

**GENELLEŐTİRİLMİŐ DOĐRUSAL MEKÂNSAL
MODELLERE KOŐULLU OTOREGRESİF MODEL
YAKLAŐIMI**

**CONDITIONAL AUTOREGRESSIVE MODEL APPROACH
TO GENERALIZED LINEAR SPATIAL MODELS**

LEYLA BAKACAK

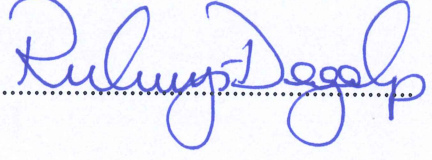
PROF. DR. SERPİL AKTAŐ ALTUNAY
Tez Danıőmanı

Hacettepe Üniversitesi
Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin
İstatistik Anabilim Dalı için Öngördüğü
YÜKSEK LİSANS TEZİ olarak hazırlanmıştır.

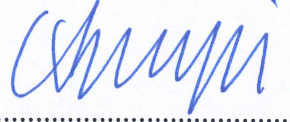
2018

LEYLA BAKACAK' in hazırladığı “Genelleştirilmiş Doğrusal Mekânsal Modellere Koşullu Otoregresif Model Yaklaşımı” adlı bu çalışma aşağıdaki jüri tarafından İSTATİSTİK ANABİLİM DALI' nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

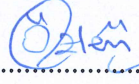
Doç. Dr. Rukiye DAĞALP
Başkan



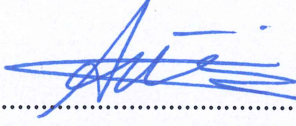
Prof. Dr. Serpil AKTAŞ ALTUNAY
Danışman



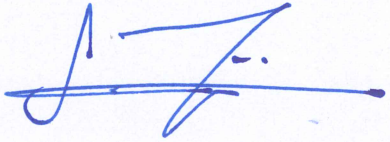
Doç. Dr. Özlem İLK DAĞ
Üye



Doç. Dr. Ayten YİĞİTER
Üye



Doç. Dr. Semra TÜRKAN
Üye



Bu tez Hacettepe Üniversitesi Fen Bilimleri Enstitüsü tarafından YÜKSEK LİSANS TEZİ olarak onaylanmıştır.

Prof. Dr. Menemşe GÜMÜŞDERELİOĞLU
Fen Bilimleri Enstitüsü Müdürü

YAYINLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanması zorunlu metinlerin yazılı izin alarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

- ☐ Tezimin/Raporumun tamamı dünya çapında erişime açılabilir ve bir kısmı veya tamamının fotokopisi alınabilir.

(Bu seçenekle teziniz arama motorlarında indekslenebilecek, daha sonra tezinizin erişim statüsünün değiştirilmesini talep etseniz ve kütüphane bu talebinizi yerine getirirse bile, tezinin arama motorlarının önbelleklerinde kalmaya devam edebilecektir.)

- ☒ Tezimin/Raporumun 08/06/2020 tarihine kadar erişime açılmasını ve fotokopi alınmasını (İç Kapak, Özet, İçindekiler ve Kaynakça hariç) istemiyorum.

(Bu sürenin sonunda uzatma için başvuruda bulunmadığım taktirde, tezimin/raporumun tamamı her yerden erişime açılabilir, kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisi alınabilir)

- ☐ Tezimin/Raporumun tarihine kadar erişime açılmasını istemiyorum, ancak kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisinin alınmasını onaylıyorum.

- ☐ Serbest Seçenek/Yazarın Seçimi

08 / 06 / 2018



Leyla BAKACAK

ETİK

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada,

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

08/06/2018

LEYLA BAKACAK

ÖZET

GENELLEŞTİRİLMİŞ DOĞRUSAL MEKÂNSAL MODELLERE KOŞULLU OTOREGRESİF MODEL YAKLAŞIMI

Leyla BAKACAK

Yüksek Lisans, İstatistik Bölümü

Tez Danışmanı: Prof. Dr. Serpil AKTAŞ ALTUNAY

Haziran 2018, 53 sayfa

Mekânsal analiz, coğrafi bölge üzerindeki mekânsal örüntüleri tanımlamak için kullanılan bir yöntemdir. Bu analiz sırasında mekânsal veriyi oluşturan nesnelerin öznitelik ve konum bilgileri kullanılır. Yanıt değişkeni, birbirine yakın olan nesnelerin benzer özellikler sergilemesinden dolayı mekânsal otokorelasyon sergiler. Bu yüzden, modelde açıklayıcı değişkenler kullanılsa bile mekânsal ilişki yakalanamamakta ve artıklarda kalmaktadır. Artıklarda kalan bu ilişkiden dolayı doğrusal model yaklaşımlarında bağımsızlık varsayımı bozulmaktadır. Bu gibi durumlarda, mekânsal ilişkiyi barındıran rasgele etkiler modele katılır ve koşullu otoregresif önselleri ile modellenir. Mekânsal veri türlerinden olan birbiriyle örtüşmeyen alansal verilerin analizinde, R programında bulunan CAR önselleri ile Bayesci mekânsal modelleme yapan CARBayes paketi kullanılır. Bu paket, Gibbs örnekleme ve Metropolis Hastings algoritmalarını kullanan Markov zinciri Monte Carlo (MCMC) benzetimine dayanmaktadır. Bu tez çalışmasında, alansal birim olarak Türkiye'nin idari bölünüş sınırları yani iller kullanılmıştır. 2016 yılında meydana gelen deprem sayıları olarak belirlenen yanıt değişkeni, Moran'ın I permütasyon testi sonucunda mekânsal otokorelasyon sergilemiştir. Bu yüzden, deprem sayılarının modellemesinde Poisson dağılımı varsayımı altında genelleştirilmiş doğrusal modeller kullanıldığında artıklarda mekânsal bilgi kalmıştır. Bunun üzerine, rasgele etkinin modelde yer aldığı Leroux koşullu

otoregresif modeli ile deprem sayıları ve deprem büyüklüğü arasındaki ilişki incelenmiştir. Model sonucunda elde edilen kestirilmiş değerler üzerinden her ile ait risk değerleri hesaplanmış ve risk haritalaması yapılmıştır.

Anahtar Kelimeler: Mekânsal Otokorelasyon, CAR Modelleri, MCMC, Genelleştirilmiş Doğrusal Mekânsal Model.



ABSTRACT

CONDITIONAL AUTOREGRESSIVE MODEL APPROACH TO GENERALIZED LINEAR SPATIAL MODELS

Leyla BAKACAK

Master, Department of Statistics

Supervisor: Prof. Dr. Serpil AKTAŞ ALTUNAY

June 2018, 53 pages

Spatial analysis is a method used to describe spatial patterns on a geographical region. During this analysis, the attributes and location information of the objects forming the spatial data are used. Response variable exhibits spatial autocorrelation because objects that are close to each other have similar characteristics rather than the objects further apart. Thus, even if explanatory variables are used in the model, the spatial relation remains in the residuals. Because of this, the assumption of independence by linear model approaches is violated. In such cases, random effects involving spatial relations are included in the model and conditional autoregressive priors are used for these effects. In the analysis of non-overlapping spatial data, the CARBayes package in R programming is used in order to set up a Bayesian spatial model with conditional autoregressive (CAR) priors. This package is based on the Markov Chain Monte Carlo (MCMC) simulation using the Gibbs sampling and Metropolis Hastings algorithms. In this study, provinces of Turkey are used as spatial areal units. The response variable, determined as the number of earthquakes in 2016, reveals spatial autocorrelation as a result of Moran's I permutation test. Hence, when generalized linear model under the assumption of Poisson distribution is used to model the number of earthquakes, spatial autocorrelation remains in the residual. On the basis of this, the relation between the number of earthquakes and magnitude is investigated with the Leroux

conditional autoregressive model in which the random effects are included in the model. The risk values of each province are calculated and risk mapping is implemented based on the fitted values obtained from the model result.

Keywords: Spatial Autocorrelation, CAR Models, MCMC, Generalized Linear Spatial Model.



TEŞEKKÜR

Tez çalışmam boyunca destekçim ve yol göstericim olan değerli danışmanım Prof. Dr. Serpil AKTAŞ ALTUNAY'a, bilgisini ve yardımını esirgemeyen, görüş ve önerileriyle her zaman yanımda olan Öğr. Üyesi Dr. Ceren Eda CAN'a, sevgileriyle beni destekleyen AİLEM'e sonsuz teşekkürlerimi sunarım.



İÇİNDEKİLER

	<u>Sayfa</u>
ÖZET	i
ABSTRACT	iii
TEŞEKKÜR	v
İÇİNDEKİLER.....	vi
ÇİZELGELER.....	vii
ŞEKİLLER	viii
SİMGELER VE KISALTMALAR	ix
1. GİRİŞ	1
1.1. Literatür Taraması	3
2. GENELLEŞTİRİLMİŞ DOĞRUSAL MODELLER.....	6
2.1. Genelleştirilmiş Doğrusal Modellerin Bileşenleri.....	7
2.2. Genelleştirilmiş Doğrusal Karma Modeller	9
3. GENELLEŞTİRİLMİŞ DOĞRUSAL MEKÂNSAL MODELLER	12
3.1. Mekânsal Ağırlık ve Mekânsal Komşuluk	15
3.2. Mekânsal Otokorelasyonun Değerlendirilmesi	18
3.2.1. Moran’ın I İstatistiği.....	19
3.3. Koşullu Otoregresif Modeller.....	22
3.3.1. CARBayes Paketi	25
3.4. Mekânsal Haritalandırma	29
4. UYGULAMA.....	32
5. SONUÇLAR	46
KAYNAKLAR.....	48
EKLER	52
ÖZGEÇMİŞ.....	53

ÇİZELGELER

Sayfa

Çizelge 2. 1. Bayesci GDKM için önsel dağılımlar	10
Çizelge 4. 1. Tanımlayıcı istatistikler	32
Çizelge 4. 2. Poisson GDM	33
Çizelge 4. 3. Sınır paylaşma durumuna göre ikili bağlantı matrisi ($\mathbf{W}_1$)	34
Çizelge 4. 4. Poisson GDM artıkları için $\mathbf{W}_1$ matrisi komşuluğunda otokorelasyon testi ..	34
Çizelge 4. 5. Yanıt değişkeni için $\mathbf{W}_1$ matrisi komşuluğunda otokorelasyon testi	34
Çizelge 4. 6. Poisson mekânsal log-doğrusal model ($\mathbf{W}_1$)	35
Çizelge 4. 7. Poisson GDMM artıkları için otokorelasyon testi ($\mathbf{W}_1$)	38
Çizelge 4. 8. Üstel Poisson GDMM ($\mathbf{W}_1$)	38
Çizelge 4. 9. Genel standartlaştırılmış matris ($\mathbf{W}_2$)	40
Çizelge 4. 10. Yanıt değişkeni için $\mathbf{W}_2$ matrisi komşuluğunda otokorelasyon testi	41
Çizelge 4. 11. Poisson GDMM artıkları için otokorelasyon testi ($\mathbf{W}_2$)	42
Çizelge 4. 12. Poisson mekânsal log-doğrusal model ($\mathbf{W}_2$)	44
Çizelge 4. 13. Üstel Poisson GDMM ($\mathbf{W}_2$)	44

ŞEKİLLER

	<u>Sayfa</u>
Şekil 3. 1. Mekânsal otokorelasyon.....	21
Şekil 4. 1. $\beta_0(a, b)$ ve $\beta_1(c, d)$ için iz ve yoğunluk grafiği ($\mathbf{W}_1$)	36
Şekil 4. 2. Otokorelasyon parametresi ρ için iz ve yoğunluk grafiği ($\mathbf{W}_1$)	37
Şekil 4. 3. Varyans parametresi τ^2 için iz ve yoğunluk grafiği ($\mathbf{W}_1$).....	37
Şekil 4. 4. 2016 yılında meydana gelen büyüklüğü 2’den fazla olan depremlerin riski ($\mathbf{W}_1$)	39
Şekil 4. 5. Riski 1’den büyük olan depremlerin sonsal olasılıkları ($\mathbf{W}_1$).....	40
Şekil 4. 6. $\beta_0(a)$ ve $\beta_1(b)$ için iz ve yoğunluk grafiği ($\mathbf{W}_2$)	42
Şekil 4. 7. Otokorelasyon parametresi ρ için iz ve yoğunluk grafiği ($\mathbf{W}_2$)	43
Şekil 4. 8. Varyans parametresi τ^2 için iz ve yoğunluk grafiği ($\mathbf{W}_2$).....	43
Şekil 4. 9. 2016 yılında meydana gelen büyüklüğü 2’den fazla olan depremlerin riski ($\mathbf{W}_2$)	44
Şekil 4. 10. Riski 1’den büyük olan depremlerin sonsal olasılıkları ($\mathbf{W}_2$)	45

SİMGELER VE KISALTMALAR

Simgeler

ρ	Mekânsal Otokorelasyon
ϕ	Mekânsal Değişim (Rasgele Etki)
ψ	Mekânsal Bileşen
θ	Mekânsal Olmayan Değişim
O	Offset Vektörü

Kısaltmalar

GDM	Genelleştirilmiş Doğrusal Model
GDKM	Genelleştirilmiş Doğrusal Karma Model
HGDM	Hiyerarşik Genelleştirilmiş Doğrusal Model
GMMM	Genelleştirilmiş Doğrusal Mekânsal Model
GMRF	Gaussian Markov Random Field (Normal Markov Rasgele Alanı)
CAR	Conditional Autoregressive Model (Koşullu Otoregresif Model)
ICAR	Intrinsic Conditional Autoregressive Model (Temel Koşullu Otoregresif Model)
BYM	Besag-York-Mollie
MCMC	Markov Chain Monte Carlo (Markov Zinciri Monte Carlo)
MALA	Metropolis Adjusted Langevin Algorithm (Metropolis Düzeltmeli Langevin Algoritması)
INLA	Integrated Nested Laplace Approximations
DIC	Deviance Information Criteria
WAIC	Watanabe-Akaike Information Criterion
LMPL	Log Marginal Predictive Likelihood
p.d.e	Percentage Deviance Explained

1. GİRİŞ

Mekânsal veri, haritada belirli bir lokasyona sahip verilerdir. Bu tip veriler, harita detayları (feature), destek (support) ve öznitelik (attribute) olmak üzere üç parçadan oluşur. Harita detayları, verinin haritada nasıl bir şekle sahip olduğunu tanımlar. Bu bağlamda veriler, nokta, çizgi, poligon veya grid olarak tanımlanabilir. Noktasal veriler, elektrik direği, ağaç veya otobüs durağı gibi haritada yalnızca bir nokta ile ifade edilen verilerdir. Çizgisel veriler, yol veya nehir gibi birbirine bağlanmış noktalarından oluşan verilerdir. Çizgilerle çevrelenmiş kapalı bir alan oluşturan veriler ise poligon tipli veriler olarak adlandırılır ve iller, ülkeler veya göller bu tip verilere örnek olarak verilebilir. Grid olarak tanımlanan veriler ise belirli bir düzen oluşturan noktalar veya hücrelerden meydana gelmektedir. Bu tip verilere dijital resimler, hava fotoğrafları veya uydu görüntüleri örnek verilebilir.

Mekânsal verilerin diğer bir parçası destek, verilerdeki her bir özneliğin şekli, boyutu ve yönü ile ilişkilidir. En küçük destek olarak, nokta desenli verilerin boyutu sıfır, şekli ve yönü bulunmamaktadır. Çizgisel veriler, belirli bir uzunluğa ve yöne sahiptir. Alansal veriler ise belirli bir alana ve sınıra sahiptir.

Son olarak mekânsal verilerin öznitelikleri, harita detayına bağlı olarak ölçülen değerlerdir. Örneğin, her bir hava gözlem istasyonundan elde edilen ozon, partiküler madde veya sülfür dioksit gibi maddelerin ölçümleri, noktasal verilerden elde edilen öznitelik verilerini oluşturur. Bir gölün tuzluluk, derinlik veya sıcaklık gibi ölçümleri ise alansal verilere ait özniteliğe örnek olarak verilebilir.

Mekânsal alansal veriler, çokgen olarak adlandırılabilen birbiriyle örtüşmeyen belirli bir sınır içerisinde gözlenmiş veriler olarak da ifade edilebilir. Bu sınırlar, yapılan çalışmanın amacına göre bazı çalışmalarda araştırmacı tarafından belirlenir, bazı çalışmalarda ise yönetim birimi, idari sınır veya coğrafi oluşumlar kullanılarak belirlenir. Araştırmacının seçimine bağlı olarak belirlenen sınırlarda, seçimler farklı sonuçlara neden olabileceği için sınırlar hassas bir şekilde belirlenmelidir. Eğer veri toplama işlemi, örneğin iş gücü piyasası verilerinin yerel iş gücü piyasaları ile eşleştirilmesi gibi gözlem değerlerini veri ile eşleştirmek için yapılmakta ise sınır seçiminin etkisi azalmaktadır. Fakat nüfus sayımları gibi sayılabilir veri üzerine yapılan bir çalışmada idari birimler direk olarak çalışma sınırları olarak kullanılır. Bu tür verilere, tarım, ekoloji, epidemiyoloji gibi alanlarda sıkça rastlanmaktadır. Bu verilerin modellenmesinde ise ekolojik regresyon veya hastalık haritalandırılması gibi bir çok yaklaşımda bulunmak mümkündür.

Mekânsal analizde, sayısal verilerin yanı sıra alan olarak tanımlanabilen mekân bilgisinin kullanılması için mekânsal otokorelasyonun hesaplanması ve bu bilginin modele katılması gerekir. Çünkü mekânsal otokorelasyona sahip birbirine yakın alanların gözlem değerlerinin mekân bilgisinden dolayı benzer özellikler göstermesi beklenir. Burada mekânsal yakınlık kavramından yola çıkarak ve otokorelasyonun hesaplanabilmesi için komşuluk kavramı ortaya çıkmaktadır. Sürekli bir yüzeyde, tüm noktalar birbirinin komşusudur, ancak birbirine yakın noktaların ağırlıkları uzak olanlara göre daha fazladır. Mozaik yüzeylerde ise, tüm alanlar komşu gözlem grubunun üyesidir ya da değildir şeklinde sınıflandırılmaktadır. Bu sınıflandırma yapılırken de komşuluk ilişkisine eşit bir ağırlık ya da mekânsal bağımlılığa bağlı olarak farklı ağırlıklandırma da yapılabilir [1].

Mekânsal veriler, birbirine yakın olan gözlemlerin benzer değerlere sahip olmasından dolayı mekânsal otokorelasyon sergiler. Mekânsal otokorelasyon, regresyon modeline açıklayıcı değişkenlerin katılmasıyla modellenabilir fakat açıklayıcı değişkenlerin etkilerinden sonra bile artıklarda mekânsal bir yapı söz konusu olmaktadır. Artıklarda kalan mekânsal otokorelasyondan dolayı birçok regresyon modelinde bağımsızlık varsayımı bozulur. Bunun bir nedeni, mekânsal açıklayıcı değişkenlerin bilinmemesi veya ölçülememesi durumunda etki karışımı (confounding) olabilir. Açıklayıcı değişkendeki bu mekânsal yapı model tarafından açıklanamadığı için yanıt değişkeninde mekânsal otokorelasyona neden olur. Artık mekânsal otokorelasyonun diğer bir nedeni ise komşuluk etkisi veya grupta etkisi olabilir. Bu yüzden, mekânsal analiz yaparken mekân bilgisinin yer aldığı modeller tercih edilmelidir.

Sayılabılır verilerin analizinde sıkça Poisson dağılımına sahip genelleştirilmiş doğrusal model (GDM) yaklaşımı kullanılır. Sayımla elde edilmiş mekânsal veri için bu model kullanıldığında artıklarda mekânsal bağımlılık söz konusu olmaktadır. Bu yüzden mekân bilgisinin komşuluk matrisi ile modele dahil edildiği ve alanlar arası otokorelasyonun rasgele etkiler ile ifade edildiği koşullu otoregresif türü modeller mekânsal verilerin analizinde kullanılmaktadır. Bu modelde, komşu olmayan alanlar koşullu bağımsız olarak modellenir iken, komşu alanlar mekânsal ilişkili olarak modellenir.

