

**ÇEVRE EPİDEMİYOLOJİSİNDE HASTALIKLARIN  
MODELLENMESİ İÇİN ZAMAN SERİSİ  
REGRESYON ANALİZİ YÖNTEMLERİNİN  
KARŞILAŞTIRILMASI**

**Mehmet KARADAĞ**  
İnönü Üniversitesi ve Mersin Üniversitesi  
Biyoistatistik ve Tıp Bilişimi Anabilim Dalı  
Ortak Doktora Programı

**Tez Danışmanı: Prof. Dr. Saim YOLOĞLU**  
**Ortak Tez Danışmanı: Prof. Dr. Bahar TAŞDELEN,**  
**İkinci Tez Danışmanı: Doç. Dr. Seval KUL**

**Doktora Tezi 2018**

**T.C.  
İNÖNÜ ÜNİVERSİTESİ  
SAĞLIK BİLİMLERİ ENSTİTÜSÜ**

**ÇEVRE EPİDEMİYOLOJİSİNDE HASTALIKLARIN MODELLENMESİ İÇİN  
ZAMAN SERİSİ REGRESYON ANALİZİ YÖNTEMLERİNİN  
KARŞILAŞTIRILMASI**

**Mehmet KARADAĞ**

**İnönü Üniversitesi ve Mersin Üniversitesi  
Biyoistatistik ve Tıp Bilişimi Anabilim Dalı  
Ortak Doktora Programı**

**Tez Danışmanı  
Prof. Dr. Saim YOLOĞLU**

**Ortak Tez Danışmanı  
Prof. Dr. Bahar TAŞDELEN**


**İkinci Tez Danışmanı  
Doç. Dr. Seval KUL**

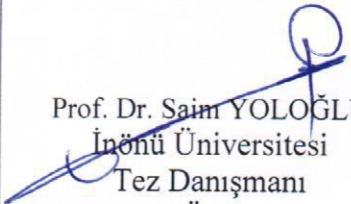
**MALATYA  
2018**

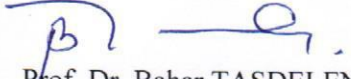
## KABUL VE ONAY SAYFASI


İnönü Üniversitesi ile Mersin Üniversitesi Sağlık Bilimleri Enstitüleri Biyoistatistik ve Tıp Bilişimi Anabilim Dalı Ortak Doktora Programı çerçevesinde yürütülmüş olan; **Mehmet KARADAĞ 'ın** “**Çevre Epidemiyolojisinde Hastalıkların Modellenmesi İçin Zaman Serisi Regresyon Analiz Yöntemlerinin Karşılaştırılması**” konulu bu çalışması, aşağıdaki jüri tarafından Doktora tezi olarak kabul edilmiştir.


Tez Savunma Tarihi: 17/10/2018

  
Prof. Dr. Cemil ÇOLAK  
İnönü Üniversitesi  
Jüri Başkanı

  
Prof. Dr. Saim YOLOĞLU  
İnönü Üniversitesi  
Tez Danışmanı  
Üye

  
Prof. Dr. Bahar TAŞDELEN  
Mersin Üniversitesi  
Ortak Tez Danışmanı  
Üye

  
Prof. Dr. Vildan SÜMBÜLOĞLU  
Sanko Üniversitesi  
Üye

  
Dr. Öğr. Üyesi Harika G.G. BAĞ  
İnönü Üniversitesi  
Üye

### ONAY

Bu tez, İnönü Üniversitesi Lisansüstü Eğitim-Öğretim Yönetmeliği'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından kabul edilmiş ve Enstitü Yönetim Kurulu'nun ..../..../2018 tarih ve 2018/.....sayılı Kararıyla da uygun görülmüştür.

Prof. Dr. Yusuf TÜRKÖZ  
Enstitü Müdürü

# İÇİNDEKİLER

ÖZET .....	vi
ABSTRACT.....	vii
SİMGELER VE KISALTMALAR DİZİNİ .....	viii
ŞEKİLLER DİZİNİ .....	ix
TABLolar DİZİNİ.....	x
1. GİRİŞ.....	1
2. GENEL BİLGİLER .....	3
2.1 Regresyon ve Genelleştirilmiş Lineer Modeller .....	3
2.2. Genelleştirilmiş Eklemeli Model (Generalized Additive Model-GAM).....	5
2.2.1. Düzgünleştirme (smoothing) Mantığı.....	7
2.2.2. GAM’da Sık Kullanılan Düzgünleştirme Fonksiyonları .....	10
2.2.2.1. Doğal Kübik Zincir Yapısı ve İnterpolasyon Mantığı .....	10
2.2.2.2. Kübik Düzgünleştirme Zincirleri-KRZ (Cubic smoothing splines) .....	11
2.2.2.3. Döngüsel Kübik Regresyon Zinciri-DKRZ (Cyclic cubic regression spline- CCRS).....	12
2.2.2.4. P-Zincirleri (P-Splines).....	13
2.2.2.5. İnce Tabaka Regresyon Zincirleri-İTRZ (Thin-Plate Regression Spline- TPRS) .....	13
2.2.2.6. Duchon Zinciri-DZ (Duchon Spline-DS) .....	14
2.2.2.7. Küre Zincirleri-KZ (Spline on the Sphere).....	15
2.2.3. Düzgünleştirme Parametresinin Seçimi .....	15
2.2.3.1. Düzgünleştirme Parametresine Çapraz Geçerlilik (CV) ve Genelleştirilmiş Çapraz Geçerlilik (GCV) Yöntemleri ile Karar Vermek.....	16
2.2.3.2. Düzgünleştirme Parametresine UBRE Yöntemi ile Karar Vermek.....	18
2.2.4. Temel Boyut ( $k$ ) Seçimi .....	18
2.2.5. Etkili Serbestlik Derecesi (ESD) .....	19
2.2.6 GAM’da Geriye Doğru En Uygun Modeli Oluşturma Algoritması .....	20
2.3. Dağıtılmış Gecikmeli Doğrusal Olmayan Modeller-DGDOM (Distributed Lag Non-Linear Models -DLNM) .....	22
2.3.1. DGDOM’de $s$ Düzgünleştirme Fonksiyonu Yapısı .....	24
2.3.2. DGDOM Cebirsel Yapısı ve Çapraz Taban Konsepti .....	25

2.4. Otoresif Terimli Genelleştirilmiş Eklemeli Modeller (Generalized Additive Model with Autoregressive Terms – GAMAR) .....	27
3. MATERYAL VE METOT .....	30
3.1. Çalışmada Kullanılan Veri Seti .....	30
3.1.1. Genelleştirilmiş Eklemeli Modeller için Kurulan Model Yapıları .....	32
3.1.2. Model Başarısının Test Edilmesi İçin Kullanılan Yöntemler .....	34
4. BULGULAR .....	37
4.1. Genelleştirilmiş Eklemeli Modellerle Yapılan Analizlerin Sonuçları .....	38
4.1.1. Astım Hastalığı ile İlgili Bulgular .....	38
4.1.2. KOAH Hastalığı ile İlgili Bulgular .....	42
4.1.3. Pnömoni Hastalığı ile İlgili Bulgular .....	46
4.1.4. Pulmoner Tromboemboli (PTE) Hastalığı ile İlgili Bulgular .....	50
4.2 Dağıtılmış Gecikmeli Doğrusal Olmayan Modellerle (DGDOM) Yapılan Analizlerin Sonuçları .....	54
4.2.1. Astım Hastalığı ile İlgili Bulgular .....	54
4.2.2. KOAH Hastalığı ile İlgili Bulgular .....	57
4.2.3. Pnömoni Hastalığı ile İlgili Bulgular .....	61
4.2.4. Pulmoner Tromboemboli (PTE) hastalığı ile ilgili bulgular .....	65
4.3. Otoresif Terimli Genelleştirilmiş Eklemeli Modellerle (GAMAR) yapılan analizlerin sonuçları .....	69
4.3.1. Astım hastalığı ile ilgili bulgular .....	69
4.3.2. KOAH hastalığı ile ilgili bulgular .....	70
4.3.3. Pnömoni hastalığı ile ilgili bulgular .....	72
4.3.4. Pulmoner Tromboemboli (PTE) hastalığı ile ilgili bulgular .....	73
5. TARTIŞMA .....	79
6. SONUÇ VE ÖNERİLER .....	84
KAYNAKLAR .....	85
EKLER .....	93
EK.1. ÖZGEÇMİŞ .....	93
EK.2. ETİK KURUL ONAY FORMU .....	96

## TEŐEKKÜR

Bu eserin ortaya ıkma s¼recinde bilgisini ve deneyimlerini benimle paylaŐan, bana her konuda destek olan ve y¼n g¼steren danıŐman hocam Sayın Prof. Dr. Saim YOLOĐLU'na ve ikinci danıŐmanım Prof. Dr. Bahar TAŐDELEN'e teŐekk¼r ve saygılarımı sunarım. İstatistik eĐitim hayatımın ilk g¼nlerinden beri gerek bilgisi gerek deneyimleri ile bana yol g¼steren, benimle birlikte sevinen ve birlikte kaygılanan ortak tez danıŐman hocam Sayın Do. Dr. Seval KUL'a teŐekk¼rlerimi sunarım. Bana her konuda destek olan ve yanımda duran eŐim Arzu ve oĐlum Deniz'e, yine hibir zaman eksikliklerini hissetmediĐim ve desteklerini esirgemeyen annem, babam ve kardeŐime sonsuz teŐekk¼rlerimi sunarım.

Mehmet KaradaĐ

## ÖZET

### Çevre Epidemiyolojisinde Hastalıkların Modellenmesi İçin Zaman Serisi Regresyon Analizi Yöntemlerinin Karşılaştırılması

**Amaç:** Bu çalışmada Astım, KOAH, Pnömoni ve Pte hastalıklarının 3 farklı sonuç değişkeni GAM, DGDOM ve GAMAR yöntemleri kullanılarak modellenmiş ve yöntemlerin performans sonuçları karşılaştırılarak en iyi model araştırılmıştır.

**Materyal ve Metot:** Gerçek verimiz, Gaziantep ili Kamu Hastaneleri Genel Sekreterliğine bağlı 3 hastaneden 01 Ocak 2009 ile 31 Mart 2014 tarihleri arasında toplam 1916 gün boyunca geriye dönük olarak izlenmesi ile elde edilen veriler kullanılarak oluşturuldu. Cevap değişkenleri; astım, KOAH, pnömoni ve pulmoner tromboemboli şikâyeti ile “acil ve göğüs hastalıkları servislerine başvurular”, “hastanede yatış” ve “mortalite” durumlarının günlük sayımı şeklindedir. Tahminlerde GAM, DGDOM ve GAMAR yöntemleri kullanılmış, aynı yöntemle kurulan 4 farklı modelden en iyi performansa sahip model, ilgili yöntem için karşılaştırma modeli olarak kullanılmıştır. Tahminlerin performansının değerlendirilmesinde AIC, düzeltilmiş AIC, BIC, fark yüzdesi ve düzeltilmiş  $R^2$  kriterleri kullanılmıştır.

**Bulgular:** Astım, KOAH, pnömoni ve PTE hastalıklarının başvuru ve hastanede yatış cevap değişkenlerinin tahmininde performans kriterlerine göre GAM ve DGDOM benzer sonuçlar gösterirken, GAMAR metodunun diğer yöntemlere göre daha üstün sonuçlar verdiği gözlenmiştir. Hastalıkların eks cevap değişkenlerinde ise; astım hastalığında DGDOM diğer modellere göre üstün performans gösterse de diğer hastalıklarda yine GAMAR’ın üstün olduğu gözlenmiştir.

**Sonuç:** İncelenen bağımlı değişkenler ile oluşturulan modeller kıyaslandığında, en iyi performansın GAMAR yönteminden elde edildiği gözlenmiştir. Bunun yanında OR grafiklerinde en düşük standart hatalı grafikler GAMAR yöntemi kullanılan modellerde gözlenmiştir.

**Anahtar Kelimeler:** GAM, DGDOM, GAMAR, Çevre Epidemiyolojisi, Hava Kirliliği.

## ABSTRACT

### **Comparison of Time Series Regression Models in Environmental Epidemiology for Diseases Modeling**

**Aim:** In this study, 3 different outcome variables of Asthma, COPD, Pneumonia and Pte diseases were modeled by using GAM, DGDOM and GAMAR methods and the best model was investigated by comparing the performance results of the methods.

**Material and Methods:** The data was obtained from 3 hospitals under the General Secretariat of Gaziantep province public hospitals retrospectively for a total of 1916 days between 01 January 2009 and 31 March 2014. Response variables were number of the admission, hospitalization, and mortality due to asthma, COPD, pneumonia and pulmonary thromboembolism. The response variables were estimated by GAM, DGDOM and GAMAR methods by building 4 different models and the performances of the models were compared.

**Results:** GAM and DGDOM showed similar results according to the performance criteria in predicting the recourse and hospitalization response variables of asthma, COPD, pneumonia and PTE diseases. It was observed that GAMAR method gave superior results to other methods. In the mortality response variables of the diseases; although DGDOM gave superior performance to other models in asthma disease, GAMAR was also superior in other diseases.

**Conclusion:** When the models created with dependent variables was compared, it was observed that the best performance was obtained from the GAMAR method. However, the lowest standard error graphics in the OR charts were observed in the models using the GAMAR method.

**Key words:** GAM, DGDOM, GAMAR, Environmental Epidemiology, Air Pollution.



## SİMGELER VE KISALTMALAR DİZİNİ

<b>AIC</b>	: Akaike Bilgi Ölçütü
<b>AICc</b>	: Düzeltilmiş Akaike Bilgi Ölçütü
<b>BIC</b>	: Bayesci Bilgi Kriteri
<b>CV</b>	: Çapraz Geçerlilik
<b>DGDOM</b>	: Dağıtılmış Gecikmeli Doğrusal Olmayan Modeller
<b>DGM</b>	: Dağıtılmış Gecikmeli Modeller
<b>DKRZ</b>	: Döngüsel Kübik Regresyon Zinciri
<b>ESD</b>	: Etkili Serbestlik Derecesi
<b>GA</b>	: Güven Aralığı
<b>GAM</b>	: Genelleştirilmiş Eklemeli Modeller
<b>GAMAR</b>	: Otopregresif Terimli Genelleştirilmiş Eklemeli Modeller
<b>GARMA</b>	: Genelleştirilmiş Otopregresif Hareketli Ortalama Serileri
<b>GCV</b>	: Genelleştirilmiş Çapraz Geçerlilik
<b>GLM</b>	: Genelleştirilmiş Lineer Modeller
<b>HKO</b>	: Hata Kareler Ortalaması
<b>HKT</b>	: Hata Kareler Toplamı
<b>İT</b>	: İnce Tabaka
<b>İTRZ</b>	: İnce Tabaka Regresyon Zincirleri
<b>KBRZ</b>	: Kübik Regresyon Zinciri
<b>OR</b>	: Odds Oranı
<b>P-IRLS</b>	: Ceza İteratif Yeniden Ağırlıklandırılmış En Küçük Kareler Yöntemi
<b>PTE</b>	: Pulmoner Tromboemboli
<b>SD</b>	: Serbestlik Derecesi
<b>UBRE</b>	: Yansız Risk Tahmini

## ŞEKİLLER DİZİNİ

<b><u>Sekil No</u></b>	<b><u>Sayfa No</u></b>
<b>Şekil 2.1.</b> Polinom temelli fonksiyon kullanarak bir düzgünleştirici fonksiyonun gösterimi .....	8
<b>Şekil 2.2.</b> Kübik zincir yapısının iki boyutlu düzlemde gösterilmesi.....	9
<b>Şekil 2.3.</b> Döngüsel kübik regresyon zincir yapısının gösterimi .....	12
<b>Şekil 4.1.</b> Astım, KOAH ve Pnömoni hastalıklarından başvuru sayıları üzerine PM10 ve gecikme etkilerinin OR (%95 GA) grafikleri .....	76
<b>Şekil 4.2.</b> Astım, KOAH, Pnömoni ve PTE hastalıklarından yatan sayıları üzerine PM10 ve gecikme etkilerinin OR (%95 GA) grafikleri .....	77
<b>Şekil 4.3.</b> Astım, KOAH, Pnömoni ve PTE hastalıklarından yatan sayıları üzerine PM10 ve gecikme etkilerinin OR (%95 GA) grafikleri .....	78

## TABLULAR DİZİNİ

<b>Tablo No</b>	<b>Sayfa No</b>
<b>Tablo 3.1.</b> Çalışmada kullanılan bağımlı değişkenlerin tanımları.....	30
<b>Tablo 3.2.</b> Çalışmada kullanılan bağımsız değişkenlerin tanımlayıcı bilgiler .....	31
<b>Tablo 4.1.</b> Çalışmada kullanılan Nümerik Ölçümlü bağımsız değişkenlerin tanımlayıcı istatistikleri.....	37
<b>Tablo 4.2.</b> Bağımlı değişkenlere ait tanımlayıcı istatistikler.....	37
<b>Tablo 4.3.</b> Astım ile hava kirliliği ilişkisinin GAM'ın farklı modelleri ile incelenmesi .....	39
<b>Tablo 4.4.</b> Astım GAM modellerinin Anova Gam değerlendirmeleri .....	39
<b>Tablo 4.5.</b> Astım hastalığı bağımlı değişkenlerinin GAM yöntemi ile en iyi sonucu veren modellerinin uyum iyiliği test sonuçları .....	41
<b>Tablo 4.6.</b> KOAH ile hava kirliliği ilişkisinin GAM'ın farklı modelleri ile incelenmesi .....	43
<b>Tablo 4.7.</b> KOAH vakasına ait Modellerin Anova Gam değerlendirmeleri .....	43
<b>Tablo 4.8.</b> KOAH hastalığı bağımlı değişkenlerinin GAM yöntemi ile en iyi sonucu veren modellerinin uyum iyiliği test sonuçları .....	45
<b>Tablo 4.9.</b> Pnömoni ile hava kirliliği ilişkisinin GAM'ın farklı modelleri ile incelenmesi .....	47
<b>Tablo 4.10.</b> Pnömoni vakasına ait modellerin Anova Gam değerlendirmeleri.....	47
<b>Tablo 4.11.</b> Pnömoni hastalığı bağımlı değişkenlerinin GAM yöntemi ile en iyi sonucu veren modellerinin uyum iyiliği test sonuçları.....	49
<b>Tablo 4.12.</b> PTE ile hava kirliliği ilişkisinin GAM'ın farklı modelleri ile incelenmesi .....	50
<b>Tablo 4.13.</b> PTE vakasına ait Modellerin Anova Gam değerlendirmeleri .....	51
<b>Tablo 4.14.</b> PTE hastalığı bağımlı değişkenlerinin GAM yöntemi ile en iyi sonucu veren modellerinin uyum iyiliği test sonuçları .....	52
<b>Tablo 4.15.</b> PM10 değişkeninin 1 mg/m <sup>3</sup> artışının GAM ile OR tahmini .....	53
<b>Tablo 4.16.</b> Astım ile hava kirliliği ilişkisinin DGDOM'nin farklı modelleri ile incelenmesi .....	55
<b>Tablo 4.17.</b> Astım DGDOM modellerinin Anova Gam değerlendirmeleri .....	55

<b>Tablo 4.18.</b> Astım hastalığı bağımlı değişkenlerinin DGDOM yöntemi ile en iyi sonucu veren modellerinin uyum iyiliği test sonuçları.....	57
<b>Tablo 4.19.</b> KOAH ile hava kirliliği ilişkisinin DGDOM'nin farklı modelleri ile incelenmesi .....	58
<b>Tablo 4.20.</b> KOAH vakası DGDOM modellerinin Anova Gam değerlendirmeleri .....	59
<b>Tablo 4.21.</b> KOAH hastalığı bağımlı değişkenlerinin DGDOM yöntemi ile en iyi sonucu veren modellerinin uyum iyiliği test sonuçları.....	60
<b>Tablo 4.22.</b> Pnömoni ile hava kirliliği ilişkisinin DGDOM'nin farklı modelleri ile incelenmesi .....	62
<b>Tablo 4.23.</b> Pnömoni vakası DGDOM modellerinin Anova Gam değerlendirmeleri ..	62
<b>Tablo 4.24.</b> Pnömoni hastalığı bağımlı değişkenlerinin DGDOM yöntemi ile en iyi sonucu veren modellerinin uyum iyiliği test sonuçları.....	64
<b>Tablo 4.25.</b> PTE ile hava kirliliği ilişkisinin DGDOM'ın farklı modelleri ile incelenmesi .....	65
<b>Tablo 4.26.</b> Pnömoni vakasına ait Modellerin Anova Gam değerlendirmeleri .....	66
<b>Tablo 4.27.</b> PTE hastalığı bağımlı değişkenlerinin DGDOM yöntemi ile en iyi sonucu veren modellerinin uyum iyiliği test sonuçları.....	67
<b>Tablo 4.28.</b> PM10 değişkeninin 1 mg/m <sup>3</sup> artışının DGDOM ile OR tahmini .....	68
<b>Tablo 4.29.</b> Astım ile hava kirliliği ilişkisinin GAMAR'nin farklı modelleri ile incelenmesi .....	69
<b>Tablo 4.30.</b> KOAH ile hava kirliliği ilişkisinin GAMAR'nin farklı modelleri ile incelenmesi .....	71
<b>Tablo 4.31.</b> Pnömoni ile hava kirliliği ilişkisinin GAMAR'nin farklı modelleri ile incelenmesi .....	72
<b>Tablo 4.32.</b> PTE ile hava kirliliği ilişkisinin GAMAR'ın farklı modelleri ile incelenmesi .....	74
<b>Tablo 4.33.</b> PM10 değişkeninin 1 mg/m <sup>3</sup> artışının GAMAR ile OR tahmini .....	75

# 1. GİRİŞ

Sağlık alanında istatistiksel modelleme yöntemlerinin kullanılmaya başlanması ve tahmin yöntemlerinde gelişmeler ile beraber birçok hastalığı etkileyen faktörler önceden tahmin edilebilmekte ve bu doğrultuda hastalıkların engellenmesi kolaylaşmaktadır. Küresel ısınma, sera gazları, çöl fırtınaları ve partiküler madde gibi hava kirliliği olayları insan sağlığını ciddi anlamda tehdit etmeye başlamıştır (1). Şehirlerimiz gittikçe kalabalıklaşıyor. Bu düzensiz kalabalıklaşma hava kirliliğini de hızlıca artırıyor. Dünya Sağlık Örgütü geçen yıl 7 milyon kişinin ölüm sebebinin hava kirliliği olduğunu raporlamıştır. Raporun Türkiye bölümünde Gaziantep ili, havası en kötü iller arasında ilk sıralarda yer almaktadır (2, 3). Tüm bölgeler iklim değişikliğinden etkilenir fakat Ortadoğu ülkeleri uzun sıcak mevsimler ve sınırlı doğal su rezervleri nedeniyle küresel iklim değişikliğine karşı en savunmasız bölgelerden biridir (4). Çalışmalar çöl fırtınalarının ve sıcaklık artışının özellikle Sahra çölü ve Akdeniz bölgesinde daha belirgin hale geldiğini göstermektedir. Sıcaklık dalgaları ve kuraklığa ek olarak, Orta Doğu, Güneydoğu, Asya'da ve Batı Pasifik'teki hızla büyüyen şehirler bu bölgedeki kirlilik düzeyini, beş ila on kat arttığı görülmektedir (5). Epidemiyolojik olarak, Ortadoğu ve Karayip toplulukları gibi çöl tozu fırtınası, sıcaklık, partiküler madde ve benzeri etkenlere maruz kalanlarda astım, KOAH, pnömoni benzeri solunum yolu hastalıkları, beklenenden daha sık görülmektedir (6). Son yıllarda hızla ivme gösteren kirlilik düzeyleri ve bu konunun insan sağlığına tehdidi, araştırmacılara bu mesele üzerinde daha fazla çalışma yapma ihtiyacı doğurmuştur (1, 4).

Literatürde hava kirliliği ile hastalıklar üzerine yapılmış birçok çalışmaya rastlanmaktadır. Yapılan çalışmalarda, genelde modelleme yöntemleri ile tahminler yapıldığı görülmüştür. Kullanılan modellerde cevap değişkeni hastaların servise başvurusu, serviste yatma veya eks olma durumlarının günlük olarak sayımları şeklindedir (7). Hava kirliliğinin insan sağlığını olumsuz etkilediği düşünülen faktörler modele alınmıştır. Bunlardan bazıları günlük ortalama sıcaklık (°C), PM10 (partikül madde) ( $\mu\text{g}/\text{m}^3$ ), nem, çöl fırtınası varlığı, ozon tabakası ölçümü ve olayın gözlemlendiği günün hafta içi ya da hafta sonu olması durumlarıdır (8). Aynı zamanda literatürde yapılan incelemede cinsiyet ve yaş grupları için modellerin etkinliklerinin karşılaştırıldığı görülmüştür (9).

Regresyon analizi ile hava kirliliği için zamanın etkisini, modele gecikmeler (lag) ekleyerek kontrol edilebilir. Böylelikle, örneğin solunan partiküler madde miktarının önceki gün etkileri ile günlük ölüm sayısı tahmin edilebilir. Literatürde cevap değişkeninin trendini açıklayıcı değişkenlerle, modelleme yeteneğine sahip birçok yöntem bulunmaktadır. Genelleştirilmiş Lineer Modeller (GLM) bu modellerden birisidir. 1990 yılında Hastie ve Tibshirani tarafından bulunmuş olan Genelleştirilmiş Eklemeli Modeller (GAM), GLM modellerinin daha gelişmiş formudur. Bu model lineer olmayan yapıları tahmin edebilme yeteneğine sahiptir. Ayrıca GAM yönteminde değişkenler düzgünleştirme fonksiyonları ile modele eklenebilmektedir (10). Dağıtılmış Gecikmeli Doğrusal Olmayan Modeller (DGDOM) yaklaşımı, son yıllarda hava kirliliğinin sağlık üzerine etkilerini ölçmek için geliştirilen istatistiksel yöntemlerdendir. Bu model mevsimsellik ve uzun vadeli faktörlerin zamanla değişen karıştırıcı etkisini gecikme yapısı ile modele tek bir tahmin edici olarak ekleme özelliği ile öne çıkmaktadır (11). Zaman serisi tahmin edicisi ardışık günlerde ölçüldüğünden cevap değişkeninin zamana karşı bir otokorelasyonuna sebep olmaktadır. Bu problem ile başa çıkabilen özelliğe sahip GAMAR (Oto-regresif Terimli Genelleştirilmiş Eklemeli Modeller) yöntemi tahmin başarısı yüksek bir modelleme yöntemidir (12).

Bu çalışmanın birincil amacı; çevre epidemiyolojisinde hastalıkları modelleyebilen GAM, DGDOM ve GAMAR yöntemlerinin gerçek veri kullanılarak, her yöntemin farklı serbestlik derecesi ve farklı düzgünleştirme fonksiyonları ile önce kendi içlerinde daha sonrada kendi aralarında performanslarını karşılaştırıp tahmin başarısı en iyi olan yöntem ve modellerine karar vermektir. İkincil amacı ise en iyi tahmin başarısına sahip modellerin, PM10 ana etkisi ve gecikmeleri üzerinden (OR %95 GA) tahmin başarılarını kıyaslamaktır.

## 2. GENEL BİLGİLER

### 2.1 Regresyon ve Genelleştirilmiş Lineer Modeller

Biyoistatistik bilim dalının da dâhil olduğu birçok farklı bilim dalı bağımlı değişken/çıktı/cevap (dependent variable/outcome/response) ile bağımsız/ortak/tahmin edici değişken (independent variable/covariate/predictor) arasındaki ilişki ile ilgilenir. Korelasyon ve regresyon kavramları ilk olarak Francis Galton tarafından 1886 yılında öne sürülmüştür (13). Galton akademik kariyeri boyunca parmak izinden meteorolojiye kadar birçok farklı alanda çalışmış olmasına karşın, ününü kalıtım alanında yaptığı çalışmalar ile kazanmıştır (14). Galton kalıtımsal olarak boy özelliklerinin nesillerden nesillere nasıl aktarıldığını anlayabilmek için, bireylerden ve ebeveynlerinden boy bilgilerini toplamıştır. Bu çalışmaları sırasında Galton, deneklerin boy uzunlukları ile ebeveynlerinin boy uzunlukları arasındaki ilişkileri basit lineer regresyon modeli kullanarak başarılı bir şekilde ortaya koymuştur (13). Bu çalışma tıp alanındaki kazandırdığı yeni bilginin yanı sıra, yıllarca kullanılacak olan regresyon analizi yöntemlerinin de başlangıcı olmuştur. Regresyon bir bağımlı değişken ile bağımsız değişken ya da değişkenler arasındaki ilişkiyi matematiksel modellerle ifade etme şekli olarak tanımlanabilir (15). Bunun yanı sıra elde edilen model yardımıyla bağımsız değişkenler kullanılarak bağımlı değişkenin değerlerinin de tahmin edilmesine de olanak tanır (16).

Aldıkları değerler bireyden bireye ve zamandan zamana farklılık gösteren özellikler, değişken olarak adlandırılır (17). Değişkenler elde edilışlerine göre nominal, ordinal, kesikli ve sürekli olarak sınıflandırılabilirler (16). İstatistiksel yöntemler bağımlı ve bağımsız değişkenlerin tipi, dağılımı ve sayısına bakılarak seçilir (18). İstatistiksel modellemede en az iki değişken arasındaki ilişki araştırılır. Bu değişkenlerden biri bağımlı değişken ve diğeri ise bağımsız değişkendir. Modelleme yapılırken bağımlı değişkenin değerlerinin bağımsız değişkenin düzeylerine göre değiştiği varsayılır (19). Bir bağımlı ve bir bağımsız değişken arasındaki ilişkinin araştırıldığı modeller basit modeller olarak adlandırılır. Bazı araştırmalarda bir bağımsız değişken bağımlı değişkeni açıklamakta yetersiz kalmaktadır. Bu durumda, birden fazla bağımsız değişken kullanılarak çoklu regresyon modelleri oluşturulmalıdır (20). Regresyon modellerinin adını belirleyen bazı önemli faktörler vardır. Bunlardan birisi yukarıda bahsedildiği gibi bağımsız değişkenin sayısıdır. Bir diğerk faktör ise bağımlı değişkenin türü veya

dağılımıdır. Örneğin bağımlı değişken binary yapıda olduğu durumda kullanılan model lojistik regresyon olarak adlandırılırken (21), bağımlı değişken ikiden çok kategoride sırasız (nominal) nitel düzeyde ölçümlere sahip olduğunda ise kullanılan çözümlene yöntemine çok değişkenli (multinomial) lojistik regresyon yöntemi denilmektedir (20). Buna ek olarak bağımlı değişkenin kesikli değerler aldığı Poisson dağılımına uyduğu modellerle de karşılaşmak mümkündür. Bu modeller Poisson regresyonu yöntemi ile çözümlenmektedir (22). Ayrıca sürekli ve normal dağılıma sahip bir bağımlı değişken için oluşturulan model ise lineer regresyon modeli olarak adlandırılır (20). Model seçiminde önemli rol oynayan diğer faktör ise bağımlı ve bağımsız değişkenler arasındaki bağlantı fonksiyonunun türüdür.

Bağımlı değişkenin normal dağılıma sahip olmadığı ve bağımsız değişken veya değişkenler ile ilişkilerinin basit lineer bir yapıda olmadığı durumlarda lineer modeller yerine Genelleştirilmiş Lineer Modeller (GLM) kullanılır (23). GLM’de en önemli varsayım, bağımlı değişkenin üstel dağılım ailesinden olan Normal, Binom, Poisson ve Gama gibi dağılımlardan birine sahip olmasıdır (24). Üstel dağılım ailesinden herhangi birine sahip bir fonksiyon düşünelim. Örneğin bu fonksiyon  $f(y; \theta) = s(y)t(\theta)e^{a(y)b(\theta)}$  şeklinde olsun bu fonksiyonda  $a, b, s$  ve  $t$  bilinen fonksiyonlar olsun, örnek verdiğimiz fonksiyonu  $f(y; \theta) = \eta = \exp[a(y)b(\theta) + c(\theta) + d(y)]$  şeklinde tekrar yazabilir. Fonksiyon yapısında  $s(y) = \exp d(y)$  ve  $t(\theta) = \exp c(\theta)$  şekline dönüşmüştür (23). Burada  $a(y) = y$  eşitliğine kanonik form denilmektedir. Link fonksiyonu ise  $g(\mu_i) = X_i^T \beta$  şeklinde olsun,  $\mu_i = E(Y_i)$  olarak tanımlanır (23). Bu  $g(\mu)$  dönüşümü link (bağlantı) fonksiyonu olarak adlandırılır. Bu adlandırmanın nedeni, modelin lineer kısmı olan  $\eta$  ile tesadüfi tarafı olan  $\mu$  arasında bir bağlantı sağlamasındandır. Link fonksiyonlara bazı regresyon modelleri için örnek verecek olursak; normal ve gama regresyon için link fonksiyonu  $g(\mu) = \mu$ , Poisson regresyon  $g(\mu) = \log(\mu)$ , lojistik regresyon için  $g(\mu) = \log(\mu/1 - \mu)$  şeklindedir (24, 25). Farklı dağılımlar için farklı link fonksiyonları tanımlanır. Bunlara birkaç örnek verilecek olursa:

- Log link :  $\ln \mu$
- Ters (inverse) link :  $1/\mu$
- Kare-kök link :  $\sqrt{\mu}$
- Logit link :  $\ln\left(\frac{\pi}{1-\pi}\right)$
- Probit link :  $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha+\beta X} \exp\left(-\frac{1}{2}Z^2\right) dZ$



- Log-log link :  $\ln[-\ln(1 - \pi)]$
- Poisson :  $\frac{\mu^y}{y!} e^{-\mu}$
- Negative binomial :  $\frac{\Gamma(y_i + \omega)}{y! \Gamma(\omega)} \frac{\mu_i^{y_i} \omega^\omega}{(\mu_i + \omega)^{\mu_i + \omega}}$

Özetle link fonksiyonu farklı dağılım sergileyen bağımlı değişkeni modele uydurmak için kullanılan bir çeşit dönüşüm aracıdır (24, 26). Genelleştirilmiş lineer modeller 3 bileşenden oluşur;

- Birincisi cevap değişkeni  $Y_1, \dots, Y_N$
- İkincisi açıklayıcı değişkenler ve  $\beta$  parametre seti,
- Üçüncüsü ise kanonik formda monoton link fonksiyonudur

Poisson regresyon için GLM de nasıl tahmin yapıldığına dair bir örnek verecek olursak;

$E(Y_i) = \mu_i = \exp(X_i^T \beta)$  Kanonik link fonksiyona sahip Poisson GLM yoğunluk fonksiyonu;

$$f(y_i/x_i) = \exp\left\{\frac{X_i^T \beta y_i - \exp(X_i^T \beta)}{\phi} + c(y_i, \phi)\right\} \text{ şeklindedir. Formüldeki } c(y_i, \phi)$$

normalleştirme katsayısıdır. Poisson GLM yöntemi ile hesaplanan  $\hat{\beta}_{PGLM}$  log olabilirlik yöntemiyle hesaplanmış  $\beta$ 'ya göre maksimum düzeye çıkararak birinci dereceden koşullar ve  $\sum_{i=1}^n \frac{1}{\phi} (y_i - \exp(X_i^T \beta)) X_i = 0$  denkleminde hesaplanmaktadır (26).

GLM ile GAM arasındaki en büyük fark: GAM sadece bilinmeyen bir fonksiyonla ilgilenmez aynı zamanda fonksiyonların toplamına ilişkin bir yakınsamaya izin verir. Bu yapı Genelleştirilmiş eklemeli modeller algoritmasının yapıtaşdır (27).

## 2.2. Genelleştirilmiş Eklemeli Model (Generalized Additive Model-GAM)

İncelediğimiz veri net bir lineer ilişki göstermediği zamanlarda, eklemeli modelleme lineer regresyona uygun bir alternatif haline gelir (28). Araştırmacıların GAM yöntemini tercih etmelerinin önemli bir sebebi, tahmin edicilerle cevap değişkeni arasındaki katı lineer varsayımları hafifletebilmesidir (29). GAM açıklayıcı değişkenler (kovariyet) ve düzgünleştirme fonksiyona (smoothing functions) sahip değişkenlerin toplamının, bir lineer tahmin edici ile Genelleştirilmiş lineer modele dâhil edilmesidir (30). Aslında bu metot düzgünleştirme eğrilerini kullanarak bağımlı değişken ile bağımsız değişken arasındaki ilişkiyi modeller (28). Bir GAM aynı bir GLM gibi açıklayıcı değişken(ler)in düzgünleştirici fonksiyonları ile cevap değişkeninin ortalaması

arasındaki ilişkiyi kurmak için bir link fonksiyonu kullanır. GAM'ın stratejisi, açıklayıcı değişken seti ve cevap değişkeni arasındaki lineer ve monoton olmayan ilişkiyi çözebilme kabiliyetidir (31). GAM'ın bu kabiliyeti verideki bazı parametrik ilişkileri göz önünde bulundurmak yerine bağımsız değişken seti ile cevap değişkeni arasındaki ilişkinin doğasını doğrudan tanımlamasından ileri gelmektedir (32). Genel model yapısı şu şekildedir;

$$g(\mu_i) = X_i^* \theta + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}, x_{4i}) + \dots$$

Burada  $\mu_i \equiv \mathbb{E}(Y_i)$ ,  $Y_i \sim$  bazı üstel fonksiyonlar.  $Y_i$  bağımlı değişkendir,  $X_i^*$  her parametrik model bileşeni için model matrisinin bir satırı,  $\theta$  parametre vektörü ve  $f_i, x_k$  kovaryetlerinin düzgün fonksiyonlarıdır (30). Bu GAM modelindeki  $f_i$  fonksiyonu ya parametrik olmayan düzgünleştirici bir fonksiyondur ya da regresyon zinciri (spline) yapısına sahiptir (29). Modeldeki  $f_i$  fonksiyonu; lokal regresyon, regresyon zinciri ve düzgünleştirici zincirler gibi parametrik olmayan eğri tahmin yöntemlerini en optimal şekilde kullanır (33). GAM'da tüm açıklayıcı değişkenlerin bağımlı değişkeni, parametrik olmayan ve lineer düzgünleştirici fonksiyonlar yoluyla etkilediği varsayılmaktadır (10). Düzgünleştirici fonksiyonlar sayesinde detaylı parametrik ilişkiler, modelde daha anlaşılır bir şekilde açıklanır. Bu durum hantal ve kullanışsız model görüntüsünü ortadan kaldırır (30). Düzgünleştirici fonksiyon seçimi ve seviyesi GAM için önemli bir adımdır (28, 32). Bu adımın en iyi şekilde çözümlenmesi, düzgünleştirici fonksiyonlarının serbestlik derecesinin doğru seçimine de bağlıdır (32).

