktasal testler.....	34
4.3.3. İki örnek için L_2 -norm tabanlı test.....	35
4.3.4. İki örnek için F -tipi test.....	37
4.3.5. İki örnek için bootstrap testleri.....	38

4.4.	Fonksiyonel Veriler için k Örnek Problemi (FANOVA)	39
4.4.1.	Ana etki testi	40
4.4.2.	Post-hoc testi	40
4.4.3.	Kontrast testi	41
4.4.4.	Grup ortalama ve kovaryans fonksiyonlarının tahmini.....	41
4.4.5.	Tek yönlü varyans analizinin varsayımları	42
4.5.	Tek Yönlü ANOVA Problemi için Testler	43
4.5.1.	Tek yönlü ANOVA problemi için noktasal F -testi.....	43
4.5.2.	Tek yönlü ANOVA problemi için L_2 -norm tabanlı testler	44
4.5.3.	Tek yönlü ANOVA problemi için F -tipi test.....	45
4.5.4.	Tek yönlü ANOVA problemi için bootstrap testleri.....	46
5.	BULGULAR VE TARTIŞMA	49
5.1.	Değişkenlerin Bireysel Fonksiyonlarının İncelenmesi	50
5.2.	Değişkenlerin Simülasyon ile Düzgünleştirme Parametrelerinin Belirlenmesi	55
5.3.	İllere Göre İklimsel Verilerin Fonksiyonel Temel Bileşen Analizi ile Değerlendirilmesi	60
5.4.	Çok Değişkenli Fonksiyonel Verilerin Model Tabanlı Kümeleme Yöntemi ile Kümelenmesi.....	73
5.5.	Çok Değişkenli Fonksiyonel Verilerin Fonksiyonel ANOVA ile İncelenmesi.....	79
6.	SONUÇ VE ÖNERİLER	82
	KAYNAKLAR	86
	ÖZGEÇMİŞ	91

SİMGELER VE KISALTMALAR

TBA	: Temel Bileşenler Analizi
FVA	: Fonksiyonel Veri Analizi
ÇDVA	: Çok Değişkenli Veri Analizi
FTBA	: Fonksiyonel Temel Bileşenler Analizi
EKK	: En Küçük Kareler
iz	: Matrisin İzi
sd	: Serbestlik Derecesi
HDDC	: Yüksek Boyutlu Veri Kümeleme
EM	: Expectation- Maximization
HKT	: Hata Kareler Toplamı
ω	: Kaydırma Parametresi
τ_i	: Düğüm Noktası
h	: Ölçek Parametresi
ψ	: Ana Dalgacık Fonksiyonu
λ	: Düzgünleştirme Parametresi
GCV	: Genelleştirilmiş Çapraz Geçerlilik Yöntemi
R	: Pürüzlü Ceza Matrisi
FANOVA	: Fonksiyonel ANOVA
n	: Gözlem Sayısı
c	: Baz Fonksiyonlarına Karşılık Gelen Katsayılar Matrisi
Φ	: Baz Fonksiyonlarından Oluşan Bir Set
$\eta(t)$: Ortalama Fonksiyonu
$\gamma(s, t)$: Kovaryans Fonksiyonu
$W(s, t)$: Wishart Süreci
ϕ	: Standart Gauss Yoğunluğu
λ	: Özdeğer
AIC	: Akaike Information Criterion
L_2N	: L_2 -Norm Tabanlı Naive Yöntem
L_2B	: L_2 -Norm Tabanlı Yanlı İndirgenmiş Yöntem
L_2b	: L_2 -Norm Tabanlı Bootstrap Testi
PF	: Noktasal F Testi
FT	: F Tipi Test
Fb	: F Tipi Bootstrap Testi

ŞEKİLLER DİZİNİ

Sayfa

Şekil 5. 1. Günlük Ortalama Sıcaklık için 81 Bireysel Fonksiyon	51
Şekil 5. 2. Günlük Toplam Yağış Miktarı için 81 Bireysel Fonksiyon	52
Şekil 5. 3. Günlük Sıcaklık Değişimi için 81 Bireysel Fonksiyon	53
Şekil 5. 4. Günlük Buharlaşma Miktarı için 81 Bireysel Fonksiyon	54
Şekil 5. 5. Günlük Güneşlenme Süresi için 81 Bireysel Fonksiyon	55
Şekil 5. 6. Günlük Ortalama Sıcaklık için GCV Simülasyonu	56
Şekil 5. 7. Günlük Toplam Yağış için GCV Simülasyonu	57
Şekil 5. 8. Günlük Sıcaklık Değişimi için GCV Simülasyonu	58
Şekil 5. 9. Günlük Buharlaşma Miktarı için GCV Simülasyonu	59
Şekil 5. 10. Günlük Güneşlenme Süresi için GCV Simülasyonu	60
Şekil 5. 11. Günlük Ortalama Sıcaklık için Temel Bileşen Fonksiyonları	61
Şekil 5. 12. Günlük Toplam Yağış için Temel Bileşen Fonksiyonları	64
Şekil 5. 13. Günlük Sıcaklık Değişimi için Ana Bileşen Fonksiyonları	66
Şekil 5. 14. Günlük Buharlaşma Miktarı için Temel Bileşen Fonksiyonları	69
Şekil 5. 15. Günlük Güneşlenme Süresi için Temel Bileşen Fonksiyonları	71
Şekil 5. 16. Günlük Ortalama Sıcaklık için Kümeleme Fonksiyonları	76
Şekil 5. 17. Günlük Toplam Yağış Miktarı için Kümeleme Fonksiyonları	77
Şekil 5. 18. Günlük Sıcaklık Değişimi için Kümeleme Fonksiyonları	77
Şekil 5. 19. Günlük Buharlaşma Miktarı için Kümeleme Fonksiyonları	78
Şekil 5. 20. Günlük Güneşlenme Süresi için Kümeleme Fonksiyonları	78

ÇİZELGELER DİZİNİ

Sayfa

Çizelge 3. 1. Model Parametre Sayısı.....	22
Çizelge 5. 1. Günlük Ortalama Sıcaklık için Simülasyon Sonuçları	56
Çizelge 5. 2. Günlük Toplam Yağış için Simülasyon Sonuçları	57
Çizelge 5. 3. Günlük Sıcaklık Değişimi için Simülasyon Sonuçları	58
Çizelge 5. 4. Günlük Buharlaşma Miktarı için Simülasyon Sonuçları	59
Çizelge 5. 5. Günlük Güneşlenme Süresi için Simülasyon Sonuçları	60
Çizelge 5. 6. Günlük Ortalama Sıcaklık için Temel Bileşen Skorları	62
Çizelge 5. 7. Günlük Toplam Yağış için Temel Bileşen Skorları	64
Çizelge 5. 8. Günlük Sıcaklık Değişimi için Temel Bileşen Skorları	67
Çizelge 5. 9. Günlük Buharlaşma Miktarı için Temel Bileşen Skorları	69
Çizelge 5. 10. Günlük Güneşlenme Süresi için Temel Bileşen Skorları	72
Çizelge 5. 11. Küme Sayısının Belirlenmesi Simülasyonu	74
Çizelge 5. 12. Model Tabanlı Kümeleme Sonuçları.....	75
Çizelge 5. 13. İllerin ait Olduğu Küme Olasılıkları.....	76
Çizelge 5. 14. Değişkenlere göre Tek Yönlü Fonksiyonel ANOVA Test İstatistiği Sonuçları ve p Değerleri	81

1. GİRİŞ

Günümüzde geniş hacimli örneklerle çalışabilmek ve bunların istatistiksel yöntemlerle analizini yapmak mümkündür. İncelenen çalışma sahası genişledikçe diğer bir ifadeyle örneğe dâhil edilen gözlem sayısı arttıkça, ayrık noktalarda gözlenen verilerin reel bir fonksiyondan örneklendiği varsayılmaktadır. Dolayısıyla bu gözlemler fonksiyonel veriler ismini almaktadır (Ramsay ve Dalzell, 1991).

Fonksiyonel veriler, bir takım denekler, zaman, mekân veya başka bir süreç boyunca gözlemlenerek elde edilir. Fonksiyonel veriler, sürekli veya kesikli hale getirilmiş bir zaman aralığında, rastgele bir değişkenin (ya da rastgele değişkenlerin) gözlemleri olan fonksiyonlarla veya eğrilerle temsil edilmektedir. Fonksiyonel veriler, çok değişkenli verilerin sonlu boyuttan sonsuz boyuta doğal bir genellemesidir. Fonksiyonel gözlem kümeleri, değişken sayısının gözlem sayısından anlamlı derecede büyük olduğu veri örnekleridir.

Belirli bir zaman aralığında örneklenen veriler, gözlemlerin kaynağı olduğu varsayılan temel bir fonksiyonun davranışını yansıtır. Fonksiyonel veri analizi (FVA) 1990’larda geliştirilmiş bir istatistiksel yöntemdir (Ramsay ve Dalzell, 1991). FVA veya diğer bir ifadeyle “eğrilerle veri analizi” istatistik alanında geniş bir uygulama alanına sahiptir. Son yıllarda, fonksiyonel verilerin analizine özel bir ilgi gösterilmektedir. FVA’da, ayrı ayrı gözlenen veri dizileri fonksiyonlarla temsil edilir ve tek bir bütün olarak analiz edilir. FVA’nın amacı, geleneksel istatistiksel yöntemlerle temel olarak aynıdır. Diğer taraftan FVA’nın avantajı, baz fonksiyonlarından ve türevlerden elde edilebilecek ek bilgilerdir. Herhangi bir istatistiksel analizde ilk adım olarak, veriler tanımlayıcı istatistiklerin bir fonksiyonel versiyonuyla ele alınır. Böylece fonksiyonel veri kümelerini tanımaya yönelik değer yapıları ortaya çıkarılabilmektedir (Ramsay ve Silverman, 2002a). FVA’nın amaçları ise genel olarak şunlardır:

1. İleri analizlere yardım edecek şekilde verileri sunmak.
2. Çeşitli özellikleri öne çıkaracak biçimde verileri sergilemek.
3. Veriler arasındaki önemli modelleri ve noktaları tanımlamak.
4. Bağımlı değişkendeki ilişkiyi bağımsız değişken yapısını kullanarak açıklamak.

5. İki veya daha fazla veri setini karşılaştırmak.

Bu amaçları gerçekleştirmek, en uygun tekniğin uygulanmasıyla mümkündür. Verilerin yeni ve farklı görünümünü ortaya çıkartmak için doğru tekniklerin seçimi son derece önemlidir. Bu aynı zamanda beklenmeyen özelliklerin ortaya çıkarılmasında önemli rol oynamaktadır.

FVA'nın ilk adımı, ayrı noktalarda gözlemlenen verilerin fonksiyonlara veya eğrilere dönüştürülmesidir. Bu nedenle, farklı eğri tahmini prosedürleri FVA'da önemli bir rol oynamaktadır. Günümüzde, birçok bilim dalında özel izleme ekipmanları kullanılmaktadır ve doğru şekilde ölçüm yapıldığında ölçüm hataları ihmal edilebilir düzeyde olmaktadır. Eğri uydurma işlemi genellikle interpolasyon yöntemi ile gerçekleştirilebilmektedir. Bununla birlikte, gözlemlenen verilerde ölçüm hataları mevcutsa, fonksiyonların oluşturulmasında düzgünleştirme (smoothing) yöntemi kullanılmaktadır (Ramsay ve Silverman, 2005).

Fonksiyonel veriler, birçok durumda çok değişkenli veriler olarak ele alınabilir, böylece klasik çok değişkenli veri analizi (ÇDVA) yöntemleri uygulanabilir. ÇDVA yöntemleri bir deneğin ya da objenin ölçümlerinin altında yatan amacın bir fonksiyon olduğunu göz ardı etmektedir. Buna ek olarak, doğrudan ÇDVA yöntemlerinin uygulanması çeşitli zorlukları da beraberinde getirmektedir. Gözlenen fonksiyonel verilerin örnekleme zaman noktaları farklı deneklerde ya da objelerde aynı değildir. Örnekleme zaman noktaları eşit aralıklı olmayabilir. Örnekleme zaman noktalarının sayısı, fonksiyonel verilerin bir örneğindeki denek sayısından daha fazla olabilir. Bu durumlarda, FVA'nın uygulanması son derece doğaldır.

Klasik ÇDVA'da, çıkarımlar genellikle normal dağılıma ve Mahalanobis uzaklığına bağlıdır. FVA'da, Mahalanobis uzaklığı yanılıcıdır, bu nedenle L_2 -normu kullanılmaktadır. Aslında, FVA'daki Gauss süreci ve L_2 -normu, klasik ÇDVA'da normal dağılım ve Mahalanobis uzaklığının rolünü oynamaktadır (Zhang ve Sun, 2010).

Özellikle Ramsay ve Silverman (2002a) ile beraber ivme kazanan FVA'nın önemi her geçen gün artmaktadır. Bunun temel nedeni, interpolasyon yöntemi sayesinde kayıp verilerle baş edilebilmesi ve oluşturulan fonksiyonların istenilen dereceden türevlerinin incelenebilmesine olanak sağlamasıdır.

FVA birçok bilim dalında kullanılmaktadır. FVA uygulamalarını değişik bilim dallarında örneklerle açıklamak mümkündür. Örneğin; psikoloji alanında araştırmacılar, müziğin hem sesli dinlenmesinin hem de video ile izlenmesinin, sadece müzik dinlemeye veya sadece video izlemeye kıyasla insanlar üzerinde nasıl bir duygusal etki bıraktığını araştırmak istemişlerdir (Ramsay, 2004). Fonksiyonel temel bileşenler analizi (FTBA) bu gruplar arasında bulunan farklılığın bileşenlerini ortaya çıkarmak için kullanılmıştır.

Başka bir örnekte nörogörüntüleme alanından verilebilir. FVA, beyin ve zihin arasındaki ilişkinin araştırılması ve belirlenmesinde kullanılmıştır (Tian, 2010). Tian, FVA metotlarını nörogörüntüleme alanında kullanımını araştırmış ve bu alanda FTBA yöntemini kullanmıştır.

Diğer bir örnekte Ramsay (2005) tarafından mühendislik alanından verilmiştir. Ramsay'ın bu modelinde Kuzey Vancouver'da yağın yağmur miktarı ile yeraltı su seviyesi, yeraltı sularının seviyesinde olan değişiklik miktarını belirlemede kullanılmıştır. Bu analiz için FVA'nın kullanılması diğer istatistiksel tekniklere göre daha avantajlıdır. Çünkü diğer teknikler yeraltı su seviyesinde meydana gelen değişikliklerin tespitinde yetersiz kalmaktadırlar. FVA, yeraltı su seviyesinde meydana gelen değişimleri incelemekte ve yeraltı su seviyesi belli bir seviyeye gelir gelmez o yörelin sel felaketine karşı önlem alınmasını öngörebilmektedir. Bu problem, fonksiyonel veriler için iki örnek problemi olarak tanımlanmaktadır.

Berkeley Sağlık Hizmetleri Üreme Epidemiyoloji Bölümünde yapılan bir araştırmada, gebeliğe elverişsiz ve gebeliğe elverişli progesteron eğrileri arasında anlamlı bir fark olup olmadığını eğer fark varsa, hangi döneme rastladığı araştırılmıştır (Ramsay vd, 2009). Bu problem de, fonksiyonel veriler için iki örnek problemi olarak tanımlanmaktadır.

Bir başka çalışmada cinsiyet farkının, bir çocuğun büyüme sürecinde etkisi olup olmadığını kontrol etmek için yapılmıştır. Kız ve erkek çocukların ortalama büyüme eğrileri arasında anlamlı bir farklılık olup olmadığını test etmek için FVA'dan yararlanılmıştır (Ramsay ve Silverman, 2002b). Bu da, fonksiyonel veriler için iki örnek problemi olarak tanımlanmaktadır.

FVA, Doğu, Batı ve Kuzey hava istasyonlarının ortalama sıcaklık eğrileri arasında bir konum etkisi olup olmadığının araştırılmasında kullanılmıştır (Ramsay

ve Silverman, 2002a). Bu problem fonksiyonel veriler için tek yönlü ANOVA problemidir.

Otomobil sürücülerinin hareketini incelemek için, Michigan Üniversitesi, Ergonomi Merkezi araştırmacıları, bir otomobil sürücüsünün bir test arabası içindeki yirmi lokasyon hareketi hakkında veri toplamıştır. Araştırmacılar, sürücünün üst ve alt kolları arasındaki sağ dirseğin oluşturduğu açıdan, diğer ölçümlerden üç kat daha fazla ölçüm yapmışlardır. Her hareket için kaydedilen veriler, kolaylık sağlamak için $[0,1]$ 'e göre yeniden ölçeklendirilmiştir (Ramsay ve Silverman, 2005). Bu problemin analizinde Fonksiyonel veriler için En Küçük Kareler (EKK) Yöntemi kullanılmıştır.

Son yıllar, fonksiyonel hipotez test yöntemlerinin geliştirilmesinde başarılı bir dönem olmuştur. Test istatistiği açısından bu test istatistikleri; L_2 - norm tabanlı testler, F -tipi testler ve Bootstrap yöntemleridir. Regresyon modelleri açısından bu test yöntemleri; tek örnek problemleri, iki örnek problemleri ve tek yönlü ANOVA'dır.

Tek değişkene sahip fonksiyonel veriler için yaygın olarak incelenen problemler; varyans analizi (ANOVA), kanonik korelasyon analizi, kümeleme analizi, temel bileşen analizi ve regresyon analizidir. Çok değişkene sahip fonksiyonel veriler için yaygın olarak incelenen problemler ise varyans analizi (MANOVA), kanonik korelasyon analizi, kümeleme analizi, doğrusal regresyon analizi ve temel bileşenler analizidir.

Kümeleme analizi, çalışılan bir olguyu anlamak ve yorumlamak amacıyla verilerin gruplandırılmasını gerektiren, birçok bilimsel disiplin tarafından kullanılan en önemli istatistiksel araçlardan biridir. Düşük ve yüksek boyutlu çok değişkenli fonksiyonel verilerin analizi için yararlı bir keşif aracıdır. Kümeleme analizi, özellikle yüksek boyutlu veriler için kullanılmaktadır. Örüntü tanıma, yapay zeka, bilgi teknolojisi, görüntü işleme, biyoloji, psikoloji, pazar araştırması, astronomi, psikiyatri, arkeoloji, biyoinformatik ve genetik gibi birçok alana dair kümeleme yöntemi, tekniği ve algoritması mevcuttur (Gan vd, 2007; Everitt vd, 2011).

Literatürde birçok fonksiyonel veri kümeleme yöntemi mevcuttur. Jacques ve Preda (2014a; 2014b), farklı fonksiyonel veri kümeleme yaklaşımlarına dair bir sınıflandırma sunmaktadır. Fonksiyonel veri kümeleme yaklaşımları dört gruba

ayrılmaktadır. Bu yaklaşımlar ham veri yöntemleri, filtreleme yöntemleri, uyarlanabilir yöntemler ve uzaklığa dayalı yöntemlerdir. Ham veri yöntemlerine örnek, Boullé (2012) tarafından gerçekleştirilen çalışmadır. Peng ve Müller (2008), Abraham vd (2003), Rossi vd (2004) ve Kayano vd (2010) FTBA katsayılarına dayanan filtreleme yöntemleri geliştirmiştir. James ve Sugar (2003), Heard vd (2006), Ray ve Mallick (2006), Samé vd (2011) ve Giacofci vd (2013) baz fonksiyon katsayılarının olasılıksal modeline dayanan farklı uyarlamalı yöntemler ortaya koymuştur. Chiou ve Li (2007), Delaige ve Hall (2010), Bouveyron ve Jacques (2011) ve Jacques ve Preda (2013), FTBA puanlarının olasılıksal modeline dayanan başka uyarlanabilir bir yöntem geliştirmiştir. Tarpey ve Kinatader (2003), Ferraty ve Vieu (2006), Tokushige vd (2007) ve Ieva vd (2013) uzaklığa dayalı farklı yöntemler geliştirmiştir.

Çok değişkenli veriler için varyans analizi, gruplar arasında ortalamalar bakımından anlamlı bir farklılık olup olmadığını araştırmak için kullanılmaktadır. Fonksiyonel veriler için ise varyans analizi oluşturulan grupların ortalama fonksiyonları bakımından anlamlı bir farklılık olup olmadığını araştırmak için kullanılmaktadır.

FVA ile ilgili daha ayrıntılı bilgi için Cravan ve Wahba (1979), Ramsay ve Dalzell (1991), Green ve Silverman (1994), Gasser ve Kenip (1995), Silverman (1995; 1996), Ramsay (2004; 2005), Ulbricht (2004), Ramsay ve Silverman (2002a; 2005), Costanzo (2005), Horvath vd (2009) ve Zhang (2013a; 2013b; 2013c) çalışmaları incelenebilir.

Son yıllarda fonksiyonel veri olarak tanımlanan periyodik yapıdaki verilerin değerlendirilmesinde özellikle tek değişkenli durumda kümeleme analiz yöntemleri yaygın olarak kullanılmaktadır. Bu tez çalışmasının amacı, çok değişkenli fonksiyonel verilerin, fonksiyonel kümeleme yöntemleriyle incelenmesidir. Türkiye’de bulunan 81 ile ait 5 iklim değişkeni dikkate alınarak, illerin mevsimsel değişkenlere göre benzerlikleri yönünden kümelene yapısı araştırılmıştır. Ayrıca kümelene sonucunda elde edilen fonksiyonların ortalamaları bakımından aralarında fark olup olmadığı belirlenmiştir.

Bu tez çalışması altı bölümden oluşmaktadır. Bölüm 2’de verilerin fonksiyonel hale getirilip modellenmesi için gerekli olan yöntemler hakkında bilgi verilmiştir.

Bölüm 3'te fonksiyonel verilerin kümelenmesi için gerekli olan yöntemler tanıtılmıştır. Bölüm 4'de oluşturulan fonksiyonların ortalamalar bakımından aralarında fark olup olmadığını araştırmada kullanılan, test teknikleri sunulmuştur. Bölüm 5'de gerçek veri seti üzerine bir uygulama yapılmış ve elde edilen bulgular değerlendirilmiştir. Bölüm 6'da ise sonuç ve önerilere yer verilmiştir.



2. FONKSİYONEL VERİLERİN MODELLENMESİ

Fonksiyonel veriler ayrık noktalarda gözlemlenir ve $x(t)$ fonksiyonunun fonksiyonel bir gözlemi n sayıda (y_j, t_j) çiftlerinden oluşmaktadır. Burada y_j, t_j zamanında $x(t)$ fonksiyonunun bir gözlemidir. Fonksiyonel model ise;

$$y_j = x(t_j) + \varepsilon_j, \quad j = 1, 2, \dots, n \quad (2.1)$$

şeklinde ifade edilmektedir. Eşitlik (2.1) ile verilen modelde, ε_j 'lerin dağılımı 0 ortalama ve σ^2 sabit varyansa sahip olup birbirinden bağımsızdır. Ayrıca t aralığı $T = [t_1, t_n]$ aralığında sınırlandırılmakta ve $x(t)$ fonksiyonu T üzerinde sürekliliği sağlamaktadır.

2.1. En Küçük Kareler Yöntemi

$x(t)$ fonksiyonunu tahmin etmenin yollarından bir tanesi, fonksiyonu K sayıda bilinen $(\Phi_1, \Phi_2, \dots, \Phi_K)$ baz fonksiyonların doğrusal bir kombinasyonu aracılığı ile

$$x(t) = \sum_{i=1}^K c_i \Phi_i(t) \quad (2.2)$$

şeklinde ifade etmektir. Burada (c_1, c_2, \dots, c_K) katsayıları Hata Kareler Toplamı (HKT) minimize edilerek belirlenmektedir. Eşitlik (2.3) ile verilen HKT;

$$HKT = \sum_{j=1}^n [y_j - \sum_{i=1}^K c_i \Phi_i(t_j)]^2 \quad (2.3)$$

olup, burada c ve y vektörleri ile Φ matrisi sırasıyla;

$$c = (c_1, c_2, \dots, c_K)^T \quad (2.4)$$

$$y = (y_1, y_2, \dots, y_n)^T \quad (2.5)$$

$$\Phi = \begin{bmatrix} \Phi_1(t_1) & \cdots & \Phi_K(t_1) \\ \vdots & \ddots & \vdots \\ \Phi_1(t_n) & \cdots & \Phi_K(t_n) \end{bmatrix}_{n \times K} \quad (2.6)$$

eşitlikleri ile ifade edilmektedir.

O halde HKT;

$$HKT = (y - \Phi c)^T (y - \Phi c) \quad (2.7)$$

şeklinde ifade edilebilir.

c katsayı vektörünün en küçük kareler tahmini, eşitlik (2.7)'nin c 'ye göre türevinin alınması ile elde edilmektedir. Bu denklem sıfıra eşitlenerek çözüldüğünde c parametre vektörünün tahmin edicisi;

$$\hat{c} = (\Phi^T \Phi)^{-1} \Phi^T y \quad (2.8)$$

olarak elde edilmektedir.

Burada, $x(t)$ fonksiyonunun tahmin edicisi, c vektörünün tahmini değerlerinin kullanılmasıyla elde edilir.

2.2. Baz Fonksiyon Yaklaşımı

FVA'da $x_i(t)$ fonksiyonunu oluşturmak için yöntemlerden bir tanesi K adet baz fonksiyonundan oluşan bir sistem seçilip $x_i(t)$ fonksiyonunu bu baz fonksiyonlarının ağırlıklandırılmış toplamı olarak ifade etmektir (Keser 2008; 2010). Bu ifade eşitlik (2.2) ile benzer olup,

$$x_i(t) = c_1 \Phi_1(t) + c_2 \Phi_2(t) + \dots + c_K \Phi_K(t) \quad (2.9)$$

şeklindedir.

Bu ifade de yer alan, $\Phi_i(t)$, i .inci baz fonksiyon ve c_i , i .inci baz fonksiyonuna karşılık gelen katsayı olarak ifade edilmektedir.

En yaygın kullanılan baz fonksiyon yaklaşımları Fourier, B-Splayn, Dalgacıklar ve Polinomlar'dır.

2.2.1. Fourier baz fonksiyon yaklaşımı

Veriler $T = [t_1, t_n]$ aralığında olmak üzere Fourier baz fonksiyonları;

$$\Phi_0(t) = \frac{1}{\sqrt{T}} \quad (2.10)$$

$$\Phi_{2r-1}(t) = \sqrt{\frac{2}{T}} \sin(rwt) \quad (2.11)$$

$$\Phi_{2r}(t) = \sqrt{\frac{2}{T}} \cos(rwt) \quad (2.12)$$

şeklinde ifade edilmektedir. Burada, ω sabiti, $\omega = 2\pi/T$ şeklinde tanımlıdır ve $T = t_n - t_1$ aralığının uzunluğuyla ilgilidir ve $r = 1, 2, \dots$ 'dir

Fourier serisi periyodik veriler için kullanışlıdır. K sayıda belirlenen baz fonksiyonu, yaklaşımın düzgünlüğünü kontrol etmektedir. Düşük bir K değeri, tahmin edilen fonksiyona ait olduğu bilinen önemli özelliklerin bulunmasını zorlaştırırken, yüksek bir K değeri verilere daha iyi uyum sağlayan bir fonksiyon sağlamaktadır. Ancak çok sayıda baz fonksiyon kullanılırsa, tahmin edilen fonksiyon gözlem noktaları arasında yüksek frekanslı salınımlar gösterebilir. Bu nedenle uygun bir baz fonksiyonunun belirlenmesi, verilerin belirgin özelliklerinin ortaya çıkarılmasında oldukça önemlidir (Ulbricht, 2004).

2.2.2. Polinom baz fonksiyon yaklaşımı

İlk K çokterimli (polynomial) baz fonksiyonu, aşağıdaki şekilde tek terimli (monomial) bazdan oluşur ve aşağıdaki şekilde gösterilir.

$$\Phi_k(t) = (t - \omega)^k, \quad k = 0, \dots, K - 1 \quad (2.13)$$

Burada, ω bir kaydırma parametresidir. Çok terimli polinom baz fonksiyonları, büyük bir K kullanmadan verilerin belirgin özelliklerini bulmakta zorlanmaktadır. Bu da zayıf türev tahminlerine neden olmaktadır. Polinomlar, aralığın ortasında, verilere çok iyi şekilde uyma eğilimindeyken, uç noktalara yakın yerlerde verilere uyumu tatmin edici değildir (Ramsay ve Silverman, 2002b)

2.2.3. B-splayn baz fonksiyon yaklaşımı

Bir Splayn, τ_i , $i = 1, 2, \dots, m$ değerlerinde polinomların düzgün bir şekilde birleştirilmesiyle tanımlanmış ve sabit bir k derecesi ile oluşturulmuş çok terimli bir fonksiyondur.

τ_i değerleri düğüm olarak adlandırılmakta ve aşağıdaki denklemi sağlamaktadır.

$$t_1 = \tau_1 < \tau_2 < \dots < \tau_m = t_n \quad (2.14)$$

Dahası, iç düğümlerde, iki bitişik polinomun, $k - 1$ olarak seçilen türevlerinin belirli bir seviyeye kadar olan değerlerinin eşleşmesi gereklidir. k dereceli splayn, güç bazlı fonksiyonların ;

$$\Phi_i(t) = \max(0, (t - \tau_i)^k), \quad i = 1, 2, \dots, m - 1 \quad (2.15)$$

doğrusal bir kombinasyonudur.

Periyodik olmayan verileri tahmin etmek için B-splayn baz fonksiyonları kullanılmaktadır. Bunlar kompakt desteklere sahiptir, her güç baz fonksiyonu sonlu bir $[\tau_p, \tau_q]$ aralığı dışında her yerde sıfırdır ve burada $0 < p < q < m$ 'dir. B-splayn baz fonksiyonu;

$$B_p(t) = \sum_{j=p}^{p+k+1} \left[\prod_{i=p, i \neq j}^{p+k+1} \frac{1}{(\tau_i - \tau_j)} \right] \cdot \max(0, (t - \tau_j)^k), \quad -\infty < t < \infty \quad (2.16)$$

şeklinde ifade edilmektedir.

$t, (\tau_p, \tau_{p+k+1})$ aralığında olduğu zaman, p alt endeksi, $B_p(t)$ 'nin sıfır olmadığını göstermektedir. B-splaynların baz fonksiyonlar olarak faydaları ise polinomların kolay hesaplanabilmesidir. İlâveten, B-splaynların kompakt desteği, splaynlara göre doğruluk kaybını önlerken bu da tüm katsayılar için hesaplama süresinin azaldığı manasına gelmektedir. Düğümün yerleştirilmesinde kullanılan yaklaşımlardan bir tanesi, yüksek eğriliğe sahip bölgelere çok sayıda düğüm yerleştirmek ve ardından makul bir pürüzsüzlük elde edilinceye kadar bunları birer birer azaltmaktır.

2.2.4. Dalgacık baz fonksiyon yaklaşımı

Dalgacık, ortalama değeri sıfır olan ve genellikle Fourier serileriyle karşılaştırılan sınırlı süreli bir dalga olarak görülebilir. Aradaki fark, dalgacıkların hem frekansta hem de zaman içinde, Fourier serilerinin ise sadece zaman içinde sınırlandırılmış olmasıdır (Fugal, 2009).

Dalgacık baz fonksiyonları, uygun bir ψ ana dalgacık fonksiyonunun genişlemeleri dikkate alınarak,

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k), \quad j, k \in Z \quad (2.17)$$

şeklinde oluşturulmaktadır.

Dalgacıklar yakın zamanda geliştirilmiş olmasına rağmen, kullanım açısından büyük potansiyele sahiptir.

2.3. Düzleştirme Yöntemleri

Düzleştirme prosedürlerinin altında yatan ana fikir verilere uygun bir eğri tahmin etmektir.

