ANADOLU ÜNİVERSİTESİ BİLİM VE TEKNOLOJİ DERGİSİ –A
Uygulamalı Bilimler ve Mühendislik

ANADOLU UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY –A
Applied Sciences and Engineering
Cilt/Vol.:14-Sayı/No: 1 : 47-54 (2013)

ARAŞTIRMA MAKALESİ / **RESEARCH ARTICLE**

SIRALI ZAMAN SERİSİ VERİLERİNİN ORANTILI ODDS MODELİ İLE MODELLENMESİ

Esra SATICI[1], Serpil AKTAŞ ALTUNAY[2]

ÖZ

Zamana bağlı açıklayıcı değişkenlere sahip kategorik zaman serileri, bağımlı değişken kategorik olduğu durumda ortaya çıkar. Kategorik zaman serileri analizi için litaratürde bir çok yöntem önerilmiştir, bunlardan bazıları Markov zincirleri modeli, tamsayılı otoregresif süreçler, kesikli ARMA modeli gibi yöntemlerdir. Genellikle modelin seçimi, ilgilenilen değişkenin sınıflayıcı, sıralı veya aralıklı ölçümüne bağlıdır. Bununla birlikte, kategorik zaman serileri analizi için genel doğrusal modellere dayalı ve kısmi olabilirlik çıkarımlı regresyon teorisi başarılı bir yaklaşımdır. Regresyon teorisinde sıralı zaman serisi modellerinden biri orantılı odds modeldir. Bu çalışmada, sıralı kategorik zaman serisi modeli için orantılı odds modeli tanıtılmış ve gerçek hava kalitesi veri kümesi üzerinde uygulama yapılarak sonuçlar tartışılmıştır.

Anahtar Kelimeler: Kategorik zaman serileri, Sıralı regresyon, Orantılı odds modeli.

MODELLING OF ORDINAL TIME SERIES BY PROPORTIONAL ODDS MODEL

ABSTRACT

Categorical time series data with random time dependent covariates often arise when the variable categories are assigned as categorical. There are several other models that have been proposed in the literature for the analysis of categorical time series. For example, Markov chain models, integer autoregressive processes, discrete ARMA models can be utilized for modeling of categorical time series. In general, the choice of model depends on the measurement of study variables: nominal, ordinal and interval. However, regression theory is successful approach for categorical time series which is based on generalized linear models and partial likelihood inference. One of the models for ordinal time series in regression theory is proportional odds model. In this study, proportional odds model approach to ordinal categorical time series is investigated based on a real air pollution data set and the results are discussed.

Keywords: Categorical time series, Ordinal regression, Proportional odds model.

[1,] Karayolları Genel Müdürlüğü, Strateji Geliştirme Dai. Başk., Yücetepe, Ankara.
E-mail: esra.satici@gmail.com
[2,] Hacettepe Üniversitesi, İstatistik Bölümü, Beytepe, Ankara.

Geliş: 24 Ağustos 2011; **Düzeltme:** 08 Şubat 2013 ; **Kabul:** 12 Mart 2013

1. GİRİŞ

Kronolojik sırayla elde edilen verilere sahip değişkenlere zaman serisi adı verilir. Zaman serisi verileri gözlendikleri aralıklara göre isimlendirildikleri gibi (yıllık, aylık, günlük, vb.) bağımlı değişkenin ölçüm türüne göre de isimlendirilmektedir. Bağımlı değişkenin kategorik yapıda ve sıralı ölçek türünde olduğu durumda sıralı kategorik zaman serisinden bahsedilmektedir. Günlük hayatta ilgilenilen bir çok değişken sıralı ölçekte olabilir. Örneğin saatlik takip edilen hastanın kan basıncı seviyesi (düşük, orta, yüksek), uyurken kişinin beyin dalga seviyeleri (uyanık, yarı uyanık, uyuyor) yada belirli periyotlarda yapılan memnuniyet anketlerinde incelenen memnuniyet dereceleri (zayıf, orta, yüksek) gibi.