GLM ve GAM çevresel epidemiyoloji araştırmalarında sıklıkla kullanılır (12). Bağımlı değişken kesikli ölçüm düzeyine sahip olduğunda (örneğin: ölüm sayısı), GAM modeli logaritmik link fonksiyonuna sahip Poisson eklemeli regresyon modeli formunu alır (12). Bu modelin en önemli özelliklerinden birisi de, bağımlı değişken ve link fonksiyonlarına geniş yelpazede dağılımlardan herhangi birini seçebilme şansı vermesidir (10). Modeldeki bağımlı değişken  $Y_i$ 'nin  $\mu_i$  ortalamalı Poisson dağıldığı varsayılır (12). Üstel dağılım ailesinden link fonksiyonuna sahip model ya lineer bir kombinasyona sahiptir (ki bu durumda GLM) ya da logaritmik fonksiyon aracılığıyla düzgünleştirme fonksiyonuna (ki bu durumda GAM) sahiptir (12). Hava kirliliğinin etkilediği hastalıkların araştırıldığı çalışmalarda mevsimsellik, uzun vadeli trendler, otokorelasyon ve hava değişimini GAM'da standart parametrelerle ifade etmek zordur (34). Genelde bir hava kirliliği ölçümünün sağlığa etkisini belirlemek için yapılan araştırmalarda, sağlık ölçütü olarak kullanılan günlük ölüm sayısı ve benzeri hastalık sonucu ortaya çıkabilecek durumlar ile hava kirliliği ölçümleri arasında bir ilişki olduğu varsayılır. Bu durumda

modelde hem nitel hem de nicel deęişkenler yer alabilir. Model formuna örnek verecek olursak;

$$E(Y/X, T) = g(\mu) = \eta = \alpha + X\beta + \sum f_i(T_i),$$

Burada  $X$  bir matris olmak üzere, hava kirlilięi ölçümü numeriktir. Hafta içi/sonu gibi kategorik düzeyde bir kukla (dummy) deęişken de modele girebilir.  $f_i(T_i)$  ise  $T_i$  kovariyeti için bir düzgünleştirici fonksiyondur (35).

### 2.2.1. Düzgünleştirme (smoothing) Mantięı

Düzgünleştirme mantıęı bir çeşit tahmin yöntemi olarak tanımlanabilir (36). Farklı amaçlarda farklı düzgünleştirme fonksiyonları vardır, bunlardan birisi de üstel düzgünleştirme fonksiyonudur (36). Düzgünleştirme mantıęı  $X_1, X_2, \dots, X_p$  şeklinde bir veya daha fazla bağımsız deęişkenin ölçümlerinin fonksiyonlar yardımı ile cevap deęişkeninin eğilimini düzgünleştirmek için kullanılan bir araç olarak tanımlayabilir (37). Bu fonksiyon cevap deęişkeninin trendini, cevap deęişkeninden daha az sayıda deęişkenle tahmin edebildięinden düzgünleştirici olarak isimlendirilir. Düzgünleştiriciler istatistiksel araç olarak oldukça kullanışlıdırlar. Düzgünleştiriciler yardımı ile saçılım grafiklerinin trendleri rahatlıkla tespit edilebilir. Ayrıca bağımlı deęişkenin ortalamasını kovaryetlere baęlı olarak tahmin edebilir (27). Düzgünleştiricilerin en önemli özellięi içlerinden bazılarının parametrik bazılarının ise parametrik olmayan yapılara sahip olmalarıdır (27). Bu yüzden cevap deęişkeni ile açıklayıcı deęişkenler arasında katı varsayımlar içermezler. Bir düzgünleştirici fonksiyonu anlamının en iyi yolu tek deęişkenli düzgünleştirici fonksiyonu içeren bir modeli incelemektir;

$$y_i = f(x_i) + \epsilon_i$$

Burada  $y_i$  cevap deęişkeni,  $x_i$  bir kovariyettir,  $f$  ise düzgünleştirici fonksiyondur ve son olarak  $\epsilon_i$  ise ortalaması sıfır, varyansı  $\sigma^2$  olan normal dağılıma sahip hata terimidir (30, 38).

Bir düzgünleştirici fonksiyonu regresyon zincirlerini kullanarak tanımlayalım. Öncelikle  $f_i$  düzgünleştirici fonksiyonunu düşünelim; daha önceden yapısı bilinen bir temel fonksiyon  $b_i(x_i)$  şeklinde olsun ve  $f_i$  fonksiyonunun aşıęıdaki yapıda olduęunu varsayalım;

$$f(x) = \sum_{i=1}^q b_i(x)\beta_i$$

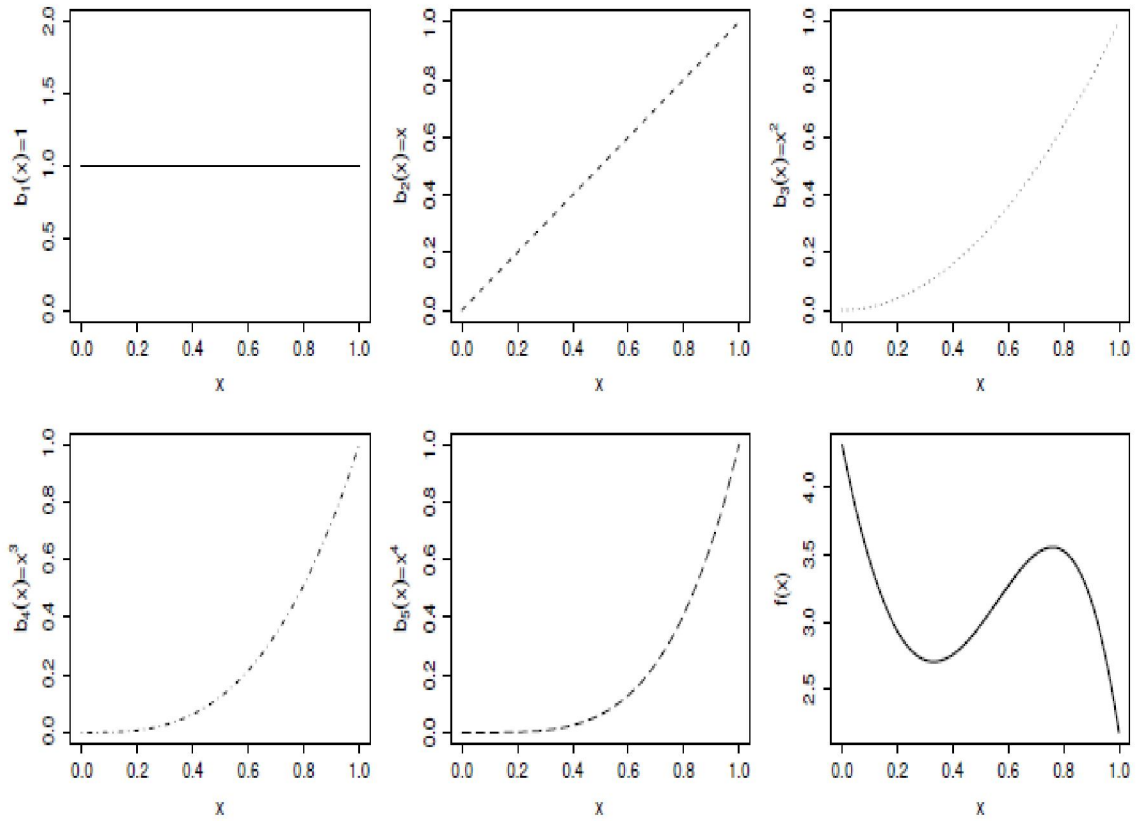
Bazı değerler için bilinmeyen parametre  $\beta_i$  olmak üzere;  $f_i$  fonksiyonunun polinom temelli bir fonksiyon olduğunu düşünelim.  $b_1(x) = 1, b_2(x) = x, b_3(x) = x^2, b_4(x) = x^3$  ve son olarak  $b_5(x) = x^4$  olacak şekilde tanımlansın;

$$f(x) = \beta_1 + x\beta_2 + x^2\beta_3 + x^3\beta_4 + x^4\beta_5,$$

Olup tek değişkenli düzgünleştirici model içeren esas modelimiz;

$$y_i = \beta_1 + x_i\beta_2 + x_i^2\beta_3 + x_i^3\beta_4 + x_i^4\beta_5 + \epsilon_i$$

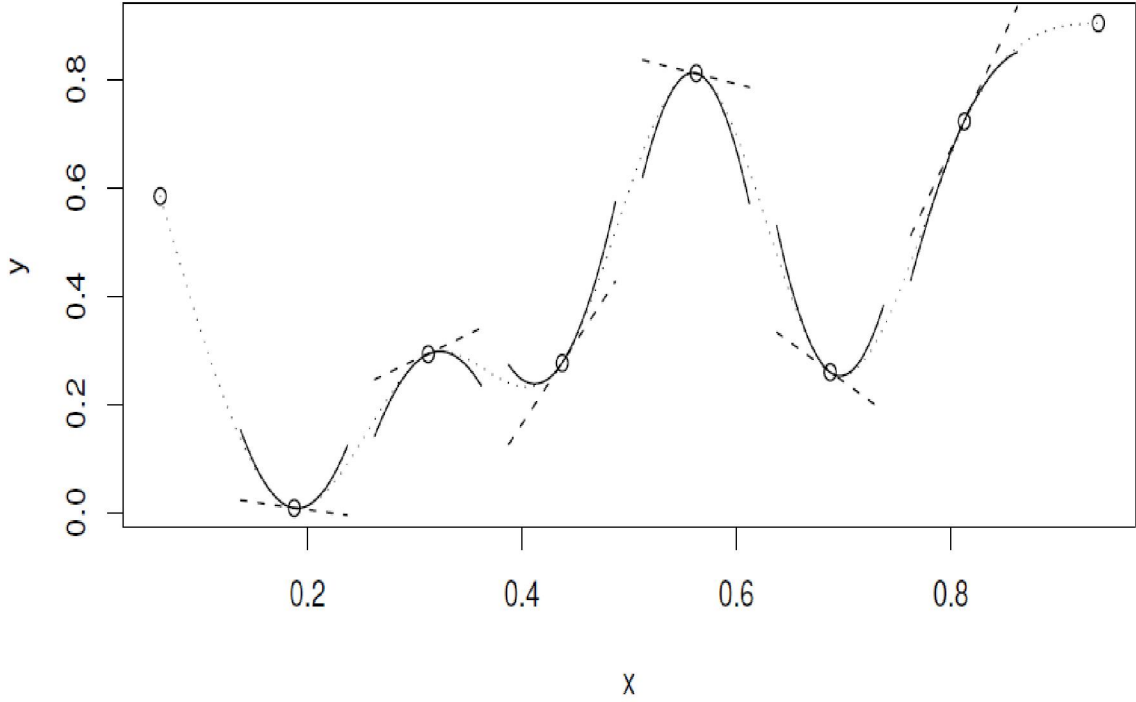
Şekline dönüşür (30).



**Şekil 2.1.** Polinom temelli fonksiyon kullanarak bir düzgünleştirici fonksiyonun gösterimi

Şekil 2.1’de polinom tabanını kullanarak bir fonksiyonun, temel fonksiyon gösterimi görülmektedir (39). Sol üst köşeden başlamak üzere ilk 5 grafikten 4 tanesi polinom tabanlı  $b_i(x)$  temel fonksiyonunu göstermektedir. Temel fonksiyonların her biri  $\beta_j$  ile çarpılıp alt en sağda bir örnek olarak gösterilen  $f(x)$  sonuç eğrisi oluşturulur. Çeşitli  $\beta_j$  değerleri ile birbirinden farklı  $f(x)$  formları üretilebilir (30).

İncelediğimiz hemen hemen tüm düzgünleştiriciler zincir (spline) içerdiğinden bu fonksiyonları teorik açıdan kavrayabilmek için incelenmesi gereken bir diğer yapı da Kübik zincirdir (cubic spline) (39).



**Şekil 2.2.** Kübik zincir yapısının iki boyutlu düzlemde gösterilmesi

Şekil 2.2 bir kübik zincir yapısı görülmektedir (30). Bir kübik zincir ikinci türevi sürekli olan, polinomial kübik bölümlerinden oluşan bir eğridir. Şekil 2.2'deki kübik zincir 7 bölümde oluşmuştur (30). Bir dizi parçalı polinomun birleşmesi sonucu bir zincir ortaya çıkar (40). Zincirde doğruları birleştiren noktalar ( $^{\circ}$ ) zincirin düğümleri olarak bilinir. Şekil 2.2 de görülen her farklı kübik bölüm farklı bir katsayıyı temsil eder. Düğümlerin üzerindeki kalın kesik çizgiler düğümlerdeki eğrinin eğimini, düz çizgi ise düğümlerdeki birinci ve ikinci türevlere karşılık gelir. Şekil 2.2'deki zincirin uç düğümleri üzerindeki ikinci türevleri sıfırdır. Bu yapıda bir zincirin doğal zincir (natural spline) olduğu söylenebilir (39).

Zincir eğrisi kavisli bir hattı izlemek için kolaylıkla bükülebilen ince bir ahşap şeride benzetilebilir (41). Regresyon zincirleri (spline) bu fikrin istatistiksel bir çevirisidir, aynı zamanda lineer olmayan ve bilinmeyen ortalama fonksiyonlarının temsilinin bir yoludur (42). Regresyon eğrileri her ne kadar deneysel çalışmalarda fazla

tercih edilmese de geleneksel parametrik olmayan regresyon mantığı ve düzgünleştirici fonksiyon türlerinin kavranması konusunda yardımcı olacaktır. Ölçüm ile elde edilmiş bir bağımlı değişkenin, bir fonksiyon yardımı ile tahmin edilmeye çalışılmasının genel gösterimi,  $Y = f(X) + \varepsilon$  şeklindedir (42). Burada  $X$  bir tahmin setidir,  $f(X)$  bilinmeyen bir fonksiyon ve  $\varepsilon$  hata terimidir. Regresyon eğrileri deneysel olarak bir  $f(X)$  fonksiyonuna ulaşmanın algoritmik bir yöntemidir. Buradaki tahmin hedefi; gerçek  $f(X)$  yerine, cevap değişkenlerine en yakın  $f(X)$ 'i tahmin etmeye çalışmaktır. Regresyon eğrileri bu tahmini lineer ve kayıp fonksiyonlu karesel tahmin yöntemi ile yapmaktadır (42).

Verilen bir düzgünleştirme fonksiyonu  $f$  olsun, bu fonksiyon  $[a,b]$  aralığında tanımlı olsun,  $f$  eğrisinin düzlüğü veya kıvrım miktarını ölçmenin birçok yolu vardır (41). İki kere diferansiyeli alınabilen  $f$  fonksiyonunun düzlük miktarını ölçebilmenin mantıklı bir yolu, fonksiyonun ikinci derece türevinin sınırları dâhilinde integralini hesaplamaktır. Bu yapı  $\int_a^b (f''(x))^2$  şeklindedir. Yapılan bu hesaplamayı başka yöntemlerle de yapmak mümkündür (41).

Bu bölümde zincir yapısına değinildi ve bu yapıya ek olarak, çevre epidemiyolojisi çalışmalarının modellenmesinde (GAM'da) en sık kullanılan düzgünleştirme fonksiyonlarını incelemek faydalı olacaktır.

### 2.2.2. GAM'da Sık Kullanılan Düzgünleştirme Fonksiyonları

Farklı modelleme görevleri için farklı düzgünleştirme fonksiyonları mevcuttur. Düzgünleştiriciler GAM içinde oldukça esnek bir şekilde kullanılabilir. GAM'da düzgünleştirme fonksiyonları farklı temellere dayalıdır (43). Cezalı regresyon düzgünleştiricileri (Penalized regression smoothers), Serpme diyagramı düzgünleştiricileri (Scatterplot smoothers), Tek boyutlu düzgünleştirme fonksiyonları (one-dimensional smoothers), Tensor düzgünleştiriciler (Tensor product smoothers) ve İzotropik düzgünleştirme (Isotropic smoothing) bu yapılar için verilecek örnekler arasındadır. Bu bölümde özellikle uygulamacılar için yararlı olacak ve R programında da sıklıkla başvurulan bazı düzgünleştiricilerden bahsedilmiştir.

#### 2.2.2.1. Doğal Kübik Zincir Yapısı ve İnterpolasyon Mantığı

Kübik zincir yapısındaki  $g(x)$  fonksiyonlarının açık yapısını ele alarak inceleyecek olursak. Gerçek  $t_1, \dots, t_n$  sayıları  $[a, b]$  aralığında olsun ve  $a < t_1 < t_2 < \dots < t_n < b$  şeklinde tanımlansın. Eğer  $g$ ,  $[a, b]$  aralığında tanımlı bir kübik zincir ise iki

önemli koşulu sağlamalıdır (41, 44). İlk koşul her bir  $(a, t_1), (t_1, t_2), (t_2, t_3), \dots, (t_n, b)$  çiftinin bir kübik polinom olması. İkincisi ise polinom parçaları  $t_i$  noktasında birlikte veriye uydurulması, aynı şekilde  $g$  tüm  $[a, b]$  aralığında her bir  $t_i$  noktasında kendisi, birinci ve ikinci türevi sürekli değerler almasıdır. Burada bahsi geçen  $t_i$ 'lere düğüm (knot) denir. Bir Kübik zinciri birçok farklı eşitlikle tanımlayabilirsiniz. Bunlardan birisi de dördüncü derece polinom ile tanımlanan yapıdır (41). Örneğin;

$$g(t_i) = d_i(t - t_i)^3 + c_i(t - t_i)^2 + b_i(t - t_i) + a_i, \quad t_i \leq t \leq t_{i+1} \text{ için}$$

Verilen  $a_i, b_i, c_i$  ve  $d_i$  için  $i = 1, \dots, n, t_0 = a$  ve  $t_{n+1} = b$  şeklinde tanımlanır (41).

Bir  $\{x_i, y_i : i = 1, \dots, n\}$  veri seti düşünelim. Burada  $x_i < x_{i+1}$  olmak üzere Doğal kübik zincir bu noktaların interpolasyonunu sağlayan  $g(x)$  fonksiyonudur (45). Böyle bir kübik zincir  $[x_i, x_{i+1}]$  her biri için tüm düzlemin ikinci türevlerinin sıfır olacağı şekilde birleştirilen bir fonksiyondur. Burada  $g(x_i) = y_i$  ve  $g''(x_i) = g''(x_n) = 0$  şeklindedir. Bu koşulları sağlayan kübik zincir yapısı Doğal kübik zincirdir (45). Fonksiyonların tamamı  $[x_1, x_n]$  sınırlarında sürekli. İlk türevleri ve  $\{x_i, y_i\}$  interpolasyonu olan  $g(x)$  fonksiyonu  $J(f) = \int_{x_1}^{x_n} (f''(x))^2 dx$  denklemi minimize eden düzgünleştiriciye sahiptir (41, 44).

### 2.2.2.2. Kübik Düzgünleştirme Zincirleri-KRZ (Cubic smoothing splines)

Bir Kübik düzgünleştirme zinciri aşağıdaki optimizasyon probleminin çözümüdür. Çözüm tüm ikinci sürekli türevli  $f(x_i)$  fonksiyonları arasında ceza en küçük kareler değerini minimize eden fonksiyondur (46).

$$\sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \int_a^b [f''(x)]^2 dx$$

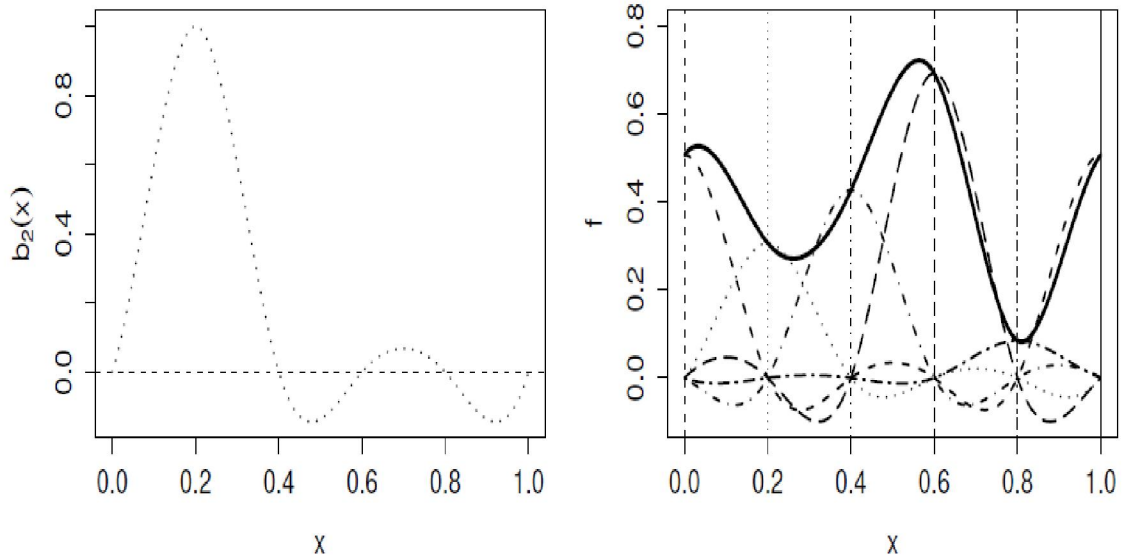
Burada  $\lambda$  ayarlanabilen bir parametredir ve  $a \leq x_1 \leq \dots \leq x_n \leq b$  sınırlarında tanımlanır. Kübik zincir ile yapılan çalışmada düğüm sayısı (parçalı polinomiyal) genelde üç değeri atandığı için ismini, bu derece değerinden almaktadır (47, 48).

Yukarıdaki denklemde ilk terim en küçük kareler metodunu temsil eder. Sadece bu parça ile hesaplama yapılırsa düzgünleştirme kullanılmadan yapılan interpolasyon sonucuna varılır. İkinci terim fonksiyonun eğriliğinde cezalandırma yaparken aynı zamanda veriye yakınlığı ölçer. Denklemde  $\int (f''(x))^2 dx$  kısmı  $f(x)$  fonksiyonunun salınım (wiggliness) ölçüsüdür (30, 49). Düzgünleştirme parametresi  $\lambda$  daima pozitif değerler alır. Veriler eşleştirilmelerle bir  $g$  fonksiyonun üretilebilmesi için göreceli ağırlıkları kontrol etmekte kullanılır (39). Elde edilen  $g(x_i)$  bir düzgünleştirilmiş

zincirdir (50). Aslında  $\lambda$  fonksiyonun salınımı ile veri uyum iyiliği arasında bir ayar yapar (51). Eğer  $\lambda \rightarrow \infty$  ise düzgünleştirme parametresi burada etkili bir çarpan olarak ceza terimindeki fonksiyonu sıfıra eşitlemeye ( $f'''(x) = 0$ ) zorlar ve sonuç en küçük kareler denkleminde dönüşür. Tam tersi durumda  $\lambda \rightarrow 0$  olursa düzgünleştirme parametresi bu etkisini kaybederek çözümü ikinci dereceden bir türev fonksiyonu ile sonuçlandırır. Bu durumda  $\lambda$  küçük değerler aldığında daha fazla kıvrıma sahip eğrilerin oluşmasına sebep olduğu söylenebilir (30).

### 2.2.2.3. Döngüsel Kübik Regresyon Zinciri-DKRZ (Cyclic cubic regression spline-CCRS)

Bir düzgünleştirme fonksiyonunun periyodik olması uygulamada sık rastlanan bir durumdur. Döngüsel kübik regresyon zinciri fonksiyonunu diğer zincirlerden ayıran en önemli özellik; fonksiyonda bir değer için aynı alt ve üst sınır değerlerinde son bulmasıdır. Bir temel fonksiyon ve düzgünleştirici fonksiyonun döngüsel kübik zincir yapısı Şekil 2.3'de görülmektedir. Ceza kübik regresyon zincirleri Şekil 2.3'deki görüntüyü veren düzgünleştirici gibi çalışması için modifiye edilebilir (39).



Şekil 2.3. Döngüsel kübik regresyon zincir yapısının gösterimi

Şekil 2.3'de sol taraftaki grafik bir temel fonksiyona  $b_2(x)$  döngüsel kübik regresyon zinciri uygulandığında nasıl bir grafik yapısına sahip olacağını göstermektedir (39). Dikkat edilecek olursa temel fonksiyon değeri ve ilk iki türevi  $x = 0$  ve  $x = 1$  olarak sıfır değerinde eşleşmiştir. Sağ taraftaki grafik ise sol taraftaki grafik gibi temel



fonksiyonların bir düzgünleştirici eğriye nasıl kombine olacağını göstermektedir (30). İki grafikte de zincirin başlangıç ve bitiş noktaları aynı değer üzerindedir. Bu yapıyı sağlayan zincir yapısı Döngüsel kübik regresyon zinciridir (39).

#### 2.2.2.4. P-Zincirleri (P-Splines)

GAM çözümlerinde Eilers ve Marx tarafından önerilen, tek değişkenli P zincirler de kullanılabilir (52). Bu yapı katsayıları direk olarak ayrı ayrı cezalar ile cezalandırabilme yeteneğine sahip B-zincirleri (B-spline) temeline dayalıdır. P-zincirleri algoritması ceza ve temel sırayı değiştirebilme esnekliği sağlar (52). P-zincirleri yöntemi, yapısı duyarlı çalışmayan düzgünleştirme fonksiyonlarını duyarlı hale getirmede diğer yöntemlere göre daha başarılıdır. Türev esaslı ceza yapıları HKO (Hata Kareler Ortalaması) değerlerini düşürmede daha iyi bir performans gösterir, çünkü iyi sonuç veren zincir yapısının teorik özelliği türev penaltılarının kullanımı ile yakından ilişkilidir (52).

P-zincirlerinin temel mantığı, k tane temel zincir ile genişleme sağlayabilen  $f(x)$  düzgünleştirme fonksiyonu;  $f(x) = \sum_{j=1}^k \beta_j B_{m_1, j}(x)$  şeklinde tanımlanabilir. Burada  $B_{m_1, j}(x)$  ise  $m_1$ 'e bağlı B-zincir temel fonksiyonudur (53, 54). P-zincirleri yapısında  $m_1$  temel zincirin düzeyini belirler. Örneğin  $m_1 = 3$  olursa bu bir kübik zincir yapısına dönüşür, eğer  $m_1 = 2$  olursa da karesel zincir (quadratic form) yapısına dönüşür. P-zincir denkleminde cezanın türev düzeyini  $m_2$  parametresi belirler (53).

$$J = \int_a^b (f^{[m_2]}(x))^2 dx = \beta^T S \beta$$

Burada  $f^{[m_2]}(x)$  kısmındaki  $m_2$   $x$ 'e bağlı  $f$  fonksiyonunun  $m_2$ -nci türevini göstermektedir. Varsayımlara göre  $m_2 \leq m_1$  koşulu yerine gelmezse ceza temel fonksiyonlar bakımından türev anlamında doğru bir şekilde tanımlanmamış olur.  $J = \beta^T S \beta$  denklemindeki  $S$  katsayıların diyagonal matrisidir (53).

#### 2.2.2.5. İnce Tabaka Regresyon Zincirleri-İTRZ (Thin-Plate Regression Spline-TPRS)

İnce Tabaka (İT) zincirleri (Duchon, 1977) özellikle modelde uç değerler içeren çoklu açıklayıcı değişken kullanıldığında, bir fonksiyonu tahmin etme de zarif ve genel bir çözüm yoludur (30). Gözlem sayısı  $n$  olan  $(y_i, x_i)$  gözlem çiftinden gelen  $g(x)$  fonksiyonunu, bir düzgünleştirme fonksiyonu ile tahmin etme problemini düşünelim;

$$y_i = g(x_i) + \epsilon_i$$

İT zincirleri  $g$  fonksiyonunu  $\hat{f}$  fonksiyonunu minimize ederek bulur.

$$\|y - f\|^2 + \lambda J_{md}(f)$$

Burada  $y$  içinde  $y_i$  değerlerini içeren bir vektördür ve  $f = [f(x_1, x_2, \dots, x_n)]^T$  şeklindedir.  $J_{md}(f)$  ise  $f$ 'in eğim miktarını ölçen bir ceza fonksiyonudur.  $\lambda$  ise düzgünleştirme fonksiyonundaki  $f$  ile veriye uydurulmuş bir düzgünleştirme parametresidir (39). İT regresyon zinciri düğüm konumları ile ilgilenmez. Bu yüzden hesaplamaları diğer zincirlere göre nispeten daha basittir. Ayrıca Lanczos iterasyonunu kullanarak temel değişim ve budama konusunda etkili çözümler sunar. İT düzgünleştiriciler büyük ölçekte verilerin modellenmesine izin verir. İTRZ yönteminde, model bir veya daha fazla sayıda kovariyet ile kurulabilir (55).

İTRZ'de kovariyetler düşük ranklı İzotropik düzgünleştiricilere sahiptir. İzotropik kovariyetlerde koordinat sistemleri rotasyona uğrasa da düzgünleştirme durumu değişmez (55). Düşük rank düzgünleştirilen veri; daha az sayıda katsayı atanması anlamına gelir (55). İT regresyon zincirleri, İT zincir bileşen eğrilerinde kıvrım noktalarının kesime uğratılması (budanması) fikrine dayalıdır. İT regresyon zincirleri çok sayıda tahmin edicinin düzgünleştiricilerini kurmada, hesaplamada ve düğümlerin hızlı bir şekilde konumlandırılmasında gayet başarılıdır (30). İTRZ yaklaşımında veriye uydurulan zincir biçiminde olabilecek değişiklikleri minimize ederken aynı zamanda zincirin uygun değerinde olabilecek hataları da minimuma indirir (30). Ayrıca bu yaklaşımın aralıklı yapısı ile zincir benzeri yapılar ürettiği göz önüne alındığında, lineer regresyon ve GLM'de cezasız zincir benzeri terimlerin modele dahil edilmesinin doğal bir yolunu da sağlamış olur. Bu açıdan bakıldığında İTRZ, lineer regresyon ve GLM'de model seçimi, çıkarsama ve teşhis noktasında kullanım olanağı sunabilir (55).

#### **2.2.2.6. Duchon Zinciri-DZ (Duchon Spline-DS)**

Duchon zinciri İnce plaka zincirlerin genelleştirilmiş bir hâlidir. Duchon zincirinin en önemli özelliği verilen herhangi sayıda kovariyet için İnce plaka zincirlerinden daha düşük cezalı türevlere izin vermesidir (55). Duchon klasik İnce plaka zincirleri algoritmasını şu şekilde genelleştirir: Duchon ceza yapısını Fourier serileri yardımıyla yeniden inşa eder. Fourier sıklıklarının öklid normları yardımıyla integral içindeki normları kareleri şeklinde ağırlıklar atayarak gücü iki katına çıkarmayı başarmıştır (56). Burada  $s$  kullanıcı tarafından seçilen sabiti ikiye bölen bir sabittir. Hesaplaması şu şekildedir:  $d$  düzgünleştirme argümanlarının sayısı olsun;  $s$  değeri

$-d/2 < s < d/2$  aralığındadır. Eğer  $s = 0$  olursa klasik ince plaka zinciri elde edilir (57). Duchon bu ceza yapısını aşağıdaki gibi ifade etmiştir:

$$\hat{f}(x) = \sum_{i=1}^n \delta_i \eta_{2m-2s-d}(\|X - X_i\|) + \sum_{j=1}^M \alpha_j \phi_j(X).$$

### 2.2.2.7. Küre Zincirler-KZ (Spline on the Sphere)

GAM izotropik düzgünleştiriciler bir küre fonksiyonu üzerine de uygulayabilir. Böyle bir düzgünleştirici için iki argümana ihtiyaç vardır. Bunlardan ilki enlem (derece olarak), ikincisi de boylamdır (derece olarak) (49).  $\theta$ : enlem,  $\phi$ : boylam olmak üzere, küre üzerindeki ikinci dereceden zincirlerin ceza analoğu aşağıdaki gibidir (39):

$$J(f) = \int_0^{2\pi} \int_0^{\pi} \left( \frac{f_{\phi\phi}}{\sin^2\theta} + \frac{\{ \sin(\theta f_{\theta}) \}_{\theta}}{\sin\theta} \right)^2 \sin\theta d\theta d\phi$$

$p = (\phi, \theta)^T$  ve  $\gamma(p, p') = \arccos\{\sin\theta\sin\theta' + \cos\theta\cos\theta'\cos(\phi - \phi')\}$  ise  $p$  ve  $p'$  arasındaki açıdır. Wendelberg küre zincirine ceza yapısını ekleyerek aşağıdaki gibi göstermiştir (57):

$$f(p) = \alpha + \sum_{i=1}^n \gamma_i R(p, p')$$

### 2.2.3. Düzgünleştirme Parametresinin Seçimi

Bir önceki başlıkta GAM'da sık kullanılan bazı düzgünleştirme fonksiyonları tanıtıldı. Bu bölümde GAM'da tahmini etkileyen diğer önemli bir parametre olan düzgünleştirme parametresi ele alınacaktır. Düzgünleştirme parametresi  $\lambda$  düzgünleştirme fonksiyonu  $f$  ile veri arasında denge sağlar (31).

$E(Y_i) = \mu_i$ , olmak üzere GAM modeli;  $g(\mu_i) = \eta_i \equiv X_i^* \beta^* + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}) + \dots$ , şeklinde tanımlansın. Burada  $g$  monoton link fonksiyonu,  $X^*$  model matrisinin  $i$ -nci satırı  $X_i^*$  olmak üzere modelin parametrik kısmını temsil eder,  $\beta^*$  ise ona karşılık gelen parametre vektörüdür,  $f_i$  ise her bir  $x_i$  kovaryetine denk gelen düzgünleştirici fonksiyon olsun (45). Bu türde modellerin çözümünde karşılaşılan problemlerden biriside aşırı uyum (over-fitting) problemidir. Bu yüzden GAM modelinde genelde ceza maksimum olabilirlik tahmini (Penalized maximum likelihood estimation) yöntemi kullanılır (45).

$$l(\eta) - \frac{1}{2} \sum_j \theta_j \int [f_j''(x)]^2 dx.$$

Burada  $l$  lineer tahmin edicinin log olabilirliği. Toplam teriminin bulunduğu bölüm ise GAM fonksiyonunun bileşenlerinin salınım ölçümüdür (39).  $\theta_j$  düzgünlük ile veriye uygunluk arasındaki dengeyi kontrol eden düzgünleştirme parametresidir. Pratikte model veriye uydurularak ceza maksimum olabilirlik problemi, ceza yenilemeli yeniden ağırlıklandırılmış en küçük kareler” ifadesi yerine “Model veriye uydurulurken Ceza Maksimum Olabilirlik yöntemi yerine Ceza Yenilemeli Yeniden Ağırlıklandırılmış En Küçük Kareler (Penalized iteratively reweighted least squares P-IRLS) yöntemi kullanılır (58). Örneğin verilen olabilirlik üzerinden; k-nci P-IRLS yenilemesi aşağıdaki ceza kareler toplamını  $\eta^{(k+1)}$  lineer tahminin  $(k + 1)$ -nci tahminini bulmak için  $\eta^{(k)}$  tahminini minimize eder (45).

$$\|W^{[k]}(z^{[k]} - \eta)\|^2 + \sum_j \theta_j \int [f_j''(x)]^2 dx.$$

$W^{[k]}$  yenileme ağırlıkları,  $z^{[k]}$  standartlaştırılmış değerlerdir.

$$W_{ii}^{[k]} = 1/\sqrt{g'(\mu_i^{[k]})^2 V_i^{[k]}}$$

Şeklinde hesaplanır,  $V_i^{[k]}$ ,  $\mu_i^{[k]}$  tahmininin  $Y_i$ 'nin varyansına oranıdır.

$$z_i^{[k]} = \eta_i^{[k]} + g'(\mu_i^{[k]})(y_i - \mu_i^{[k]})$$

şeklinde hesaplanır (45). Ceza olabilirlik yaklaşımının kullanılmasıyla bariz olarak ortaya çıkan güçlük düzgünleştirme parametresinin tahminidir (59).  $\lambda$  veri analizi ile seçilen negatif olmayan bir parametredir.  $\lambda \rightarrow \infty$  olması durumunda  $f$  doğrusal bir çizgi şeklini alır,  $\lambda = 0$  olarak hesaplandığında ise  $f$  cezasız parçalı lineer regresyon tahmini ile aynı sonucu verir.