2.3.1. Kernel düzleştirme

Birbirine yakın zaman değerlerinin fonksiyon değerleri genellikle birbirine benzer olduğundan bir kernel düzleştirici, genellikle pürüzsüz bir şekilde ağırlıkları hedef noktadan uzaklaştıkça tanımlamaktadır. $\{S_j(t)\}_{j=1}^n$ bir ağırlık dizisi olmak üzere, belirli bir t için tahminler, y_j gözlemlerinin doğrusal bir birleşimi olarak,

$$\hat{x}(t) = \sum_{j=1}^n S_j(t)y_j \quad (2.18)$$

şeklinde ifade edilir.

$S_j(t)$ ağırlıkları ya da Nadaraya-Watson Kernel tahmincisi, Kernel (K) adı verilen yoğunluk fonksiyonlarıyla ve h ölçek parametresiyle oluşturulmaktadır. h parametresi, sabit olabileceği gibi değişken de olabilir ve

$$S_j(t) = \frac{K\left(\frac{t_j - t}{h}\right)}{\sum_{j=1}^N K\left(\frac{t_j - t}{h}\right)} \quad (2.19)$$

şeklinde hesaplanmaktadır (Genton, 2001).

En sık kullanılan kernel düzleştiriciler aşağıdaki gibidir.

$$\begin{aligned} \text{Uniform: } K(u) &= \begin{cases} 0,5, & |u| \leq 1 \\ 0, & \text{dd} \end{cases} \\ \text{Kuadratik: } K(u) &= \begin{cases} 0,75, & |1 - u^2| \leq 1 \\ 0, & \text{dd} \end{cases} \\ \text{Gauss: } K(u) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \end{aligned} \quad (2.20)$$

Kernel düzleştiriciler kullanıldığı takdirde $x(t)$ 'ye ait türevlerin tahmini iyi bir fikir değildir. Bunun nedeni, çekirdekleri ayırt etmenin her durumda mümkün olmamasıdır. Bununla birlikte kernel fonksiyonları, fonksiyonları ve türevleri tahmin etmek için kullanılabilir. Bu yöntemle bir baz fonksiyon tahmin edicisi, EKK kriteri genişletilerek elde edilebilir (Chen vd, 2009). Bu genişletimde HKT ,

$$HKT = \sum_{j=1}^n w_j(t) \left[y_j - \sum_{k=1}^K c_k \Phi_k(t_j) \right]^2 \quad (2.21)$$

olup, burada $w_j(t)$ ağırlık fonksiyonları, kernel fonksiyonları tarafından

$$w_j(t) = K \left(\frac{t_j - t}{h} \right) \quad (2.22)$$

şeklinde tanımlanmaktadır. Matris gösterimi ise,

$$HKT = (y - \Phi c)^T W(t) (y - \Phi c) \quad (2.23)$$

şeklinde ifade edilmektedir. Burada, $W(t)$ bir köşegen matris olup,

$$W(t) = \begin{bmatrix} w_1(t) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_n(t) \end{bmatrix} \quad (2.24)$$

şeklinde tanımlıdır. Bu durumda c katsayısı vektörünün en küçük kareler tahmini;

$$\hat{c} = (\Phi^T W(t) \Phi)^{-1} \Phi^T W(t) y \quad (2.25)$$

şeklinde ifade edilmektedir.

2.3.2. Pürüzlülük cezası

Ayrık verileri bir fonksiyonla temsil etmenin başka bir yolu da pürüzlülük cezası yaklaşımıdır. Pürüzlülük cezası yaklaşımı, verinin aşırı uyumunu önlemekte ve çok sayıda baz fonksiyon kullanılarak düzgünleştirme miktarının kontrol edilmesine olanak sağlamaktadır. Bir fonksiyonun pürüzlülüğü, o fonksiyonun ikinci mertebeden türevinin karesinin zaman değişkenine göre integrasyonu olup,

$$PEN_2(x) = \int [x''(t)]^2 dt \quad (2.26)$$

eşitliği ile ifade edilmektedir (Kadir ve Keser, 2007).

Bu ifade toplam eğrilik fonksiyonu olarak da bilinmektedir. Başka bir ifadeyle $x(t)$ 'nin düz bir çizgiden ayrıldığı miktar olarak görülmektedir. Bu nedenle, değişken fonksiyonların $PEN_2(x)$ değerlerinin yüksek olması beklenebilir. Tahmini ise eğriliği en aza indiren $x(t)$ fonksiyonunun bulunmasıyla elde edilir. Söz konusu fonksiyonun bulunabilmesi için,

$$SSE_{PEN} = \sum_{j=1}^n [y_j - x(t_j)]^2 + \lambda PEN_2(x) \quad (2.27)$$

eşitliğinin en küçük yapılması gerekmektedir. Eşitlik (2.27)'deki λ düzgünleştirme parametresi, toplam eğrilik üzerindeki ceza miktarını belirler. $\lambda \rightarrow 0$ ise, tahmin, (y_j, t_j) gözlemlerine mümkün olduğunca yakın olacaktır. $\lambda \rightarrow \infty$ ise, tahmin, verilere en iyi uyan düz bir çizgi olacaktır.

İkinci türevi baz alan pürüzlülük cezası en yaygın kullanılan yöntem olmasına rağmen, yöntem cezalandırmak için diğer türevleri de kullanabilmektedir.

2.4. Düzgünleştirme Parametresinin Belirlenmesi

λ düzgünleştirme parametresinin belirlenmesinde farklı yaklaşımlar mevcuttur. İlk yaklaşımda, en iyi sonucu ve görünümü veren düzgünleştirme parametresi doğrudan seçilebilir (Green ve Silverman, 1994).

Yapılan analizlerde λ için 10^{-2} , 10^{-3} ve 10^{-4} değerlerinin iyi sonuçlar verdiği görülmüştür (Ramsay ve Li, 1998).

λ , gözlem sayısının baz fonksiyon sayısından küçük olduğu durumlarda;

$$\lambda = 10^{-4} \left[\frac{iz(B^T B)}{izR} \right] \quad (2.28)$$

şeklinde ifade edilmektedir (Ulbricht, 2004).

İkinci yaklaşım, düzgünleştirme parametresinin belirlenmesinde genelleştirilmiş çapraz geçerlilik yönteminin (GCV) kullanılmasıdır (Craven ve Wahba, 1979). GCV;

$$GCV = \frac{n^{-1}HKT}{[n^{-1}iz(I-S_{\phi,\lambda})]^2} \quad (2.29)$$

şeklinde ifade edilmektedir.

Burada $S_{\phi,\lambda}$ düzgünleştirme matrisi olmak üzere;

$$S_{\phi,\lambda} = \Phi(\Phi^T \Phi)^{-1} \Phi \quad (2.30)$$

şeklinde tanımlanmaktadır. Bu matrisin izi için $sd(\lambda) = iz(S_{\phi,\lambda})$ olmak üzere GCV,

$$GCV_{\lambda} = \left(\frac{n}{n - sd(\lambda)} \right) \left(\frac{HKT}{n - sd(\lambda)} \right) \quad (2.31)$$

şeklinde de ifade edilmektedir. Amaç λ 'ya göre GCV'nin minimizasyonudur (Ramsay ve Silverman, 2005).

R , pürüzlü ceza matrisi olmak üzere hesaplamalar yapıldıktan sonra CKT_{λ} 'yı minimize eden pürüzlü ceza tahminleri, baz fonksiyon yaklaşımına göre

$$\hat{c} = (\Phi^T \Phi + \lambda R)^{-1} \Phi^T y \quad (2.32)$$

şeklinde tanımlanmaktadır.



3. FONKSİYONEL VERİLERİN KÜMELENMESİNDE KULLANILAN YÖNTEMLER

Fonksiyonel verileri kümelemek, verilerin ait olduğu sonsuz boyutlu uzay nedeniyle genellikle zor bir süreçtir. Fonksiyonel verileri kümelemek için birçok farklı yaklaşım önerilmiştir. En popüler yaklaşım ise, elemanlar sonlu boyutlu uzaya yaklaştırılarak, sonsuz boyutlu bir problem sonlu boyutlu bir probleme indirgenir. O zaman, sonlu boyutlu veriler için kümeleme algoritmaları uygulanabilir. Öte yandan, kümeleme için parametrik olmayan yöntemler genellikle fonksiyonel veriler için belirli mesafeleri veya farklılıkları tanımlamaya ve daha sonra kümeleme algoritmalarını uygulamaya dayanmaktadır. Son zamanlarda, fonksiyonel veriler için model tabanlı algoritmalar geliştirilmiştir (Barigally, 2016).

Fonksiyonel verilerdeki model tabanlı kümeleme yöntemleri, ilk önce sonsuz boyutlu problemi sonlu boyuta dönüştürmeye ve ardından sonlu boyutlu veriler için model tabanlı bir kümeleme yöntemini kullanmaya dayanmaktadır.

3.1. Ham Veri Yöntemleri

Ham veri yöntemleri, fonksiyonel veri kümeleme yöntemleri arasında en basit yöntemdir. Çünkü eğrileri, doğrudan değerlendirme noktalarına dayalı ya da temel veriye dayalı olarak kümelendirmektedir (Jacques ve Preda, 2014a). Bu metotlar verilerin fonksiyonel formunun yeniden yapılandırılmasına, baz fonksiyon katsayılarına veya FTBA gibi filtreleme tekniklerine ihtiyaç duymamaktadır.

Bu yöntemlerin değerlendirme puanları ile ilgili dezavantajlarından biri de, verilerin süreklilik ve türev gibi fonksiyonel özelliklerini dikkate almamalarıdır. Ayrıca, eğriler farklı değerlendirme noktalarında gözlemlendiğinde, bu yöntem onu da göz ardı etmektedir (Boullé, 2012).

Yüksek boyutlu veri kümeleme (HDDC) yöntemi Bouveyron vd (2007) tarafından geliştirilmiştir. Bu yöntem, veriler orijinal alanın boyutundan daha düşük bir boyuta sahip alt alanlarda olduğunda, Gaussian karışım modeline dayanan model tabanlı bir kümeleme yöntemidir. Diğer model tabanlı kümeleme yöntemleri gibi,

modelin parametrelerini tahmin etmek için EM (Expectation-Maximization) algoritması kullanılmaktadır.

3.2. Filtreleme Yöntemleri

Filtreleme yöntemlerinde, eğriler bilinen sınırlı bir fonksiyona yaklaştırılmaktadır ve daha sonra baz fonksiyon yaklaşımına göre hesaplanan katsayılar veya FTBA'dan elde edilen skorlar kullanılarak kümeleme yaklaşımı gerçekleştirilmektedir. Filtreleme olarak bilinen ilk adımda (James ve Sugar, 2003), eğriler sınırlı bir fonksiyona yaklaştırılarak verinin boyutu azaltılmaktadır. Kümeleme adımı olarak bilinen ikinci adımda ise ilk adımda elde edilen sonlu boyutlu veriler için kümeleme yöntemleri kullanılmaktadır.

Filtreleme adımında genellikle iki farklı yöntem kullanılmaktadır. Birincisi, eğrilerin fonksiyonel yapısını, temel bileşen puanlarının sınırlı bir sayısına yaklaştırmak suretiyle yeniden yapılandırmasını gerektiren FTBA'dır. İkinci yöntem, verinin özelliklerine uygun bir baz fonksiyonunun belirlenmesidir (Wahba, 1990).

3.2.1. FTBA puanlarında filtreleme yöntemi

Fonksiyonel verilerin analizinde, modellenmesinde ve sınıflandırılmasında kullanılabilir temel araçlardan biri olan FTBA'nın kullanımı son yıllarda artmıştır. FTBA, çok değişkenli temel bileşenler analizinden (TBA) farklıdır. Çünkü verilerin ana değişkenlik modlarını sergileyen öz fonksiyonlar da birer fonksiyondur ve zaman içinde değişen, değişkenlik modları olarak yorumlanabilirler (Ramsay ve Silverman, 2002b).

FTBA'nın amacı, fonksiyonel verileri eğrilerdeki ana değişkenlik özelliklerini yakalayan küçük boyutlu bir alanda temsil etmektir. H , T üzerindeki gerçek fonksiyonların Hilbert uzayı olmak üzere her $t \in T$ için, f' 'yi $f(t)$ ile ilişkilendiren, $L_t f \rightarrow f(t)$, sınırlı bir lineer fonksiyonelliktedir (Ingrassia ve Costanzo, 2005). O halde fonksiyonel temel bileşenin hedefi, varyans fonksiyonunun ortogonal bir şekilde ayrıştırılmasıdır. Fonksiyonel verilerde varyans fonksiyonu,

$$v(t, u) = \frac{1}{n-1} \sum_{i=1}^n \{x_i(t) - \bar{x}(t)\} \{x_i(u) - \bar{x}(u)\} \quad (3.1)$$

eşitliği ile verilir.

Fonksiyonel veriler için varyans fonksiyonunun ayrışması ise;

$$v(t, u) = \sum_j \lambda_j \xi_j(t) \xi_j(u) \quad (3.2)$$

şeklinde ifade edilmektedir. Burada, ξ_j temel bileşen ağırlık fonksiyonları olarak adlandırılmaktadır.

$\lambda_j = \int_T \xi_j(t) v(t, u) \xi_j(u) dt du$ olup, pozitif ve azalmayandır. Diğer taraftan, öz fonksiyonların ise $\int_T \xi_j^2(t) dt = 1$ ve $\int_T \xi_j \xi_i(t) dt = 0$ ($i < j$) kısıtlarını yerine getirmesi gerekmektedir. O zaman veri kümesindeki birimlerin temel bileşen skorları S_i değerleri olup,

$$S_i^{(j)} = \langle x_i, \xi_j(t) \rangle = \int_T \xi_j(t) x_i(t) dt \quad (3.3)$$

eşitliği ile verilir.

Eşitlik (3.2) ile verilen varyans fonksiyonu ayrışması sayesinde, v kovaryans fonksiyonuna EKK yaklaşımı uygulanabilir. Böylece, ξ öz fonksiyonları, x_i örnek fonksiyonları arasında değişimin temel bileşenlerini tanımlamaktadır. $\{C_j\}_{j \geq 1}$ temel bileşenleri elde etmek için ortalama $\mu(t)$ fonksiyonu kullanılmaktadır (Mardia vd, 1989). $j \geq 1$ olmak üzere C_j temel bileşeni,

$$C_j = \int_0^T (X(t) - \mu(t)) \xi_j(t) dt \quad (3.4)$$

şeklinde tanımlanır. Burada, $\mu = \{\mu(t) = E(X)\}_{t \in T}$, temel bileşenler $\{C_j\}_{j \geq 1}$ ise λ_j varyanslı ve 0 ortalamalı ilişkisiz rastgele değişkenlerdir. $X(t)$ ise L_2 -normlu sürekli bir stokastik süreç olmak üzere $\forall t \in T$ için,

$$\lim_{h \rightarrow 0} E[|X(t+h) - X(t)|^2] = 0 \quad (3.5)$$

şeklinde ifade edilmektedir.

O zaman, Karhunen-Loeve genişlemesi,

$$X(t) = \mu(t) + \sum_{j \geq 1} C_j \xi_j(t), t \in T \quad (3.6)$$

eşitliği ile verilmektedir (Karhunen, 1947; Loève, 1945).

$X(t)$ için L_2 -normundaki en iyi yaklaşım, eşitlik (3.6)'nın ilk q terimi kırılarak elde edilen ve $X^q(t)$ ile gösterilen yaklaşımdır (Saporta, 1981). Böylece $X^q(t)$ 'yi,

$$X^q(t) = \mu(t) + \sum_{j=1}^q C_j \xi_j(t), t \in T \quad (3.7)$$

şeklinde ifade etmek mümkündür.

3.2.2. Baz fonksiyon katsayılarını filtreleme yöntemi

FVA'nın ayırt edici bir özelliği de gözlemlerin sonsuz boyutlu bir alana ait olduğu varsayımdır. Bununla birlikte, her bir sonlu $\{t_{ij}: j = 1, 2, \dots, m_i\}$ düğümünde $X_i(t)$ için ayrık X_{ij} gözlemleri vardır. Bu durum ise verilerin fonksiyonel formunun X_{ij} ayrık gözlemleriyle yeniden oluşturulması gerektiği anlamına gelmektedir (Baragilly, 2016).

3.3. Uyarlanabilir Yöntemler

Uyarlanabilir yöntemler, kümelere bağlı olarak verilerin fonksiyonel bir sunumunu ifade etmektedir. Bu yöntemler boyut küçültme ve kümelemeyi aynı anda gerçekleştirmektedir. Uyarlanabilir yöntemlerin arkasındaki ana fikir, baz fonksiyon katsayılarının ve FTBA skorlarının, filtreleme yöntemlerinde olduğu gibi parametreler yerine rastgele değişkenler olarak kabul edilmesidir. Ayrıca uyarlanabilir yöntemlerde, bu rastgele değişkenlerin kümeye özgü bir olasılık dağılımına sahip olduğu varsayılmaktadır (Jacques ve Preda, 2014a). Bu yöntemde, baz fonksiyon katsayılarının olasılıksal modeli bulunmaktadır ve ardından aynı anda boyut küçültme ve kümeleme gerçekleştirilmektedir.

James ve Sugar (2003), eğrilerin baz fonksiyon katsayılarına kümelenmeye özgü bir olasılık dağılımı vererek, seyrek örneklenen fonksiyonel veriler için kümeleme yaklaşımı geliştirmiştir. Her kümeye ve genel varyansa göre farklı araçlarla Gaussian dağılımının bir karışımı kullanılmaktadır ve bu yöntemde $c_i \sim N(\mu_k, \Sigma)$ 'dir. Baz fonksiyon katsayılarının rastgele değişkenler olduğunu varsayıp Gaussian dağılımındaki parametrelere dayanarak, temel kümeleme modeli oluşturmak mümkündür. Bu yöntem fcust yöntemi olarak bilinmektedir.

Ray ve Mallick (2006), fonksiyonel eğrilerin baz fonksiyon katsayıları için hiyerarşik bir Bayesian modeli önermiştir. FTBA'nın olasılık modeline dayalı uyarlanmış yöntemler aynı anda boyut küçültmeyi ve kümelemeyi gerçekleştirir. Örneğin, Chiou ve Li (2007), Karhunen-Loeve açılımlarının kırılmaları arasındaki L_2 uzaklığına bağlı olarak K -ortalamalar algoritmasını önermiştir. Delaigle ve Hall (2010), fonksiyonel rastgele değişkenler için olasılık yoğunluğunun yaklaşık olarak tahmin edilmesinden sonra FTBA skorlarının olasılıklı bir modelini geliştirmiştir.

Bouveyron ve Jacques (2011) tarafından geliştirilen FunHDDC yöntemi, Bouveyron vd (2007) tarafından daha önce geliştirilen bir HDDC uzantısıdır. HDDC gibi, FunHDDC de Gaussian karışım modeline dayanan model tabanlı bir kümeleme yöntemidir. FunHDDC, FTBA yaklaşımında kullanılan baz fonksiyonlarını kullanmaktadır ve varyansa ilişkin varsayımları temel almaktadır. Baz fonksiyonlarının her biri, belirli bir grup içinde modellenerek fonksiyonel verileri kümelemektedir.

Jacques ve Preda (2013), temel bileşenlerin Gaussian dağılımına dayanan benzer bir algoritma geliştirmiş ve olasılıksal bir model tabanlı kümeleme tekniği tanımlamıştır. Diğer model tabanlı kümeleme yöntemleri gibifunclust algoritması, modelin parametrelerini tahmin etmek için EM algoritmasını kullanmaktadır.

3.4. Uzaklığa Dayalı Yöntemler

Uzaklığa dayalı yöntemler, fonksiyonel veriler için bazı farklılık ölçülerine ve uzaklık fonksiyonlarına dayalı kümeleme algoritmalarını kullanan yöntemlerdir. Fonksiyonel verilerin kümelenmesi için farklı uzaklıklara dayalı yöntemler mevcuttur. Jacques ve Preda (2014a)'ya göre, $X_i(t)$ ve $X_j(t)$ eğrileri arasındaki uzaklık ölçümlerinin genel formu;

$$d_l(X_i(t), X_j(t)) = \int_T (X_i^{(l)}(t) - X_j^{(l)}(t))^2 dt)^{\frac{1}{2}} \quad (3.8)$$

şeklinde ifade edilmektedir.

Burada $X_i^{(l)}(t)$, $X_i(t)$ 'nin l -inci türevidir. Buna göre d_0 ve d_1 uzaklıkları sırası ile;

$$d_0(X_i(t), X_j(t)) = \int_T (X_i(t) - X_j(t))^2 dt)^{\frac{1}{2}} \quad (3.9)$$

$$d_1(X_i(t), X_j(t)) = \int_T (X_i^{(1)}(t) - X_j^{(1)}(t))^2 dt)^{\frac{1}{2}} \quad (3.10)$$

eşitlikleri ile verilir.

3.5. Model Tabanlı Kümeleme Yöntemleri

Model tabanlı kümeleme yönteminin amacı, bir karışım yoğunluk modelinden örneklenen homojen veri gruplarını tanımlamaktır. $x_i = \{x_i(t)\}_{t \in T}$, $1 \leq i \leq n$, i -nci eğri fonksiyonunu göstermek üzere $\{x_1, x_2, \dots, x_n\}$ eğriler kümesini göz önüne alalım. Bu eğrileri K tane homojen gruba kümelemek isteyelim. X 'in grup üyeliğini gösteren bir gözlemlenmemiş $Z = (Z_1, Z_2, \dots, Z_K) \in \{0,1\}^K$ rastgele vektörünün var olduğunu kabul edelim. Burada, X 'in grup üyeliği; eğer X , K -ıncı gruba aitse $Z_K = 1$, aksi takdirde $Z_K = 0$ 'dır. Bu nedenle kümeleme görevi, gözlemlenen her bir x_i eğrisi için Z 'nin $z_i = (z_{i1}, z_{i2}, \dots, z_{iK})$ değerini tahmin etmeyi amaçlamaktadır (Jacques ve Preda, 2014b).

Her bir x_i eğrisi, X 'in γ_i katsayı vektörü ile bir baz genişlemesini ifade eden bir örnek yörüngesi olduğu varsayalım. Burada γ_i katsayı vektörünün dağılımı;

$$p(\gamma) = \sum_{k=1}^K \pi_k \phi(\gamma; \mu_k, \Sigma_k) \quad (3.11)$$

yoğunluğu ile Gaussianların bir karışımıdır. Eşitlik (3.11)'de ϕ , $\mu_k = U_k m_k$ ortalamalı, $\Sigma_k = Q_k \Delta_k Q_k^T$ kovaryans matrisli standart Gaussian yoğunluk fonksiyonu ve $\pi_k = P(Z_k = 1)$, k -ıncı gruba ait olma önsel olasılığıdır. Bu fonksiyonel karışım modeli Çizelge (3.1)'deki gibi özetlenebilir.

Çizelge 3. 1. Model Parametre Sayısı

Model	Parametre Sayısı
$a_{kj} b_k Q_k d_k$	$p + \tau + 2K + D$

Burada;

a_{kj} : Grupların parametre sayısı ($k = 1, 2, \dots, K; j = 1, 2, \dots, d_k$)

b_k : Grupların gürültüsü

Q_k : Σ_k matrisinin özvektörler matrisi

d_k : Grupların boyutları ($d_k \in \{1, 2, \dots, p - 1\}$)

şeklinde ifade edilmektedir.

$p = Kp + K - 1$, ortalamaların ve oranların tahmini için gereken parametrelerin sayısı iken $\tau = \sum_{k=1}^K d_k [p - (d_k + 1)/2]$, özvektörler matrislerinin (Q_k) ve $D = \sum_{k=1}^K d_k$ 'nın tahmini için gereken parametrelerin sayısıdır.

3.5.1. Karışım modeli

Bu model her bir (X_i, Z_i) çiftinin, (X, Z) rastgele vektörünün bağımsız bir gerçekleşişi olduğunu varsaymaktadır. Burada X , ait olduğu gruba bağlı olarak yaklaşık bir olasılık yoğunluk fonksiyonuna sahiptir (Baragilly, 2016). Bu olasılık yoğunluk fonksiyonu X , k -ıncı gruba aitken,

$$f_{x|z_k=1}^{(q_k)} = \prod_{j=1}^{q_k} f_{c_{j|z_k=1}}(c_{j,k}(x); \sigma_{j,k}^2) \quad (3.12)$$

eşitliği ile verilsin. Burada q_k , k grup için ilk temel bileşenlerin sayısıdır. $c_{j,k}(x)$ ise $X = x$ için, $x|z_k = 1$ 'in j -nci temel bileşen skorudur, $f_{c_{j|z_k=1}}$ ise olasılık yoğunluk fonksiyonudur. Σ_k ise köş $(\sigma_{1,k}^2, \sigma_{2,k}^2, \dots, \sigma_{q_k,k}^2)$ matrisinin j -nci köşegen elemanıdır.

Koşullu olarak, grup üzerinde, X 'in j -nci ana bileşeninin $f_{c_{j|z_k=1}}$ olasılık yoğunluğu, sıfır ortalamaya ve σ^2 varyansına sahip tek değişkenli Gaussian fonksiyonudur.

$Z = (Z_1, Z_2, \dots, Z_K)$ vektörünün dağılımı çok terimli dağılım olup;

$$z \sim M_1(\pi_1, \pi_2, \dots, \pi_K) \quad (3.13)$$

olarak gösterilmektedir. Burada $\pi_1, \pi_2, \dots, \pi_K$ olasılıkları $\sum_{k=1}^K \pi_k = 1$ şartını sağlar. Buradan, X 'in koşulsuz (yaklaşık) fonksiyonu;

$$f_X^{(q)}(x; \theta) = \sum_{k=1}^K \pi_k \prod_{j=1}^{q_k} f_{c_{j|z_k=1}}(c_{j,k}(x); \sigma_{j,k}^2) \quad (3.14)$$

şeklinde ifade edilmektedir.

Burada, $\theta = (\pi_k, \sigma_{1,k}^2, \sigma_{2,k}^2, \dots, \sigma_{q_k,k}^2)_{1 \leq k \leq K}$ tahmin edilmelidir ve $q = (q_1, q_2, \dots, q_K)$ 'dir. Sonlu boyutlu ortamlarda olduğu gibi, X eğrilerinin yaklaşık olasılık fonksiyonu;

$$l^{(q)}(\theta; \underline{X}) = \prod_{i=1}^n \sum_{j=1}^{q_k} \pi_k \frac{1}{\sqrt{2\pi}\sigma_{j,k}} \exp - \frac{1}{2} \left(\frac{C_{i,j,k}}{\sigma_{j,k}} \right)^2 \quad (3.15)$$

eşitliği ile verilir.

EM algoritması, yaklaşık olarak tanımlanmış eşitlik (3.15)'in log olasılığını en büyük yapmaya dayanmaktadır ve;

$$L_c^{(q)}(\theta; \underline{X}, \underline{Z}) = \sum_{i=1}^n \sum_{k=1}^K Z_{i,k} (\log \pi_k + \sum_{j=1}^{q_k} \log f_{C_{j|z_k=1}}(C_{i,j,k})) \quad (3.16)$$

şeklinde ifade edilmektedir.

E adımı: $Z_{i,k}$ grubu göstergeleri bilinmediğinde, E adımı, yaklaşık olarak tamamlanmış log olasılığının koşullu beklenen değerinin hesaplanmasına dayanmaktadır ve;

$$Q(\theta; \theta^h) = E_{\theta^h} [L_c^{(q)}(\theta; \underline{X}, \underline{Z}) | \underline{X} = \underline{x}] = \sum_{i=1}^n \sum_{k=1}^K t_{i,k} (\log \pi_k + \sum_{j=1}^{q_k} \log f_{C_{j|z_k=1}}(c_{i,j,k})) \quad (3.17)$$

olarak ifade edilir.

Burada $t_{i,k}$, $C_{i,j,k} = c_{i,j,k}$ koşulu altında X_i eğrisinin k grubuna ait olma olasılığı olup,

$$t_{i,k} = E_{\theta^h} [Z_{i,k} | \underline{X} = \underline{x}] \approx \frac{\pi_k \sum_{j=1}^{q_k} f_{C_{j|z_{i,k}=1}}(c_{i,j,k})}{\sum_{k=1}^K \pi_k \sum_{j=1}^{q_k} f_{C_{j|z_{i,k}=1}}(c_{i,j,k})} \quad (3.18)$$

şeklinde ifade edilmektedir.

Asıl puan güncelleme adımı: FTBA öz fonksiyonlarının ve verilen bir küme içindeki puanların hesaplanması, bazı yaklaşımlara ihtiyaç duymaktadır. En yaygın olan yaklaşım, eğrinin $\phi = (\phi_1, \phi_2, \dots, \phi_K)$ fonksiyonlarına bir genişleme getirmesidir. Γ , $n \times K$ boyutlu katsayılar matrisi ve $W = \int \phi \phi'$ baz fonksiyonlar arasındaki iç çarpımlar matrisidir. Burada, k grubundaki X_i eğrisinin $C_{i,j,k}$ temel bileşen puanlarının hesaplanması, önceki E adımında hesaplanan mevcut $t_{i,k}$ koşullu olasılığına bağlı olarak güncellenmektedir. Bu hesaplamanın, temel bileşenlerin $T_k = \text{köş}(t_{1,k}, t_{2,k}, \dots, t_{n,k})$ koşullu olasılıkları ile yapılması her eğrinin önemi düşünülerek gerçekleştirilmektedir. Sonuç olarak, ilk adım, $t_{i,k}$ 'lar kullanılarak

hesaplanan ortalama eğriyi çıkartarak k grup içindeki X^i eğrisini merkezlemektir (Jacques ve Preda, 2014b). $C_{i,j,k}$ temel bileşen puanları ise;

$$C_{i,j,k} = (\lambda_{j,k})^{-\frac{1}{2}} \lambda_{i,k} W B_{j,k} \quad (3.19)$$

olup, burada $B_{j,k} = W^{-\frac{1}{2}} u_{j,k}$ 'dir.

$u_{j,k}$ ve $\lambda_{j,k}$ ise $n^{-1} W^{\frac{1}{2}} \Gamma^T T_k \Gamma W^{\frac{1}{2}}$ matrisinin j -inci öz vektörleri ve öz değerleridir.

M adımı: Bu adım, $Q(\theta; \theta^h)$ 'yı en büyük yapan $\theta^{(h+1)}$ modeli parametrelerinin hesaplanmasına dayanmaktadır. Sırası ile $\pi_k^{(h+1)}$ ve $\sigma_{j,k}^{2(h+1)}$;

$$\pi_k^{(h+1)} = \frac{1}{n} \sum_{i=1}^n t_{i,k}, \quad 1 \leq j \leq q_k \quad (3.20)$$

$$\sigma_{j,k}^{2(h+1)} = \lambda_{j,k}, \quad 1 \leq j \leq q_k \quad (3.21)$$

eşitlikleri ile ifade edilmektedir.

Burada $\lambda_{j,k}$, k kümesinin temel bileşeninin varyansıdır.

EM algoritması, iki ardışık adımın yaklaşık olabilirlik değeri farkı, verilen bir ε eşik değerinden düşük olduğu zaman durmaktadır.

4. FONKSİYONEL VERİLER İÇİN ÖRNEK PROBLEMLER

4.1. Stokastik Süreçler

T , sonlu bir aralık olsun ve $\|f\|$, $t \in T$ için $f(t)$ fonksiyonunun, L_2 -normunu gösterebiliriz. Bu takdirde söz konusu normu;

$$\|f\| = \left[\int_T f^2(t) dt \right]^{\frac{1}{2}} \quad (4.1)$$

şeklinde tanımlayabiliriz.

Eğer $\|f\| < \infty$ ise, bu durum $t \in T$ için $f(t)$ 'nin karesi alınmış integrallenebilir bir fonksiyon olduğunu belirtmektedir. Bu durumda $f(t) \in L_2(T)$ olur. Burada $L_2(T)$, T aralığı üzerinde karesi integrallenebilir fonksiyonlar tarafından oluşturulan Hilbert uzayını göstermektedir. Bu uzayda ilişkili iç çarpım fonksiyonu, $f(t), g(t) \in L_2(T)$ olmak üzere;

$$\langle f, g \rangle = \int_T f(t)g(t) dt \quad (4.2)$$

şeklinde tanımlanır.