Son otuz yılda kategorik zaman serilerinin modellenmesi için markov zincirleri modeli, integer otoregresif süreçler, kesikli ARMA modelleri gibi çok sayıda farklı yöntem geliştirilmiştir. Bununla birlikte kategorik zaman serilerinin modellemesinde genel doğrusal model teorisinden faydalanılmaktadır. Sıralı zaman serilerinin modellenmesi için kullanılan genel doğrusal modellerden birisi de orantılı odds modeldir.

Lojistik regresyon ve Loglinear modeller gibi kategorik veri analiz yöntemleri ilk olarak 1960 ve 1970'lerde geliştirilmiştir. Sıralı cevap değişkenleri için ise ilk ciddi çalışma McCullagh (1980) tarafından birikimli olasılıkların logit modellemesi üzerine yapılmış ve birikimli odds modelleri sunulmuştur. McCullagh (1980) tarafından sıralı cevap değişkenlerinin regresyon modellemesi üzerinde yapılan bu çalışmadan sonra, çoğunlukla orantılı odds model olarak anılan birikimli logit modeller üzerine popüler çalışmalar yapılmıştır. Fokianos ve Kedem (2003), cevap değişkeninin ve açıklayıcı değişkenlerin gecikme terimlerini ekleyerek orantılı odds modeli kategorik zaman serileri analizi için uyarlamışlardır. Bru *et.al.* (2007), sıralı zaman serisi analizi için Fokianos ve Kedem (2003) tarafından sunulan regresyon modelini ve Jacobs ve Levis (1978) tarafından sunulan kesikli otoregresif modeli, ekolojik veriler üzerine uygulamış, her iki modelinde avantaj ve dezavantajlı yönlerini uygulama üzerinde tartışmıştır.

Bu çalışmada ilk olarak sıralı kategorik zaman serileri analizi için Orantılı Odds Modeli incelenmiş daha sonra günlük ortalama değerleri cinsinden derlenen hava kalitesi ölçüm verileri üzerine uygulaması yapılmış ve sonuçlar tartışılmıştır.

2. ORANTILI ODDS MODEL (OOM)

Birikimli odds model olarakda adlandırılan Orantılı Odds Model (OOM), sıralı bağımlı değişkenin, kesikli ve sürekli değişkenler (açıklayıcı değişkenler) ile modellenmesi için kullanılan, genel doğrusal modellerin bir sınıfı olarak tanımlanabilir (Agresti,2002). Örneğin Y, 1,…,k (k≥2) sınıflı cevap değişkeni ve $\gamma_j = P(Y \leq j|x)$ birikimli cevap olasılıkları olmak üzere, j. birikimli cevap olasılığı için doğrusal lojistik modelin en genel formu aşağıdaki gibidir,

$$logit(\gamma_j) = \alpha_j - \beta_j' X, \qquad (1)$$

burada,α_j, j. kategoriye bağlı kesim parametresi (intercept), ya da eşik (threshold) parametresi; β_j ise j. kategoriye bağlı regresyon ya da eğim katsayısı (konum parametresi olarak da adlandırılır) parametreleridir. OOM'de kesme parametresi j. kategoriye bağlıdır fakat eğimler yani regresyon katsayıları cevap kategorisinden bağımsızdır. Yani OOM, bağımsız değişkenlerin etkisini farklı düzeylerde sabit olmasını sağlamak için kolaylaştırılmış bir modeldir. Buna göre model yapısı,

$$logit(\gamma_j) = \alpha_j - \beta' X, \qquad j=1,\ldots,k\text{-}1. \qquad (2)$$

olur (Liu ve Agresti, 2005). Bu ifade grafiksel olarak,$\alpha_1,\ \alpha_2,\ \alpha_3, \ldots, \alpha_{k-1}$ kesim parametreleri ile x'e karşılık (k-1) tane doğrunun paralel olması anlamına gelmektedir. Bu nedenle, OOM'nin uygulanabilmesi için,"bağımsız değişkenler ortak bir eğim parametresine sa-

hiptir" varsayımının (paralellik varsayımı) sağlanması gereklidir.