Tez çalışmasında, mekâna bağlı sayım verilerinin modellenmesinde mekânsal otokorelasyondan dolayı GDM'lerin yetersiz kaldığı gösterilmiştir. Bu tür verilerin analizinde, genelleştirilmiş doğrusal mekânsal modellerin (GDMM) uygulanabilirliği incelenmiş ve gerçek bir sayım verisi üzerinde uygulaması gösterilmiştir. Bu bağlamda, BÖLÜM 2'de, sırasıyla genelleştirilmiş doğrusal modeller ve rasgele etkilerin yer aldığı

genelleştirilmiş doğrusal karma modellerin (GDKM) özel bir hali olan hiyerarşik genelleştirilmiş doğrusal modellere (HGDM) yer verilmiştir.

BÖLÜM 3'te alansal verilerdeki mekânsal bilgisinin de modele katılmasıyla genelleştirilmiş doğrusal mekânsal modeller (GDMM) tanıtılmıştır. Alt bölümlerde mekânsal bağımlılığın tanımlanması için komşuluk matrisine değinilmiş ve mekânsal verilerde mevcut olan mekânsal otokorelasyonun testi için sıklıkla kullanılan yaklaşımlar incelenmiştir. Daha sonra, komşuluk bilgisinin ve mekânsal otokorelasyonun rasgele etkiler yardımıyla modelde yer aldığı koşullu otoregresif (CAR) modeller anlatılmıştır. R programında bulunan ve CAR önselleri ile Bayesci mekânsal modelleme yapan CARBayes paketi hakkında bilgi verilmiş ve diğer yardımcı paketlere değinilmiştir. Son olarak da risk haritalandırılması yapılabilmesi için mekânsal haritalandırma hakkında bilgi verilmiştir.

BÖLÜM 4'te verinin tanımı yapılarak mekânsal analizde kullanılmak üzere iki farklı komşuluk matrisi tanımlanmıştır. Bu matrislerin temel alındığı mekânsal otokorelasyon testi yapılmış ve çıkan sonuç doğrultusunda Poisson mekânsal log-doğrusal modeli kurularak çıktıları yorumlanmıştır. Modelden elde edilen kestirilmiş değerlerin beklenen değerlere kıyaslanması ile risk haritaları oluşturulmuştur.

BÖLÜM 5'te elde edilen sonuçlar özetlenmiş ve tartışılmıştır.

1.1. Literatür Taraması

Seçim bölgeleri gibi bitişik mekânsal birimlerle ilgili veri modellemesi, tarımsal alan denemeleri, görüntü analizi, coğrafi analiz veya hastalıkların haritalandırılması gibi alanlarda sıkça karşılaşılan bir durumdur. Bu uygulamalardaki yanıt değişkenleri mekânsal bağımlılık gösterir. Yani birbirine yakın birimlerden elde edilen gözlemler, birbirine uzak olanlara göre daha benzer olma eğilimi gösterir. Bu tarz verilerde, gözlem değerleri arasındaki mekânsal otokorelasyonu modellemek için çeşitli istatistiksel modeller geliştirilmiştir. Bunlardan bazıları; jeoistatistiksel model [2], eşzamanlı otoregresif modeller [3] ve koşullu otoregresif modellerdir [4].

Cressie [5], Schabenberger ve Gotway [6], ve Waller ve Gotway [7] mekânsal otoregresif modeller üzerine yoğunlaşmıştır. Wall [8], alansal verilerdeki mekânsal süreçlerin modellemesinde kullanılan yöntemlerin karşılaştırmasını yapmıştır. Banerjee et al. [9] temel özelliklerin hiyerarşik modellere geçişi üzerine odaklanmıştır. Fortin ve Dale [10], mekânsal otoregresif modellerin ekoloji alanındaki kullanımına değinmiştir.

Mekânsal olarak yapılandırılmış modellerin analizinde doğrusal model yaklaşımı kullanıldığında artıklarda mekânsal bağımlılık söz konusu olmaktadır. Artıklardaki mekânsal otokorelasyon için en yaygın çözümlerden biri Bayesci hiyerarşik modellerdir.- Bu modellerde, mekânsal ilişkili rasgele etki kümesi yer almakta ve model farklı düzeylerle ifade edilmektedir. Bu bağlamda, alansal verilerde var olabilecek mekânsal otokorelasyon hiyerarşinin ikinci düzeyinde bir dizi rasgele etkiyle modellenir. Böylece, Bayesci hiyerarşinin her bir düzeyi farklı değişim kaynaklarını açıklamak için kullanılır. Modelde açıklayıcı değişkenlerin yanı sıra komşuluk bilgisinin de rasgele etkiler ile modele katılmasıyla daha tutarlı tahmin değerleri elde edilir. Bu modellerin haritalandırmada kullanılması için, Schabenberger ve Gotway [6], Waller ve Gotway [7], Banerjee et al. [9], Haining [11] ve Lawson et al. [12] tarafından modeller öne sürülmüştür.

GDMM'lerdeki rasgele etkiler genellikle Markov koşullu otoregresif önsel dağılım ile temsil edilir. Bu modellerde mekânsal otokorelasyon, alansal birimlerin komşuluk yapılarıyla ilişkilendirilmektedir. Literatürde birçok koşullu otoregresif model önerilmiştir. Bunlara, temel koşullu otoregresif modeli (intrinsic CAR, ICAR) ve Besag-York-Mollie [13] modelleri, Leroux, Lei ve Breslow [14] ve Stern ve Cressie [15] tarafından öne sürülen modeller örnek verilebilir. Bu modellerin önselleri, bağımsızlıktan mekânsal ilişkili yüzeylere geçiş yaparken rasgele etkileri tek bir genel mekânsal otokorelasyon düzeyi ile ifade etmektedir. Lee [16] ise bu modellerin hem teorik özelliklerini hem de pratik performanslarını karşılaştırmış ve Leroux modelinin en iyi model olduğu sonucuna ulaşmıştır.

Bazen tüm bölge için tek bir düzeyde mekânsal düzleştirme (smoothing) yapılması yeterli olmayabilir. Bu durumda, mekânsal otokorelasyonun alt alanlarını da dikkate almak gerekir. Bu tarz yerleştirilmiş mekânsal düzgünlük, yoksul ve zengin kesimlerin yan yana olduğu alanlarda meydana gelebilir. Bu bağlamda, yanıt değişkeni her iki kesimin bulunduğu sınır değerindeki ani bir değişim ile her kesim kendi içinde düzgün bir şekilde evrimleşebilir. Yerleştirilmiş mekânsal düzgünlüğün dikkate alındığı bu modeller, koşullu otoregresif modellerin genişletilmiş hali olarak öne sürülmüştür. Lawson ve Clark [17] ICAR modelini sıçrama bileşeni ile birleştirmiş, Lu et al. [18] lojistik regresyon kullanarak alansal birimlerin komşuluk yapısını modellemiş, Reich ve Hodges [19] mekânsal-zamansal çerçevede mekânsal olarak değişen bir varyansla değişken düzgünleştirme yapmış, ve Lee ve Mitchell [20] ise komşu alan birimlerindeki rasgele etkiler arasındaki kısmi korelasyonu birbirinden farklı olmalarının bir fonksiyonu olarak modellemiştir. Bu modeller, Markov zinciri Monte

Carlo (MCMC) benzetimine dayalı çıkarımların yapıldığı Bayes ortamında uygulanmaktadır.

Bu çalışma kapsamında, rasgele etkiler genel mekânsal otokorelasyon ile ifade edilmiştir. Alansal mekânsal veri, koşullu otoregresif önselleri kullanılarak modellenmiştir.



2. GENELLEŞTİRİLMİŞ DOĞRUSAL MODELLER

Doğrusal modeller, Y 'nin $n \times 1$ veri vektörü, X 'in $n \times p$ açıklayıcı değişken vektörü, β 'nin $p \times 1$ bilinmeyen parametre vektörü olduğu durumda,

$$Y = X\beta + \varepsilon$$

şeklinde tanımlanır. Hata terimine ait beklenen değer $E(\varepsilon_i) = 0$, $i = 1, \dots, n$ ve $\text{var}(\varepsilon_i) = \sigma^2 I$ olarak varsayılır. Bu modeldeki amaç parametre tahmini yapmaktır. Modeldeki çıkarımlar Normal dağılıma dayanır. Fakat sayım verilerinde veya ikili tanımlı verilerde normallik varsayımı sağlanamamaktadır. Bu gibi durumlarda, ortalamanın kendisini modellemek yerine bilinmeyen parametrelerin doğrusal fonksiyonu olarak ortalamanın fonksiyonu modellenir (Eşitlik 2.1). Burada kullanılan fonksiyon link veya bağ fonksiyonu olarak adlandırılır. Oluşturulan modeller genelleştirilmiş doğrusal model (GDM) adını alır ve yaygın olarak kullanılan log-doğrusal ve logit modeller bu modellere örnek olarak verilebilir.

$$E(Y) = \mu$$

$$g(\mu) = X\beta \tag{2.1}$$

Yanıt değişkeninin sürekli olduğu regresyon modelleri gibi modeller, genelleştirilmiş doğrusal modellerin özel bir halidir. Bu modellerde, yanıt değişkeninin dağılımı Normal dağılım ise, bağ fonksiyonu ($g(\cdot)$) olarak *birim* fonksiyon kullanılır. Bir GDM, klasik doğrusal modelleri iki şekilde genelleştirir. Birincisi, Y 'nin normalden başka bir dağılıma sahip olmasını sağlar. İkincisi ise ortalamanın bağ fonksiyonları sayesinde modellenmesine izin verir.

Tarihsel olarak, normal olmayan yanıt değişkenlerinin analizi yapılırken Y 'ye dönüştürme uygulanmış ve sabit bir varyans ile birlikte yaklaşık bir Normal dağılım elde edilmiştir. Daha sonra en küçük kareleri kullanan klasik regresyon yöntemleri uygulanmıştır. Fakat GDM'nin teorisi, yukarıda bahsedilen dönüşüme gerek kalmadığını göstermiştir. Yanıt değişkenin normal dağılıma uyma zorunluluğu olmamakla birlikte, verilerin üstel dağılım ailesinden gelmesi gerekir. GDM'deki bağ fonksiyonu seçimi, rasgele bileşen seçiminden farklıdır ve normalliğin sağlanması ve varsayımının sabitlenmesi gibi nedenlere göre seçilmemektedir.

GDM, normal olmayan verilerin analizi için ilk defa Nelder ve Wedderburn [21] tarafından ortaya atılmış ve McCullagh ve Nelder [22] tarafından geliştirilmiştir.

2.1. Genelleştirilmiş Doğrusal Modellerin Bileşenleri

Genelleştirilmiş doğrusal modeller (GDM), rasgele bileşen, sistematik bileşen ve bağ fonksiyonu olmak üzere üç bileşenden oluşur. Rasgele bileşen, Y yanıt değişkenini tanımlar ve bir olasılık dağılımı olduğunu varsayar. Sistematik bileşen, modelde yer alan açıklayıcı değişkenleri, bağ fonksiyonu ise yanıt değişkeninin beklenen değerinin bir fonksiyonunu belirtir.

Rasgele Bileşen: Bir GDM'nin rasgele bileşeni, yanıt değişkenini tanımlar ve bunun için bir olasılık dağılımı seçer. Yanıt değişkeninin gözlemleri $(y_1, \dots, y_n)$ olarak belirtilsin. Standart GDM'ler yanıt değişkeninin bağımsız olmasını varsayar.

Birçok uygulamada, yanıt değişkeni ile ilgili gözlemler “başarı” veya “başarısızlık” gibi ikili tanımlanır. Daha genel olarak, her bir Y_i değeri, $i = 1, \dots, n$, sabit sayıda denemeden elde edilen başarıların sayısı olabilir. Her iki durumda da yanıt değişkeni için binom dağılımı olduğu varsayılır. Bazı uygulamalarda ise her gözlem sayımla elde edilir. Bu tür verilerde yanıt değişkeni negatif olmayan tam sayı değerleri alır ve dağılımının Poisson veya negatif binom dağılımı olduğu varsayılır. Eğer gözlemler sürekli yapıda ise, yanıt değişkeni için normal dağılım varsayımı yapılır.

Sistematik Bileşen: Bu bileşen modelde yer alan açıklayıcı değişkenleri belirtir. Bu değişkenler, model denkleminin sağ tarafında doğrusal tahmin ediciler olarak yer alır. Formülde $\{x_j\}$, $j = 1, \dots, k$, olarak ifade edilen değişkenler sabit sayı ile birlikte

$$\alpha + \beta_1 x_1 + \dots + \beta_k x_k$$

şeklinde tanımlanan sistematik bileşeni oluşturmaktadır.

Bazı $\{x_j\}$ 'ler modeldeki diğerleri tarafından elde edilebilir. Örneğin, $x_3 = x_1 x_2$ olarak tanımlanır ise, x_1 ve x_2 arasındaki etkileşimin yanıt değişkeni üzerindeki etkisi test edilmiş olur veya $x_3 = x_1^2$ şeklinde bir tanımlama yapıldığında modelde x_1 'in eğrisel etkisine yer verilmiş olur.

Bağ Fonksiyonu: Yanıt değişkeninin beklenen değeri yani olasılık dağılımının ortalaması, $\mu = E(Y)$ ile gösterilsin. $g(\cdot)$ fonksiyonu sayesinde μ , doğrusal olarak tahmin edicilere bağlanır (Eşitlik 2.1).

$$g(\mu) = \alpha + \beta_1 x_1 + \dots + \beta_k x_k \quad (2.1)$$

Burada $g(\cdot)$, rasgele ve sistematik bileşeni birbirine bağlayan fonksiyon olarak ifade edilebilir.

En basit bağ fonksiyonu $g(\mu) = \mu$ ile ifade edilir. Bu fonksiyon, ortalamayı direk olarak modeller ve *birim* fonksiyon olarak adlandırılır. Böylece, ortalama yanıt doğrusal bir model ile belirlenir (Eşitlik 2.2).

$$\mu = \alpha + \beta_1 x_1 + \dots + \beta_k x_k \quad (2.2)$$

Bu model, sürekli yanıt değişkeni için klasik regresyon modelini ifade etmektedir.

Diğer bağ fonksiyonları, μ 'yü doğrusal olmayan şekilde tahmin edicilere bağlar. Örneğin, $g(\mu) = \log(\mu)$, ortalamayı logaritmik fonksiyonla modeller. Buradaki *log* fonksiyonu, negatif değer alamayan sayısal veriler için uygundur. *log* bağlantısını kullanan bir GDM, log doğrusal model olarak adlandırılır ve Eşitlik (2.3)'teki gibi gösterilir.

$$\log(\mu) = \alpha + \beta_1 x_1 + \dots + \beta_k x_k \quad (2.3)$$

Diğer bir bağ fonksiyonu *logit* fonksiyon, lojistik regresyon modelinin oluşturulmasında kullanılır. Bağ fonksiyonu $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$ şeklinde tanımlanır ve odds değerlerinin logaritmasını modeller. μ , bir olasılık gibi 0 ile 1 arasında değer aldığından yani Binom dağılımına sahip olduğunda uygun bir modeldir.

Tez çalışmasında belirli bir süre içerisinde elde edilen sayım verileri üzerinden analiz yapıldığı için Poisson dağılımına sahip GDM modeli aşağıda verilmiştir.

Poisson Dağılımlı Genelleştirilmiş Doğrusal Model: Poisson dağılımı, pozitif bir ortalamaya sahiptir ve genel olarak ortalamanın logaritması alınarak modellenir. Ortalamanın logaritması, doğrusal tahmin edici $\alpha + \beta x$ gibi herhangi bir gerçek sayı değerini alabilir. Poisson log doğrusal modeli, y için Poisson dağılımını varsayar ve bağ fonksiyonu olarak *log* kullanır.

Tek bir açıklayıcı değişken x için, Poisson log doğrusal modeli Eşitlik (2.4)'teki gibidir.

$$\log(\mu) = \alpha + \beta x \quad (2.4)$$

Ortalama (μ), üstel ilişkiyi karşılamaktadır.

$$\mu = \exp(\alpha + \beta x) = e^\alpha (e^x)^\beta \quad (2.5)$$

x 'teki bir birimlik artış ortalama üzerinde çarpımsal etkiye sahiptir. Yani, $x+1$ düzeyindeki Y 'nin ortalaması, x 'in e^β ile çarpımındaki Y 'nin ortalamasına eşittir. Eğer $\beta = 0$ ise, $e^\beta = e^0 = 1$ ve çarpım faktörü 1 olur. Bu yüzden, Y 'nin ortalaması x değıştikçe değışmeyecektir. Eğer $\beta > 0$ ise, $e^\beta > 1$ olur ve Y 'nin ortalaması x arttıkça artmaktadır. Tam tersi $\beta < 0$ olduđu durumda, x azaldıkça Y 'nin ortalaması da azalacaktır.

2.2. Genelleştirilmiş Doğrusal Karma Modeller

Genelleştirilmiş doğrusal karma modeller (GDKM), doğrusal tahmin edicide sabit etkilerin yanı sıra rasgele etkileri de içeren GDM'lerdir. GDM'lerde olduđu gibi bu modeller de üç adımda tanımlanabilir. Öncelikle, rasgele etkiler üzerinde koşullu bağımlı değışkenin dağılımının üstel aile olması gerekir. Sabit ve rasgele etkileri içeren doğrusal tahmin edici tanımlanır ve kullanılan dağılımlara göre uygun bağ fonksiyonu seçilir.

GDKM'lerde yer alan rasgele etkiler arasındaki otokorelasyonu belirlemek için normal dağılım uygun olsa da, rasgele etkiler için farklı dağılımlar model sınıfını büyük ölçüde zenginleştirmektedir. Lee ve Nelder [23], rasgele etkilerin dağılımını karma modellerden hiyerarşik GDM'lere genişletmiştir. Başka bir deyişle, Hiyerarşik GDM'ler, normal doğrusal karma modeller ve GDM'lerin sentezinden elde edilen modellerdir.

Lee ve Nelder, HGDM'leri Eşitlik (2.6)'da verilen beklenen değeri ve varyans ile Eşitlik (2.7)'deki gibi tanımlanmıştır.

(i) Yanıt değışkeni (Y), rasgele etkilere (u) koşullu olarak GDM ailesine sahiptir. Bu yüzden, ortalama ve varyansı bu koşula bağılı olarak

$$E(Y|u) = \mu \text{ ve } \text{var}(Y|u) = \phi V(\mu) \quad (2.6)$$

şeklinde hesaplanır. ϕ yayılım parametresini ifade eder.

Doğrusal tahmin edici Eşitlik (2.7)'deki gibi tanımlanır.

$$\eta = g(\mu) = X\beta + Z\mathcal{G} \quad (2.7)$$

Monoton fonksiyon $\mathcal{G} = \mathcal{G}(u)$ ve Z rasgele etkileri, β ve X ise sabit etkileri ifade eder.

(ii) Rasgele bileşen u , λ parametresine sahip bir GDM eşlenik dağılımına sahiptir. Önsel ve sonsal dağılımların aynı aileden geliyor olması önsel dağılımın eşlenik önsel olduđu

anlamına gelir. Hesaplamalarda kolaylık sağlaması açısından eşlenik önsel dağılımlar uygulamada sıkça tercih edilir [24].

Normal doğrusal karma modeller, normal-normal HGDM'lere örnek olarak verilebilir çünkü bu model,

(i) $Y|u$ GDM dağılımına sahip olduğunda

$$\text{var}(Y|u) = \varphi, \quad \varphi = \sigma^2 \quad \text{ve} \quad V(\mu)=1 \quad \text{iken}$$

$$\eta = \mu = X\beta + Z\vartheta, \quad \vartheta = u \quad \text{iken}$$

(ii) $u \sim N(0, \lambda), \quad \lambda = \sigma_g^2$

şeklinde tanımlanır.