### 2.2.3.1. Düzgünleştirme Parametresine Çapraz Geçerlilik (CV) ve Genelleştirilmiş Çapraz Geçerlilik (GCV) Yöntemleri ile Karar Vermek

$\lambda$  parametresi seçiminde kullanılan yöntemlerden birisi de çapraz geçerlilik (cross validation-CV) yöntemidir (10).  $\lambda$  parametresi gereğinden yüksek bir değer olarak atanırsa veride aşırı düzgünleşme (over-smoothed) problemi ile karşılaşılır, aksi durumda  $\lambda$  gereğinden düşük bir değer olarak seçilirse de düşük düzgünleşme (under-smoothed) sorunu ortaya çıkar. Bu iki durum da gerçek  $f$  fonksiyonunu tahmin etmeye çalışan  $\hat{f}$

fonksiyonunda sapmalar yaşanmasına neden olacaktır (30). İdeal olan ise olabilecek gerçek  $f$  fonksiyonuna en yakın  $\hat{f}$  fonksiyonuna optimal bir  $\lambda$  seçimi ile ulaşmaktır. Buradan yola çıkarak uygun  $\lambda$  seçim aşamalarına aşağıdaki denklemi minimize ederek başlanır (60).

$$M = \frac{1}{n} \sum_{i=1}^n (\hat{f}_i - f_i)^2,$$

Denklemden  $f$  fonksiyonu bilinmediğinden  $M$  direk olarak hesaplanamaz fakat  $\mathbb{E}(M) + \sigma^2$ 'nin bir tahminini türetmek mümkündür. Bu da yeni bir değişken tahmininin standart hatasıdır (58). Tüm verideki beklenen  $y_i$  değerini  $\hat{f}^{[-i]}$  ile veriye uygun hale getirilerek çapraz geçerlilik skorunu şu şekilde tanımlanabilir:

$$V_o = \frac{1}{n} \sum_{i=1}^n (\hat{f}^{[-i]} - y_i)^2.$$

Bu CV skoru sırayla her değeri veriden çıkarıp modeli veriden geri kalan değerlerle tekrar uydurup, eksik veri ile öngörülen değerler arasındaki karesel farkın hesaplanması ile sonuçlanır (39).  $y_i = f_i + \epsilon_i$  eşitliğini CV denkleminde yerine koyarsak:

$$V_o = \frac{1}{n} \sum_{i=1}^n (\hat{f}^{[-i]} - f_i)^2 - 2(\hat{f}^{[-i]} - f_i)\epsilon_i + \epsilon_i^2$$

$\mathbb{E}(\epsilon_i) = 0$  olup  $\epsilon_i$  ve  $\hat{f}^{[-i]}$  bağımsız olduğundan denklemin son hali:

$$\mathbb{E}(V_o) = \frac{1}{n} \mathbb{E} \left( \sum_{i=1}^n (\hat{f}^{[-i]} - y_i)^2 \right) + \sigma^2$$

Büyük örnekte  $\hat{f}^{[-i]} \approx \hat{f}$  denkliği sağlanır ve böylece  $\mathbb{E}(V_o) \approx \mathbb{E}(M) + \sigma^2$  denkliği de sağlanmış olur (39).  $V_o$  eşitliğini minimize etmek için seçilen  $\lambda$  değeri aynı zamanda  $M$  eşitliğini de minimize eden değer olacaktır. Her seferinde veriden bir gözlemin eksiltilmesi suretiyle yapılan CV hesabı bazı zorluklar ortaya çıkardığından bu hesaplama yerine  $n$  tane gözlem için aynı anda hesaplama imkânı sunan GCV metodunun kullanılması daha yaygındır (45). GCV formülüne aşağıdaki denklem yardımıyla ulaşılabilir;

$$V_o = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}_i)^2 / (1 - A_{ii})^2$$

Burada  $\hat{f}$  tüm veriden yapılan tahmin,  $A$  etki matrisidir. Hesaplamalarda  $A$  etki matrisi ortalaması alınır. Yani denklemden yer alan  $A_{ii}$  matrisi yerine  $iz(A)/n$  yer alır. Buna göre Genelleştirilmiş çapraz geçerlilik formülü;

$$V_g = \frac{n \sum_{i=1}^n (y_i - \hat{f}_i)^2}{[n - iz(A)]^2}$$

Şeklinde hesaplanır (46).

### 2.2.3.2. Düzgünleştirme Parametresine UBRE Yöntemi ile Karar Vermek

Düzgünleştirme parametresine karar vermek için sıklıkla kullanılan bir diğer yöntem de UBRE (Yansız risk tahmini – Unbiased risk estimator) yöntemidir. UBRE yaklaşımı 1979 yılında Craven ve Wahba tarafından önerilmiştir (61, 62). Bilinen sabit bir varyans yardımıyla basit eklemeli bir modelde düzgünleştirme parametresi tahmin ettiğimizi düşünelim. Akla en uygun yaklaşım  $\mu \equiv \mathbb{E}(y)$  gerçek değerine en yakın  $\hat{\mu}$  tahmin etmeye çalışmaktır. Bu yakınlığı ölçmek için kullanılacak en uygun ölçüm Hata kareler ortalaması (HKO) ölçümünün tahmini olan  $M$  değerini hesaplamaktır (30, 61, 62).

$$M = \mathbb{E} \left( \|\mu - X\hat{\beta}\|^2 \right) / n - \sigma^2 + 2iz(A)\sigma^2 / n$$

$M$  eşitliğini minimize etmek suretiyle seçilecek düzgünleştirme parametresi  $V_u(\lambda)$  olsun bu  $\lambda$ 'nın UBRE yöntemiyle tahmin edilmesi anlamına gelmektedir (61).

$$V_u(\lambda) = \|\mu - Ay\|^2 / n - \sigma^2 + \frac{2iz(A)\sigma^2}{n}$$

Bu denklem literatürde Mallow'un  $C_p$  istatistiği olarak da geçmektedir. Eğer  $\sigma^2$  biliniyor ise  $V_u$  denklemini minimize eden  $\lambda$  seçimleri daha iyi sonuçlar verir.

$$\mathbb{E}(\|y - Ay\|^2) = \sigma^2 \{n - iz(A)\}$$

Bu denklemden yola çıkarak  $M$  eşitliğini hesaplayacak olursak;

$$M = \mathbb{E} \left( \|\mu - X\hat{\beta}\|^2 \right) / n = \frac{iz(A)}{n} \sigma^2$$

Ve hata kareler ortalaması tahmini  $\hat{M} = iz(A)\hat{\sigma}^2 / n$  şeklinde hesaplanır (39).

### 2.2.4. Temel Boyut ( $k$ ) Seçimi

Ceza regresyon zincirleri kullanılarak yapılan modellemede, temel boyut seçimi model yapım aşamasının bir parçasıdır. Temel boyut seçimi düzgünleştirme fonksiyonunun hesaplama gücünü etkilemektedir. Bir zincirin veride var olandan daha fazla katsayıya ihtiyaç duyması durumunda, hesaplamalarda ciddi bir hata yapılmaktadır. Örneğin bir zincir 100 düğüme ihtiyacı varken, siz 40 düğüm olarak karar verirsiniz bu durum ilgili zincirin hesaplamalarında hataya neden olacaktır (30). Kim ve Gu'nun düzgünleştirme zincirlerini kullanarak yaptıkları simülasyonlarda, örneklem sayısı  $n$

olmak üzere, temel boyutun yaklaşık olarak  $n^{2/9}O$  olması gerektiğini söylemişlerdir (63). Bazı farklı görüşler temel boyut sayısı hesaplanırken örneklem sayısına ( $n$ ) tahmin edici sayısının eklenip hesaplamasını tavsiye etmişlerdir (30).

Pratikte temel boyut seçimi model belirlemenin bir parçasıdır. Düzgünleştirme parametresi esasında etkili serbestlik derecesi hesabıyla kontrol edilir. Ceza regresyon zinciri yapısında etkili serbestlik derecesinin alacağı maksimum değer ( $k-1$ ) dir (62). Seçilen  $k$  değeri bir anlamda serbestlik derecesinin seçimi anlamına gelmektedir. Diğer taraftan temel boyut seçiminin genel bir kriteri olmamakla birlikte serbestlik derecesini makul düzeyde temsil edecek kadar büyük olmalıdır. Seçilen  $k$  değerinin büyük ya da küçük olması, problemin özelliğine de bağlı olacağından daha önce yapılmış çalışmalar referans alınarak da bir seçim yapılabilir (64). Bu yüzden uydurulan modelde hesaplanan temel boyut çok küçük olmadığı sürece herhangi bir problem ortaya çıkmaz (62). Ayrıca R yazılımında farklı temel boyut seçimlerinde düzgünleştirme parametrelerinin anlamlılıklarını kıyaslamamıza imkân sunan yazılımlar da mevcuttur (65).

### 2.2.5. Etkili Serbestlik Derecesi (ESD)

Düzgünleştirme parametresinin bir modeli veriye daha iyi uydurabilmek için serbestlik derecesinin doğru atanması gerekir. Aynı zamanda etkili serbestlik derecesi düzgünleştirme miktarının bir göstergesidir. Bir düzgünleştirme operatörü  $S_\lambda$  olsun  $S_\lambda = \sum_j \lambda_j S_j$  şeklindedir. ESD aşağıdaki algoritma ile tayin edilir (39).

$$\tau = iz\{(X^T W X + S_\lambda)^{-1} X^T W X\}$$

ESD'nin tanımlanmasında alternatif bir yol daha vardır.  $A = X(X^T X + S_\lambda)^{-1} X^T$  ve  $F = (X^T X + S_\lambda)^{-1} X^T X$  etki matrisleridir. Bir model için hata kareler ortalamasını düşünelim:

$$E(\|y - Ay\|^2) = \sigma^2\{n - 2iz(A) + iz(AA)\}b^T b$$

Burada  $b = \mu - A\mu$  düzgünleştirmenin yanlılığının hesabıdır ve  $\hat{b} = \hat{\mu} - A\hat{\mu}$  ise  $b$ 'nin tahminidir. Buradan alternatif bir varyans tahminine de ulaşmak mümkündür.

$$\sigma^2 = \frac{\|y - Ay\|^2 - \hat{b}^T \hat{b}}{n - 2iz(A) + iz(AA)}$$

Buradaki varyans denkleminde paydada yer alan  $\tau = 2iz(A) - iz(AA) = 2iz(F) + iz(FF)$  bir modelin etkili serbestlik derecesi olarak karşımıza çıkmaktadır. Diğer taraftan  $A$  matrisinin Hat matrisi denkleminde  $H = X(X^T X)^{-1} X^T$  bir etkisi olduğu düşünülür. Bu düşünce öyle bir  $\lambda$  değeri atayalım ki elde edilen  $A$  matrisinin izi arzu edilen serbestlik derecesine eşit olsun (65).

Düzenleme içeren kovaryetlerin serbestlik derecesine karar vermek tahminin yanlılığı bakımından hayati derecede önemlidir. Düzenleme fonksiyonunun temel fonksiyon sayısı yüksek olur ise bu durum tahminde aşırı uyum problemini ortaya çıkarmaktadır. Düzenlemede düğüm sayısını kontrol etmenin bir yöntemi de sadece düzenlemeye sahip bağımsız değişkenin yer alındığı modelle bağımlı değişkenin tahmininin saçılım grafiğinin incelenmesidir. Eğer tahmin edilen bağımlı değişkenin trend özelliği kayboluyorsa seçilen serbestlik derecesi gereğinden büyüktür ve yanlılığa sebep olacaktır (12, 66, 67).

### 2.2.6 GAM'da Geriye Doğru En Uygun Modeli Oluşturma Algoritması

Genelleştirilmiş eklemeli modellerin tahmini konusunda birçok yaklaşım mevcuttur. Hastie ve Tibshirani tarafından önerilen geriye doğru en uygun modeli oluşturma yöntemi, herhangi bir modelde kullanılabilen hemen hemen tüm düzenleme tekniklerini modele ekleme şansı verir ve birçok avantaja sahiptir. Dezavantajı ise algoritmasında bir modelin serbestlik derecesi tahmininin yer almamasıdır (10, 39).

Genel fikir eklemeli modelden gelen düzeltmelerin kısmi artıklarının yenilme tekniği ile eklemeli modelin düzeltme fonksiyonları üzerinde tahminler yapılmasına imkân sağlamasıdır. Bu yöntemle tüm düzeltme fonksiyonlarının tahminini yapmak mümkündür. Geriye doğru en uygun modeli oluşturma algoritması yenileme tabanlıdır. Başlangıç fonksiyonu  $f_0$ 'dan, ...,  $f_p$ 'ye kadar yinelemeyle fonksiyonlar boyunca dolaşarak bireysel düzenleme bileşenlerini kısmi artıkları dikkate alarak uydurulan bir tekrar yapısına sahiptir (59). Formülasyonu  $j$ -nci düzenleme parametresinin kısmi artıklarının,  $j$ -nci düzenleme terimi ile tahmin edilen cevap değişkeninin son model terimi tahmininden çıkarılması ile elde edilen sonuçların eşit olması prensibine dayalıdır (30).

$j$ -nci kısmi artıkları şu şekilde tanımlayalım;

$$R_j = Y - f_0 - \sum_{k \neq j} f_k(X_k)$$

Burada  $E(R_j/X_j) = f_j(X_j)$  eşitliğidir. Bu yapı her düzeltme fonksiyonu  $\hat{f}_j$  için bir tahmin yapabilme yolunu açar (59). Bilinen bir  $(x_i, y_i)$  gözlemi için Ceza kareler toplamı aşağıdaki formül yardımıyla hesaplanır;



$$\sum_{j=1}^p (y_j - f_0 - f_j(x_j))^2 + \sum_{j=1}^p \lambda_j \int [f''(t_j)]^2 dt_j$$

Eklemler için Geriye doğru en uygun modeli oluşturma algoritması aşağıdaki gibidir:

1. Başlangıç:

$$\hat{f}_0 = \frac{1}{n} \sum_{i=1}^n y_i, \hat{f}_j^1 \equiv 0, m = 1$$

2. Yenileme:

$$m \leftarrow m + 1, j = 1, 2, \dots, p$$

$$f_j^m(x_j) \leftarrow S_j \left[ \left\{ y_i - \hat{f}_0 - \sum_{k=1}^{j-1} \hat{f}_k^m(x_k) - \sum_{k=j+1}^p \hat{f}_k^{m-1}(x_k) \right\}_1^n \right]$$

3. Sonlandırma kriteri:

$$HKT = S \left( Y - \hat{f}_0 - \sum_{k=1}^p \hat{f}_k^m(x_k) \right)^2$$

HKT değerinin azalması hatalıdır,  $\hat{f}_j$  fonksiyonundaki değişim önceden belirlenen bir eşğin altında olduğu anlamına gelir. Yukarıdaki yenileme algoritmasındaki  $\hat{f}_j^m(\cdot)$ ,  $f_j(\cdot)$  fonksiyonunun  $m - nci$  yenilemedeki tahminidir. GAM prosedürü Geriye doğru en uygun modeli oluşturma algoritmasında yakınsama kriteri olarak aşağıdaki algoritmayı kullanır (47):

$$\frac{\sum_{j=1}^p \sum_{i=1}^n (f_j^{m-1}(x_{ij}) - f_j^m(x_{ij}))^2}{\sum_{j=1}^p \sum_{i=1}^n f_j^{m-1}(x_{ij})^2} \leq \varepsilon$$

Burada  $\varepsilon = 10^{-8}$  dir. Geriye doğru en uygun modeli oluşturma yöntemi yapısını örneklerle detaylandırırız:

Örneğin  $m = 1 + 1, j = 1$  olsun;

$$f_1^2(x_1) = S_1 \left[ \left\{ y_i - \hat{f}_0 - \sum_{k=1}^{1-1} \hat{f}_1^2(x_1) - \sum_{k=1+1}^p \hat{f}_2^1(x_k) \right\}_1^n \right] = S_1 \left[ \left\{ y_i - \frac{1}{n} \sum_{i=1}^n y_i \right\}_1^n \right]$$

Şeklinde hesaplanır.

Örneğin  $m = 1 + 1, j = 2$  olsun;

$$f_2^2(x_2) = S_2 \left[ \left\{ y_i - \hat{f}_0 - \sum_{k=1}^{2-1} \hat{f}_1^2(x_1) - \sum_{k=3}^p \hat{f}_3^1(x_k) \right\}_1^n \right] = S_2 \left[ \left\{ y_i - \frac{1}{n} \sum_{i=1}^n y_i - \hat{f}_1^2(x_{i1}) \right\}_1^n \right]$$

Bu yapı tekrarlanarak 2.yenilemede düzeltme fonksiyonlarının tümünden  $\hat{f}_j^2(x_j)$  tahmin edilebilir.

Algoritmamızı bir adım daha yenilersek  $m = 2 + 1 = 3, j = 1$  olacak ve aşağıdaki gibi hesaplanacaktır;

$$f_1^3(x_1) = S_1 \left[ \left\{ y_i - \hat{f}_0 - \sum_{k=1}^{1-1} \hat{f}_1^3(x_1) - \sum_{k=3}^p \hat{f}_k^2(x_k) \right\}_1^n \right]$$

Burada  $\sum_{k=2}^p \hat{f}_k^2(x_k)$  bir önceki adımda hesaplanmıştı, fakat  $\sum_{k=1}^{1-1} \hat{f}_1^3(x_1) = \sum_{k=1}^0 \hat{f}_1^3(x_1)$  eşitliği mevcut değildir. Benzer döngüyle  $m = 2$ , ikinci yenileme devam edildiğinde  $\hat{f}_1^3(x_1)$  hesaplanabilecektir. Bu yenilemeyi devam ettirerek tüm düzgülleştirme fonksiyonları hesaplanabilecektir (47).

### 2.3. Dağıtılmış Gecikmeli Doğrusal Olmayan Modeller-DGDOM (Distributed Lag Non-Linear Models -DLNM)

Dağıtılmış Gecikmeli Modeller (DGM) ilk olarak 1965 yılında Almon tarafından ekonometri alanında yapılan çalışmalarda kullanıldı, daha sonra 2000 yılında Schwartz tarafından epidemiyolojik çalışmalarda yerini aldı (68). Dağıtılmış gecikmeli modellerin zaman serisi verilerine uygulanmasını ilk olarak 2006 yılında Armstrong gerçekleştirdi. 2010 yılında Gasparrini'nin yaptığı çalışmalarla, DGM modellerinin genişletilmiş hali olan, doğrusal olmayan yapılara uygun DGDOM modelinin ortaya çıkmasını sağlamıştır (68).

İstatistiksel regresyon modelinin asıl amacı bir cevap değişkeni ile tahmin ediciler seti arasındaki ilişkiyi tanımlamak ve sonrasında ilişkiye yön veren faktörleri tahmin etmektir. Fakat bir modelde bazı gecikmiş etkilere dayalı yapılar mevcutsa kompleks bir durumla karşılaşılır. Burada tahmin edicinin özel bir durumu söz konusudur. Buna maruz kalma olayı denir ve sonuçlar bir olayın görülme zamanının ötesinde bir zaman aralığı ile

ilişkilendirilir. Bu çerçevenin en önemli özelliği, ilişkiyi karakterize etmek için ek bir boyutun tanımlanmasıdır; buda gecikme ölçeği üzerindeki maruziyet ve sonuç arasındaki zamansal bağımlılığı ifade eder. Zaman serileri literatüründen ödünç alınan gecikme (lag) terimi, gecikme etkisi değerlendirilmek istendiğinde cevap ile maruz kalınan olay arasındaki zaman aralığıdır. Belirli bir maruziyet olayının, gelecekteki çok sayıda sonuç üzerinde etki yarattığını söylenebilir veya alternatif olarak, geçmişte birçok maruziyet olayının sonuca katkı sağladığı düşünülebilir (11). DGDOM gecikme yapısına maruz kalma-tepki ilişkisi olarak da adlandırılan, doğrusal olmayan gecikmeye dayanan yapıları tanımlamak için kurulan bir modelleme yöntemidir. DGDOM doğrusal yapıları olan DGM model yapısındaki katı doğrusal varsayımların yumuşatılması fikri ile ortaya çıkmış ve daha esnek uygulama alanları bulmuştur. DGDOM ve DGM metodolojisi aslında zaman serileri için geliştirilmiştir. Fakat son zamanlarda kohort, vaka-kontrol, boylamsal ve retrospektif analizlerde de sıklıkla kullanılmaya başlanmıştır (69).

Bazen maruz kalınan olayın etkisi sadece gözlemlendiği süre ile sınırlı olmayabilir, etkinin net olarak tespit edilmesi için değişkeni, gecikmeli olarak ölçülmüş zamanla ilişkilendirmek gerekir. Bu dönüşüm maruz kalınan etki ile tahmin edilen sonuç dizisi arasındaki modelleme sorununu ortaya koyar (11). DGDOM modelleri doğrusal olmayan modellerdir ve gecikmelerin etkisini açıklayabilen bir modelleme yapısına sahiptir (68). Çevre epidemiyolojisi zaman serisi çalışmalarında yüksek hava kirliliği ya da aşırı sıcaklara maruz kalanların sağlıklarının, birkaç gün süren bir zaman zarfında etkilendiği gözlenmiştir (70, 71). Çevresel stres ölçerlerinden olan maruz kalma-tepki ilişkisinin boyutunu, zamanı dahil ederek (geçmiş zamanın etkisini) tanımlama esnekliği sağlayan istatistiksel modeller son zamanlarda sıklıkla kullanılmaya başlanmıştır (72). DGDOM modeli son yıllarda hava kirliliğinin kısa vadeli etkilerini ölçmek için tasarlanmış istatistiksel yöntemlerin önemli gelişmelerden yararlanmıştır (73, 74). Gecikmeli etki ile başa çıkabilen çeşitli yöntemler arasında hatırı sayılır bir itibara kavuşmuş olan DGDOM yöntemine, hava kirliliği ve sıcaklığın insan sağlığı ile ilişkisinin değerlendirilmesinde sıklıkla başvurulmuştur (71, 75, 76).

DGDOM yöntemi aynı anda hem maruz kalma-tepki ilişkisini hem de geçmiş zaman etkisini modelleyen bir çerçevede geliştirilmiştir (72). DGDOM'nin metodolojisi, çapraz-taban (cross-basis) adı verilen hem tahmin uzayını hem de oluşan gecikme etki (lag effect) boyutunun ilişki şeklini aynı anda tanımlayabilen iki boyutlu bir uzaya dayalıdır (72). Bu bakış açısıyla DGDOM, daha önce bu model gibi kullanılan model

yelpazesi için birleşik bir çerçeve oluşturmanın yanı sıra daha esnek varyasyonlar sağlayan yeni bir model haline gelmiştir (72).

DGDOM'de temel model aşağıdaki gibi tanımlanır. Bağımlı değişken  $Y_t$  için ( $t = 1, \dots, n$ ) olmak üzere genel algoritması aşağıdaki gibidir:

$$g(\mu_t) = \alpha + \sum_{j=1}^J s_j(x_{tj}; \beta_j) + \sum_{k=1}^K \gamma_k u_{tk}$$

Burada  $\mu \equiv E(Y)$ ,  $g$  tek düze link fonksiyonudur ve  $Y$ 'nin üstel dağılıma sahip olduğu varsayılır (77, 78).  $s_j$  ile gösterilen fonksiyon düzgünleştirici fonksiyondur ve  $\beta_j$  parametre vektörü tarafından tanımlanan lineer tahmin ediciler ile  $x_j$  değişkeni arasındaki düzgünleştirilmiş ilişkileri belirtir (72).  $u_k$  değişkenleri  $\gamma_k$  katsayıları ise ilişkileri tanımlanan diğer lineer tahmin edicileri içerir (73).  $s_j$  fonksiyonları Genelleştirilmiş eklemeli modellere dayalı parametrik olmayan metod yapıları ile benzerdir (51, 72). Çevresel faktörlerin zaman serileri analizlerinde cevap değişkeni  $Y_t$  ortalaması  $\mu$ , varyansı  $\emptyset\mu$  olan doymuş Poisson dağıldığı varsayılır, kanonik link fonksiyonuna sahiptir ve genellikle günlük sayımlar halindedir (72). Mevsimsellik ve uzun vadeli eğilimler olarak tanımlanan, zamanla değişen karıştırıcıların etkisini elde etmek için modele bir zaman düzgün fonksiyonu ve bunun yansira lineer olmayan partiküler madde, nem, ve sıcaklık gibi meteorolojik faktörler dahil edebilir (79). Bunlara ek olarak yaş grupları ya da haftanın günleri (hafta içi/sonu) gibi kategorik değişkenlerin faktör olarak modele eklendiği örneklere de rastlamak mümkündür (79).

### 2.3.1. DGDOM'de s Düzgünleştirme Fonksiyonu Yapısı

$x$  ile  $g(\mu)$  arasındaki ilişki lineer terimlerin bir toplamı olarak genelleştirilmiştir.  $s_j$  fonksiyonu ise  $\beta_j$  parametre vektörü tarafından tanımlanan lineer tahmin ediciler ile  $x_j$  arasındaki ilişkiyi tanımlar (11). Lineer olmayan maruz kalma-cevap yapısı regresyon modeline, uygun seçilmiş  $s$  fonksiyonu sayesinde dahil edilmiş olur. Bu fonksiyon GAM'daki düzgünleştirme fonksiyonuna benzer. Bu fonksiyonlar modele çeşitli temel fonksiyon uygulamalarından biri seçilerek eklenir (30). Birçok farklı temel fonksiyon yapısı mevcuttur. Çevresel faktörlerin doğrusal olmayan sağlık etkilerini açıklamak üzere seçilen temel fonksiyonlar araştırmanın spesifik yapısı, yorumu ve ilişki şekli hakkındaki varsayımlara bağlı olarak yapılır. Örneğin çevresel faktörlerin ozon ve sıcaklık ölçümleri olduğu bir çalışma düşünelim. Ozon ve sıcaklığın doğrusal olmayan gecikmeli etkileri, tahmin edici ve gecikmelerin iki boyutu boyunca ilişkiyi tanımlayabilen  $s_j$  fonksiyonları

yardımı ile modelle eklenir. Temel fonksiyonlar; tipik olarak, polinomial veya eğri fonksiyonları gibi düzgünleştirme eğrilerinden oluşan veya kesikli doğrusal bir fonksiyonla temsil edilen doğrusal bir eşik parametre kullanılmasına dayanır (11, 80). Yukarıda açıklanan basit temel fonksiyonun genel bir sunumu aşağıdaki gibidir;

$$s(x_t; \beta) = z_t^T \beta$$

Temel fonksiyonların orijinal  $x$  vektörüne uygulanması ile elde edilen  $(n \times v_x)$  boyutlu  $Z$  temel matrisinin t-nci satırı  $z_t$  dir. Burada  $v_x$  serbestlik derecesi mantığında çalışır ve fonksiyonun esneklik derecesinin göstergesidir.  $g(\mu_t)$  denklemindeki model matrisine  $Z$  dahil edilerek bilinmeyen  $\beta$  parametreleri tahmin edilebilir (11). Açık kodlu R yazılımında yer alan “dlnm” paketi kullanıcısına, farklı tipte temel fonksiyon yapıları (Ceza zincirler, Doğal kübik veya B zincirleri gibi) seçme imkânı sunar. Dummy değişkenler ise kategoriler özelleştirme veya doğrusal bir eşik parametreleştirme teknikleri kullanarak seçebilir (11). Bu fonksiyonların tümü modele doğrusal terimler olarak dahil edilecek bir dizi orijinal tahmin edicinin dönüşümün sağlamak içindir (30).

### 2.3.2. DGDOM Cebirsel Yapısı ve Çapraz Taban Konsepti

Çapraz taban gecikmeler ve tahmin ediciler uzayı boyunca bağımlılık yapısını tanımlayan iki boyutlu uzayda tanımlanan fonksiyondur. Çapraz temelli fonksiyonlar var olan veya kullanıcı tarafından tanımlanan regresyon zinciri, polinom ve doğrusal eşik gibi fonksiyonları uygulayarak üretilen iki boyutta temel fonksiyonları birleştirerek oluşturulmuştur (11). Çapraz taban yöntemi ile değişkenler tüm gecikmeleriyle birlikte modele tek bir tahmin edici olarak eklenebilir. Dönüştürülmüş değişkenlerin çapraz temel matrisi, ilişkili parametreleri tahmin etmek için bir regresyon modelinin formülüne dahil edilir. Doğrusal olmayan etkileri açıklayabilmek için çeşitli yapılar önerilmiştir. Bunlardan bazıları, parçalı (piecewise) veya polinomial parametrize yöntemleridir. Her bir terime bir eşik uygulamak suretiyle doğrusal terimleri ya da parçalı fonksiyon yardımıyla kuadratik terimleri bir  $C$  sabitler matrisine uygulayarak bir DGM kurulabilir.  $C$  belirli temel fonksiyonların  $\ell$  gecikme vektörüne uygulanmasından türeyen  $(L + 1) \times v_\ell$  boyutlu bir matristir (81). Gecikmeler boyunca tahmin uzayını ilişkilendirebilen doğrusal olmayan yeni bir model çerçevesi düşünüldüğünde karşımıza DGDOM yapısı çıkmaktadır (72).

DGDOM cebirsel gösterimi üç boyutlu dizilimler içeren oldukça kompleks bir yapıya sahip olsa da çapraz-taban yapısının temel kavramları basittir. Çapraz-tabanın  $x$  ve onun gecikmeli dağılımının ilişki şeklini eş zamanlı tanımlayabilen iki boyutlu bir

işleve sahiptir. Bir çapraz taban seçmek, çapraz temel fonksiyonlar üretmek üzere birleştirilecek iki temel fonksiyon setini seçmektir (72). Bu iki alanın her birinin ilişki yapısını modelleyebilmek için iki dönüşümü aynı anda uygulamamız gerekir. İlk olarak  $Z$ 'yi açıklayan tahmin uzayını tanımlayabilmek için bir  $x$  tabanı seçilir. Daha sonra  $Z$ 'de yer alan  $x$  değişkeninin gecikme boyutları bu yapıya eklenir (72, 81)

$x$ 'in temel değişkenlerinin her birinin gecikmelerini temsil eden  $(L + 1) \times (v_l \times n)$  boyutlu simetrik bir matris olan  $\dot{R}$  matrisi üretilir. Bu yapı temel fonksiyonların,  $x$  değişkenlerinin gecikmelerinin yer aldığı  $Q$  matrisinin her sütununa direk olarak uygulanmasına da imkân tanır (82). Buradan yola çıkarak DGDOM aşağıdaki gibi tanımlanabilir:

$$s(x_t; \eta) = \sum_{j=1}^{v_x} \sum_{k=1}^{v_l} r_{tj}^T \cdot c_{.k} \eta_{jk} = w_t^T \cdot \eta$$

Formülde yer alan  $r_{tj}$ ,  $j$ -nci temel fonksiyonun  $t$  zaman için dönüştürülmüş gecikmeli vektörüdür.  $w_t$  vektörü ise  $v_x \times v_l$  uygulanarak elde edilen  $x_t$  için çapraz tabanlı fonksiyondur ve  $W$  çapraz taban matrisinin  $t$ -nci satırıdır.  $W$  matrisi özetle değişkenlerin uygun dönüşümlerle farklı matrislere dönüşümünü sağlayan  $Z$  matrisi ile değişkenlerin gecikmeleri ile yeniden boyutlanmasını sağlayan  $C$  matrisinin kombine edilmesi ile oluşturulur (11).  $W$  matrisini elde etmek için tensor bir yapıya ihtiyacımız vardır.  $P_{i,j}$   $i \times j \times 1$  boyutunda bir matrisi  $i \times j$  boyutunda bir matrise dönüştürebilen bir operatör olduğunu varsayalım;

$$\dot{A} = (1^T \otimes \dot{R}) \odot (1 \otimes P_{1,3}(C) \otimes 1^T)$$

Formüldeki 1 uygun boyutta bir değerlerinden oluşan vektördür.  $\otimes$  Kronecker çarpımını  $\odot$  ise Hadamard çarpımını temsil etmektedir.  $n \times (v_l \cdot v_l) \times (L + 1)$  boyutlu  $\dot{A}$  matrisi üçüncü boyut gecikmeleri boyunca toplanarak çapraz-tabanın final matrisi  $W$  elde edilir. Yukarıdaki eşitlik Tensör yapısı yardımıyla çok boyutlu bölmeler üzerinde düzgünleştirmeyi uygulamaya fırsat verir. Çapraz-taban yaklaşımını diğer yöntemlerden ayıran en önemli nokta boyutların her birini farklı bir model olarak ele almasıdır. Klasik modelleme yaklaşımlarında iki farklı değişken uzayda düzgün bir yüzeyde tanımlar, DGDOM'de ise farklı olarak bir değişkenin uzaydaki ve gecikme boyutundaki etkilerini eş zamanlı olarak tanımlanabilir (82, 83).

## 2.4. Otoregresif Terimli Genelleştirilmiş Eklemeli Modeller (Generalized Additive Model with Autoregressive Terms – GAMAR)

Zaman serisi yöntemi ile tahmin etme ve modelleme birçok pratik alanda esaslı bir öneme sahiptir. Bu yüzden son yıllarda zaman serileri üzerine çok sayıda önemli araştırma hâlihazırda devam etmektedir (84). Zaman ile değişim gösteren değişkenlerin karakteristiği ve istatistiksel anlamlılığını ortaya koyabilen yöntemler zaman serisi analizleridir. Zaman serisi modellemesinin temel amacı; bir zaman dizisinin geçmiş gözlemlerini toplayıp inceledikten sonra, serilerin yapısını tanımlayan uygun bir model geliştirmektir. Böylece Zaman serisi analizi geçmişini anlayarak geleceği tahmin etme eylemi olarak da tanımlanabilir (85).

Zaman serisi modelleri stokastik süreçlerin farklı sunum ve formlarına sahiptir. Literatürde genel anlamda iki lineer zaman serisi modelinin kullanımına sıklıkla rastlanır bunlar; Otoregresif terim (AR), Hareketli ortalama (MA) ve bu iki yapının birleşimi olan ARMA (Otoregresif hareketli ortalama serileri) (86, 87).

AR modeli,

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + e_t$$

Bir AR modelini bağımlı değişkenin gecikme değerlerinin bir fonksiyonu olarak tanımlanabilir (88). Yukarıdaki modelde yer alan  $e$  hata terimini ifade eder. Bu model  $p$ -nci derecede bir AR modeline örnektir gösterimi AR(p) şeklindedir (89).

MA modeli;

$$X_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

Şeklinde ifade edilir. Eğer seride gecikmeli hata terimi şu an ki hata terimi ile ilişkili ise hareketli ortalama süreci tanımlanabilir. Bir hareketli ortalama serisinde değişken tahmin değeri hata terimlerinin tahmini ile yapılır (88, 90). Yukarıdaki model  $q$ 'ncü dereceden hareketli ortalama serisine bir örnek gösterimdir. Burada  $e \sim WN(0, \sigma^2)$  beyaz gürültü (white noise) serisine sahiptir. Beyaz gürültü sıfır ortalama ve sabit varyans anlamına gelmektedir. Dağılım olarak Normal dağılım gösterirse Gausscu beyaz gürültü olarak adlandırılır (91).

ARMA modeli;

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

Çoğu zaman serisi hem AR hem de MA sürecini birlikte içermektedir. Bu süreçleri modellemek için Otoregresif hareketli ortalama serileri (ARMA) modeli türetilmiştir. Bu model AR ve MA modellerinin matematiksel birleşimi olduğu görülmektedir (92).

Zaman serileri analizinin uygulanabilmesi için bazı ön şartların sağlanması gereklidir. Bunlar beyaz gürültü ve durağanlık ön şartlarıdır. Zaman serisi durağan olmaz ise deterministik ya da stokastik trend içerecektir. Bu durum da öngörme ve regresyon tahmininde yanlışlıklara sebep olacaktır (92).

GAM çevresel faktörlerin sağlık üzerine etkilerini doğrusal olmayan zaman serisi teknikleri ile etkili ve esnek bir şekilde modelleyebilir (12). GAM hataların karşılıklı olarak bağımsız olduğunu varsayar fakat zaman serilerinde ardışık zaman noktaları birbiri ile ilişkili olabilir. Bu ilişki durumuna bir düzeltme getirmek ve bu konudaki eksikliği gidermek amacıyla otoregresif terime (AR) sahip bir GAM modeli önerilmiştir (12). Bu modele GAMAR (Oto-regresif terimli genelleştirilmiş eklemeli model) adı verilmektedir. GAMAR modelleri GARMA (Genelleştirilmiş Oto-regresif Hareketli Ortalama-Generalized Autoregressive Moving Average) modelden türetilmiştir (12).

GARMA modelleri 2003 yılında Benjamin tarafından bulunmuştur (93). Benjamin tek değişkenli zaman serisi modelini, Gaussian olmayan zaman serilerine uyarlayarak GARMA modellerini buldu (93). GARMA modellerinin tahmin yeteneği klasik ARMA modellerinden çok daha iyidir (94). Model tahmini, iterasyonlu ağırlıklandırılmış en küçük kareler algoritması kullanılarak gerçekleştirilir (93). Bu model dış etkili değişkenleri içerebildiğinden dolayı durağan olamayan model gibi davranır (95). GARMA modeli zamana bağlı Poisson, Negatif binom veya Binom dağılımı gibi kesikli ya da Gama dağılımlı sürekli cevap değişkenlerini ya da açıklayıcı değişkenleri modellemek için kullanılabilir (95).