Bir stokastik süreç, zaman veya mekan üzerinde rastgele değişkenler dizisi veya vektörler olarak tanımlanmaktadır. Zaman içerisinde stokastik süreç, zaman serisi olarak bilinmektedir (Xu vd, 2011).

4.1.1. Gauss süreçleri

Gauss süreçleri özel stokastik süreçlerdir. ÇDVA'da, normal dağılım önemli bir rol oynamaktadır. FVA'da ise Gauss süreçleri, ÇDVA'daki normal dağılım ile benzerlik göstermektedir.

Bir $y(t)$ süreci, $t \in T$ olmak üzere, ortalama fonksiyonu $\eta(t), t \in T$ ve kovaryans fonksiyonu $\gamma(s, t), s, t \in T$ ile bir Gauss sürecidir ve $GP(\eta, \gamma)$ şeklinde gösterilir. $t_j, (j = 1, 2, \dots, p)$ herhangi bir p zaman noktası için, rastgele vektör $[y(t_1), y(t_2), \dots, y(t_p)]^T$ olmak üzere çok değişkenli normal dağılım $N_p(\eta, \Gamma)$ şeklinde ifade edilmektedir.

Burada η ve Γ sırasıyla ortalama vektörü ve varyans-kovaryans matrisi olup;

$$\eta = [\eta(t_1), \eta(t_2), \dots, \eta(t_p)]^T: p \times 1 \quad (4.3)$$

$$\Gamma = (\gamma(t_i, t_j)): p \times p \quad (4.4)$$

eşitlikleri ile verilir.

4.1.2. Wishart süreçleri

Wishart süreçleri, Wishart rastgele matrislerinin genelleştirilmiş halidir. $WP(n, \gamma)$, n serbestlik dereceli ve $\gamma(s, t)$ kovaryans fonksiyonlu bir Wishart sürecini belirtmek için kullanılmaktadır. Genel bir Wishart süreci $W(s, t) \sim WP(n, \gamma)$ şeklinde ifade edilmektedir.

$$W(s, t) = \sum_{i=1}^n W_i(s, t) = \sum_{i=1}^n v_i(s)v_i(t) \quad (4.5)$$

Burada, $i = 1, 2, \dots, n$ için;

$W_i(s, t) = v_i(s)v_i(t) \sim WP(1, \gamma)$ ve $v_i(t) \sim GP(0, \gamma)$ 'dir.

4.1.3. χ^2 -tipi karışımlar

$W(s, t) \sim WP(n, \gamma)$ olduğu zaman, $iz(W)$ χ^2 -tipi bir karışımdır. Genel bir χ^2 -tipi karışım;

$$T = \sum_{r=1}^q c_r A_r, \quad A_r \sim \chi_{d_r}^2(u_r^2) \quad (4.6)$$

şeklinde ifade edilmektedir.

Burada, $c_r, r = 1, 2, \dots, q$, sıfır olmayan reel katsayılarıdır ve $u_r^2, r = 1, 2, \dots, q$ ise χ^2 değişkenleri $A_r, r = 1, 2, \dots, q$ 'nin merkez dışı parametreleridir. $u_r^2 = 0$ olduğu zaman, T merkezi bir χ^2 -tipi karışım olarak adlandırılmaktadır (Zhang, 2005).

χ^2 -tipi bir karışımın dağılımı, fonksiyonel veriler için varyans analizinde, parametrik olmayan uyum testlerinde (Zhang, 2005) ve klasik varyans analizinde (Satterthwaite, 1946) kullanılmaktadır.

4.1.4. Kümülantlar

χ^2 -tipi bir karışımın dağılımını hesaplamak için, rastgele bir değişkenin kümülantları kavramına ihtiyaç vardır. X , rastgele bir değişken olmak üzere X 'in karakteristik fonksiyonu $\psi_X(t)$ olarak gösterilmektedir. $\log(\psi_X(t))$ ise;

$$\log(\psi_X(t)) = \sum_{l=1}^{\infty} K_l(X) \frac{(it)^l}{l!} \quad (4.7)$$

şeklinde ifade edilmektedir.

$K_l(X)$, $l = 1, 2, \dots, \infty$ sabitleri, X 'in kümülantları olarak bilinmektedir (Muirhead, 1982). İlk dört kümülant:

$$K_1(X) = E(X) \quad (4.8)$$

$$K_2(X) = Var(X) \quad (4.9)$$

$$K_3(X) = E(X - EX)^3 \quad (4.10)$$

$$K_4(X) = E(X - EX)^4 - 3Var^2(X) \quad (4.11)$$

eşitlikleri ile verilir.

Kümülantları kullanarak, X 'in çarpıklığı ve basıklığı;

$$sk(T) = \frac{K_3(X)}{\frac{3}{K_2^2(X)}} \quad (4.12)$$

$$ku(T) = \frac{K_4(X)}{K_2^2(X)} \quad (4.13)$$

şeklinde gösterilir.

X rastgele değişkeninin dağılımının tanımlanmasının zor olduğu durumlarda, benzer dağılıma sahip ve dağılımı daha kolay elde edilebilen bir R rastgele değişkeni kullanılarak X 'in dağılımı yaklaşık olarak türetilebilir. Bunun için X ve R rastgele değişkenlerinin ilk iki veya üç kümülantını eşleştirmek yeterli olacaktır (Zhang, 2005). Böylece,

$$P(X \leq x) \approx P(R \leq x) \quad (4.14)$$

yazılabilir.

χ^2 -tipi karışım'da T 'nin kümülantları;

$$K_l(X) = 2^{l-1}(l-1)! \sum_{r=1}^q c_r^l (d_r + lu_r^2), \quad l = 1, 2, \dots, \infty \quad (4.15)$$

eşitliği ile elde edilmektedir.

T 'nin ilk 4 kümülantı, eşitlik (4.15) kullanıldığında;

$$K_1(T) = \sum_{r=1}^q c_r (d_r + u_r^2) \quad (4.16)$$

$$K_2(T) = 2 \sum_{r=1}^q c_r^2 (d_r + 2u_r^2) \quad (4.17)$$

$$K_3(T) = 8 \sum_{r=1}^q c_r^3 (d_r + 3u_r^2) \quad (4.18)$$

$$K_4(T) = 48 \sum_{r=1}^q c_r^4 (d_r + 4u_r^2) \quad (4.19)$$

olarak bulunur.

4.1.5. F -tipi karışımlar

Fonksiyonel veriler için istatistiksel çıkarımlarda, F -tipi karışım olarak bilinen iki bağımsız χ^2 -tipi karışımın oranı ile karşılaşmak mümkündür (Zhang, 2011a). F -tipi karışım;

$$F = \frac{\sum_{r=1}^{q_1} c_r A_r}{\sum_{s=1}^{q_2} h_s B_s} = \frac{A}{B}, \quad A_r \sim \chi_{a_r}^2(u_r^2), \quad B_s \sim \chi_{b_s}^2(v_s^2) \quad (4.20)$$

şeklinde ifade edilmektedir. Burada, A_r ve B_s birbirinden bağımsızdır, $h_s > 0$ olarak kabul edilir. Serbestlik dereceleri a_r ve b_s pozitif tam sayılardır. Merkez dışı parametreler u_r^2 ve v_s^2 ise negatif olmayan parametrelerdir.

4.2. Fonksiyonel Veriler için Tek Örnek Problemi

Bir fonksiyonel veri, $t \in T$ için ortalama fonksiyonu $\eta(t)$ ve kovaryans fonksiyonu $\gamma(s, t)$ olan $y(t) \sim SP(\eta, \gamma)$ süreci ile verilsin. Bu sürecin ortalama fonksiyonu hakkında bir hipotez testi ile bilgi edinebiliriz. Bu amaçla test edilecek olan hipotezler,

$$H_0: \eta(t) \equiv \eta_0(t), \quad t \in T \quad (4.21)$$

$$H_1: \eta(t) \neq \eta_0(t)$$

biçiminde oluşturulur. H_0 hipotezini test etmek için n birimlik bir rastgele örnek $y_1(t), y_2(t), \dots, y_n(t)$ olsun. $i = 1, 2, \dots, n$ için $y_i(t) \sim SP(\eta, \gamma)$ olup, bağımsız rastgele fonksiyonlardır. Fonksiyonel verilerde stokastik sürecin ortalama fonksiyonu ile ilgili H_0 hipotezinin test edilmesinde kullanılan yöntemler mevcuttur. İzleyen kısımlarda bu yöntemlerden bahsedilecektir.

Burada $\eta_0(t)$, ilgili fizik kuramlarına, geçmiş deneyimlere veya geçmişteki deneysel sonuçlara dayanılarak önceden belirlenmiş bir fonksiyondur. $\eta_0(t)$, veri üzerinde gürültü olup olmadığını veya T üzerinde zaman etkisinin olup olmadığını test etmek için 0 olarak tanımlanmaktadır.

$\eta(t)$ ve $\gamma(s, t)$ 'nin yansız tahmin edicileri, yani örnek ortalama ve kovaryans fonksiyonları sırasıyla;

$$\hat{\eta}(t) = \bar{y}(t) = n^{-1} \sum_{i=1}^n y_i(t) \quad (4.22)$$

$$\hat{\gamma}(s, t) = (n - 1)^{-1} \sum_{i=1}^n [y_i(s) - \hat{\eta}(s)][y_i(t) - \hat{\eta}(t)] \quad (4.23)$$

şeklinde ifade edilir.

4.2.1. Tek örnek için noktasal testler

Noktasal testler, her zaman noktası $t \in T$ için tek örnek probleminde sıfır hipotezini test etmeyi amaçlamaktadır. Herhangi bir sabit t noktası için hipotezler,

$$H_{0t}: \eta(t) = \eta_0(t), \quad t \in T$$

$$H_{1t}: \eta(t) \neq \eta_0(t)$$

olacaktır. Noktasal testler için test istatistiği;

$$z(t) = \frac{\Delta(t)}{\sqrt{\hat{\gamma}(t, t)}} = \frac{\sqrt{n}[\bar{y}(t) - \eta_0(t)]}{\sqrt{\hat{\gamma}(t, t)}} \quad (4.24)$$

şeklinde tanımlıdır.

Fonksiyonel veri Gaussian dağılımına sahip olduğunda, test istatistiğinin örnekleme dağılımı her bir $t \in T$ için,

$$z(t) \sim t_{n-1} \quad (4.25)$$

olacaktır (Zhang, 2013c).

Noktasal t-testi, eşitlik (4.25) dağılımına bağlı olarak her t zaman noktasında H_{0t} için bir t-testi gerçekleştirmeyi amaçlamaktadır. Gauss varsayımı geçerli olmadığında, büyük hacimli veriler için noktasal z-testi kullanılabilir. $n \rightarrow \infty$ iken herhangi bir $t \in T$ için,

$$z(t) \xrightarrow{d} N(0,1), \quad t \in T \quad (4.26)$$

şeklinde ifade edilmektedir (Zhang, 2011b).

Eğer $|z(t)| > t_{n-1; 1-\frac{\alpha}{2}}$ veya $|z(t)| > z_{1-\alpha/2}$ ise H_{0t} hipotezi reddedilir, aksi takdirde ret edilemez. Ortalama fonksiyonu için % 100(1 - α) güven aralığı;

$$\bar{y}(t) \pm t_{n-1; 1-\frac{\alpha}{2}} \sqrt{\hat{\gamma}(t, t)/n}, \quad t \in T \quad (4.27)$$

veya

$$\bar{y}(t) \pm z_{1-\alpha/2} \sqrt{\hat{\gamma}(t, t)/n}, \quad t \in T \quad (n \rightarrow \infty) \quad (4.28)$$

olarak verilir.

Fonksiyonel veri Gaussian dağılımına sahip olmadığında ve örneklem hacmi (n) küçük olduğunda, kritik değerleri yeniden örneklemek için parametrik olmayan bir bootstrap yaklaşımı kullanılmaktadır. $y_i^*(t)$, $i = 1, 2, \dots, n$ 'den rastgele oluşturulmuş bir bootstrap örneğidir. Örnek ortalama ve kovaryans fonksiyonları, bootstrap örneğine göre, $\bar{y}^*(t)$ ve $\hat{\gamma}^*(s, t)$ olarak oluşturulmaktadır. Bootstrap test istatistiği $z^*(t)$, eşitlik (4.24)'de olduğu gibi hesaplanmaktadır. Fakat $\bar{y}(t)$ ve $\hat{\gamma}(t, t)$, sırasıyla $\bar{y}^*(t)$ ve $\hat{\gamma}^*(t, t)$ ile yer değiştirmektedir (Zhang ve Liang, 2013).

4.2.2. Tek örnek için L_2 -norm tabanlı test

H_0 hipotezi altında, Gauss varsayımı geçerli olduğunda, test fonksiyonu $\Delta(t) \sim GP(0, \gamma)$ 'dir. Gauss varsayımı geçerli olmayıp n büyük olduğunda, test fonksiyonu $\Delta(t) \rightarrow GP(0, \gamma)$ 'dir. L_2 -norm tabanlı test, test istatistiği olarak

$\Delta(t)$ 'nin karesi alınmış olan L_2 -normunu kullanmaktadır. L_2 -norm tabanlı test için test istatistiği;

$$T_n = \|\Delta\|^2 = n \int_T [\bar{y}(t) - \eta_0(t)]^2 dt \quad (4.29)$$

şeklinde ifade edilmektedir.

T_n değerinin, H_0 hipotezi altında küçük ve alternatif hipotezler altında büyük olması beklenmektedir. T_n istatistiği,

$$T_n \cong \sum_{r=1}^m \lambda_r A_r, \quad A_r \sim \chi_1^2 \quad (4.30)$$

eşitliği ile verilmektedir.

Gauss varsayımı n büyük olduğunda asimptotik olarak geçerlidir. T_n 'in dağılımı, Welch-Satterthwaite χ^2 yaklaşımına göre $T_n \sim \beta \chi_d^2$ olup parametreler;

$$\beta = \frac{iz(\gamma^{\otimes 2})}{iz(\gamma)} \quad (4.31)$$

$$d = \frac{iz^2(\gamma)}{iz(\gamma^{\otimes 2})} \quad (4.32)$$

olarak verilir (Zhang vd, 2010).

β ve d parametreleri, fonksiyonel veriler için hipotezlere dayalı olarak tahmin edilir. Bunun için $iz(\gamma)$, $iz^2(\gamma)$, $iz(\gamma^{\otimes 2})$ ve $\gamma(s, t)$ kendi yansız tahmin edicileri ile yer değiştirir. Böylece parametreler sırasıyla;

$$\hat{\beta} = \frac{iz(\hat{\gamma}^{\otimes 2})}{iz(\hat{\gamma})} \quad (4.33)$$

$$\hat{d} = \frac{iz^2(\hat{\gamma})}{iz(\hat{\gamma}^{\otimes 2})} \quad (4.34)$$

olarak tahmin edilir. Burada $\hat{\gamma}(s, t)$ eşitlik (4.23) ile hesaplanır. Diğer taraftan T_n istatistiği yaklaşık olarak,

$$T_n \sim \hat{\beta} \chi_{\hat{d}}^2 \quad (4.35)$$

şeklinde bir dağılım göstermektedir (Zhang, 2013a).

$iz^2(\hat{\gamma})$ ve $iz(\hat{\gamma}^{\otimes 2})$ sırasıyla $iz^2(\gamma)$ ve $iz(\gamma^{\otimes 2})$ için sapmalı olduğundan bu yöntem, tek örnek problemi için naive yöntem olarak bilinmektedir (Zhang, 2011a). $iz^2(\gamma)$ ve $iz(\gamma^{\otimes 2})$ 'nin yansız tahmin edicileri,

$$\widehat{iz^2}(\gamma) = \frac{(n-1)n}{(n-2)(n+1)} \left[iz^2(\hat{\gamma}) - \frac{2iz(\hat{\gamma}^{\otimes 2})}{n} \right] \quad (4.36)$$

$$iz(\widehat{\gamma}^{\otimes 2}) = \frac{(n-1)^2}{(n-2)(n+1)} \left[iz(\hat{\gamma}^{\otimes 2}) - \frac{iz^2(\hat{\gamma})}{n-1} \right] \quad (4.37)$$

eşitlikleri ile verilir.

$iz^2(\gamma)$ ve $iz(\gamma^{\otimes 2})$, kendi yansız tahmin edicileri ile değiştirilerek β ve d parametrelerinin tahmin edilmesine, yanlış indirgenmiş yöntem adı verilir. Böylece parametrelerin tahmin edicileri,

$$\hat{\beta} = \frac{iz(\widehat{\gamma}^{\otimes 2})}{iz(\hat{\gamma})} \quad (4.38)$$

$$\hat{d} = \frac{\widehat{iz^2}(\gamma)}{iz(\widehat{\gamma}^{\otimes 2})} \quad (4.39)$$

şeklinde ifade edilmektedir (Zhang ve Liang, 2013).

4.2.3. Tek örnek için F -tipi test

F -tipi testin iki avantajı vardır. Birincisi, ölçek değişmezliğidir. Yani, fonksiyonel veriler sıfırdan farklı gerçel sayı ile çarpıldığında F -tipi test değişmezdir. İkincisi, F -tipi testin dağılımı, serbestlik derecesi (d parametresi) ile orantılı, bir F dağılımına yaklaştırılabilir. Bu parametre uygun şekilde tahmin edildiğinde, F -tipi test kolay ve hızlı uygulanmaktadır (Zhang vd, 2010).

F -tipi test için test istatistiği;

$$F_n = \frac{\|\Delta\|^2}{iz(\hat{\gamma})} = \frac{n \int_T [\bar{y}(t) - \eta_0(t)]^2 dt}{iz(\hat{\gamma})} \quad (4.40)$$

şeklinde ifade edilmektedir.

$iz(\hat{\gamma})$ varyansı hesaba katılmadığında, test istatistiği F_n , L_2 -norm tabanlı test istatistiği T_n 'e eşdeğerdir. H_0 hipotezi altında,

$$\|\Delta\|^2 = \sum_{r=1}^m \gamma_r A_r, \quad A_r \sim \chi_1^2 \quad (4.41)$$

$$iz(\hat{\gamma}) = \sum_{r=1}^m \gamma_r E_r, \quad E_r \sim \chi_1^2 \quad (4.42)$$

şeklinde ifade edilmektedir.

Burada A_r ve E_r bağımsızdır. Dolayısıyla, H_0 hipotezi altında, F_n , F -tipi bir karışımdır. Yani, F_n 'nin dağılımı;

$$F_n \sim F_{\hat{d}; (n-1)\hat{d}} \quad (4.43)$$

şeklinde olacaktır (Zhang, 2013c).

Belirli bir anlamlılık düzeyi α için $F_n > F_{\hat{d}; (n-1)\hat{d}; (1-\alpha)}$ olduğunda, H_0 hipotezi ret edilmektedir.

4.2.4. Tek örnek için bootstrap testi

L_2 -norm tabanlı test için bootstrap test istatistiği;

$$\Delta^*(t) = \sqrt{n}[\bar{y}^*(t) - \bar{y}(t)] \quad (4.44)$$

şeklinde ifade edilmektedir.

Bu işlem, çok sayıda tekrarlanmaktadır, böylece T_n 'in $100(1 - \alpha)$ yüzdesini tahmin etmek için kullanılabilir bir T_n^* bootstrap örneği elde edilmektedir.

Benzer şekilde, F -tipi test için bootstrap test istatistiği;

$$F_n^* = \frac{\|\Delta^*\|}{iz(\hat{\gamma}^*)} \quad (4.45)$$

şeklinde ifade edilmektedir (Zhang, 2013b).

Bu işlem çok sayıda tekrarlanmaktadır, böylece F_n 'nin $100(1 - \alpha)$ yüzdesini tahmin etmek için kullanılabilir bir F_n^* bootstrap örneği elde edilmektedir.

Bootstrap testlerini uygulamak kolaydır, ancak zaman alıcıdır. Örnek boyutu küçük olduğunda kullanışlıdır.

4.3. Fonksiyonel Veriler için İki Örnek Problemi

Ortak bir kovaryans fonksiyonuna sahip fonksiyonel veriler için iki örnek problemi;

$$y_{11}(t), y_{12}(t), \dots, y_{1n_1}(t) \sim SP(\eta_1, \gamma)$$

$$y_{21}(t), y_{22}(t), \dots, y_{2n_2}(t) \sim SP(\eta_2, \gamma)$$

şeklinde ifade edilmektedir.

Burada $\eta_1(t)$ ve $\eta_2(t)$ bilinmeyen ortalama fonksiyonlardır. $\gamma(s, t)$ ise bilinmeyen ortak kovaryans fonksiyonudur. Test edilmek istenen hipotezler;

$$H_0: \eta_1(t) \equiv \eta_2(t), \quad t \in T \quad (4.46)$$

$$H_1: \eta_1(t) \neq \eta_2(t)$$

şeklinde kurulmaktadır.

Burada T , söz konusu bir zaman aralığı olup, $-\infty < a < b < \infty$ olacak şekilde sonlu bir $[a, b]$ aralığına kısıtlanabilir.

İki fonksiyonel örneklem temel alınarak, $\eta_1(t), \eta_2(t)$ ortalama fonksiyonlarının ve $\gamma(s, t)$ ortak kovaryans fonksiyonlarının yansız kestiricileri sırasıyla,

$$\hat{\eta}_i(t) = \bar{y}_i(t) = n_i^{-1} \sum_{j=1}^{n_i} y_{ij}(t), \quad i = 1, 2 \quad (4.47)$$

$$\hat{\gamma}(s, t) = (n-1)^{-1} \sum_{i=1}^2 \sum_{j=1}^{n_i} [y_{ij}(s) - \bar{y}_i(s)][y_{ij}(t) - \bar{y}_i(t)] \quad (4.48)$$

şeklinde ifade edilmektedir. Burada $n = n_1 + n_2$, iki verinin toplam boyutunu ifade etmektedir.

4.3.1. İki örnek için eksensel test fonksiyonu

İki örnek problemi test etmek için kullanılan eksensel test fonksiyonu;

$$\Delta(t) = \sqrt{\frac{n_1 n_2}{n}} [\bar{y}_1(t) - \bar{y}_2(t)] \quad (4.49)$$

şeklinde ifade edilmektedir. Bu test fonksiyonu iki verinin ölçekli ortalama fonksiyonunun farkıdır. Sıfır hipotezi kabul edildiğinde, bu fark küçük olacaktır. Benzer şekilde sıfır hipotezi ret edildiğinde fark büyük olacaktır. $\Delta(t)$ test fonksiyonunun ortalama ve kovaryans fonksiyonları sırasıyla;

$$\eta_{\Delta}(t) = E\Delta(t) = \sqrt{\frac{n_1 n_2}{n}} [\eta_1(t) - \eta_2(t)]$$

$$\gamma(s, t) = Cov[\Delta(s), \Delta(t)]$$

şeklinde ifade edilmektedir.

Sıfır hipotezi altında $E\Delta(t) \equiv 0, t \in T$ olacaktır.

4.3.2. İki örnek için noktasal testler

Noktasal bir testin amacı, sıfır hipotezini her zaman noktası $t \in T$ 'de test etmektir.

Herhangi bir sabit $t \in T$ için, test edilmek istenen hipotezler;

$$H_{0t}: \eta_1(t) = \eta_2(t), \quad t \in T$$

$$H_{1t}: \eta_1(t) \neq \eta_2(t)$$

şeklinde kurulmaktadır.

Örnek ortalama fonksiyonlarına ve toplanmış örnek kovaryans fonksiyonuna dayanarak eksensel test istatistiği;

$$z(t) = \frac{\Delta(t)}{\sqrt{\hat{\gamma}(t, t)}} = \frac{[\bar{y}_1(t) - \bar{y}_2(t)]}{\sqrt{(1/n_1 + 1/n_2) \hat{\gamma}(t, t)}} \quad (4.50)$$

şeklinde ifade edilmektedir. Eğer örnekler Gauss sürecine aitse, H_{0t} hipotezi altında test istatistiğinin örnekleme dağılımı,

$$z(t) \sim t_{n-2}, \quad t \in T \quad (4.51)$$

olacaktır, burada $n = n_1 + n_2$ 'dir (Zhang vd, 2010).

Noktasal t -testi için $|z(t)| > t_{n-2; 1-\alpha/2}$ olduğunda H_{0t} hipotezi ret edilir. Ortalama fonksiyon farkı $\eta_1(t) - \eta_2(t)$ için $100(1 - \alpha)$ güven aralığı;

$$[\bar{y}_1(t) - \bar{y}_2(t)] \pm t_{n-2; 1-\alpha/2} \sqrt{(1/n_1 + 1/n_2) \hat{\gamma}(t, t)}, \quad t \in T \quad (4.52)$$

şeklinde ifade edilir.

Burada, $t_{n-2; 1-\alpha/2}$, $n - 2$ serbestlik dereceli t -dağılımının $100(1 - \alpha/2)$ yüzdeleri değerini gösterir. Noktasal t -testi, eşitlik (4.51) ile verilen t -dağılımına dayalı olarak her bir $t \in T$ zaman noktasında t -testini gerçekleştirmeyi amaçlar.

Gaussian varsayımı sağlanmadığında; büyük veriler için noktasal z-testi kullanılabilir. $n_1, n_2 \rightarrow \infty$ olduğunda; her sabit $t \in T$ için,

$$z(t) \xrightarrow{d} N(0,1)$$

olduğu görülebilir (Zhang, 2013c).

Böylece, noktasal z-testi için $|z(t)| > z_{1-\alpha/2}$ olduğunda H_{0t} hipotezi ret edilecektir.

n_1, n_2 küçük ve örnekler Gaussian dağılımına sahip olmadığında, noktasal t ve z testleri tercih edilmez. Bu durumda, noktasal bootstrap testi uygulanır. $v_{ij}^*(t)$ ($j = 1, 2, \dots, n_i; i = 1, 2$) değeri, $\hat{v}_{ij}(t) = y_{ij}(t) - \hat{\eta}_i(t)$, ($j = 1, 2, \dots, n_i; i = 1, 2$) tahmini konu-etki fonksiyonlarından elde edilen bootstrap örnekleridir. Bu durumda iki grup tahmini için bootstrap örnekleme;

$$y_{ij}^*(t) = \hat{\eta}_i(t) + v_{ij}^*(t), \quad j = 1, 2, \dots, n_i; i = 1, 2 \quad (4.53)$$

şeklinde ifade edilir.

Örnek ortalama fonksiyonları $(\bar{y}_i^*(t), i = 1, 2)$ ve toplanmış örnek kovaryans fonksiyonunu $\hat{\gamma}^*(s, t)$, eşitlik (4.53)'de verilen bootstrap örneklerine dayalı olarak eşitlik (4.48)'deki gibi hesaplanabilmektedir. Noktasal bootstrap testi için,

$$z^*(t) = \frac{([\bar{y}_1^*(t) - \bar{y}_2^*(t)] - [\bar{y}_1(t) - \bar{y}_2(t)])}{\sqrt{(1/n_1 + 1/n_2)\hat{\gamma}^*(t, t)}} \quad (4.54)$$

değeri hesaplanır.

$z^*(t)$ 'nin dağılımı $z(t)$ 'nin dağılımına benzetmek için $z^*(t)$ 'nin hesaplanmasında orijinal örneklere ait ortalama fonksiyonları farkının $(\bar{y}_1(t) - \bar{y}_2(t))$ bootstrap örneklerine ait ortalama fonksiyonları farkından $(\bar{y}_1^*(t) - \bar{y}_2^*(t))$ çıkarılması önemlidir. $z^*(t)$ 'nin örnekleme dağılımından $100(1 - \alpha/2)$ yüzdelik değeri hesaplanır. Daha sonra noktasal bootstrap testi uygulanır ve noktasal bootstrap güven aralıkları kurulabilir.

4.3.3. İki örnek için L_2 -norm tabanlı test

İki örnek problemi için L_2 -norm tabanlı test, $\Delta(t)$ 'nin karesi alınmış L_2 -normunu test istatistiği olarak kullanmaktadır. Test istatistiği;

$$T_n = \|\Delta\|^2 = \frac{n_1 n_2}{n} \int_T [\bar{y}_1(t) - \bar{y}_2(t)]^2 dt \quad (4.55)$$

şeklinde ifade edilmektedir.

H_0 hipotezi altında, iki fonksiyonel veri Gaussian olduğunda, $\Delta(t) \sim GP(0, \gamma)$ elde edilmektedir. n_1 ve n_2 büyük olduğunda asimptotik olarak $\Delta(t) \sim GP(0, \gamma)$ elde edilmektedir. H_0 hipotezi altında T_n test istatistiği;

$$T_n \cong \sum_{r=1}^m \lambda_r A_r, \quad A_r \sim \chi_1^2 \quad (4.56)$$

olarak yazılabilmektedir. Burada $r = 1, 2, \dots, m$ için λ_r 'ler ortak kovaryans fonksiyonu olan $\gamma(s, t)$ 'nin pozitif özdeğerleri ve A_r 'ler bağımsız, aynı dağılımlı rastgele değişkenlerdir. T_n 'in dağılımı Welch-Satterthwaite Ki-Kare yaklaşımına göre $T_n \sim \beta \chi_d^2$ 'dir. Burada parametreler eşitlik (4.31) ve eşitlik (4.32) ile aynıdır (Zhang vd, 2010).

β parametresi, L_2 -normuna dayalı olan T_n test istatistiğinin ölçeğiyle ilişkili olduğu için genellikle ölçek parametresi olarak adlandırılmaktadır. d parametresi ise, ortak kovaryans fonksiyonu olan $\gamma(s, t)$ 'ye bağlı olan serbestlik derecesi olarak adlandırılmaktadır (Shen ve Faraway 2004). Uygulamada, β ve d parametrelerinin iki örneğe dayanılarak tahmin edilmesi gerekmektedir. Bunun için $iz(\gamma)$, $iz^2(\gamma)$ ve $iz(\gamma^{\otimes 2})$ 'nin tahmin edicileri ile yer değiştirmesi gerekir. Burada parametreler eşitlik (4.33) ve eşitlik (4.34) ile aynıdır.

Burada $\hat{\gamma}(s, t)$ eşitlik (4.48)'de verilen toplanmış örnek kovaryans fonksiyonudur. Bu durumda eşitlik (4.35) yazılabilir (Zhang, 2013a).

$iz^2(\hat{\gamma})$ ve $iz(\hat{\gamma}^{\otimes 2})$ sırasıyla $iz^2(\gamma)$ ve $iz(\gamma^{\otimes 2})$ için sapmalı olduğundan bu yöntem, iki örnek problemi için naive yöntem olarak bilinmektedir (Zhang, 2011b). $iz^2(\gamma)$ ve $iz(\gamma^{\otimes 2})$ 'nin yansız tahmin edicileri sırasıyla;

$$\widehat{iz^2}(\gamma) = \frac{(n-2)(n-1)}{(n-3)n} \left[iz^2(\hat{\gamma}) - \frac{2iz(\hat{\gamma}^{\otimes 2})}{n-1} \right] \quad (4.57)$$

$$iz(\widehat{\gamma^{\otimes 2}}) = \frac{(n-2)^2}{(n-3)n} \left[iz(\hat{\gamma}^{\otimes 2}) - \frac{iz^2(\hat{\gamma})}{n-2} \right] \quad (4.58)$$

şeklinde ifade edilmektedir.