$Y|u$ 'nun Poisson dağılımına sahip olduğu durumda,

$$\mu = E(Y|u) = \exp(X\beta)u$$

ortalamaya sahiptir ve model $\log$ bağlantı fonksiyonu ile Eşitlik (2.8)'deki gibi tanımlanır.

$$\eta = \log(\mu) = X\beta + \vartheta, \quad \vartheta = \log u \quad \text{iken} \quad (2.8)$$

Burada eğer u 'nun dağılımı Gamma ise, ϑ log-gamma dağılımına sahiptir ve model Poisson-gamma model olarak adlandırılır. GDKM'ler, ϑ için normal dağılımı varsayar ve u 'nun dağılımı log-normal olur. $\vartheta = \log u$ dönüşümü altında ilgili Poisson GDKM, Poisson log-normal HGDM olarak adlandırılabilir. u için Normal bir dağılım, $Y|u$ için normalin eşleniğidir ve Gamma dağılımı da Poisson için eşleniktir. Yani her iki Poisson-gamma modeli ve Poisson GDKM'ler HGDM'lerin bir sınıfına aittir. Diğer eşlenik dağılımlara Çizelge 2.1'de yer verilmiştir.

Çizelge 2. 1. Bayesci GDKM için önsel dağılımlar

Olabilirlik Fonksiyonu	Önsel Dağılım	Sonsal Dağılım
Binom	Beta	Beta
Negatif Binom	Beta	Beta
Normal	Normal	Normal
Poisson	Gamma	Gamma
Üstel	Gamma	Gamma
Gamma	Gamma	Gamma

Doğrusal modellerde ve doğrusal karma modellerde kısıtlar $E(\varepsilon)=0$ ve $E(\eta)=0$ gibi rasgele etkiler üzerinde konulmuştur. GDKM’lerde genellikle $E(\eta)=0$ olduğu varsayılır.

Lee ve Nelder [23], birden fazla rasgele bileşenli modeller kurulduğunda, rasgele etkiler üzerinde kısıtlar koymanın daha uygun olduğunu belirtmiştir. Poisson-gamma modeli için $E(u)=1$ kısıtı konulabilir. Bu kısıtlar, kısıtların koyulduğu parametrelerin tahminlerini etkiler. Bu yüzden farklı HGDM’ler karşılaştırılırken neler üzerinden kısıt koyulduğuna dikkat edilerek karşılaştırma yapılmalıdır.

GDKM’lerde rasgele etkilerin parametre tahminleri yapılırken maksimum olabilirlik tahminlerini yapmak çok boyutlu integrallerden dolayı zor olmaktadır. Bu yüzden daha karmaşık yapıların çözümlenmesine imkân sağlayan Bayesci yaklaşımla tahminler elde edilir. Bu yaklaşımda, bütün parametreler rasgele değişken gibi düşünülür ve dağılım önselleri tanımlanır. Parametre tahminleri için hedef dağılımından örneklemeler elde eden MCMC yöntemi kullanılır.

3. GENELLEŞTİRİLMİŞ DOĞRUSAL MEKÂNSAL MODELLER

Alansal verilerdeki mekânsal bağımlılık, istatistiksel modeller kullanılarak farklı şekilde modellenebilir. Doğrusal modellerde genellikle, gözlemlerin bağımsız ve aynı dağıldığı varsayımı üzerinden modelleme yapılırken mekânsal verilerde bu durum söz konusu olmayabilir. Çünkü komşu bölgeler arasında korelasyon mevcuttur ve bu da bağımsızlığı bozmaktadır. Bu durumda veri veya model artıklarındaki mekânsal korelasyonun varlığı geriye kalan değişimi modellemenin tek yoludur.

Mekânsal analizde çalışma bölgesi S , ilgili yanıt değişkeni $Y = (Y_1, \dots, Y_K)$ ve offset vektörüne $O = (O_1, \dots, O_K)$ bağlanmış, birbiriyle örtüşmeyen K alana bölünmüştür, $S = \{S_1, \dots, S_K\}$. Burada sözü geçen offset değişkeni GDM'in doğrusal bir bileşenidir. Özellikle Poisson modelinde ortalamanın modellenmesini $\log$ bağ fonksiyonu ile ölçeklendirmek için kullanılır [25]. Yanıt değişkeninde oluşabilecek kayıp değerlere modelde yer verilebilir. Yanıt değişkenindeki mekânsal değişim, açıklayıcı değişken matrisi $X = (X_1, \dots, X_K)$ ve mekânsal bileşen $\psi = (\psi_1, \dots, \psi_K)$ tarafından modellenir. Mekânsal bileşen, açıklayıcı değişkenlerin etkisinden sonra artıklarda kalabilecek mekânsal otokorelasyonu ayrıştırmak için modele dâhil edilir. Her alansal birim S_k için açıklayıcı değişken vektörü, $X_k = (1, x_{k1}, \dots, x_{kp})$ olarak tanımlanır ve vektördeki 1 sabit terimi ifade eder.

Mekânsal alansal veriler için genelleştirilmiş doğrusal karma model (Eşitlik 3.1) ve parametrelere ait önseller aşağıda verildiği gibidir.

$$Y_k | \mu_k \sim f(y_k | \mu_k, \nu^2), \quad k = 1, \dots, K$$

$$\nu^2 \sim \text{Ters Gamma}(a; b)$$

$$\beta \sim N(\mu_\beta; \Sigma_\beta)$$

$$g(\mu_k) = X_k^T \beta + O_k + \psi_k \quad (3.1)$$

Yanıt değişkeni Y_k üstel dağılım ailesinden gelir ve Bayesci koşullu otoregresif modeller için de Normal, Poisson veya Binom dağılımına sahip olabilir. Y_k 'nın beklenen değeri $E(Y_k) = \mu_k$ ile gösterilir. ν^2 sadece Normal dağılımda kullanılan ek bir ölçek parametresini

ifade eder ve ters gamma dağılımına sahiptir. Yanıt değişkeninin beklenen değeri, doğrusal tahmin ediciye bir bağ fonksiyonu $g(.)$ üzerinden bağlanır.

Regresyon parametre vektörü $\beta = (\beta_0, \beta_1, \dots, \beta_p)$ ile gösterilir ve normal dağılıma sahiptir. Burada μ_β , $1 \times p$ ortalama vektörü, Σ_β ise $p \times p$ köşegen varyans matrisidir. Ayrıca, doğal kübik eğrisel çizgi veya polinom temelli fonksiyonlar sayesinde doğrusal olmayan açıklayıcı değişkenler de modele eklenebilir.

Sürekli ve sayılabilir mekânsal veriler için Normal, Binom ve Poisson dağılımlarına sahip mekânsal modeller tanımlanabilir.

Mekânsal Doğrusal Normal Model: Sürekli veriler için mekânsal doğrusal model, Eşitlik (3.2)'deki gibi tanımlanır.

$$Y_k \sim N(\mu_k, \sigma^2)$$

$$\mu_k = X_k^T \beta + O_k + \psi_k \quad (3.2)$$

$$\mu_k = \beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp} + \phi_k \quad (3.3)$$

Yanıt değişkeni Y_k 'nın modelde normal dağılıma sahip olduğu varsayılır. σ^2 açıklanamayan rasgele değişim miktarını ifade eder. Offset vektörünün olmadığı ve mekânsal bileşenin koşullu otoregresif modellerden gelen rasgele etkiye eşit olduğu durumda ($\psi_k = \phi_k$) model Eşitlik (3.3)'teki gibi tanımlanabilir.

Eşitlik (3.3)'te yer alan ϕ_k , rasgele etkiler olarak adlandırılır ve modelde yer alan her alan birimi için bir tane yani toplamda K tane bulunmaktadır. ϕ_k değeri, tahmini riske veya kestirilmiş değerlere düzenleme sağlar ve regresyon bileşenine, $\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp}$, eklenir. Rasgele etkilerin kümesi mekânsal otokorelasyon ile sınırlıdır ve haritalandırma yapıldığında mekânsal olarak pürüzsüz bir yüzey oluşturmaktadır. Ayrıca, verilerde ölçülemeyen etki karışımı (confounding) gibi nedenlerden dolayı oluşabilecek mekânsal otokorelasyon ve değişime izin verir.

Binom Mekânsal Lojistik Model: Sayılabilir veriler için kullanılabilecek mekânsal lojistik model, Eşitlik (3.4) veya offset vektörünün olmadığı ve $\psi_k = \phi_k$ olduğu durumda $Y_k \sim \text{Binom}(n_k, \xi_k)$ olmak üzere Eşitlik (3.5)'teki gibi tanımlanabilir.

$$\ln\left(\frac{\xi_k}{1-\xi_k}\right) = X_k^T \beta + O_k + \psi_k \quad (3.4)$$

$$\ln\left(\frac{\xi_k}{1-\xi_k}\right) = \beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp} + \phi_k \quad (3.5)$$

Modelde yer alan n_k , k. alandaki deneme sayısını, ξ_k ise bir denemedeki başarı olasılığını ifade eder. Mekânsal değişim, açıklayıcı değişkenler ve açıklanamayan mekânsal değişim tarafından modellenir.

Poisson Mekânsal Log-Doğrusal Model: Sayılabilir veriler için kullanılan diğer bir model ise Poisson mekânsal log-doğrusal modelidir ve $Y_k \sim \text{Poisson}(\mu_k)$ olmak üzere Eşitlik (3.6) veya Eşitlik (3.7)'deki gibi tanımlanır.

$$\ln(\mu_k) = X_k^T \beta + O_k + \psi_k \quad (3.6)$$

$$\ln(\mu_k) = \beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp} + \phi_k \quad (3.7)$$

Y_k , k. alanda meydana gelen olay sayısıdır.

Önsel Dağılımlar: Her bir regresyon parametresi β , $(\beta_1, \dots, \beta_p)$, için bağımsız bir Normal önseli tanımlanmıştır, $\beta \sim N(\mu_\beta, \Sigma_\beta)$. Sadece Normal dağılımda yer alan ölçek parametresi ν^2 , tek biçimli önsel dağılımına sahiptir. Varyans parametresi için yaygın olarak kullanılan diğer bir önsel dağılım ise ters gamma dağılımıdır, $\nu^2 \sim \text{Ters Gamma}(a; b)$. Bu dağılımda hiperparametrelerin seçimi ν^2 'nin küçük değerleri için kolay olmamaktadır.

Genel Koşullu Otoresif Önselleri: Literatürde mekânsal otokorelasyonu modellemek için dört model yaygın olarak kullanılmaktadır. Bunlar ICAR modeli, Besag et al. [13] tarafından önerilen Besag-York-Mollie (BYM) veya konvolüsyon modeli, Stern ve Cressie [15] tarafından öne sürülen model, Leroux et al. [14] tarafından önerilen leroux modelidir. Her model Gauss Markov Rasgele Alanı (GMRF)'nin özel bir örneğidir.

Yukarıdaki modellerde yer alan ve Eşitlik (3.1)'de belirtilen mekânsal bileşen ψ , koşullu otoresif modellerden gelen bir dizi rasgele etki içerir, $\phi = (\phi_1, \dots, \phi_K)$. Koşullu otoresif modeller, ϕ 'nin mekânsal otokorelasyon olarak modellenmesi için kullanılan en yaygın modellerdendir. Bu modeller, zaman serileri modellemesindeki otoresif modellerin mekânsal bir uyarlamasıdır ve ϕ 'ye mekânsal otokorelasyonu adapte edebilmek için

komşuluk matrisini, $\mathbf{W} = [w_{kj}]_{K \times K}$, kullanır. Rasgele etkiler, $\phi \sim N(0, \tau^2 Q(\mathbf{W}, \rho)^{-1})$ şeklinde genel bir formda yazılabilir. Buradaki τ^2 , rasgele etkiler arasındaki değişimi kontrol eden varyans parametresi, $Q(\mathbf{W}, \rho)$ ise tekil olabilen bir kesinlik (precision) matrisidir. Bu matris, rasgele etkilerin mekânsal otokorelasyonunu kontrol eder ve komşuluk veya ağırlık matrisi olarak adlandırılan negatif olmayan $K \times K$ boyutlu $\mathbf{W}$ matrisine ve mekânsal bağımlılık parametresine (ρ) dayanır.

3.1. Mekânsal Ağırlık ve Mekânsal Komşuluk

Mekânsal ağırlıkların oluşturulması alansal verilerin analizinde önemli bir adımdır. İlk adım, gözlemler arasındaki ilişkilere bağlı olarak hangilerine sıfırdan farklı bir ağırlık verileceğini tanımlamak yani komşu alanları belirlemektir. İkinci adım ise komşuluk ilişkisi olanlara ağırlıkları tanımlamaktır. Komşuluk ve ağırlıkların tanımlanması için literatürde birçok yaklaşım mevcuttur.

Mekansal ağırlıkların tanımlanması Cressie [5], Schabenberger ve Gotway [6], Fortin ve Dale [10], O’Sullivan ve Unwin [26], Ward ve Gleditsch [27], Chun ve Griffith [28], ve Banerjee et al. [9] tarafından incelenmiştir.

Mekansal ağırlıklar, komşu alanların oluşturduğu komşuluk listeleri tarafından belirlenmiş ağırlıklar listesi olarak görülebilir ve her alanın kendine ait komşuluk listesi bulunmaktadır. S_k ve S_j alanları arasındaki ağırlık w_{kj} , k . alana ait ağırlıklar listesi bileşeninin k_{kj} . ögesini ifade eder. Eğer herhangi bir S_j alanı S_k ’nın komşuluk listesinde değil ise, S_j , S_k ’nın komşusu değildir ve komşuluk matrisindeki w_{kj} ağırlığı bu alanlar için sıfır olarak tanımlanır.

Alansal verilerdeki mekânsal bağımlılığın tanımlanması için mekânsal ağırlıklardan elde edilen komşuluk matrisinin, $\mathbf{W}$, tanımlanması gerekir. Bu matris mekânsal yakınlık, mekânsal link veya mekânsal ağırlık matrisi olarak da ifade edilebilir [29], [30], [31]. $\mathbf{W}$ matrisinin k . ve j . elemanı w_{kj} , S_k ve S_j alanları arasındaki mekânsal yakınlığı nicelleştirmekte ve tüm alan üzerinde komşuluk yapısı tanımlamaktadır. Böylece, K tane alanın birbirine göre mekânsal olarak nasıl konumlandığını gösteren $K \times K$ boyutlu bir matris elde edilir.

$$\mathbf{W} = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1K} \\ w_{21} & w_{22} & \cdots & w_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ w_{K1} & w_{K2} & \cdots & w_{KK} \end{pmatrix}_{K \times K}$$

En yalın komşuluk tanımı, ikili bağlantı matrisi olarak tanımlanır.

$$w_{kj} = \begin{cases} 1, & \text{eğer } S_k \text{ ve } S_j \text{ sınır paylaşıyor ise} \\ 0, & \text{diğer} \end{cases} \quad (3.8)$$

Eğer herhangi iki alan (S_k, S_j) mekânsal olarak birbirine yakın ve ortak bir sınır paylaşıyor ise bu alanlar komşu olarak adlandırılır ve $\mathbf{W}$ matrisindeki w_{kj} değeri 1 olarak tanımlanır. Eğer alanlar mekânsal olarak birbirine yakın değil yani ortak bir sınırları mevcut değil ise komşu alanlar değildir ve $\mathbf{W}$ matrisindeki w_{kj} değeri 0 olarak tanımlanır. Ayrıca, her alan kendisinin komşusu olamayacağı için w_{kk} ağırlığı 0 olarak tanımlanır. Bu yakınlık ölçüsünün seçimi ile simetrik bir komşuluk matrisi elde edilir, $w_{kj} = w_{jk}$.

Mekânsal yakınlık ölçümleri için başka seçenekler de mevcuttur. Örneğin, alanlar arasında sınır durumu olmadan yakın çevredeki bölgeleri kapsayacak bir komşuluk matrisi oluşturulabilir.

$$w_{kj} = \begin{cases} 1, & \text{eğer } S_k \text{'nin merkez noktası, } S_j \text{'nin merkez noktasına en yakın } d \text{ alan içinde ise} \\ 0, & \text{diğer} \end{cases} \quad (3.9)$$

S_j alanı, S_k 'nın d en yakın komşusundan biri ise, matristeki bu alanlara ait mekânsal ağırlık değeri 1 olarak tanımlanır, $w_{kj} = 1$. Bu şekilde elde edilen matris, simetrik olma özelliğini sağlamayabilir, yani $w_{kj} \neq w_{jk}$.

Belirli sayıda en yakın komşuyu belirtmek yerine, komşuluklar mesafenin bazı parametrik fonksiyonlarıyla tanımlanabilir. Örneğin, eğer d_{kj} , S_k ve S_j alanlarının merkez noktaları arasındaki mesafe (Öklid, şehir bloğu veya başka bir mesafe metriği) ise w_{kj} ağırlıkları Eşitlik (3.10) ve Eşitlik (3.11)'deki gibi tanımlanabilir.

$$w_{kj} = \begin{cases} 1, & \text{eğer } d_{kj} < \delta \\ 0, & \text{diğer} \end{cases} \quad (3.10)$$

veya

$$w_{kj} = \begin{cases} d_{kj}^{-\alpha}, & \text{eğer } \alpha > 0 \\ 0, & \text{diğer} \end{cases} \quad (3.11)$$

δ , komşuluk için tanımlanan maksimum uzaklık miktarını, α ise gücü ifade etmektedir. Her iki yaklaşım da simetrik ağırlık vermektedir.

Son yaklaşım olarak, S_k 'nın S_j ile paylaştığı sınırın oranı ile elde edilen komşuluk yapısı verilebilir ve ortaya çıkan komşuluk matrisi simetrik değildir.

$$w_{kj} = \begin{cases} \frac{l_{kj}}{l_k}, & \text{eğer } S_k \text{ ve } S_j \text{ sınır paylaşıyor ise} \\ 0, & \text{diğer} \end{cases} \quad (3.12)$$

Burada l_{kj} , S_k ve S_j arasındaki ortak sınırın uzunluğudur ve l_k , S_k 'nin çevre uzunluğunu temsil etmektedir [29]. Böyle bir yapı, malların, insanların veya hastalıkların akışının modellenmesinde ortaya çıkabilir. Eğer, bir bölgeden başka bir bölgeye ciddi bir akış varsa, ortak sınırın daha büyük bir bölümünü kapsayan komşu bölgeler, daha fazla bölgenin çıktısını alacaktır.

Bütün bu komşuluk matrislerinin yanı sıra, her bölgedeki toplam komşu sayısı üzerinden düzeltme yapmak istenebilir. Bunun için her w_{kj} değeri, S_k 'nin komşu ağırlıklarının toplamına bölünerek standartlaştırılmış satır matrisi ($\mathbf{W}_{std}$, std :standartlaştırılmış) elde edilebilir. Böylece, az sayıda komşuluğu olan alanlardan gelen bağlantıların ağırlıkları, daha çok komşusu olan alanlara göre daha büyük olacaktır.

$$\mathbf{W}_{std} = \frac{w_{kj}}{\sum_{j=1}^N w_{kj}} \quad (3.13)$$

Örneğin, S_k 'nin 4 komşusu olduğunu varsayalım, bu alanın komşularıyla oluşturduğu her bir ağırlık $1/4$ ağırlığına sahip olacaktır. $\mathbf{W}_{std}$ matrisi için de simetrik olma gerekliliği bulunmamaktadır ve çoğu düzensiz bölgelerde de simetrik değildir.

Yukarıdaki komşuluk matrislerinin tanımında yer alan d uzaklığının nasıl seçileceği açık bir konu olmadığı için ikili bağlantı matrisi yaygın olarak kullanılmaktadır. Yani, S_k ve S_j ortak bir sınırı paylaşıyor ($k \sim j$) ise $w_{kj} = 1$, aksi halde $w_{kj} = 0$ olarak tanımlanır. Buradan hareketle de coğrafi olarak bitişik alanlardaki rasgele etkiler (ϕ_k, ϕ_j) otokorelasyon ile ilişkilendirilir iken, bitişik olmayan alanlardaki rasgele etkiler geri kalan rasgele etkilerin değerleri göz önüne alındığında koşullu bağımsızdırlar. Başka bir deyişle, $\mathbf{W}$ matrisinde

$w_{kj}=1$ olarak tanımlanan komşu alanlardaki veriler mekânsal olarak, $w_{kj}=0$ olan diğer alanlar ise koşullu bağımsız olarak modellenmektedir.

3.2. Mekânsal Otokorelasyonun Değerlendirilmesi

Klasik regresyon modellerinde (doğrusal modeller, lojistik regresyon vb.), modelden gelen hataların (artıkların) bağımsız olduğu varsayılır. Bu varsayım mekânsal verilerin kullanıldığı modellerde mekânsal ilişkiden dolayı bozulmaktadır ve böylece verideki mekânsal otokorelasyonun varlığını test etmek gerekir.