GARMA modelleri  $t = 1, \dots, n$  olmak üzere  $y_t$  bağımlı değişkeni, önceki bilgi kümesi  $H_t = \{X_1, \dots, X_t, y_1, \dots, y_{t-1}\}$  ve kovariyet vektörünün  $X_t = (X_{t1}, \dots, X_{tm})$  aynı üstel dağılıma sahip olduğunu varsayar (95). GLM'de olduğu gibi koşullu ortalama  $\mu_t$ , iki kere türevlenebilen monotonik  $g$  link fonksiyonu uygulamasına tabidir. Link fonksiyonuna sahip GARMA, GLM yapısından farklı olarak Oto-regresif terim ve Hareketli ortalama bileşenlerinin formüle eklenmesine izin verir (95).

$$g(\mu_t) = \sum_{i=1}^m X_{ti}\beta_i + \sum_{j=1}^p c_j \left( g(y_{t-j}) - \sum_{i=1}^m X_{t-j,i}\beta_i \right) + \sum_{j=1}^q d_j \left( g(y_{t-j}) - g(\mu_{t-j}) \right)$$

Burada  $\sum_{j=1}^p c_j \left( g(y_{t-j}) - \sum_{i=1}^m X_{t-j,i}\beta_i \right)$  kısmı Oto-regresif terim  $\sum_{j=1}^q d_j \left( g(y_{t-j}) - g(\mu_{t-j}) \right)$  bölümü Hareketli ortalama'dır. Sayı ile elde edilen verinin Poisson dağıldığı varsayalım. Buna göre Poisson GARMA alt modeli;



$$\ln(E(y_t)) = \ln(\mu_t) = \sum_{i=1}^m X_{ti}\beta_i + \sum_{j=1}^p c_j \left( \ln(y_{t-j}^*) - \sum_{i=1}^m X_{t-j,i}\beta_i \right) + \sum_{j=1}^q d_j (\ln(y_{t-j}^*/\mu_{t-j}))$$

Şeklinde. Denklemdaki  $y_t^* = \max(y_t, \tau)$ ,  $\tau$  pozitif bir eşik parametresidir.  $\ln()$  yapısı 0 veya negatif değerlerde tanımsız olduğundan, sıfır ya da negatif değerli  $y$  değeri yerine  $\tau$  kullanılır (12).

GAMAR modelleri; GARMA'dan MA terimi hariç bırakılarak türetilmiştir ve algoritması ise aşağıdaki gibidir:

$$g(E(y_t)) = g(\mu_t) = \sum_{i=1}^n X_{ti}\beta_i + \sum_{i=1}^m s_i(X_{ti}) + \sum_{j=1}^p c_j \left( g(y_{t-j}) - \sum_{i=1}^m s_i(X_{t-j,i}) \right)$$

Formüldeki  $\sum_{i=1}^m s_i(X_{ti})$  bağımsız değişkenlerin düzleştirme fonksiyonlarıdır,  $\sum_{j=1}^p c_j (g(y_{t-j}) - \sum_{i=1}^m s_i(X_{t-j,i}))$  kısmı ise otoregresif terimdir. GAMAR formülleri GAM ile karşılaştırıldığında otoregresif terimin link tahmin edicisinin içine eklendiği görülmektedir (12). Bağımlı değişken sayılarından oluştuğunda Poisson alt model;

$$\ln(E(y_t)) = \ln(\mu_t) = \sum_{i=1}^n X_{ti}\beta_i + \sum_{i=1}^m s_i(X_{ti}) + \sum_{j=1}^p c_j \left( \ln(y_{t-j}^*) - \sum_{i=1}^m s_i(X_{t-j,i}) \right)$$

Burada  $y_t^* = \max(y_t, \tau)$ ,  $\tau$  pozitif bir eşik parametresidir. GAMAR parametreleri kısmi en çok olabilirlik yöntemi ile tahmin edilebilir (12).

### 3. MATERYAL VE METOT

#### 3.1. Çalışmada Kullanılan Veri Seti

Bu Çalışma İnönü Üniversitesi Bilimsel Araştırmalar Etik Kurulu (Sağlık Bilimleri Girişimsel Olmayan Klinik Araştırmalar Etik Kurulu) 2018/2-8 protokol numaralı izni ile onaylanmıştır (Ek-2). Bu çalışmada gözlemlerimiz Gaziantep ili Kamu Hastaneleri Genel Sekreterliğine bağlı Dr. Ersin Arslan Eğitim ve Araştırma Hastanesi, 25 Aralık Devlet Hastanesi ve Gaziantep Üniversitesi Tıp Fakültesi Hastanesinin Acil servis ve Göğüs hastalıkları servisine başvuran 16 yaş ve üzeri hastaların ilgili “servislere başvuruları”, “hastanede yatış” durumları ve “mortalite” durumları 01 Ocak 2009 ile 31 Mart 2014 tarihleri arasında toplam 1916 gün boyunca geriye dönük olarak elde edilen veriler kullanılarak oluşturulmuştur (4). Çalışmamızda kullanılan bağımlı değişkenlere ait tanımlayıcı bilgiler Tablo 3.1’de sunulmuştur.

**Tablo 3.1.** Çalışmada kullanılan bağımlı değişkenlerin tanımları

<b>Değişkenler</b>	<b>Değişkenin Ölçüm Tipi</b>	<b>Değişkenin Aldığı Değerler/Birimi</b>	<b>Değişken Rolü</b>
Astım şikayeti ile başvuran hasta sayısı	Nümerik	Doğal Sayı	Bağımlı/Hedef
Astım şikayeti ile yatan hasta sayısı	Nümerik	Doğal Sayı	Bağımlı/Hedef
Astım nedeniyle eks olan sayısı	Nümerik	Doğal Sayı	Bağımlı/Hedef
KOAH şikayeti ile başvuran hasta sayısı	Nümerik	Doğal Sayı	Bağımlı/Hedef
KOAH nedeniyle yatan hasta sayısı	Nümerik	Doğal Sayı	Bağımlı/Hedef
KOAH nedeniyle eks olan sayısı	Nümerik	Doğal Sayı	Bağımlı/Hedef
Pnömoni şikayeti ile başvuran hasta sayısı	Nümerik	Doğal Sayı	Bağımlı/Hedef
Pnömoni nedeniyle yatan hasta sayısı	Nümerik	Doğal Sayı	Bağımlı/Hedef
Pnömoni nedeniyle eks olan sayısı	Nümerik	Doğal Sayı	Bağımlı/Hedef
Pulmoner Tromboemboli nedeni ile yatan hasta sayısı	Nümerik	Doğal Sayı	Bağımlı/Hedef
Pulmoner Tromboemboli nedeniyle eks olan sayısı	Nümerik	Doğal Sayı	Bağımlı/Hedef

Çalışmamızda kullanılan bağımsız değişkenlere ait tanımlayıcı bilgiler ise Tablo 3.2’de sunulmuştur.

**Tablo 3.2:** Çalışmada kullanılan bağımsız değişkenlerin tanımlayıcı bilgiler

<b>Değişkenler</b>	<b>Ölçüm Tipi</b>	<b>Aldığı Değerler/Birimi</b>	<b>Değişken Rolü</b>
Günlük ortalama sıcaklık	Nümerik	°C	Bağımsız/Tahminleyici
Partikül madde (PM10)	Nümerik	µg/m <sup>3</sup>	Bağımsız/Tahminleyici
Zaman	Nümerik	Doğal sayı	Bağımsız/Tahminleyici
Çöl fırtınası varlığı	Kategorik	Var/Yok	Bağımsız/Tahminleyici
Haftanın günleri	Kategorik	Hafta içi/sonu	Bağımsız/Tahminleyici

Birçok epidemiyolojik çalışma günlük mortalite ile havadaki partikül madde (PM) arasındaki ilişkiyi kanıtlamaktadır. Partiküller kurum, sülfat ve nitrat partikülleri, asidik gazlar, ulaşım sanayi ve enerji üretimi için kullanılan fosil yakıtlardan oluşmaktadır. Partiküller akciğerin en uç kısımlarına kadar inhale edildikleri için bu durum insan sağlığını tehdit eden ciddi bir sağlık sorunu haline gelmektedir (96).

Hava kirliliğinin akut reaksiyonlarının sağlık üzerine etkilerini araştıran birçok çalışma haftanın günleri, grip salgını ve tatil günleri gibi değişkenleri modellerine eklemiştir (97). Çalışmalar haftanın farklı günlerinde kirlilik konsantrasyonun farklılık gösterebileceğini tespit etmiştir. Benzer şekilde hafta sonu hava kirliliğinin düşük seviyede olduğu fakat hafta içi özellikle pazartesi gününde yüksek seviyelerde seyrettiğini gözlenmiştir. Buradan yola çıkarak hava kirliliğine maruz kalma değişikliğinin hastalığı tetiklemede etkili olduğu düşünülerek haftanın günleri kategorik değişkeni modele dâhil edilmiştir (98).

Kirlilik kaynaklı mortalite (günlük) tahminleri yapılırken zamanla değişen hava durumu değişkenleri, mevsim ve grip salgını gibi faktörler yanlılığa sebep olmaktadır. Bu tip karıştırıcı yanlılıklara neden olan faktörleri elemine etmenin yolu modele zaman veya sıcaklık değişkeninin eklenmesidir (99). Çalışmamızda modelimize zaman değişkeni eklenmiş bu sayede yanlılık elemine edilmiştir. Çevresel epidemiyoloji çalışmalarında değişkenler mevsimsel etkiden arındırılmalıdır. Tahmin doğruluğunu etkilediğini düşündüğümüz mevsimsellik sorunu da zaman değişkenini düzgünleştirme fonksiyonu yardımıyla modele ekleyerek elemine edilmiştir.

Çalışmamızda Genelleştirilmiş eklemeli modelleri analiz etmek için R programının “mgcv” paketi kullanılmıştır. Bu paket herhangi bir değişkende kayıp gözlem varsa kayıp gözlemin olduğu satırı değerlendirmeye almaz. Kayıp gözlemler için daha önce yapılmış çalışmalar incelendiğinde iki yol izlendiği görülmüştür. Birincisi kayıp olan gözlemlere ait değerlere BM yöntemi kullanılarak atama yapılması, ikincisi ise kayıp olan gözlemlere ait satırın silinmesi seçeneğidir. Çalışmamızda bahsi geçen iki yöntemle de denemeler yapılmış olup hangi yaklaşımın modelleme yeteneğinin daha iyi olduğu belirlenmiştir. Buna göre AIC bakımından kayıp gözlemler silinerek yapılan analizin, kayıp gözlem yerine atamalar yapılan analize göre daha üstün sonuçlar verdiği gözlenmiştir

### 3.1.1. Genelleştirilmiş Eklemeli Modeller için Kurulan Model Yapıları

Çalışmamızda gerçek cevap değişkenine en yakın tahmini yapabilen GAM modelini önermek için farklı GAM modelleri denenmiş ve bu modellerin tahmin performansları kıyaslanmıştır. Aşağıda bağımlı değişken için dört farklı model denemesi yapılmıştır. Tüm bağımlı değişkenlerde bu dört model yapısı uygulanmıştır. Model yapıları aşağıdaki gibidir:

İlk model denemesi tüm bağımsız değişkenler ve PM10 değişkeninin gecikmelerinin modele eklenmesi ile yapılmıştır. Daha önce yapılan çalışmalarda PM10 değişkeni modele dört gecikme ile eklendiği gözlenmiştir (7). Model I yapısı aşağıdaki gibidir:

$$\text{Model I: } \log\{E[\text{Bağımlı değişken}_i]\} = (\text{Haftanın günleri}_i) + (\text{Fırtına}_i) + (\text{Pm10}_i) + (\text{Pm10}_{1i}) + (\text{Pm10}_{2i}) + (\text{Pm10}_{3i}) + (\text{Pm10}_{4i}) + (\text{Günlük ort sıcaklık}_i) + (\text{Zaman}_i)$$

İkinci model; zaman değişkeni düzgünleştirme fonksiyonu yardımıyla eklenerek düzeltilmiştir. Zaman değişkenindeki düzgünleştirme fonksiyonunun temel fonksiyon sayısı  $k$  üzerinde de düzeltme yapılmıştır. Düzgünleştirme fonksiyonu  $k$  serbestlik derecesini kontrol eder ve  $k = sd - 1$  şeklindedir. R programında “mgcv” paketinde gam fonksiyonunda temel fonksiyon sayısı  $k$  varsayılan değeri 10 dur. Daha önce yapılan çalışmalarda zaman değişkeninin temel fonksiyon sayısı  $k$  belirlenirken verideki mevsim sayısı dikkate alınarak düzeltme yapılmıştır (39). Bizim verimizde toplamda 21 mevsim mevcuttur. Zaman değişkeni düzgünleştirme fonksiyonu yardımı ile modele eklenmiş ve düğüm sayısı 21 olarak düşünülmüştür. Fakat 21 düğümlü tahmin edici ile bağımlı değişken tahmininin zamana karşı çizilen saçılım grafiğinde  $k$  değerinin büyük

olmasından kaynaklı aşırı uyum problemi olduğu gözlenmiştir. Daha önce yapılan çalışmalarda referans alınarak zaman değişkeni için serbestlik derecesi iki olarak belirlenmiştir (12). Ayrıca modelimizde çevresel epidemiyoloji çalışmalarında sıklıkla kullanılan üç farklı düzgünleştirme fonksiyonun performansı denenmiş ve en iyi performansı gösteren modelin ikinci model olmasına karar verilmiştir. Denemesi yapılan düzgünleştirme fonksiyonları; Döngüsel kübik regresyon zinciri (Dkrz), Kübik regresyon zinciri (Krz) ve İnce tabaka regresyon zinciridir (İtrz). Model II olmasına karar verilen model aşağıdaki gibidir:

$$\begin{aligned} \text{Model II: } \log\{\mathbb{E}[\text{Bağımlı değişken}_i]\} &= (\text{Haftanın günleri}_i) + (\text{Fırtına}_i) + \\ &+ (\text{Pm10}_i) + (\text{Pm10}_{1i}) + (\text{Pm10}_{2i}) + (\text{Pm10}_{3i}) + (\text{Pm10}_{4i}) + \\ &+ (\text{Günlük ort sıcaklık}_i) + \\ &f(\text{Zaman}_i, sd: 2, \text{"en iyi performansı gösteren düzgünleştirme fonksiyonu"}) \end{aligned}$$

Üçüncü modelimizde ikinci modelde denemesi yapılan düzgünleştirme fonksiyonları üzerinde Wang ve arkadaşlarının gerçek veri setinde model tahmin başarısını düzelttiğini savundukları serbestlik derecesi olan altı değeri kullanılarak oluşturulmuştur (100, 101). Model ikide uygulanan süreç model III için de uygulanmış altı serbestlik derecesine sahip üç farklı düzgünleştirme fonksiyonlu modelden en iyi performansa sahip model seçilmiştir. Yapılan denemeler sonucunda Model III olmasına karar verilen Genelleştirilmiş eklemeli model aşağıdaki gibidir:

$$\begin{aligned} \text{Model III: } \log\{\mathbb{E}[\text{Bağımlı değişken}_i]\} &= (\text{Haftanın günleri}_i) + (\text{Fırtına}_i) + \\ &+ (\text{Pm10}_i) + (\text{Pm10}_{1i}) + (\text{Pm10}_{2i}) + (\text{Pm10}_{3i}) + (\text{Pm10}_{4i}) + \\ &+ (\text{Günlük ort sıcaklık}_i) + \\ &f(\text{Zaman}_i, sd: 6, \text{"en iyi performansı gösteren düzgünleştirme fonksiyonu"}) \end{aligned}$$

Dördüncü modelimizde ikinci modelde denemesi yapılan düzgünleştirme fonksiyonları daha önceki yapılan çalışmalar referans alınarak serbestlik derecesi 12 değeri ile düzeltilmiştir (102). Model ikide uygulanan süreç model III için de uygulanmış 12 serbestlik derecesine sahip üç farklı düzgünleştirme fonksiyonlu modelden en iyi performansa sahip model seçilmiştir. Yapılan denemeler sonucunda Model IV olmasına karar verilen Genelleştirilmiş eklemeli model aşağıdaki gibidir:

$$\begin{aligned} \text{Model VI: } \log\{\mathbb{E}[\text{Bağımlı değişken}_i]\} &= (\text{Haftanın günleri}_i) + (\text{Fırtına}_i) + \\ &+ (\text{Pm10}_i) + (\text{Pm10}_{1i}) + (\text{Pm10}_{2i}) + (\text{Pm10}_{3i}) + (\text{Pm10}_{4i}) + \\ &+ (\text{Günlük ort sıcaklık}_i) + \\ &f(\text{Zaman}_i, sd: 12, \text{"en iyi performansı gösteren düzgünleştirme fonksiyonu"}) \end{aligned}$$

Düzeltilen bu dört model çalışmamızda yer alan tüm bağımlı değişkenler için uygulanmıştır. Ayrıca GAM ve DGDOM yöntemleri için denenmiş ve hangi model ve yöntemin daha iyi performans sergilediği aşağıda bahsedilen model performans kriterleri esas alınarak karşılaştırılmıştır.

GAMAR modelleri ile yapılan simülasyon ve gerçek veri çalışmalarında düzgünleştirme fonksiyonu olarak doğal zincir fonksiyonu kullanılmış ve otoregresif terim 4 hesaplanmıştır (12, 66, 67). Çalışmamızda GAMAR modellerinde zaman değişkeninin düzgünleştirmesinde doğal zincir fonksiyonu kullanılmıştır ve otoregresif terim 4 kabul edilmiştir. GAM modellerindeki yapıya benzer olarak ilk modelden itibaren PM10 değişkeninin dört gecikmesi modele dahil edilmiş, sonraki modellerde zaman değişkeni doğal zincir fonksiyonu ile düzgünleştirilip model II serbestlik derecesi 2, model III serbestlik derecesi 6 ve model IV serbestlik derecesi 12 olarak belirlenmiştir. Tüm modellerde düzgünleştirme parametresine Genelleştirilmiş çapraz karşılaştırma yöntemi ile karar verilmiştir.

Tüm istatistiksel testler iki yönlü ve  $P < 0.05$  değerleri istatistiksel olarak anlamlı kabul edilmiştir. R yazılımı 3.2. sürümünde yer alan “mgcv” ve “dlnm” paketleri GAM ve DGDOM modellerini uydurmak, regresyon katsayıları ve standart hataların tahmini için kullanılmıştır. R yazılımı 3.4.4 sürümünde yer alan “gam”, “splines” paketleri ve “gam with AR” fonksiyonu kullanılarak GAMAR modelleri veriye uydurulmuştur (12).

### **3.1.2. Model Başarısının Test Edilmesi İçin Kullanılan Yöntemler**

Başlıca model seçim kriterlerinden bazıları; 1974 yılında Akaike tarafından bulunmuş olan Akaike Bilgi Kriteri (AIC) ve 1978 yılında Schwarz tarafından bulunmuş olan Bayesci Bilgi Kriteridir (BIC) (103). Bu kriterlere ek olarak minimum Artıkların Kısmi Oto-korelasyon Fonksiyonlarının (PACF) ve genelleştirilmiş çapraz geçerlilik teknikleri de karşılaşılan diğer model sınaama kriterleri olarak literatürde yerini almıştır (104). Eklemeli modelleme ve yukarıda bahsi geçen modellerin başarılarının kıyaslanmasındaki mantık klasik karşılaştırma yöntemlerinden biraz farklıdır (97). Modelleme ve eğri tipi çerçevesinden bağımsız olarak asıl düşünülmesi gereken konu karıştırıcı hataları önlemek için düzgünleştirme miktarlarının seçimidir (104). Bu seçim, zaman için eğride optimal (etkili) serbestlik derecesinin (sd) seçilmesine karşılık gelir (104). Veri tabanlı yöntem kullanımlarında genel strateji belirli bir kriteri optimize eden serbestlik derecesini seçmektir (102). Bu yaklaşımda bir dizi regresyon modeli bir

serbestlik derecesi aralığı kullanılarak çözümlenir (102). Özet olarak uygun serbestlik derecesine karar vermek için aşağıdaki kriterler değerlendirilmiştir:

- i. Serbestlik derecesinin sabitlenmesi: Literatür bilgi ve önceki çalışmalara dayalı olarak bir serbestlik derecesi belirlenmelidir. Serbestlik derecesine göre  $\hat{\beta}$  değişkenliğinin araştırılması için bir duyarlılık (sensitivity) analizinin yapılması (102) ve duyarlılık analizine göre  $\beta$  'yı  $sd = 1, 2, \dots, 20$  değerleri için tahmin edilmesi şeklinde uygulamalarına rastlanmıştır (97).
- ii. Akaike Bilgi Kriteri: Bu metottaki mantık AIC'yi minimum yapan serbestlik derecesi değerinin seçilmesidir. AIC genelde belirli kovariyetlerin ve zamanın düzgünleştirme fonksiyonlarının seçilmesinde kullanılır (102). Bir modelin serbestlik derecesi için AIC aşağıdaki şekilde hesaplanır;

$$AIC(df) = -2(\text{maksimum log olabilirlik}) + 2sd$$

- iii. Bayesci Bilgi Kriteri: Bu metot BIC'yi minimum yapan  $sd$  değerinin seçilmesi şeklinde uygulanır (102). Bu model yaklaşık bir sonsal model ağırlığını kullanır.  $n$  gözlem sayısı olmak üzere BIC aşağıdaki gibi hesaplanır;

$$BIC(df) = -2(\text{maksimum log olabilirlik}) + \log(n)sd$$

- iv. Minimum Oto-korelasyon Artıkları: Oto-korelasyon artıklarını minimum yapan  $sd$  değeri seçilir (97). Pratikte sabit sayıdaki gecikmeler için artıkların kısmi Oto-korelasyon Fonksiyonunun (PACF) mutlak değerinin toplamını en aza indirilebilir (97). Alternatif olarak uygulanan bir teknik ise artıkların içindeki beyaz gürültü değerleri için bir test kullanarak  $sd$  değerinin seçilmesidir (102).

Yukarıda bahsi geçen teknikler aynı zamanda modellerin tahmin başarılarını ölçmede de kullanılabilir. Örneğin bir istatistiksel model düşünelim. Bu modelin tahmin edici parametre sayısını  $k$  olduğunu varsayalım. Model için olabilirlik fonksiyonunun maksimum değeri de  $l$  olsun.  $AIC = -2l + 2k$  şeklinde hesaplanmaktadır. Veriler için bir dizi aday model hesaplandığında aralarından tercih edilecek model AIC değeri olarak en düşük değere sahip model olmalıdır (105). Denemelerde küçük AIC, muhtemelen tahmin edilen parametre sayısı en yüksek olan modele ait olacaktır. Bu durum hesaplanan AIC değerinde aşırı uyum problemi ile karşılaşılması anlamına gelmektedir. Düzeltilmiş AIC istatistiksel modele bağlı hesaplamalar yaparak bu sorun hakkında bir çözüm sağlamıştır (106). Formülü aşağıdaki gibidir.

$$AIC_c = AIC + \frac{2k^2 + 2k}{n - k - 1} = 2k \frac{n}{n - k - 1} - 2\ln(l)$$

Aynı şekilde  $BIC = -2l + k \log n$  şeklinde hesaplanır (103). Bunların yanı sıra ROC (Receiver Operating Characteristic) eğrisi altında kalan alan tahmin edicinin performansı değerlendirilirken duyarlılık, seçicilik, pozitif kestirim oranı ve negatif kestirim oranı gibi ifadeleri ayrı ayrı değerlendirmek yerine tek bir değerle ifade eden ölçüttür. Bu ölçütler kullanım yerlerine göre doğruluk (accuracy), C istatistiği, Odds Oranı istatistikleri olarak karşımıza çıkmaktadır (20). Üstel dağılım ailesine sahip dağılımlarda doymuş model  $\hat{\mu} = y$  olabilirlik hesaplanır. Fark (D) hesaplaması  $D = 2[l(\hat{\beta}_{max}) - l(\hat{\beta})]\phi$  şeklindedir (108). Formülde geçen  $l(\hat{\beta}_{max})$  doymuş modelin en çok olabilirliği,  $l(\hat{\beta})$  ise hesaplanmasını istediğiniz modelin en çok olabilirliğidir.  $l(\hat{\beta}_{max})$  verideki değişkenlerin hepsini içeren bir modeldir.  $\phi$  ise varyanstır (107).

Veriye uydurulan modelin anlamlılığını test etmek için birçok yöntem vardır. Bir veya birden fazla modelin istatistiksel anlamlılığını test etmek için başvurulan diğer bir yöntem de Anova gam'dır. Bu yöntem tek bir model için bakıldığında her bir parametre ve düzgülleştirme teriminin anlamlılığını wald istatistik yöntemi ile test ederken. Birden fazla model için bakıldığında ise veriye uydurulan modellerin performanslarının istatistiksel olarak anlamlı bir farklılık gösterip göstermediğini test eder (109).



## 4. BULGULAR

Çalışmamız 01 Ocak 2009 ile 31 Mart 2014 tarih aralığındaki kayıtları kapsamaktadır. Gözlemler tam sayım yöntemi ile çalışmaya dahil edilmiştir. Kayıp gözlemler çalışman dışında bırakılmıştır ve geriye kalan 1810 günün; 81 (%4.50) günde toz fırtınası olduğunu, 1290 (%71.30) günün hafta içi, 520 (%28.70) günün hafta sonu olduğu gözlenmiştir. Toplam 513841 hasta başvurusundan 197720 (%38.48) hastaya astım, 136016 (%26.47) hastaya KOAH, 177471 (%34,54) hastaya pnömoni, 2634 (%0.51) hastaya ise pulmoner tromboemboli teşhisi konulduğu gözlenmiştir. 1810 güne ait PM10 ve ortalama sıcaklık, değerlerine ait tanıttıcı istatistikler Tablo 4.1’de özetlenmiştir. Bağımlı değişkenlere ait tanıttıcı istatistikler ise Tablo 4.2’de sunulmuştur.

**Tablo 4.1.** Çalışmada kullanılan Nümerik Ölçümlü bağımsız değişkenlerin tanıttıcı istatistikleri

Değişkenler ( <i>n</i> =1810)	$\bar{X}$ (SS)	Min	M [Q1/Q3]	Maks	IQR
PM10 ( $\mu\text{g}/\text{m}^3$ )	88.6 (58.7)	11.0	72.0 [48.0/115.0]	631.0	67.0
Günlük Ort. Sıc. ( $^{\circ}\text{C}$ )	15.7 (9.1)	-2.6	14.5 [7.6/24.5]	34.9	16.9

SS: Standart Sapma, M: Medyan, IQR: Çeyrekler arası fark ( $Q_3-Q_1$ )

**Tablo 4.2.** Bağımlı değişkenlere ait tanıttıcı istatistikler

Değişkenler	$\bar{X}$ (SS)	Min	M [Q1/Q3]	Maks	IQR	
Astım	<i>Başvuru</i>	110 (80)	3	93 [45/151]	449	106
	<i>Yatan</i>	1 (2)	0	0 [0/2]	13	2
	<i>Eks</i>	0 (0)	0	0 [0/0]	1	0
KOAH	<i>Başvuru</i>	76 (40)	20	75 [35/104]	257	69
	<i>Yatan</i>	2 (3)	0	2 [0/3]	20	3
	<i>Eks</i>	1 (1)	0	1 [0/1]	6	1
Pnömoni	<i>Başvuru</i>	98 (67)	7	81 [49/129]	439	80
	<i>Yatan</i>	6 (5)	0	5 [1/8]	48	7
	<i>Eks</i>	0 (0)	0	0 [0/0]	3	0
Pulmoner Tromboemboli	<i>Yatan</i>	1 (2)	0	1 [0/2]	10	2
	<i>Eks</i>	0 (0)	0	0 [0/0]	2	0

SS: Standart Sapma, M: Medyan, IQR: Çeyrekler arası fark ( $Q_3-Q_1$ )

Verimizdeki nicel bağımsız değişkenler incelendiğinde; PM10 değişkeni dağılımı (86.6±58.7) olduğu ve en yüksek 631.0 µg/m<sup>3</sup> değeri aldığı gözlenmiştir. Ortalama sıcaklık değişkeni dağılımı (15.7±9.1) ve aldığı en yüksek değer 34.9 °C olarak gözlenmiştir (Tablo 4.1). Bağımlı değişkenler incelendiğinde en çok başvuru astım (110±80) şikayetinden olmuştur, astım başvurusunun maksimum değeri 449 hasta olarak gözlenmiştir. Başvuru bakımından astım değişkenini takip eden pnömoni (98±67) değişkenidir. Pnömoni değişkeninin aldığı en yüksek değer 439 hasta olarak gözlenmiştir.

#### **4.1. Genelleştirilmiş Eklemeli Modellerle Yapılan Analizlerin Sonuçları**

Gerçek cevap değişkenine en yakın tahmini yapabilen GAM modelini önermek için Bölüm 3.1.1’de bahsi geçen farklı dört GAM modeli belirlenmiş ilgili modellerin tahmin performansları kıyaslanmıştır. Modellerde bağımlı değişkenler astım, KOAH, pnömoni ve pulmoner tromboemboli vakaları nedeniyle hastanelerin acil ve ilgili servisine başvuran hasta sayısı, servise yatan hasta ve eks olan hasta sayılarının günlük değerleridir. Dört modelin GAM ile analizleri yapılmıştır ve modellerin AIC, düzeltilmiş AIC, BIC, fark yüzdesi ve düzeltilmiş R<sup>2</sup> gibi model tahmin performans değerleri incelenmiştir.

##### **4.1.1. Astım Hastalığı ile İlgili Bulgular**

Çalışmamızda astım hastalığı bağımlı değişkenleri ile ilgili gerçek cevap değişkenine en yakın tahmini yapabilen GAM modelini önermek için dört farklı GAM modeli denenmiş ve bu modellerin tahmin performansları kıyaslanmıştır. Astım hastalığı modellerinde bağımlı değişkenlerimiz astım vakası nedeniyle hastanelerin acil ve ilgili servisine başvuran hasta sayısı, servise yatan hasta ve eks olan sayılarının günlük değerleridir. Bölüm 3.1.1’de tanımlanan ve GAM’da modellenen dört farklı modelin AIC, düzeltilmiş AIC, BIC, düzeltilmiş R<sup>2</sup> ve fark yüzdesi değerleri Tablo 4.3’de sunulmuştur. Hangi modelin hava kirliliği ile astım vakası nedeniyle başvuran, yatan ve eks olan ilişkisini tahmin etmede daha başarılı olacağı araştırılmıştır.

**Tablo 4.3.** Astım ile hava kirliliği ilişkisinin GAM'ın farklı modelleri ile incelenmesi

Değişken		AIC	AICc	BIC	Fark (%)	Düz. R <sup>2</sup>	
ASTİM	Başvuru	<i>Model I</i>	35792.53	32792.63	35842.04	0.7456	0.7135
		<i>Model II</i>	35623.01	35623.11	35672.52	0.7474	0.7176
		<i>Model III</i>	32095.70	32095.80	32142.21	0.7841	0.7602
		<i>Model IV</i>	29574.22	29574.32	29623.73	0.8104	0.7911
	Yatan	<i>Model I</i>	4422.92	4423.02	4472.43	0.4605	0.4378
		<i>Model II</i>	4399.32	4399.42	4449.37	0.4668	0.4332
		<i>Model III</i>	4322.10	4322.20	4371.61	0.4883	0.4554
		<i>Model IV</i>	4280.63	4280.73	4330.14	0.5004	0.4647
	Eks	<i>Model I</i>	292.62	292.72	342.13	0.0716	0.0178
		<i>Model II</i>	291.37	291.47	340.88	0.0827	0.0206
		<i>Model III</i>	290.49	290.56	339.97	0.1125	0.0271
		<i>Model IV</i>	291.54	291.64	341.05	0.0815	0.0212

Düz. R<sup>2</sup>: Düzeltilmiş çoklu açıklayıcılık katsayısı, AIC: Akaike bilgi kriteri, AICc: Düzeltilmiş Akaike bilgi kriteri, BIC: Bayesian bilgi kriteri

Astım vakasının bağımlı değişkeni Bölüm 3.1.1'de bahsi geçen dört farklı modelle tahmin edilmiştir. Modellerin fark istatistikleri Tablo 4.4'de sunulmuştur.

**Tablo 4.4.** Astım GAM modellerinin Anova Gam değerlendirmeleri

Değişken		Artık SD	Artık fark	P	
ASTİM	Başvuru	<i>Model I</i>	1796	24482.0	Temel Model
		<i>Model II</i>	1795	24311.0	<0.001
		<i>Model III</i>	1790	20774.0	<0.001
		<i>Model IV</i>	1785	18242.0	<0.001
	Yatan	<i>Model I</i>	1796	2185.6	Temel Model
		<i>Model II</i>	1795	2160.1	<0.001
		<i>Model III</i>	1790	2073.1	<0.001
		<i>Model IV</i>	1785	2024.1	<0.001
	Eks	<i>Model I</i>	1796	261.6	Temel Model
		<i>Model II</i>	1802	214.1	0.092
		<i>Model III</i>	1801	207.1	0.115
		<i>Model IV</i>	1799	213.7	0.095

SD: Serbestlik derecesi.

Tablo 4.3’de bağımlı değişkenin astım hastalığı nedeniyle hastaneye başvuran hasta sayısı olduğu modellerin GAM yöntemi ile çözümleri incelendiğinde; “itrz” düzgünleştirme fonksiyonuna sahip model IV AIC (29574.22), Düzeltilmiş AIC (29574.32), BIC (29623.73), fark yüzdesi (0.8104) ve düzeltilmiş  $R^2$  (0.7911) değerlerinin diğer modellerin ilgili performans değerlerinden daha üstün olduğu gözlenmiştir. Ayrıca Tablo 4.4’de bağımlı değişkenin başvuru olduğu bölüm incelendiğinde; model II’den başlayarak bir önceki modele göre istatistiksel anlamlılığı incelenmiştir. Birbirinden farklı bu dört modelin her yeni modelinin bir önceki modele göre daha iyi sonuçlar verdiğini ve Anova Gam analizi sonucunda bir önceki modele göre istatistiksel olarak anlamlı düzeyde farklı oldukları gözlenmiştir ( $P<0.001$ ).

Tablo 4.3’de bağımlı değişkenin astım hastalığı nedeniyle hastanede yatan hasta sayısı olduğu modeller incelendiğinde; “itrz” düzgünleştirme fonksiyonuna sahip model IV AIC (4280.63), Düzeltilmiş AIC (4280.73), BIC (4330.14), fark yüzdesi (0.5004) ve düzeltilmiş  $R^2$  (0.4647) değerlerinin diğer modellerin ilgili performans değerlerinden daha üstün olduğu gözlenmiştir. Her yeni model bir önceki modele göre daha üstün performans göstermiştir ve son model en iyi performansı gösteren model olmuştur. Ayrıca Tablo 4.4’de bağımlı değişkenin yatan olduğu bölüm incelendiğinde; tüm modellerin Anova Gam analizine göre bir önceki modelle kıyaslandığında istatistiksel olarak anlamlı düzeyde farklı oldukları gözlenmiştir ( $P<0.001$ ).

Tablo 4.3’de bağımlı değişkenin astım hastalığı nedeniyle eks olan hasta sayısı olduğu modeller incelendiğinde; “itrz” düzgünleştirme fonksiyonuna sahip model III AIC (290.49), Düzeltilmiş AIC (290.56), BIC (339.97), fark yüzdesi (0.1125) ve düzeltilmiş  $R^2$  (0.0271) değerlerinin diğer modellerin ilgili performans değerlerinden daha üstün olduğu gözlenmiştir. Model IV, model III kadar iyi bir performans gösterememiştir. Tablo 4.4’de yer alan Anova Gam analizi sonucunda bir önceki modele göre kıyaslandığında modellerin birbirinden istatistiksel olarak benzer oldukları gözlenmiştir ( $P>0.05$ ).

**Tablo 4.5.** Astım hastalığı bağımlı değişkenlerinin GAM yöntemi ile en iyi sonucu veren modellerinin uyum iyiliği test sonuçları

	<b>Değişken</b>	<b>Ki-Kare</b>	<b>P</b>	
<b>ASTİM</b>	<b>Başvuru</b>	<i>Hafta sonu/içi</i>	29740.4	<b>0.001</b>
		<i>Fırtına</i>	5.4	<b>0.021</b>
		<i>Ort. Sıcaklık</i>	480.1	<b>0.001</b>
		<i>Zaman</i>	26096.0	<b>0.001</b>
	<b>Yatan</b>	<i>Hafta sonu/içi</i>	392.3	<b>0.001</b>
		<i>Fırtına</i>	14.2	<b>0.004</b>
		<i>Ort. Sıcaklık</i>	40.9	<b>0.001</b>
		<i>Zaman</i>	821.7	<b>0.001</b>
	<b>Eks</b>	<i>Hafta sonu/içi</i>	3.2	0.072
		<i>Fırtına</i>	6.3	<b>0.011</b>
		<i>Ort. Sıcaklık</i>	2.2	0.139
		<i>Zaman</i>	6.5	0.306

Astım hastalığı bağımlı değişkenleri ile uydurulan ve model tahmin performansları bakımından en iyi sonucu veren modellerin uyum iyiliği test bulguları Tablo 4.5’de sunulmuştur. Buna göre bağımlı değişkenin astım hastalığı şikâyeti ile başvuru yapan hasta sayısı olan model incelendiğinde en iyi tahmin performansı gösteren model IV olmuştur. Bu modelin uyum iyiliği testine göre; hafta sonu/içi değişkeni, ortalama sıcaklık, fırtına varlığı ve zaman değişkenlerinin modelde anlamlı oldukları gözlenmiştir ( $P<0.05$ ). Bağımlı değişkenin astım hastalığı ile yatan hasta sayısı olduğu model incelendiğinde en iyi tahmin performansı model IV göstermiştir. Bu modelin uyum iyiliği testine göre; hafta sonu/içi değişkeni, ortalama sıcaklık, fırtına varlığı ve zaman değişkenlerinin modelde anlamlı oldukları gözlenmiştir ( $P<0.05$ ). Bağımlı değişkenin astım hastalığı ile eks olan hasta sayısı olduğu model incelendiğinde en iyi tahmin performansı model III göstermiştir. Bu modelin uyum iyiliği testine göre; Sadece fırtına varlığı değişkenlerinin modelde anlamlı olduğu gözlenmiştir ( $P<0.05$ ).