$(n - 2) \gamma (s, t) \sim WP (n - 2, \gamma)$ 'dır. $iz^2(\gamma)$ ve $iz(\gamma^{\otimes 2})$ 'nin kendi yansız tahmin edicileri ile yer değiştirilerek, β ve d parametrelerinin tahmin edilmesine yanlı indirgenmiş yöntem adı verilir. Ortaya çıkan tahmin ediciler;

$$\hat{\beta} = \frac{\frac{(n-2)^2}{(n-3)n} \left[iz(\hat{\gamma}^{\otimes 2}) - \frac{iz^2(\hat{\gamma})}{n-2} \right]}{iz(\hat{\gamma})} \quad (4.59)$$

$$\hat{d} = \frac{(n-1) \left[iz^2(\hat{\gamma}) - \frac{2iz(\hat{\gamma}^{\otimes 2})}{n-1} \right]}{(n-2) \left[iz(\hat{\gamma}^{\otimes 2}) - \frac{iz^2(\hat{\gamma})}{n-2} \right]} \quad (4.60)$$

eşitlikleri ile verilir (Zhang vd, 2010).

n_1 ve n_2 büyük olduğunda, yanlı indirgenmiş yöntemle elde edilen tahmin ediciler $\hat{\beta}$ ve \hat{d} , naive yöntemle elde edilen tahmin edicilere yaklaşık olarak eşittir. Ayrıca, örneklem boyutları $n_1, n_2 \rightarrow \infty$ iken $\hat{\beta}$ ve \hat{d} tahmin edicileri tutarlıdır.

$n \rightarrow \infty$ için $iz(\hat{\gamma}) \rightarrow iz(\gamma)$ ve $iz(\hat{\gamma}^{\otimes 2}) \rightarrow iz(\gamma^{\otimes 2})$ olmaktadır. Ayrıca, $n \rightarrow \infty$ iken;

$$\hat{\beta} \xrightarrow{p} \beta, \hat{d} \xrightarrow{p} d$$

olmaktadır (Zhang, 2013c). Burada $\hat{\beta}$ ve \hat{d} sırasıyla β ve d 'nin yanlı indirgenmiş tahmin edicileridir.

L_2 -norm tabanlı test için herhangi bir α anlamlılık seviyesinde $T_n > \hat{\beta} \chi_{\hat{d}, 1-\alpha}^2$ olduğunda H_0 hipotezi ret edilir.

4.3.4. İki örnek için F -tipi test

H_0 hipotezi ve Gauss varsayımı altında, $\Delta(t) \sim GP(0, \gamma)$ ve $(n - 2) \gamma (s, t) \sim WP (n - 2, \gamma)$ 'dir. Ayrıca,

$$E\|\Delta\|^2 = iz(\gamma) \quad (4.61)$$

$$Eiz(\hat{\gamma}) = iz(\gamma) \quad (4.62)$$

olmaktadır.

F -tipi test istatistiği;

$$F_n = \frac{\|\Delta\|^2}{iz(\hat{\gamma})} = \frac{\frac{n_1 n_2}{n} \int_T [\bar{y}_1(t) - \bar{y}_2(t)]^2 dt}{iz(\hat{\gamma})} \quad (4.63)$$

eşitliği ile verilir.

$iz(\hat{\gamma})$ değişimi hesaba katılmadığında, F_n dağılımı L_2 -norm tabanlı test istatistiği T_n ile aynıdır. Bu değişimi hesaba katmak için Gaussian varsayımı yeterlidir. Gaussian varsayımı ve H_0 hipotezi altında,

$$\|\Delta\|^2 = \sum_{r=1}^m \lambda_r A_r, \quad A_r \sim \chi_1^2 \quad (4.64)$$

$$iz(\hat{\gamma}) = \left(\sum_{r=1}^m \lambda_r E_r \right) / (n-2), \quad E_r \sim X_{n-2}^2 \quad (4.65)$$

eşitlikleri yazılabilir. Burada $r = 1, 2, \dots, m$ için A_r ve E_r 'ler bağımsız, aynı dağılımlı rastgele değişkenler ve λ_r 'ler ise $\gamma(s, t)$ kovaryans fonksiyonunun pozitif öz değerleridir.

Eşitlik (4.64) ve eşitlik (4.65) dikkate alındığında eşitlik (4.63) ile verilen F -tipi test istatistiği;

$$F_n = \frac{\sum_{r=1}^m \lambda_r A_r}{\sum_{r=1}^m (\lambda_r E_r) / (n-2)} \quad (4.66)$$

biçiminde de yazılabilmektedir (Zhang, 2013c).

Gauss varsayımı ve H_0 hipotezi altında, F_n test istatistiği F -tipi bir karışımdır. Dolayısıyla, F_n test istatistiğinin örnekleme dağılımı yaklaşık olarak,

$$F_n \sim F_{d; (n-2)d}$$

yazılabilmektedir (Zhang vd, 2010). Burada d , eşitlik (4.32)'deki gibi tanımlıdır.

Anlamlılık seviyesi α olmak üzere, eğer $F_n > F_{\hat{d}; (n-2)\hat{d}; 1-\alpha}$ ise H_0 hipotezi ret edilir. Aksi takdirde H_0 hipotezi ret edilemez.

4.3.5. İki örnek için bootstrap testleri

İki fonksiyonel örnek Gaussian dağılımlı olmadığında, n_1 ve n_2 örneklem hacimleri küçük olduğunda, L_2 -norm tabanlı testin ve F -tipi testin dağılımlarını bulmak kolay değildir. Bu durumda, L_2 -norm tabanlı bootstrap testi veya F -tipi bootstrap testi uygulanmalıdır. Bootstrap örnekleme ile elde edilen örneğe ait, ortalama

fonksiyonları $\bar{y}_1^*(t)$, $\bar{y}_2^*(t)$ ve örnek kovaryans fonksiyonu $\hat{\gamma}^*(s, t)$ hesaplanmış olsun.

L_2 -norm tabanlı bootstrap testi için test istatistiği $T_n^* = \|\Delta^*\|^2$ 'dir. Burada $t \in T$ için Δ^* istatistiği;

$$\Delta^*(t) = \sqrt{\frac{n_1 n_2}{n}} [(\bar{y}_1^*(t) - \bar{y}_2^*(t)) - (\bar{y}_1(t) - \bar{y}_2(t))] \quad (4.67)$$

şeklinde ifade edilmektedir. Tanımlanan bu bootstrap işleminin çok sayıda tekrarlanması gerekmektedir. $T_n = \|\Delta\|^2$ 'nin $100(1 - \alpha)$ yüzde değerini tahmin etmek için kullanılacak bir T_n^* bootstrap örneği elde edilebilmektedir. Benzer şekilde, F -tipi bootstrap testi için test istatistiği $F_n^* = \frac{\|\Delta^*\|}{iz(\hat{\gamma}^*)}$ 'dir. Bu işlem, F_n^* 'nin $100(1 - \alpha)$ yüzdesinin tahminine izin veren bir bootstrap örneği elde edebilmek için çok sayıda tekrarlamak gerekmektedir (Zhang, 2013c).

4.4. Fonksiyonel Veriler için k Örnek Problemi (FANOVA)

Klasik istatistikte, tek yönlü varyans analizi (ANOVA), üç veya daha fazla grubun ortalamalarını karşılaştırmak için kullanılan bir tekniktir. Gruplar, faktör olarak bilinen kategorik bir değişkene göre oluşturulur. Faktörün aldığı farklı değerlerin sayısına faktörün seviyesi denir. Fonksiyonel veriler için tek yönlü ANOVA problemi şu şekilde tanımlanabilir. k sayıda bağımsız örnekler,

$$y_{i1}(t), y_{i2}(t), \dots, y_{in_i}(t), \quad i = 1, 2, \dots, k \quad (4.68)$$

olmak üzere, bu örnekler;

$$y_{ij}(t) = \eta_i(t) + v_{ij}(t), \quad v_{ij}(t) \sim SP(0, \gamma), \quad j = 1, 2, \dots, n_i; i = 1, 2, \dots, k \quad (4.69)$$

eşitliğini sağlar.

Burada $\eta_1(t), \eta_2(t), \dots, \eta_k(t)$, k adet örneklemin bilinmeyen grup ortalama fonksiyonları, $v_{ij}(t)$, ($j = 1, 2, \dots, n_i; i = 1, 2, \dots, k$) konu-etki fonksiyonları olup, bunlar birbirinden bağımsız ve aynı dağılımlı rastgele süreçlerdir ve $\gamma(s, t)$ ortak kovaryans fonksiyonudur. Tek yönlü ANOVA test problemi için test edilmesi gereken hipotezler;

$$H_0: \eta_1(t) \equiv \eta_2(t) \equiv \dots \equiv \eta_k(t), \quad t \in T \quad (4.70)$$

$$H_1: \exists \eta_i(t) \text{ diğ}erlerinden \text{ farklıdır}$$

şeklinde kurulmaktadır.

Burada T , $-\infty < a < b < \infty$ olmak üzere $[a, b]$ olarak belirtilen belirli bir zaman aralığıdır. Tek yönlü ANOVA problemi, fonksiyonel veriler için k -örnek problemi olarak da bilinmektedir ve fonksiyonel veriler için iki örnek probleminin bir genişlemesidir.

Tek yönlü ANOVA test problemi, bir faktörün etkisinin istatistiksel olarak anlamlı olup olmadığını kontrol etmeyi amaçlamaktadır. Bu faktör genellikle, bireysel fonksiyonları çeşitli örneklere, gruplara veya kategorilere ayırmak için kullanılmaktadır. Faktör fonksiyonel veriler üzerinde ciddi bir etkiye sahipse, Tek yönlü ANOVA problemi istatistiksel olarak anlamlı olacaktır. Tek yönlü ANOVA problemi için üç ana test türü ile ilgilenilir.

4.4.1. Ana etki testi

$\eta_i(t) = \eta(t) + \alpha_i(t)$, ($i = 1, 2, \dots, k$), doğrusal modeli verilsin. Burada $\eta(t)$, k tane örneğin genel ortalama fonksiyonudur ve $\alpha_i(t)$, $i = 1, 2, \dots, k$, için i -inci ana etki fonksiyonudur. Fonksiyonel veriler için tek yönlü ANOVA modeli:

$$y_{ij}(t) = \eta(t) + \alpha_i(t) + v_{ij}(t), \quad j = 1, 2, \dots, n_i; \quad i = 1, 2, \dots, k \quad (4.71)$$

şeklinde ifade edilmektedir.

Bu formülasyonda H_0 hipotezi, ana etki fonksiyonlarının aynı olup olmadığını veya 0'a eşit olup olmadığını test etmek için,

$$H_0: \alpha_1(t) \equiv \alpha_2(t) \equiv \dots \equiv \alpha_k(t) \equiv 0, \quad t \in T \quad (4.72)$$

şeklinde kurulmaktadır.

4.4.2. Post-hoc testi

H_0 hipotezi ret edilmediğinde, tek yönlü ANOVA modeli anlamlı değildir. Ret edildiğinde ise daha fazla araştırma yapılmalıdır. Herhangi iki ana etki fonksiyonu $\alpha_i(t)$ 'nin ve $\alpha_j(t)$ 'nin aynı olup olmadığını öğrenmek için hipotezler;

$$H_0: \alpha_i(t) \equiv \alpha_j(t), \quad t \in T \quad (4.73)$$

$$H_1: \alpha_i(t) \neq \alpha_j(t)$$

şeklinde kurulmaktadır (Gorecki ve Smaga, 2015).

Bu test, post-hoc testi olarak bilinir. Eş değer olarak hipotezler;

$$H_0: \eta_i(t) \equiv \eta_j(t), \quad t \in T \quad (4.74)$$

$$H_1: \eta_i(t) \neq \eta_j(t)$$

şeklinde kurulmaktadır.

4.4.3. Kontrast testi

Post-hoc testler kontrast testlerinin özel halidir. a_1, a_2, \dots, a_k toplamı 0 olup yani, $\sum_{i=1}^k a_i = a^T \mathbf{1}_k = 0$, $a = [a_1, a_2, \dots, a_k]^T$ ve $\mathbf{1}_k$ ise boyutu k olan bir sütun vektörüdür. Bir kontrast, $\sum_{i=1}^k a_i \alpha_i(t) = a^T \alpha(t)$ ana-etki fonksiyonlarının $\alpha_i(t)$, $i = 1, 2, \dots, k$, doğrusal bir kombinasyonu olarak tanımlanmaktadır; burada, $\alpha(t) = [\alpha_1(t), \alpha_2(t), \dots, \alpha_k(t)]^T$ ana-etki fonksiyonlarının tümünü içermektedir. Belirli bir $a \in \mathbb{R}^k$ için $a^T \mathbf{1}_k = 0$ olacak şekilde, bir kontrast testi;

$$H_0: a^T \alpha(t) \equiv 0, \quad t \in T \quad (4.75)$$

$$H_1: a^T \alpha(t) \neq 0$$

şeklinde tanımlanmaktadır (Gorecki ve Smaga, 2017).

$a^T \mathbf{1}_k = 0$ olduğunda, alternatif hipotezler;

$$H_0: a^T \eta(t) \equiv 0, \quad t \in T \quad (4.76)$$

$$H_1: a^T \eta(t) \neq 0$$

şeklinde kurulmaktadır.

Burada $\eta(t) = [\eta_1(t), \eta_2(t), \dots, \eta_k(t)]^T$ 'dir.

4.4.4. Grup ortalama ve kovaryans fonksiyonlarının tahmini

FVA'da tek yönlü ANOVA için k tane örnek veri yardımıyla grup ortalama fonksiyonları $\eta_i(t)$ 'ler ($i = 1, 2, \dots, k$) ve ortak kovaryans fonksiyonunun ($\gamma(s, t)$) yansız tahmin edicileri elde edilebilir. Bu yansız tahmin ediciler sırasıyla;

$$\hat{\eta}_i(t) = \bar{y}_i(t) = n_i^{-1} \sum_{j=1}^{n_i} y_{ij}(t), \quad i = 1, 2, \dots, k \quad (4.77)$$

$$\hat{\gamma}(s, t) = (n - k)^{-1} \sum_{i=1}^k \sum_{j=1}^{n_i} [y_{ij}(s) - \bar{y}_i(s)][y_{ij}(t) - \bar{y}_i(t)] \quad (4.78)$$

eşitlikleri ile verilmiştir. Burada $n = \sum_{i=1}^k n_i$, toplam veri sayısını belirtmektedir. Tahmini kovaryans fonksiyonu $\hat{\gamma}(s, t)$ örnek kovaryans fonksiyonu olarak da bilinmektedir. $\hat{\eta}_i(t)$ 'ler bağımsız ve aynı zamanda toplanmış,

$$E\hat{\eta}_i(t) = \eta_i(t)$$

$$\text{cov}[\hat{\eta}_i(s), \hat{\eta}_i(t)] = \frac{\gamma(s, t)}{n_i}, \quad i = 1, 2, \dots, k$$

özelliklerine sahiptir.

$\hat{\eta}(t) = [\hat{\eta}_1(t), \hat{\eta}_2(t), \dots, \hat{\eta}_k(t)]^T$ şeklinde alalım. Bu $\hat{\eta}(t)$, $\eta(t)$ 'nin yansız bir tahmin edicisidir. $E\hat{\eta}(t) = \eta(t)$ ve $\text{cov}[\hat{\eta}(s), \hat{\eta}(t)] = \gamma(s, t)D$ şeklinde ifade edilir. Burada $D = \text{köş}(1/n_1, 1/n_2, \dots, 1/n_k)$ olan köşegen bir matristir (Zhang, 2013c).

4.4.5. Tek yönlü varyans analizinin varsayımları

FVA'da tek yönlü varyans analizi probleminin çözümü için gerekli olan varsayımlar aşağıdaki gibidir.

1. k tane gruba ait ortalama fonksiyonları Hilbert uzayına ait olmalı, yani $i = 1, 2, \dots, k$ için $\eta_i(t) \in L_2(T)$ ve ortak kovaryans fonksiyonu için $\text{iz}(\gamma) < \infty$ olmalıdır.
2. k tane grubun her biri Gaussian dağılımlı olmalıdır.
3. $n \rightarrow \infty$ iken, k tane örnek için örnek hacimleri $i = 1, 2, \dots, k$ olmak üzere $\tau_i \in (0, 1)$ olacak şekilde $n_i/n \rightarrow \tau_i$ olmalıdır.
4. Konu-etki fonksiyonları $v_{ij}(t) = y_{ij}(t) - \eta_i(t)$ ($j = 1, 2, \dots, n_i$; $i = 1, 2, \dots, k$) olmak üzere birbirinden bağımsız ve aynı dağılıma sahip olmalıdır.
5. Konu-etki fonksiyonu $v_{11}(t)$, $E\|v_{11}\|^4 < \infty$ olmalıdır.
6. Maksimum varyans $\rho = \max_{t \in T} \gamma(t, t) < \infty$ olmalıdır.
7. $E[v_{11}^2(s), v_{11}^2(t)]$ beklenen değeri düzgün sınırlı olmalıdır.

4.5. Tek Yönlü ANOVA Problemi için Testler

4.5.1. Tek yönlü ANOVA problemi için noktasal F -testi

Noktasal F -testi, klasik F -testinin fonksiyonel veri analizi için genişletilmiş halidir.

Noktasal F -testi, için noktasal F istatistiği;

$$F_n(t) = \frac{\frac{GAKT_n(t)}{k-1}}{\frac{HKT_n(t)}{n-k}} = \frac{GAKO_n(t)}{HKO_n(t)} \quad (4.79)$$

şeklinde ifade edilir. Burada k tane örnek Gaussian dağılımlı olduğunda H_0 hipotezi altında, $F_n(t) \sim F_{k-1, n-k}$, $t \in T$ olacaktır (Zhang ve Liang, 2013).

Noktasal F -testi α anlamlılık seviyesinde $t \in T$ için $F_n(t) > F_{k-1; n-k; (1-\alpha)}$ olduğunda H_0 hipotezi ret edilir, aksi takdirde ret edilemez.

Örnek grupları Gaussian dağılımına sahip olmadığında, büyük örnekler için, noktasal χ^2 -testi kullanılmaktadır. $n_{min} = \min\{n_1, n_2, \dots, n_k\} \rightarrow \infty$ olduğunda, asimptotik olarak;

$$F_n(t) \sim \frac{\chi_{k-1}^2}{k-1}, \quad t \in T \quad (4.80)$$

şeklinde ifade edilebilmektedir (Zhang vd, 2018).

Bunun nedeni, $n_{min} \rightarrow \infty$ olduğunda, eşitlik (4.79)'da verilen test istatistiği $F_n(t)$ 'nin payında yer alan $\frac{GAKT_n(t)}{k-1}$ terimi, $\gamma(t, t)\chi_{k-1}^2 / (k-1)$ terimine yaklaşırken paydasında bulunan $\frac{HKT_n(t)}{n-k}$ ise $\hat{\nu}(t, t)$ terimine yaklaşmaktadır. Böylece noktasal χ^2 -testi için herhangi bir α anlamlılık seviyesinde, her $t \in T$ noktasında $F_n(t) > \chi_{k-1; (1-\alpha)/(k-1)}^2$ olduğunda H_0 hipotezi ret edilir, aksi takdirde ret edilemez.

Örnek grupları Gaussian dağılımlı olmadığında ve örnek büyüklükleri (n_1, n_2, \dots, n_k) küçük olduğunda noktasal F ve χ^2 testleri tercih edilmemektedir. Bu durumda bootstrap testine başvurulmaktadır.

Tahmin edilen konu-etki fonksiyonları $\hat{\nu}_{ij}(t) = y_{ij}(t) - \hat{\eta}(t_{ij})$, ($j = 1, 2, \dots, n_{ij}$; $i = 1, 2, \dots, k$) değerlerinden rastgele olarak türetilen k tane bootstrap örneği,

$$v_{ij}^*(t), \quad j = 1, 2, \dots, n_{ij}; \quad i = 1, 2, \dots, k \quad (4.81)$$

olsun. Bootstrap örneği kullanılarak, bootstrap örnek ortalama fonksiyonları ve genel ortalama fonksiyonları sırasıyla,

$$\bar{v}_i^*(t) = n_i^{-1} \sum_{j=1}^{n_i} v_{ij}^*(t) \quad (4.82)$$

$$\bar{v}_{..}^*(t) = n^{-1} \sum_{i=1}^k \sum_{j=1}^{n_i} v_{ij}^*(t) \quad (4.83)$$

eşitliklerinden hesaplanabilir. Böylece bootstrap örnekleme yardımıyla $GAKT_n^*(t)$ ve $HKT_n^*(t)$ istatistikleri sırasıyla;

$$GAKT_n^*(t) = \sum_{i=1}^k n_i [\bar{v}_i^*(t) - \bar{v}_{..}^*(t)]^2 \quad (4.84)$$

$$HKT_n^*(t) = \sum_{i=1}^k \sum_{j=1}^{n_i} [v_{ij}^*(t) - \bar{v}_i^*(t)]^2 \quad (4.85)$$

ile bulunur. Sonuç olarak noktasal bootstrap test istatistiği,

$$F_n^*(t) = \frac{\frac{GAKT_n^*(t)}{k-1}}{\frac{HKT_n^*(t)}{n-k}} \quad (4.86)$$

eşitliği ile verilir.

Burada verilen bootstrap örnekleme süreci çok sayıda tekrarlanır. $F_n(t)$ istatistiğinin $100(1 - \alpha)$ yüzdelerik değerini tahmin etmek için kullanılan $F_n^*(t)$ istatistiğinin $100(1 - \alpha)$ yüzdelerik değeri, örnekleme dağılımından hesaplanır.

4.5.2. Tek yönlü ANOVA problemi için L_2 -norm tabanlı testler

FVA'da tek yönlü ANOVA için L_2 -norm tabanlı test, iki örnek problemi için verilen L_2 -norm tabanlı testin, k örnek problemine genişletilmiş halidir. İlgili test istatistiği, konular arası noktasal değişimlerin integrali olarak,

$$T_n = \int_T GAKT_n(t) dt = \sum_{i=1}^k n_i \int_T [\bar{y}_i(t) - \bar{y}_{..}(t)]^2 dt \quad (4.87)$$

eşitliği ile verilir.

Örneklem grupları Gaussian dağılımlı ve örneklem hacmi büyük olduğunda test istatistiği H_0 hipotezi altında asimptotik olarak;

$$T_n = \sum_{r=1}^m \lambda_r A_r, \quad A_r \sim \chi_{k-1}^2 \quad (4.88)$$

istatistiğine yaklaşıır (Zhang, 2013c). Burada $r = 1, 2, \dots, m$ için λ_r 'ler ortak kovaryans fonksiyonu olan $\gamma(s, t)$ 'nin pozitif özdeğerleri ve A_r 'ler bağımsız aynı dağılımlı rastgele değişkenlerdir. T_n 'in dağılımı, Welch-Satterthwaite χ^2 -yaklaşım metoduna göre $T_n \sim \beta \chi_{(k-1)d}^2$ olacaktır. Burada β ve d parametreleri sırasıyla eşitlik (4.31) ve eşitlik (4.32)'de tanımlandığı gibidir.

β parametresi, L_2 -normuna dayalı T_n test istatistiğinin ölçek parametresi olarak adlandırılmaktadır. d parametresi ise, kovaryans fonksiyonu olan $\gamma(s, t)$ 'ye bağlı serbestlik derecesi olarak adlandırılmaktadır (Shen ve Faraway 2004).

Uygulamada bu iki parametre, fonksiyonel verilere dayalı olarak tahmin edilmelidir. Bu amaçla kullanılabilir iki yöntem mevcuttur. Birinci yöntem naive yöntem (L_2N), ikinci yöntem ise yanlı indirgenmiş yöntem (L_2B) denilmektedir. Naive yöntemle parametreler sırasıyla, eşitlik (4.33) ve eşitlik (4.34)'deki gibi tahmin edilir.

Yanlı indirgenmiş yöntemle ise;

$$\widehat{iz}^2(\gamma) = \frac{(n-k)(n-k+1)}{(n-k-1)(n-k+2)} \left[iz^2(\hat{\gamma}) - \frac{2iz(\hat{\gamma}^{\otimes 2})}{n-k+1} \right] \quad (4.89)$$

$$iz(\widehat{\gamma}^{\otimes 2}) = \frac{(n-k)^2}{(n-k-1)(n-k+2)} \left[iz(\hat{\gamma}^{\otimes 2}) - \frac{iz^2(\hat{\gamma})}{n-k} \right] \quad (4.90)$$

olmak üzere parametrelerin tahmin edicileri eşitlik (4.38) ve eşitlik (4.39)'daki gibi tahmin edilir.

L_2 -norm tabanlı test, herhangi bir anlamlılık düzeyi α için için $T_n > \hat{\beta} \chi_{(k-1)d; (1-\alpha)}^2$ olduğunda H_0 hipotezi ret edilir.

4.5.3. Tek yönlü ANOVA problemi için F -tipi test

Örnek grupları Gaussian dağılımlı olduğunda, eşitlik (4.70)'de verilen ana etki testi için bir F -tipi test yapılabilmektedir. F -tipi test için test istatistiği,

$$F_n = \frac{\int_T \frac{GAKT_n(t)dt}{k-1}}{\int_T \frac{HKT_n(t)dt}{n-k}} \quad (4.91)$$

şeklinde ifade edilir.

H_0 hipotezi altında;

$$F_n = \frac{\sum_{r=1}^m \frac{\lambda_r A_r}{k-1}}{\sum_{r=1}^m \frac{\lambda_r E_r}{n-k}} \quad (4.92)$$

olur. $r = 1, 2, \dots, m$ için λ_r 'ler ortak kovaryans fonksiyonu olan $\gamma(s, t)$ 'nin pozitif özdeğerleri ve A_r ile E_r 'ler birbirinden bağımsız rastgele değişkenlerdir. Böylece F_n istatistiğinin örnekleme dağılımı,

$$F_n \sim F_{(k-1)\hat{a}; (n-k)\hat{a}} \quad (4.93)$$

şeklinde ifade edilir (Zhang, 2013c).

Herhangi bir α anlamlılık düzeyi için $F_n > F_{(k-1)\hat{a}; (n-k)\hat{a}; 1-\alpha}$ olduğunda, H_0 hipotezi ret edilir, aksi taktirde ret edilemez.

4.5.4. Tek yönlü ANOVA problemi için bootstrap testleri

FVA'da tek yönlü ANOVA için k tane örnek grup Gaussian dağılımlı olmadığında, F -tipi test uygun olmayabilir. Bu durumda bootstrap örnek hacimleri (n_1, n_2, \dots, n_k) büyük olduğunda, H_0 hipotezini test etmek için bootstrap testleri uygulanabilmektedir. H_0 hipotezi altında $n \rightarrow \infty$ iken bootstrap test istatistiği T_n olup;

$$T_n = \int_T GAKT_n(t)dt \xrightarrow{d} \sum_{i=1}^{k-1} \int_T w_i^2(t)dt \quad (4.94)$$

şeklinde ifade edilmektedir. Burada $w_i(t)$ 'ler, $i = 1, 2, \dots, k-1$, $w(t) \sim GP_{k-1}(0, \gamma I_{k-1})$ sürecinin $k-1$ tane bileşenidir. Yani $i = 1, 2, \dots, k-1$ için $w_i(t)$ 'ler bağımsız, aynı dağılımlı ve $\gamma(s, t)$ kovaryans fonksiyonu hariç bilinen $GP(0, \gamma)$ süreçleridir. $\gamma(s, t)$ kovaryans fonksiyonunun yansız tahmin edicisi $\hat{\gamma}(s, t)$ eşitlik (4.78)'de verilmiştir. Bu yöntemin amacı $GP(0, \hat{\gamma})$ sürecinden $i = 1, 2, \dots, k-1$ için $w_i(t)$ Gauss süreçlerini, H_0 hipotezi altında T_n istatistiğinin örnekleme dağılımının bulunabileceği kadar örneklemeektir.

H_0 hipotezi altında T_n istatistiğinin asimptotik rastgele ifadesi,

$$T_n^* = \sum_{r=1}^m \lambda_r A_r, \quad A_r \sim \chi_{k-1}^2 \quad (4.95)$$

eşitliği ile verilen χ^2 -tipi bir karışımdır. Burada $\gamma(s, t)$ kovaryans fonksiyonunun bilinmeyen öz değerleri olan λ_r 'ler ($r = 1, 2, \dots, m$) hariç, T_n^* istatistiğinin dağılımı bilinmektedir. Çünkü $r = 1, 2, \dots, m$ için A_r 'ler bağımsız ve aynı dağılımlı rastgele değişkenlerdir. Öz değerleri λ_r ve kovaryans fonksiyonu $\gamma(s, t)$, yansız tahmin edicileri $\hat{\gamma}(s, t)$ ve $\hat{\lambda}_r$ ($r = 1, 2, \dots, \hat{m}$) ile tahmin edilmektedir. Burada \hat{m} , $\hat{\gamma}(s, t)$ 'nin pozitif özdeğerlerinin sayısıdır.

Örnekleme grupları Gaussian dağılımlı olmadığında ve örnek hacimleri (n_1, n_2, \dots, n_k) küçük olduğunda L_2 -norm tabanlı test, F -tipi test ve parametrik bootstrap yöntemleri tercih edilmez. Bu durumda L_2 -norm tabanlı bootstrap testi ve F -tipi bootstrap testi, T_n ve F_n istatistiklerinin bootstrap kritik değerlerini elde etmek için kullanılabilir (Zhang, 2013c).

$v_{ij}^*(t)$, $j = 1, 2, \dots, n_i$; $i = 1, \dots, k$ bootstrap örnekleri rastgele olarak $\hat{v}_{ij}(t) = y_{ij}(t) - \hat{\eta}_i(t)$, $j = 1, 2, \dots, n_i$; $i = 1, 2, \dots, k$ şeklinde gösterilen tahmini konu-etki fonksiyonlarından türetilmiş olsun. Bu takdirde bootstrap yöntemi,

$$y_{ij}^*(t) = \hat{\eta}_i(t) + v_{ij}^*(t), \quad j = 1, 2, \dots, n_i; i = 1, 2, \dots, k \quad (4.96)$$

şeklinde ifade edilir.

Böylece k tane bootstrap örneklerine dayalı olarak k tane örnek grubunun ortalama fonksiyonları $\bar{y}_1^*(t), \bar{y}_2^*(t), \dots, \bar{y}_k^*(t)$, örnek genel ortalama fonksiyonu $\bar{y}_{..}^*(t)$ ve toplanmış örnek kovaryans fonksiyonu $\hat{\gamma}^*(s, t)$ hesaplanabilmektedir. Daha sonra yine bootstrap örneği kullanılarak $GAKT_n^*(t)$ ve $HKT_n^*(t)$ test istatistikleri sırasıyla,

$$GAKT_n^*(t) = \sum_{i=1}^k n_i \{ [\bar{y}_i^*(t) - \bar{y}_{..}^*(t)] - [\bar{y}_i(t) - \bar{y}_{..}(t)] \}^2 \quad (4.97)$$

$$HKT_n^*(t) = (n - k) \hat{\gamma}^*(t, t) \quad (4.98)$$

eşitlikleri ile hesaplanabilir.

Bu durumda L_2 -norm tabanlı bootstrap testi veya F -tipi bootstrap testi için test istatistikleri sırasıyla,

$$T_n^* = \int_T GAKT_n^*(t) dt \quad (4.99)$$

$$F_n^* = \frac{\int_T \frac{GAKT_n^*(t) dt}{k-1}}{\int_T \frac{HKT_n^*(t) dt}{n-k}} \quad (4.100)$$

şeklinde ifade edilir.

Bu süreç, T_n veya F_n istatistiklerinin örnekleme dağılımına ait $100(1 - \alpha)$ yüzdelerik değerini tahmin etmek için kullanılacak olan T_n^* veya F_n^* istatistiklerini elde etmek için birçok kez tekrarlanır.

5. BULGULAR VE TARTIŞMA

Bu çalışmada Türkiye'nin 81 ili için, Meteoroloji Genel Müdürlüğü'nden temin edilen bazı iklimsel değişkenlere ait 2012-2018 yılları arası günlük verilerden oluşturulan fonksiyonel veriler, FVA teknikleri ile incelenmiştir. Söz konusu iklim değişkenleri şunlardır:

$$X_1(t) = \text{Günlük ortalama sıcaklık } (^{\circ}\text{C})$$

$$X_2(t) = \text{Günlük toplam yağış miktarı } (kg/m^2)$$

$$X_3(t) = \text{Günlük sıcaklık değişimi } (^{\circ}\text{C})$$

$$X_4(t) = \text{Günlük buharlaşma miktarı } (mm)$$

$$X_5(t) = \text{Günlük güneşlenme süresi } (saat)$$

Burada $t \in T$ olup, $T = [0,2557]$ 'dir.