Açıklayıcı değişkenler üzerinde gizli (latent) değişkenin bağımlılığı, doğrusal veya doğrusal olmayan modeller ile gösterilebilir (McCullagh, 2005). Doğrusal model yapısında, ε, F birikimli dağılım fonksiyonuna sahip bir raslantı değişkeni olmak üzere, $Z = \beta' X + \varepsilon$ olur. Gizli değişken ve cevap arasındaki ilişki aşağıdaki yapıdadır,

$$\gamma_j = P(Y \leq j) = P\left(Z \leq \alpha_j\right) = F(\alpha_j - \beta' X) \quad (3)$$

veya doğrusal formda ifade edilecek olursa

$$F^{-1}\left(\gamma_j\right) = \alpha_j - \beta' X \quad (4)$$

olur. Eğer, $F(z) = e^z/(1 + e^z)$ ise yani F lojistik dağılıyorsa, Eş.(4)'de gösterilen yapı OOM olarak adlandırılmaktadır. F'nin farklı dağılımları için aynı şekilde farklı genel doğrusal modellere ulaşılır (örneğin ε normal dağılıyorsa Probit modeli, exp(ε) üstel veya Weibull dağılıma sahip ise Orantılı Hazards Modeli veya tamamlayıcı log-log model gibi). Buna göre, OOM için olasılıklar birikimli olarak aşağıdaki eşitliğe göre elde edilir,

$$P(Y \leq j) = \frac{1}{1+\exp[-(\alpha_j - \beta X)]} \quad (5)$$

OOM'nin anlamı, $Y \leq j$ olayının oddsunu modelden elde etmesine imkan sağlamasından gelmektedir. Buna göre $Y \leq j$ için odds,

$$odds(Y \leq j|x) = \exp(\alpha_j - \beta^T x) \quad (6)$$

ile elde edilir. Buna benzer olarak x_0 ve x_1 için $Y \leq j$ olayının odds oranı ise aşağıdaki eşitlik ile ifade edilir,

$$\frac{odds\ (Y \leq j|x_1)}{odds\ (Y \leq j|x_0)} = \exp(-\beta^T(x_1 - x_0)). \quad (7)$$

Açıklayıcı değişken değerlerine karşılık gelen ilgilenilen değişkene ait odds miktarı ise

$$\exp(-\beta). \quad (8)$$

eşitliği ile elde edilir (McCullagh, 2005).

3. UYGULAMA

Bu bölümde, sıralı kategorik zaman serisi yaklaşımlı regresyon modeli, Çevre, Orman ve Şehircilik Bakanlığı'na bağlı hava kalitesi izleme istasyonları web sitesinden (www.havaizleme.gov.tr) alınan ölçüm verileri üzerine uygulanmıştır. Diğer kirleticiler ile birlikte tepkileşme gösterdiğinden beş temel kirleticiden en önemli ikinci kirletici olarak kabul edilen Ozon (O_3) bakımından hava kalitesi indeksi OOM ile modellenmiştir.

Yaşamımızın en önemli kaynağı olan havanın kalitesi, hayatımızın kalitesini direkt olarak etkileyen en önemli unsurlardan biridir. Hava kirliliğinin doğru bir şekilde ölçülmesi, hava kirliliği ile ilgili önlemler alınarak daha iyi duruma getirilebilmesi amacıyla bakanlığa bağlı hava kirliliği ölçüm istasyonlarında otomatik olarak ölçümler yapılmakta aynı zamanda toplanan veriler ilgili internet sitesinde yayınlanmaktadır.