Sonuçlar değerlendirildiğinde hava kirliliği ile astım hastalığı nedeniyle başvuran, yatan ve eks olan hasta sayısı arasındaki ilişkiyi GAM yöntemi kullanılarak çözümlendiğinde serbestlik dereceleri ve temel fonksiyon yapıları bakımından değerlendirildiğinde tahmin performansı en iyi modeller aşağıdaki gibi tanımlanmıştır:

$$\log\{\mathbb{E}[Astım başvuru_i]\} = (Haftanın günleri_i) + (Fırtına_i) + (Pm10_i) + (Pm10_{1i}) + (Pm10_{2i}) + (Pm10_{3i}) + (Pm10_{4i}) + (Günlük ort sıcaklık_i) + f(Zaman_i, sd = 12, "itrz")$$

$$\log\{\mathbb{E}[Astım yatan_i]\} = (Haftanın günleri_i) + (Fırtına_i) + (Pm10_i) + (Pm10_{1i}) + (Pm10_{2i}) + (Pm10_{3i}) + (Pm10_{4i}) + (Günlük ort sıcaklık_i) + f(Zaman_i, sd = 12, "krz")$$

$$\log\{\mathbb{E}[Astım eks_i]\} = (Haftanın günleri_i) + (Fırtına_i) + (Pm10_i) + (Pm10_{1i}) + (Pm10_{2i}) + (Pm10_{3i}) + (Pm10_{4i}) + (Günlük ort sıcaklık_i) + f(Zaman_i, sd = 6, "itrz")$$

#### 4.1.2. KOAH Hastalığı ile İlgili Bulgular

Çalışmamızda KOAH hastalığı bağımlı değişkenleri ile ilgili gerçek cevap değişkeni değerlerine en yakın tahmini yapabilen GAM modelini önermek için dört farklı GAM modeli denenmiş ve bu modellerin tahmin performansları kıyaslanmıştır. KOAH hastalığı incelenmesinde bağımlı değişkenler KOAH vakası nedeniyle hastanelerin acil ve ilgili servisine başvuran hasta sayısı, servise yatan hasta ve eks olan hasta sayılarının günlük değerleridir. Bölüm 3.1.1'de bahsi geçen model denemeleri KOAH bağımlı değişkenleri ile denenmiştir. GAM yöntemi ile analizi yapılan modellerin AIC, düzeltilmiş AIC, BIC, düzeltilmiş R<sup>2</sup> ve fark yüzdesi değerleri Tablo 4.6'de sunulmuştur. Ayrıca modellerin Anova Gam testi ile farklılıkları ve model uyum iyiliği testleri yapılmıştır.

**Tablo 4.6.** KOAH ile hava kirliliği ilişkisinin GAM'ın farklı modelleri ile incelenmesi

Değişken		AIC	AICc	BIC	Fark (%)	Düz. R <sup>2</sup>	
KOAH	Başvuru	<i>Model I</i>	23999.35	23999.45	24048.86	0.6610	0.6000
		<i>Model II</i>	23690.82	23690.92	23740.33	0.6690	0.6124
		<i>Model III</i>	22556.72	22556.82	22606.23	0.6985	0.6520
		<i>Model IV</i>	21542.05	21542.15	21591.56	0.7249	0.6854
	Yatan	<i>Model I</i>	6699.61	6499.71	6549.12	0.4314	0.4333
		<i>Model II</i>	6495.28	6495.38	6544.79	0.4327	0.4282
		<i>Model III</i>	6209.43	6209.53	6258.94	0.4913	0.4869
		<i>Model IV</i>	6177.99	6178.09	6227.50	0.4997	0.4919
	Eks	<i>Model I</i>	4460.61	4460.71	4510.12	0.0271	0.0261
		<i>Model II</i>	4334.99	4335.09	4384.50	0.0835	0.0868
		<i>Model III</i>	4298.02	4298.12	4347.53	0.1041	0.1010
		<i>Model IV</i>	4290.68	4290.78	4340.19	0.1113	0.1055

Düz. R<sup>2</sup>: Düzeltilmiş çoklu açıklayıcılık katsayısı, AIC: Akaike bilgi kriteri, AICc: Düzeltilmiş Akaike bilgi kriteri, BIC: Bayesian bilgi kriteri

KOAH vakasının bağımlı değişkenleri bölüm 3.1.1'de bahsi geçen dört farklı modelle tahmin edilmiştir. Modellerin fark istatistikleri Tablo 4.7'de sunulmuştur.

**Tablo 4.7.** KOAH vakasına ait Modellerin Anova Gam değerlendirmeleri

Değişken		Artık SD	Artık fark	P	
KOAH	Başvuru	<i>Model I</i>	1796.0	13143.0	Temel Model
		<i>Model II</i>	1795.0	12871.0	<0.001
		<i>Model III</i>	1790.0	11689.0	<0.001
		<i>Model IV</i>	1785.0	10664.0	<0.001
	Yatan	<i>Model I</i>	1796.0	2852.9	Temel Model
		<i>Model II</i>	1795.0	2846.9	0.014
		<i>Model III</i>	1790.2	2552.3	<0.001
		<i>Model IV</i>	1785.1	2510.5	<0.001
	Eks	<i>Model I</i>	1796.0	2198.8	Temel Model
		<i>Model II</i>	1795.0	2071.2	<0.001
		<i>Model III</i>	1790.0	2024.6	<0.001
		<i>Model IV</i>	1785.2	2008.4	<0.001

SD: Serbestlik derecesi.

Tablo 4.6’da bağımlı değişkenin KOAH hastalığı nedeniyle hastaneye başvuran hasta sayısı olduğu modeller incelendiğinde; “itrz” düzgünleştirme fonksiyonuna sahip model IV AIC (21542.05), düzeltilmiş AIC (21542.15), BIC (21591.56), fark yüzdesi (0.7249) ve düzeltilmiş  $R^2$  (0.6854) değerlerinin diğer dört modelin ilgili değerinden daha üstün olduğu gözlenmiştir. Ayrıca Tablo 4.7’de bağımlı değişkenin başvuru olduğu bölüm incelendiğinde; her yeni model bir önceki modele göre daha iyi sonuçlar vermiş ve modellerin Anova Gam analizine göre bir önceki modele göre istatistiksel olarak anlamlı düzeyde farklı oldukları gözlenmiştir ( $P<0.001$ ).

Tablo 4.6.’de bağımlı değişkenin KOAH hastalığı nedeniyle yatan hasta sayısı olduğu modeller incelendiğinde; “krz” düzgünleştirme fonksiyonuna sahip model IV AIC (6177.99), düzeltilmiş AIC (6178.09), BIC (6227.50), fark yüzdesi (0.4997) ve düzeltilmiş  $R^2$  (0.4919) değerlerinin diğer dört modelin ilgili değerinden daha üstün olduğu gözlenmiştir. Ayrıca Tablo 4.7’de bağımlı değişkenin başvuru olduğu bölüm incelendiğinde; her yeni model bir önceki modele göre daha iyi sonuçlar vermiş ve modellerin Anova Gam analizine göre bir önceki modele göre istatistiksel olarak anlamlı düzeyde farklı oldukları gözlenmiştir ( $P<0.05$ ).

Tablo 4.6.’de bağımlı değişkenin KOAH hastalığı nedeniyle eks olan hasta sayısı olduğu modeller incelendiğinde; “itrz” düzgünleştirme fonksiyonuna sahip model IV AIC (4290.68), düzeltilmiş AIC (4290.78), BIC (4340.19), fark yüzdesi (0.1113) ve düzeltilmiş  $R^2$  (0.1055) değerlerinin diğer dört modelin ilgili değerinden daha üstün olduğu gözlenmiştir. Ayrıca Tablo 4.7’de bağımlı değişkenin başvuru olduğu bölüm incelendiğinde; her yeni model bir önceki modele göre daha iyi sonuçlar vermiş ve modellerin Anova Gam analizine göre bir önceki modele göre istatistiksel olarak anlamlı düzeyde farklı oldukları gözlenmiştir ( $P<0.001$ ).



**Tablo 4.8.** KOAH hastalığı bağımlı değişkenlerinin GAM yöntemi ile en iyi sonucu veren modellerinin uyum iyiliği test sonuçları

	<b>Değişken</b>	<b>Ki-Kare</b>	<b>P</b>	
<b>KOAH</b>	<b>Başvuru</b>	<i>Hafta sonu/içi</i>	16978.1	<b>0.001</b>
		<i>Fırtına</i>	0.335	0.563
		<i>Ort. Sıcaklık</i>	59.8	<b>0.001</b>
		<i>Zaman</i>	2512.0	<b>0.001</b>
	<b>Yatan</b>	<i>Hafta sonu/içi</i>	695.0	<b>0.001</b>
		<i>Fırtına</i>	0.9	0.340
		<i>Ort. Sıcaklık</i>	44.2	<b>0.001</b>
		<i>Zaman</i>	1027.0	<b>0.001</b>
	<b>Eks</b>	<i>Hafta sonu/içi</i>	2.5	0.111
		<i>Fırtına</i>	3.6	<b>0.049</b>
		<i>Ort. Sıcaklık</i>	1.4	0.243
		<i>Zaman</i>	168.6	<b>0.001</b>

KOAH hastalığı ile ilgili bağımlı değişkenler en iyi performansı sağlamış olan IV. modeller uygun görülmüştür. Tablo 4.8 de modellerin uyum iyiliği test sonuçları yer almaktadır. Buna göre bağımlı değişkenin KOAH hastalığı ile başvuran hasta sayısı olduğu model incelendiğinde uyum iyiliği testine göre fırtına (P=0.563) değişkeni istatistiksel olarak anlamlı bulunamamış iken hafta sonu/içi değişkeni, ortalama sıcaklık ve zaman değişkenlerinin modelde anlamlı oldukları gözlenmiştir (P<0.05). Bağımlı değişkenin KOAH hastalığı ile yatan hasta sayısı olduğu model incelendiğinde uyum iyiliği testine göre fırtına (P=0.340) değişkeni istatistiksel olarak anlamlı bulunamamış iken hafta sonu/içi değişkeni, ortalama sıcaklık ve zaman değişkenlerinin modelde anlamlı oldukları gözlenmiştir (P<0.05). Bağımlı değişkenin KOAH hastalığı ile eks olan hasta sayısı olduğu model incelendiğinde uyum iyiliği testine göre hafta sonu/içi (P=0.111), ortalama sıcaklık (P=0.243) değişkenleri istatistiksel olarak anlamlı bulunamamış iken fırtına ve zaman değişkenlerinin modelde anlamlı oldukları gözlenmiştir (P<0.05).

Sonuçlar değerlendirildiğinde hava kirliliği ile KOAH hastalığı nedeniyle başvuran, yatan ve eks olan hasta sayısı arasındaki ilişkiyi GAM yöntemi kullanılarak

çözümlendiğinde serbestlik dereceleri ve temel fonksiyon yapıları bakımından iyi açıklayan modeller aşağıdaki gibi tanımlanmıştır:

$$\log\{\mathbb{E}[Koah\ başvuru_i]\} = (Haftanın\ günleri_i) + (Fırtına_i) + (Pm10_i) + (Pm10_{1i}) + (Pm10_{2i}) + (Pm10_{3i}) + (Pm10_{4i}) + (Günlük\ ort\ sıcaklık_i) + f(Zaman_i, sd = 12, "itrz")$$

$$\log\{\mathbb{E}[Koah\ yatan_i]\} = (Haftanın\ günleri_i) + (Fırtına_i) + (Pm10_i) + (Pm10_{1i}) + (Pm10_{2i}) + (Pm10_{3i}) + (Pm10_{4i}) + (Günlük\ ort\ sıcaklık_i) + f(Zaman_i, sd = 12, "krz")$$

$$\log\{\mathbb{E}[Koah\ eks_i]\} = (Haftanın\ günleri_i) + (Fırtına_i) + (Pm10_i) + (Pm10_{1i}) + (Pm10_{2i}) + (Pm10_{3i}) + (Pm10_{4i}) + (Günlük\ ort\ sıcaklık_i) + f(Zaman_i, sd = 12, "itrz")$$

#### 4.1.3. Pnömoni Hastalığı ile İlgili Bulgular

Çalışmamızda Pnömoni hastalığı bağımlı değişkenleri ile ilgili gerçek cevap değişkenine en yakın tahmini yapabilen GAM modelini önermek için farklı GAM modelleri denenip bu modellerin tahmin performanslarını kıyaslanmıştır. Pnömoni hastalığı ile ilgili bağımlı değişkenlerimiz pnömoni vakası nedeniyle hastanelerin acil ve ilgili servisine başvuran hasta sayısı, servise yatan hasta ve eks olan hasta sayılarının günlük değerleridir. Yapılan model denemelerinde Bölüm 3.1.1’de belirtilen dört modelde bulunan tüm değişkenler ve uygulamaları pnömoni hastalığının bağımlı değişkenleri kullanılarak tekrar modellenmiştir. GAM yöntemi ile analiz edilen bu dört farklı modelin AIC, düzeltilmiş AIC, BIC, düzeltilmiş  $R^2$  ve fark yüzdesi değerleri Tablo 4.9’de sunulmuştur. Ayrıca modellerin Anova Gam testi ile farklılıkları ve model uyum iyiliği testleri yapılmıştır.

**Tablo 4.9.** Pnömoni ile hava kirliliği ilişkisinin GAM'ın farklı modelleri ile incelenmesi

Değişken		AIC	AICc	BIC	Fark (%)	Düz R <sup>2</sup>	
PNÖMONİ	Başvuru	<i>Model I</i>	66642.09	66642.19	66691.60	0.2268	0.2069
		<i>Model II</i>	68052.08	68052.18	68101.59	0.2879	0.2327
		<i>Model III</i>	59048.34	59048.44	59097.85	0.3671	0.3277
		<i>Model IV</i>	51557.48	51557.58	51606.99	0.4664	0.4571
	Yatan	<i>Model I</i>	14062.34	14062.44	14111.85	0.0566	0.0445
		<i>Model II</i>	14062.11	14062.21	14111.62	0.0567	0.0444
		<i>Model III</i>	11310.45	11310.55	11359.96	0.3504	0.2850
		<i>Model IV</i>	10755.71	10755.81	10805.22	0.4104	0.3822
	Eks	<i>Model I</i>	2116.22	2116.32	2165.73	0.0228	0.0065
		<i>Model II</i>	2068.85	2068.95	2118.36	0.0588	0.0354
		<i>Model III</i>	2063.22	2063.32	2113.96	0.0696	0.0399
		<i>Model IV</i>	2064.62	2064.72	2114.21	0.0679	0.0391

Düz. R<sup>2</sup>: Düzeltilmiş çoklu açıklayıcılık katsayısı, AIC: Akaike bilgi kriteri, AICc: Düzeltilmiş Akaike bilgi kriteri, BIC: Bayesian bilgi kriteri.

Pnömoni vakasının bağımlı değişkenleri Bölüm 3.1.1'de bahsi geçen dört farklı modelle tahmin edilmiştir. Modellerin fark istatistikleri Tablo 4.10'da sunulmuştur.

**Tablo 4.10.** Pnömoni vakasına ait modellerin Anova Gam değerlendirmeleri

Değişken		Artık SD	Artık fark	P	
PNÖMONİ	Başvuru	<i>Model I</i>	1796	58439.0	Temel Model
		<i>Model II</i>	1795	56847.0	<0.001
		<i>Model III</i>	1790	47833.0	<0.001
		<i>Model IV</i>	1785	40332.0	<0.001
	Yatan	<i>Model I</i>	1796	8874.0	Temel Model
		<i>Model II</i>	1795	8872.8	0.196
		<i>Model III</i>	1790	6110.1	<0.001
		<i>Model IV</i>	1785	5545.4	<0.001
	Eks	<i>Model I</i>	1796	1339.7	Temel Model
		<i>Model II</i>	1795	1290.5	<0.001
		<i>Model III</i>	1790	1275.6	<0.001
		<i>Model IV</i>	1789	1278.0	<0.001

SD: Serbestlik derecesi.

Tablo 4.9’da bağımlı değişkenin Pnömoni hastalığı nedeniyle hastaneye başvuran hasta sayısı olduğu modeller incelendiğinde; düzgünleştirme fonksiyonu “krz” olan model IV AIC (51557.48), düzeltilmiş AIC (51557.58), BIC (51606.99), fark yüzdesi (0.4664) ve düzeltilmiş  $R^2$  (0.4571) değerlerinin diğer modellerin ilgili performans değerinden daha üstün olduğu gözlenmiştir. Ayrıca Tablo 4.10’da bağımlı değişkenin başvuru olduğu bölüm incelendiğinde; birbirinden farklı bu dört modelin her yeni modelinin bir önceki modele göre daha iyi sonuçlar verdiğini ve bu modellerin Anova Gam analizi sonucunda bir önceki modele göre istatistiksel olarak anlamlı düzeyde farklı oldukları gözlenmiştir ( $P<0.001$ ).

Tablo 4.9’da bağımlı değişkenin pnömoni hastalığı nedeniyle hastanede yatan hasta sayısı olduğu modeller incelendiğinde; düzgünleştirme fonksiyonu “itrz” olan model IV ise AIC (10755.71), Düzeltilmiş AIC (10755.81), BIC (10805.22), fark yüzdesi (0.4104) ve düzeltilmiş  $R^2$  (0.3822) değerlerinin diğer altı modelin ilgili performans değerinden daha üstün olduğu gözlenmiştir. Buna göre Model IV tüm kriter bakımından diğer modellere göre daha yüksek bir performans sergilediği gözlenmiştir. Ayrıca Tablo 4.10’da bağımlı değişkenin yatan olduğu bölüm incelendiğinde; model II haricindeki tüm modellerin Anova Gam analizi sonucunda bir önceki modele göre kıyaslandığında istatistiksel olarak anlamlı düzeyde farklı oldukları gözlenmiştir ( $P<0.001$ ).

Tablo 4.9.’da bağımlı değişkenin Pnömoni hastalığı nedeniyle eks olan hasta sayısı olduğu modeller incelendiğinde; düzgünleştirme fonksiyonu “krz” olan model III AIC (2063.22), Düzeltilmiş AIC (2063.32), BIC (2113.96) fark yüzdesi (0.0696) ve düzeltilmiş  $R^2$  (0.0399) değerlerinin diğer dört modelin ilgili değerinden daha üstün olduğu gözlenmiştir. Buna göre Model III tüm kriter bakımından diğer modellere göre daha yüksek bir performans sergilediği gözlenmiştir. Ayrıca Tablo 4.10’da bağımlı değişkenin eks olduğu bölüm incelendiğinde, modellerin Anova Gam analizi sonucunda bir önceki modele göre kıyaslandığında istatistiksel olarak anlamlı düzeyde farklı oldukları gözlenmiştir ( $P<0.05$ ).

**Tablo 4.11.** Pnömoni hastalığı bağımlı değişkenlerinin GAM yöntemi ile en iyi sonucu veren modellerinin uyum iyiliği test sonuçları

		<b>Değişken</b>	<b>Ki-Kare</b>	<b>P</b>
<b>PNÖMONİ</b>	<b>Başvuru</b>	<i>Hafta sonu/içi</i>	2.6	0.107
		<i>Fırtına</i>	460.3	<b>0.001</b>
		<i>Ort. Sıcaklık</i>	36.7	<b>0.001</b>
		<i>Zaman</i>	29310.0	<b>0.001</b>
	<b>Yatan</b>	<i>Hafta sonu/içi</i>	0.9	0.338
		<i>Fırtına</i>	0.6	0.453
		<i>Ort. Sıcaklık</i>	4.5	<b>0.033</b>
		<i>Zaman</i>	1574.0	<b>0.001</b>
	<b>Eks</b>	<i>Hafta sonu/içi</i>	0.7	0.415
		<i>Fırtına</i>	0.7	0.415
		<i>Ort. Sıcaklık</i>	3.09	0.078
		<i>Zaman</i>	75.6	<b>0.001</b>

Pnömoni hastalığı ile hava kirliliği arasındaki ilişkiyi en iyi açıklayan modeller: eks bağımlı değişkeni için model III, başvuru ve yatan değişkenleri için model IV olduğu görülmüştür. Tablo 4.11’de bu modellere ait uyum iyiliği test sonuçları yer almaktadır. Buna göre bağımlı değişkenin pnömoni hastalığı ile başvuran hasta sayısı olduğu model incelendiğinde uyum iyiliği testine göre hafta sonu/içi (P=0.107) değişkeni istatistiksel olarak anlamlı bulunamamış iken, fırtına değişkeni (P=0.001) ve ortalama sıcaklık değişkenlerinin modelde anlamlı olduğu gözlenmiştir (P=0.001). Bağımlı değişkenin pnömoni hastalığı ile yatış alan hasta sayısı olduğu model incelendiğinde uyum iyiliği testine göre hafta içi/sonu değişkeni (P=0.338) ve fırtına değişkeni (P=0.453) istatistiksel olarak anlamlı bulunamazken, ortalama sıcaklık değişkenini (P=0.033) anlamlı olduğu gözlenmiştir. Pnömoni hastalığı ile eks olan hasta sayısı bağımlı değişkeni modellendiğinde uyum iyiliği testine göre hafta sonu/içi, fırtına (P=0.415) değişkenleri ve ortalama sıcaklık değişkeninin (P=0.078) istatistiksel olarak anlamlılıklarına rastlanmadı. Zaman değişkeninin tüm modellerde anlamlı olduğu gözlenmiştir (P<0.001).

Bu sonuçlar değerlendirildiğinde hava kirliliği ile pnömoni hastalığı nedeniyle başvuran, yatan ve eks olan hasta sayısı arasındaki ilişkiyi en iyi açıklayan modeller aşağıdaki gibi tanımlanmıştır:

$$\log\{\mathbb{E}[Pn\ddot{o}moni\ bařvuru_i]\} = (Haftanın\ g\ddot{u}nleri_i) + (Firtına_i) + (Pm10_i) + (Pm10_{1i}) + (Pm10_{2i}) + (Pm10_{3i}) + (Pm10_{4i}) + (G\ddot{u}nl\ddot{u}k\ ort\ sıcaklık_i) + f(Zaman_i, sd = 12, "krz")$$

$$\log\{\mathbb{E}[Pn\ddot{o}moni\ yatan_i]\} = (Haftanın\ g\ddot{u}nleri_i) + (Firtına_i) + (Pm10_i) + (Pm10_{1i}) + (Pm10_{2i}) + (Pm10_{3i}) + (Pm10_{4i}) + (G\ddot{u}nl\ddot{u}k\ ort\ sıcaklık_i) + f(Zaman_i, sd = 12, "itrz")$$

$$\log\{\mathbb{E}[Pn\ddot{o}moni\ eks_i]\} = (Haftanın\ g\ddot{u}nleri_i) + (Firtına_i) + (Pm10_i) + (Pm10_{1i}) + (Pm10_{2i}) + (Pm10_{3i}) + (Pm10_{4i}) + (G\ddot{u}nl\ddot{u}k\ ort\ sıcaklık_i) + f(Zaman_i, sd = 6, "krz")$$

#### 4.1.4. Pulmoner Tromboemboli (PTE) Hastalığı İle İlgili Bulgular

Çalışmamızda pulmoner tromboemboli hastalığı bağımlı değişkenleri ile ilgili gerçek cevap değişkenine en yakın tahmini yapabilen GAM modelini önermek için farklı GAM modelleri denenmiş ve bu modellerin tahmin performansları kıyaslanmıştır. Pulmoner tromboemboli hastalığı ile ilgili bağımlı değişkenlerimiz pulmoner tromboemboli vakası nedeniyle hastanelerin ilgili servisinde yatan hasta ve eks olan hasta sayılarının günlük değerleridir. Bölüm 3.1.1’de yer alan dört model pulmoner tromboemboli vakasının iki bağımlı değişkeni için denenmiştir. GAM yöntemi ile analiz edilen bu dört farklı modelin AIC, düzeltilmiş AIC, BIC, düzeltilmiş R<sup>2</sup> ve fark yüzdesi değerleri Tablo 4.12’de sunulmuştur. Ayrıca modellerin Anova Gam testi ile farklılıkları ve model uyum iyiliği testleri yapılmıştır. Bu analiz sonuçları sırası ile Tablo 4.13 ve Tablo 4.14’de sunulmuştur.

**Tablo 4.12.** PTE ile hava kirliliği ilişkisinin GAM’ın farklı modelleri ile incelenmesi

	Değişken	AIC	AICc	BIC	Fark (%)	Düz. R <sup>2</sup>	
PTE	Yatan	Model I	5247.13	5247.23	5296.64	0.3677	0.3011
		Model II	5233.98	5234.08	5283.49	0.3716	0.2893
		Model III	4989.87	4989.97	5039.38	0.4372	0.3565
		Model IV	4967.31	4967.41	5016.82	0.4453	0.3627
	Eks	Model I	1215.64	1215.74	1265.15	0.0157	0.0038
		Model II	1206.75	1206.85	1256.36	0.0281	0.0102
		Model III	1207.31	1207.41	1256.83	0.0279	0.0098
		Model IV	1205.64	1205.74	1255.15	0.0491	0.0146

Düz. R<sup>2</sup>: Düzeltilmiş çoklu açıklayıcılık katsayısı, AIC: Akaike bilgi kriteri, AICc: Düzeltilmiş Akaike bilgi kriteri, BIC: Bayesian bilgi kriteri.

Pulmoner Tromboemboli vakasının bağımlı değişkenleri Bölüm 3.1.1’de bahsi geçen dört farklı modelle tahmin edilmiştir. Modellerin fark istatistikleri Tablo 4.13’de sunulmuştur.

**Tablo 4.13.** PTE vakasına ait Modellerin Anova Gam değerlendirmeleri

	Değişken	Artık SD	Artık fark	P	
PTE	Yatan	<i>Model I</i>	1796.0	841.4	Temel Model
		<i>Model II</i>	1795.1	831.2	<0.001
		<i>Model III</i>	1794.6	831.5	<0.001
		<i>Model IV</i>	1786.2	812.7	0.020
	Eks	<i>Model I</i>	1796.0	2448.0	Temel Model
		<i>Model II</i>	1795.0	2433.0	<0.001
		<i>Model III</i>	1790.0	2178.9	<0.001
		<i>Model IV</i>	17851	2147.5	<0.001

SD: Serbestlik Derecesi.

Tablo 4.12’de bağımlı değişkenin pulmoner tromboemboli hastalığı nedeniyle hastanede yatan hasta sayısı olduğu modeller incelendiğinde; düzgünleştirme fonksiyonu “krz” olan model IV ise AIC (4967.31), düzeltilmiş AIC (4967.41), BIC (5016.82), fark yüzdesi (0.4453) ve düzeltilmiş  $R^2$  (0.3627) değerleri elde edilmiştir. Bu değerlere göre model IV AIC, Düzeltilmiş AIC, fark yüzdesi ve Düzeltilmiş  $R^2$  değerleri bakımından diğer modellerden üstün olduğu görülmektedir. Ayrıca Tablo 4.13’da bağımlı değişkenin yatan olduğu bölüm incelendiğinde; tüm modellerin Anova Gam analizine göre birbirlerinden istatistiksel olarak anlamlı düzeyde farklı oldukları gözlenmiştir ( $P<0.001$ ).

Tablo 4.12’de bağımlı değişkenin pulmoner tromboemboli hastalığı nedeniyle eks olan hasta sayısı olduğu modeller incelendiğinde; düzgünleştirme fonksiyonu “dkrz” olan model IV AIC (1205.64), Düzeltilmiş AIC (1205.74), BIC (1255.15) fark yüzdesi (0.0491) ve düzeltilmiş  $R^2$  (0.0146) değerleri gözlenmiştir. Buna göre PTE eks bağımlı değişkeni için kurulan modellerde yine genel anlamda performansı yüksek olan model IV diğer modellerin performans kriteri değerlerine göre üstünlük sağlayarak öne çıkmayı başarmıştır. Ayrıca Tablo 4.13’da bağımlı değişkenin eks olduğu bölüm incelendiğinde Anova Gam analizine göre; model IV, model III’den, model III, model II’den ve son olarak model II’de model I’den istatistiksel olarak anlamlı düzeyde farklı oldukları gözlenmiştir ( $P<0.001$ ).

**Tablo 4.14.** PTE hastalığı bağımlı değişkenlerinin GAM yöntemi ile en iyi sonucu veren modellerinin uyum iyiliği test sonuçları

	<b>Değişken</b>	<b>Ki-Kare</b>	<b>P</b>	
<b>PTE</b>	<b>Yatan</b>	<i>Hafta sonu/içi</i>	512.7	<b>0.001</b>
		<i>Fırtına</i>	0.1	0.695
		<i>Ort. Sıcaklık</i>	0.1	0.813
		<i>Zaman</i>	292.3	<b>0.001</b>
	<b>Eks</b>	<i>Hafta sonu/içi</i>	0.5	0.470
		<i>Fırtına</i>	2.4	0.125
		<i>Ort. Sıcaklık</i>	0.2	0.264
		<i>Zaman</i>	23.8	<b>0.008</b>

Pulmoner tromboemboli hastalığı ile yatan ve eks olan hasta sayısı bağımlı değişkenleri model IV uygun bulunmuştur. Tablo 4.14’de değişkenlere ilişkin uyum iyiliği test sonuçları yer almaktadır. Buna göre bağımlı değişkenin pulmoner tromboemboli hastalığı ile yatan hasta sayısı olduğu model incelendiğinde uyum iyiliği testine göre hafta sonu/içi ve zaman değişkenleri istatistiksel olarak anlamlı bulunmuştur (P=0.001). Diğer değişkenler fırtına (P=0.695) ve ortalama sıcaklık değişkeni (P=0.813) modelde istatistiksel olarak anlamlı olmadıkları gözlenmiştir. Bağımlı değişkenin pulmoner tromboemboli hastalığı ile eks alan hasta sayısı olduğu model incelendiğinde uyum iyiliği testine göre hafta içi/sonu değişkeni (P=0.470), ortalama sıcaklık değişkeni (P=0.264) ve fırtına değişkeninin (P=0.125) istatistiksel olarak anlamlı olmadıkları gözlenmiştir. Zaman değişkeninin istatistiksel olarak anlamlı olduğu gözlendi (P=0.008).

Bu sonuçlar değerlendirildiğinde hava kirliliği ile Pulmoner Tromboemboli hastalığı nedeniyle yatan ve eks olan hasta sayısı arasındaki ilişkiyi en iyi açıklayan modeller aşağıdaki gibi tanımlanmıştır

$$\log\{\mathbb{E}[Pn\ddot{o}moni\ yatan_i]\} = (Haftanın\ g\ddot{u}nleri_i) + (Fırtına_i) + (Pm10_i) + (Pm10_{1i}) + (Pm10_{2i}) + (Pm10_{3i}) + (Pm10_{4i}) + (G\ddot{u}nl\ddot{u}k\ ort\ sıcaklık_i) + f(Zaman_i, sd = 12, "krz")$$

$$\log\{\mathbb{E}[PTeks_i]\} = (Haftanın\ g\ddot{u}nleri_i) + (Fırtına_i) + (Pm10_i) + (Pm10_{1i}) + (Pm10_{2i}) + (Pm10_{3i}) + (Pm10_{4i}) + (G\ddot{u}nl\ddot{u}k\ ort\ sıcaklık_i) + f(Zaman_i, sd = 12, "dkrz")$$



**Tablo 4.15.** PM10 değişkeninin 1 mg/m<sup>3</sup> artışının GAM ile OR tahmini

	Başvuru			Yatan		Eks	
	Değ.	OR [%95 GA]	P	OR [%95 GA]	P	OR [%95 GA]	P
Astım	<i>Lag0</i>	1.0007 [1.0006-1.0008]	<b>0.017</b>	1.0011 [1.0001-1.0021]	<b>0.035</b>	1.0014 [0.9957-1.0073]	0.651
	<i>Lag1</i>	0.9998 [0.9997-0.9999]	<b>0.001</b>	0.9990 [0.9977-1.0003]	0.125	1.0027 [0.9956-1.0102]	0.483
	<i>Lag2</i>	1.0002 [1.0001-1.0003]	<b>0.014</b>	1.0004 [0.9992-1.0016]	0.514	0.9983 [0.9892-1.0074]	0.681
	<i>Lag3</i>	1.0002 [1.0001-1.0003]	<b>0.025</b>	1.0006 [0.9994-1.0017]	0.322	1.0030 [0.9951-1.0108]	0.500
	<i>Lag4</i>	0.9999 [0.9998-0.9999]	<b>0.002</b>	1.0008 [0.9998-1.0019]	0.110	1.0014 [0.9947-1.0082]	0.712
KOAH	<i>Lag0</i>	1.0003 [1.0002-1.0005]	<b>0.001</b>	1.0007 [0.9999-1.0015]	0.061	0.9994 [0.9982-1.0006]	0.313
	<i>Lag1</i>	0.9999 [0.9997-1.0000]	0.114	0.9993 [0.9983-1.0003]	0.175	1.0005 [0.9991-1.0018]	0.483
	<i>Lag2</i>	0.9998 [0.9997-0.9999]	<b>0.007</b>	1.0002 [0.9992-1.0011]	0.754	1.0003 [0.9989-1.0016]	0.653
	<i>Lag3</i>	1.0001 [0.9999-1.0002]	0.310	1.0003 [0.9994-1.0012]	0.505	1.0012 [1.0000-1.0025]	<b>0.049</b>
	<i>Lag4</i>	0.9998 [0.9999-0.9999]	<b>0.006</b>	1.0002 [0.9994-1.0010]	0.566	0.9992 [0.9980-1.0004]	0.210
Pnömoni	<i>Lag0</i>	1.0002 [1.0001-1.0003]	<b>0.003</b>	1.0001 [0.9996-1.0006]	0.684	1.0018 [0.9996-1.0040]	0.101
	<i>Lag1</i>	0.9997 [0.9995-0.9998]	<b>0.001</b>	0.9996 [0.9990-1.0001]	0.177	0.9999 [0.9971-1.0027]	0.948
	<i>Lag2</i>	0.9999 [0.9998-1.0001]	0.785	0.9998 [0.9993-1.0004]	0.560	1.0008 [0.9979-1.0035]	0.587
	<i>Lag3</i>	1.0001 [1.0000-1.0002]	<b>0.049</b>	0.9991 [0.9986-0.9998]	<b>0.007</b>	1.0006 [0.9978-1.0034]	0.657
	<i>Lag4</i>	1.0000 [0.9999-1.0001]	0.754	1.0004 [0.9999-1.0009]	0.107	1.0005 [0.9981-1.0029]	0.675
PTE	<i>Lag0</i>			1.0005 [0.9998-1.0018]	0.266	1.0031 [0.9999-1.0063]	0.058
	<i>Lag1</i>			0.9981 [0.9967-0.9996]	<b>0.002</b>	1.0027 [0.9992-1.0064]	0.134
	<i>Lag2</i>			1.0000 [0.9989-1.0012]	0.953	0.9957 [0.9907-1.0007]	0.095
	<i>Lag3</i>			1.0008 [0.9998-1.0018]	0.129	1.0010 [0.9964-1.0056]	0.666
	<i>Lag4</i>			0.9996 [0.9986-1.0006]	0.400	0.9997 [0.9959-1.0035]	0.887

PTE: Pulmoner Tromboemboli, GA: Güven aralığı, OR: Odds Oranı,

GAM yöntemi ile modellenmiş dört hastalıkta en iyi tahmin başarısına sahip modellerde PM10 ana etkisi ve gecikmelerinin 1 mg/m<sup>3</sup> artışı üzerinden tahmin edilen OR ve güven aralıkları Tablo 4.15’de incelenmiştir.

## **4.2 Dağıtılmış Gecikmeli Doğrusal Olmayan Modellerle (DGDOM) Yapılan Analizlerin Sonuçları**

Çalışmamızın bu bölümünde Dağıtılmış gecikmeli doğrusal olmayan modellerin hava kirliliği ile astım, KOAH, pnömoni ve pulmoner tromboemboli vakalarının ilişkisini tahmin etme başarısı incelenmiştir. Bölüm 3.1.1’deki dört model DGDOM yöntemi kullanılarak tekrar analiz edilmiştir. Yapılan analizlerde GAM modellerindeki en iyi performansı gösteren düzgünleştirme fonksiyonlarının, DGDOM modelleri için de aynı sonuçları verdikleri gözlenmiştir. Analizler: astım, KOAH, pnömoni ve pulmoner tromboemboli vakaları nedeniyle hastanelerin acil ve ilgili servisine başvuran hasta sayısı, servise yatan hasta ve eks olan hasta sayılarının günlük değerlerini sırayla bağımlı değişken kabul edilerek yapılmıştır. Modellerin AIC, düzeltilmiş AIC, BIC, Fark yüzdesi ve R<sup>2</sup> gibi performans kriterleri incelenmiştir.

### **4.2.1. Astım Hastalığı ile İlgili Bulgular**

Astım hastalığı bağımlı değişkenleri ile ilgili gerçek cevap değişkenine en yakın tahmini yapabilen DGDOM modelini önermek için dört farklı DGDOM modeli denenmiş ve bu modellerin tahmin performansları kıyaslanmıştır. Astım hastalığı ile ilgili modellerimizde bağımlı değişkenlerimiz astım vakası nedeniyle hastanelerin acil ve ilgili servisine başvuran hasta sayısı, servise yatan hasta ve eks olan sayılarının günlük değerleridir. Bölüm 3.1.1’de tanımlanan ve DGDOM yöntemi ile modellenen dört farklı modelin AIC, düzeltilmiş AIC, BIC, düzeltilmiş R<sup>2</sup> ve fark yüzdesi değerleri Tablo 4.16’de sunulmuştur. Hangi modelin hava kirliliği ile astım vakası nedeniyle başvuran, yatan ve eks olan ilişkisini tahmin etmede daha başarılı olacağı araştırılmıştır.