Bu verilerin analiz edilmesinde kullanılan FVA teknikleri; Fourier baz fonksiyon yaklaşımı, pürüzlü ceza yaklaşımı, FTBA, fonksiyonel model tabanlı kümeleme ve FANOVA'dır. Bu veri analizi yöntemleri ile ilgili detaylı bilgiler önceki bölümlerde verilmiştir. Bu bölümde bahsedilen yöntemlere göre analizlerden elde edilen bulgular değerlendirilecektir.

Bu çalışmanın ilk kısmında, günlük ortalama sıcaklık, günlük toplam yağış miktarı, günlük sıcaklık değişimi, günlük buharlaşma miktarı ve günlük güneşlenme süresi verilerinin incelenmesi gereken baz fonksiyon sayısı belirlenmiştir. Bu kapsamda verileri açıklaması beklenen baz fonksiyon sayısı 365 olarak tespit edilmiştir. 81 il için 365 ayırık noktada ölçülen çok değişkenli verilere ait fonksiyonların oluşturulmasında Fourier baz fonksiyon ve pürüzlü ceza yaklaşımı kullanılmıştır. Bu yaklaşımlar sayesinde her değişken için 81 adet bireysel fonksiyon oluşturulmuştur. Verilere ait düzgülleştirme parametrelerinin belirlenmesinde GCV kullanılmıştır. Oluşturulan fonksiyonlara FTBA uygulanmış ve FTBA'ya göre değişkenlerin varyans açıklama oranı ile ana bileşen skorları bulunmuştur. Buna paralel olarak veriler için değişkenliğin neden kaynaklandığını gösteren grafikler elde edilmiş ve ana bileşen skorları ile desteklenmiştir. Fonksiyonel model tabanlı kümeleme yöntemi kullanılarak, çok değişkenli fonksiyonel veriler yardımıyla illerin

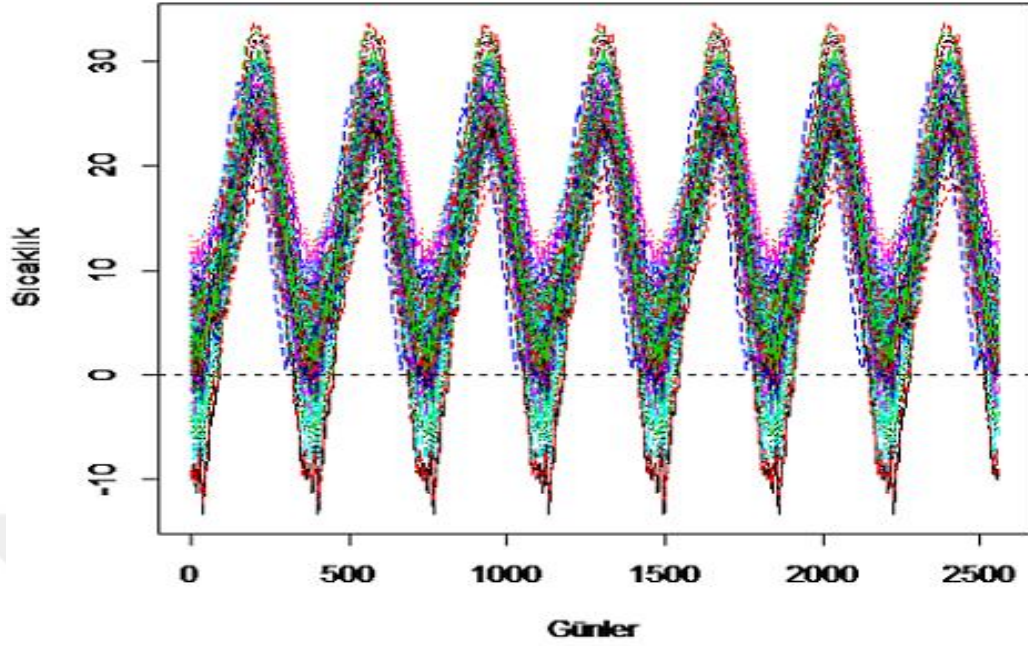
kümelenme yapısı belirlenmiştir. Her bir fonksiyonel veri seti için FANOVA kullanılarak, kümelenen fonksiyonların ortalamaları bakımından anlamlı bir farklılık olup olmadığı test edilmiştir. Bu uygulamalar, R programı kullanılarak yapılmıştır. Analizde `fda`, `fda.usc`, `funHDDC`, `fdANOVA` paketleri kullanılmıştır (URL-1; URL-2; URL-3; URL-4).

5.1. Değişkenlerin Bireysel Fonksiyonlarının İncelenmesi

Çalışmanın bu kısmında, ele alınan beş iklimsel değişkenle ilgili olarak bireysel fonksiyonlar oluşturulmuş ve değerlendirilmiştir. Söz konusu bireysel fonksiyonların oluşturulmasında Fourier baz fonksiyonları ve pürüzlü ceza yaklaşımı kullanılmıştır. Beş iklimsel değişkenin her biri için 81 tane bireysel fonksiyon elde edilmiştir. Elde edilen fonksiyonların şekilsel gösterimleri incelenerek, bireysel davranışları hakkında yorum yapılmıştır. Şekillere göre fonksiyonların sinüsoidal bir yapı gösterdiklerini söylemek mümkündür.

Günlük ortalama sıcaklık değişkenine ($X_1(t)$) ait bireysel fonksiyonlar Şekil 5.1'de gösterilmektedir. Şekil 5.1'de en yüksek sıcaklıkların yaz aylarına ve en düşük sıcaklıkların ise kış aylarına denk geldiği görülmektedir. Yaz aylarında sıcaklıklar tepe noktasına ulaştıktan sonra azalmaktadır. Kış aylarında ise sıcaklıklar dip noktasına ulaştıktan sonra artmaktadır. Yaz aylarında sıcaklıkları pozitif yönde yukarı çeken uç değerlerin, kış aylarında ise sıcaklıkları negatif yönde aşağı çeken uç değerlerin mevcut olduğu görülmektedir.

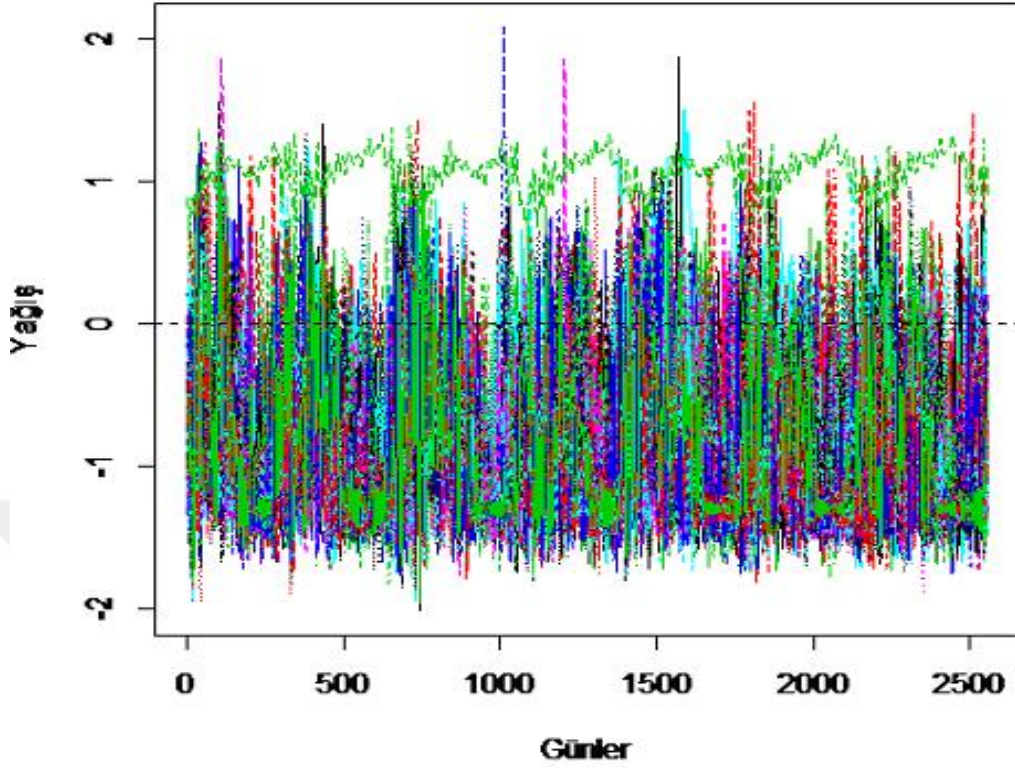
Sıcaklık için Bireysel Fonksiyonlar



Şekil 5. 1. Günlük Ortalama Sıcaklık için 81 Bireysel Fonksiyon

Günlük toplam yağış miktarına ($X_2(t)$) ait bireysel fonksiyonlar Şekil 5.2’de verilmiştir. Şekil 5.2’de en yüksek yağış miktarının kış aylarına ve en düşük yağış miktarının ise yaz aylarına denk geldiği görülmektedir. Kış aylarında yağış miktarı tepe noktasına ulaştıktan sonra azalmaktadır. Yaz aylarında ise yağış miktarı dip noktasına ulaştıktan sonra artmaktadır. Kış aylarında yağış miktarını pozitif yönde yukarı çeken uç değerlerin, yaz aylarında ise yağış miktarını negatif yönde aşağı çeken uç değerlerin mevcut olduğu görülmektedir.

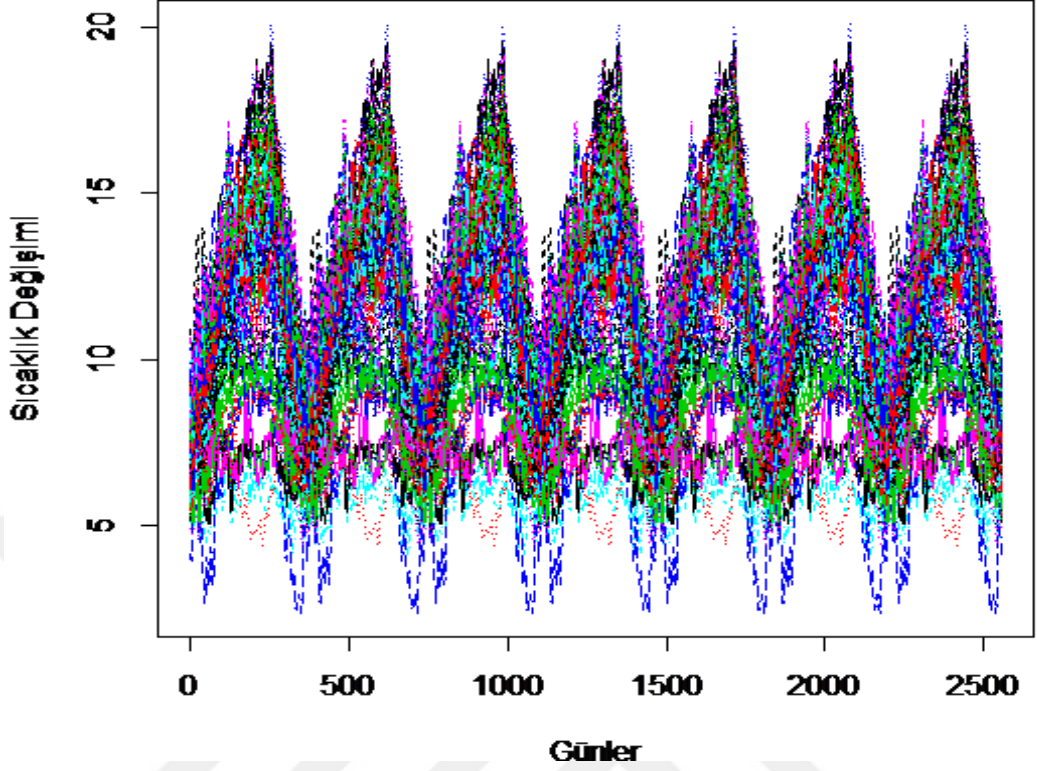
Yağış için Bireysel Fonksiyonlar



Şekil 5. 2. Günlük Toplam Yağış Miktarı için 81 Bireysel Fonksiyon

Günlük sıcaklık değişimine ($X_3(t)$) ait bireysel fonksiyonlar Şekil 5.3'de gösterilmiştir. Şekil 5.3'de en yüksek sıcaklık değişim değerlerinin yaz aylarına ve en düşük sıcaklık değişim değerlerinin ise kış aylarına denk geldiği görülmektedir. Yaz aylarında sıcaklık değişim değerleri tepe noktasına ulaştıktan sonra azalmaktadır. Kış aylarında ise sıcaklık değişim değerleri dip noktasına ulaştıktan sonra artmaktadır. Yaz aylarında sıcaklık değişim değerlerini pozitif yönde yukarı çeken uç değerlerin, kış aylarında ise sıcaklık değişim değerlerini negatif yönde aşağı çeken uç değerlerin mevcut olduğu görülmektedir.

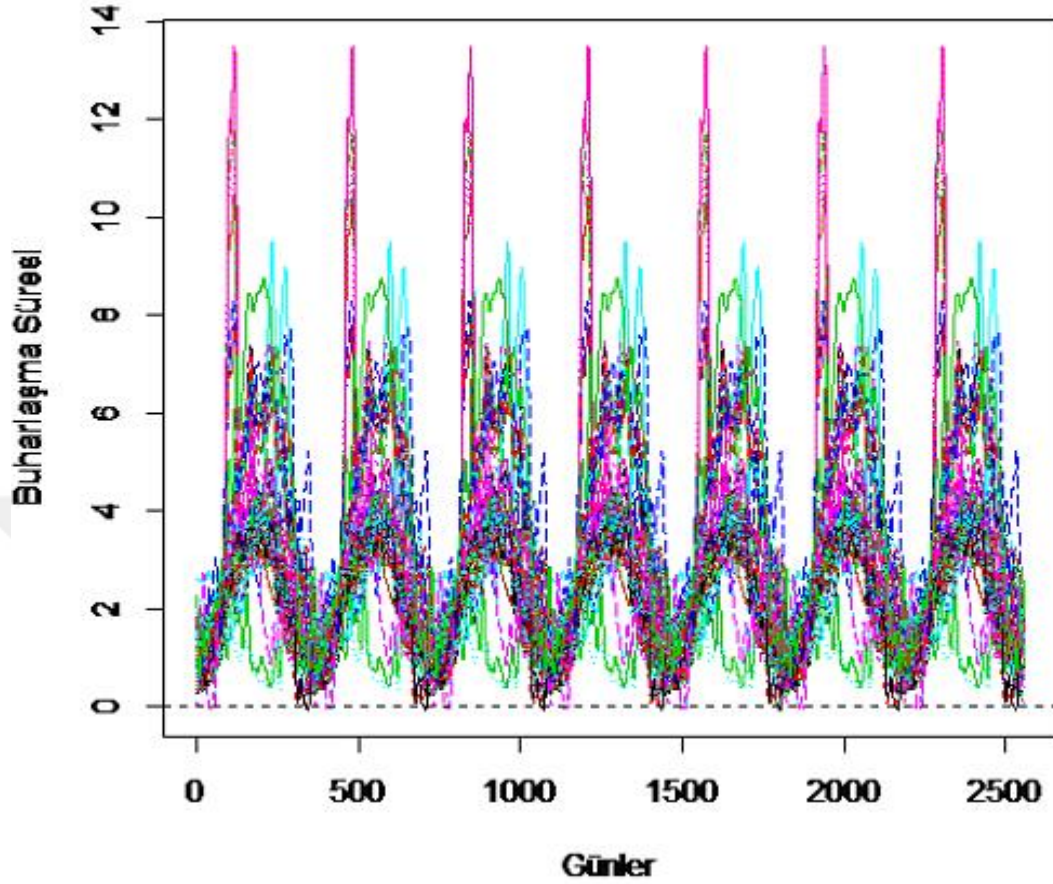
Sıcaklık Değişimi için Bireysel Fonksiyonlar



Şekil 5. 3. Günlük Sıcaklık Değişimi için 81 Bireysel Fonksiyon

Günlük buharlaşma miktarına ($X_4(t)$) ait bireysel fonksiyonlar Şekil 5.4'de verilmiştir. Şekil 5.4'de en yüksek buharlaşma miktarının yaz aylarına ve en düşük buharlaşma miktarının ise kış aylarına denk geldiği görülmektedir. Yaz aylarında buharlaşma miktarı tepe noktasına ulaştıktan sonra azalmaktadır. Kış aylarında ise buharlaşma miktarı dip noktasına ulaştıktan sonra artmaktadır. Yaz aylarında buharlaşma miktarını pozitif yönde yukarı çeken uç değerlerin, kış aylarında ise buharlaşma miktarını negatif yönde aşağı çeken uç değerlerin mevcut olduğu görülmektedir.

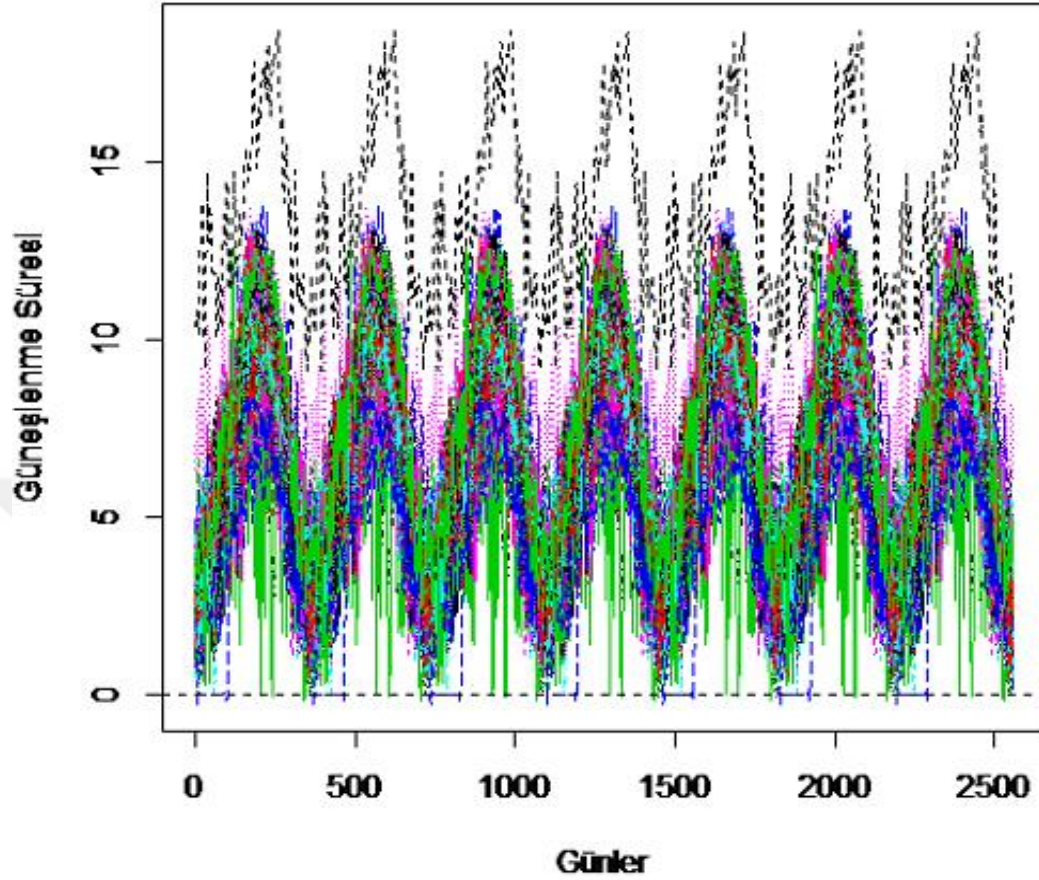
Buharlařma iin Bireysel Fonksiyonlar



řekil 5. 4. Günlük Buharlařma Miktarı iin 81 Bireysel Fonksiyon

Günlük güneřlenme süresine ($X_5(t)$) ait bireysel fonksiyonlar řekil 5.5'de gösterilmektedir. řekil 5.5'de en fazla güneřlenme süresinin yaz aylarına ve en az güneřlenme süresinin ise kış aylarına denk geldiđi görülmektedir. Yaz aylarında güneřlenme süreleri tepe noktasına ulařtıktan sonra azalmaktadır. Kış aylarında ise güneřlenme süreleri dip noktasına ulařtıktan sonra artmaktadır. Yaz aylarında güneřlenme sürelerini pozitif yönde yukarı eken uç deđerlerin, kış aylarında ise güneřlenme sürelerini negatif yönde ařađı eken uç deđerlerin mevcut olduđu görülmektedir.

Güneşlenme için Bireysel Fonksiyonlar

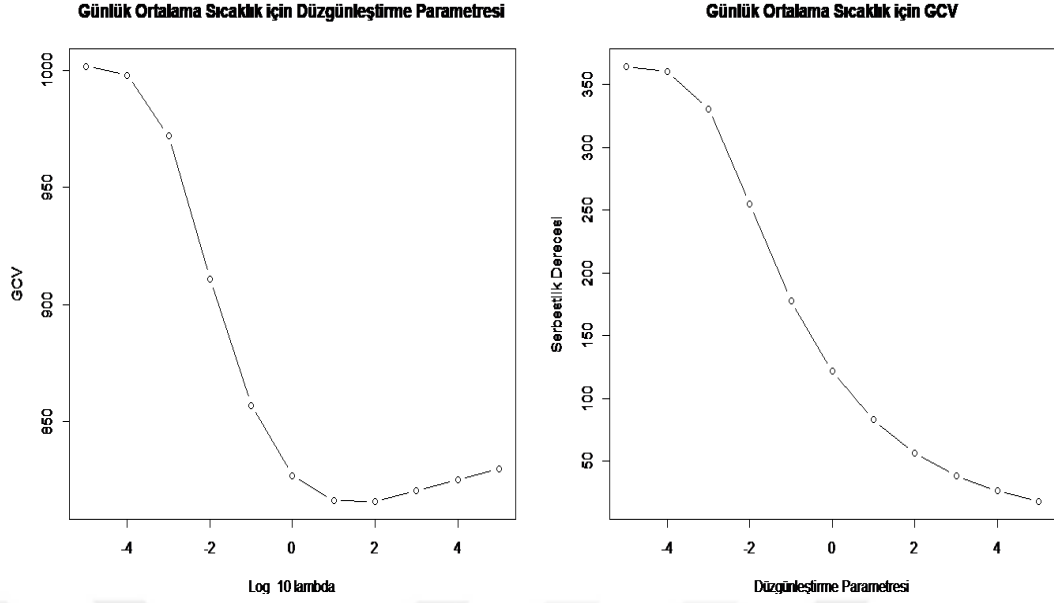


Şekil 5. 5. Günlük Güneşlenme Süresi için 81 Bireysel Fonksiyon

5.2. Değişkenlerin Simülasyon ile Düzgünleştirme Parametrelerinin Belirlenmesi

Simülasyon çalışması, kullanılan iklimsel değişkenler için düzgünleştirme parametrelerinin belirlenmesi amacıyla yapılmıştır. Her bir iklimsel değişken için ayrı ayrı uygulanan simülasyon çalışmaları sonucunda elde edilen düzgünleştirme parametreleri Şekil 5.6-5.10'da ve simülasyon sonuçları ise Çizelge 5.1-5.5'de verilmiştir. Düzgünleştirme parametresi belirlenirken 365 baz fonksiyonu kullanılmıştır.

Şekil 5.6 günlük ortalama sıcaklık verilerine ait düzgünleştirme parametresinin belirlenebilmesi için yapılan simülasyon çalışmasına ait grafikleri, Çizelge 5.1 ise simülasyon sonuçlarını göstermektedir. Çalışmada $[-5,5]$ değerleri için simülasyon yapılmış ve minimum GCV değeri düzgünleştirme parametresi olarak belirlenmiştir. Bu değer günlük ortalama sıcaklık verileri için 2 olarak alınmıştır.

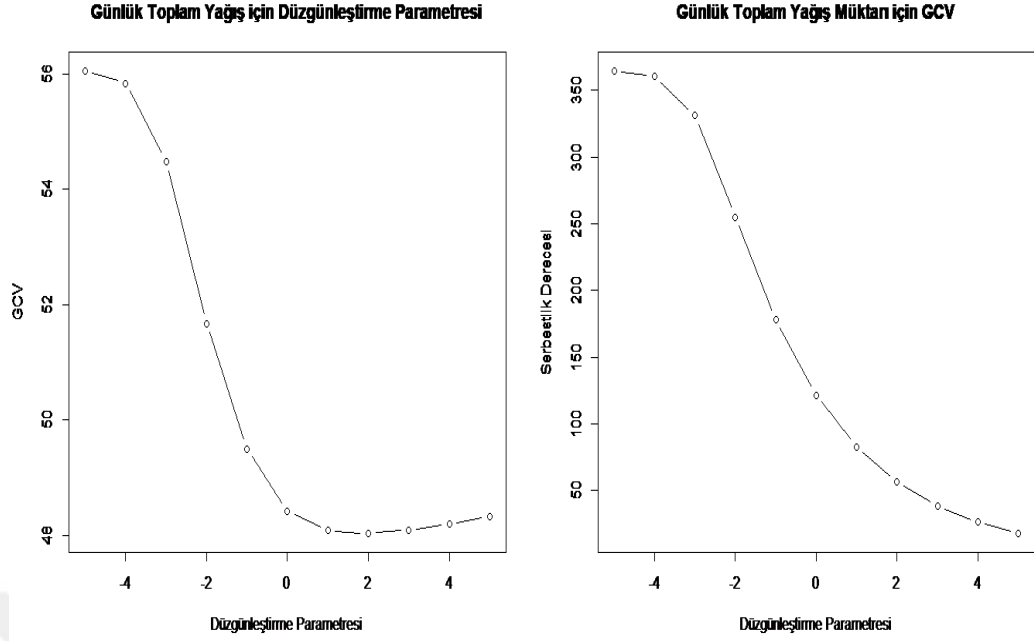


Şekil 5. 6. Günlük Ortalama Sıcaklık için GCV Simülasyonu

Çizelge 5. 1. Günlük Ortalama Sıcaklık için Simülasyon Sonuçları

Değerler	Serbestlik Derecesi	GCV
-5	364,5013	1001,76
-4	360,231	997,8755
-3	330,8371	971,9769
-2	254,8712	910,8631
-1	177,8344	857,1082
0	121,6026	826,9709
1	82,90069	816,5622
2	56,49807	816,0434
3	38,51261	820,5284
4	26,26832	825,4053
5	17,9403	829,75

Şekil 5.7 günlük toplam yağış miktarı verilerine ait düzgünleştirme parametresinin belirlenebilmesi için yapılan simülasyon çalışmasına ait grafikleri, Çizelge 5.2 ise simülasyon sonuçlarını göstermektedir. Çalışmada simülasyon aralığı [-5,5] olarak alınmış ve minimum GCV değeri düzgünleştirme parametresi olarak belirlenmiştir. Bu değer günlük toplam yağış verileri için 2 olarak elde edilmiştir.

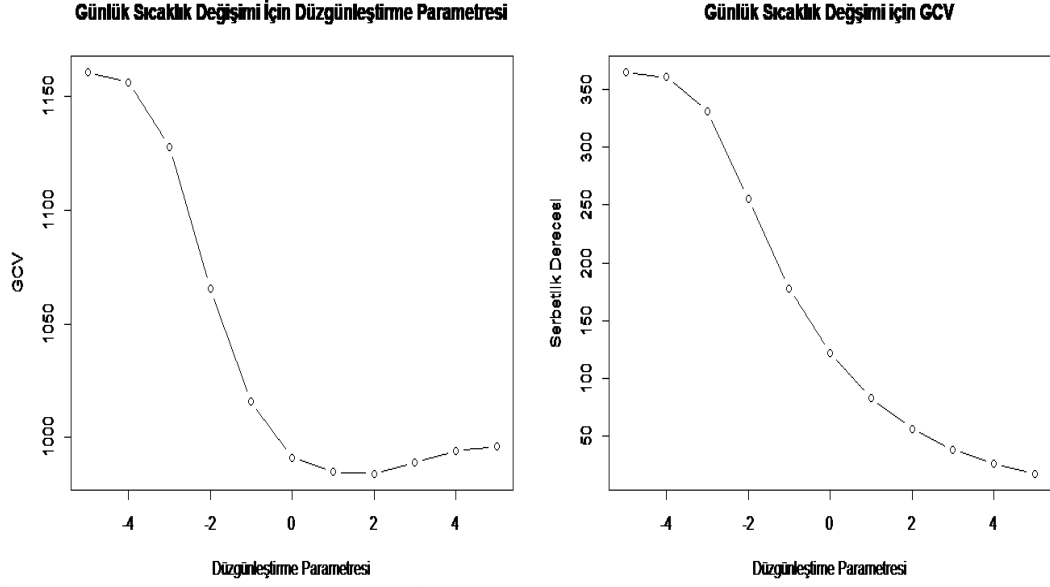


Şekil 5. 7. Günlük Toplam Yağış için GCV Simülasyonu

Çizelge 5. 2. Günlük Toplam Yağış için Simülasyon Sonuçları

Değerler	Serbestlik Derecesi	GCV
-5	364,5013	56,05391
-4	360,231	55,83885
-3	330,8371	54,4874
-2	254,8712	51,67523
-1	177,8344	49,49375
0	121,6026	48,41473
1	82,90069	48,08578
2	56,49807	48,03698
3	38,51261	48,09322
4	26,26832	48,19835
5	17,9403	48,33542

Şekil 5.8 günlük sıcaklık değişimi verilerine ait düzgünleştirme parametresinin belirlenebilmesi için yapılan simülasyon çalışmasına ait grafikleri, Çizelge 5.3 ise simülasyon sonuçlarını göstermektedir. Simülasyon aralığı [-5,5] alınmış olup Şekil 5.8'e göre minimum GCV değeri düzgünleştirme parametresi olarak alınmıştır. Bu değer günlük sıcaklık değişimi verileri için 2 olarak belirlenmiştir.