Bu çalışmada, Ankara-Sincan istasyonunun 31/05/2008-29/09/2008 yaz dönemine ait günlük ölçüm verileri esas alınmıştır. Bu istasyonda otomatik olarak, sicaklık (SIC), basınç (BP), azot dioksit (NO2) ve nem oranı (RH) ölçümleri yapılmaktadır.

Köln Üniversitesi-Rhen Çevre Araştırmaları Enstitüsü tarafından belirlenen Ozon (O_3) için belirlenmiş hava kalitesi indeksi sınıfları için sınır değerleri Tablo 1'de verilmiştir.

Tablo 1'de verilen sınır değerleri esas alınarak elde edilen ozon bakımından hava kalitesi indeksi için, sıcaklık, basınç, azot dioksit ve nem oranı açıklayıcı değişkenleri ve gecikme terimleri kullanılarak, birbirinden farklı 50'den fazla OOM incelenmiştir. Bu modellerden, OOM'nin kullanılabilmesi için sağlanması gereken paralellik varsayımını sağlayan ve aynı zamanda değerlendirilen değişkenlerin modele katkılarının anlamlı olduğu, uyum iyiliği hipotezinin red edilemediği (p>0,05) modellere ilişkin, uyum iyiliği testi sonuçları, serbestlik dereceleri ve Akaike Bilgi Kriteri(ABK) değerleri Tablo 2'de verilmiştir.

Tablo 1. Ozon (O_3) için belirlenmiş hava kalitesi indeksi sınır değerleri

Sınıf	Ozon 24 Saatlik Ortalama (mg/m^3)
Çok iyi	<17
İyi	17-33
Yeterli	34-60
Orta	61-90
Kötü	91-120
Çok kötü	>120

Tablo 2. Paralellik varsayımını sağlayan ve açıklayıcı değişkenlerin katkılarının anlamlı olduğu OOM'lerin karşılaştırılması

Model	Açıklayıcı Değişkenler	Serbestlik Derecesi	Ki-Kare Değeri	ABK
1	$1 + Y_{t-2} + SIC_{t-1} + BP_{t-1} + NO2_{t-1}$	350	143,288	-556,712
2	$1 + Y_{t-2} + BP_{t-1} + NO2 + NO2_{t-1}$	347	136,981	-557,019
3	$1 + Y_{t-2} + SIC + BP_{t-1} + NO2 + NO2_{t-1}$	352	133,473	-570,527
4	$1 + NO2 + RH$	358	129,876	-586,124
5	$1 + Y_{t-2} + SIC_{t-1} + RH + NO2$	353	110,153	-595,847
6	$1 + Y_{t-1} + SIC + RH + NO2$	356	108,389	-603,611
7	$1 + Y_{t-2} + SIC_{t-1} + RH + BP_{t-1} + NO2$	352	98,461	-605,539
8	$1 + Y_{t-2} + SIC + BP_{t-1} + \mathrm{NO2}$	352	92,352	-611,648
9	$1 + Y_{t-1} + Y_{t-2} + SIC + RH + BP_{t-1} + NO2$	351	89,494	-612,506
10	$1 + Y_{t-2} + SIC + RH + BP_{t-1} + NO2 + NO2_{t-1}$	351	85,204	-616,796

Tablo 2'de açıklayıcı değişkenlerin alt indisinde verilen t-1 ve t-2, ilgili değişkenlerin sırasıyla birinci ve ikinci gecikmeli serilerini göstermektedir. Tablo 2'de ABK değerlerinden görüldüğü gibi, zamana bağlı gözlenen hava kalitesi düzeylerinin modellenmesinde, gecikmeli serilerin dahil edilmesi hatayı azaltmıştır.