**Tablo 4.16.** Astım ile hava kirliliği ilişkisinin DGDOM'nin farklı modelleri ile incelenmesi

	<b>Değişken</b>	<b>AIC</b>	<b>AICc</b>	<b>BIC</b>	<b>Fark (%)</b>	<b>Düz. R<sup>2</sup></b>	
<b>ASTİM</b>	<b>Başvuru</b>	<i>Model I</i>	35792.53	35792.56	35820.04	0.7456	0.7135
		<i>Model II</i>	35623.01	35623.04	35650.52	0.7474	0.7176
		<i>Model III</i>	32095.70	32095.73	32123.21	0.7841	0.7602
		<i>Model IV</i>	29574.22	29574.25	29601.73	0.8104	0.7912
	<b>Yatan</b>	<i>Model I</i>	4422.92	4422.95	4450.43	0.4605	0.4378
		<i>Model II</i>	4399.32	4399.35	4426.83	0.4669	0.4332
		<i>Model III</i>	4334.28	4334.31	4361.79	0.4853	0.4514
		<i>Model IV</i>	4285.11	4285.14	4312.62	0.4992	0.4633
	<b>Eks</b>	<i>Model I</i>	292.62	292.65	320.13	0.0716	0.0179
		<i>Model II</i>	291.40	291.43	318.91	0.0825	0.0204
		<i>Model III</i>	290.48	290.51	317.99	0.1129	0.0275
		<i>Model IV</i>	291.54	291.57	319.05	0.0841	0.0212

Düz. R<sup>2</sup>: Düzeltilmiş çoklu açıklayıcılık katsayısı, AIC: Akaike bilgi kriteri, AICc: Düzeltilmiş Akaike bilgi kriteri, BIC: Bayesian bilgi kriteri.

Astım vakasının bağımlı değişkenleri Bölüm 3.1.1'de bahsi geçen dört farklı modelle tahmin edilmiştir. Modellerin fark istatistikleri Tablo 4.17'de sunulmuştur.

**Tablo 4.17.** Astım DGDOM modellerinin Anova Gam değerlendirmeleri

	<b>Değişken</b>	<b>Artık SD</b>	<b>Artık fark</b>	<b>P</b>	
<b>ASTİM</b>	<b>Başvuru</b>	<i>Model I</i>	1796	24482	Temel Model
		<i>Model II</i>	1795	24311	<0.001
		<i>Model III</i>	1790	20774	<0.001
		<i>Model IV</i>	1785	18242	<0.001
	<b>Yatan</b>	<i>Model I</i>	1796	216	Temel Model
		<i>Model II</i>	1795	214	0.094
		<i>Model III</i>	1791	207	0.113
		<i>Model IV</i>	1794	213	0.095
	<b>Eks</b>	<i>Model I</i>	1796	216	Temel Model
		<i>Model II</i>	1795	214	0.094
		<i>Model III</i>	1791	207	0.113
		<i>Model IV</i>	1794	213	0.095

SD: Serbestlik derecesi

Tablo 4.16’de bağımlı değişkenin Astım hastalığı nedeniyle hastaneye başvuran hasta sayısı olduğu modellerin DGDÖM yöntemi ile çözümleri incelendiğinde; Model IV AIC (29574.22), Düzeltilmiş AIC (29574.25), BIC (29601.73), fark yüzdesi (0.8104) ve düzeltilmiş  $R^2$  (0.7912) değerlerinin diğer modellerin tahmin performanslarından daha üstün olduğu gözlenmiştir. Ayrıca Tablo 4.17’de bağımlı değişkenin başvuru olduğu bölüm incelendiğinde; Model II’den başlayarak bir önceki modele göre istatistiksel anlamlılığına bakılmıştır birbirinden farklı bu dört modelin her yeni modelinin bir önceki modele göre daha iyi sonuçlar verdiğini ve Anova Gam analizi sonucunda bir önceki modele göre istatistiksel olarak anlamlı düzeyde farklı oldukları gözlenmiştir ( $P < 0.001$ ).

Tablo 4.16’da bağımlı değişkenin Astım hastalığı nedeniyle hastanede yatan hasta sayısı olduğu modeller incelendiğinde; Model IV AIC (4285.11), Düzeltilmiş AIC (4285.14), BIC (4312.62), fark yüzdesi (0.4992) ve düzeltilmiş  $R^2$  (0.4633) değerlerinin diğer üç modelin ilgili değerinden daha üstün olduğu gözlenmiştir. Buna göre model IV’ün tüm kriter bakımından diğer modellere göre daha yüksek bir performans sergilediği gözlenmiştir. Ayrıca Tablo 4.17’de bağımlı değişkenin yatan olduğu bölüm incelendiğinde; tüm modellerin Anova Gam analizine göre bir önceki modelle kıyaslandığında istatistiksel olarak anlamlı düzeyde farklılığı gözlenmemiştir ( $P > 0.05$ ).

Tablo 4.16’da bağımlı değişkenin astım hastalığı nedeniyle eks olan sayısı olduğu modeller incelendiğinde; model III AIC (290.48), Düzeltilmiş AIC (290.51), BIC (317.99), fark yüzdesi (0.1129) ve düzeltilmiş  $R^2$  (0.0275) değerlerinin diğer üç modelin ilgili değerinden daha üstün olduğu gözlenmiştir. Buna göre model III, tüm kriter bakımından diğer modellere göre daha yüksek bir performans sergilediği gözlenmiştir. Ayrıca Tablo 4.17’de bağımlı değişkenin eks olduğu bölüm incelendiğinde; tüm modellerin Anova Gam analizine göre bir önceki modelle kıyaslandığında istatistiksel olarak anlamlı düzeyde farklılığı gözlenmemiştir ( $P > 0.05$ ).

**Tablo 4.18.** Astım hastalığı bağımlı değişkenlerinin DGDOM yöntemi ile en iyi sonucu veren modellerinin uyum iyiliği test sonuçları

	<b>Değişken</b>	<b>Kovariyet</b>	<b>Ki-kare</b>	<b>P</b>
<b>ASTİM</b>	<b>Başvuru</b>	<i>Hafta sonu/içi</i>	2974.4	< <b>0.001</b>
		<i>Fırtına</i>	5.4	<b>0.021</b>
		<i>Ort. Sıcaklık</i>	480.1	< <b>0.001</b>
		<i>Zaman (d)</i>	277.8	< <b>0.001</b>
	<b>Yatan</b>	<i>Hafta sonu/içi</i>	392.9	< <b>0.001</b>
		<i>Fırtına</i>	15.0	< <b>0.001</b>
		<i>Ort. Sıcaklık</i>	16.4	< <b>0.001</b>
		<i>Zaman (d)</i>	12.7	<b>0.026</b>
	<b>Eks</b>	<i>Hafta sonu/içi</i>	3.3	0.071
		<i>Fırtına</i>	5.9	<b>0.014</b>
		<i>Ort. Sıcaklık</i>	1.5	0.206
		<i>Zaman (d)</i>	1.9	0.424

d: Düzenleştirme fonksiyonu

Astım hastalığı bağımlı değişkenleri tahmin edilen model IV ve III'e ilişkin uyum iyiliği test bulguları Tablo 4.18'de sunulmuştur. Buna göre bağımlı değişkenin astım hastalığı ile başvuran hasta sayısı olduğu model incelendiğinde uyum iyiliği testine göre fırtına değişkeni (P=0.021), ortalama sıcaklık ve zaman değişkenlerinin modelde istatistiksel olarak anlamlı oldukları gözlenmiştir (P<0.001). Bağımlı değişkenin astım hastalığı ile yatan hasta sayısı olduğunda incelenen model uyum iyiliği testine göre hafta içi/sonu, fırtına, ortalama sıcaklık (P<0.001) ve zaman (P=0.026) değişkenlerinin modelde istatistiksel olarak anlamlı bulunmuştur. Astım hastalığı ile eks olan hasta sayısı bağımlı değişken olduğunda incelenen model uyum iyiliği testine göre fırtına (P=0.014) değişkeni istatistiksel olarak anlamlı bulunurken diğer değişkenlerin istatistiksel anlamlılığına rastlanmamıştır (P>0.05).

#### **4.2.2. KOAH Hastalığı ile İlgili Bulgular**

KOAH hastalığı bağımlı değişkenleri ile ilgili gerçek cevap değişkenine en yakın tahmini yapabilen DGDOM modelini önermek için dört farklı DGDOM modeli denenmiş ve bu modellerin tahmin performansları kıyaslanmıştır. KOAH hastalığı ile ilgili modellerde bağımlı değişkenler; KOAH vakası nedeniyle hastanelerin acil ve ilgili servisine başvuran hasta sayısı, servise yatan hasta ve eks olan sayılarının günlük

değerleridir. Bölüm 3.1.1’de tanımlanan ve DGDOM yöntemi ile modellenen dört farklı modelin AIC, düzeltilmiş AIC, BIC, düzeltilmiş  $R^2$  ve fark yüzdesi değerleri Tablo 4.19’de sunulmuştur. Hangi modelin hava kirliliği ile KOAH vakası nedeniyle başvuran, yatan ve eks olan ilişkisini tahmin etmede daha başarılı olacağı araştırılmıştır.

**Tablo 4.19.** KOAH ile hava kirliliği ilişkisinin DGDOM’nin farklı modelleri ile incelenmesi

Değişken		AIC	AICc	BIC	Fark (%)	Düz. $R^2$	
KOAH	Başvuru	<i>Model I</i>	23737.69	23737.79	23787.20	0.6621	0.6030
		<i>Model II</i>	23542.09	23542.19	23591.60	0.6673	0.6099
		<i>Model III</i>	22444.06	22444.16	22493.57	0.6963	0.6490
		<i>Model IV</i>	21432.48	21432.58	21481.99	0.7230	0.6826
	Yatan	<i>Model I</i>	6468.90	6469.00	6518.41	0.4344	0.4348
		<i>Model II</i>	6461.58	6461.68	6511.09	0.4362	0.4282
		<i>Model III</i>	6197.05	6197.15	6246.56	0.4906	0.4853
		<i>Model IV</i>	6170.66	6170.76	6220.17	0.4981	0.4909
	Eks	<i>Model I</i>	4462.28	4462.38	4511.79	0.0220	0.0170
		<i>Model II</i>	4337.18	4337.28	4386.69	0.0780	0.0780
		<i>Model III</i>	4294.34	4294.44	4343.85	0.1013	0.0966
		<i>Model IV</i>	4285.33	4285.43	4334.84	0.1096	0.1018

Düz.  $R^2$ : Düzeltilmiş çoklu açıklayıcılık katsayısı, AIC: Akaike bilgi kriteri, AICc: Düzeltilmiş Akaike bilgi kriteri, BIC: Bayesian bilgi kriteri.

KOAH vakasının bağımlı değişkenleri Bölüm 3.1.1’de bahsi geçen dört farklı modelle tahmin edilmiştir. Modellerin fark istatistikleri Tablo 4.20’de sunulmuştur.

**Tablo 4.20.** KOAH vakası DGDOM modellerinin Anova Gam değerlendirmeleri

Değişken		Artık SD	Artık fark	P	
KOAH	Başvuru	Model I	1790	12921	Temel Model
		Model II	1789	12724	<0.001
		Model III	1784	11616	<0.001
		Model IV	1779	10594	<0.001
	Yatan	Model I	1790	2826	Temel Model
		Model II	1789	2817	0.002
		Model III	1785	2545	<0.001
		Model IV	1779	2508	<0.001
	Eks	Model I	1790	2204	Temel Model
		Model II	1789	2077	<0.001
		Model III	1784	2024	<0.001
		Model IV	1779	2006	0.002

SD: Serbestlik derecesi

Tablo 4.19’de bağımlı değişkenin KOAH hastalığı nedeniyle hastaneye başvuran hasta sayısı olduğu modellerin DGDOM yöntemi ile çözümleri incelendiğinde; model IV AIC (21432.48), Düzeltilmiş AIC (21432.58), BIC (21481.99), fark yüzdesi (0.7230) ve düzeltilmiş  $R^2$  (0.6826) değerlerinin diğer modellerin tahmin performans sonuçlarından daha üstün olduğu gözlenmiştir. Ayrıca Tablo 4.20’de bağımlı değişkenin başvuru olduğu bölüm incelendiğinde; Model II’den başlayarak bir önceki modele göre ardışık olarak istatistiksel anlamlılığı incelenmiştir. Birbirinden farklı bu dört modelin her yeni modelinin bir önceki modele göre daha iyi sonuçlar verdiğini ve Anova Gam analizine göre bir önceki modele göre istatistiksel olarak anlamlı düzeyde farklı oldukları gözlenmiştir ( $P<0.001$ ).

Tablo 4.19’de bağımlı değişkenin KOAH yatan hasta sayısı olduğu modellerin DGDOM yöntemi ile çözümleri incelendiğinde; model IV AIC (6170.66), Düzeltilmiş AIC (6170.76), BIC (6220.17), fark yüzdesi (0.4981) ve düzeltilmiş  $R^2$  (0.4909) değerlerinin diğer modellerin tahmin performans sonuçlarından daha üstün olduğu gözlenmiştir. Ayrıca Tablo 4.20’de bağımlı değişkenin başvuru olduğu bölüm incelendiğinde; model II’den başlayarak bir önceki modele göre ardışık olarak istatistiksel anlamlılığına bakılmıştır. Birbirinden farklı bu dört modelin her yeni modelinin bir önceki modele göre daha iyi sonuçlar verdiğini ve Anova Gam analizi

sonucunda bir önceki modele göre istatistiksel olarak anlamlı düzeyde farklı oldukları gözlenmiştir ( $P<0.05$ ).

Tablo 4.19’da bağımlı değişkenin KOAH eks olan hasta sayısı olduğu modellerin DGDOM yöntemi ile çözümleri incelendiğinde; model IV AIC (4285.33), Düzeltilmiş AIC (4285.43), BIC (4334.84), fark yüzdesi (0.1096) ve düzeltilmiş  $R^2$  (0.1018) değerlerinin diğer modellerin tahmin performans sonuçlarından daha üstün olduğu gözlenmiştir. Ayrıca Tablo 4.20’de bağımlı değişkenin başvuru olduğu bölüm incelendiğinde; model II’den başlayarak bir önceki modele göre ardışık olarak istatistiksel anlamlılığına bakılmış, birbirinden farklı bu dört modelin her yeni modelinin bir önceki modele göre daha iyi sonuçlar verdiğini ve Anova Gam analizi sonucunda bir önceki modele göre istatistiksel olarak anlamlı düzeyde farklı oldukları gözlenmiştir ( $P<0.05$ ).

**Tablo 4.21.** KOAH hastalığı bağımlı değişkenlerinin DGDOM yöntemi ile en iyi sonucu veren modellerinin uyum iyiliği test sonuçları

	<b>Değişken</b>	<b>Kovariyet</b>	<b>Ki-kare</b>	<b>P</b>
<b>KOAH</b>	<b>Başvuru</b>	<i>Hafta sonu/içi</i>	16879.2	<b>&lt;0.001</b>
		<i>Fırtına</i>	2.9	0.087
		<i>Ort. Sıcaklık</i>	58.9	<b>&lt;0.001</b>
		<i>Zaman (d)</i>	2387.0	<b>&lt;0.001</b>
	<b>Yatan</b>	<i>Hafta sonu/içi</i>	7005	<b>&lt;0.001</b>
		<i>Fırtına</i>	1.8	0.174
		<i>Ort. Sıcaklık</i>	23.9	<b>&lt;0.001</b>
		<i>Zaman (d)</i>	969.0	<b>&lt;0.001</b>
	<b>Eks</b>	<i>Hafta sonu/içi</i>	2.4	0.115
		<i>Fırtına</i>	3.8	<b>0.049</b>
		<i>Ort. Sıcaklık</i>	0.3	0.562
		<i>Zaman (d)</i>	0.6	0.962

d: Düzgünleştirme fonksiyonu

KOAH hastalığı bağımlı değişkenleri ile tahmin edilen model IV’e ilişkin uyum iyiliği test bulguları Tablo 4.21’de sunulmuştur. Buna göre bağımlı değişkenin KOAH hastalığı ile başvuran hasta sayısı olduğu model incelendiğinde uyum iyiliği testine göre



fırtına deęişkeni istatistiksel olarak anlamlı bulunamamıştır ( $P=0.087$ ). Hafta sonu/içi, ortalama sıcaklık, PM10 ve zaman deęişkenlerinin modelde anlamlı olduęu gözlenmiştir ( $P<0.05$ ). Benzer şekilde bağımlı deęişkenin KOAH hastalığı ile yatan hasta sayısı olduęu model incelendiğinde uyum iyilięi testine göre fırtına deęişkeni istatistiksel olarak anlamlı bulunamamıştır ( $P=0.174$ ). Hafta sonu/içi, ortalama sıcaklık, PM10 ve zaman deęişkenleri modelde anlamlı olduęu gözlenmiştir ( $P<0.001$ ). Bağımlı deęişkenin KOAH hastalığı ile yatan hasta sayısı olduęu model incelendiğinde uyum iyilięi testine göre fırtına deęişkeni istatistiksel olarak anlamlı bulunurken ( $P=0.049$ ). Hafta sonu/içi, ortalama sıcaklık, PM10 ve zaman deęişkenleri modelde anlamlı olmadıkları gözlenmiştir ( $P>0.05$ ).

#### **4.2.3. Pnömoni Hastalığı ile İlgili Bulgular**

Çalışmamızda Pnömoni hastalığı bağımlı deęişkenleri ile ilgili gerçek cevap deęişkenine en yakın tahmini yapabilen DGDOM modelini önermek için farklı model yapılarının tahmin performansları kıyaslanmıştır. Pnömoni hastalığı ile ilgili bağımlı deęişkenlerimiz pnömoni vakası nedeniyle hastanelerin acil ve ilgili servisine başvuran hasta sayısı, servise yatan hasta ve eks olan hasta sayılarının günlük deęerleridir. Yapılan model denemelerinde Bölüm 3.1.1’de belirtilen dört modelde bulunan tüm deęişkenler ve uygulamaları pnömoni hastalığının bağımlı deęişkenleri kullanılarak tekrar modellenmiştir. DGDOM yöntemi ile analiz edilen bu dört farklı modelin AIC, düzeltilmiş AIC, BIC, düzeltilmiş  $R^2$  ve fark yüzdesi deęerleri Tablo 4.22’de sunulmuştur. Ayrıca modellerin Anova Gam testi ile farklılıkları ve model uyum iyilięi testleri yapılmıştır.

**Tablo 4.22.** Pnömoni ile hava kirliliği ilişkisinin DGDOM'nin farklı modelleri ile incelenmesi

	<b>Değişken</b>	<b>AIC</b>	<b>AICc</b>	<b>BIC</b>	<b>Fark (%)</b>	<b>Düz. R<sup>2</sup></b>	
<b>PNÖMONİ</b>	<b>Başvuru</b>	<i>Model I</i>	69690.15	69690.25	69739.66	0.2249	0.2056
		<i>Model II</i>	68170.45	68170.55	68219.96	0.2450	0.2308
		<i>Model III</i>	59383.44	59383.54	59432.95	0.3615	0.3272
		<i>Model IV</i>	51889.13	51889.29	51938.64	0.4610	0.4515
	<b>Yatan</b>	<i>Model I</i>	13987.60	13987.70	14037.11	0.0577	0.0467
		<i>Model II</i>	13987.61	13987.71	14037.12	0.0578	0.0467
		<i>Model III</i>	11321.20	11321.30	11370.71	0.3446	0.2810
		<i>Model IV</i>	10740.25	10740.35	10789.76	0.4078	0.3811
	<b>Eks</b>	<i>Model I</i>	2104.43	2104.46	2131.94	0.0274	0.0100
		<i>Model II</i>	2061.81	2061.84	2089.32	0.0600	0.0354
		<i>Model III</i>	2051.24	2051.27	2078.75	0.0745	0.0442
		<i>Model IV</i>	2053.17	2053.20	2080.68	0.0700	0.0415

Düz. R<sup>2</sup>: Düzeltilmiş çoklu açıklayıcılık katsayısı, AIC: Akaike bilgi kriteri, AICc: Düzeltilmiş Akaike bilgi kriteri, BIC: Bayesian bilgi kriteri.

Pnömoni vakasının bağımlı değişkenleri Bölüm 3.1.1'de bahsi geçen dört farklı modelle tahmin edilmiştir. Modellerin fark istatistikleri Tablo 4.23'de sunulmuştur.

**Tablo 4.23.** Pnömoni vakası DGDOM modellerinin Anova Gam değerlendirmeleri

	<b>Değişken</b>	<b>Artık SD</b>	<b>Artık fark</b>	<b>P</b>	
<b>PNÖMONİ</b>	<b>Başvuru</b>	<i>Model I</i>	1790	58524	Temel Model
		<i>Model II</i>	1789	57002	<0.001
		<i>Model III</i>	1784	48205	<0.001
		<i>Model IV</i>	1779	40701	<0.001
	<b>Yatan</b>	<i>Model I</i>	1790	8799	Temel Model
		<i>Model II</i>	1789	8799	0.139
		<i>Model III</i>	1784	6120	<0.001
		<i>Model IV</i>	1779	5529	<0.001
	<b>Eks</b>	<i>Model I</i>	1790	58524	Temel Model
		<i>Model II</i>	1789	57002	<0.001
		<i>Model III</i>	1784	48205	<0.001
		<i>Model IV</i>	1779	40701	<0.001

SD: Serbestlik derecesi

Tablo 4.22’de bağımlı değişkenin pnömoni hastalığı nedeniyle hastaneye başvuran hasta sayısı olduğu modellerin DGDOM yöntemi ile çözümleri incelendiğinde; Model IV AIC (51889.13), Düzeltilmiş AIC (51889.29), BIC (51938.64), fark yüzdesi (0.4610) ve düzeltilmiş  $R^2$  (0.4515) değerlerinin diğer modellerin tahmin performansı sonuçlarından daha üstün olduğu gözlenmiştir. Ayrıca Tablo 4.23’de bağımlı değişkenin başvuru olduğu bölüm incelendiğinde; Model II’den başlayarak bir önceki modele göre ardışık olarak istatistiksel anlamlılığına bakılmıştır. Birbirinden farklı bu dört modelin her yeni modelinin bir önceki modele göre daha iyi sonuçlar verdiğini ve Anova Gam analizi sonucunda bir önceki modele göre istatistiksel olarak anlamlı düzeyde farklı oldukları gözlenmiştir ( $P<0.001$ ).

Tablo 4.22’de bağımlı değişkenin pnömoni hastalığı nedeniyle hastanede yatan hasta sayısı olduğu modellerin DGDOM yöntemi ile çözümleri incelendiğinde; model IV AIC (10740.25), Düzeltilmiş AIC (10740.35), BIC (10789.76), fark yüzdesi (0.4078) ve düzeltilmiş  $R^2$  (0.3811) değerlerinin diğer modellerin tahmin performansı sonuçlarından daha üstün olduğu gözlenmiştir. Ayrıca Tablo 4.23’de bağımlı değişkenin başvuru olduğu bölüm incelendiğinde; modellerin ardışık olarak istatistiksel anlamlılığına bakılmıştır. Birbirinden farklı bu dört denemede model II haricinde her yeni modelinin bir önceki modele göre daha iyi sonuçlar verdiğini ve Anova Gam analizi sonucunda bir önceki modele göre istatistiksel olarak anlamlı düzeyde farklı oldukları gözlenmiştir ( $P<0.001$ ).

Tablo 4.22’de bağımlı değişkenin pnömoni hastalığı nedeniyle eks olan sayısı olduğu modellerin DGDOM yöntemi ile çözümleri incelendiğinde; model III AIC (2051.24), Düzeltilmiş AIC (2051.27), BIC (2078.75), fark yüzdesi (0.0745) ve düzeltilmiş  $R^2$  (0.0442) değerlerinin diğer modellerin tahmin performansı sonuçlarından daha üstün olduğu gözlenmiştir. Ayrıca Tablo 4.23’de bağımlı değişkenin başvuru olduğu bölüm incelendiğinde; modellerin ardışık olarak istatistiksel anlamlılığına bakılmıştır. Birbirinden farklı bu dört denemede model II haricinde her yeni modelinin bir önceki modele göre daha iyi sonuçlar verdiğini ve Anova Gam analizi sonucunda bir önceki modele göre istatistiksel olarak anlamlı düzeyde farklı oldukları gözlenmiştir ( $P<0.001$ ).

**Tablo 4.24.** Pnömoni hastalığı bağımlı değişkenlerinin DGDOM yöntemi ile en iyi sonucu veren modellerinin uyum iyiliği test sonuçları

	<b>Değişken</b>	<b>Kovariyet</b>	<b>Ki-kare</b>	<b>P</b>
<b>PNÖMONİ</b>	<b>Başvuru</b>	<i>Hafta sonu/içi</i>	1.9	0.178
		<i>Fırtına</i>	407.6	<0.001
		<i>Ort. Sıcaklık</i>	183.5	<0.001
		<i>Zaman (d)</i>	2387.0	<0.001
	<b>Yatan</b>	<i>Hafta sonu/içi</i>	0.9	0.345
		<i>Fırtına</i>	0.2	0.635
		<i>Ort. Sıcaklık</i>	0.9	0.325
		<i>Zaman (d)</i>	160	<0.001
	<b>Eks</b>	<i>Hafta sonu/içi</i>	1.9	<0.001
		<i>Fırtına</i>	407.6	<0.001
		<i>Ort. Sıcaklık</i>	183.5	<0.001
		<i>Zaman (d)</i>	2878	<0.001

d: Düzgünleştirme fonksiyonu

Pnömoni hastalığı bağımlı değişkenleri ile en iyi sonuçları elde ettiğimiz model IV ve model III'e ilişkin uyum iyiliği test bulguları Tablo 4.24'de sunulmuştur. Buna göre bağımlı değişkenin pnömoni hastalığı ile başvuran hasta sayısı olduğu model IV incelendiğinde uyum iyiliği testine göre hafta sonu/içi değişkeni istatistiksel olarak anlamlı bulunamamıştır (P=0.178). Fırtına, ortalama sıcaklık, PM10 ve zaman değişkenleri modelde anlamlı olduğu gözlenmiştir (P<0.05). Bağımlı değişkenin pnömoni hastalığı ile başvuran hasta sayısı olduğu model IV incelendiğinde uyum iyiliği testine göre hafta sonu/içi, fırtına ve ortalama sıcaklık değişkenleri istatistiksel olarak anlamlı bulunamamıştır (P>0.05). Sadece zaman değişkeninin modelde anlamlı olduğu gözlenmiştir (P<0.001). Bağımlı değişkenin pnömoni hastalığı ile eks olan sayısı olduğu model III incelendiğinde uyum iyiliği testine göre; hafta içi/sonu, fırtına, ortalama sıcaklık, PM10 ve zaman değişkenleri modelde anlamlı olduğu gözlenmiştir (P<0.001).

#### 4.2.4. Pulmoner Tromboemboli (PTE) hastalığı ile ilgili bulgular

Çalışmamızda pulmoner tromboemboli hastalığı bağımlı değişkenleri ile ilgili gerçek cevap değişkenine en yakın tahmini yapabilen DGDOM modelini önermek için farklı modeller denenmiş ve bu modellerin tahmin performanslarını kıyaslanmıştır. PTE hastalığı ile ilgili bağımlı değişkenlerimiz PTE vakası nedeniyle hastanelerin ilgili servisinde yatan hasta ve eks olan hasta sayılarının günlük değerleridir. Bölüm 3.1.1’de yer alan dört model PTE vakasının iki bağımlı değişkeni için denenmiştir. DGDOM yöntemi ile analiz edilen bu dört farklı modelin AIC, düzeltilmiş AIC, BIC, düzeltilmiş  $R^2$  ve fark yüzdesi değerleri Tablo 4.25’de sunulmuştur. Ayrıca modellerin Anova Gam testi ile farklılıkları ve model uyum iyiliği testleri yapılmıştır. Bu analiz sonuçları sırası ile Tablo 4.26 ve Tablo 4.27’de sunulmuştur.

**Tablo 4.25.** PTE ile hava kirliliği ilişkisinin DGDOM’ın farklı modelleri ile incelenmesi

	Değişken	AIC	AICc	BIC	Fark (%)	Düz. $R^2$	
PTE	Yatan	Model I	5243.60	5243.70	5293.11	0.3657	0.2952
		Model II	5231.73	5231.83	5281.24	0.3693	0.2924
		Model III	4998.62	4998.72	5048.13	0.4324	0.3486
		Model IV	4976.66	4976.76	5026.17	0.4400	0.3556
	Eks	Model I	1215.26	1215.29	1242.77	0.0156	0.0060
		Model II	1206.06	1206.09	1233.57	0.0350	0.0105
		Model III	1206.97	1207.01	1234.48	0.0322	0.0100
		Model IV	1205.23	1204.67	1233.44	0.0512	0.0155

Düz.  $R^2$ : Düzeltilmiş çoklu açıklayıcılık katsayısı, AIC: Akaike bilgi kriteri, AICc: Düzeltilmiş Akaike bilgi kriteri, BIC: Bayesian bilgi kriteri.

Pulmoner Tromboemboli vakasının cevap değişkenleri Bölüm 3.1.1’de bahsi geçen dört farklı modelle tahmin edilmiştir. Modellerin fark istatistikleri Tablo 4.26’da sunulmuştur.

**Tablo 4.26.** Pnömoni vakasına ait Modellerin Anova Gam değerlendirmeleri

	<b>Değişken</b>	<b>Artık SD</b>	<b>Artık fark</b>	<b>P</b>	
<b>PTE</b>	<b>Yatan</b>	<i>Model I</i>	1796.0	841.4	Temel Model
		<i>Model II</i>	1795.1	831.2	<b>&lt;0.001</b>
		<i>Model III</i>	1794.6	831.5	<b>&lt;0.001</b>
		<i>Model IV</i>	1786.2	812.7	<b>0.020</b>
	<b>Eks</b>	<i>Model I</i>	1796.0	2448.0	Temel Model
		<i>Model II</i>	1795.0	2433.0	<b>&lt;0.001</b>
		<i>Model III</i>	1790.0	2178.9	<b>&lt;0.001</b>
		<i>Model IV</i>	17851	2147.5	<b>&lt;0.001</b>

SD: Serbestlik Derecesi

Tablo 4.25’de bağımlı değişkenin PTE hastalığı nedeniyle hastanede yatan hasta sayısı olduğu modeller incelendiğinde; Model IV AIC (4976.66), düzeltilmiş AIC (4976.76), BIC (5026.17), fark yüzdesi (0.4400) ve düzeltilmiş  $R^2$  (0.3556) değerleri elde edilmiştir. Bu değerlere göre model IV AIC, Düzeltilmiş AIC, fark yüzdesi ve Düzeltilmiş  $R^2$  değerleri bakımından diğer modellerden daha üstün olduğu görülmektedir. Ayrıca Tablo 4.26’da bağımlı değişkenin yatan olduğu bölüm incelendiğinde; tüm modellerin Anova Gam analizine göre birbirlerinden istatistiksel olarak anlamlı düzeyde farklı oldukları gözlenmiştir ( $P<0.05$ ).

Tablo 4.12’de bağımlı değişkenin pulmoner tromboemboli hastalığı nedeniyle eks olan hasta sayısı olduğu modeller incelendiğinde; model IV AIC (1205.23), Düzeltilmiş AIC (1204.67), BIC (1233.44) fark yüzdesi (0.0512) ve düzeltilmiş  $R^2$  (0.0155) değerleri gözlenmiştir. Buna göre PTE eks bağımlı değişkeni için kurulan modellerde yine genel anlamda performansı yüksek olan model IV diğer modellerin performans kriteri değerlerine göre üstünlük sağlayarak öne çıkmayı başarmıştır. Ayrıca Tablo 4.13’da bağımlı değişkenin eks olduğu bölüm incelendiğinde Anova Gam analizine göre; model IV, model III’den, model III, model II’den ve son olarak model II’de model I’den istatistiksel olarak anlamlı düzeyde farklı oldukları gözlenmiştir ( $P<0.001$ ).

**Tablo 4.27.** PTE hastalığı bağımlı değişkenlerinin DGDOM yöntemi ile en iyi sonucu veren modellerinin uyum iyiliği test sonuçları

	<b>Değişken</b>	<b>Ki-Kare</b>	<b>P</b>	
<b>PTE</b>	<b>Yatan</b>	<i>Hafta sonu/içi</i>	512.7	<b>&lt;0.001</b>
		<i>Fırtına</i>	0.1	0.695
		<i>Ort. Sıcaklık</i>	0.1	0.813
		<i>Zaman</i>	292.3	<b>&lt;0.001</b>
	<b>Eks</b>	<i>Hafta sonu/içi</i>	0.5	0.470
		<i>Fırtına</i>	2.4	0.125
		<i>Ort. Sıcaklık</i>	0.2	0.264
		<i>Zaman</i>	23.8	<b>0.008</b>

Pulmoner tromboemboli hastalığı ile yatan ve eks olan hasta sayısı bağımlı değişkenleri model IV ile tahmin edilmiştir. Tablo 4.27’de değişkenlere ilişkin uyum iyiliği test sonuçları yer almaktadır. Buna göre bağımlı değişkenin pulmoner tromboemboli hastalığı ile yatan hasta sayısı olduğu model incelendiğinde uyum iyiliği testine göre hafta sonu/içi ve zaman değişkenleri istatistiksel olarak anlamlı bulunmuştur ( $P<0.001$ ). Diğer değişkenler fırtına ( $P=0.695$ ) ve ortalama sıcaklık değişkeni ( $P=0.813$ ) modelde istatistiksel olarak anlamlı olmadıkları gözlenmiştir. Bağımlı değişkenin pulmoner tromboemboli hastalığı ile eks alan hasta sayısı olduğu model incelendiğinde uyum iyiliği testine göre hafta içi/sonu değişkeni ( $P=0.470$ ), ortalama sıcaklık değişkeni ( $P=0.264$ ) ve fırtına değişkeninin ( $P=0.125$ ) istatistiksel olarak anlamlı olmadıkları gözlenmiştir. Zaman değişkeninin istatistiksel olarak anlamlı olduğu gözlendi ( $P=0.008$ ).

DGDOM yöntemi ile modellenmiş dört hastalıkta en iyi tahmin başarısına sahip modellerde PM10 ana etkisi ve gecikmelerinin  $1 \text{ mg/m}^3$  artışı üzerinden tahmin edilen OR ve güven aralıkları Tablo 4.28’de incelenmiştir.

**Tablo 4.28.** PM10 deęişkeninin 1 mg/m<sup>3</sup> artışının DGDOM ile OR tahmini

	Başvuru			Yatan			Eks		
	Deę.	OR [%95 GA]	P	OR [%95 GA]	P	OR [%95 GA]	P		
Astım	<i>Lag0</i>	1.0007 [1.0006-1.0008]	< <b>0.001</b>	1.0001 [0.9999-1.0020]	0.064	1.0015 [0.9956-1.0074]	0.625		
	<i>Lag1</i>	0.9905 [0.9870-0.9940]	< <b>0.001</b>	0.9727 [0.9376-1.0091]	0.139	1.0715 [0.8520-1.3475]	0.555		
	<i>Lag2</i>	1.0350 [1.0170-1.0533]	< <b>0.001</b>	1.1179 [0.9302-1.3434]	0.234	0.6714 [0.1981-2.2757]	0.522		
	<i>Lag3</i>	0.9581 [0.9318-0.9852]	<b>0.003</b>	0.8610 [0.6438-1.1515]	0.313	1.9345 [0.2648-9.1280]	0.515		
	<i>Lag4</i>	1.0173 [1.0034-1.0315]	<b>0.015</b>	1.0680 [0.9252-1.2327]	0.369	0.7184 [0.2651-1.9466]	0.516		
KOAİ	<i>Lag0</i>	1.0005 [0.9852-1.0161]	0.946	0.9025 [0.8222-0.9905]	<b>0.031</b>	1.0488 [0.9068-1.2131]	0.521		
	<i>Lag1</i>	0.9885 [0.8970-1.0892]	0.815	1.9262 [1.0729-3.4582]	<b>0.028</b>	0.7394 [0.2959-1.9852]	0.518		
	<i>Lag2</i>	1.0466 [0.8393-1.3052]	0.686	0.2219 [0.0586-0.8404]	<b>0.027</b>	1.9996 [0.2483-9.8959]	0.515		
	<i>Lag3</i>	0.9384 [0.7562-1.1651]	0.566	4.3921 [1.1922-9.1812]	<b>0.026</b>	0.5041 [0.0652-3.8959]	0.511		
	<i>Lag4</i>	1.0293 [0.9530-1.1118]	0.462	0.5898 [0.3704-0.9390]	<b>0.026</b>	1.2794 [0.6165-2.6547]	0.508		
Pnömoni	<i>Lag0</i>	0.9859 [0.9719-1.0001]	0.052	1.0414 [0.9824-1.1039]	0.172	1.0625 [0.8051-1.4022]	0.668		
	<i>Lag1</i>	1.0886 [0.9948-1.1912]	0.065	0.7657 [0.5304-1.1053]	0.154	0.6792 [0.1186-3.8884]	0.664		
	<i>Lag2</i>	0.8326 [0.6782-1.0222]	0.080	1.8943 [0.8205-4.3733]	0.134	2.4997 [0.0471-98.8886]	0.651		
	<i>Lag3</i>	1.1843 [0.9687-1.4480]	0.099	0.5173 [0.2277-1.1749]	0.115	0.3873 [0.0079-18.8886]	0.632		
	<i>Lag4</i>	0.9450 [0.8796-1.0152]	0.122	1.2807 [0.9550-1.7158]	0.098	1.4356 [0.3593-5.7353]	0.609		
PTE	<i>Lag0</i>			0.9119 [0.8113-1.0250]	0.716	1.0157 [0.6692-1.5414]	0.942		
	<i>Lag1</i>			1.7950 [0.8624-3.7362]	0.451	0.9443 [0.0689-12.9326]	0.966		
	<i>Lag2</i>			0.2631 [0.0498-1.3901]	0.088	1.0379 [0.0027-89.5111]	0.990		
	<i>Lag3</i>			3.6811 [0.7236-9.7749]	0.256	1.0464 [0.0031-88.9882]	0.988		
	<i>Lag4</i>			0.6307 [0.3535-1.1252]	0.235	0.9614 [0.1222-7.5632]	0.970		

PTE: Pulmoner Tromboemboli, GA: Güven aralığı, OR: Odds Oranı.