Yakın alanlar arasında gözlemlenen mekânsal benzerliğin kapsamını açıklayan çeşitli yöntemler mevcuttur. Griffith [32], otokorelasyon kavramı da dahil olmak üzere mekânsal korelasyon üzerine çeşitli yorumlama yapmıştır. Bu yorumlamada, mekânsal otokorelasyon terimi, farklı yerlerden alınan aynı tür ölçümler arasındaki korelasyon anlamına gelmektedir. Genel mekânsal otokorelasyon indeksi, komşu alanlar arasında gözlenen mekânsal benzerlik ile tüm çalışma üzerinde özet bir bilgi verir. Otokorelasyonun istatistiksel indekslerine çeşitli uygulamalarda rastlanmaktadır ve bu indekslerin teori ve uygulaması üzerine Cliff ve Ord [29], [33] referans isimler olarak verilebilir.

Yaygın olarak kullanılan Moran'ın I istatistiği gibi genel mekânsal otokorelasyon indekslerinin amacı, komşu alanların gözlem değerlerinin benzer olma eğilimini özetlemektir. Böylece, bir yöndeki indeksin uç değerleri pozitif mekânsal otokorelasyonu gösterirken, zıt yöndeki uç değerler negatif mekânsal otokorelasyonu gösterir.

Çoğu otokorelasyon indeksi ortak bir temel yapıyı paylaşır. Bu yapıda, k ve j konumlarındaki değerlerin benzerliği hesaplanır ve bu benzerlik konumların yakınlığına göre ağırlıklandırılır. Yüksek ağırlığa sahip yüksek benzerlikler (yakın alanlardaki benzer değerler) indeksin yüksek değerler almasına neden olur iken, yüksek ağırlıktaki düşük benzerlikler (yakın alanlardaki benzer olmayan değerler) indeksin düşük olmasına neden olmaktadır.

Yokluk Hipotezi Dağılımları (Null Distributions): Genel mekânsal otokorelasyon indeksinin çıkarımı, yokluk hipotezi altındaki test istatistiği dağılımından (yokluk dağılımı) gelmektedir. Bu dağılımın kuyruklarında yer alan indeksin gözlenen değerleri, anlamlı mekânsal otokorelasyon olduğunu gösterir. Bu nedenle, yokluk dağılımının belirlenmesi doğru istatistiksel sonuçlar elde edilmesinde önemli rol oynamaktadır.

Cliff ve Ord [33], iki farklı varsayım altında bağımsız gözlemlerin yokluk hipotezi altında genel mekânsal otokorelasyon indeksleri için dağılımsal özellikleri tanımlamıştır. Normallik varsayımı olarak adlandırılan ilk varsayım, gözlemlerin aynı ve bağımsız Normal dağılımına sahip olması koşuluna dayanmaktadır. Bu varsayım, sonsuz sayıdaki olası değerler kümesindeki tek bir gözlemi temsil eder. Böyle bir sürecin benzetimi yapıldığında (her bir lokasyonda bağımsız Normal gözlemleri oluşturmak), değerlerin toplamı sınırlandırılmaz ve dolayısıyla normallik varsayımı, normallik örnekleme veya serbest örnekleme olarak adlandırılır. Rasgeleleştirme olarak adlandırılan ikinci varsayım, konumlara gözlenen değerlerin rasgele atanması altında otokorelasyon indeksinin dağılımını değerlendirir. Her iki durumda da, bölge çapında verilerin ortalama ve varyanslarının sabit olduğu ek bir varsayım altında, Cliff ve Ord [33] genel mekânsal otokorelasyon indekslerinin bölge sayısı arttıkça asimptotik Normal dağılımına sahip olduğunu kanıtlamıştır. Upton ve Fingleton [34], asimptotik yaklaşımların bölge sayısının 20’den büyük olduğu durumlar için doğru olduğunu öne sürmektedir. Tiefelsdorf ve Boots [35] ise asimptotik dağılımın uygunluğunun, mekânsal yakınlık matrisi ($\mathbf{W}$) ve aynı zamanda bölgelerin sayısının bir fonksiyonu olduğunu belirtmiştir.

Kümelenmenin yani mekânsal ilişkinin olmadığını ifade eden yokluk hipotezi, bölgesel sayımlar arasında varsayılan bağımsızlığa ek olarak Poisson dağılımlı bölgesel sayıları heterojen beklenen değerler ile varsayar. Verilerin kesikli yapısı ve küçük sayım değerlerinin normallik varsayımını bozması beklenir. Ayrıca, sayımların beklenen değerleri ve varyanslarındaki değişim rasgeleleştirme varsayımını da bozmaktadır.

Mekânsal otokorelasyonun genel indekslerinin yokluk dağılımına ilişkin standart yaklaşımlar, heterojen olarak dağılmış Poisson verilerinin kümelenmesini değerlendirmek için uygun olamayacağından, Besag ve Newell [36] tarafından tanımlanan Monte Carlo hipotez testlerine geçiş yapılır.

3.2.1. Moran’ın I İstatistiği

Veride var olabilecek mekânsal otokorelasyon için kullanılan testlerden birisi Moran’ın I istatistiğidir. $y = (y_1, \dots, y_K)$, K tane lokasyonda ölçülen veriler olsun. Moran’ın I istatistiği Eşitlik (3.14)’de verildiği gibi, y_k ’nın yakınlarındaki yerlerde kendisiyle ilişkilendirilip ilişkilendirilmediğini test etmek için kullanılmaktadır.

$$I = \frac{K \sum_{k=1}^K \sum_{j=1}^K w_{kj} (y_k - \bar{y})(y_j - \bar{y})}{\left(\sum_{k=1}^K \sum_{j=1}^K w_{kj} \right) \left(\sum_{k=1}^K (y_k - \bar{y})^2 \right)}, \quad k=1, \dots, K, j=1, \dots, K \quad (3.14)$$

Burada y_k , y_j , sırasıyla k. ve j. alana ait gözlem değeri, $\bar{y}$ yanıt değişkeninin ortalaması,

$\bar{y} = \frac{\sum_{i=1}^K y_i}{K}$; w_{kj} ise k. ve j. alanlar arasındaki bağlantının mekânsal ağırlığıdır. Ortalamanın

merkezleştirilmesi, doğru modelin sabit bir ortalamaya sahip olduğunu ifade eder ve merkezleştirmeden sonra kalan herhangi bir örüntü mekânsal ağırlıklarla kodlanmış mekânsal ilişkilerden kaynaklanır.

Mekânsal otokorelasyon,

$$\sum_{k=1}^K \sum_{j=1}^K w_{kj} (y_k - \bar{y})(y_j - \bar{y})$$

şeklinde tanımlanan Moran'ın I istatistiğinin pay kısmı tarafından belirlenir. Eğer veriler pozitif mekânsal ilişkili ise, bu miktar pozitif değere sahip olacaktır. Çünkü mekânsal olarak yakın veri noktaları ($w_{kj}=1$), benzer değerler olarak ortalamanın altında veya üstünde yer alacaktır.

I, y_i 'ler arasındaki etkileşim ve yanıt değişkeninin dağılımı tarafından tanımlanan bir dağılıma sahip rasgele değişkendir. Komşu bölgeler benzer değerlere sahip olma eğilimi gösterdiğinde, I pozitif değer alacaktır. Eğer komşu bölgeler farklı değerlere sahip olma eğiliminde ise I negatif değer alacaktır. Komşu değerler arasında bir otokorelasyon olmadığında, I'nın beklenen değeri, $E(I) = -\frac{1}{K-1}$, bölge sayısı arttıkça sıfıra yaklaşmaktadır.

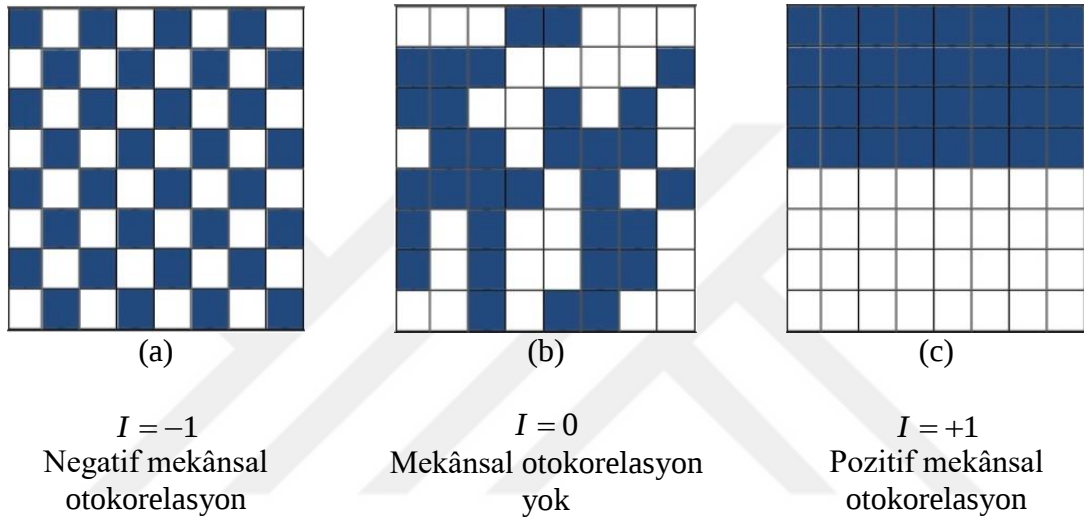
Moran'ın I istatistiği, herhangi iki X ve Y rastlantı değişkenlerinin gözlenen değerleri arasındaki ilişkiyi ölçen Pearson korelasyon katsayısına benzemektedir. Yani Moran I, mekânsal ağırlıklı bir Pearson korelasyon katsayısını ifade eder.

Moran'ın I istatistiği korelasyon katsayısı gibi -1 ile 1 arasında değer almaktadır.

- $I = -1$ ise güçlü negatif otokorelasyon vardır. Uzayda birbirine yakın olan veri kümeleri farklı değerler almaktadır.
- $I = 0$, bağımsızlığı yani mekânsal otokorelasyonun olmadığını ifade eder.

- $I = 1$, güçlü pozitif otokorelasyon varlığını göstermektedir. Uzayda birbirine yakın olan veri kümeleri benzer değerler almaktadır.

Moran'ın I istatistiğinin bu değerleri için grafiksel bir yaklaşım Şekil 3.1'de verilmiştir. (a) incelendiğinde gözlem değerleri mükemmel bir şekilde dağılım göstermiş ve komşu alanların gözlem değerleri birbirinden zıt değerler almıştır. (b)'de rasgelelik söz konusudur ve bu yüzden mekânsal otokorelasyon 0 değerini almıştır. (c)'de ise kümelenmenin yani pozitif mekânsal otokorelasyonun varlığı görülmektedir.



Şekil 3. 1. Mekânsal otokorelasyon

Moran'ın I istatistiğinin 0,5'ten büyük olduğu durumlara nadiren rastlandığı için 0,2'lik bir I değeri pozitif mekânsal otokorelasyonu ifade eder [37].

Mekânsal otokorelasyonun önem kontrolü hipotez testi ile değerlendirilebilir.

H_0 : Verilerde mekânsal ilişki yoktur.

H_1 : Verilerde mekânsal ilişki vardır.

Bu hipotezi test etmek için Moran'ın I istatistiği kullanılır ve anlamlılık değeri (p) değeri permütasyon testi yardımıyla hesaplanabilir.

I 'nın gözlenen herhangi bir değerinin önem kontrolü, beklenen değeri ile kıyaslanarak da yapılabilir. Burada dağılım varsayımlarının doğru tanımlanması önem kazanmaktadır. Rasgeleleştirme varsayımıyla, K lokasyon arasında veri değerleri yeniden atanır ve gözlenen değer için rasgeleleştirme dağılımı sağlanır. Eğer gözlenen değer bu dağılımın

kuyruklarında yer alırsa gözlemler arasındaki bağımsızlık varsayımı reddedilir. Böylece verilerde mekânsal otokorelasyon olduğu sonucuna varılır. Rasgeleleştirme yerine normallik varsayımına dayanacak olursa, z değeri, $z = \frac{I - E(I)}{\sqrt{Var(I)}}$, standart normal dağılım ile karşılaştırılır. Bu durumda I 'ya ait varyans değeri Eşitlik (3.15)'teki gibi elde edilir.

$$Var(I) = \frac{K^2 S_1 - K S_2 + 3 S_0^2}{(K-1)(K+1) S_0^2} - \left(\frac{1}{K-1} \right)^2 \quad (3.15)$$

Varyansın bulunmasında kullanılan yardımcı fonksiyonlar,

$$S_0 = \sum_{k=1}^K \sum_{j=1}^K w_{kj}, \quad S_1 = \frac{1}{2} \sum_{k=1}^K \sum_{j=1}^K (w_{kj} + w_{jk})^2, \quad S_2 = \sum_{k=1}^K (w_{k+} + w_{+j})^2$$

$$w_{k+} = \sum_{j=1}^K w_{kj}, \quad w_{+j} = \sum_{k=1}^K w_{kj} \quad (3.16)$$

şeklinde tanımlanır [7].

3.3. Koşullu Otoregresif Modeller

Koşullu otoregresif modeller (Conditional Autoregressive Models, CAR Models), çalışma bölgesinde bulunan alanlar arası otokorelasyonu rasgele etkiler üzerinden yakalayan GDMM'dir. Bu modelde komşu alanlar mekânsal modellenirken komşu olmayan alanlar koşullu bağımsız olarak modellenir. Parametre tahminleri, Gibbs örnekleme ve Metropolis Hastings algoritmalarına dayanan MCMC metodu kullanarak Bayesci yaklaşımla elde edilir.

Rasgele etkiler genellikle K tek değişkenli koşullu dağılımlar kümesi olarak tanımlanmaktadır, $\phi_{-k} = (\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_K)$. En basit koşullu otoregresif modeli, temel CAR (ICAR) modeli olarak adlandırılır ve Eşitlik (3.17)'deki gibi tanımlanır. Bütün CAR modelleri Eşitlik (3.1)'de verilen GDMM modeli altında tanımlanmış ve Eşitlik (3.17)'de olduğu gibi CAR önselleri ilgili model tanımında verilmiştir.

$$\left. \begin{aligned} \psi_k &= \phi_k \\ \phi_k | \phi_{-k}, \mathbf{W} &\sim N \left(\frac{\sum_{j=1}^K w_{kj} \phi_j}{\sum_{j=1}^K w_{kj}}; \frac{\tau^2}{\sum_{j=1}^K w_{kj}} \right) \\ \tau^2 &\sim \text{Ters Gamma}(a; b) \end{aligned} \right\} \quad (3.17)$$

Her bir ϕ_k , komşuluk bilgisine yani mekânsal yakınlığa bağlı bir ortalama ve varyansa sahip normal dağılımlı olarak modellenmektedir. Böylece modeldeki mekânsal otokorelasyon beklenen değer ve varyans üzerinden yakalanmaktadır.

- ϕ_k 'nın beklenen değeri $\frac{\sum_{j=1}^K w_{kj} \phi_j}{\sum_{j=1}^K w_{kj}}$ şeklinde ifade edilir, yani komşu bölgelerdeki

rasgele etkilerin ortalamasıdır. ($w_{kj}=1$ olanlar için)

- ϕ_k 'nın varyansı $\frac{\tau^2}{\sum_{j=1}^K w_{kj}}$ ile elde edilir ve komşu bölgelerin sayısı ile ters orantılıdır.

Bu varyans yapısı, güçlü mekânsal otokorelasyon varlığında, daha fazla komşunun, bir alanın rasgele etkisinin değeri hakkında daha fazla bilgiye sahip olduğu anlamına gelmektedir. Böylece k alanına yakın olan ve ϕ_k 'ya benzer değerlere sahip olan daha fazla alan, ϕ_k hakkında daha fazla bilgi olduğu anlamına gelerek belirsizliği azaltmaktadır.

Varyans parametresi τ^2 ise rasgele etkiler arasındaki değişim miktarını kontrol eder.

$\sum_{j=1}^K w_{kj}$, S_k 'nın komşuluk listesinde yer alan, mekânsal olarak komşu alanların sayısını belirtmektedir.

Bu modeldeki sınırlama, yalnızca güçlü mekânsal otokorelasyonun yakalanabilmesidir. Ayrıca, modelde mekânsal bağımlılık parametresi de bulunmamaktadır. Bu yüzden güçlü mekânsal otokorelasyon yapıları için iyi bir model olsa da zayıf ilişkilerde uygun değildir. Bunun üzerine Besag et al.[13] tarafından yaygın olarak kullanılan Besag-York-Mollie (BYM, konvolüsyon) modeli öne sürülmüştür (Eşitlik 3.18). Bu modelde hem zayıf hem de güçlü mekânsal otokorelasyona izin verecek bir uzantı önerilmiştir.

Bir bölgedeki değişkenlik, komşuluk bilgisine dayanan mekânsal olarak ilişkili bir etki ve alandan bağımsız etki olarak iki parçaya ayrılmıştır. Her ne kadar modeldeki değişkenlerin tahminleri direk olarak yapılamasa da, Bayesci hiyerarşik modeller kullanılarak MCMC yöntemleri ile sonsal dağılımları elde edilir. MCMC yöntemleri, temel olarak uygun bir yanma (burn-in) periyodundan sonra sonsal dağılımların elde edilişi olarak model parametrelerinin benzetimlerini üretir.

BYM modelinde [13], ICAR modeli bir dizi rasgele etki ile birleştirilmiştir.

$$\left. \begin{aligned} \psi_k &= \phi_k + \theta_k \\ \phi_k | \phi_{-k}, \mathbf{W}, \tau^2 &\sim N \left(\frac{\sum_{j=1}^K w_{kj} \phi_j}{\sum_{j=1}^K w_{kj}}; \frac{\tau^2}{\sum_{j=1}^K w_{kj}} \right) \\ \theta_k &\sim N(0; \sigma^2) \\ \tau^2, \sigma^2 &\sim \text{Ters Gamma}(a; b) \end{aligned} \right\} \quad (3.18)$$

Bu modelde ψ_k , mekânsal değişim (ϕ_k) ve mekânsal olmayan değişim (θ_k) olarak iki parçaya ayrılmıştır. Mekânsal otokorelasyon ϕ tarafından modellenir iken, θ rasgele etkileri sıfır ortalama ve sabit varyans ile bağımsızdır. Mekânsal etkiler için koşullu beklenen değer, bölgelerdeki rasgele etkilerin ortalaması ve varyans ise komşuların sayısı ile ters orantılıdır. Bu durumda, rasgele etkiler güçlü mekânsal otokorelasyona sahip ise çok fazla komşu, bir alanın komşularından gelen bilginin o kadar fazla olduğu anlamına gelmektedir. BYM modelinde, verideki her bir nokta için tahmin edilecek iki rasgele etki gerekmektedir. Bunun üzerine Leroux et al. [14] tarafından mekânsal otokorelasyon parametresi de dâhil edilen sadece tek bir rasgele etki kümesi içeren bir model öne sürülmüştür (Eşitlik 3.19).

$$\left. \begin{aligned} \psi_k &= \phi_k \\ \phi_k | \phi_{-k}, \mathbf{W}, \tau^2, \rho &\sim N \left(\frac{\rho \sum_{j=1}^K w_{kj} \phi_j}{\rho \sum_{j=1}^K w_{kj} + 1 - \rho}; \frac{\tau^2}{\rho \sum_{j=1}^K w_{kj} + 1 - \rho} \right) \\ \tau^2 &\sim \text{Ters Gamma}(a; b) \\ \rho &\sim \text{Tek Biçimli}(0; 1) \end{aligned} \right\} \quad (3.19)$$

Stern ve Cressie [15] tarafından önerilen modelde rasgele etkiler, Leroux modelinden farklı olarak Eşitlik (3.20)'deki gibi tanımlanır.

$$\phi_k | \phi_{-k} \sim N \left(\frac{\rho \sum_{j=1}^K w_{kj} \phi_j}{\sum_{j=1}^K w_{kj}}; \frac{\tau^2}{\sum_{j=1}^K w_{kj}} \right) \quad (3.20)$$

Her iki modelde de τ^2 , ters Gamma dağılımına sahiptir. ρ ise mekânsal otokorelasyon parametresidir ve tek biçimli dağılıma sahiptir. Böylelikle ρ değeri, modelde verilerden

tahmin edilmek yerine $[0,1]$ aralığında sabitlenir. ρ 'nin bu aralıkta değer alıyor olması, iki özel duruma neden olmaktadır.