Paralellik varsayımını sağlayan (p=0,999) ve açıklayıcı değişkenlerin katkılarının anlamlı olduğu modeller arasında ABK değeri en düşük olan onuncu Model incelemeye esas alınmıştır. Buna göre, ozon bakımından hava kalitesi düzeyi tahmininde iki gün önceki hava kalitesi düzeyi (bir gün önceki hava kalitesi düzeyide modele dahil edilmiş fakat bu model varsayımları sağlamadığı için değerlendirilmemiştir), sıcaklık, nem oranı, bir gün önceki basınç, azot dioksit ve bir gün önceki azot dioksit değerleri etkilidir. Model 10 için parametre tahminleri, standart hataları ve önem düzeyleri Tablo 3'de verilmiştir. Ele alınan veri kümesinde dört ozon kategorisi gözlendiği için bunlara ilişkin eşik değerleri (k-1=3) yer almaktadır.

Tablo 3. Model 10'a ait parametre tahminleri, standart hataları ve önem düzeyleri

			Tahmin	Standart Hata	Önem Düzeyi (p)
Eşik (Threshold) Parametreleri	[Ozon=1]	α_1	-435,706	120,647	0,000
	[Ozon=2]	α_2	-428,397	119,951	0,000
	[Ozon=3]	α_3	-418,53	119,073	0,000
Konum (Location) Parametreleri	SIC	β_1	-0,535	0,157	0,001
	NO2	β_2	-0,289	0,066	0,000
	RH	β_3	-0,374	0,075	0,000
	Y_{t-2}	β_4	2,207	0,564	0,000
	BP_{t-1}	β_5	-0,415	0,124	0,001
	$NO2_{t-1}$	β_6	-0,109	0,043	0,012

Tablo 3'deki sonuçlara göre hem kesim, hem de eğime göre elde edilen tüm parametre tahminlerinin istatistiksel olarak anlamlı olduğu görülmektedir ($p<0.05$).

Ayrıca Tablo 3'de verilen parametre tahminlerine göre, açıklayıcı değişkenlerin yaz dönemine ait aşağıda verilen ortalama değeri için birikimli olasılık tahminleri, Eş.(5) kullanılarak aşağıdaki gibi elde edilmiştir:

	SIC	RH	NO2	BP_{t-1}	Y_{t-2}	$NO2_{t-1}$
Ortalama:	25,4656	46,2746	28,2541	933,9008	3,0833	28,2562

$$P(Y \leq 1) = \frac{1}{1+\exp\{-[-435{,}706-(-0{,}535*25{,}4656-0{,}289*28{,}2541-0{,}374*46{,}2746+2{,}207*3{,}0833-0{,}415*933{,}9008-0{,}109*28{,}2562)]\}}$$

$$= 0{,}000003$$

$P(Y \leq 2) = 0{,}004$
$P(Y \leq 3) = 0{,}987.$

Birikimli olasılık tahminlerinden her bir hava kirliliği düzeyine ait olasılık değerleri,

Çok iyi → P(Y=1)=0
İyi → P(Y=2)=0,004
Yeterli → P(Y=3)=0,983
Orta→ P(Y=4)=0,013

olarak hesaplanır. Buna göre yeterli düzeyde olan ozon bakımından hava kalitesinin, yaz dönemindeki sıcaklık, basınç, azot dioksit ve nem oranı değerleri düşünüldüğünde, gün aşırı olarak iyi yönde iyileşmesi beklenmemekte, %99 ihtimalle düzeyini koruyacağı düşünülmektedir.

Daha öncede bahsedildiği gibi OOM'in uygulamada ki zenginliklerinden biri odds değerlerini yorumlamaya imkan vermesidir. Açıklayıcı değişkenlere ait odds değerleri Eş.(8) kullanılarak aşağıdaki gibi elde edilmiştir:

SIC : exp(0,0535) = 1,71 Hava sıcaklığındaki 1 birimlik artış, hava kirliliğinin 1 kademe azalması olasılığını 1,71 kat etkilemektedir.

NO2 : exp(0,289)=1,34 Havadaki azot dioksit miktarındaki 1 birimlik artış, hava kirliliğinin 1 kademe azalması olasılığını 1,34 kat etkilemektedir.