### 4.3. Otoregresif Terimli Genelleştirilmiş Eklemeli Modellerle (GAMAR) yapılan analizlerin sonuçları

Gerçek cevap değişkenine en yakın tahmini yapabilen GAMAR modelini önermek için Bölüm 3.1.1’de bahsi geçen dört model uygulanmış, ilgili modellerin tahmin performansları kıyaslanmıştır. Modellerde bağımlı değişkenler astım, KOAH, pnömoni ve pulmoner tromboemboli vakaları nedeniyle hastanelerin acil ve ilgili servisine başvuran hasta sayısı, servise yatan hasta ve eks olan hasta sayılarının günlük değerleridir. Dört modelin GAMAR yöntemi ile analizleri yapılmıştır ve modellerin AIC, düzeltilmiş AIC, BIC, fark yüzdesi ve düzeltilmiş  $R^2$  gibi model tahmin performansı kriterleri incelenmiştir.

#### 4.3.1. Astım hastalığı ile ilgili bulgular

Çalışmamızda astım hastalığı bağımlı değişkenleri ile ilgili gerçek cevap değişkenine en yakın tahmini yapabilen GAMAR modelini önermek için dört farklı model denemesi yapılmış ve bu modellerin tahmin performansları kıyaslanmıştır. Astım hastalığı modellerinde bağımlı değişkenlerimiz astım vakası nedeniyle hastanelerin acil ve ilgili servisine başvuran hasta sayısı, servise yatan hasta ve eks olan sayılarının günlük değerleridir. Bölüm 3.1.1’de tanımlanan dört farklı modelin AIC, düzeltilmiş AIC, BIC, düzeltilmiş  $R^2$  ve fark yüzdesi değerleri Tablo 4.28’de sunulmuştur. Hangi GAMAR modelinin hava kirliliği ile astım vakası nedeniyle başvuran, yatan ve eks olan ilişkisini tahmin etmede daha başarılı olacağı araştırılmıştır.

**Tablo 4.29.** Astım ile hava kirliliği ilişkisinin GAMAR’nin farklı modelleri ile incelenmesi

Değişken		AIC	AICc	BIC	Fark	Düz. $R^2$
<b>Başvuru</b>	<i>Model I</i>	28564.44	28564.54	28613.95	0.8320	0.7877
	<i>Model II</i>	28461.95	28462.05	28511.46	0.8348	0.7882
	<i>Model III</i>	28165.92	28166.02	28215.43	0.8573	0.7981
	<i>Model IV</i>	27487.19	27487.29	27536.70	0.8602	0.8011
<b>ASTİM</b> <b>Yatan</b>	<i>Model I</i>	4304.73	4304.83	4354.24	0.4853	0.4677
	<i>Model II</i>	4271.03	4271.13	4320.54	0.5015	0.4744
	<i>Model III</i>	4420.44	4420.54	4469.95	0.4604	0.4305
	<i>Model IV</i>	4201.26	4201.36	4250.77	0.5862	0.5001
<b>Eks</b>	<i>Model I</i>	298.28	298.48	347.89	0.0545	0.0098
	<i>Model II</i>	297.38	297.48	346.89	0.0599	0.0101
	<i>Model III</i>	296.55	296.65	346.06	0.0620	0.1100
	<i>Model IV</i>	294.00	294.10	343.51	0.0700	0.1251

Düz.  $R^2$ : Düzeltilmiş çoklu açıklayıcılık katsayısı, AIC: Akaike bilgi kriteri, AICc: Düzeltilmiş Akaike bilgi kriteri, BIC: Bayesian bilgi kriteri.

Tablo 4.28’de bağımlı değişkenin astım hastalığı nedeniyle hastaneye başvuran hasta sayısı olduğu modellerin GAMAR yöntemi ile çözümleri incelendiğinde; zaman değişkeninin mevsimsel etkisinin doğal zincir düzgünleştirme fonksiyonu ile düzgünleştirildiği model IV AIC (27487.19), düzeltilmiş AIC (27487.29), BIC (27536.70), fark yüzdesi (0.8602) ve düzeltilmiş  $R^2$  (0.8011) değerlerinin diğer modellerin ilgili performans değerlerinden daha üstün olduğu gözlenmiştir.

Tablo 4.28’de bağımlı değişkenin astım hastalığı nedeniyle yatan hasta sayısı olduğu modellerin GAMAR yöntemi ile çözümleri incelendiğinde; zaman değişkeninin mevsimsel etkisinin doğal zincir düzgünleştirme fonksiyonu ile düzgünleştirildiği model IV AIC (4201.26), düzeltilmiş AIC (4201.36), BIC (4250.77), fark yüzdesi (0.5862) ve düzeltilmiş  $R^2$  (0.5001) değerlerinin diğer modellerin ilgili performans değerlerinden daha üstün olduğu gözlenmiştir.

Tablo 4.28’de bağımlı değişkenin astım hastalığı nedeniyle eks olan sayısı olduğu modellerin GAMAR yöntemi ile çözümleri incelendiğinde; zaman değişkeninin mevsimsel etkisinin doğal zincir düzgünleştirme fonksiyonu ile düzgünleştirildiği model IV AIC (294.00), düzeltilmiş AIC (294.10), BIC (343.51), fark yüzdesi (0.0700) ve düzeltilmiş  $R^2$  (0.1251) değerlerinin diğer modellerin ilgili performans değerlerinden daha üstün olduğu gözlenmiştir.

#### **4.3.2. KOAH hastalığı ile ilgili bulgular**

KOAH hastalığı bağımlı değişkenleri ile ilgili gerçek cevap değişkenine en yakın tahmini yapabilen GAMAR modelini önermek için dört farklı GAMAR modeli denenmiş ve bu modellerin tahmin performansları kıyaslanmıştır. KOAH hastalığı ile ilgili modellerde bağımlı değişkenler; KOAH vakası nedeniyle hastanelerin acil ve ilgili servisine başvuran hasta sayısı, servise yatan hasta ve eks olan sayılarının günlük değerleridir. Bölüm 3.1.1’de tanımlanan ve GAMAR yöntemi ile modellenen dört farklı modelin AIC, düzeltilmiş AIC, BIC, düzeltilmiş  $R^2$  ve fark yüzdesi bulguları Tablo 4.29’da sunulmuştur. Hangi modelin hava kirliliği ile KOAH vakası nedeniyle başvuran, yatan ve eks olan ilişkisini tahmin etmede daha başarılı olacağı araştırılmıştır.

**Tablo 4.30.** KOAH ile hava kirliliği ilişkisinin GAMAR'nin farklı modelleri ile incelenmesi

Değişken		AIC	AICc	BIC	Fark (%)	Düz. R <sup>2</sup>	
<b>Başvuru</b>	<i>Model I</i>	21550.18	21550.28	21599.69	0.6992	0.6681	
	<i>Model II</i>	21506.09	21506.19	21555.60	0.7145	0.6856	
	<i>Model III</i>	21240.19	21240.29	21289.70	0.7468	0.6986	
	<i>Model IV</i>	20848.38	20848.48	20897.89	0.7652	0.7225	
<b>KOAH</b>	<b>Yatan</b>	<i>Model I</i>	6302.11	6302.21	6351.62	0.5004	0.4994
		<i>Model II</i>	6296.22	6296.32	6345.73	0.5125	0.5089
		<i>Model III</i>	6162.27	6162.37	6211.78	0.6124	0.6001
		<i>Model IV</i>	6136.14	6136.24	6185.65	0.6201	0.6142
<b>Eks</b>	<i>Model I</i>	4400.55	4400.65	4450.06	0.0365	0.0456	
	<i>Model II</i>	4325.36	4325.46	4374.87	0.0965	0.0868	
	<i>Model III</i>	4303.33	4303.43	4352.84	0.1021	0.0998	
	<i>Model IV</i>	4290.49	4290.69	4340.30	0.1213	0.1155	

Düz. R<sup>2</sup>: Düzeltilmiş çoklu açıklayıcılık katsayısı, AIC: Akaike bilgi kriteri, AICc: Düzeltilmiş Akaike bilgi kriteri, BIC: Bayesian bilgi kriteri.

Tablo 4.29'da bağımlı değişkenin KOAH hastalığı nedeniyle hastaneye başvuran hasta sayısı olduğu modellerin GAMAR yöntemi ile çözümleri incelendiğinde; zaman değişkeninin mevsimsel etkisinin doğal zincir düzgünleştirme fonksiyonu ile düzgünleştirildiği model IV AIC (20848.38), Düzeltilmiş AIC (20848.48), BIC (20897.89), fark yüzdesi (0.7652) ve düzeltilmiş R<sup>2</sup> (0.7225) değerlerinin diğer modellerin ilgili performans değerlerinden daha üstün olduğu gözlenmiştir. Tablo 4.29'da bağımlı değişkenin KOAH hastalığı nedeniyle hastanede yatan hasta sayısı olduğu modellerin GAMAR yöntemi ile çözümleri incelendiğinde; zaman değişkeninin mevsimsel etkisinin doğal zincir düzgünleştirme fonksiyonu ile düzgünleştirildiği model IV AIC (6136.14), Düzeltilmiş AIC (6136.24), BIC (6185.65), fark yüzdesi (0.6201) ve düzeltilmiş R<sup>2</sup> (0.6142) değerlerinin diğer modellerin ilgili performans değerlerinden daha üstün olduğu gözlenmiştir. Tablo 4.28'de bağımlı değişkenin KOAH hastalığı nedeniyle eks olan sayısı olduğu modellerin GAMAR yöntemi ile çözümleri incelendiğinde; zaman değişkeninin mevsimsel etkisinin doğal zincir düzgünleştirme fonksiyonu ile düzgünleştirildiği model IV AIC (4290.49), Düzeltilmiş AIC (4290.69),

BIC (4340.30), fark yüzdesi (0.1213) ve düzeltilmiş  $R^2$  (0.1155) değerlerinin diğer modellerin ilgili performans değerlerinden daha üstün olduğu gözlenmiştir

#### 4.3.3. Pnömoni hastalığı ile ilgili bulgular

Çalışmamızda pnömoni hastalığı bağımlı değişkenleri ile ilgili gerçek cevap değişkenine en yakın tahmini yapabilen GAMAR modelini önermek için farklı model yapılarının tahmin performanslarını kıyaslanmıştır. Pnömoni hastalığı ile ilgili bağımlı değişkenlerimiz pnömoni vakası nedeniyle hastanelerin acil ve ilgili servisine başvuran hasta sayısı, servise yatan hasta ve eks olan hasta sayılarının günlük değerleridir. Yapılan model denemelerinde Bölüm 3.1.1’de belirtilen dört modelde bulunan tüm değişkenler ve uygulamaları pnömoni hastalığının bağımlı değişkenleri kullanılarak tekrar modellenmiştir. GAMAR yöntemi ile analiz edilen bu dört farklı modelin AIC, düzeltilmiş AIC, BIC, düzeltilmiş  $R^2$  ve fark yüzdesi değerleri Tablo 4.30’da sunulmuştur.

**Tablo 4.31.** Pnömoni ile hava kirliliği ilişkisinin GAMAR’nin farklı modelleri ile incelenmesi

	Değişken	AIC	AICc	BIC	Fark (%)	Düz. $R^2$	
PNÖMONİ	Başvuru	Model I	56658.90	56659.00	56708.41	0.3253	0.3157
		Model II	56464.36	54646.46	56513.87	0.3442	0.3407
		Model III	54190.44	54190.54	54239.95	0.4612	0.4265
		Model IV	48796.63	48796.73	48846.14	0.5623	0.5154
	Yatan	Model I	11745.64	11745.74	11795.15	0.3810	0.2991
		Model II	11747.41	11747.51	11796.92	0.3815	0.3050
		Model III	11281.28	11281.38	11330.79	0.4001	0.3856
		Model IV	10682.67	10682.77	10732.18	0.5150	0.4846
	Eks	Model I	2110.49	2110.59	2160.00	0.0260	0.0070
		Model II	2072.59	2072.69	2122.10	0.0456	0.0266
		Model III	2062.45	2062.55	2111.96	0.0701	0.0422
		Model IV	2074.39	2074.49	2123.90	0.0685	0.0388

Düz.  $R^2$ : Düzeltilmiş çoklu açıklayıcılık katsayısı, AIC: Akaike bilgi kriteri, AICc: Düzeltilmiş Akaike bilgi kriteri, BIC: Bayesian bilgi kriteri.

Tablo 4.30’da bağımlı değişkenin Pnömoni hastalığı nedeniyle hastaneye başvuran hasta sayısı olduğu modellerin GAMAR yöntemi ile çözümleri incelendiğinde; Model IV AIC (48796.63), Düzeltilmiş AIC (48796.73), BIC (48846.14), fark yüzdesi (0.5623) ve düzeltilmiş  $R^2$  (0.5154) değerlerinin diğer modellerin tahmin performansı sonuçlarından daha üstün olduğu gözlenmiştir. Tablo 4.30’da bağımlı değişkenin pnömoni hastalığı nedeniyle hastanede yatan hasta sayısı olduğu modellerin GAMAR yöntemi ile çözümleri incelendiğinde; zaman değişkeninin mevsimsel etkisinin doğal zincir düzgünleştirme fonksiyonu ile düzgünleştirildiği model IV AIC (10682.67), Düzeltilmiş AIC (10682.77), BIC (10732.18), fark yüzdesi (0.5150) ve düzeltilmiş  $R^2$  (0.4846) değerlerinin diğer modellerin ilgili performans değerlerinden daha üstün olduğu gözlenmiştir. Tablo 4.30’da bağımlı değişkenin pnömoni hastalığı nedeniyle eks olan sayısı olduğu modellerin GAMAR yöntemi ile çözümleri incelendiğinde; zaman değişkeninin mevsimsel etkisinin doğal zincir düzgünleştirme fonksiyonu ile düzgünleştirildiği model IV AIC (2074.39), Düzeltilmiş AIC (2074.49), BIC (2123.90), fark yüzdesi (0.0685) ve düzeltilmiş  $R^2$  (0.0388) değerlerinin diğer modellerin ilgili performans değerlerinden daha üstün olduğu gözlenmiştir

#### **4.3.4. Pulmoner Tromboemboli (PTE) hastalığı ile ilgili bulgular**

Çalışmamızda PTE hastalığı bağımlı değişkenleri ile ilgili gerçek cevap değişkenine en yakın tahmini yapabilen GAMAR modelini önermek için farklı modeller denenmiş ve bu modellerin tahmin performanslarını kıyaslanmıştır. Pulmoner tromboemboli hastalığı ile ilgili bağımlı değişkenlerimiz PTE vakası nedeniyle hastanelerin ilgili servisinde yatan hasta ve eks olan hasta sayılarının günlük değerleridir. Bölüm 3.1.1’de yer alan dört model PTE vakasının iki bağımlı değişkeni için denenmiştir. GAMAR yöntemi ile analiz edilen bu dört farklı modelin AIC, düzeltilmiş AIC, BIC, düzeltilmiş  $R^2$  ve fark yüzdesi değerleri Tablo 4.31’de sunulmuştur.

**Tablo 4.32.** PTE ile hava kirliliği ilişkisinin GAMAR'ın farklı modelleri ile incelenmesi

	<b>Değişken</b>	<b>AIC</b>	<b>AICc</b>	<b>BIC</b>	<b>Fark (%)</b>	<b>Düz. R<sup>2</sup></b>	
<b>PTE</b>	<b>Yatan</b>	<i>Model I</i>	5178.98	5179.08	5228.49	0.3854	0.3001
		<i>Model II</i>	5172.67	5172.78	5222.18	0.4103	0.3104
		<i>Model III</i>	4983.11	4983.21	5032.62	0.4402	0.3665
		<i>Model IV</i>	4958.13	4958.23	5007.64	0.4503	0.3702
	<b>Eks</b>	<i>Model I</i>	1210.76	1210.86	1260.27	0.0110	0.0080
		<i>Model II</i>	1203.93	1204.03	1253.44	0.0651	0.0205
		<i>Model III</i>	1209.91	1210.01	1259.42	0.0200	0.0110
		<i>Model IV</i>	1206.16	1206.26	1255.67	0.0281	0.0140

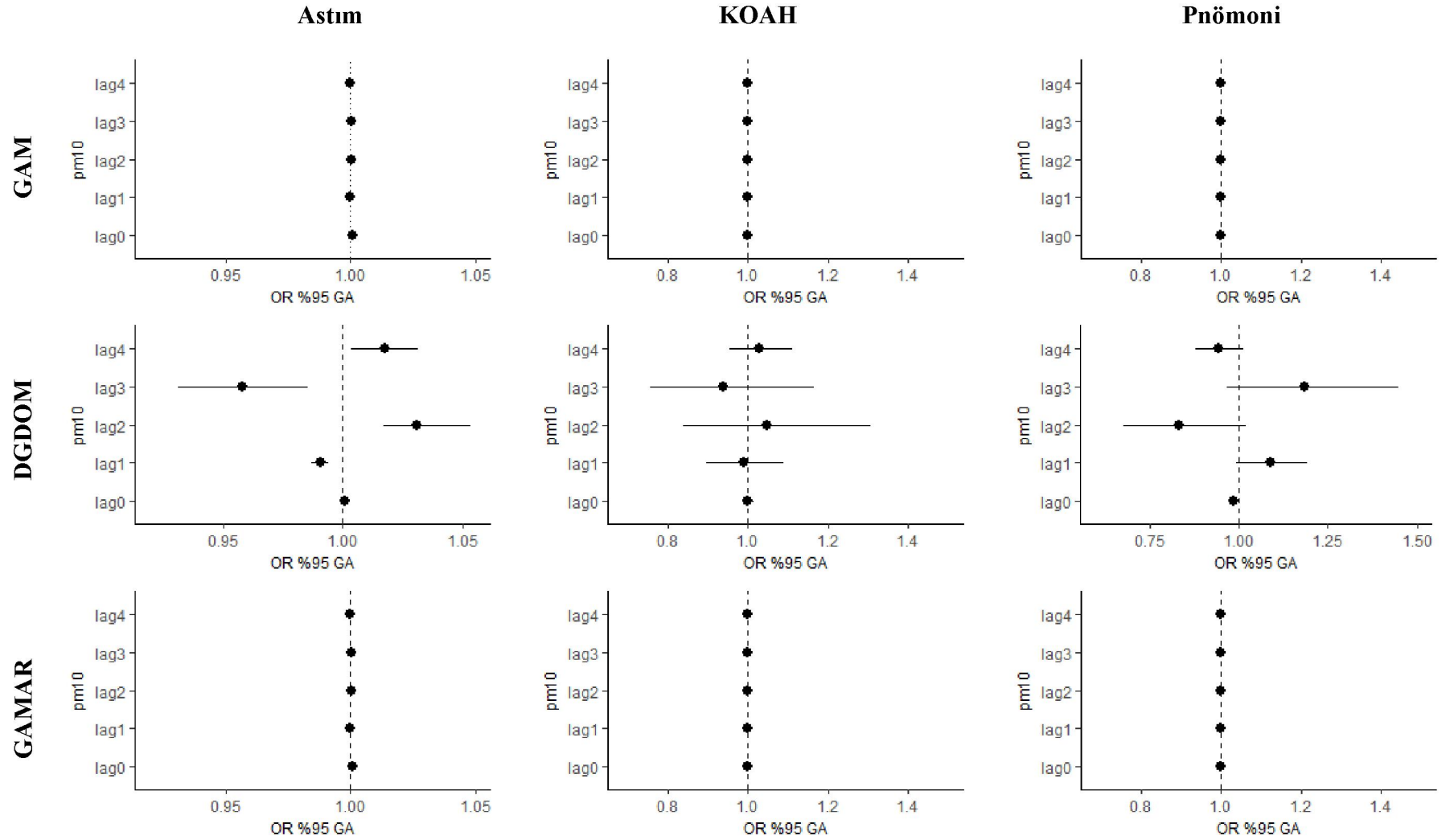
Düz. R<sup>2</sup>: Düzeltilmiş çoklu açıklayıcılık katsayısı, AIC: Akaike bilgi kriteri, AICc: Düzeltilmiş Akaike bilgi kriteri, BIC: Bayesian bilgi kriteri.

Tablo 4.31'de bağımlı değişkenin pulmoner tromboemboli hastalığı nedeniyle hastanede yatan hasta sayısı olduğu modeller incelendiğinde; model IV AIC (4958.13), düzeltilmiş AIC (4958.23), BIC (5007.64), fark yüzdesi (0.4503) ve düzeltilmiş R<sup>2</sup> (0.3702) değerleri elde edilmiştir. Bu değerler ışığında Model IV'ün AIC, düzeltilmiş AIC, fark yüzdesi ve düzeltilmiş R<sup>2</sup> değerleri bakımından diğer modellerden üstün olduğu görülmüştür. Tablo 4.31'de bağımlı değişkenin pulmoner tromboemboli hastalığı nedeniyle eks olan sayısı olduğu modeller incelendiğinde; model II'nin AIC (1203.93), düzeltilmiş AIC (1204.03), BIC (1253.44), fark yüzdesi (0.0651) ve düzeltilmiş R<sup>2</sup> (0.0205) değerleri elde edilmiştir. Bu değerler ışığında Model II AIC, düzeltilmiş AIC, fark yüzdesi ve düzeltilmiş R<sup>2</sup> değerleri bakımından diğer modellerden üstün olduğu gözlenmiştir.

**Tablo 4.33.** PM10 deęişkeninin 1 mg/m<sup>3</sup> artışının GAMAR ile OR tahmini

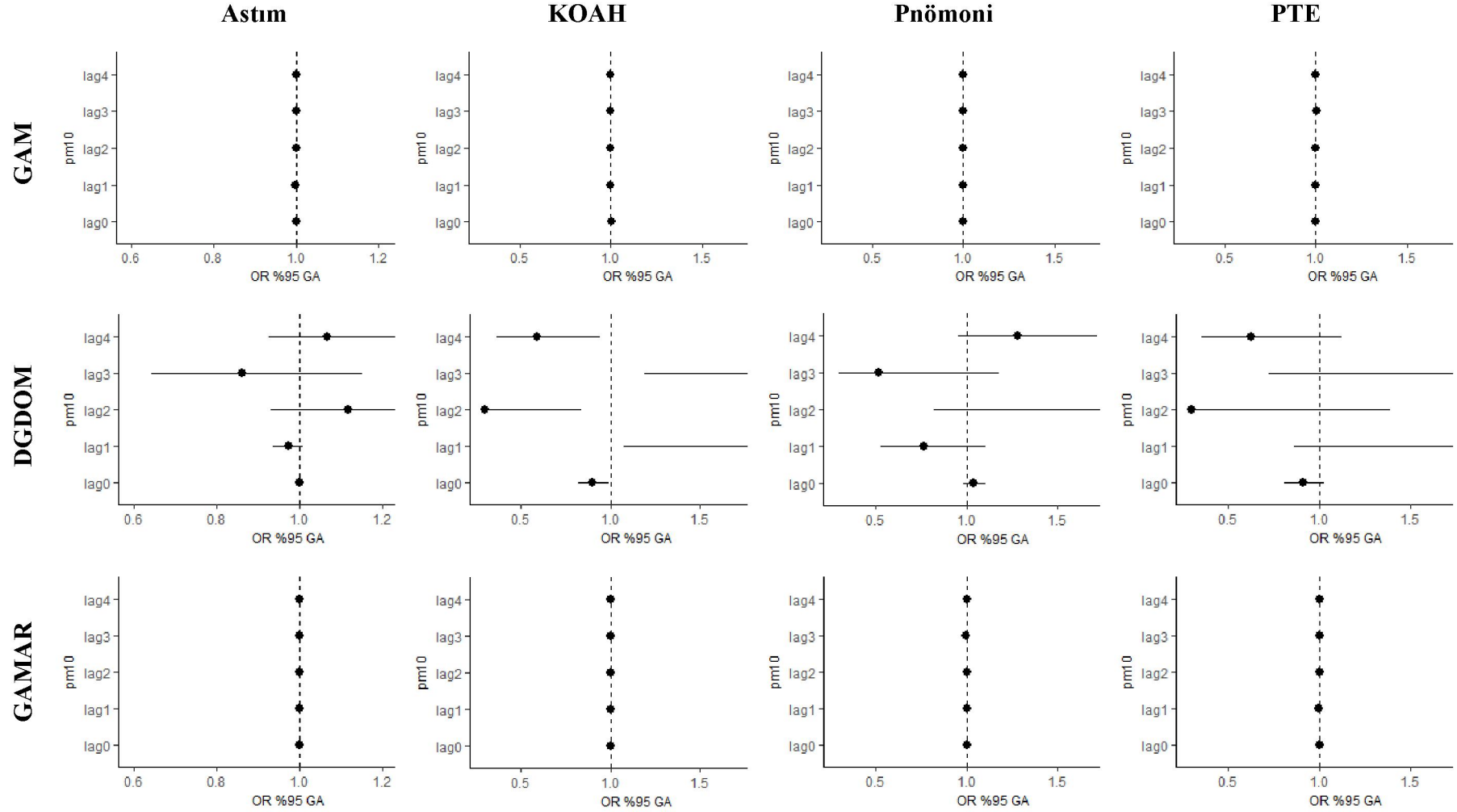
	Başvuru	Yatan	Eks	
	Deę.	OR [%95GA]	OR [%95GA]	
Astım	<i>Lag0</i>	1.0007 [1.0006-1.0007]	1.0009 [0.9999-1.0019]	1.0030 [0.9979-1.0081]
	<i>Lag1</i>	0.9998 [0.9997-0.9999]	0.9993 [0.9982-1.0005]	1.0014 [0.9964-1.0064]
	<i>Lag2</i>	1.0002 [1.0001-1.0003]	1.0004 [0.9993-1.0015]	1.0009 [0.9960-1.0068]
	<i>Lag3</i>	1.0002 [1.0001-1.0003]	1.0005 [0.9995-1.0016]	1.0019 [0.9945-1.0079]
	<i>Lag4</i>	0.9999 [0.9999-0.9999]	1.0008 [0.9996-1.0018]	0.9998 [0.8279-1.0052]
KOAH	<i>Lag0</i>	1.0005 [1.0003-1.0006]	1.0008 [1.0003-1.0016]	0.9996 [0.9983-1.0007]
	<i>Lag1</i>	0.9999 [0.9998-1.0000]	0.9998 [0.9994-1.0003]	1.0005 [0.9992-1.0018]
	<i>Lag2</i>	0.9998 [0.9997-0.9999]	1.0002 [0.9997-1.0011]	1.0004 [0.9990-1.0016]
	<i>Lag3</i>	1.0001 [0.9999-1.0002]	1.0002 [0.9999-1.0011]	1.0012 [0.9999-1.0024]
	<i>Lag4</i>	0.9998 [0.9997-0.9999]	1.0002 [0.9994-1.0010]	0.9994 [0.9991-1.0039]
Prümoni	<i>Lag0</i>	1.0001 [1.0000-1.0003]	1.0002 [0.9998-1.0007]	1.0018 [0.9975-1.0027]
	<i>Lag1</i>	0.9999 [0.9998-1.0001]	0.9997 [0.9992-1.0003]	1.0001 [0.9986-1.0038]
	<i>Lag2</i>	0.9999 [0.9997-1.0000]	0.9999 [0.9994-1.0005]	1.0012 [0.9980-1.0033]
	<i>Lag3</i>	1.0001 [0.9999-1.0002]	0.9993 [0.9987-0.9998]	1.0065 [0.9983-1.0031]
	<i>Lag4</i>	1.0001 [0.9997-1.0002]	1.0005 [0.9981-1.0009]	1.0006 [0.9981-1.0030]
PTE	<i>Lag0</i>		1.0005 [0.9995-1.0015]	1.0020 [0.9992-1.0047]
	<i>Lag1</i>		0.9981 [0.9969-0.9994]	1.0038 [1.0009-1.0067]
	<i>Lag2</i>		0.9998 [0.9987-1.0009]	0.9992 [0.9958-1.0026]
	<i>Lag3</i>		1.0008 [0.9997-1.0017]	1.0024 [0.9993-1.0056]
	<i>Lag4</i>		0.9996 [0.9985-1.0006]	1.0018 [0.9987-1.0049]

GAMAR yöntemi ile modellenmiş dört hastalıkta en iyi tahmin başarısına sahip modellerde PM10 ana etkisi ve gecikmelerinin 1 mg/m<sup>3</sup> artışı üzerinden tahmin edilen OR ve güven aralıkları Tablo 4.32’de incelenmiştir.

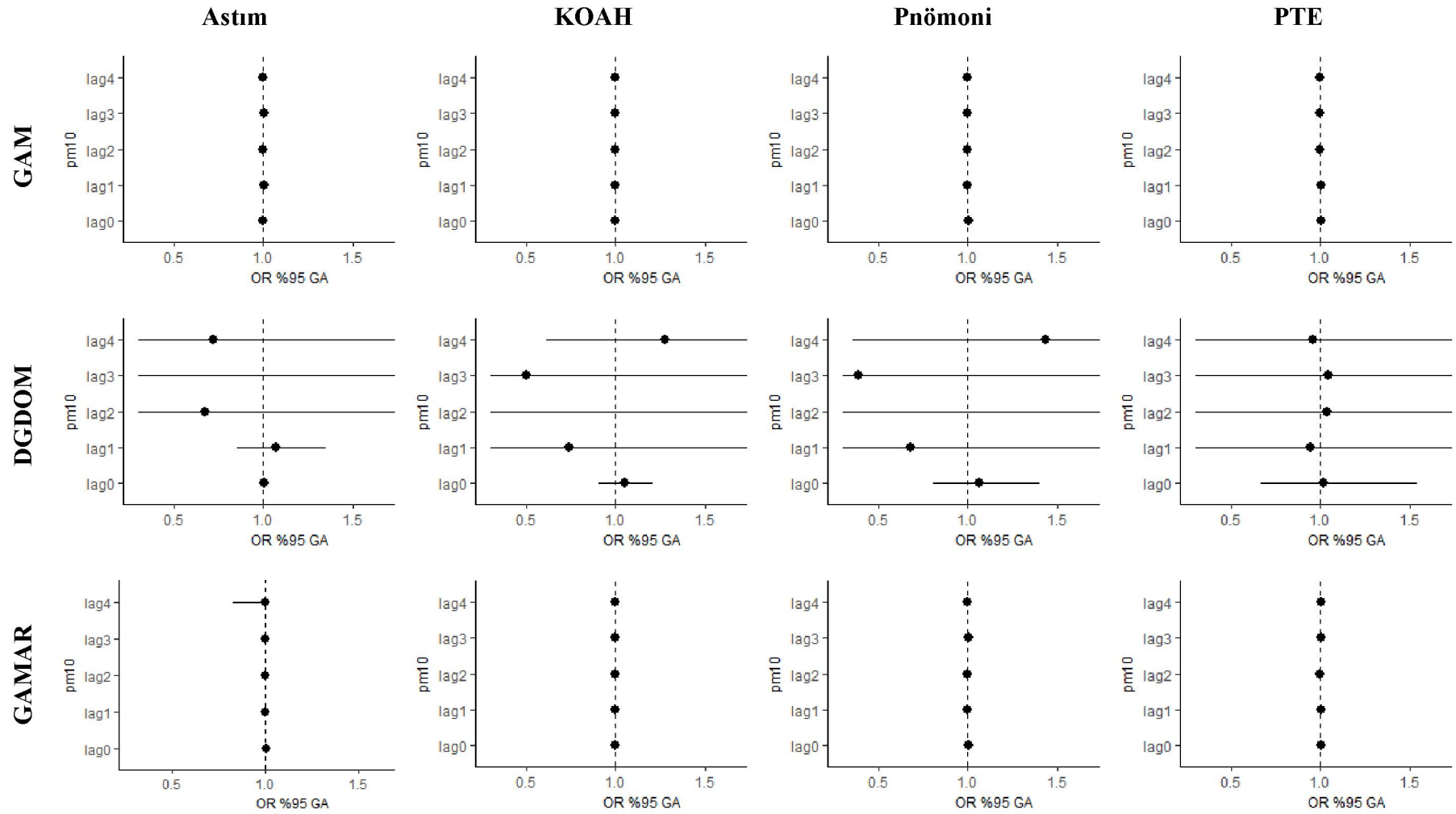


Şekil 4.1. Astım, KOAH ve Pnömoni hastalıklarından başvuru sayıları üzerine PM10 ve gecikme etkilerinin OR (%95 GA) grafikleri





Şekil 4.2. Astım, KOAH, Pnömoni ve PTE hastalıklarından yatan sayıları üzerine PM10 ve gecikme etkilerinin OR (%95 GA) grafikleri



Şekil 4.3. Astım, KOAH, Pnömoni ve PTE hastalıklarından yatan sayıları üzerine PM10 ve gecikme etkilerinin OR (%95 GA) grafikleri

## 5. TARTIŞMA

Gaziantep ili Türkiye'nin güneydoğusunda yer alır, yazları sıcak ve kuru bir hava sıcaklığına maruz kalır. Ayrıca bu bölge Arabistan ve Afrika'dan gelen çöl fırtınalarının taşıdığı tozdan ciddi anlamda etkilenir. Sık aralıklarla meydana gelen çöl fırtınalarının tozuna eşlik eden sıcaklık, insan sağlığı üzerinde önemli bir tehdit oluşturmaktadır (4). Bu sebepten çevre faktörlerinin insan sağlığı üzerine etkilerini incelemek için Gaziantep ili uygun bir örnek olacağı düşünülmüştür.

Bu çalışmada Gaziantep ilinin farklı hastanelerinden elde ettiğimiz gerçek veri ile 4 farklı hastalığın 3 farklı sonuç değişkeni 3 farklı yöntem kullanılarak modellenmiş ve yöntemlerin performans sonuçları karşılaştırılarak en iyi model araştırılmıştır.

Üç farklı yöntem karşılaştırılırken her bir sonuç değişkeninde farklı düzleştirme fonksiyonuna ve serbestlik derecesine sahip 4 farklı model kurulmuştur. Aynı yöntemle kurulan 4 farklı modelden en iyi performansa sahip model ilgili yöntem için karşılaştırma modeli olarak kullanılmıştır. Tahminlerin performansının değerlendirilmesinde AIC, düzeltilmiş AIC, BIC, fark yüzdesi ve düzeltilmiş  $R^2$  kriterleri kullanılmıştır.

Astım nedeni ile hastaneye başvuran sayılarının üzerine çevresel faktörlerin ve mevsimsellik etkisinin değerlendirilmesinde GAM, DGDOM ve GAMAR yöntemleri kullanılmıştır. Model performans kriterleri bakımından GAM ve DGDOM benzer sonuçlar verirken daha iyi tahmin başarısına sahip yöntemin GAMAR olduğu gözlenmiştir. En düşük AIC (27487.19), düzeltilmiş AIC (27487.29), BIC (27536.70) değerlerini ve en yüksek fark yüzdesi (0.8602) ve  $R^2$  (0.8011) değerleri GAMAR'dan elde edilmiştir. PM10 ana etkisi ve gecikmeleri üzerinden tahmin edilen OR grafik sonuçları değerlendirdiğinde (Şekil 4.1); astım başvurusu cevap değişkeni tahmininde GAM ve GAMAR yöntemlerinin daha düşük standart hataya dolayısıyla daha dar güven aralıklarına sahip oldukları gözlenmiştir. Buna karşın DGDOM grafiklerinde OR güven aralıklarında bozulmalar olduğu gözlenmiştir. Ayrıca istatistik anlamlılık bakımından 3 yöntem sonucu karşılaştırıldığında DGDOM yönteminin sonuçlarının anlamlılığının diğer yöntemlerden farklı olduğu gözlenmiştir.

KOAH ve pnömoni nedeni ile hastaneye başvuru sayıları üzerine çevresel faktörlerin ve mevsimsellik etkisinin değerlendirilmesinde performans kriterlerine göre tahmin başarısı en iyi olan yöntemin GAMAR olduğu gözlenmiştir. KOAH ve pnömoni

nedeni ile acil servise başvuran hasta sayısının PM10 ana etkisi ve gecikmeleri üzerinden tahmin edilen OR grafik sonuçları incelendiğinde (Şekil 4.1); KOAH hastalığı için GAM ve GAMAR'ın benzer sonuçlar verdiği fakat DGDOM yönteminin risk katsayılarının yönleri ve güven aralığı genişliği bakımından diğer metotlardan farklı sonuçlar verdiği gözlenmiştir. Pnömoni ile ilgili tahminlerdeki OR grafikleri incelendiğinde en tutarlı görüntünün GAM ve GAMAR yöntemlerinde olduğu, DGDOM yönteminin OR güven aralıklarının daha geniş dolayısıyla standart hatalarının daha yüksek olduğu gözlenmiştir.

Ma ve arkadaşlarının başvuran hasta sayıları ile hava sıcaklığı arasındaki ilişkiyi araştırdıkları makalelerinde çalışmamıza benzer sonuçlar gözlemlemişlerdir (66, 67). Bu çalışmada yapılan başvurunun tipik bir zaman serisine ve güçlü bir otokorelasyon sahip olduğunu düşünmüşlerdir. GAM ve otoregresif terime sahip GAM (GAMAR) modelleri ile yaptıkları karşılaştırma analizlerinde tahmin başarısı yönünden otoregresif terimli GAM yöntemini daha başarılı bulmuşlardır. Çalışmamızın sonuçlarına benzer bir diğer çalışmada ise Gasparini tarafından 2011 yılında yapılmıştır. Çalışmada DGDOM ile ilgili uygulamada ceza zincir yapısına dayalı düzgünleştirme fonksiyonlarına sahip GAM (Wood 2006) modellerinin DGDOM'ye iyi bir alternatif olacağı sonucuna varmışlardır (11).