- $\rho = 1$ olması güçlü mekânsal bağımlılığa karşılık gelir ve Eşitlik (3.19) ve Eşitlik (3.20)'de tanımlanan modelleri temel koşullu otoregresif modele (ICAR) indirger.
- $\rho = 0$ ise bağımsızlığı ifade etmektedir ve $\phi_k \sim N(0, \tau^2)$ dağılımına sahip olur. Bu durumda komşuluk matrisi modelin bir parçası olmadığı için, $K \times K$ boyutlu sıfırlardan oluşan bir $\mathbf{W}$ matrisi tanımlanabilir.

Yukarıdaki modellerde tanımlanan koşullu otoregresif önselleri, rasgele etkiler kümesi için tek bir düzeyde mekânsal düzgünleştirme uygular. Eşitlik (3.19)'da tanımlanan Leroux modelinde bu düzgünleştirme otokorelasyon parametresi ρ , tarafından kontrol edilir ve kısmi otokorelasyon yapısı (ϕ_k, ϕ_j) için,

$$COR(\phi_k, \phi_j | \phi_{-kj}, \mathbf{W}, \rho) = \frac{\rho w_{kj}}{\sqrt{\left(\rho \sum_{i=1}^K w_{ki} + 1 - \rho \right) \left(\rho \sum_{i=1}^K w_{ji} + 1 - \rho \right)}} \quad (3.21)$$

şeklinde gösterilir. Komşu olmayan alanlar ($w_{kj} = 0$) için rasgele etkiler koşullu bağımsız iken, komşu alanlardaki ($w_{kj} = 1$) kısmi otokorelasyon ρ tarafından kontrol edilir.

Bu modellerin karşılaştırılması Lee [16] tarafından yapılmış ve Leroux et al. [14] tarafından önerilen model hem teorik hem de pratik anlamda en iyi model olarak değerlendirilmiştir.

Bu model, belirsizliğin modelde doğru şekilde yayılmasına olanak sağladığı için en çok olabilirlik yerine Bayesci yöntemle modellenir. Bu yüzden, parametre tahminleri için aralıklar, % 95 güven aralığı yerine % 95 güvenilirlik aralığı (credible interval) olarak adlandırılır ve parametrenin tahmin edilen aralığa düşmesi olasılığı % 95'tir şeklinde yorumlanır. Parametre tahmini MCMC algoritması olarak adlandırılan benzetim tabanlı bir yaklaşımla yapılır.

3.3.1. CARBayes Paketi

CARBayes paketi, Duncan Lee tarafından alansal veriler için tek değişkenli ve çok değişkenli mekânsal genelleştirilmiş doğrusal karma modellerin uygulamasını içerir. MCMC benzetimi kullanılarak Bayesci yaklaşımla çıkarım yapılır. Yanıt değişkeni Binom, Normal veya Poisson dağılımına sahip olabilir. Mekânsal otokorelasyon, koşullu otoregresif

önsel dağılımlara dayanan bir dizi rasgele etki ile modellenir. Tek değişkenli mekânsal verilere, rasgele etkinin olmadığı modeller ve farklı CAR önsellerine sahip modeller uygulanabilir. Çok değişkenli mekânsal veriler için çok değişkenli CAR modeli, alanlardaki her bir birey için iki aşamalı hiyerarşik model olarak tanımlanmıştır.

En son sürümü olan 5.0'da, rasgele etki içermeyen basit bir model (*S.glm*), bölgelerdeki bireylerin modellenmesi için iki aşamalı hiyerarşik model, rasgele yürüyüş Metropolis adımları veya Metropolis düzeltmeli Langevin algoritması kullanılarak rasgele etkilerin güncellenmesi için bir argüman ve model tarafından açıklanan sapma yüzdesinin hesaplanması gibi model ve fonksiyonlar mevcuttur.

Bölüm 3.3'te yer verilen ICAR veya BYM gibi temel koşullu otoregresif modelleri R programında uygulayabilmek için **BayesX** [38], **hSDM** [39], **spatcount** [40], **spdep** [41] ve **INLA** [42] gibi paketler mevcuttur. Fakat bu paketler sadece kısıtlı CAR modellerini uygulamaktadır. Bu da CARBayes paketinin geliştirilmesine öncü olmuştur. Bu paketin avantajı, CAR modellerinin kolaylıkla uygulanabilir olmasıdır. Mekân bilgisi, komşuluk matrisi ile kolayca temsil edilebilmekte ve bu komşuluk matrisi verildiğinde R programında tek bir komut ile modeller uygulanabilmektedir. Ayrıca, yukarıda belirtilen paketlerde mevcut olandan daha çok model, bu paket sayesinde mekânsal alansal verilere uygulanabilmektedir. Mekânsal verilerin yanı sıra mekânsal zamansal veriler için de **CARBayesST** [43] paketi alansal birimler için geliştirilmiştir.

Tek değişkenli mekânsal verilerin analizinde CARBayes paketinde mekânsal bileşen ψ için çeşitli rasgele etki modelleri geliştirilmiştir.

Rasgele Etkinin Olmadığı Model

- *S.glm()*, rasgele etkinin olmadığı model, genelleştirilmiş doğrusal model için kullanılan fonksiyondur. Eşitlik (3.1)'de tanımlanan modelde $\psi_k = 0$ olduğu durum için elde edilen modeldir ve karmaşık modellerin karşılaştırılmasında kullanılmaktadır.

Genel Düzgünleştirilmiş CAR Modelleri

- *S.CARbym()*, konvolüsyon veya BYM olarak adlandırılan modelin uygulanmasında kullanılan fonksiyondur. Genel otokorelasyona sahip CAR modellerinden biri olan bu model, Eşitlik (3.18)'de tanımlanan forma sahiptir. τ^2 ve σ^2 için ters gamma

önselinde tanımlanan a ve b varsayılan değerleri sırası ile 1 ve 0.01 şeklinde tanımlanır.

- *S.CARleroux()*, Leroux et al. [14] tarafından önerilen Eşitlik (3.19)'da verilen genel düzgünleştirilmiş CAR modeli için kullanılır. Bu model otokorelasyon parametresinin 1 ve 0 değerlerini alma durumuna göre ICAR ve bağımsız rasgele etkiler modeline indirgenebilir.

Yerel Düzgünleştirilmiş CAR Modelleri

- *S.CARDissimilarity()*, Lee ve Mitchell [20] tarafından önerilen yerelleştirilmiş mekânsal otokorelasyon modeli için kullanılır. Bu modelde komşular, klasik yaklaşımdaki gibi komşu alanların benzer özellikler sergilemesinin aksine, sosyal (sigara içme oranı, gelir durumu vb.) veya fiziksel (nehir, demiryolu hattı vb.) nedenlerden dolayı farklı özellik göstermektedir. Böylece, alanlar arasındaki ilişki üzerinden benzersizlik ölçütü kullanılarak komşuluk matrisinin ağırlıkları tanımlanır ve ona göre analiz edilir.
- *S.CARlocalised()*, Lee ve Sarran [44] tarafından önerilen yerelleştirilmiş mekânsal otokorelasyon modeli için kullanılan diğer bir fonksiyondur. Benzersizlik modelinde komşuluk matrisinin ağırlıklarındaki düzeltme ile yerel mekânsal otokorelasyona izin verilir iken bazı alanlar arasında kısmi otokorelasyonda azalma meydana gelebilir. Bu gibi durumlara neden olmaması ve alanların gözlem değerleri arasında meydana gelebilecek daha büyük sıçramaları yakalayabilmesi için bu model tercih edilir.

CARBayes paketinde komşuluk matrisinin tanımlanmasında *S.CARDissimilarity* modelinin kullanımı dışında Eşitlik (3.8)'de verilen ikili bağlantı matrisi (B) kullanılmak zorunda değildir. Paket kapsamında tanımlanan diğer komşuluk matrisleri ise ikili bağlantı matrisi referans olarak düzenleme sağlamaktadır. Bunlar; standartlaştırılmış satır matrisi (R), genel standartlaştırılmış matris (C), ve Tiefelsdorf et al. [45] tarafından önerilen varyans durağanlaştırması (S) matrisleridir. R matrisi, k alanının komşuluk listesinde yer alan toplam komşu sayısının oranından elde edilir ve sınır paylaştığı tüm komşularına bu ağırlık tanımlanır. C matrisi, bölgedeki tüm bağlantıların toplamının kullanıldığı matristir. Sınır paylaşan tüm alanların mekânsal ağırlıkları, toplam alan sayısının toplam bağlantı sayısına oranından elde edilir. Başka bir matris U ise, C'ye benzer bir matristir fakat oran doğrudan toplam bağlantı sayısından elde edilir. Bölüm 4'te verilen uygulama alanında R ve C matrisleri için örnekler verilmiştir.

Çok değişkenli mekânsal analiz için *MVS.CARleroux()* ve her bölgedeki farklı birimlerde meydana gelen iki düzeyli verilerin analizinde *S.CARmultilevel()* kullanılır.

Eşitlik (3.1)'de tanımlanan regresyon parametre vektörü için çok değişkenli Normal dağılımlı önsel kullanıcı tarafından belirlenebilir. Yazılım tarafından belirlenen varsayılan değerler, sabit sıfır ortalama vektörü ve Σ_β 'nın köşegen elemanları için 100 değeridir. Ölçek parametresi ν^2 için önsel dağılım $\nu^2 \sim \text{Ters Gamma}(1; 0, 01)$ şeklinde tanımlanır.

Bu paketteki tüm modeller, Gibbs örnekleme ve Metropolis adımlarının kombinasyonu ile MCMC benzetimi kullanarak Bayes tahminlerine dayanmaktadır. Rasgele etkiler ve regresyon parametreleri için Metropolis düzeltmeli Langevin algoritması (MALA) kullanılır [46]. Rasgele etkiler için ise basit rasgele yürüyüş Metropolis adımları MALA=FALSE ile tercih edilebilir.

CARBayes paketi, istatistiksel yazılım programı R için bir eklenti paketidir ve R arşivinden ücretsiz olarak indirilebilir. Paket, 3.0.0 ve daha yüksek sürüm gerektirmekte ve **MASS** [11], **Rcpp** [12] paketlerine bağlı olarak çalışmaktadır. Ek olarak, **CARBayesdata**, **coda**, **matrixcalc**, **MCMCpack**, **spam**, **sp**, **spdep**, **stats**, **truncnorm** ve **utils** gibi paketlere de ihtiyaç duyulmaktadır. Bunların yanı sıra, **maptools**, **shapefiles**, **sp** ve **spdep** gibi paketlerin de yüklenmesi gerekir. Bu paketlerden shapefiles, mekân bilgisinin analizde kullanılması amacı ile *shapefile* yani mekân bilgisinin bulunduğu dosyanın R programına aktarılması için kullanılır. Bu *shapefile* dosyasında, alanların coğrafi koordinatları, yüz ölçümleri gibi bilgiler dbf, shp ve shx uzantılı dosyalarda yer almaktadır. Mekânsal analizde komşuluk matrisinin yaratılması veya mekânsal otokorelasyonun varlığının kontrolü için testlerin yapılması amacıyla **sp** ve **spdep** paketlerinden yararlanılır. Kestirilmiş değerlerin ve artıkların mekânsal haritalarının çıkarılması için **maptools** paketi kullanılır [47]. Ayrıca, **ggplot2** ve **rgeos** paketleri haritalandırmada kullanılan diğer yardımcı paketlerdir.

Haritalandırma yapılırken kullanılan renk paletlerinin yer aldığı **RColorBrewer** paketi, Brewer et al. [48] tarafından geliştirilmiştir. Brewer et al. [49] tarafından, Amerika Birleşik Devletleri'nde bir hastalık haritası oluşturması için yapılan çalışma öncü olmuştur. Brewer ve Pickle [50], değişken aralıklarının ve renklerinin haritada nasıl algılandığı üzerine çalışmıştır. Olson ve Brewer [51] ise renk körü insanlar için hastalık haritalamasında kullanılması amacıyla bir dizi renk paleti geliştirmiştir.

Model tanımlaması yaparken örneklem sayısı, yanma ve inceltme (thinning) miktarının tanımlanması gerekir. Yanma, başlangıç değeri etkisinin azaltılması için zincirin başlangıç

bölümünün atılması olarak ifade edilebilir. Böylece, sonsal çıkarımlar için iyi örneklemeler kullanılmış olur. İnceltme ise örneklemdeki bağımlılığı yani otokorelasyonu azaltmak için kullanılır.

Paket kapsamında oluşturulan modeller sonunda elde edilen çıktılar aşağıda tanımlanmıştır.

Çıktı tablosunda, sonsal medyan değeri, % 95 güvenirlilik aralığı, üretilen örneklem sayısı, Markov zincirleri için kabul edilme oranları ve bu zincirlerin yakınsaklığının kontrolü için Geweke değerleri yer almaktadır.

Geweke [52] tarafından geliştirilen bu değer, Markov zincirinin ilk ve son bölümünün karşılaştırılmasına dayanan bir skordur. Zincirin başından ve sonundan yüzde kaçlık bir karşılaştırma yapılacağı ile ilgili varsayılan değerler, sırasıyla % 10 ve % 50 olarak belirlenmiştir. Eğer ortalamalar birbirine yakın ise zincir yakınsaktır. Geweke değerinin skor testi olmasından ötürü, -1,96 ile 1,96 arasında yer alan test istatistikleri yakınsamayı ifade ederken, daha aşırı değerler yakınsama eksikliğini işaret etmektedir. Zincirlerin yakınsama durumunun kontrol edilebilmesi için diğer bir yol ise parametrelerin iz ve yoğunluk grafiğinin incelenmesidir. Eğer iz grafiğinde herhangi bir trend veya büyük dalgalanma yok ise yakınsama gerçekleşmiş ve hedef dağılıma ulaşılmıştır.

Model sonucunda her bir alan için sonsal medyan değerine bağlı olarak kestirilmiş değerler elde edilir. Model artıkları ise “response”, “Pearson” ve “deviance” olarak 3 farklı şekilde elde edilir. Çıktıdan elde edilen artık matrisinde her satır bir veri noktasını ifade ederken her sütunda üç farklı artık tipi yer almaktadır.

Model bilgi kriteri olarak, etkin parametre sayısı ile birlikte sapma bilgi kriteri (Deviance Information Criterion, DIC) [53], Watanabe-Akaike bilgi kriteri (WAIC) [54], LMPL (Log Marginal Predictive Likelihood) [55] ve açıklanan sapma yüzdesi (percentage deviance explained, p.d.e) verilir. En uygun model, DIC ve WAIC değerlerini minimum, LMPL’yi maksimum yapan modeldir.

3.4. Mekânsal Haritalandırma

Risk haritası, belirli bir sınır içerisinde tanımlanmış ve birbiriyle örtüşmeyen alanlardan meydana gelen bir bölge üzerinden riskin mekânsal dağılımının belirlenebilmesi için oluşturulur. Ayrıca, hangi bölgelerin riski artırdığı da bu haritalar yardımı ile incelenebilir. Mekânsal analizin yaygın olarak kullanıldığı alanlardan birisi olan epidemiyolojide hastalık

haritalarına sıkça rastlanmaktadır. Bu yüzden literatürde hastalık haritalaması olarak yerini almıştır.

Hastalık riski, risk altındaki kitle için belirli bir süre içerisinde meydana gelen hastalık veya ölüm sayılarından elde edilen oranı yansıtmaktadır. Bir bölgede meydana gelen hastalık sayısı, riski tanımlamak için yeterli olmamaktadır. Bu yüzden, hastalık riskine ilişkin bir tahmin elde etmek için gözlenen vaka sayısı ile beklenen vaka sayısı karşılaştırılmalıdır.

Herhangi bir k alanı için beklenen vaka sayısı,

$$E_k = P_k r_+ \quad (3.22)$$

şeklinde hesaplanır ve dolaylı standartlaştırma olarak adlandırılır [7]. Burada P_k , k alanında risk altındaki kitleyi ve r_+ ise genel ölüm (insidans) oranını ifade eder. Bu ölüm oranı ise

$$r_+ = \frac{Y_+}{P_+} \quad (3.23)$$

eşitliği ile hesaplanır. Y_+ , tüm vaka sayısını, P_+ ise risk altındaki tüm kitleyi temsil eder.

Hastalık riski, standartlaştırılmış ölüm oranı (Standardised Mortality Ratio, SMR) ile tahmin edilir ve k alanı için Eşitlik (3.24)'teki gibi elde edilir.

$$SMR_k = \frac{Y_k}{E_k} \quad (3.24)$$

$SMR_k > 1$ ise, yüksek düzeyde hastalık riskine sahip alanları temsil eder ve bu alanlardaki ölüm oranları beklenenden daha fazla gerçekleşmiştir. Bunun aksine, $SMR_k < 1$ ise nispeten sağlıklı alanları temsil eder ve bu alanlardaki ölüm oranının beklenenden daha düşük olduğu söylenebilir. Fakat SMR bazı nedenlerden dolayı hastalık riskinin tahmini için yetersiz kalmaktadır. İlgilenilen hastalığın nadir bir hastalık veya risk altındaki popülasyonun küçük olduğu durumlarda, eğer E_k küçük ise yüksek riskler rasgele gerçekleşebilir. Komşu alanların değerlerinden etkilenmediği için yine yüksek riskler rasgele gerçekleşebilir. Ayrıca, herhangi bir açıklayıcı değişken etkisi de içermemektedir. Tüm bu nedenlerden dolayı SMR, mekânsal dağılımı yanlış gösterebilmektedir. Çünkü, modelde hem açıklayıcı değişkenler yer almakta hem de alanlar arasındaki ilişkiyi açıklayan komşuluk bilgisi kullanılmaktadır. Bu durum karşısında genelleştirilmiş doğrusal mekânsal karma modeller bir alternatif olmaktadır. Bu modelin en genel hali Eşitlik (3.1)'de tanımlanmıştır.

Haritalamada ise model sonucunda elde edilen kestirilmiş deęerler ve beklenen vaka sayısının oranından risk haritaları oluřturulur.



4. UYGULAMA

Bu çalışmada, 2016 yılında (01.01.2016 – 31.12.2016) Türkiye'nin 81 ilinde meydana gelen deprem sayıları için modelleme ve risk haritalandırılması yapılmıştır. Bu bağlamda, modelde kullanılacak yanıt değişkeni, her bir ilde meydana gelen depremlerin büyüklüğü 2'den büyük olan deprem sayılarıdır (Ek 1). Açıklayıcı değişken ise her il için büyüklüğü 2'den büyük olan depremlerin ortalama deprem büyüklüğüdür. Veriler, Boğaziçi Üniversitesi Kandilli Rasathanesi BDTİM deprem sorgulama sisteminden sarsıntı türü sadece deprem seçilerek 26-45 boylam ve 35-42 enlemleri arasında kalan bölgeye göre elde edilmiştir [56].

Burada çalışma bölgesi S , Türkiye sınırı olarak tanımlanmış ve her bir idari bölünüş olan iller alansal birimleri oluşturmaktadır, $S = (S_1, \dots, S_{81})$. 2016 yılı içerisindeki depremler baz alındığı için modellemelerin Poisson dağılımı üzerinden yapılması gerekir. Her bir ilde meydana gelen deprem sayılarını açıklamak için ortalama deprem büyüklüğünün etkisinin yanı sıra konum etkisinin de olmasından ötürü, mekânsal analize geçilmesi gerekir. Bu bağlamda, hem açıklayıcı değişkenin etkisinin olduğu hem de mekânsal bilginin yer aldığı genelleştirilmiş doğrusal mekânsal modeller (GDMM) önem kazanmaktadır.

Yanıt ve açıklayıcı değişken için tanımlayıcı istatistikler Çizelge 4.1'de verildiği gibidir. 2016 yılında meydana gelen depremler içerisinde, Kırklareli, Sinop ve Zonguldak'ta büyüklüğü 2'den büyük olan deprem meydana gelmediği için minimum deprem sayısı ve büyüklüğü 0 olarak tanımlanmıştır. En çok deprem olan il ise büyüklüğü 2'den büyük olan 669 deprem ile Manisa ili olmuştur. Ortalama deprem sayısı 62 ve deprem büyüklüğü 2,3 olarak elde edilmiştir. Poisson dağılımının özelliklerinden biri olan ortalama ve varyans değerlerinin eşit olması durumu burada sağlanamamış, varyans değerinin ortalamadan büyük olmasından dolayı aşırı yayılım durumu söz konusu olmaktadır. Literatürde bu durumda negatif binom dağılımına dayalı analizler yapılmıştır fakat genelleştirilmiş doğrusal mekânsal modeller aşırı yayılım durumunu çözmekte ve Poisson dağılımlı olarak modellenmesine izin vermektedir [16].