RH :exp(0,374)=1,45 Havadaki nem oranındaki %1'lik değişim, hava kirliliğinin 1 kademe artması olasılığını 1,45 kat etkilemektedir.

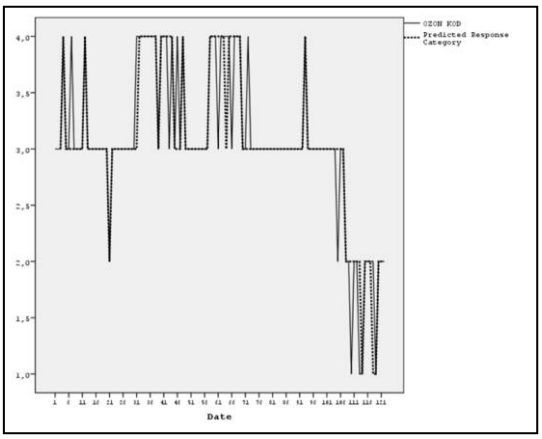
Y_{t-2} : exp(-2,207)=0,11 İki gün önceki ozon bakımından hava kirliliğindeki artış, iki gün sonraki hava kirliliğini 0,11 kat etkilemektedir.

BP_{t-1} : exp(0,415)=1,5 Bir gün önceki hava basıncının 1 gün sonraki hava kirliliğine etkisi 1,5 kattır.

$NO2_{t-1}$: exp(1,109)=1,12 Bir gün önceki azot dioksitin 1 gün sonraki hava kirliliğine etkisi 1,12 kattır.

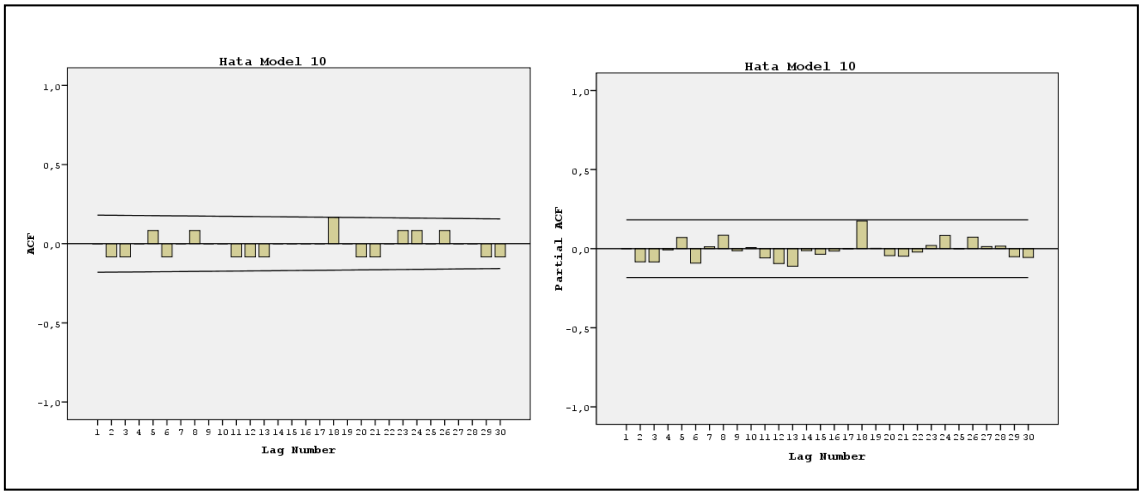
Sonuç olarak, hava kirliliğinin 1 kademe artması olasılığını, en çok hava sıcaklığının etkilediği söylenebilir. Bunun yanında, o günkü nem oranının ve bir önceki günün hava basıncının da, hava kirliliğinin artmasında etkili değişkenler olduğu görülmüştür.

Şekil 1'de, gözlenen birimlere karşılık, Model 10'a göre tahmin edilen ozon bakımından hava kalitesi düzeylerinin zaman serisi grafikleri verilmiştir.