Astım nedeni ile serviste yatan hasta sayılarının üzerine çevresel faktörlerin ve mevsimsellik etkisinin değerlendirilmesinde GAM, DGDOM ve GAMAR yöntemleri kullanılmıştır. Performans kriterleri bakımından GAM ve DGDOM benzer sonuçlar verirken daha iyi tahmin başarısına sahip yöntemin yine GAMAR olduğu gözlenmiştir. En düşük AIC (4201.26), düzeltilmiş AIC (4201.36), BIC (4250.77) değerlerini ve en yüksek fark yüzdesi (0.5862) ve  $R^2$  (0.5001) değerleri GAMAR'dan elde edilmiştir. PM10 ana etkisi ve gecikmeleri üzerinden tahmin edilen OR grafik sonuçları değerlendirdiğinde (Şekil 4.2); astım yatan cevap değişkeni tahmininde GAM ve GAMAR yöntemlerinin OR grafiklerinin daha dar başka bir deyişle standart hatası düşük grafikler olduğu görülmüştür. Fakat DGDOM yönteminin yüksek standart hataya sahip tahminler yaptığı gözlenmiştir.

KOAH, pnömoni ve PTE nedeni ile acil serviste yatan hasta sayıları üzerine çevresel faktörlerin ve mevsimsellik etkisinin değerlendirilmesinde performans kriterlerine göre tahmin başarısı en iyi olan yöntemin GAMAR olduğu gözlenmiştir. Tahmin başarı kriterlerinden fark yüzdesi ve  $R^2$  gibi kriterlerdeki düşüşlerin sebebinin yatan hasta sayılarında sıfır gözlemlerinin varlığının fazla olmasından kaynaklandığı düşünülmektedir. KOAH, pnömoni ve PTE nedeni ile acil servise başvuran hasta

sayısının PM10 ana etkisi ve gecikmeleri üzerinden tahmin edilen OR grafik sonuçları incelendiğinde (Şekil 4.2); GAM ve GAMAR yöntemleri dar güven aralıklarına sahip OR grafikleri ortaya koymuşken DGDOM yönteminin geniş güven aralıklarına sahip OR tahminleri yaptığı gözlenmiştir.

Astım nedeni ile eks olan hasta sayılarının üzerine çevresel faktörlerin ve mevsimsellik etkisinin değerlendirilmesinde performans kriterleri bakımından daha iyi tahmin başarısına sahip yöntemin DGDOM olduğu gözlenmiştir. En düşük AIC (290.48), düzeltilmiş AIC (290.51), BIC (317.99) değerlerini ve en yüksek fark yüzdesi (0.1129) ve  $R^2$  (0.0275) değerleri DGDOM'den elde edilmiştir. Eks olan hasta tahmin başarı kriterlerinden fark yüzdesi ve  $R^2$  gibi kriterlerdeki düşük değerlerin sebebi, eks olan hasta sayılarının da yatan hasta sayılarında olduğu gibi sıfır gözlemlerinin varlığının fazla olmasından kaynaklandığı düşünülmektedir. Bu sonuçlara göre eks bağımlı değişkenleri için denenen modellerin hiçbiri tahmin başarı gösteremediği gözlenmiştir. PM10 ana etkisi ve gecikmeleri üzerinden tahmin edilen OR grafik sonuçları değerlendirdiğinde (Şekil 4.3); astım başvurusu cevap değişkeni için DGDOM yönteminde standart hatası yüksek dolayısı ile geniş güven aralıklarına sahip grafikler gözlenmiştir. GAM ve GAMAR yöntemleri ise diğer yöntemlere göre daha dar güven aralıklı OR grafikleri ortaya koymuştur.

KOAH ve pnömoni nedeni ile eks olan sayıları üzerine çevresel faktörlerin ve mevsimsellik etkisinin değerlendirilmesinde performans kriterlerine göre tahmin başarısı en iyi olan yöntemin DGDOM olduğu gözlenmiştir. PTE nedeni ile eks olan cevap değişkeni ile modellenen yapıların performans değerleri incelendiğinde en iyi tahminin GAMAR yöntemi ile elde edildiği görülmüştür. KOAH, pnömoni ve PTE nedeni ile eks olan hasta sayısının PM10 ana etkisi ve gecikmeleri üzerinden tahmin edilen OR grafik sonuçları incelendiğinde (Şekil 4.3); astım eks bağımlı değişkenine benzer sonuçlar gözlenmiştir. DGDOM yönteminin OR grafik güven aralıklarının geniş olduğu GAM ve GAMAR yöntemlerinin ise daha isabetli tahminler yaptıkları gözlenmiştir.

Bağımlı değişkenlerin tümü dikkate alındığında (eks bağımlı değişkenleri hariç) tahmin performansları bakımından en iyi tahmini GAMAR modelinin yaptığı görülmüştür. Ayrıca GAMAR ile tahmin edilen değişkenlerin Pearson artıkları ile çizilen otokorelasyon ve kısmi otokorelasyon fonksiyonu grafiklerinde GAMAR'ın GAM dan daha üstün olduğu görülmüştür. Çalışmamıza benzer sonuçlar, otokorelasyon ile ilgili problemle baş etmek için 2012 yılında Yang. ile arkadaşları tarafından önerilen GAMAR modeli kullanılan çalışmada görülmüştür (12). Çalışmalarında 2001 yılı ile 2004 yılları

arasında Şangay’da günlük sıcaklık değişimlerinin ölümler üzerindeki etkisini GAM ve GAMAR modellerini kullanarak tahmin etmişlerdir. Yaptıkları simülasyon ve gerçek veri denemelerinde bağımlı değişkenin otokorelasyona sahip olduğu durumlarda GAMAR modelleri GAM modellerine kıyasla daha iyi performans göstermiş ve Pearson artıkları ile çizilen otokorelasyon ve kısmi otokorelasyon fonksiyonu grafiklerinde GAMAR yönteminin görüntülerinin daha başarılı olduğunu gözlemlemişlerdir.

Çalışmamıza benzer diğer bir sonuç ise İtalya’nın Milan kentinde hava kirliliğine bağlı ölümler üzerine yapılan geniş çapta bir araştırmada da gözlemlenmiştir. Araştırmada ölümlerin aynı gün partikül maddeye maruz kalınma sebebiyle gerçekleştiği gibi birkaç gün öncesinden maruz kalınan partikül madde ile günlük ölümler arasında güçlü bir ilişki tespit edilmiştir. Tüm ölümlerde bu tip karışık etkilerin sebebinin maruziyet ve sonuç arasında dağıtılmış bir gecikme ilişkisinin doğasından kaynaklandığı öngörülmüştür (110). Yapılan çalışmalarda araştırmacılar GAM ve Genelleştirilmiş lineer modelleri kullanırken değişkenlerin gecikmelerini ayrı ayrı bir değişken gibi modele ekledikleri ve GAM model sonuçlarının tahmin başarısının üstün olduğunu gözlemlemişlerdir (30, 111, 112)

### **Model performansları**

Bağımlı değişkenlerin GAM, DGDOM ve GAMAR yöntemleri ile modellendiği bölüm bulguları incelendiğinde her hastalık için denenen dörtlü model yapılarında genelde IV. modellerin performans bakımından üstün oldukları gözlemlendi. İstisna olarak astım eks ve pnömoni eks vakalarının günlük sayımı olan bağımlı değişkenlerde serbestlik derecesi 6 olan model III performans açısından diğer modellerden üstün olduğu gözlemlenmiştir.

Cevap değişkenleri ile yapılan model denemelerinde serbestlik derecesi arttıkça model tahmin sonuçları olumlu şekilde etkilemiştir. Fakat gereğinden yüksek serbestlik derecesi tahmin yanlılığına sebep olduğu gözlemlenmiştir. Bu çalışma ile aynı sonuçları vermediği gözlenen Wood’un Genelleştirilmiş eklemeli modeller kitabında değişkenleri düzgünleştirmek için kullandığı ceza düzgünleştirme fonksiyonlarında serbestlik derecesinin bir eksiği olan ve düğüm olarak bilinen k sayılarını belirlerken 50,100 ve 200 gibi değerler atamayı tercih etmiştir. Bu atadığı değerlerle modellerin tahmin başarılarının arttığını savunmuştur (30).

Çalışmamızda zaman bir düzgünleştirme fonksiyonu yardımıyla modele eklenmiş ve yıllık mevsim sayısı üzerinden serbestlik derecesi seçildiğinde bu serbestlik

derecesinin gereğinden yüksek olduğu saçılım grafiklerinde görülmüştür ve yüksek serbestlik derecesinin yanlış tahmine sebep olduğu gözlenmiştir. Bu çalışmaya benzer sonuçlar veren bir araştırmada, zamana bağlı hasta sayılarının tahminini yapan araştırmacılar zamanı düzgünleştiren fonksiyonda düşük serbestlik derecesi kullanmanın yanlış tahmin yapma ihtimalini azalttığını öngörmüşlerdir (67). Çalışmamıza benzer sonuçlar veren Ma ve arkadaşlarının yaptığı çalışmada serbestlik derecesini mevsim sayısı üzerinden seçtiklerinde düzeltilmiş cevap değişkeni ile zaman arasındaki saçılım grafiklerindeki görüntünün mevsimsel yapıya uygun olmadığını görmüşlerdir (66). Benzer sonuçlar Peng ve arkadaşlarının yaptığı çalışmada gözlenmiştir. Peng ve arkadaşları zamanı düzgünleştirme fonksiyonu yardımıyla modele alıp serbestlik derecesi belirlemek için düşük AIC yöntemini kullanmışlar ve yüksek değerlere sahip serbestlik dereceli düzgünleştirme fonksiyonlarının model tahmin yanlılığına neden olacağını bildirmişlerdir (102).

Yukarıda bahsi geçen çalışmalar çevre ve mevsimsel etkinin hastalıklarla olan ilişkisini değişik teknikler ve yaklaşımlarla incelemişlerdir. İncelemeler sonucunda bu yöntemler arasında performans bakımından farklılıkların olduğu hipotezi doğrulanmıştır. Ayrıca daha önce yapılmış çalışmalarda GAM, DGDOM ve GAMAR yöntemlerinin bir karşılaştırmasına rastlanmamıştır. Araştırmamız bu yönüyle literatüre önemli katkılar sağlamaktadır.

## 6. SONUÇ VE ÖNERİLER

Sonuç olarak; astım, KOAH, pnömoni ve PTE nedeni ile hastaneye başvuran serviste yatan ve eks olan sayılarının üzerine çevresel faktörlerin ve mevsimsellik etkisinin değerlendirilmesinde GAM, DGDOM ve GAMAR yöntemleri performansları ve OR %95 GA tahmin başarıları bakımından karşılaştırılmıştır. GAMAR yönteminin hem performans kriterleri bakımından hem de düşük standart hataya sahip OR tahminleri bakımından diğer yöntemlere göre üstünlük sağladığı ortaya konmuştur. Burada GAMAR modelinin başarılı olmasında önemli bir etken; cevap değişkeninin ardışık değerleri arasındaki bağımlılık yapısına getirdiği çözümdür (12). Bu sonuçta dikkat çekici bir diğer husus, DGDOM yönteminin model performansı bakımından diğer yöntemlere yakın sonuçlar elde etmesine rağmen OR tahmininde ortaya çıkan zayıf performansıdır. Tahmin performanslarını etkileyen bir diğer konu ise doğru serbestlik derecesinin tayinidir ve serbestlik derecesi seçilmesi konusunda birçok farklı görüş mevcuttur (102, 66, 97). Etkili serbestlik derecesi gözlenen zaman serisi mevsim sayısı, kullanılan tahmin yöntemi ve düzgünleştirme fonksiyonu çeşidi gibi birçok kritere bağlıdır (97). Çalışmamızda her bir yöntem için farklı serbestlik derecelerine sahip modeller denenerek en iyi serbestlik derecesine karar verilmiştir. Birkaç istisna haricinde serbestlik derecesinin 12 olarak atanmış ve bu değer yanlılık ve tahmin başarısı arasındaki dengeyi kurma niteliği sağlayan bir değer olduğu gözlenmiştir. Çalışmamızdaki bulgulardan yola çıkarak çevre epidemiyolojisinde okuyuculara araştırılması için önerilecek yöntemler; ceza zincirleri yöntemlerini içinde barındıran GAMAR yöntemi, Otoregresif hareketli ortalamalar modeli (ARIMA) ve Sezonsal otoregresif hareketli ortalamalar modelidir. Ayrıca bu modellerin GAMAR ile karşılaştırılmasının incelendiği bir çalışma araştırmacılara önerilebilir.



## KAYNAKLAR

1. Özşahin E, Erođlu İ, Pektezel H. Keşan'da (Edirne) Hava Kirliliđi. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi* 2016 (36): 83-100.
2. HEAL, Air Pollution and Health in Turkey Facts, [https://www.env-health.org/IMG/pdf/150220\\_factsheet\\_air\\_and\\_health\\_turkey\\_en\\_final.pdf](https://www.env-health.org/IMG/pdf/150220_factsheet_air_and_health_turkey_en_final.pdf) (1 Şubat 2018).
3. WHO, Air Pollution, <http://www.who.int/airpollution/en/> (1 Şubat 2018).
4. Bođan M. İklim deđişiklikleri, çöl tozları ve hava kirliliđinin insan sađlıđı üzerindeki etkilerinin araştırılması. Tıp Fakültesi, Acil Tıp Anabilim Dalı. Tıpta Uzmanlık Tezi, Gaziantep: Gaziantep Üniversitesi 2016.
5. Türkeş M. Türkiye'de gözlenen ve öngörölen iklim deđişikliđi, kuraklık ve çölleşme. *Ankara Üniversitesi Çevre Bilimleri Dergisi* 2012, 4(2): 1-32.
6. Bayram H, Bogan M, Kul S, Oktay MM, Akpınar-Elci M, Al B. Effects Of Desert Dust Storms And Meteorological Variables On Emergency Room Visits And Hospitalization Due To COPD In South East Turkey. *Am J Resp Crit Care Med* 2015, A6167-A.
7. Bogan M, Kul S, Oktay M, Akpınar-Elci M, Al B, Bayram H. Effects Of Desert Dust Storms And Meteorological Variables On Emergency Room Visits And Hospitalization Due To COPD In South East Turkey. *Am J Respir Crit Care Med* 2015, 191: A6167.
8. Xia Y, Tong H. Cumulative effects of air pollution on public health. *Stat Med* 2006, 25(20): 3548-59.
9. Anlar O, Tombul T, Unal O, Kayan M. Seasonal and environmental temperature variation in the occurrence of ischemic strokes and intracerebral hemorrhages in a Turkish adult population. *Int J Neurosci* 2002, 112(8): 959-63.
10. Hastie TJ, Tibshirani RJ. *Generalized Additive Models*, Taylor & Francis 1990.
11. Gasparrini A. Distributed lag linear and non-linear models in R: the package dlnm. *J Stat Softw* 2011, 43(8): 1.
12. Yang L, Qin G, Zhao N, Wang C, Song G. Using a generalized additive model with autoregressive terms to study the effects of daily temperature on mortality. *BMC Med Res Methodol* 2012, 12(1): 165.

13. Galton F. Regression towards mediocrity in hereditary stature. *J R Anthropol Inst* 1886, 15: 246-63.
14. Allen MP. *Understanding Regression Analysis*, Springer US 2007.
15. Forthofer RN, Lee ES, Hernandez M. *Biostatistics: A Guide to Design, Analysis and Discovery*, Elsevier Science 2006.
16. Rosner B. *Fundamentals of Biostatistics*, Cengage Learning 2010.
17. Çelik MY. *Nasıl?: biyoistatistik bilimsel araştırma SPSS : yöntemlerin uygulamalı ve görsel anlatımı, araştırma projesi ve bilimsel dergilere (SCI-SCI Expanded...) yayın tasarlamak yazmak ve sunmak: tüm bilim dalları için çok değişkenli veri çözümleme yöntemleri*, M. Yusuf Çelik 2011.
18. Campbell MJ, Swinscow TDV. *Statistics at Square One*, Wiley 2009.
19. Campbell MJ. *Statistics at Square Two: Understanding Modern Statistical Applications in Medicine*, Wiley 2013.
20. Alpar R. *Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş-1*, Nobel Yayın Dağıtım 2003.
21. Hosmer DW, Lemeshow S. *Applied Logistic Regression*, Wiley 2004.
22. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*, Springer New York 2012.
23. Dobson AJ. *An Introduction to Generalized Linear Models*, Second Edition, Taylor & Francis 2010.
24. Montgomery DC, Peck EA, Vining GG. *Introduction to Linear Regression Analysis*, Wiley 2012.
25. McCullagh P, Nelder JA. *Generalized Linear Models*, Second Edition, Taylor & Francis 1989.
26. Cameron AC, Trivedi PK. *Regression Analysis of Count Data*, Cambridge University Press 1998.
27. Liu H. *Generalized additive model*, Department of Mathematics and Statistics, University of Minnesota Duluth, Duluth, MN 2008.
28. Zuur A, Ieno EN, Smith GM. *Analyzing Ecological Data*, Springer New York 2007.
29. Austin PC. A comparison of regression trees, logistic regression, generalized additive models, and multivariate adaptive regression splines for predicting AMI mortality. *Stat Med* 2007, 26(15): 2937-57.

30. Wood S. *Generalized Additive Models: An Introduction with R*, Taylor & Francis 2006.
31. Guisan A, Edwards TC, Hastie T. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecol Modell* 2002, 157(2): 89-100.
32. Yee TW, Mitchell ND. Generalized additive models in plant ecology. *J Veg Sci* 1991, 2(5): 587-602.
33. Higdon R. Generalized Additive Models. In: Dubitzky W, Wolkenhauer O, Cho K-H, Yokota H (eds). *Encyclopedia of Systems Biology*, New York, Springer 2013: 814-5.
34. Schwartz J. Air pollution and daily mortality in Birmingham, Alabama. *Am J Epidemiol* 1993, 137(10): 1136-47.
35. He S. Generalized additive models for data with concurvity: statistical issues and a novel model fitting approach. Graduate School of Public Health. Doctoral Dissertation, Pittsburgh, PA: University of Pittsburgh 2004.
36. Hyndman R, Koehler AB, Ord JK, Snyder RD. *Forecasting with Exponential Smoothing: The State Space Approach*, Springer Berlin Heidelberg 2008.
37. Härdle W. *Applied Nonparametric Regression*, Cambridge University Press 1990.
38. Bühlmann P, Van de Geer S. *Statistics for High-Dimensional Data: Methods, Theory and Applications*, Springer Berlin Heidelberg 2011.
39. Wood SN. *Generalized Additive Models: An Introduction with R*, Second Edition, CRC Press 2017.
40. Lin J. Topics in Application of Nonparametric Smoothing Splines. Master's Thesis, North Carolina: North Carolina State University 2005.
41. Green PJ, Silverman BW. *Nonparametric Regression and Generalized Linear Models: A roughness penalty approach*: Taylor & Francis 1993.
42. Berk RA. *Statistical Learning from a Regression Perspective*, Springer International Publishing 2016.
43. Wahba G, Wendelberger J. Some new mathematical methods for variational objective analysis using splines and cross validation. *Mon Weather Rev* 1980, 108(8): 1122-43.
44. Schoenberg IJ. Spline functions and the problem of graduation. *Proc Natl Acad Sci U S A* 1964, 52(4): 947-50.
45. Wood SN. Stable and efficient multiple smoothing parameter estimation for generalized additive models. *J Am Stat Assoc* 2004, 99(467): 673-86.

46. Wahba G. *Spline Models for Observational Data*, Society for Industrial and Applied Mathematics 1990.
47. Liu H. Generalized additive model, Duluth - MN, University of Minnesota Duluth 2008.
48. Hastie T, Tibshirani R. Generalized Additive Models, Cubic Splines and Penalized Likelihood, Stanford Univ Ca Dept Of Statistics 1987.
49. Wahba G. Spline interpolation and smoothing on the sphere. *SIAM J Sci Comput* 1981, 2(1): 5-16.
50. Reinsch CH. Smoothing by spline functions. *Numer Math (Heidelb)* 1967, 10(3): 177-83.
51. Hastie T, Tibshirani R. Generalized additive models: some applications. *J Am Stat Assoc* 1987, 82(398): 371-86.
52. Eilers PH, Marx BD. Flexible smoothing with B-splines and penalties. *Statist Sci* 1996, 11(2): 89-102.
53. Wood SN. P-splines with derivative based penalties and tensor product smoothing of unevenly distributed data. *Stat Comput* 2017, 27(4): 985-9.
54. Eilers PH, Marx BD, Durbán M. Twenty years of P-splines. *Sort (Barc)* 2015, 39(2): 149-86.
55. Wood SN. Thin plate regression splines. *J R Stat Soc Series B Stat Methodol* 2003, 65(1): 95-114.
56. Duchon J. *Splines minimizing rotation-invariant semi-norms in Sobolev spaces. Constructive Theory of Functions of Several Variables*, Springer, Berlin 1977.
57. Wendelberger JG. *The Computation of Laplacian Smoothing Splines with Examples*, Madison Dept Of Statistics. Technical Report, Wisconsin: Wisconsin University 1981.
58. Wood SN, Pya N, Säfken B. Smoothing parameter and model selection for general smooth models. *J Am Stat Assoc* 2016, 111(516): 1548-63.
59. Xiang D (ed) *Fitting generalized additive models with the GAM procedure*, SUGI Proceedings, SAS Institute, Inc Cary, NC 2001.
60. Wahba G. A Comparison of GCV and GML for Choosing the Smoothing Parameter in the Generalized Spline Smoothing Problem. *Ann Stat* 1985, 13(4): 1378-402.
61. Craven P, Wahba G. Smoothing noisy data with spline functions. *Numer Math (Heidelb)* 1978, 31(4): 377-403.

62. Koehn S. Generalized additive models in the context of shipping economics. Department of Economics. PhD Thesis, Leicester: University of Leicester 2008.
63. Kim YJ, Gu C. Smoothing spline Gaussian regression: more scalable computation via efficient approximation. *J R Stat Soc Series B Stat Methodol* 2004, 66(2): 337-56.
64. Wood S. Package ‘mgcv’, <https://cran.r-project.org/web/packages/mgcv/mgcv.pdf> (1 Şubat 2018).
65. Rodriguez G. Smoothing and non-parametric regression 2001.
66. Ma W, Sun X, Song Y, Tao F, Feng W, He Y, Zhao N, Yuan Z. Applied mixed generalized additive model to assess the effect of temperature on the incidence of bacillary dysentery and its forecast. *PloS One* 2013, 8(4): e62122.
67. Ma W-p, Gu S, Wang Y, Zhang X-j, Wang A-r, Zhao N-q, Song Y-y. The use of mixed generalized additive modeling to assess the effect of temperature on the usage of emergency electrocardiography examination among the elderly in Shanghai. *PloS One* 2014, 9(6): e100284.
68. Gasparrini A, Scheipl F, Armstrong B, Kenward MG. A penalized framework for distributed lag non- linear models. *Biometrics* 2017, 73(3): 938-948.
69. Gasparrini A. Package ‘dlnm’, <https://cran.r-project.org/web/packages/dlnm/dlnm.pdf> (12 Şubat 2018).
70. Zanobetti A, Schwartz J, Samoli E, Gryparis A, Touloumi G, Atkinson R, Le Tertre A, Bobros J, Celko M, Goren A, Forsberg B, Michelozzi P, Rabczenko D, Aranguiz Ruiz E, Katsouyanni K. The temporal pattern of mortality responses to air pollution: a multicity assessment of mortality displacement. *Epidemiology* 2002, 13(1): 87-93.
71. Braga ALF, Zanobetti A, Schwartz J. The time course of weather-related deaths. *Epidemiology* 2001, 12(6): 662-7.
72. Gasparrini A, Armstrong B, Kenward MG. Distributed lag non- linear models. *Stat Med* 2010, 29(21): 2224-34.
73. Dominici F, Sheppard L, Clyde M. Health effects of air pollution: a statistical review. *Int Stat Rev* 2003, 71(2): 243-76.
74. Touloumi G, Atkinson R, Tertre AL, Samoli E, Schwartz J, Schindler C, et al. Analysis of health outcome time series data in epidemiological studies. *Environmetrics* 2004, 15(2): 101-17.
75. Schwartz J. The distributed lag between air pollution and daily deaths. *Epidemiology* 2000, 11(3): 320-6.

76. Zanobetti A, Wand M, Schwartz J, Ryan L. Generalized additive distributed lag models: quantifying mortality displacement. *Biostatistics* 2000, 1(3): 279-92.
77. McCullagh P. Generalized linear models. *Eur J Oper Res* 1984, 16(3): 285-92.
78. Dobson AJ, Barnett A. *An introduction to generalized linear models*, CRC Press 2008.
79. Bhaskaran K, Gasparrini A, Hajat S, Smeeth L, Armstrong B. Time series regression studies in environmental epidemiology. *Int J Epidemiol* 2013, 42(4): 1187-95.
80. Daniels MJ, Dominici F, Samet JM, Zeger SL. Estimating particulate matter-mortality dose-response curves and threshold levels: an analysis of daily time-series for the 20 largest US cities. *Am J Epidemiol* 2000, 152(5): 397-406.
81. Roberts S, Martin MA. A distributed lag approach to fitting non-linear dose-response models in particulate matter air pollution time series investigations. *Environ Res* 2007, 104(2): 193-200.
82. Eilers PH, Currie ID, Durbán M. Fast and compact smoothing on large multidimensional grids. *Comput Stat Data Anal* 2006, 50(1): 61-76.
83. Eilers PH, Gampe J, Marx BD, Rau R. Modulation models for seasonal time series and incidence tables. *Stat Med* 2008, 27(17): 3430-41.
84. Adhikari R, Agrawal R. *An introductory study on time series modeling and forecasting*. LAP Lambert Academic Publishing, Germany 2013.
85. Erdoğan T. Performance of time series decomposition for different type of time series. Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı. Yüksek Lisans Tezi, İzmir: Dokuz Eylül Üniversitesi 2017.
86. Box G, Jenkins G. *Times series Analysis Forecasting and Control*, Holden-Day San Francisco 1970.
87. Hipel KW, McLeod AI. *Time series modelling of water resources and environmental systems*, Elsevier 1994.
88. Hanedar A, Akkaya O, Bizim Ç. *Durağanlık Analizi, Birim Kök Testleri ve Trend*, Akademik Proje 2014.
89. Wei WWS. *Time Series Analysis: Univariate and Multivariate Methods*, Pearson Addison Wesley 2006.
90. Sandy R. *Statistics for Business and Economics*, Random House Inc 1989.
91. Montgomery DC, Johnson LA, Gardiner JS. *Forecasting and time series analysis*, McGraw-Hill 1990.
92. Cryer JD, Chan KS. *Time Series Analysis: With Applications in R*, Springer New York 2008.

93. Benjamin MA, Rigby RA, Stasinopoulos MD, (eds). Fitting non-Gaussian time series models, COMPSTAT, Springer 1998.
94. Pillai TR (ed). *Generalized autoregressive moving average models: an application to GDP in Malaysia*. Third Malaysia Statistics Conference—MYSTATS 2015.
95. Benjamin MA, Rigby RA, Stasinopoulos DM. Generalized autoregressive moving average models. *J Am Stat Assoc* 2003, 98(461): 214-23.
96. Yang C-Y, Cheng M-H, Chen C-C. Effects of Asian dust storm events on hospital admissions for congestive heart failure in Taipei, Taiwan. *J Toxicol Environ Health A* 2009, 72(5): 324-8.
97. Touloumi G, Samoli E, Pipikou M, Le Tertre A, Atkinson R, Katsouyanni K. Seasonal confounding in air pollution and health time-series studies: Effect on air pollution effect estimates. *Stat Med* 2006, 25(24): 4164-78.
98. Schwartz J, Spix C, Touloumi G, Bachárová L, Barumamdzadeh T, le Tertre A, Piekarksi T, Ponce de Leon A, Pönkä A, Rossi G, Saez M, Schouten JP. Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions. *J Epidemiol Community Health* 1996, 50(Suppl 1): 3-11.
99. Dominici F, McDermott A, Hastie TJ. Improved semiparametric time series models of air pollution and mortality. *J Am Stat Assoc* 2004, 99(468): 938-48.
100. Wang Q, Gao C, Wang H, Lang L, Yue T, Lin H. Ischemic stroke hospital admission associated with ambient temperature in Jinan, China. *PloS One* 2013, 8(11): e80381.
101. Lin H, Zou H, Wang Q, Liu C, Lang L, Hou X, et al. Short-term effect of El Nino-Southern Oscillation on pediatric hand, foot and mouth disease in Shenzhen, China. *PLoS One* 2013, 8(7): e65585.
102. Peng RD, Dominici F, Louis TA. Model choice in time series studies of air pollution and mortality. *J R Stat Soc Ser A Stat Soc* 2006, 169(2): 179-203.
103. Koç T, Cengiz MA. Genelleştirilmiş Lineer Karma Modellerde Tahmin Yöntemlerinin Uygulamalı Karşılaştırılması. *Karaelmas Science & Engineering Journal* 2012, 2(2): 47-52.
104. Gasparrini A, Armstrong B. Time series analysis on the health effects of temperature: advancements and limitations. *Environmental research* 2010, 110(6): 633-8.
105. Aho K, Derryberry D, Peterson T. Model selection for ecologists: the worldviews of AIC and BIC. *Ecology* 2014, 95(3): 631-6.
106. Cavanaugh JE. Unifying the derivations for the Akaike and corrected Akaike information criteria. *Stat Probab Lett* 1997, 33(2): 201-8.

107. Stack Exchange Network, How I can interpret GAM results? <https://stats.stackexchange.com/questions/190172/how-i-can-interpret-gam-results> (3 Nisan 2018)
108. Sheiner LB, Beal SL. Some suggestions for measuring predictive performance. *J Pharmacokinet Biopharm* 1981, 9(4): 503-12.
109. Wood S.N. Approximate hypothesis tests related to GAM fits, <https://stat.ethz.ch/R-manual/R-devel/library/mgcv/html/anova.gam.html> (2 Mart 2018).
110. Rossi G, Vigotti MA, Zanobetti A, Repetto F, Gianelle V, Schwartz J. Air pollution and cause-specific mortality in Milan, Italy, 1980–1989. *Arch Environ Health* 1999, 54(3): 158-64.
111. Dominici F, McDermott A, Zeger SL, Samet JM. On the use of generalized additive models in time-series studies of air pollution and health. *Am J Epidemiol* 2002, 156(3): 193-203.
112. Bayram H, Bogan M, Kul S, Oktay M, Akpınar-Elci M, Al B, (eds). Effects Of Desert Dust Storms And Meteorological Variables On Emergency Room Visits And Hospitalization Due To COPD In South East Turkey. *Am J Respir Crit Care Med* 2015.



## EKLER

### EK.1. ÖZGEÇMİŞ

**Adı Soyadı** : Mehmet Karadağ  
**e-posta** : [mkarad@gmail.com](mailto:mkarad@gmail.com)  
**Doğum Tarihi** : 05.10.1981  
**Unvanı** : Biyoistatistik Uzmanı  
**Öğrenim Durumu** : Doktora Devam

Derece	Alan	Üniversite	Yıl
Lisans	İstatistik	Ondokuz Mayıs Üniversitesi	1999-2003
Yüksek Lisans	Biyoistatistik ve Tıp Bilişimi	Trakya Üniversitesi	2011-2014
Doktora	Biyoistatistik ve Tıp Bilişimi	İnönü Üniversitesi	2014 -Devam

#### **Yüksek Lisans Tez Başlığı ve Tez Danışmanı:**

Karar ağaçları ile lojistik regresyon analizinin performanslarının simülasyon çalışması ile karşılaştırılması, (Danışman: Prof. Dr. Necdet Süt)

#### **Görevler:**

Görev Unvanı	Görev Yeri	Yıl
İş Akış Yönetmeni	T. Garanti Bankası A.Ş. Erzincan Şb.	2007-2015

#### **Projelerde Yaptığı Görevler:**

Karar ağaçları ile lojistik regresyon analizinin performanslarının simülasyon çalışması ile karşılaştırılması (BAP), Araştırmacı, 06.06.2012 – 05.08.2014

#### **Eserler**

#### **i. Uluslararası hakemli dergilerde yayımlanan makaleler (SCI & SSCI & Arts and Humanities)**

1. Uysal E, Dokur M, Altınay S, Saygılı I, Batcıoğlu K, Ceylan M. S, Kazımoğlu H, Uyumlu A.B, **Karadağ M.** Investigation of the Effect of Milrinone on Renal

Damage in an Experimental Non-Heart Beating Donor Model. J Invest Surg. 2017. DOI:10.1080/08941939.2017.1343880.

2. Dokur M, Petekkaya E, **Karadağ M**. Analysis of selfie-related injuries and deaths. Turkish Journal of Trauma and Emergency Surgery. DOI:10.5505/tjtes.2017.83103.

## ii. Uluslararası diğer hakemli dergilerde yayınlanan makaleler

1. Uysal E, Dokur M, Kırdak T, Kurt A, **Karadağ M**. Evaluation of the effect of adipose-derived mesenchymal stem cell on intraperitoneal adhesions. Turkish Journal of Surgery. DOI: 10.5152/turkjsurg.2017.3860.

## iii. Uluslararası bilimsel toplantılarda sunulan ve bildiri kitabında (Proceeding) basılan bildiriler

1. **Karadağ M**, Yoloğlu S, Özturacı RO. Karar Ağaçlarından CART ve J48'in Performanslarının Karşılaştırılması: Bir Simülasyon Çalışması. 17. Uluslararası Ekonometri, Yöneylem Araştırması ve İstatistik Sempozyumu (EYİ-2016),2-4 Haziran 2016, Sivas.
2. **Karadağ M**, Kul S, Yoloğlu S, Karar Ağaçlarından CHAID ve C4.5'in Performanslarının Karşılaştırılması: Bir Simülasyon Çalışması. 18. Ulusal Biyoistatistik Kongresi ve 1. Uluslararası Biyoistatistik Kongresi, 26 – 29 Ekim 2016, Belek Antalya.
3. Seval Kul, **Mehmet K**, Saim Yoloğlu, Mustafa Boğan. Genelleştirilmiş Additive Poisson Regresyon Modeli Yardımıyla Çevre kirliliği ve Meteorolojik Değişkenlerin Hastane Başvuruları ile İlişkilendirilmesi: Astım Nedenli Acil Başvurularının Modellenmesi. 19. Ulusal Biyoistatistik Kongresi ve 2. Uluslararası Biyoistatistik Kongresi, 25 – 28 Ekim 2017, Belek Antalya.

## Diğer Bilimsel Etkinlikler:

### Mesleki Sertifikasyonlar:

1. Temel iyi klinik uygulamalar kursu (Klinik Araştırmalar Derneği, Online eğitim) (21.04.2016)
2. İleri iyi klinik uygulamalar kursu (Klinik Araştırmalar Derneği, Online eğitim) (21.04.2016)
3. Yönetici eğitimi (01.12.2015)
4. İnsan kaynakları eğitimi (30.10.2016)

**Verdiği Kurslar:**

1. “SPSS Uygulamalı Temel İstatistik Kursu” Harran Üniversitesi Tıp Fakültesi (26/27 Mayıs 2017)

**Hakemlik Yaptığı Dergiler**

- Sağlık Akademisi Kastamonu Dergisi

## EK.2. ETİK KURUL ONAY FORMU

T.C. İNÖNÜ ÜNİVERSİTESİ BİLİMSEL ARAŞTIRMA VE YAYIN ETİĞİ KURULU (Sağlık Bilimleri Girişimsel Olmayan Klinik Araştırmalar Etik Kurulu)			
Oturum Tarihi	Oturum Sayısı	Karar Sayısı	
16.01.2018	2	2018/2-8	
<p><b>Karar No: 2018/2-8:</b> Sağlık Bilimleri Girişimsel Olmayan Klinik Araştırmalar Etik Kurulu 16.01.2018 tarihinde Tıp Fakültesi Etik Kurul Salonunda toplandı. İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi AD'nin Prof. Dr. Saim YOLOĞLU'nun sorumlu araştırmacı olduğu; Doktora Öğrencisi Mehmet KARADAĞ; Gaziantep Üniversitesi Tıp Fakültesi Biyoistatistik AD'nin Doç. Dr. Seval KUL'ın; Gaziantep Üniversitesi Tıp Fakültesi Acil Tıp AD'nin Prof. Dr. Behçet AL'ın; Gaziantep Şehitkamil Devlet Hastanesi Uzman Dr. Mustafa BOĞAN'ın; Koç Üniversitesi Göğüs Hastanesi Prof. Dr. Hasan BAYRAM'ın "<b>Çevre Epidemiyolojisinde Hastahkların Modellenmesi İçin Zaman Serisi Regresyon Analizi Yöntemlerinin Karşılaştırılması</b>" başlıklı çalışması Üniversitemiz Bilimsel Araştırma ve Yayın Etiği Yönergesi açısından uygun olup-olmadığı hususundaki başvurusuna ilişkin raportör raporu görüşüldü. Çalışma Bilimsel Araştırma ve Yayın Etiği Yönergesi açısından değerlendirildiğinde; çalışmanın <b>etik açıdan uygun olduğuna; oy birliği ile karar verilmiştir.</b></p>			
Prof. Dr. Osman CELBİŞ Etik Kurul Başkanı 			
Prof. Dr. Kadir ERTEM Etik Kurul Başkan Yrd.	KATILDI	Prof. Dr. Gülsen GÜNEŞ Etik Kurul Üyesi	KATILDI
Prof. Dr. Cemşit KARAKURT Etik Kurul Üyesi	KATILDI	Prof. Dr. Yüksel SEÇKİN Etik Kurul Üyesi	KATILDI
Prof. Dr. Erkan KARATAŞ Etik Kurul Üyesi	KATILDI	Prof. Dr. Yılmaz TABEL Etik Kurul Üyesi	KATILDI