Çizelge 4. 1. Tanımlayıcı istatistikler

	Minimum	Maksimum	Ortalama	Varyans
Deprem Sayısı	0	669	62	10027,2
Ortalama Deprem Büyüklüğü	0	2,7	2,3	0,2

Genelleştirilmiş doğrusal mekânsal modellere geçmeden önce, Poisson dağılımı altında veriye genelleştirilmiş doğrusal model uygulaması yapılmış ve sonuçları Çizelge 4.2’de verilmiştir.

$$Y_k \sim \text{Poisson} (\mu_k), k = 1, \dots, 81$$

$$\ln(\mu_k) = \beta_0 + \beta_1 * \text{ortalama deprem büyüklüğü}$$

Çizelge 4. 2. Poisson GDM

Parametre	Tahmin	Std. Hata	Z değeri	P değeri
β_0	1,524	0,184	8,269	<2e-16
β_1	1,104	0,076	14,413	<2e-16

AIC: 7718,6

Parametre tahminlerinin önem kontrolü yapıldığında, ortalama deprem büyüklüğü anlamlı bir değişkendir ($p < 0,05$ veya $Z > 1,96$). Ortalama deprem büyüklüğü arttıkça, deprem sayısı da artmaktadır ve aralarında aynı yönlü bir ilişki mevcuttur. Temel olarak oluşturulan bu modelde mekânsal bağımlılığın varlığının testi için artıklarının kontrol edilmesi gerekir. Bu test için ilk önce bir komşuluk matrisinin tanımlanması gerekir.

İdari bölünüşlere göre her bir il, alansal bir birim oluşturmaktadır. Mekânsal bağımlılık, birbirine komşu olan iller tarafından açıklanır ve her il için komşuluk listesi oluşturulur. Örneğin, Ankara ilinin komşuluk listesinde, Çankırı, Kırıkkale, Kırşehir, Aksaray, Konya, Eskişehir ve Bolu bulunmaktadır. İkili tanımlı komşuluk matrisi oluşturulurken de Ankara ilinin komşuluk listesinde bulunan illere 1 ($w_{kj} = 1$), bulunmayanlara 0 ($w_{kj} = 0$) değeri atanır. Komşuluk matrisi tanımından gelen, Ankara kendisinin komşusu olamayacağı için $w_{kk} = 0$ değeri atanır.

81 il için oluşturulan 81×81 boyutlu simetrik komşuluk matrisi, ortak sınıra sahip olma durumuna göre w_{kj} değerleri üzerinden Çizelge 4.3’teki gibi tanımlanmıştır.

Çizelge 4. 3. Sınır paylaşma durumuna göre ikili bağlantı matrisi ($\mathbf{W}_1$)

	<i>Adana</i>	...	<i>Ankara</i>	...	...	...	<i>Konya</i>	...	...	...	<i>Düzce</i>
<i>Adana</i>	0	.	0	.	.	.	0	.	.	.	0
$\vdots$	.	.	.	.	.	.	.	.	.	.	.
<i>Ankara</i>	0	.	0	.	.	.	1	.	.	.	0
$\vdots$	.	.	.	.	.	.	.	.	.	.	.
$\vdots$	.	.	.	.	.	.	.	.	.	.	.
<i>Konya</i>	0	.	1	.	.	.	0	.	.	.	0
$\vdots$	.	.	.	.	.	.	.	.	.	.	.
<i>Niğde</i>	1	.	0	.	.	.	1	.	.	.	0
$\vdots$	.	.	.	.	.	.	.	.	.	.	.
$\vdots$	.	.	.	.	.	.	.	.	.	.	.
<i>Düzce</i>	0	.	0	.	.	.	0	.	.	.	0

81x81

Çizelge 4.2’de verilen Poisson GDM modelinin artıkları için mekânsal otokorelasyon testi sonucu Çizelge 4.4’teki gibidir.

Çizelge 4. 4. Poisson GDM artıkları için $\mathbf{W}_1$ matrisi komşuluğunda otokorelasyon testi

Moran’ın I İstatistiği	p değeri
0,231	0,0013

“Verilerde mekânsal ilişki yoktur.” şeklindeki yokluk hipotezi permütasyon testi ile test edilmiş, artıklarda mekânsal otokorelasyon bulunduğu sonucuna ulaşılmıştır ($p < 0,05$). Bu yüzden, bu tür mekânsal bağımlılığın olduğu verilerin modellenmesinde klasik GDM’ler yetersiz kalmaktadır. Böylece genelleştirilmiş doğrusal mekânsal modellere geçiş yapılması gerekir.

Poisson GDM artıklarının mekânsal testinin sonucuna destek olması açısından, iller arasında deprem sayıları bakımından otokorelasyon varlığının testi için yanıt değişkeni incelenebilir (Çizelge 4.5).

Çizelge 4. 5. Yanıt değişkeni için $\mathbf{W}_1$ matrisi komşuluğunda otokorelasyon testi

Moran’ın I İstatistiği	p değeri
0,227	0,0016

İller arasında deprem sayıları bakımından % 5 anlamlılık düzeyinde mekânsal ilişki bulunmaktadır ($p < 0,05$). Yani her bir ildeki gözlem değerleri o alanın mekân bilgisine bağlı olarak değişmektedir. Bu yüzden, deprem sayısı ile ortalama deprem büyüklüğü arasındaki ilişkiyi incelemek için GDM’ler yeterli olmayıp mekânsal GDMM’lere geçiş yapılması gerekir. Ayrıca GDMM’ler, aşırı yayılım durumunun da üstesinden gelmektedir.

Deprem sayılarının analizinde, Eşitlik (3.19)’da tanımlanan Leroux koşullu otoregresif modeli kullanılmıştır. Model R programında tanımlanırken S.CARleroux() fonksiyonu ile 100.000 örneklem üretilmiş ve yanma olarak zincirin ilk 20.000’i analizden çıkarılmıştır. Böylece, 80.000 örneklem üzerinden analiz yapılmış ve sonuçlar Çizelge 4.6’da verilmiştir.

$$Y_k \sim \text{Poisson} (\mu_k), \quad k = 1, \dots, 81$$

$$\ln(\mu_k) = \beta_0 + \beta_1 * \text{ortalama deprem büyüklüğü} + \phi_k$$

$$\phi_k | \phi_{-k}, \mathbf{W}_1, \tau^2, \rho \sim N \left(\frac{\rho \sum_{j=1}^{81} w_{kj} \phi_j}{\rho \sum_{j=1}^{81} w_{kj} + 1 - \rho}; \frac{\tau^2}{\rho \sum_{j=1}^{81} w_{kj} + 1 - \rho} \right)$$

$$\nu^2 \sim \text{Ters Gamma}(1; 0,01)$$

$$\rho \sim \text{Tek Biçimli}(0;1)$$

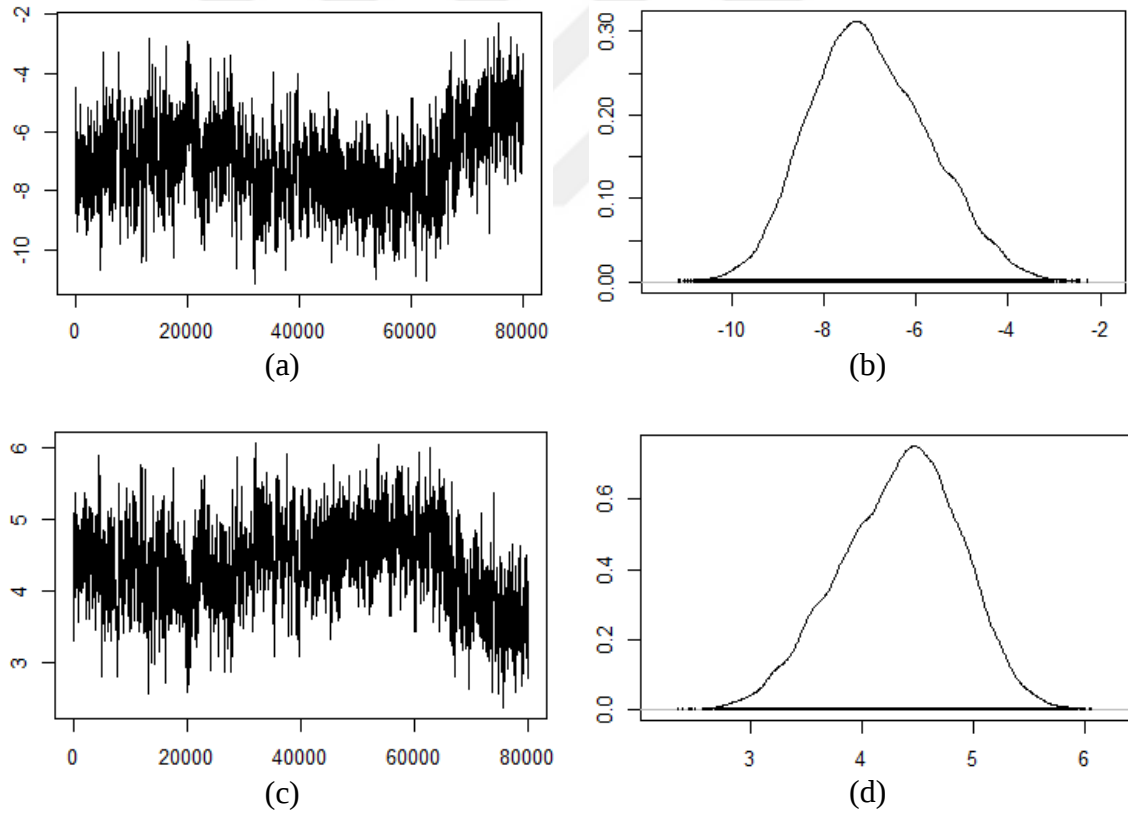
Çizelge 4. 6. Poisson mekânsal log-doğrusal model ($\mathbf{W}_1$)

Parametre	Medyan	% 95 güvenirlilik aralığı		Geweke değerleri
		% 2,5	% 97,5	
β_0	-7,088	-9,305	-4,318	0,2
β_1	4,380	3,225	5,308	-0,2
τ^2	3,465	2,011	6,065	-0,3
ρ	0,336	0,087	0,729	-0,2

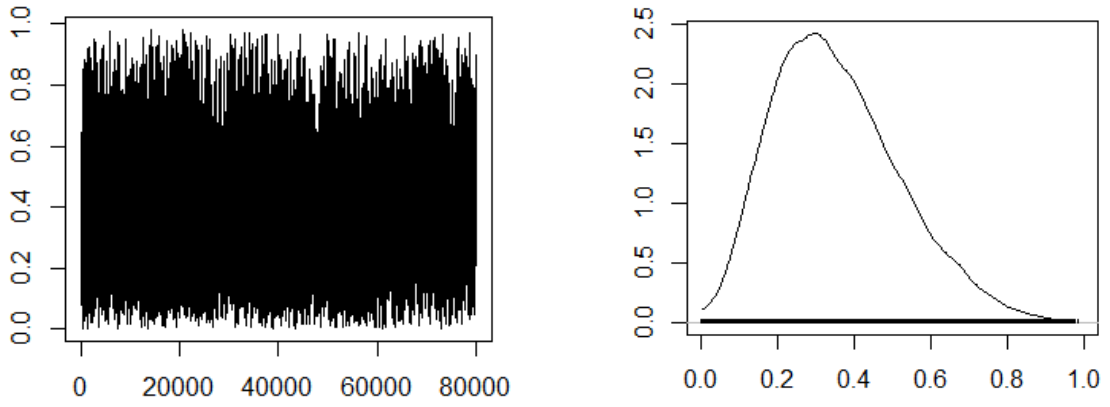
Çizelge 4.6 ‘daki medyan değeri, üretilen örneklemelerin değeridir ve parametreler için nokta tahminini ifade eder. (% 2,5 ; % 97,5) aralığı Bayesci modellemekten dolayı parametreler için güvenirlilik aralığını temsil eder. Geweke değerleri, Geweke [52] tarafından önerilen Markov zincirinin yakınsayıp yakınsamadığını kontrol etmek için kullanılan bir skor

değeridir. Standartlaştırılmış normal dağılım tablosu değerleri (Z değerleri) ile karşılaştırıldığında, (-1,96 ; 1,96) arasında kalan değerler yakınsamanın gerçekleştiğini ifade eder. Bu bağlamda, açıklayıcı değişken ortalama deprem büyüklüğü ve diğer parametrelere ait zincirlerden elde edilen Geweke değerleri bu sınırlar içerisinde kaldığı için yakınsama gerçekleşmiştir. Zincirin yakınsamasının gerçekleşip gerçekleşmediğinin kararına varılabilecek başka bir yöntem ise iz grafiğidir. Şekil 4.1, Şekil 4.2 ve Şekil 4.3'te β , ρ ve τ^2 için iz grafikleri ve örneklemelerin yoğunluk grafikleri verilmiştir. Bütün parametrelerin iz grafikleri incelendiğinde herhangi bir trende rastlanmadığı için her bir parametreye ait zincirlerin yakınsadığı sonucuna ulaşılır.

Açıklayıcı değişkenin, yanıt değişkeni üzerinde anlamlı ve pozitif bir etkisi olduğu yoğunluk grafiğinden de görülebilir (Şekil 4.1-(d)). Çünkü, bu grafikteki sonsal ortalama sıfırdan farklı bir değer almakta ve güvenilirlik aralığı sıfırı kapsamamaktadır.

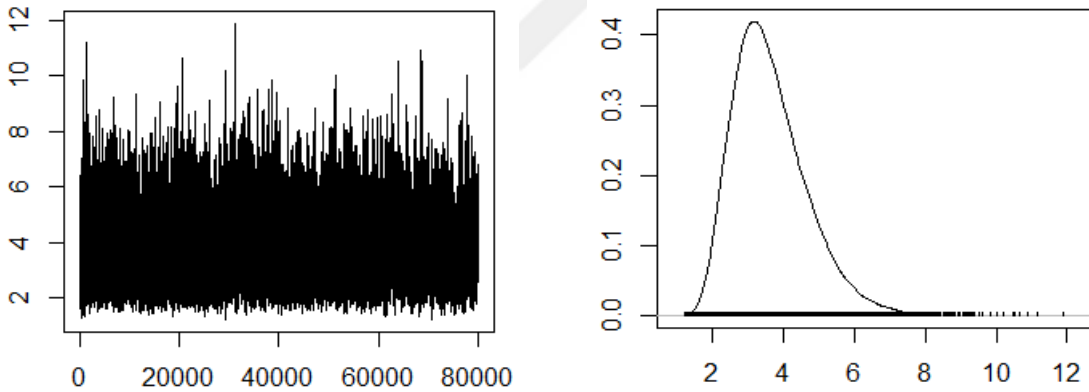


Şekil 4. 1. β_0 (a, b) ve β_1 (c, d) için iz ve yoğunluk grafiği (W_1)



Şekil 4. 2. Otokorelasyon parametresi ρ için iz ve yoğunluk grafiği ($\mathbf{W}_1$)

Şekil 4.2’de verilen otokorelasyon iz grafiği incelendiğinde yakınsamanın gerçekleştiği sonucuna ulaşılmıştır. Ayrıca, yoğunluk grafiğinde ρ parametresine ait sonsal ortalama değerinin sıfırdan büyük olması ve % 95 güvenirlilik aralığının sıfırı kapsamamasından dolayı anlamlı ve pozitif bir otokorelasyon parametresi olduğu sonucuna ulaşılır.



Şekil 4. 3. Varyans parametresi τ^2 için iz ve yoğunluk grafiği ($\mathbf{W}_1$)

Hem Geweke değeri, hem de iz grafiklerinden kontrol edilen Markov zincirlerinin yakınsamalarının sağlanmasından sonra, parametre tahminlerine geçebilmek için modelin artıklarında mekân bilgisinin varlığı test edilmelidir. Çizelge 4.7’de Poisson GDMM artıklarının sonucu incelendiğinde, mekân bilgisinin model tarafından açıklandığı ve artıkların mekân bilgisinden bağımsız olduğu sonucuna ulaşılır ($p > 0,05$).

Çizelge 4. 7. Poisson GDMM artıkları için otokorelasyon testi (W_1)

Moran'ın I İstatistiği	p değeri
0,009	0,336

Artıkların kontrolünün de yapılmasından sonra parametre tahminlerinin yorumuna geçilebilir. Poisson log-doğrusal mekânsal modelin uygulanmasında, yanıt değişkeninin ortalaması üzerinden model kurulurken *log* bağ fonksiyonu kullanıldığı için parametre tahminleri yapılırken test sonucunda elde edilen değerlerin üstelinin alınması gerekir (Çizelge 4.8).

Çizelge 4. 8. Üstel Poisson GDMM (W_1)

	Parametre	Medyan	% 95 güvenirlilik aralığı	
			% 2,5	% 97,5
Sabit Terim	e^{β_0}	0,0008	9,09e-05	0,013
Ortalama Deprem Büyüklüğü	e^{β_1}	79,838	25,151	201,945
	e^{τ^2}	31,983	7,471	430,781
	e^{ρ}	1,399	1,091	2,073

DIC=1178,635, p.d.e=88,48

β_1 'in tahmini medyan değeri pozitif geldiği için her bir ilde meydana gelen büyüklüğü 2'den büyük olan depremlerin ortalama büyüklükleri arttıkça, o ilde meydana gelen deprem sayısı da artmaktadır (Çizelge 4.8). Parametrelerin anlamlılıklarını test etmek için % 95 güvenirlilik aralıkları incelendiğinde, bütün parametrelerin anlamlı olduğu sonucuna ulaşılır (Çizelge 4.8). Ayrıca modelin açıklanabilirlik oranı (p.d.e) yaklaşık olarak % 88'dir.