Şekil 1. Gözlenen ve tahmin edilen hava kalitesi düzeylerinin zaman serisi grafikleri

Şekil 1 incelendiğinde, gözlenen ve tahmin değerleri arasında grafiksel uyum olduğu görülmektedir. Grafiksel olarak tam bir uyum olmasına rağmen ele alınan modelin istatistiksel olarak anlamlı ve tahminlerinin güvenilir olması için, hata serisinin tamamıyla rasgele hareketlere sahip olması ve ele alınan zaman serisi ile ilgili hiçbir bilgi taşımaması, dolayısıyla hata serisinin akgürültü serisi olması gereklidir. Ele alınan Model 10 için elde edilen hataların otokorelasyon ve kısmi otokorelasyon grafikleri Şekil 2'de verilmiştir.



Şekil 2. Model 10 hata serisine ait ACF ve PACF grafikleri

Şekil 2'de verilen grafiklere ve Box-Ljung istatistiklerine ($p>0,05$) göre hatalar akgürültüdür. Sonuç olarak, OOM için geçerli varsayımların sağlanmış olması, katsayıların ve buna bağlı olarak modelin anlamlı olması ve son olarak hataların akgürültü serisi olması, Model 10'un ele alınan hava kalitesi veri kümesi için istatistiksel olarak anlamlı olduğu söylenebilir. Model 10, varsayım koşullarını sağlayan modeller arasında, hatası en düşük olan model olduğu için tercih edilmiştir.

4. SONUÇ

Bu çalışmada, sıralı zaman serisi modelleri için OOM tanıtılmış ve model gerçek hava kalitesi verilerine uygulanmıştır. Ozon bakımından hava kalitesi düzeyi, sıcaklık, basınç, nem oranı açıklayıcı değişkenlerin yanında gecikmeli terimlerde dahil edilerek OOM ile modellenmiştir. Buna bağlı olarak birikimli olasılıklar ve odds miktarları elde edilmiş ve sonuçlar yorumlanmıştır. Ele alınan veri kümesine göre, ozon bakımından hava kalitesi düzeyinde sıcaklığın yanı sıra nem oranının ve bir gün önceki hava basıncının daha etkili olduğu görülmüştür.

Tahmin edilen model, ilgilenilen istasyon verisi için öngörü modeli olarak kullanılabilir. Ayrıca, farklı istasyon verileri ve farklı zaman dilimlerinde, farklı değişkenlerin dahil edilmesi ile çalışma genişletilerek genelleştirilebilir.

KAYNAKLAR

Agresti, A. (2002). *Categorical Data Analysis*, 2nd ed., New Jersey:John Wiley.

Bru, N., Despres, L. and Paroissin, C.A. Comparison of Statistical Models for Short Categorical or Ordinal Time Series with Applications in Ecology,*arxiv.math/0702706v1*.

Fokianos, K. and Kedem, B. (2003). Regression Theory for Categorical Time Series, *Statistical Science*. 18, 3, 357-376.

Jacobs, P. and Lewis, P. (1978). Discrete Time Series Generated by Mixtures i: Correlation and Runs Properties, *Journal of The Royal Statistical Society (Series B)*, 40(1), 94-105.

Liu, L. and Agresti, A. (2005). The Analysis of Ordered Categorical data: An Overview and A Survey of Recent Developments, *Sociedad de Estadıstica e Investigacion Operativa Test*, 14, 1-73.

McCullagh, P. (1980). Regression Models for Ordinal Data, *Journal of the Royal Statistical Society - Series B* 42, 109-142.

McCullagh, P. (2005). The proportional Odds model, *The Encyclopedia of Biostatistics* (Editor: Armitage, P., Colton, T.), Wiley, NewYork.

SPSS Inc. Released 2008. SPSS Statistics for Windows, Version 17.0. Chicago: SPSS Inc.

www.havaizleme.gov.tr (Erişim Tarihi 09/08/2011)