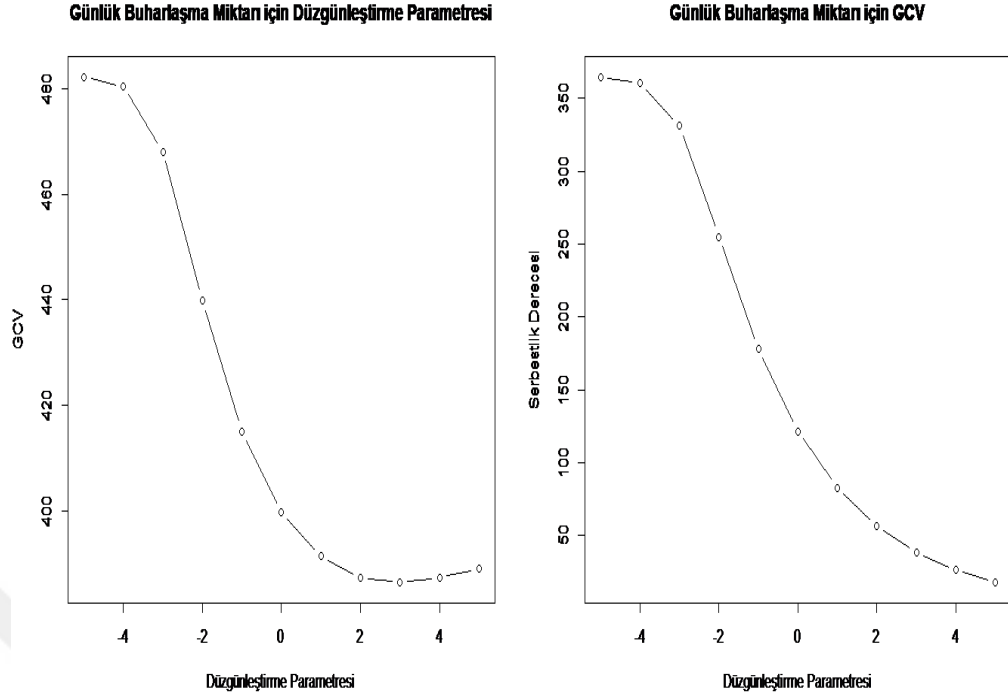


Şekil 5. 8. Günlük Sıcaklık Değişimi için GCV Simülasyonu

Çizelge 5. 3. Günlük Sıcaklık Değişimi için Simülasyon Sonuçları

Değerler	Serbestlik Derecesi	GCV
-5	364,5013	1160,777
-4	360,231	1156,309
-3	330,8371	1127,672
-2	254,8712	1065,507
-1	177,8344	1015,771
0	121,6026	991,0959
1	82,90069	984,7931
2	56,49807	984,1055
3	38,51261	988,8072
4	26,26832	993,9249
5	17,9403	996,1273

Şekil 5.9 günlük buharlaşma süresi verilerine ait düzgünleştirme parametresinin belirlenebilmesi için simülasyon çalışmasına ait grafikleri, Çizelge 5.4 ise simülasyon sonuçlarını göstermektedir. Simülasyon aralığı [-5,5] alınmış olup Şekil 5.9'a göre minimum GCV değeri düzgünleştirme parametresi olarak alınmıştır. Bu değer günlük sıcaklık değişimi verileri için 3 olarak belirlenmiştir.

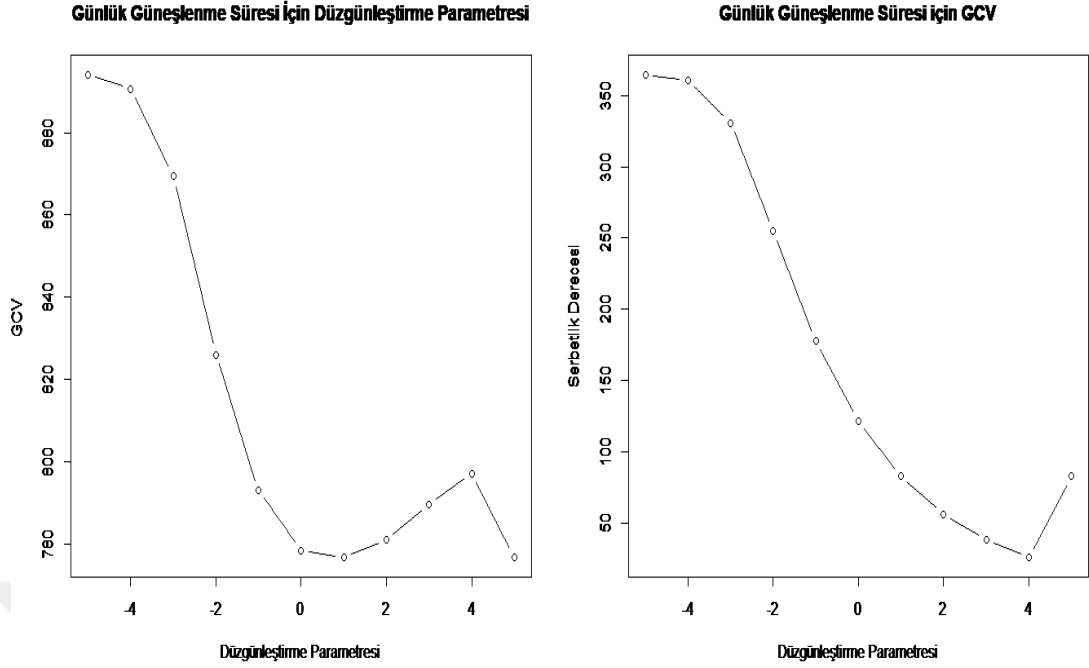


Şekil 5. 9. Günlük Buharlaşma Miktarı için GCV Simülasyonu

Çizelge 5. 4. Günlük Buharlaşma Miktarı için Simülasyon Sonuçları

Değerler	Serbestlik Derecesi	GCV
-5	364,5013	482,1474
-4	360,231	480,2832
-3	330,8371	468,0453
-2	254,8712	439,8923
-1	177,8344	415,0447
0	121,6026	399,7455
1	82,90069	391,4655
2	56,49807	387,3497
3	38,51261	386,5378
4	26,26832	387,4616
5	17,9403	389,0577

Şekil 5.10 günlük güneşlenme süresi verilerine ait düzgünleştirme parametresinin belirlenebilmesi için simülasyon çalışmasına ait grafikleri, Çizelge 5.5 ise simülasyon sonuçlarını göstermektedir. Simülasyon aralığı $[-5,5]$ alınmış olup, minimum GCV değeri düzgünleştirme parametresi olarak belirlenmiştir. Bu değer günlük güneşlenme süresi verileri için 1 olarak belirlenmiştir.



Şekil 5. 10. Günlük Güneşlenme Süresi için GCV Simülasyonu

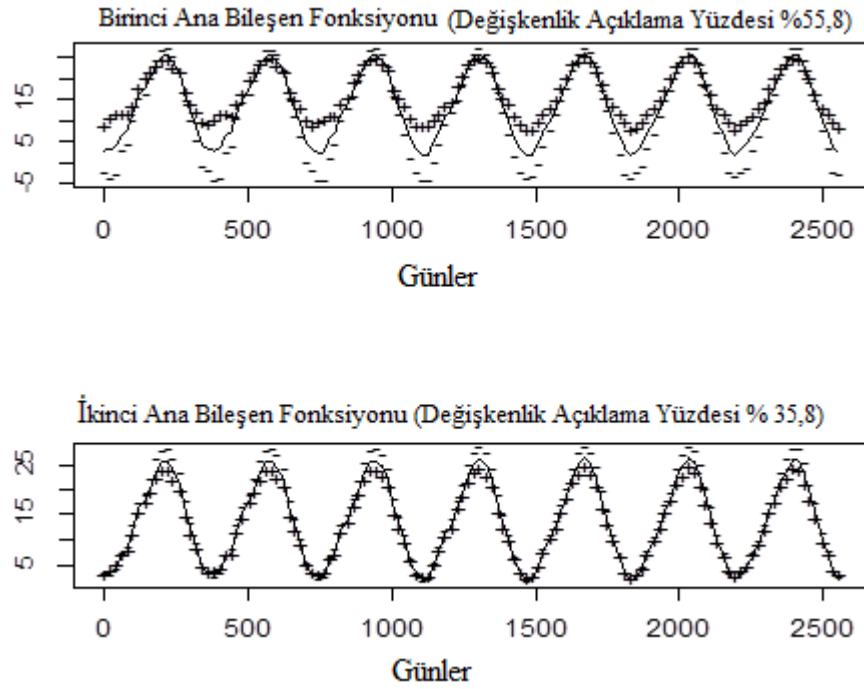
Çizelge 5. 5. Günlük Güneşlenme Süresi için Simülasyon Sonuçları

Değerler	Serbestlik Derecesi	GCV
-5	364,5013	894,0939
-4	360,231	890,6692
-3	330,8371	869,3652
-2	254,8712	826,0291
-1	177,8344	792,9708
0	121,6026	778,4261
1	82,90069	776,7679
2	56,49807	781,1089
3	38,51261	789,639
4	26,26832	797,0927
5	82,90069	776,7679

5.3. İllere Göre İklimsel Verilerin Fonksiyonel Temel Bileşen Analizi ile Değerlendirilmesi

İklimsel değişkenlere ait verilerdeki değişkenliğin sebebini araştırmak için FTBA kullanılmıştır. Temel bileşenler için oluşturulan ortalama fonksiyonlarını temsil eden eğriler etrafındaki (+) ve (-) işaretler, bu ortalama fonksiyonların özdeğerlerinin eklenmesinin ve çıkarılmasının etkisini göstermektedir. Bu işaretler ortalama fonksiyonlardan ne kadar uzaklaşırsa, sapmalar da o kadar yüksek olmaktadır. Ortalamadan sapmaların yüksek olması değişkenliğin sebebini belirlemede kullanılmaktadır. İklimsel değişkenlere göre ilk iki temel bileşen fonksiyonları üzerinde sapmaların belirlenmesi ve bu temel bileşenlerin değişkenlik açıklama oranları Şekil 5.11-5.15’de, FTBA için skorlar ise Çizelge 5.6-5.10’da verilmiştir.

Şekil 5.11’de günlük ortalama sıcaklık verileri için, birinci temel bileşen fonksiyonu üzerinde ortalama sapmaların kış aylarına rastladığı gözlemlenmektedir. Başka bir ifadeyle %55,8 değişkenlik açıklama gücüyle günlük ortalama sıcaklık açısından, fonksiyonlar arasındaki birinci temel değişkenliğin kış aylarından kaynaklandığı görülmektedir. İkinci temel bileşen fonksiyonu üzerinde ortalama sapmaların yaz aylarına rastladığı gözlemlenmektedir. Yani %35,8 değişkenlik açıklama gücüyle günlük ortalama sıcaklık açısından, fonksiyonlar arasındaki temel değişkenliğin yaz aylarından kaynaklandığı görülmektedir.



Şekil 5. 11. Günlük Ortalama Sıcaklık için Temel Bileşen Fonksiyonları

Çizelge 5.6 incelendiğinde temel bileşen skorları %55,8 değişkenlik açıklayabilme gücüyle birinci temel bileşen için değerlendirildiğinde Elazığ, Giresun, Ordu, Rize, Samsun, Sinop, Zonguldak, Bursa, Kocaeli, Sakarya, Yalova, Balıkesir, Tekirdağ, İstanbul, Aydın, Denizli, İzmir, Manisa, Muğla, Adana, Antalya, Hatay, Mersin, Kahramanmaraş, Osmaniye, Adıyaman, Kilis ve Şanlıurfa illerinin birinci temel bileşen üzerinde etkili olduğu ve yüksek temel bileşen skoruna sahip oldukları görülmektedir. Değişkenliğin sebebinin kış ayları olduğunu gösteren birinci temel bileşen için veriler incelendiğinde kış aylarındaki en yüksek sıcaklıkların bu illere ait olduğu görülmektedir. İkinci temel bileşen açısından değerlendirildiğinde Ağrı, Ardahan, Bayburt, Erzurum, Kars, Elazığ, Bitlis, Tunceli, Artvin, Gümüşhane, Trabzon, Bartın, Çankırı, Çorum, Kastamonu, Tokat, Bolu, Eskişehir, Afyon, Uşak,

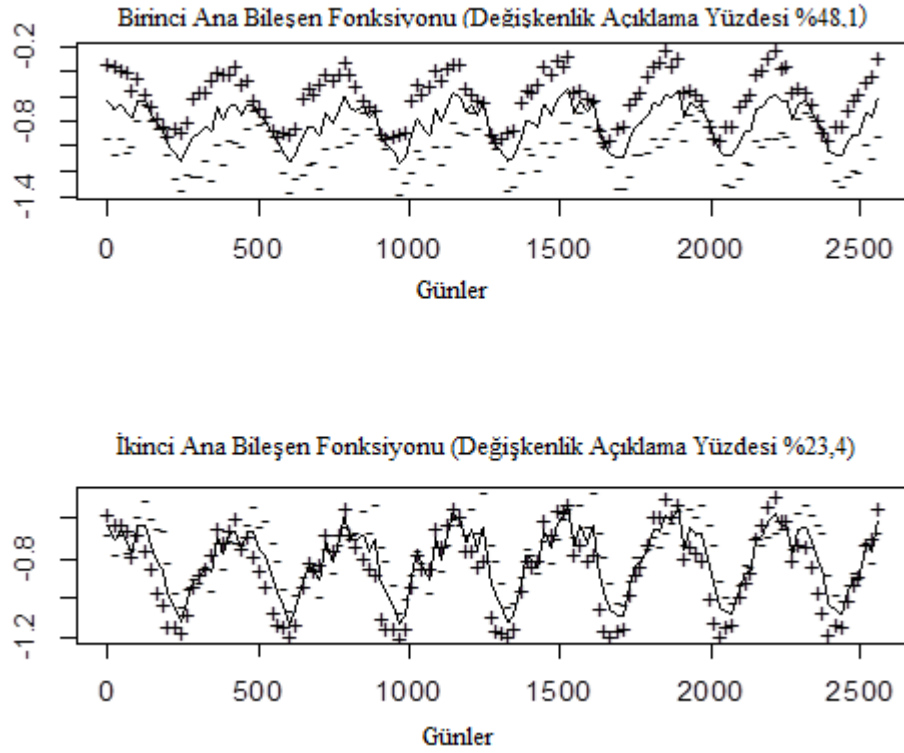
Kayseri ve Nevşehir illerinin etkili olduğu ve yüksek bir temel bileşen skoruna sahip oldukları görülmektedir. Değişkenliğin %35,8'ini açıklama gücüne sahip ikinci temel bileşen için veriler incelendiğinde, yaz aylarındaki en yüksek sıcaklıkların bu illere ait olduğu görülmektedir.

Çizelge 5. 6. Günlük Ortalama Sıcaklık için Temel Bileşen Skorları

İller	Birinci Temel Bileşen Skorları	İkinci Temel Bileşen Skorları
Ağrı	-371,333	148,4286
Ardahan	-370,082	286,8941
Bayburt	-235,328	185,0184
Erzincan	-118,618	26,72335
Erzurum	-283,026	181,0255
İğdır	-109,914	-51,1457
Kars	-353,949	244,3627
Bingöl	-109,991	-35,3781
Bitlis	-246,007	122,4029
Elazığ	83,89337	120,4828
Hakkari	-171,529	12,37877
Malatya	-30,9869	-97,1962
Muş	-238,327	-2,57662
Tunceli	-353,949	244,3627
Van	-118,618	26,72335
Artvin	10,64114	77,32119
Giresun	138,4644	-2,56892
Gümüşhane	-118,228	133,1189
Ordu	128,1139	-0,13545
Rize	110,3657	-3,40667
Trabzon	10,64114	77,32119
Amasya	28,70938	-7,12093
Bartın	30,36894	61,55894
Çankırı	-92,5375	69,80659
Çorum	-83,9435	117,8886
Karabük	20,22826	8,709424
Kastamonu	-101,748	152,0578
Samsun	131,6901	3,075403
Sinop	108,0111	22,16805
Tokat	-83,9435	117,8886
Zonguldak	91,75443	54,77748
Bilecik	12,50586	43,09911
Bolu	-47,6462	131,7843
Bursa	89,96479	-24,5967
Düzce	43,22716	50,88832
Eskişehir	-73,6528	102,3419
Kocaeli	122,8035	-26,9902
Sakarya	126,126	-20,3757
Yalova	113,9357	-9,86661
Çanakkale	114,7512	-49,4796
Edirne	30,9974	-24,3049
Kırklareli	30,45616	-5,79129
Balıkesir	84,45016	-26,7721
Tekirdağ	90,44902	-19,186
İstanbul	124,9351	-47,3969
Afyonkarahisar	-55,9553	70,96115
Aydın	175,4692	-144,356
Denizli	118,2738	-107,789
İzmir	197,1292	-141,19
Kütahya	-56,4287	101,8481

Manisa	119,0317	-122,094
Muğla	244,5682	-177,894
Uşak	-55,9553	70,96115
Ankara	-39,1921	23,35903
Karaman	-39,4175	30,3757
Konya	-33,2098	8,609689
Aksaray	-34,0529	11,88015
Kayseri	-95,9876	78,74226
Kırıkkale	-57,5819	15,92039
Kırşehir	-76,1034	43,3476
Nevşehir	-73,016	87,70753
Niğde	-66,1804	59,20416
Sivas	-143,743	128,75
Yozgat	-123,636	155,6946
Adana	249,9168	-168,381
Antalya	273,615	-179,62
Burdur	1,749526	-11,0119
Hatay	316,3474	-206,622
Isparta	-32,2873	48,63957
Mersin	220,4405	-140,802
Kahramanmaraş	107,8594	-164,281
Osmaniye	215,5446	-157,656
Adıyaman	91,49265	-206,379
Batman	10,92073	-162,395
Diyarbakır	55,45064	-183,465
Gaziantep	50,86642	-127,096
Kilis	131,2746	-159,831
Mardin	53,11091	-175,907
Siirt	35,8052	-186,941
Şanlıurfa	139,0481	-254,377

Şekil 5.12’de günlük toplam yağış verileri için, birinci temel bileşen fonksiyonu üzerinde ortalama sapmaların kış aylarına rastladığı gözlemlenmektedir. Başka bir ifadeyle %48,1 değişkenlik açıklama gücüyle günlük toplam yağış açısından, fonksiyonlar arasındaki birinci temel değişkenliğin kış aylarından kaynaklandığı görülmektedir. İkinci temel bileşen fonksiyonu üzerinde ortalama sapmaların ilkbahar aylarına rastladığı gözlemlenmektedir. Yani %23,4 değişkenlik açıklama gücüyle günlük toplam yağış açısından, fonksiyonlar arasındaki ikinci temel değişkenliğin yaz aylarından kaynaklandığı görülmektedir.



Şekil 5. 12. Günlük Toplam Yağış için Temel Bileşen Fonksiyonları

Çizelge 5.7 incelendiğinde temel bileşen skorları, %48,1 değişkenlik açıklayabilme gücüyle birinci temel bileşen için değerlendirildiğinde, özellikle Diyarbakır ilinin birinci temel bileşen üzerinde etkili olduğu ve yüksek ana bileşen skoruna sahip olduğu görülmektedir. Değişkenliğin sebebinin kış ayları olduğunu gösteren birinci temel bileşen için veriler incelendiğinde, kış aylarındaki en yüksek yağışın bu ile ait olduğu görülmektedir. İkinci temel bileşen açısından değerlendirildiğinde, özellikle Doğu ve Güneydoğu Anadolu Bölgesinde yer alan illerin etkili olduğu ve yüksek bir ana bileşen skoruna sahip oldukları görülmektedir. Değişkenliğin %23,4'ünü açıklama gücüne sahip ikinci temel bileşen için veriler incelendiğinde yaz aylarında en düşük yağışların bu illere ait olduğu görülmektedir.

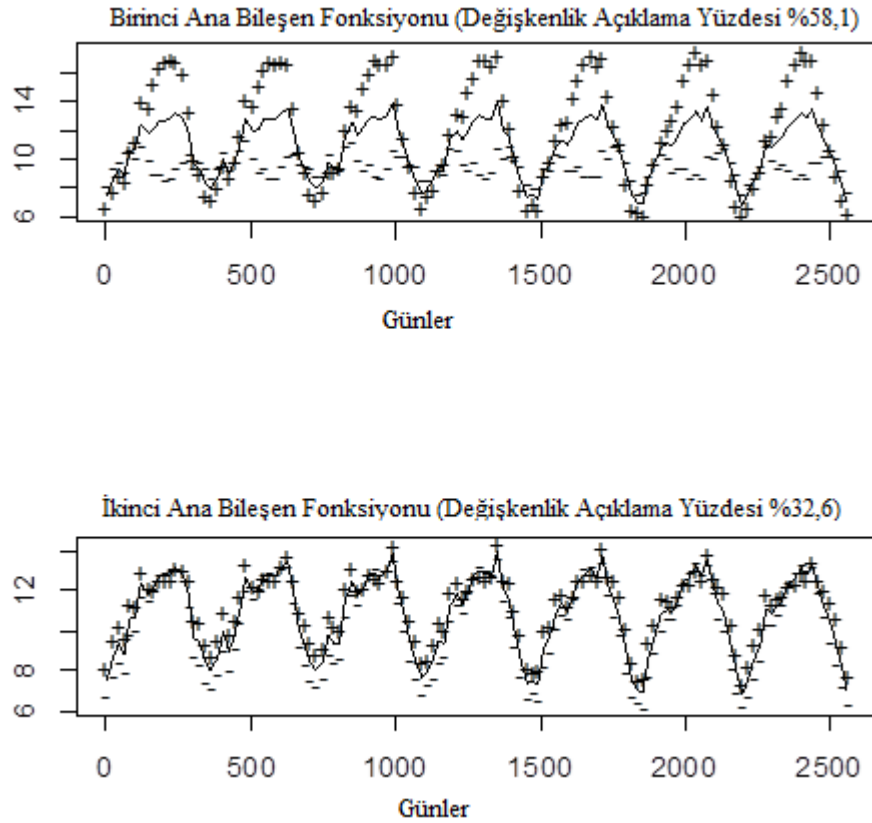
Çizelge 5. 7. Günlük Toplam Yağış için Temel Bileşen Skorları

İller	Birinci Temel Bileşen Skorları	İkinci Temel Bileşen Skorları
Ağrı	-0,43946	-2,771
Ardahan	-1,18358	-15,0178
Bayburt	-3,09364	-4,06381
Erzincan	-3,49074	-0,21725
Erzurum	-6,1721	-7,51391
Iğdır	-8,89437	-2,01308
Kars	-3,14883	-8,36653
Bingöl	2,720144	5,758561
Bitlis	12,31718	6,402775
Elazığ	-1,22877	8,10097
Hakkari	3,799882	3,236409
Malatya	-7,49299	2,34121

Muş	-2,45996	1,291885
Tunceli	1,342466	-1,06775
Van	-4,87235	0,814713
Artvin	7,189518	-5,02915
Giresun	15,95214	-5,4996
Gümüşhane	-2,02422	-1,00607
Ordu	-2,47157	1,308039
Rize	23,59193	-10,5261
Trabzon	8,133829	-7,57582
Amasya	-1,91617	1,906487
Bartın	11,92102	1,582296
Çankırı	-3,20887	-1,36107
Çorum	-4,03711	-1,04517
Karabük	1,068567	-3,87066
Kastamonu	-2,76572	-10,4634
Samsun	3,002592	-3,64882
Sinop	4,871589	-0,25374
Tokat	-6,07807	-4,35236
Zonguldak	8,900508	-0,10962
Bilecik	-1,82868	1,568992
Bolu	2,034784	-1,34195
Bursa	3,684358	3,226959
Düzce	7,093092	-1,07524
Eskişehir	-1,35037	-4,32328
Kocaeli	4,175584	-0,47916
Sakarya	5,24001	-2,54071
Yalova	0,617615	1,587756
Çanakkale	-1,78182	7,100262
Edirne	-0,18349	0,643754
Kırklareli	-4,67953	-1,33155
Balıkesir	3,535732	4,758666
Tekirdağ	-5,76117	1,464609
İstanbul	-0,13512	1,3697
Afyonkarahisar	-2,88426	0,773931
Aydın	-1,84479	8,706679
Denizli	-2,57596	4,46146
İzmir	-4,5749	5,196021
Kütahya	-1,35037	-4,32328
Manisa	-4,69902	2,602903
Muğla	-1,68908	7,587497
Uşak	-4,31823	-2,52774
Ankara	-3,02618	1,913735
Karaman	-8,93545	1,508425
Konya	-7,94157	-0,65778
Aksaray	-6,11138	3,278877
Kayseri	-6,08388	-0,57893
Kırıkkale	-6,82755	-1,87573
Kırşehir	-9,00697	-2,09218
Nevşehir	-6,68103	-0,35023
Niğde	-7,88556	-0,26102
Sivas	-5,05125	-2,56168
Yozgat	-2,82555	-3,88002
Adana	-1,64821	6,776662
Antalya	-1,89689	9,645151
Burdur	-4,05462	1,33127
Hatay	-1,64821	6,776662
Isparta	-3,49149	-3,03633
Mersin	-1,89689	9,645151
Kahramanmaraş	-3,03738	3,365169
Osmaniye	-1,40499	0,844147

Adıyaman	0,585116	9,31383
Batman	-1,74173	8,179005
Diyarbakır	81,00727	-49,8084
Gaziantep	-1,09289	9,219251
Kilis	-5,04881	6,753461
Mardin	-3,63196	4,563533
Siirt	-1,02233	2,163498
Şanlıurfa	-6,34311	6,669898
Şırnak	0,186226	3,077754

Şekil 5.13’de günlük sıcaklık değişimi için, birinci temel bileşen fonksiyonu üzerinde ortalama sapmaların ilkbahar ve yaz aylarına rastladığı gözlemlenmektedir. Başka bir ifadeyle %58,1 değişkenlik açıklama gücüyle günlük sıcaklık değişimi açısından, fonksiyonlar arasındaki birinci temel değişkenliğin ilkbahar ve yaz aylarından kaynaklandığı görülmektedir. İkinci temel bileşen fonksiyonu üzerinde ortalama sapmaların kış aylarına rastladığı gözlemlenmektedir. Yani %32,6 değişkenlik açıklama gücüyle günlük sıcaklık değişimi açısından, fonksiyonlar arasındaki temel değişkenliğin kış aylarından kaynaklandığı görülmektedir.



Şekil 5. 13. Günlük Sıcaklık Değişimi için Ana Bileşen Fonksiyonları

Çizelge 5.8 incelendiğinde, Ana Bileşen Skorları %58,1 değişkenlik açıklayabilme gücüyle birinci temel bileşen için değerlendirildiğinde; Ağrı, Ardahan, Bayburt, Kars, Erzincan, Erzurum, Bingöl, Elazığ, Muş, Tunceli, Çankırı, Çorum, Burdur, Karabük, Kastamonu ve Sivas illerinin birinci ana bileşen üzerinde etkili

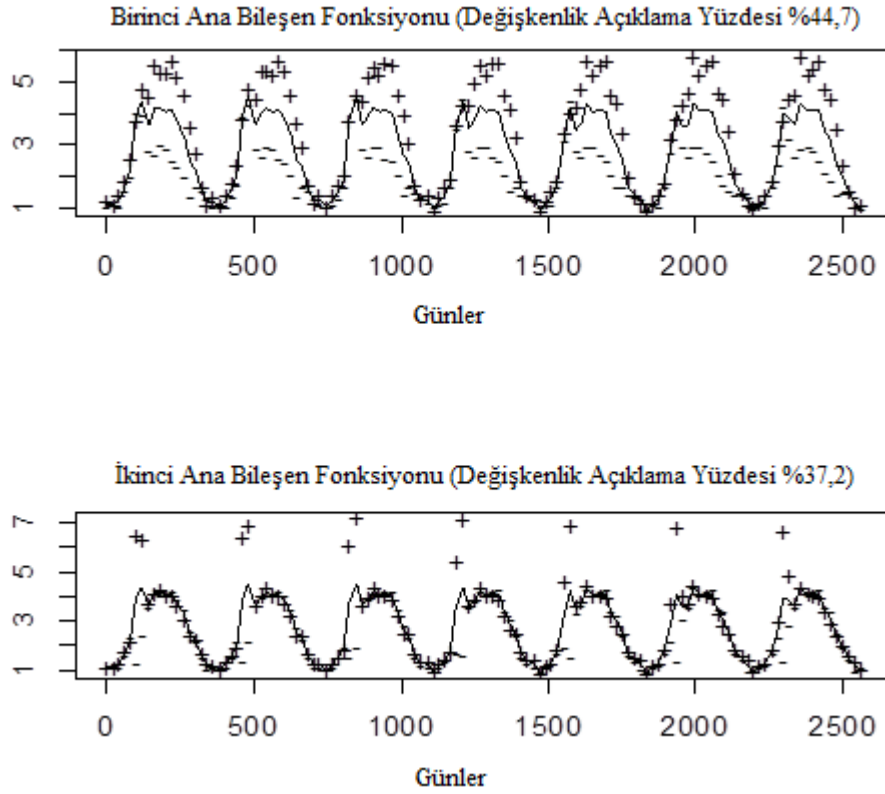
olduğu ve yüksek temel bileşen skoruna sahip oldukları görülmektedir. Değişkenliğin sebebinin ilkbahar ve yaz ayları olduğunu gösteren birinci temel bileşen için veriler incelendiğinde, yaz ve ilkbahar aylarındaki en yüksek sıcaklık değişiminin bu illere ait olduğu görülmektedir. İkinci temel bileşen açısından değerlendirildiğinde Ardahan, Kars, Çorum, Kayseri, Osmaniye, Mersin ve Isparta illerinin etkili olduğu ve yüksek bir temel bileşen skoruna sahip oldukları görülmektedir. Değişkenliğin %32,6'sını açıklama gücüne sahip ikinci temel bileşen için veriler incelendiğinde, kış aylarında en yüksek sıcaklık değişimlerinin bu illere ait olduğu görülmektedir.

Çizelge 5. 8. Günlük Sıcaklık Değişimi için Temel Bileşen Skorları

İller	Birinci Temel Bileşen Skorları	İkinci Temel Bileşen Skorları
Ağrı	139,2383	64,27275
Ardahan	73,09543	71,32479
Bayburt	117,3727	52,01881
Erzincan	96,69941	39,92822
Erzurum	73,94607	10,16191
İğdır	42,32168	45,01523
Kars	105,2438	96,45177
Bingöl	74,42523	24,04265
Bitlis	-2,69752	-48,9707
Elazığ	70,92827	26,31089
Hakkari	-23,5942	-55,5103
Malatya	36,77127	-22,0761
Muş	80,40727	-43,2595
Tunceli	129,1122	66,85772
Van	-0,25655	4,689204
Artvin	-103,634	-65,1879
Giresun	-218,582	-151,21
Gümüşhane	66,0743	33,38917
Ordu	-187,902	-112,451
Rize	-183,27	-87,7907
Trabzon	-183,27	-87,7907
Amasya	69,92126	46,40323
Bartın	-0,97386	60,16096
Çankırı	140,5331	75,511
Çorum	130,0062	84,54537
Karabük	91,23502	67,15505
Kastamonu	94,67842	56,13948
Samsun	-187,206	-111,715
Sinop	-206,883	-171,928
Tokat	46,90474	23,72161
Zonguldak	-178,251	-139,611
Bilecik	-6,98685	-42,0534
Bolu	45,20618	40,76695
Bursa	-4,69513	19,6236
Düzce	-20,5264	13,94009
Eskişehir	53,51338	50,85568
Kocaeli	-71,6165	-58,7829
Sakarya	-38,3039	-12,5481
Yalova	-89,3056	-98,5001
Çanakkale	-106,499	-197,881
Edirne	65,76327	8,801579

Kırklareli	10,11873	-26,2765
Balıkesir	-80,3863951	-8,27852085
Tekirdağ	-117,2106072	-80,20385261
İstanbul	-100,3695733	-73,32123615
Afyonkarahisar	47,3590232	23,50707138
Aydın	46,5693997	38,04324389
Denizli	30,2033614	18,36447173
İzmir	-84,3757444	-58,22950154
Kütahya	46,9942492	35,40471873
Manisa	4,5111216	5,68212971
Muğla	-42,260452	36,70883367
Uşak	48,8787715	30,95407235
Ankara	35,9508762	11,14745109
Karaman	59,4816691	58,63030793
Konya	-3,3710437	9,50098205
Aksaray	33,0748253	41,74783199
Kayseri	133,206483	103,3965336
Kırıkkale	53,5492015	26,94125594
Kırşehir	47,0880413	56,70432077
Nevşehir	24,73517	-18,53791137
Niğde	47,3400404	56,92617598
Sivas	101,4970982	59,43196434
Yozgat	10,8510462	8,48020595
Adana	-68,8377232	41,73251003
Antalya	-167,5845422	-61,75268187
Burdur	90,9004282	64,13766692
Hatay	-216,2999417	-74,76657255
Isparta	68,4115767	70,85081757
Mersin	-67,2116978	79,7562735
Kahramanmaraş	15,3426835	16,36310109
Osmaniye	-47,0223042	88,10341234
Adıyaman	21,8120758	-6,27885027
Batman	37,5848424	-22,42410353
Diyarbakır	53,6229993	39,06383032
Gaziantep	16,8238774	21,98337308
Kilis	25,734269	-33,17228925
Mardin	-95,2569757	-104,5463341
Siirt	9,1818995	-10,25047102
Şanlıurfa	-7,797184	-0,02751526
Şırnak	-86,0459605	-81,35996913

Şekil 5.14’de günlük buharlaşma miktarı için, birinci temel bileşen fonksiyonu üzerinde ortalama sapmaların ilkbahar ve yaz aylarına rastlandığı gözlemlenmektedir. Başka bir ifadeyle %44,7 değişkenlik açıklama gücüyle günlük buharlaşma miktarı açısından, fonksiyonlar arasındaki birinci temel değişkenliğin ilkbahar ve yaz aylarından kaynaklandığı görülmektedir. İkinci temel bileşen fonksiyonu üzerinde ortalama sapmaların ilkbahar aylarına rastlandığı gözlemlenmektedir. %37,2 değişkenlik açıklama gücüyle günlük buharlaşma miktarı açısından, fonksiyonlar arasındaki temel değişkenliğin ilkbahar aylarından kaynaklandığı anlamına gelmektedir.