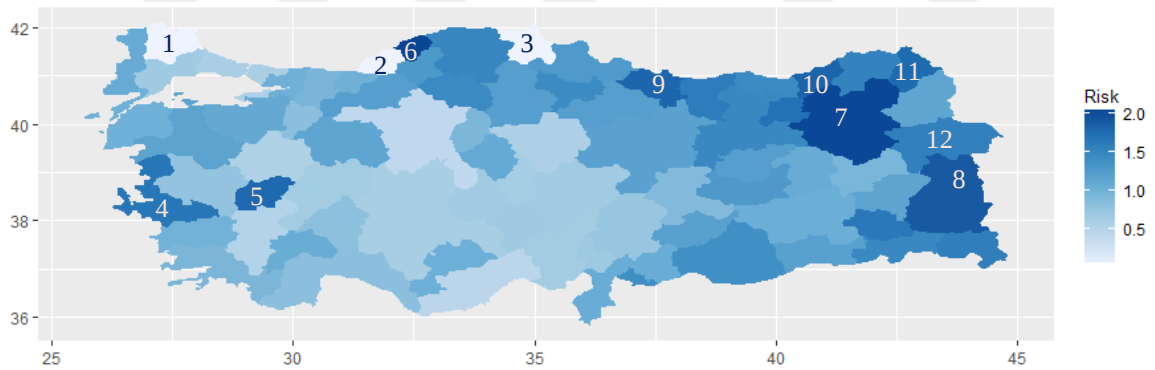
Risk haritalarının oluşturulmasında, modellerin hem açıklayıcı değişkenlere yer vermesi hem de alansal birimler arasındaki mekân bilgisinin komşuluk matrisi ile modele katılıyor olması daha tutarlı sonuçlara neden olmaktadır. Bu yüzden, Poisson GDMM'sinden elde edilen kestirilmiş değerlerin beklenen deprem sayısına bölümünden risk değerleri tahmin edilmektedir.

$$\text{Risk} = \frac{\text{Kestirilmiş değer}}{\text{Beklenen değer}} = \frac{\hat{Y}_k}{E_k}, \quad k = 1, \dots, K \quad (4.1)$$

Her bir ilde meydana gelebilecek beklenen deprem sayıları şu şekilde hesaplanmıştır;

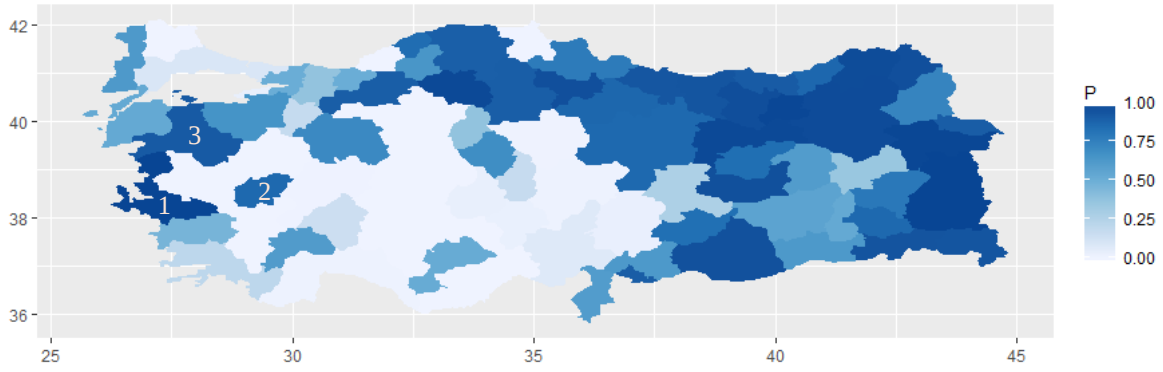
- 2016 yılında her bir ilde meydana gelen deprem sayıları ve toplam deprem sayısı (A) elde edilmiş.
- 2016 yılında her bir ilde meydana gelen büyüklüğü 2’den büyük olan deprem sayıları ve büyüklüğü 2’den büyük olan toplam deprem sayısı (B) elde edilmiş.
- $\frac{B}{A}$ oranı ile, 2016 yılında meydana gelen depremlerde, büyüklüğü 2’den büyük olanlar için bir oran elde edilmiş.
- Bu oranın her bir ilde meydana gelen deprem sayıları ile çarpımından elde edilen sayılar, o iller için büyüklüğü 2’den büyük olan beklenen deprem sayısını ifade etmektedir.

Eşitlik (4.1)’de verilen risk formülüne göre her bir il için elde edilen risk ölçümlerinin haritası Şekil 4.4’te verilmiştir. Haritalandırma **RColorBrewer** paketinde mavi rengi derecelendirme vermesi açısından farklı tonlarına göre seçilmiştir. Etiketle var olduğu üzere rengin açıktan koyuya doğru gitmesi riskin veya olasılığın arttığı anlamına gelmektedir.



Şekil 4. 4. 2016 yılında meydana gelen büyüklüğü 2’den fazla olan depremlerin riski (W_1)

Kırklareli (1), Zonguldak (2) ve Sinop (3) illerinde 2016 yılında çok fazla deprem olmadığı için bu alanların renkleri çok açık bir mavi renge sahiptir. Başka bir deyişle risksiz alanlar olarak adlandırılır. İzmir (4), Uşak (5), Bartın (6), Erzurum (7) ve Van (8) illeri birinci dereceden riskli iller olarak ifade edilebilir. 2016 yılında bu alanlarda beklenenden daha fazla deprem olmuştur. İkinci dereceden riskli grupta ise Ordu (9), Rize (10), Ardahan (11) ve Ağrı (12) illeri yer almaktadır.



Şekil 4. 5. Riski 1’den büyük olan depremlerin sonsal olasılıkları (W_1)

Risk haritasına alternatif olarak riskin belirli bir değerden büyük olduğu durumlar için sonsal olasılık haritası yapılabilir. Şekil 4.5’te verilen, riski 1’den büyük olan depremlerin yüksek olasılığı doğu illerinde gözlenmektedir. İç Anadolu bölgesi, nispeten az riskli bölge olarak tanımlanabilir. Ülkenin batı tarafında yer alan İzmir (1), Uşak (2) ve Balıkesir (3), bu bölgede deprem riski yüksek alanlar olarak değerlendirilebilir.

İkinci bir komşuluk matrisi olarak genel standartlaştırılmış matris elde edilmiştir (Çizelge 4.9). Bölüm 3.3.1’de yer verilen diğer komşuluk matrisleri (R, U ve S) simetrik olmadığı için Leroux modelinin koşulunu sağlamamaktadır. Bu yüzden modelleme yapılırken iki komşuluk matrisi (B ve C) dikkate alınarak inceleme yapılmıştır.

Çizelge 4. 9. Genel standartlaştırılmış matris (W_2)

	Adana	...	Ankara	...	...	...	Konya	...	...	...	Düzce
Adana	0	.	0	.	.	.	0	.	.	.	0
...	.	.	.	.	.	.	.	.	.	.	.
Ankara	0	.	0	.	.	.	0,203	.	.	.	0
...	.	.	.	.	.	.	.	.	.	.	.
...	.	.	.	.	.	.	.	.	.	.	.
Konya	0	.	0,203	.	.	.	0	.	.	.	0
...	.	.	.	.	.	.	.	.	.	.	.
Niğde	0,203	.	0	.	.	.	0,203	.	.	.	0
...	.	.	.	.	.	.	.	.	.	.	.
...	.	.	.	.	.	.	.	.	.	.	.
Düzce	0	.	0	.	.	.	0	.	.	.	0

81x81

Bu komşuluk matrisi oluşturulurken bütün illerin komşuluk listesi kullanılarak, tüm alansal birimler arasındaki toplam bağlantı sayısı elde edilir. Burada toplam bağlantı sayısı 398'dir. Çalışma bölgesinde 81 tane il yani alansal birim bulunmaktadır. Böylece, genel standartlaştırılmış bu matrisin ağırlıkları toplam alan sayısının toplam bağlantı sayısına oranından,

$$w_{kj} = \frac{\text{Toplam Alan Sayısı}}{\text{Toplam Bağlantı Sayısı}} = \frac{81}{398} \cong 0,203$$

şeklinde elde edilir.

Komşuluk matrisinin elde edilmesinden sonra, yanıt değişkeni için W_2 komşuluk matrisi altında otokorelasyon testi yapılır (Çizelge 4.10). Test sonucunda yine mekânsal bir bağımlılığın söz konusu olduğu görülmektedir ($p < 0,05$). Ayrıca, Moran'ın I istatistiğinin yaklaşık olarak 0.23 gelmesinden dolayı, alanlar arasında pozitif bir ilişki olduğu da söylenebilir. Bu yüzden bu komşuluk matrisi altında tekrar bir Leroux modeli kurularak parametre tahminleri yorumlanıp risk haritası oluşturulmuştur.

Çizelge 4. 10. Yanıt değişkeni için W_2 matrisi komşuluğunda otokorelasyon testi

Moran'ın I İstatistiği	p değeri
0,227	0,0015

W_2 komşuluk matrisi altında kurulan Leroux modeli ve parametreler için tanımlanan önsel dağılımlar aşağıda verilmiştir.

$$Y_k \sim \text{Poisson}(\mu_k), \quad k=1, \dots, 81$$

$$\ln(\mu_k) = \beta_0 + \beta_1 * \text{ortalama deprem büyüklüğü} + \phi_k$$

$$\phi_k | \phi_{-k}, \mathbf{W}_2, \tau^2, \rho \sim N \left(\frac{\rho \sum_{j=1}^{81} w_{kj} \phi_j}{\rho \sum_{j=1}^{81} w_{kj} + 1 - \rho}; \frac{\tau^2}{\rho \sum_{j=1}^{81} w_{kj} + 1 - \rho} \right)$$

$$\nu^2 \sim \text{Ters Gamma}(1; 0,01)$$

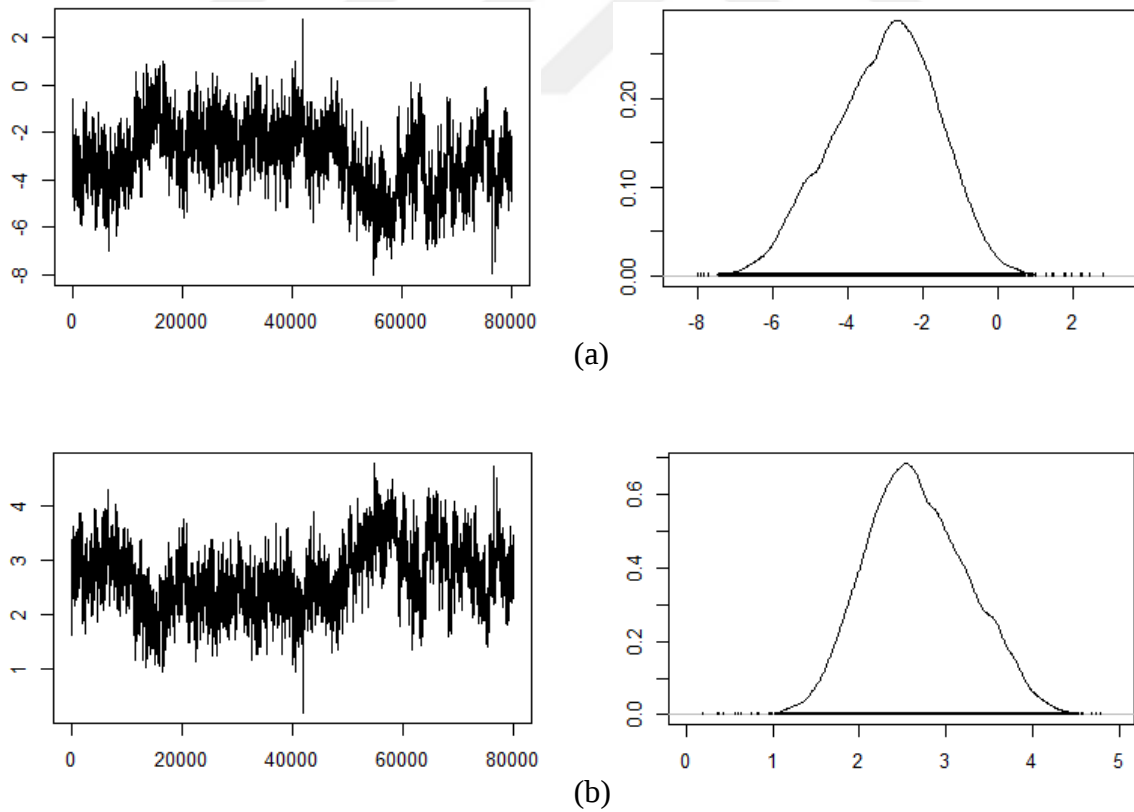
$$\rho \sim \text{Tek Biçimli}(0;1)$$

Modelin yorumlamasına geçebilmek için artıklarda mekânsal otokorelasyon olup olmadığının testi yapılır. Çizelge 4.11’de belirtilen test sonucuna göre model mekân bilgisini açıklamakta ve artıklarda herhangi bir mekânsal bağımlılık söz konusu olmamaktadır.

Çizelge 4. 11. Poisson GDMM artıkları için otokorelasyon testi (W_2)

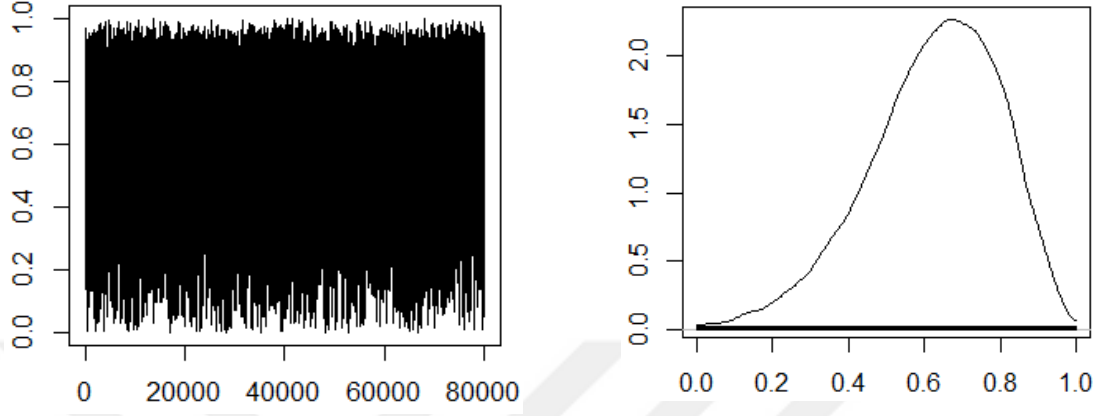
Moran’ın I İstatistiği	p değeri
-0,031	0,6825

Parametre tahminlerine geçmeden önce Markov zincirlerinin yakınsaklığına bakmak için Geweke değerleri incelenir (Çizelge 4.12). Bütün parametreler için bu değerler, (-1,96 ; 1,96) aralığına düştüğü için zincirlerde yakınsama gerçekleşmiştir. Yakınsamaya görsel bir alternatif sağlaması açısından parametrelere ait iz grafikleri incelendiğinde, bu grafiklerde herhangi bir trende rastlanmadığı için yakınsaklığın sağlandığı sonucuna ulaşılabilir (Şekil 4.6, Şekil 4.7 ve Şekil 4.8).

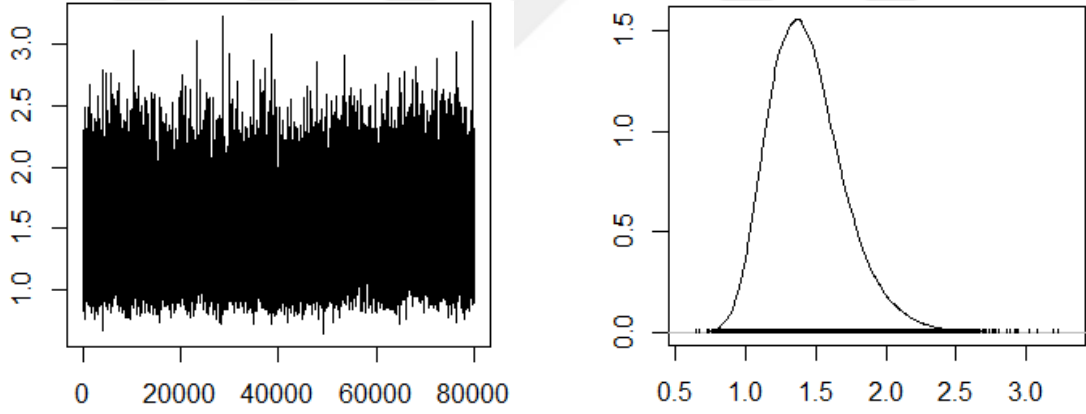


Şekil 4. 6. β_0 (a) ve β_1 (b) için iz ve yoğunluk grafiği (W_2)

Ortalama deprem büyüklüğünün katsayısı olan β_1 , yoğunluk grafiğinde sıfırdan büyük değer aldığı için ve bu yüzden güvenilirlik aralığı sıfırı içermediği için bu parametre anlamlıdır. Ayrıca, yanıt değişkeni üzerinde pozitif bir etkiye sahiptir.



Şekil 4. 7. Otokorelasyon parametresi ρ için iz ve yoğunluk grafiği ($\mathbf{W}_2$)



Şekil 4. 8. Varyans parametresi τ^2 için iz ve yoğunluk grafiği ($\mathbf{W}_2$)

Böylece parametre tahminlerine Çizelge 4.12 değerlerinin üsteli alınarak geçilebilir (Çizelge 4.13).

Çizelge 4. 12. Poisson mekânsal log-doğrusal model (W_2)

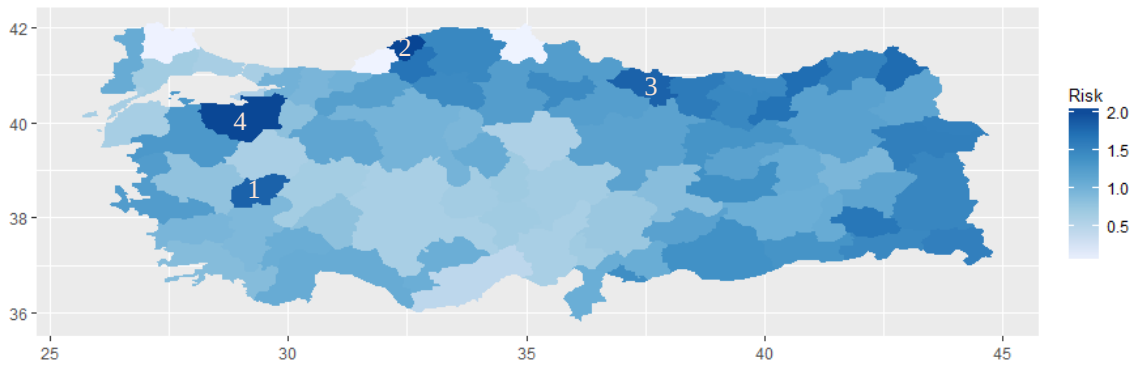
Parametre	Medyan	% 95 güvenirlilik aralığı		Geweke değerleri
		% 2,5	% 97,5	
β_0	-2,913	-5,824	-0,490	-0,6
β_1	2,631	1,613	3,8539	0,6
τ^2	1,412	0,998	2,049	-1,3
ρ	0,643	0,234	0,909	-0,3

DIC=893,415, p,d,e=91,93

Parametre tahminlerinin güvenirlilik aralığına bakıldığında bütün parametreler Poisson GDMM’de anlamlı bir etkiye sahiptir (Çizelge 4.11). W_1 matrisinden elde edilen modelin sonuçlarına paralel olarak, ortalama deprem büyüklüğü arttıkça alanlardaki deprem sayısı da artmaktadır (Çizelge 4.13). Modelin açıklanabilirlik oranı ise yaklaşık olarak % 91’dir.

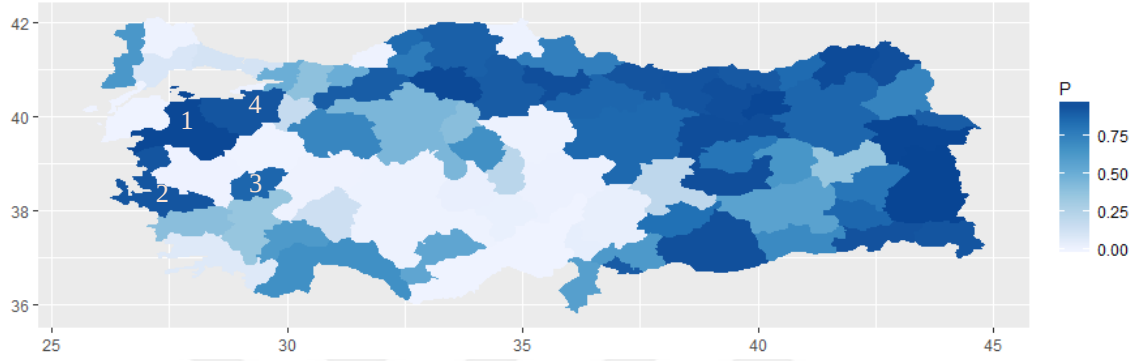
Çizelge 4. 13. Üstel Poisson GDMM (W_2)

	Parametre	Medyan	% 95 güvenirlilik aralığı	
			% 2,5	% 97,5
Sabit Terim	e^{β_0}	0,054	0,003	0,612
Ortalama Deprem Büyüklüğü	e^{β_1}	13,887	5,018	47,176
	e^{τ^2}	4,104	2,713	7,764
	e^{ρ}	1,902	1,263	2,483



Şekil 4. 9. 2016 yılında meydana gelen büyüklüğü 2’den fazla olan depremlerin riski (W_2)

Genel standartlaştırılmış komşuluk matrisi dahilinde kurulan modelden elde edilen risk tahminleri Şekil 4.9’da verilmiştir. W_1 matrisi dahilinde elde edilen risk haritasında olduğu gibi Uşak (1), Bartın (2) ve Ordu (3) riskli iller arasındadır. Balıkesir (4) ise W_2 matrisi öncülüğünde kurulan modelde yüksek riske sahiptir. Ayrıca kıyı doğu illerinde riskli alanlar da mevcuttur.