Şekil 5. 14. Günlük Buharlaşma Miktarı için Temel Bileşen Fonksiyonları

Çizelge 5.9 incelendiğinde, temel bileşen skorları %44,7 değişkenlik açıklayabilme gücüyle birinci temel bileşen için değerlendirildiğinde Malatya, Burdur, Mersin, Diyarbakır, Gaziantep, Mardin, Kilis, Siirt, Şanlıurfa ve Şırnak illerinin birinci temel bileşen üzerinde etkili olduğu ve yüksek temel bileşen skoruna sahip oldukları görülmektedir. Değişkenliğin sebebinin ilkbahar ve yaz ayları olduğunu gösteren birinci temel bileşen için veriler incelendiğinde, ilkbahar ve yaz aylarındaki en yüksek buharlaşmaların bu illere ait olduğu görülmektedir. İkinci temel bileşen açısından değerlendirildiğinde Gümüşhane, Karabük, Bilecik, Bolu, Bursa, Yalova, Antalya, Hatay ve Adana illerinin etkili olduğu ve yüksek bir temel bileşen skoruna sahip oldukları görülmektedir. Değişkenliğin %37,2'sini açıklama gücüne sahip ikinci temel bileşen için veriler incelendiğinde, ilkbahar aylarında en yüksek buharlaşmaların bu illere ait olduğu görülmektedir.

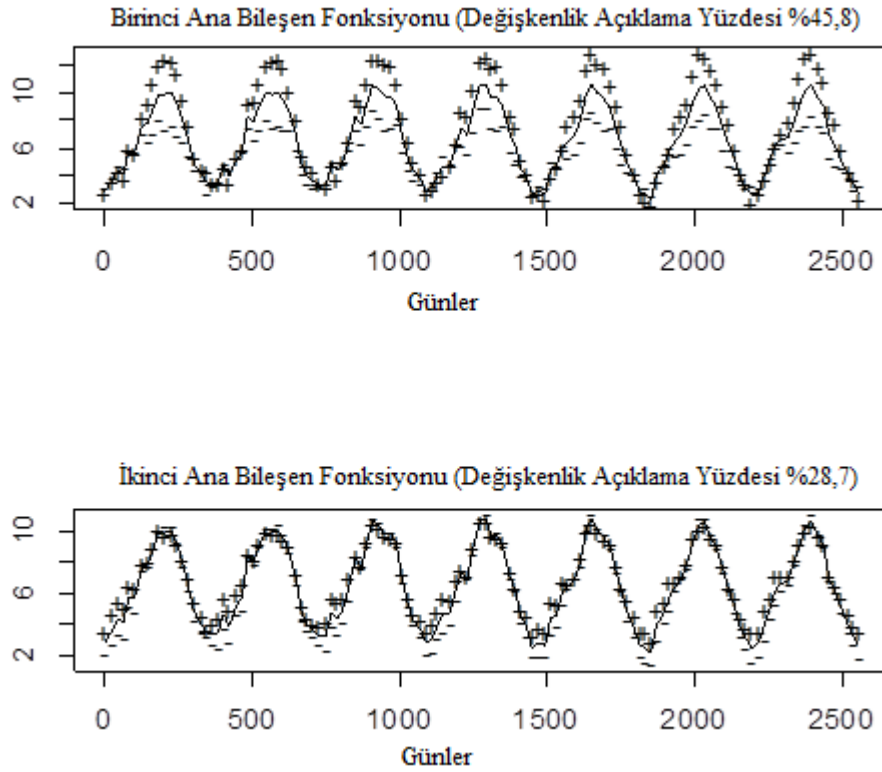
Çizelge 5. 9. Günlük Buharlaşma Miktarı için Temel Bileşen Skorları

İller	Birinci Temel Bileşen Skorları	İkinci Temel Bileşen Skorları
Ağrı	5,947979	-21,3928
Ardahan	-5,07439	-19,8688
Bayburt	5,947979	-21,3928
Erzincan	11,08317	-5,25883
Erzurum	-5,07439	-19,8688
İğdır	-16,0414	-16,5274
Kars	-5,07593	-19,8706

Bingöl	11,08317	-5,25883
Bitlis	-16,2113	-26,4866
Elazığ	9,788396	-26,1162
Hakkari	56,77819	-40,3643
Malatya	84,37522	-7,11355
Muş	-16,2113	-26,4866
Tunceli	11,08317	-5,25883
Van	60,29209	-8,67579
Artvin	-91,9727	-35,8488
Giresun	-5,77283	-24,8742
Gümüşhane	-5,52755	90,92301
Ordu	-38,6274	-25,63
Rize	-5,77283	-24,8742
Trabzon	-91,9727	-35,8488
Amasya	11,08317	-5,25883
Bartın	-99,5368	-41,0851
Çankırı	11,08317	-5,25883
Çorum	11,08317	-5,25883
Karabük	-40,721	82,16052
Kastamonu	-31,3556	-25,8116
Samsun	-35,7074	-31,5017
Sinop	-35,7074	-31,5017
Tokat	11,08317	-5,25883
Zonguldak	-38,6274	-25,63
Bilecik	-24,3825	101,9396
Bolu	-40,721	82,16052
Bursa	-24,3825	101,9396
Düzce	-38,6274	-25,63
Eskişehir	-13,371	-9,69522
Kocaeli	-13,6837	-20,7209
Sakarya	-38,6274	-25,63
Yalova	-24,3825	101,9396
Çanakkale	-7,2228	55,3613
Edirne	-18,3632	-17,1507
Kırklareli	-25,103	-17,4366
Balıkesir	-16,4871	-24,8658
Tekirdağ	-16,4871	-24,8658
İstanbul	-13,6837	-20,7209
Afyonkarahisar	-20,8633	44,67033
Aydın	-10,7341	-9,03146
Denizli	-4,98753	-18,3284
İzmir	-23,6934	-19,0662
Kütahya	-20,8633	44,67033
Manisa	-4,98753	-18,3284
Muğla	-23,6934	-19,0662
Uşak	-12,064	-18,1058
Ankara	-13,5164	-16,9828
Karaman	75,16492	-3,58828
Konya	49,79768	-5,55702
Aksaray	49,79768	-5,55702
Kayseri	-13,5164	-16,9828
Kırıkkale	-52,6196	-11,7972
Kırşehir	-52,6196	-11,7972
Nevşehir	-23,6934	-19,0662
Niğde	-12,064	-18,1058
Sivas	-13,5164	-16,9828
Yozgat	-13,5164	-16,9828
Adana	20,99793	118,146
Antalya	20,99793	118,146
Burdur	71,95801	-2,67916

Hatay	20,99793	118,146
Isparta	-21,0698	-15,7955
Mersin	74,92057	4,307298
Kahramanmaraş	-21,0698	-15,7955
Osmaniye	-1,38268	-9,01828
Adıyaman	-21,0698	-15,7955
Batman	-29,6287	-19,1061
Diyarbakır	76,99748	3,038168
Gaziantep	71,95801	-2,67916
Kilis	71,95801	-2,67916
Mardin	76,99748	3,038168
Siirt	125,7159	17,774
Şanlıurfa	76,99748	3,038168
Şırnak	125,7159	17,774

Şekil 5.15’de günlük güneşlenme süresi için birinci temel bileşen fonksiyonu üzerinde ortalama sapmaların ilkbahar ve yaz aylarına rastladığı gözlemlenmektedir. Başka bir ifadeyle %45,8 değişkenlik açıklama gücüyle günlük güneşlenme süresi açısından, fonksiyonlar arasındaki birinci temel değişkenliğin ilkbahar ve yaz aylarından kaynaklandığı görülmektedir. İkinci temel bileşen fonksiyonu üzerinde ortalama sapmaların kış aylarına rastladığı gözlemlenmektedir. %28,7 değişkenlik açıklama gücüyle günlük güneşlenme süresi açısından, fonksiyonlar arasındaki temel değişkenliğin kış aylarından kaynaklandığı görülmektedir.



Şekil 5. 15. Günlük Güneşlenme Süresi için Temel Bileşen Fonksiyonları

Çizelge 5.10 incelendiğinde, temel bileşen skorları %45,8 değişkenlik açıklayabilme gücüyle birinci temel bileşen için değerlendirildiğinde; Kars, Elazığ,

İzmir ve Mardin illerinin birinci temel bileşen üzerinde etkili olduğu ve yüksek temel bileşen skoruna sahip oldukları görülmektedir. Değişkenliğin sebebinin ilkbahar ve yaz ayları olduğunu gösteren birinci temel bileşen için veriler incelendiğinde, ilkbahar ve yaz aylarındaki en yüksek güneşlenme sürelerinin bu illere ait olduğu görülmektedir. İkinci temel bileşen açısından değerlendirildiğinde, özellikle Akdeniz, Doğu Anadolu ve Güneydoğu Anadolu Bölgesindeki illerin etkili olduğu ve yüksek bir temel bileşen skoruna sahip oldukları görülmektedir. Değişkenliğin %28,7'sini açıklama gücüne sahip ikinci temel bileşen için veriler incelendiğinde, kış aylarında en yüksek güneşlenme sürelerinin bu illere ait olduğu görülmektedir.

Çizelge 5. 10. Günlük Güneşlenme Süresi için Temel Bileşen Skorları

İller	Birinci Temel Bileşen Skorları	İkinci Temel Bileşen Skorları
Ağrı	0,148573	2,310227
Ardahan	-80,1546	-25,2327
Bayburt	31,46266	-29,8757
Erzincan	-44,6033	0,600664
Erzurum	0,953825	13,25794
Iğdır	-51,5624	-16,6867
Kars	246,061	271,1972
Bingöl	-68,186	-2,53547
Bitlis	55,88766	-51,3249
Elazığ	75,58043	-95,2064
Hakkari	40,22892	30,885
Malatya	40,14646	30,97266
Muş	44,04708	-48,4219
Tunceli	39,24204	5,600102
Van	56,90207	64,53652
Artvin	-91,928	-23,6989
Giresun	-83,068	-41,3036
Gümüşhane	-35,0428	-39,2796
Ordu	-136,309	-37,599
Rize	-70,4708	-13,8847
Trabzon	-115,104	-19,904
Amasya	-65,2432	-27,7378
Bartın	-21,2056	-31,4688
Çankırı	-1,28811	-17,6562
Çorum	-23,459	-34,9044
Karabük	-46,8969	-45,5134
Kastamonu	-78,7726	-36,9285
Samsun	-47,4806	-37,9505
Sinop	-41,2158	-50,9251
Tokat	-33,6791	-17,1871
Zonguldak	-31,5703	-51,6414
Bilecik	1,802047	-4,14483
Bolu	-82,0009	-43,8101
Bursa	-52,4193	-37,4444
Düzce	-52,5437	-58,2679
Eskişehir	-11,627	-8,07763
Kocaeli	-40,0053	-24,4942
Sakarya	-41,6608	-39,4421
Yalova	20,93074	-20,3515
Çanakkale	-2,57999	8,812605
Edirne	33,97944	-11,595

Kırklareli	-49,7175	-46,3776
Balıkesir	-19,637	-24,675
Tekirdağ	4,0358	-12,2483
İstanbul	-18,38	-7,41993
Afyonkarahisar	4,125997	5,615968
Aydın	-12,2552	43,28312
Denizli	19,04877	13,17852
İzmir	74,38399	56,01453
Kütahya	-9,60131	-18,0054
Manisa	7,986139	-23,3626
Muğla	58,75294	43,57946
Uşak	15,88578	3,722948
Ankara	-21,1474	-14,8375
Karaman	60,01156	27,735
Konya	-2,61282	0,553773
Aksaray	47,11822	16,25736
Kayseri	31,73368	-3,46001
Kırıkkale	32,32038	2,52373
Kırşehir	34,57939	21,19083
Nevşehir	58,11876	-3,2286
Niğde	33,87013	24,55377
Sivas	52,06158	-12,3192
Yozgat	30,03656	13,13129
Adana	12,40237	56,56892
Antalya	64,4953	60,33144
Burdur	39,61371	30,39334
Hatay	-36,8543	48,73057
Isparta	-6,43112	25,33269
Mersin	-109,576	-59,1288
Kahramanmaraş	4,352894	3,138281
Osmaniye	56,78108	45,49668
Adıyaman	-38,3973	-3,8577
Batman	13,68901	12,12578
Diyarbakır	59,2971	25,31728
Gaziantep	-63,3932	32,24179
Kilis	69,69039	50,5426
Mardin	85,6851	94,45247
Siirt	46,54831	24,61239
Şanlıurfa	66,21087	31,17378
Şırnak	67,87087	33,44384

5.4. Çok Değişkenli Fonksiyonel Verilerin Model Tabanlı Kümeleme Yöntemi ile Kümelenmesi

Fonksiyonel verilerin kümelenmesinde, model tabanlı kümeleme yöntemi kullanılmıştır. Model tabanlı kümelemede, küme sayısının belirlenmesi için BIC değeri kullanılmaktadır. Fonksiyonel verileri sınıflandırmak için FunHDDC paketi kullanılmıştır. Bu paket, fonksiyonel verileri, kümeye özgü fonksiyonel alt uzaylara uyarlayan bir pakettir. Kullanılan model $a_{kj}b_kQ_kd_k$ 'dir. İklimsel değişkenlerin hepsi bir bütün olarak düşünülerek kümeleme yapısı ortaya konmuştur. Bu çalışmada, hem model parametrelerini hem de gruba özgü fonksiyonel alt uzayları belirlemek için EM algoritması kullanılmıştır. EM algoritması, iki ardışık adımın yaklaşık

olabilirlik değeri farkının, belirlenen bir ε eşik değerinden düşük olduğu zaman durmaktadır. Bu çalışmada, $\varepsilon = e^{-6}$ olarak seçilmiştir. Verilerin kümeleneşmesi için yapılan iterasyon sayısı 200 olarak belirlenmiştir. Burada ortalama fonksiyonu, çok deęişkenli normal dağılım tarafından üretilmektedir. Kovaryans fonksiyonu tüm örnek için aynıdır. Algoritma aynı model için 20 defa çalıştırılmıştır.

Çizelge 5.11’de $K = 2: 10$ için simülasyonlar yapılmış ve en iyi BIC değeri küme sayısının belirlenmesinde kullanılmıştır. Simülasyon sonucunda $BIC = -852432,9$ bulunmuş ve buna göre uygun küme sayısı, $K = 3$ olarak belirlenmiştir. $K = 3$ için kullanılan parametre sayısı 12779’dur.

Çizelge 5. 11. Küme Sayısının Belirlenmesi Simülasyonu

Model	K	Eşik	BIC
$a_{kj}b_kQ_kd_k$	2	0,2	-895676,97
$a_{kj}b_kQ_kd_k$	3	0,2	-852432,9
$a_{kj}b_kQ_kd_k$	4	0,2	-882699,97
$a_{kj}b_kQ_kd_k$	5	0,2	-893239,81
$a_{kj}b_kQ_kd_k$	6	0,2	-900501,54
$a_{kj}b_kQ_kd_k$	7	0,2	-915744,96
$a_{kj}b_kQ_kd_k$	8	0,2	-920961,45
$a_{kj}b_kQ_kd_k$	9	0,2	-927554,87
$a_{kj}b_kQ_kd_k$	10	0,2	-930631,85

Belirlenen küme sayısına göre, model tabanlı kümeleme yöntemi ile 81 ilin iklimsel deęişkenler yönünden kümelendirilmesi sonuçları, Çizelge 5.12’de gösterilmiştir. Bu sonuçlara göre birinci kümede 2 il, ikinci kümede 56 il ve üçüncü kümede 23 il yer almıştır. Model tabanlı kümeleme yöntemine göre kümeleme sonuçlarına bakıldığında yoğunlaşmanın ikinci kümede olduğu görülmektedir. Kümelerin eleman sayıları bakımından değerlendirildiğinde, küme yoğunlaşma oranları $(\frac{n_i}{n}, i = 1,2,3)$ Çizelge 5.13’deki gibidir. Bu oranlara göre ele alınan iklimsel deęişkenler yönünden, Türkiye’deki illerin hemen hemen %69’u birbirine benzerlik gösterirken yaklaşık %31’nin bu illerden farklılaştığını söylemek mümkündür. Bu %31’lik dilimin içerisinde yer alan ve iki ilden (Kars ve Çanakkale) oluşan birinci kümedeki iller özellikle coęrafi yerleri bakımından dikkat çekici bulunmuştur. Çünkü coęrafi konum bakımından biri ülkenin doğusunda yer alırken, dięeri batısında yer almaktadır. Dolayısıyla bu iki ilin aynı kümede yer alması

beklenmeyen bir durum olup, bu illerin aykırı gözlemler olarak ele alınabileceği düşünülmektedir.

Çizelge 5. 12. Model Tabanlı Kümeleme Sonuçları

Gruplar			
Sıra No	1	2	3
1	Çanakkale	Adıyaman	Adana
2	Kars	Afyonkarahisar	Ağrı
3		Aksaray	Antalya
4		Amasya	Ardahan
5		Ankara	Aydın
6		Artvin	Balıkesir
7		Bartın	Bayburt
8		Batman	Bitlis
9		Bilecik	Bursa
10		Bingöl	Elazığ
11		Bolu	Erzurum
12		Burdur	Eskişehir
13		Çankırı	Gümüşhane
14		Çorum	Hakkâri
15		Denizli	Hatay
16		Diyarbakır	Iğdır
17		Düzce	Karabük
18		Edirne	Mersin
19		Erzincan	Muğla
20		Gaziantep	Muş
21		Giresun	Osmaniye
22		Isparta	Tunceli
23		İstanbul	Van
24		İzmir	
25		Kahramanmaraş	
26		Karaman	
27		Kastamonu	
28		Kayseri	
29		Kırıkkale	
30		Kırklareli	
31		Kırşehir	
32		Kilis	
33		Kocaeli	
34		Konya	
35		Kütahya	
36		Malatya	
37		Manisa	
38		Mardin	
39		Nevşehir	
40		Niğde	
41		Ordu	
42		Rize	
43		Sakarya	
44		Samsun	
45		Siirt	
46		Sinop	
47		Sivas	
48		Şanlıurfa	
49		Şırnak	
50		Tekirdağ	
51		Tokat	
52		Trabzon	

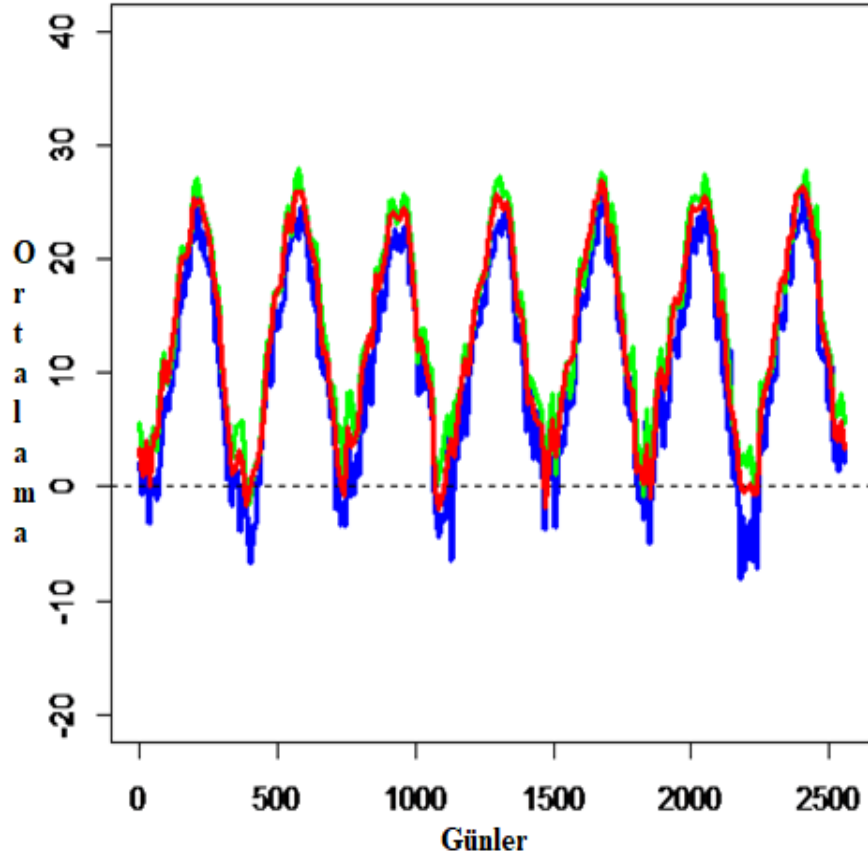
53	Uşak
54	Yalova
55	Yozgat
56	Zonguldak

Çizelge 5. 13. İllerin ait Olduğu Küme Olasılıkları

Küme	1	2	3
Olasılık	0,024691	0,691358	0,283951

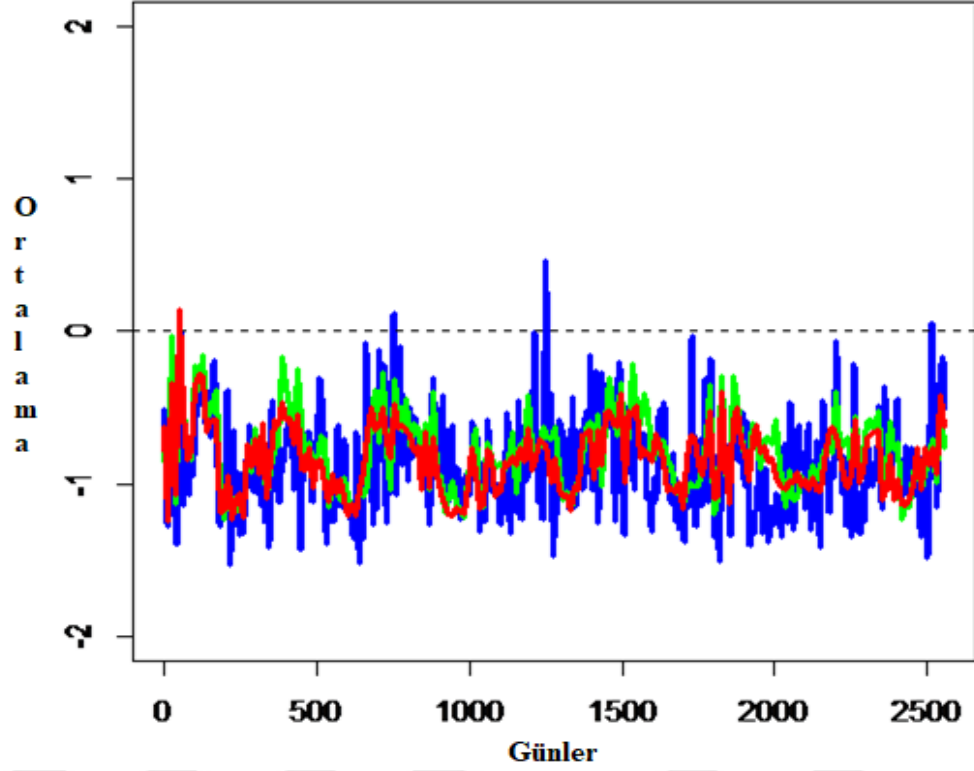
Çalışmada ele alınan iklimsel değişkenlere göre illerin kümelenmesinde kümeleme fonksiyonları Şekil 5.16 - 5.20’de gösterilmiştir. Şekil 5.16 ile verilen günlük ortalama sıcaklık ve Şekil 5.17’de verilen günlük toplam yağış verisi hariç, diğer şekillerde kümeleri oluşturulan ortalama fonksiyonlar arasında fark olduğu net bir şekilde görülmektedir. Kırmızı çizgiler birinci kümeye ait ortalama fonksiyonunu, mavi çizgiler ikinci kümeye ait ortalama fonksiyonunu ve yeşil çizgiler ise üçüncü kümeye ait ortalama fonksiyonunu göstermektedir.

Meteorolojik veriler incelendiğinde 2016-2017 yılları arasında, ikinci kümede yer alan Karadeniz Bölgesindeki illere ait sıcaklıkların düşük olduğu görülmüştür.



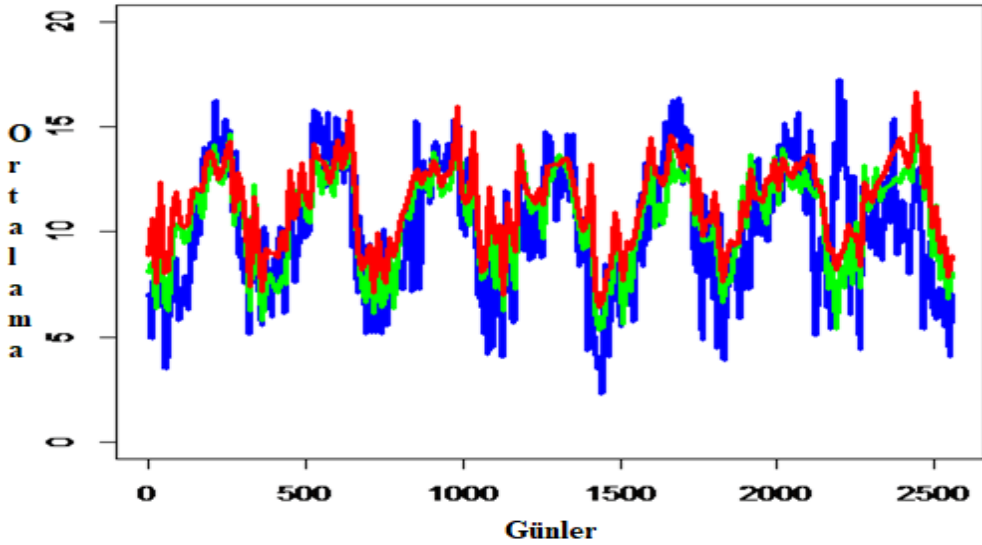
Şekil 5. 16. Günlük Ortalama Sıcaklık için Kümeleme Fonksiyonları

Meteorolojik veriler incelendiğinde 2015-2016 yılları arasında, ikinci kümede yer alan Karadeniz Bölgesindeki illerin yüksek yağış aldığı ve sel felaketi ile karşılaştığı görülmüştür.



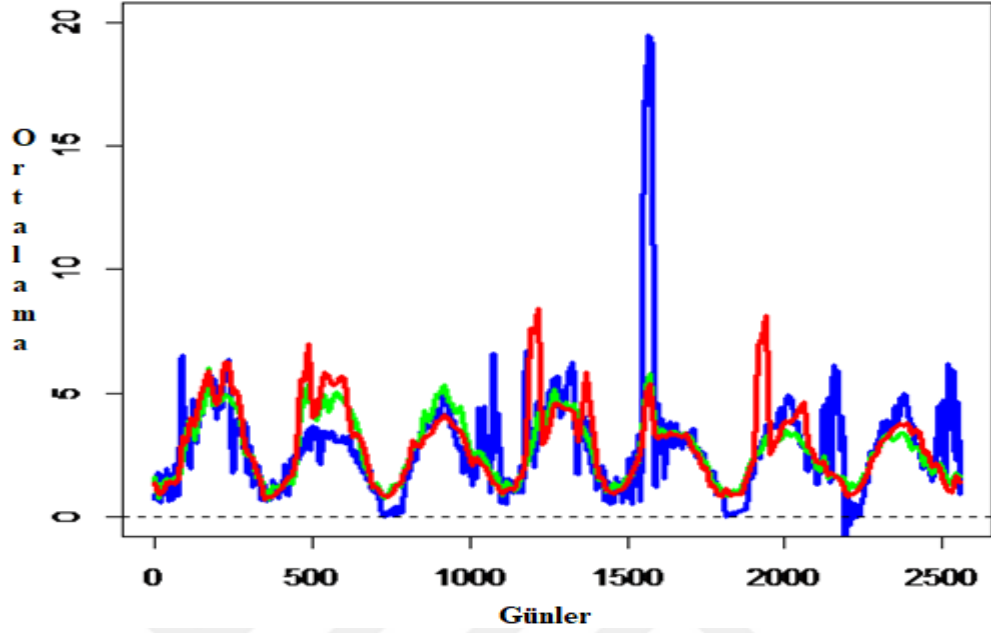
Şekil 5. 17. Günlük Toplam Yağış Miktarı için Kümeleme Fonksiyonları

Meteorolojik veriler incelendiğinde 2012-2018 yılları arasında, ikinci kümede yer alan Doğu Anadolu Bölgesindeki illere ait sıcaklık değişimlerinin yüksek, Karadeniz Bölgesindeki illere ait sıcaklık değişimlerinin ise düşük olduğu görülmüştür.



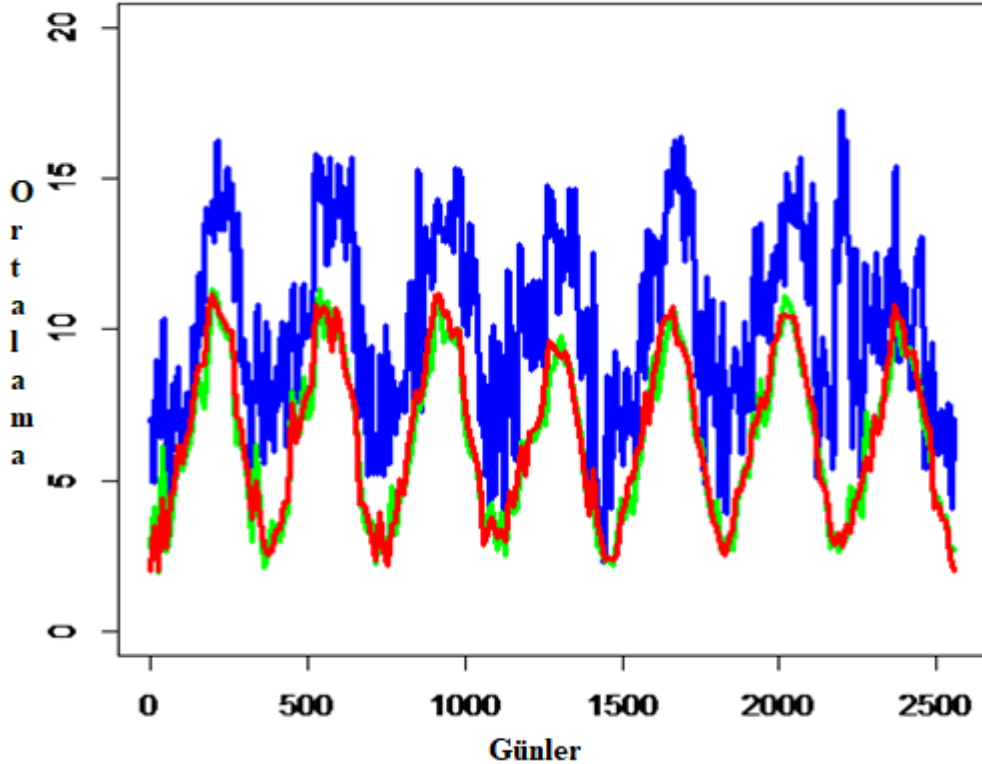
Şekil 5. 18. Günlük Sıcaklık Değişimi için Kümeleme Fonksiyonları

Meteorolojik veriler incelendiğinde 2016-2017 yılları arasında, ikinci kümede yer alan Güneydoğu Anadolu Bölgesindeki illere ait buharlaşmanın yüksek olduğu görülmüştür.