Şekil 4. 10. Riski 1’den büyük olan depremlerin sonsal olasılıkları (W_2)

Şekil 4.10’da verilen riski 1’den büyük olanların haritalaması incelendiğinde, genel anlamda doğuda bulunan illerin olasılıkları yüksek çıkmıştır. Ayrıca ülkenin kuzeyinde de bu yüksek deprem riski mevcuttur. Ülkenin İç Anadolu ve güney kesimi, W_1 ’den elde edilen model sonuçlarına nazaran daha çok risk taşıdığı anlamına gelmektedir. Batı kesimlerinde daha önceki haritalarda olduğu Balıkesir (1), İzmir (2), Uşak (3) ve ayrıca Bursa (4) yüksek deprem riski olan alanlar olarak tanımlanır.

5. SONUÇLAR

Mekânsal veri türlerinden birisi olan alansal verilerde gözlem değerleri mekâna bağlı olarak değişir. Bu da değişkenlerin gözlem değerleri arasında mekânsal otokorelasyona neden olmaktadır. Bu tür verilerin analizinde mekân bilgisinin modele katılabilmesi için alanların ilişkilerini tanımlayan komşuluk matrisinin oluşturulması gerekir. Böylece model, açıklayıcı değişken ve mekân bilgisinin yer aldığı rasgele etkilerle tanımlanır.

Normal olmayan verilerin analizinde kullanılan doğrusal modellerin genişletilmiş hali olan GDM'ler, üstel dağılım ailesine uyan verilerin analizinde kullanılır. Modelde sadece sabit etkilerin değil de rasgele etkilerin de yer alması isteniyorsa, GDKM'lere geçiş yapılması gerekir. Bu bağlamda, sayım verilerinin analizinde Poisson GDM modeli kullanılırken, alansal sayım verilerinde Poisson dağılımlı GDMM modeli kullanılır.

Tez çalışmasında 2016 yılında Türkiye'de meydana gelen büyüklüğü 2'den büyük olan deprem sayıları üzerine analiz yapılmıştır. Alansal birim olarak iller alınmış ve komşuluk matrisleri oluşturulurken idari bölünüş sınırları dikkate alınmıştır. Bu çalışmada iki komşuluk matrisi tanımlanmıştır. Birincisi ikili bağlantı matrisi, ikincisi ise genel standartlaştırılmış matristir. Veride genel mekânsal otokorelasyon testi için Moran'ın I istatistiği kullanılmıştır. Oluşturulan her iki komşuluk matrisi üzerinden kurulan permütasyon testi sonucunda deprem sayılarında mekânsal ilişki çıkmıştır. Bu yüzden, deprem sayıları ile ortalama deprem büyüklüğü arasındaki ilişki için model kurulurken GDMM'lere başvurulmuştur. Bu modelde mekân bilgisi komşuluk matrisi yardımı ile rasgele etki olarak modele katılmıştır.

GDMM'lerde, komşu alanlardaki veriler mekânsal olarak modellenirken, komşu olmayan alanlar koşullu bağımsız olarak modellenmektedir. Böylece rasgele etkiler, koşullu otoregresif önsel dağılımları ile temsil edilmiştir. CAR modelleri arasında teorik ve pratik anlamda güzel sonuçlar veren Leroux modeli parametre tahminleri ve risk haritalaması için kullanılmıştır. Parametre tahminleri, MCMC yardımı ile Bayesci yaklaşımla elde edilmiştir.

Uygulamada GDMM'lere neden ihtiyaç olduğunu göstermek amacıyla ilk önce Poisson GDM modeli kurulmuş ve parametre tahminleri anlamlı bulunmuştur. Fakat artıklarda bağımlılığın olup olmadığı test edildiğinde mekânsal ilişki çıkmıştır. Bu görüşü desteklemek amacıyla yanıt değişkeni üzerinden de mekânsal otokorelasyon testi yapıldığında mekânsal ilişkinin var olduğu sonucuna ulaşılmıştır. Bu yüzden deprem sayılarını modellemek için GDM'ler yeterli olmayıp GDMM'ler tercih edilmiştir.

İkili tanımlı komşuluk matrisi üzerinden kurulan Leroux modelinin zincirlerinde yakınsama gerçekleşmiş ve parametre tahminleri anlamlı bulunmuştur. Modelden elde edilen artıklarda mekânsal bağımlılığa rastlanmamıştır. Bu yüzden model yorumuna geçerek, büyüklüğü 2'den büyük olan depremlerde, büyüklük arttıkça deprem sayısı da artmaktadır. Modelden elde edilen kestirilmiş değerlerin beklenen deprem sayısına oranıyla risk haritası oluşturulmuştur. Çok az sayıda deprem olması nedeniyle Kırklareli, Zonguldak ve Sinop risksiz alanlar olarak tanımlanmıştır. İzmir, Uşak, Bartın, Erzurum ve Van'da beklenenden daha fazla deprem olduğu için riski yüksek alanlar olduğu sonucuna ulaşılmıştır. Bu illeri daha sonra Ordu, Rize, Ardahan ve Ağrı takip etmiştir. Ayrıca riski birden büyük alanlar için sonsal olasılık haritası yapılmıştır. Bu haritada doğu illerindeki yüksek risk oranı yoğun bir şekilde gözlenmiştir. İç Anadolu bölgesi az riskli bölge olarak tanımlanırken, İzmir, Uşak ve Balıkesir batı bölgesinde birden büyük riske sahip olma olasılığı yüksek iller olarak ifade edilebilir.

Genel standartlaştırılmış komşuluk matrisi ile kurulan modelde zincirlerin yakınsaması sağlandıktan sonra artıklar mekânsal otokorelasyon testine sokulduğunda mekânsal bağımlılığa rastlanmamıştır. Diğer modelde olduğu gibi bu modelde de parametreler anlamlı gelmiş ve ortalama deprem büyüklüğü ile deprem sayısı arasında aynı yönlü bir ilişki olduğu sonucuna ulaşılmıştır. Bu modelde elde edilen risk haritasında, kıyı doğu illerinin yanı sıra Balıkesir, Uşak, Bartın ve Ordu riskli iller olarak değerlendirilmiştir. Riski 1'den büyük olanların sonsal olasılık haritası incelendiğinde, doğu illerinde yüksek oran gözlenmiştir. Batı tarafta ise Balıkesir, İzmir ve Uşak birden büyük risk olasılığı yüksek iller olarak değerlendirilmiştir.

Uygulamada kullanılan iki komşuluk matrisi ile elde edilen Leroux modeli sonuçları incelendiğinde, ortalama deprem büyüklüğü ile deprem sayısı arasında aynı yönlü bir ilişki olduğu sonucuna ulaşılmıştır. Risk haritaları incelendiğinde ise küçük farklılıklara rağmen riskli alanlar için büyük ölçüde benzer sonuçlar elde edilmiştir.

Bundan sonra yapılacak çalışmalarda farklı modeller için farklı ağırlık matrisleri kullanılarak çözümler yapılabilir. Ayrıca zaman boyutu dikkate alınarak zamana bağlı tahminler elde edilebilir. Benzetim çalışması ile en iyi modellere ulaşılabilir.

KAYNAKLAR

- [1] Bivand, S.R., Pebesma, E., Gomez-Rubio, V., *Applied Spatial Data Analysis with R*, Second Edition, Springer, **2013**.
- [2] Biggeri, A., Dreassi, E., Catelan, D., Rinaldi, L., Laglazio, C., Cringoli, G., Disease mapping in veterinary epidemiology: a Bayesian geostatistical approach. *Stat Methods Med Res*, 15:337–52, **2006**.
- [3] Kissling, W., Carl, G., Spatial autocorrelation and the selection of simultaneous autoregressive models, *Global Ecol Biogeogr*, 17, 59–71, **2008**.
- [4] MacNab, Y., Hierarchical Bayesian modelling of spatially correlated health service outcome and utilization rates, *Biometrics*, 59, 305–16, **2003**.
- [5] Cressie, N., *Statistics for Spatial Data*, Revised Edition, John Wiley & Sons, New York, **1993**.
- [6] Schabenberger, O., Gotway, C.A., *Statistical Methods for Spatial Data Analysis*, Chapman & Hall/CRC, Boca Raton/London, **2005**.
- [7] Waller, L.A., Gotway, C.A., *Applied Spatial Statistics for Public Health Data*, John Wiley & Sons, Hoboken, NJ, **2004**.
- [8] Wall, M. M., A close look at the spatial structure implied by the CAR and SAR models, *Journal of Statistical Planning and Inference*, 121, 311–324, **2004**.
- [9] Banerjee, S., Carlin, B.P., Gelfand, A.E., *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC, Boca Raton/London, **2004**.
- [10] Fortin, M.J., Dale, M., *Spatial Analysis: A Guide for Ecologists*, Cambridge University Press, Cambridge, **2005**.
- [11] Haining, R.P., *Spatial Data Analysis: Theory and Practice*, Cambridge University Press, Cambridge, **2003**.
- [12] Lawson, A.B., Browne, W. J., Rodeiro, C.L.V., *Disease Mapping with WinBUGS and MLwiN*, Wiley, Chichester, **2003**.
- [13] Besag, J., York, J., Mollie, A., Bayesian image restoration, with two applications in spatial statistics, *Annals of the Institute of Statistical Mathematics*, 43, 1–59, **1991**.
- [14] Leroux, B., Lei, X., Breslow, N., Estimation of disease rates in small areas: a new mixed model for spatial dependence, *Statistical models in epidemiology, the environment and clinical trials*, New York, Springer-Verlag, 135–78, **1999**.
- [15] Stern, H., Cressie, N., Posterior predictive model checks for disease mapping models, *Stat Med*, 19, 2377–97, **2000**.
- [16] Lee, D., A comparison of conditional autoregressive models used in Bayesian disease mapping, *Spatial and Spatio-temporal Epidemiology*, 2, 79-89, **2011**.
- [17] Lawson, A., Clark, A., Spatial mixture relative risk models applied to disease mapping, *Statistics in Medicine*, 21, 359–370, **2002**.
- [18] Lu, H., Reilly, C., Banerjee, S., Carlin, B., Bayesian areal wombling via adjacency modelling, *Environmental and Ecological Statistics*, 14, 433–452, **2007**.

- [19] Reich, B., Hodges, J., Modeling longitudinal spatial periodontal data: a spatially-adaptive model with tools for specifying priors and checking fit, *Biometrics*, 64, 790–799, **2008**.
- [20] Lee, D., Mitchell, R., Boundary detection in disease mapping studies, *Biostatistics*, 13, 415–426, **2012**.
- [21] Nelder, J. A., Wedderburn, R.W.M., Generalized linear models, *Journal of the Royal Statistical Society, Series A*, 135, 370-384, **1972**.
- [22] McCullagh, P., Nelder, J.A., *Generalized Linear Models*, Second Edition, New York, Chapman & Hall, **1989**.
- [23] Lee, Y., Nelder, J.A., Hierarchical generalised linear models (with discussion). *Journal of the Royal Statistical Society B*, 58, 619-656, **1996**.
- [24] Lee, Y., Nelder, J.A., Pawitan, Y., *Generalized Linear Models with Random Effects*, Chapman & Hall, **2006**.
- [25] Spinelli, J.J., Lockhart, R.A., Stephens, M.A., Tests for the response distribution in a Poisson regression model, *Journal of Statistical Planning and Inference*, Volume 18, Issues 1-2, 137-154, **2002**.
- [26] O’Sullivan, D., Unwin, D.J., *Geographical Information Analysis*, Wiley, Hoboken, NJ, 200-204, **2003**.
- [27] Ward, M.D., Gleditsch, K.S., *Spatial regression models*, Sage, Thousand Oaks, CA, 14-22, **2008**.
- [28] Chun, Y., Griffith, D.A., *Spatial Statistics & Geostatistics*, Sage, Thousand Oaks, CA, 57-59, **2013**.
- [29] Cliff, A.D., Ord, J.K., *Spatial Processes: Models and Applications*, Pion, London, **1981**.
- [30] Haining, R., *Spatial Data Analysis in the Social and Environmental Sciences*, Cambridge, Cambridge University Press, **1990**.
- [31] Bailey, T.C., Gatrell, A.C., *Interactive Spatial Data Analysis*, England, Addison Wesley Longman, **1995**.
- [32] Griffith, D., What is spatial autocorrelation?, *L’Espace géographique* 21, 265–280, **1992**.
- [33] Cliff, A.D., Ord, J.K., *Spatial Autocorrelation*, Pion, London, **1973**.
- [34] Upton, G.J.G., Fingleton, B, *Spatial data analysis by example, Vol. I: Point Pattern and Quantitative Data*, New York, John Wiley & Sons, **1985**.
- [35] Tiefelsdorf, M., Boots, B., The exact distribution of Moran’s I, *Environment and Planning A* 27, 985–999, **1995**.
- [36] Besag, J., Newell, J., The detection of clusters in rare diseases, *Journal of the Royal Statistical Society, Series A* 154, 327–333, **1991**.
- [37] Lee, D., Jack, E., Modelling spatial data in R with CARBayes - Part 1: Introduction and exploratory analysis, GEOMED, 07-09 September, Porto, Portugal, **2017**.
- [38] Belitz, C., Brezger, A., Kneib, T., Lang, S., BayesX - Software for Bayesian Inference in Structured Additive Regression Models, **2009**.

- [39] Vieilledent, G., Merow, C., Guélat, J., Latimer, A.M., Kéry, M., Gelfand, A.E., Wilson, A.M., Mortier, F., Silander, J.A., hSDM: hierarchical Bayesian species distribution models, R package version 1.4, <https://CRAN.R-project.org/package=hSDM>, **2014**.
- [40] Schabenberger, H., spatcounts: Spatial count regression, R package version 1.1, <http://CRAN.R-project.org/package=spatcounts>, **2009**.
- [41] Bivand, R., spdep: Spatial Dependence: Weighting Schemes, Statistics and Models. R package version 0.5-56, <http://CRAN.R-project.org/package=spdep>, **2013**.
- [42] Rue, H., Martino, S., Chopin, N., Approximate bayesian inference for latent Gaussian models using integrated nested laplace approximations (with discussion), *Journal of the Royal Statistical Society Series B*, 71, 319–392, **2009**.
- [43] Lee, D., Rushworth, A., Napier, G., CARBayesST: spatio-temporal generalised linear mixed models for areal unit data, R package version 2.4, **2016**.
- [44] Lee, D., Sarra, C., Controlling for unmeasured confounding and spatial misalignment in long-term air pollution and health studies, *Environmetrics*, 26, 477–487, **2015**.
- [45] Tiefelsdorf, M., Griffith, D. A., Boots, B., A variance-stabilizing coding scheme for spatial link matrices, *Environment and Planning A*, 31, 165–180, **1999**.
- [46] Roberts, G., Rosenthal, J., Optimal scaling of discrete approximations to the Langevin diffusions, *Journal of the Royal Statistical Society Series B*, 60, 255–268, **1998**.
- [47] Lee, D., CARBayes: An R Package for Bayesian Spatial Modeling with Conditional Autoregressive Priors, *Journal of Statistical Software*, Volume 55, Issue 13, **2013**.
- [48] Brewer, C.A., Hatchard, G.W., Harrower, M.A., ColorBrewer in print: a catalog of color schemes for maps. *Cartography and Geographic Information Science*, 30, 5–32, **2003**.
- [49] Brewer, C.A., MacEachren, A.M., Pickle, L.W., Herrmann, D.J., Mapping mortality: Evaluating color schemes for choropleth maps, *Annals of the Association of American Geographers*, 87, 411–438, **1997**.
- [50] Brewer, C.A., Pickle, L., Comparison of methods for classifying epidemiological data on choropleth maps in series, *Annals of the Association of American Geographers*, 92, 662–681, **2002**.
- [51] Olson, J.M., Brewer, C.A., An evaluation of color selections to accommodate map users with color-vision impairments, *Annals of the Association of American Geographers*, 87, 103–134, **1997**.
- [52] Geweke, J., Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, *IN BAYESIAN STATISTICS*, University Press, 169-193, **1992**.
- [53] Spiegelhalter, D., Best, N., Carlin, B., Van der Linde, A., Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society Series B*, 64, 583–639, **2002**.
- [54] Watanabe, S., Asymptotic equivalence of the Bayes cross validation and widely applicable information criterion in singular learning theory, *Journal of Machine Learning Research*, 11, 3571–3594, **2010**.

- [55] Congdon, P., *Bayesian models for categorical data*, 1st edition, John Wiley & Sons, **2005**.
- [56] B.Ü. KRDAE Bölgesel Deprem-Tsunami İzleme ve Değerlendirme Merkezi, <http://www.koeri.boun.edu.tr/sismo/zeqdb/>



EKLER

Ek 1. 2016 yılında meydana gelen büyüklüğü 2’den büyük olan deprem sayıları

Adana	80	Edirne	21	Kütahya	234	Uşak	2
Adıyaman	31	Elazığ	102	Malatya	97	Van	326
Afyonkarahisar	37	Erzincan	120	Manisa	669	Yozgat	16
Ağrı	49	Erzurum	111	Kahramanmaraş	44	Zonguldak	0
Amasya	25	Eskişehir	18	Mardin	2	Aksaray	19
Ankara	215	Gaziantep	42	Muğla	352	Bayburt	22
Antalya	252	Giresun	11	Muş	38	Karaman	8
Artvin	25	Gümüşhane	24	Nevşehir	3	Kırıkkale	15
Aydın	48	Hakkari	12	Niğde	24	Batman	5
Balıkesir	241	Hatay	27	Ordu	7	Şırnak	18
Bilecik	22	Isparta	26	Rize	2	Bartın	1
Bingöl	109	İçel	57	Sakarya	33	Ardahan	10
Bitlis	55	İstanbul	5	Samsun	9	Iğdır	9
Bolu	60	İzmir	217	Siirt	4	Yalova	33
Burdur	46	Kars	22	Sinop	0	Karabük	1
Bursa	110	Kastamonu	10	Sivas	70	Kilis	17
Çanakkale	120	Kayseri	31	Tekirdağ	9	Osmaniye	40
Çankırı	57	Kırklareli	0	Tokat	34	Düzce	15
Çorum	61	Kırşehir	41	Trabzon	15		
Denizli	136	Kocaeli	5	Tunceli	40		
Diyarbakır	23	Konya	59	Şanlıurfa	51		

ÖZGEÇMİŞ

Kimlik Bilgileri

Adı Soyadı : Leyla BAKACAK

Doğum Yeri : Ankara

Medeni Hali : Bekâr

E-posta : leylabakacak@hacettepe.edu.tr

Adresi : Hacettepe Üniversitesi, İstatistik Bölümü, Beytepe, Çankaya, Ankara

Eğitim

Lisans : Orta Doğu Teknik Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü

Yabancı Dil ve Düzeyi

İngilizce : YÖKDİL (82,5)

İş Deneyimi

Araştırma Görevlisi (2016- halen), Hacettepe Üniversitesi, İstatistik Bölümü

Deneyim Alanları

-

Tezden Üretilmiş Projeler ve Bütçesi

-

Tezden Üretilmiş Yayınlar

-

Tezden Üretilmiş Tebliğ ve/veya Poster Sunumu ile Katıldığı Toplantılar

-



HACETTEPE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
YÜKSEK LİSANS TEZ ÇALIŞMASI ORJİNALLİK RAPORU

HACETTEPE ÜNİVERSİTESİ
FEN BİLİMLER ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI BAŞKANLIĞI'NA

Tarih: 18/06/2018

Tez Başlığı / Konusu: Genelleştirilmiş Doğrusal Mekânsal Modellere Koşullu Otoregresif Model Yaklaşımı

Yukarıda başlığı/konusu gösterilen tez çalışmamın a) Kapak sayfası, b) Giriş, c) Ana bölümler d) Sonuç kısımlarından oluşan toplam 47 sayfalık kısmına ilişkin, 18/06/2018 tarihinde şahsım/tez danışmanım tarafından Turnitin adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı % 2 'dir.

Uygulanan filtrelemeler:

- 1- Kaynakça hariç
- 2- Alıntılar hariç/dâhil
- 3- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü Tez Çalışması Orjinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve bu Uygulama Esasları'nda belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

18.06.2018

Adı Soyadı: Leyla BAKACAK

Öğrenci No: N16124317

Anabilim Dalı: İstatistik

Programı: İstatistik

Statüsü: ☒ Y.Lisans ☐ Doktora ☐ Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.

Prof. Dr. Serpil AKTAŞ ALTUNAY