Şekil 5. 19. Günlük Buharlaşma Miktarı için Kümeleme Fonksiyonları

Meteorolojik veriler incelendiğinde 2016-2017 yılları arasında, ikinci kümede yer alan Güneydoğu Anadolu Bölgesindeki illere ait güneşlenme sürelerinin yüksek olduğu görülmüştür.



Şekil 5. 20. Günlük Güneşlenme Süresi için Kümeleme Fonksiyonları

5.5. Çok Değişkenli Fonksiyonel Verilerin Fonksiyonel ANOVA ile İncelenmesi

Günlük ortalama sıcaklık ($X_1(t)$), günlük toplam yağış miktarı ($X_2(t)$), günlük sıcaklık değişimi ($X_3(t)$), günlük buharlaşma miktarı ($X_4(t)$) ve günlük güneşlenme süresi ($X_5(t)$) olmak üzere, değişkenler için ortalama fonksiyonları bakımından aralarında anlamlı bir farklılık olup olmadığının test edilmesinde FANOVA kullanılmıştır. Çizelge 5.14 FANOVA için test istatistikleri ve p değerlerini göstermektedir. Değişkenler için test edilmek istenen hipotezler ise;

$$H_0: \eta_1(t) \equiv \eta_2(t) \equiv \eta_3(t)$$

$$H_1: \exists \eta_i(t) \text{ diğerlerinden farklıdır } (i = 1,2,3)$$

şeklindedir.

Günlük ortalama sıcaklık için Çizelge 5.14 incelendiğinde, kümelerin ortalama fonksiyonları bakımından anlamlı bir farklılık taşımadığı görülmektedir. Şekil 5.16'da görsel olarak, sonuçlar desteklenmektedir. L_2 -norm tabanlı testler incelendiğinde, yanlı indirgenmiş yöntem (L_2B) ve naive yöntemin (L_2N) birbirine benzer sonuçlar verdiği görülmektedir. Ayrıca, iki yöntemin p değerleri tam olarak aynı olmasa da tutarlı oldukları görülmektedir. Bootstrap örneklerinin sayısı 10000 olarak alınmış ve iyi sonuçlar vermediği görülmüştür. L_2 -norm tabanlı bootstrap testinin, F -tipi bootstrap testine göre daha iyi sonuç verdiği görülmektedir.

Günlük toplam yağış miktarı için Çizelge 5.14 incelendiğinde, kümelerin ortalama fonksiyonları bakımından anlamlı bir farklılık taşımadığı görülmektedir. Şekil 5.17'de görsel olarak sonuçlar desteklenmektedir. L_2 -norm tabanlı testler incelendiğinde, yanlı indirgenmiş yöntem (L_2B) ve naive yöntemin (L_2N) birbirine benzer sonuçlar verdiği görülmektedir. Ayrıca iki yöntemin p değerleri tam olarak aynı olmasa da tutarlı oldukları görülmektedir. Bootstrap örneklerinin sayısı 10000 olarak alınmış ve iyi sonuçlar vermediği görülmüştür. L_2 -norm tabanlı bootstrap testinin, F -tipi bootstrap testine göre daha iyi sonuç verdiği görülmektedir.

Günlük ortalama sıcaklık değişimi için Çizelge 5.14 incelendiğinde, kümelerin ortalama fonksiyonları bakımından anlamlı bir farklılık taşıdığı görülmektedir. Şekil 5.18'de görsel olarak sonuçlar desteklenmektedir. L_2 -norm tabanlı testler incelendiğinde, yanlı indirgenmiş yöntem (L_2B) ve naive yöntemin (L_2N) birbirine benzer sonuçlar verdiği görülmektedir. Ayrıca iki yöntemin p değerleri tam olarak

aynı olmasa da tutarlı oldukları görülmektedir. L_2 -norm tabanlı testin, F-tipi testten daha iyi sonuç verdiği görülmektedir. Bootstrap örneklerinin sayısı 10000 olarak alınmış ve iyi sonuçlar vermediği görülmüştür. L_2 -norm tabanlı bootstrap testinin F-tipi bootstrap testine göre daha iyi sonuç verdiği görülmektedir.

Günlük buharlaşma miktarı için Çizelge 5.14 incelendiğinde, kümelerin ortalama fonksiyonları bakımından anlamlı bir farklılık taşıdığı görülmektedir. Şekil 5.19'da görsel olarak sonuçlar desteklenmektedir. L_2 -norm tabanlı testler incelendiğinde, yanlı indirgenmiş yöntem (L_2B) ve naive yöntemin (L_2N) birbirine benzer sonuçlar verdiği görülmektedir. Ayrıca iki yöntemin p değerleri tam olarak aynı olmasa da tutarlı oldukları görülmektedir. L_2 -norm tabanlı testin, F-tipi testten daha iyi sonuç verdiği görülmektedir. Bootstrap örneklerinin sayısı 10000 olarak alınmış ve iyi sonuçlar vermediği görülmüştür. L_2 -norm tabanlı bootstrap testinin, F-tipi bootstrap testine göre daha iyi sonuç verdiği görülmektedir.

Günlük güneşlenme süresi için Çizelge 5.14 incelendiğinde, kümelerin ortalama fonksiyonları bakımından anlamlı bir farklılık taşıdığı görülmektedir. Şekil 5.20'de görsel olarak sonuçlar desteklenmektedir. L_2 -norm tabanlı testler incelendiğinde, yanlı indirgenmiş yöntem (L_2B) ve naive yöntemin (L_2N) birbirine benzer sonuçlar verdiği görülmektedir. Ayrıca iki yöntemin p değerleri tam olarak aynı olmasa da tutarlı oldukları görülmektedir. L_2 -norm tabanlı testin, F-tipi testten daha iyi sonuç verdiği görülmektedir. Bootstrap örneklerinin sayısı 10000 olarak alınmış ve iyi sonuçlar vermediği görülmüştür. L_2 -norm tabanlı bootstrap testinin, F-tipi bootstrap testine göre daha iyi sonuç verdiği görülmektedir.

Çizelge 5. 14. Değişkenlere göre Tek Yönlü Fonksiyonel ANOVA Test İstatistiği Sonuçları ve p Değerleri

Testler	Değişkenler				
	Günlük Ortalama Sıcaklık	Günlük Toplam Yağış Miktarı	Günlük Sıcaklık Değişimi	Günlük Güneşlenme Süresi	Günlük Buharlaşma Miktarı
L_2N	$T_n = 209589,1$	$T_n = 3476,598$	$T_n = 129841,9$	$T_n = 138542,8$	$T_n = 60409,82$
	$\hat{\beta} = 22043,98$	$\hat{\beta} = 59,80377$	$\hat{\beta} = 6571,005$	$\hat{\beta} = 1204,793$	$\hat{\beta} = 2180,092$
	$\hat{d} = 4,577366$	$\hat{d} = 52,67402$	$\hat{d} = 10,90537$	$\hat{d} = 46,97927$	$\hat{d} = 13,79395$
	$p = 0,0712876$	$p = 0,2812835$	$p = 0,04688431$	$p = 1,248358e^{-7}$	$p = 0,01421683$
L_2B	$T_n = 209589,1$	$T_n = 3476,598$	$T_n = 129841,9$	$T_n = 138542,8$	$T_n = 60409,82$
	$\hat{\beta} = 21133,17$	$\hat{\beta} = 39,12208$	$\hat{\beta} = 6036,247$	$\hat{\beta} = 831,5826$	$\hat{\beta} = 1962,804$
	$\hat{d} = 4,723361$	$\hat{d} = 80,46857$	$\hat{d} = 11,82021$	$\hat{d} = 68,01206$	$\hat{d} = 15,26971$
	$p = 0,06626803$	$p = 0,2443702$	$p = 0,04034888$	$p = 3,033204e^{-10}$	$p = 0,01057928$
L_2b	$T_n^* = 209589,1$	$T_n^* = 3476,598$	$T_n^* = 129841,9$	$T_n^* = 138542,8$	$T_n^* = 60409,82$
	$p = 0,607$	$p = 0,2165$	$p = 0,7386$	$p = 0,3374$	$p = 0,1265$
FT	$F_n = 2,077127$	$F_n = 1,103645$	$F_n = 1,811936$	$F_n = 2,44774$	$F_n = 2,008834$
	$p = 0,07639064$	$p = 0,2844771$	$p = 0,05046737$	$p = 2,38694e^{-7}$	$p = 0,01604624$
Fb	$F_n^* = 2,077127$	$F_n^* = 1,103645$	$F_n^* = 1,811936$	$F_n^* = 2,44774$	$F_n^* = 2,008834$
	$p = 0,7615$	$p = 0,3313$	$p = 0,8918$	$p = 0,7856$	$p = 0,1741$
PF	$F_n(t) = 1,94262$	$F_n(t) = 1,091962$	$F_n(t) = 2,125067$	$F_n(t) = 2,723363$	$F_n(t) = 1,345824$
	$\hat{\beta} = 0,2053331$	$\hat{\beta} = 0,02039162$	$\hat{\beta} = 0,07188914$	$\hat{\beta} = 0,0201402$	$\hat{\beta} = 0,04550941$
	$\hat{d} = 4,998297$	$\hat{d} = 50,33027$	$\hat{d} = 14,27637$	$\hat{d} = 50,95856$	$\hat{d} = 22,55173$
	$p = 0,09195204$	$p = 0,3518099$	$p = 0,009936569$	$p = 1,458928e^{-9}$	$p = 0,1467945$

6. SONUÇ VE ÖNERİLER

Küreselleşen dünyada gelişen teknoloji, büyük hacimli veri setleri ile çalışabilmeye olanak sağlamaktadır. Bu gelişmelerle beraber yüksek boyutlu veri setleri ile çalışırken, verilerin analiz sürecinde klasik tek değişkenli ve çok değişkenli metodların kullanılması yerine FVA'nın kullanılması daha doğru bir tercih olacaktır. Çünkü inceleme yapılan gözlem sayısı arttıkça bu verilerin reel bir fonksiyondan örneklendiği varsayılmaktadır. Ayrıca klasik tek değişkenli ve çok değişkenli yöntemler, verilerin fonksiyonel yapısını görmezden gelmektedir. FVA verilerin analizi sürecinde önemli bir teknik olup, araştırmacılar bu yöntemle özel bir ilgi göstermektedirler. İnterpolasyon yöntemiyle kayıp gözlemler tahmin edilebilir ve oluşturulan fonksiyonların istenilen dereceden türevleri alınabilir. FVA, fark edilmesi zor olan ilişkileri rahat bir biçimde ortaya çıkarmaktadır. FVA'da analiz sürecinde bireysel fonksiyonlardan, kovaryans ve korelasyon yüzeylerinden, FTBA'dan, kümeleme yöntemlerinden ve FANOVA'dan yararlanılmaktadır.

Kümeleme analizi, çalışılan bir olguyu anlamak ve yorumlamak amacıyla verilerin kümelenmesini sağlayan en önemli istatistiksel sınıflama yöntemlerinden biridir. Fonksiyonel verileri kümeleme konusu son yıllarda araştırmacıların oldukça ilgisini çekmektedir. Bununla beraber fonksiyonel verilerin sınıflandırılması için çeşitli yöntemler geliştirilmiş ve geliştirilmeye devam edilmektedir.

Varyans analizi, gruplar ya da kümeler arasında ortalamalar bakımından istatistiksel olarak anlamlı bir farklılık olup olmadığını test etmek için kullanılmaktadır. FANOVA ise oluşturulan grupların ya da kümelerin ortalama fonksiyonları bakımından istatistiksel olarak anlamlı bir farklılık olup olmadığını test etmek için kullanılmaktadır.

Bu tez çalışmasında iklimsel değişkenler kullanılarak illerin kümelenme yapıları belirlenmiştir. Türkiye'de bulunan 81 ilin 2012-2018 yıllarındaki günlük ortalama sıcaklık, günlük toplam yağış miktarı, günlük sıcaklık değişimi, günlük güneşlenme süresi ve günlük toplam buharlaşma miktarı için Fourier baz fonksiyon yaklaşımı ve pürüzlülük cezası yaklaşımı kullanılarak her bir değişken için 81 bireysel fonksiyon oluşturulmuştur. Fonksiyonların oluşturulması için 365 baz fonksiyonu kullanılmıştır. Değişkenlere ait düzgülendirme parametresinin

bulunmasında GCV kullanılmıştır. Verilere ait deęişkenlięin neden kaynaklandığıın bulanabilmesi için FTBA kullanılmıştır. Bu alıřmada örneklem hacmi deęişken sayısından küçük olduęundan, veriler arasındaki deęişkenlik yapısını ortaya koymak için klasik çok deęişkenli temel bileşenler analizini kullanmak doęru bir tercih deęildir. Model tabanlı kümeleme yöntemi ile fonksiyonların kümelenme yapısı ortaya çıkarılmıştır. Oluřturulan kümelerin ise ortalama fonksiyonları bakımından aralarında anlamlı bir farklılık olup olmadığını test etmek için FANOVA kullanılmıştır.

Bireysel fonksiyonlar yönünden deęerlendirildięinde,

- ✓ Günlük ortalama sıcaklık ile günlük toplam yaęış miktarının negatif yönde bir ilişkiye sahip olduęunu,
- ✓ Günlük ortalama sıcaklık ile günlük sıcaklık deęişiminin pozitif yönde bir ilişkiye sahip olduęunu,
- ✓ Günlük toplam yaęış miktarı ile günlük sıcaklık deęişiminin negatif yönde bir ilişkiye sahip olduęunu,
- ✓ Günlük ortalama sıcaklık ile günlük buharlaşma miktarının pozitif yönde bir ilişkiye sahip olduęunu,
- ✓ Günlük toplam yaęış miktarı ile günlük buharlaşma miktarının negatif yönde bir ilişkiye sahip olduęunu,
- ✓ Günlük sıcaklık deęişimi ile günlük buharlaşma miktarının pozitif yönde bir ilişkiye sahip olduęunu,
- ✓ Günlük ortalama sıcaklık ile günlük güneşlenme süresinin pozitif yönde bir ilişkiye sahip olduęunu,
- ✓ Günlük toplam yaęış miktarı ile günlük güneşlenme süresinin negatif yönde bir ilişkiye sahip olduęunu,
- ✓ Günlük sıcaklık deęişimi ile günlük güneşlenme süresinin pozitif yönde bir ilişkiye sahip olduęunu,
- ✓ Günlük buharlaşma miktarı ile günlük güneşlenme süresinin pozitif yönde bir ilişkiye sahip olduęunu,

söylemek mümkündür.

FTBA kullanılarak iklim deęişkenleri arasındaki deęişkenlięin neden kaynaklandığı araştırılmıştır. %55,8 deęişkenlik açıklama gücüyle günlük ortalama

sıcaklık verileri açısından, fonksiyonlar arasındaki birinci temel değişkenliğin kış aylarından kaynaklandığı görülmüştür. %35,8 değişkenlik açıklama gücüyle günlük ortalama sıcaklık verileri açısından, fonksiyonlar arasındaki ikinci temel değişkenliğin yaz aylarından kaynaklandığı görülmüştür. %48,1 değişkenlik açıklama gücüyle günlük toplam yağış miktarı açısından, fonksiyonlar arasındaki birinci temel değişkenliğin kış aylarından kaynaklandığı görülmüştür. %23,4 değişkenlik açıklama gücüyle günlük toplam yağış miktarı açısından, fonksiyonlar arasındaki ikinci temel değişkenliğin yaz aylarından kaynaklandığı görülmüştür. %58,1 değişkenlik açıklama gücüyle günlük sıcaklık değişimi açısından, fonksiyonlar arasındaki birinci temel değişkenliğin ilkbahar ve yaz aylarından kaynaklandığı görülmüştür. %32,6 değişkenlik açıklama gücüyle günlük sıcaklık değişimi açısından, fonksiyonlar arasındaki ikinci temel değişkenliğin kış aylarından kaynaklandığı görülmüştür. %44,7 değişkenlik açıklama gücüyle günlük buharlaşma miktarı açısından, fonksiyonlar arasındaki birinci temel değişkenliğin ilkbahar ve yaz aylarından kaynaklandığı görülmüştür. %37,2 değişkenlik açıklama gücüyle günlük toplam buharlaşma miktarı açısından, fonksiyonlar arasındaki ikinci temel değişkenliğin ilkbahar aylarından kaynaklandığı görülmüştür. %45,8 değişkenlik açıklama gücüyle günlük güneşlenme süresi açısından, fonksiyonlar arasındaki birinci temel değişkenliğin ilkbahar ve yaz aylarından kaynaklandığı görülmüştür. %28,7 değişkenlik açıklama gücüyle günlük güneşlenme süresi açısından, fonksiyonlar arasındaki ikinci temel değişkenliğin kış aylarından kaynaklandığı görülmüştür.

Verilerin kümelenmesinde model tabanlı kümeleme yöntemi kullanılmıştır. Kullanılan model $a_{kj}b_kQ_kd_k$ 'dir. Model parametrelerinin tahmininde ve gruba özgü fonksiyonel alt uzayları belirlemede EM algoritması kullanılmıştır. $\epsilon = e^{-6}$ olarak belirlenmiştir. Kümeleme işlemi için iterasyon sayısı 200 olarak alınmıştır. Algoritma aynı model için 20 defa çalıştırılmıştır. *BIC* değeri küme sayısının belirlenmesi işleminde kullanılmıştır. $K = 3$ için en iyi değer $BIC = -852432,9$ olarak bulunmuştur. Bu sebeple küme sayısı, $K = 3$ olarak alınmıştır. Kullanılan parametre sayısı ise 12779'dur. İllerin yoğun bir şekilde ikinci kümede yer aldığı görülmüştür.

Kümelerin, ortalama fonksiyonları bakımından anlamlı bir farklılık olup olmadığını test etmek için FANOVA kullanılmıştır. Günlük ortalama sıcaklık için,

kümelerin ortalama fonksiyonları bakımından anlamlı bir farklılık taşımadığı görülmüştür. Günlük toplam yağış miktarı için, kümelerin ortalama fonksiyonları bakımından anlamlı bir farklılık taşımadığı görüldü. Günlük ortalama sıcaklık farkı için, kümelerin ortalama fonksiyonları bakımından anlamlı bir farklılık taşıdığı görülmüştür. Günlük buharlaşma miktarı için, kümelerin ortalama fonksiyonları bakımından anlamlı bir farklılık taşıdığı görülmektedir. Günlük güneşlenme süresi için, kümelerin ortalama fonksiyonları bakımından anlamlı bir farklılık taşıdığı görülmektedir. Bootstrap örneklerinin sayısı 10000 olarak alınmıştır ve bootstrap testlerinin iyi sonuç vermedikleri görülmüştür. L_2 -norm tabanlı bootstrap testinin, F -tipi bootstrap testine göre daha iyi sonuç verdiği görülmüştür. L_2 -norm tabanlı testler incelendiğinde, yanlış indirgenmiş yöntem ile naive yöntemin benzer sonuçlar verdiği görülmüştür.

İleriki çalışmalarda, çok değişkenli fonksiyonel verilerin analizinde fonksiyonel MANOVA (FMANOVA) kullanılabilir. Bu sayede tüm değişkenleri bir arada değerlendirme fırsatı elde edilerek, daha farklı ve ayrıntılı sonuçlara ulaşılabilir.

KAYNAKLAR

- Abraham, C., Cornillon, P. A., Matzner-Løber, E. and Molinari, N. 2003. Unsupervised curve clustering using B-splines. *Scandinavian Journal of Statistics, Theory and Applications*, 30:3, 581 - 595.
- Baragilly, M. 2016. Clustering Multivariate and Functional Data Using Spatial Rank Functions. Doctoral Dissertation, The University of Birmingham.
- Boullé, M. 2012. Functional data clustering via piecewise constant nonparametric density estimation. *Pattern Recognition*, 45:12, 4389 - 4401.
- Bouveyron, C. and Jacques, J. 2011. Model-based clustering of time series in group specific functional subspaces. *Advances in Data Analysis and Classification*, 5:4, 281 - 300.
- Bouveyron, C., Girard, S. and Schmid, C. 2007. High dimensional data clustering. *Computational Statistics and Data Analysis*, 52, 502 - 519.
- Chen, Y., Dang, X., Peng, H. and Bart, J. 2009. Outlier detection with the kernelized spatial depth function. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 31:2, 288 - 305.
- Chiou, J. M. and Li, P. L. 2007. Functional clustering and identifying substructures of longitudinal data. *Journal of the Royal Statistical Society*, 69:4, 679 - 699.
- Costanzo, G. D. 2005. *Functional principal component analysis of financial time series*. Cnam-Paris.
- Craven, P. and Wahba, G. 1979. Smoothing noisy data with spline functions. *Numerische Mathematik*, 31.
- Delaigle, A. and Hall, P. 2010. Defining probability density for a distribution of random functions. *Annals of Statistics*, 38, 1171 - 1193.
- Everitt, B. S., Landau, S., Leese, M. and Stahl, D. 2011. *Cluster analysis* (fifth Edition). John Wiley and Sons, Chichester.
- Ferraty, F. and Vieu, P. 2006. *Nonparametric functional data analysis*. Springer, New York.
- Fugal, D. 2009. *Conceptual wavelets in digital signal processing*. Space and Signals Technical Publishing.
- Gan, G., Ma, C. and Wu, J. 2007. *Data clustering theory, algorithms, and applications*. Philadelphia, PA:SIAM.
- Gasser, T. and Kneip, A. 1995. Searching for the structure in curve sample. *Journal of the American Statistical Association*, doi: 10.1080/01621459.1995.10476624

- Genton, M. G. 2001. Classes of kernels for machine learning: A statistics perspective. *Journal of Machine Learning Research*, 2, 299 - 312.
- Giacofci, M., Lambert-Lacroix, S., Marot, G. and Picard, F. 2013. Wavelet-based clustering for mixed-effects functional models in high dimension. *Biometrics*, 69:1, 31 - 40.
- Gorecki, T. and Smaga, L. 2015. A comparison of tests for the one-way ANOVA problem for functional data. *Computational Statistics*, 30,987–1010.
- Gorecki, T. and Smaga, L. 2017. Multivariate analysis of variance for functional data. *Journal of Applied Statistics*, 44, 2172–2189.
- Green, P. J. and Silverman, B. W. 1994. *Nonparametric regression and generalized linear models: a roughness penalty approach*. Chapman and Hall, London.
- Heard, N., Holmes, C. and Stephens, D. 2006. A quantitative study of gene regulation involved in the immune response of anopheline mosquitoes: an application of Bayesian hierarchical clustering of curves. *Journal of the American Statistical Association*, 101:473, 18 - 29.
- Horvath, L., Kokoszka, P. and Reimherr, M. 2009. Two sample inference in functional linear models. *Canadian Journal of Statistics*, 37:4, 571–591.
- Ieva, F., Paganoni, A., Pigoli, D. and Vitelli, V. 2013. Multivariate functional clustering for the morphological analysis of electrocardiograph curves. *Journal of the Royal Statistical Society: Series C*, 62:3, 401 - 418.
- Ingrassia, S. and Costanzo, G. D. 2005. Functional Principal Component Analysis of Financial Time Series In: *New developments in Classification and Data Analysis*. Proceedings of the Meeting of the Classification and Data Analysis Group (CLADAG) of the Italian Statistical Society, 351 – 358, University of Bologna,
- Jacques, J. and Preda, C. 2013. Funclust: a curves clustering method using functional random variable density approximation. *Neurocomputing*, 112, 164 - 171.
- Jacques, J. and Preda, C. 2014a. Functional data clustering: a survey. *Advances in Data Analysis and Classification*, 8:3, 231 – 255.
- Jacques, J. and Preda, C. 2014b. Model based clustering for multivariate functional data. *Computational Statistics and Data Analysis*, 71, 92-106.
- James, G. and Sugar, C. 2003. Clustering for sparsely sampled functional data. *Journal of the American Statistical Association*, 98:462, 397 - 408.
- Kadir, E. ve Keser, İ. 2007. Düzgünleştirilmiş fonksiyonel ana bileşenler analizi ile İMKB verilerinin İncelenmesi. *İstanbul Üniversitesi İktisat Fakültesi Ekonometri ve İstatistik Dergisi*, 8, 1-32.

- Karhunen, K. 1947. Über lineare methoden in der wahrscheinlichkeitsrechnung. *Annales Academiae scientiarum Fennicae. Series A. I, Mathematica-physica*, 37, 79.
- Kayano, M., Dozono, K. and Konishi, S. 2010. Functional cluster analysis via orthonormalized gaussian basis expansions and its application. *Journal of Classification*, 27, 211 - 230.
- Keser, İ. 2008. Karşılaştırmalı olarak fonksiyonel ana bileşenler analizi ve GSYİH verilerinin incelenmesi. *Ege Akademik Bakış*, 8, 915 - 928.
- Keser, İ. 2010. Ege bölgesi yağış verilerinin fonksiyonel veri analizi ile incelenmesi. *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 1, 41 - 67.
- Loève, M. 1945. Fonctions aléatoires de second ordre. *Comptes Rendus de Academie Sciences*, 220, 469.
- Mardia, K. V., Kent, J. T. and Biby, J. M. 1989. *Multivariate analysis*. Academic Press, London.
- Muirhead, R. J. 1982. *Aspects of multivariate statistical theory*. John Wiley and Sons, New York.
- Peng, J. and Müller, H. G. 2008. Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions. *Annals of Applied Statistics*, 2:3, 1056 - 1077.
- Ramsay J.O., 2004. Functional data analysis. *Encyclopedia of Statistics in Behavioral Science*. doi:10.1002/0471667396.ess3138
- Ramsay J.O., 2005. Functional data analysis. *Encyclopedia of Statistics in Behavioral Science*. doi: 10.1002/047013192.bsa239
- Ramsay, J. O. and Dalzell, C. 1991. Some tools for functional data analysis. *Journal of the Royal Statistical Society: Series B*, 53:3, 539-572.
- Ramsay, J. O. and Li, X. 1998. Curve registration, *Journal of the Royal Statistical Society, Series B*, 60(2), 351-363.
- Ramsay, J. O. and Silverman, B. W. 2002a. *Applied functional data analysis: methods and case studies*. Springer, New York.
- Ramsay, J. O. and Silverman, B. W. 2002b. *Functional data analysis*. Springer, New York.
- Ramsay, J. O. and Silverman, B. W. 2005. *Applied functional data analysis: methods and case studies* (second edition). Springer, New York.
- Ramsay, J. O., Hooker, G. and Graves, S. 2009. *Functional data analysis with R and MATLAB*. Springer, New York.

- Ray, S. and Mallick, B. 2006. Functional clustering by Bayesian wavelet methods. *Journal of the Royal Statistical Society. Series B*, 68:2, 305 - 332.
- Rossi, F., Conan-Guez, B. and GolliA, E. L. 2004. Clustering Functional Data with the Som Algorithm, In: Proceedings of ESANN, 305 – 312, Bruges.
- Samé, A., Chamroukhi, F., Govaert, G. and Aknin, P. 2011. Model-based clustering and segmentation of times series with changes in regime. *Advances in Data Analysis and Classification*, 5:4, 301 - 322.
- Saporta, G. 1981. Méthodes exploratoires d'analyse de données temporelles. *Cahiers du BURU*, 37 - 38.
- Satterthwaite, F. E. 1946. An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2:6, 110–114.
- Shen, Q. and Faraway, J. 2004. An F-test for linear models with functional responses. *Statistica Sinica*, 14:4, 1239–1257.
- Silverman, B. W. 1995. Incorporating parametric effects into functional principal components analysis. *Journal of the Royal Statistical Society, Series B*, 57, 673-689.
- Silverman, B. W. 1996. Smoothed functional principal component analysis by choice of norm, *The Annals of Statistics*, 24(1).
- Tarpey, T. and Kinader, K. 2003. Clustering functional data. *Journal of Classification*, 20:1, 93 - 114.
- Tian, S. 2010. Functional data analysis in brain imaging studies. *Quantitative Psychology and Measurement*, 1:35. doi: 10.3389/fpsyg.2010.00035
- Tokushige, S., Yadohisa, H. and Inada, K. 2007. Crisp and fuzzy k-means clustering algorithms for multivariate functional data. *Computational Statistics*, 22, 1 - 16.
- Ulbricht, J. 2004. Representing Functional Data as Smooth Functions. Master Thesis, Institute of Statistics and Econometrics, Humboldt University, Berlin.
- Wahba, G. 1990. *Spline models for observational data*. Society for Industrial and Applied Mathematics, Philadelphia.
- Xu, H., Shen, Q., Yang, X., and Shoptaw, S. 2011. A quasi F-test for functional linear models with functional covariates and its application to longitudinal data. *Statistics in Medicine*, 30:23, 2842 – 2853.
- Zhang, C., Peng, H., and Zhang, J.-T. 2010. Two sample tests for functional data. *Communications in Statistics–Theory and Methods*, 39:4, 559–578.
- Zhang, J. T. 2005. Approximate and asymptotic distributions of chi-squared type mixtures with applications. *Journal of the American Statistical Association*, 100, 273–285.

- Zhang, J. T. 2011a. Statistical inferences for linear models with functional responses. *Statistica Sinica*, 21, 1431–1451.
- Zhang, J. T. 2011b. Two-way MANOVA with unequal cell sizes and unequal cell covariance matrices. *Technometrics*, 53:4, 426–439.
- Zhang, J. T. 2013a. Functional linear models with time-dependent covariates (unpublished)
- Zhang, J. T. 2013b. Multi-sample equal covariance function testing for functional data (unpublished)
- Zhang, J. T. 2013c. *Analysis of variance for functional data*. Monographs on Statistics and Applied Probability 127.
- Zhang, J. T. and Liang, X. 2013. One-way ANOVA for functional data viaglobalizing the pointwise F-test (unpublished)
- Zhang, J. T. and Sun, Y. 2010. Two-sample test for equal covariance function for functional data. *Oriental Journal of Mathematics*, 4, 1–22.
- Zhang, J. T., Cheng, M. Y., Wu, H. T. and Zhou, B. 2018. A new test for functional one-way ANOVA with applications to ischemic heart screening. *Computational Statistics and Data Analysis*. doi:10.1016/j.csda.2018.05.004
- Zhang, J. T., Liang, X., and Xiao, S. 2010. On the two-sample Behrens-Fisher problem for functional data. *Journal of Statistical Theory and Practice*, 4:4, 571–587.
- Günlük Ortalama Sıcaklık, Günlük Toplam Yağış Miktarı, Günlük Ortalama Sıcaklık Farkı, Günlük Güneşlenme Süresi, Günlük Buharlaştırma Miktarı, 2018. Ankara Meteoroloji Genel Müdürlüğü.
- URL-1: <https://cran.r-project.org/web/packages/fda/index.html> (Erişim Tarihi: 15.05.2018)
- URL-2: <https://cran.r-project.org/web/packages/fda.usc/index.html> (Erişim Tarihi: 15.05.2018)
- URL-3: <https://cran.r-project.org/web/packages/funHDDC/index.html> (Erişim Tarihi: 30.08.2018)
- URL-4: <https://cran.r-project.org/web/packages/fdANOVA/index.html> (Erişim Tarihi: 10.12.2018)

ÖZGEÇMİŞ

Adı ve Soyadı : Çağlar SÖZEN

Doğum Yeri : Üsküdar

Doğum Tarihi : 06/03/1988

Yabancı Dili : İngilizce

Eğitim Durumu

Lise : Çorum Atatürk Lisesi(2006)

Lisans : Ondokuz Mayıs Üniversitesi Fen Edebiyat Fakültesi
İstatistik Bölümü (2007-2011)

Yüksek Lisans : Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı (Eylül 2012 – Ağustos 2014)

Çalıştığı Kurum/Kurumlar ve Yıl

Giresun Üniversitesi Bankacılık ve Finans Bölümü, Öğretim Görevlisi (2015 - ...)