

HAVA KİRLİLİĞİNİN SAĞLIK ÜZERİNDEKİ ETKİSİNE PARAMETRİK OLMAYAN REGRESYON YAKLAŞIMI

Rabia Ece OMA^(*), Memmedağa MEMMEDLİ

Anadolu Üniversitesi Fen Fakültesi, İstatistik Bölümü, Eskişehir

ÖZET

Son zamanlarda, genelleştirilmiş toplamsal modeller (GAM), hava kirliliği ve sağlık çalışmalarındaki en önemli istatistiksel yaklaşım olarak kullanılmaktadır. Bu nedenle, bu konuda yapılmış bir çok çalışma incelenmiş, hava kirliliği ve sağlık ile ilgili temel regresyon modelleri ve bahsedilen modellerle ilişkili olarak ortaya koyulan problemler hakkında bilgi verilmiştir. Bir çok çalışmada, partikül madde, ozon, kükürt dioksit gibi kirletici değişkenler genelleştirilmiş kısmi parametrik (semiparametrik) modelin doğrusal kısmında yer almaktadır. Diğer taraftan, bazı çalışmalarda, bu değişkenlerin modelin parametrik olmayan kısmında da kullanımı ve ince tabakalı splayn şeklinde kirletici ve diğer zaman ve meteoroloji değişkenlerin etkileşiminin de modelde ele alınması önerilmektedir.

ABSTRACT

Recently, generalized additive models (GAMs) are used as an important statistical approach in studies of air pollution and health. Therefore, many papers have been reviewed about GAMs for studies of air pollution and health, and then information about basic regression models associated with air pollution and health, and problems, which arise according to these models, has been given. In many studies, pollutant variables, such as particulate matter (PM), ozone, sulphur dioxide, take part linear component of generalized semiparametric model. On the other hand, in some studies, using these variables at nonparametric component of the model and handling interaction of these variables and, time and weather variables as a thin plate spline have been suggested.

ANAHTAR SÖZCÜKLER

Hava Kirliliği, Ölüm, Genelleştirilmiş Toplamsal Model, Kısmi Parametrik Regresyon, Zaman Serisi, Etki Karışımı Yanlılığı

GİRİŞ

1950'li yıllarda hava kirliliği nedeniyle meydana gelen bazı yaygın ve yüksek oranda ölüm vakalarından sonra, hava kirliliğinin insan sağlığı üzerindeki etkisi ile ilgili ilk çalışmalar başlamıştır. Hava kirliliği ile yasalar yoluyla baş edebilmek için A.B.D ve Avrupa'da Temiz Hava Antlaşması (Clean Air Act) yürürlüğe girmiştir (Smith, 2007). Yüksek hava kirliliği olaylarını ortadan kaldırmak için yapılan bu ilk girişimlerin başarısına ve ardından yapılan bir çok çalışmaya rağmen hava kirliliği problemi günümüze kadar gelmiştir. Hala her yıl hava kirliliğinden dünyada 3 milyon insan ölmektedir ve bu değer dünyadaki toplam ölümün

* reayar@anadolu.edu.tr

(ortalama 55 milyon) %5'ini oluşturmaktadır. Ölümlerin %90'ı gelişmekte olan ülkelerde görülmektedir (Öztürk, 2005).

Türkiye'de ise hava kirliliği özellikle 1950'lerden sonra hızlı nüfus artışı, hızlı kentleşme ve endüstrileşme sonucu yoğun enerji kullanımı nedeniyle bir halk sağlığı sorunu olmaya başlamıştır. Endüstrileşme başladıktan sonra nüfusun %60'ı şehirlerde yaşamaya başlamıştır. Petrol ve kömür gibi fosil yakıtların aşırı tüketimi, yanlış kentleşme, uygunsuz ve yetersiz yakma teknikleri, yeşil alanların azalması, motorlu araç sayısındaki artış gibi etkenler de hava kirliliğini daha da arttırmıştır. Epidemiyolojik çalışmalar ve hava kalitesini izleme verileri, Türkiye'deki hava kirliliği göstergelerinden SO₂ ve PM düzeylerinin özellikle kış aylarında, halen kabul edilebilir düzeylerin üzerinde olduğunu göstermektedir. Diğer taraftan, Başta İstanbul, Ankara ve İzmir olmak üzere metropollerde hava kirliliği düzeyini aşağılara çekme çabalarının başarılı olmasına karşın orta ve küçük kentlerde hava kalitesinin düzelmesi için daha fazla çaba ve ilgi gerekmektedir (Bayram, 2005).

Bu çalışmanın amacı, hava kirliliğinin insan sağlığı üzerindeki etkisine farklı regresyon yaklaşımlarını inceleyerek Türkiye'deki hava kirliliği çalışmalarına bu yönde katkı sağlayabilmektir. Hava kirliliği ve sağlık ile ilgili regresyon yaklaşımları ilk olarak sadece doğrusal ve genelleştirilmiş doğrusal modelleri kapsamaktaydı. Fakat, özellikle son 10 yıldır, hava kirliliği ve sağlık problemine bu doğrusal yaklaşımların yerini neredeyse tamamen parametrik olmayan (nonparametrik) regresyon yaklaşımları almıştır ve bu yönde A.B.D. ve Avrupa'da hava kirliliği ve sağlık problemi ile ilgili çok fazla çalışma yapılmıştır. Bu çalışmalardan bazılarında değinilmiş ve önemli noktalar açıklanmıştır. Daha sonra ise hava kirliliği ve sağlık ile ilgili kullanılan regresyon modelleri kısaca açıklanmıştır.

HAVA KİRLİLİĞİNİN SAĞLIK ÜZERİNDEKİ ETKİSİNİ İNCELEMELİK İÇİN YAPILAN İSTATİSTİKSEL ÇALIŞMALARA GENEL BİR BAKIŞ

Hava kirliliği ile yasalar yolu ile baş etme konusunda yapılmış olan ilk girişimlerin (1956 yılında (British Clean Air Act) ve 1970 yılında (U.S. Clean Air Act) başarılı olmasına karşın, hava kirliliğinin erken ölümler gibi kötü sağlık etkilerinden sorumlu olduğu yönünde devam eden endişeler nedeniyle Londra'da kaydedilen uzun-dönem verilerinin analizleri, araştırmacıların bazı A.B.D. şehirlerinden alınan zaman serilerini derlemeye ve analiz etmeye başlamasına sebep olmuştur (Schwartz ve Dockery 1992, Schwartz 1993). Bu çalışmaların çoğu partikül madde (PM) ve ölüm arasındaki ilişkiyi ortaya koymuştur. Bu çalışmaların ardından çok sayıda büyük ölçekli araştırmalar yapılmaya başlanmıştır. Bunların içinde en iyi bilinenleri NMMAPS (National Mortality and Morbidity Air Pollution Study)(Samet vd. 2000a,b) ve APHEA (Air Pollution and Health: A European Approach) dır (Katsouyanni vd. 2001; Samoli vd. 2002). NMMAPS ve APHEA, A.B.D. ve Avrupa'daki bir çok merkezden ölüm, kirlilik ve meteoroloji ile ilgili zaman serisi verileri toplamaktadır. Bu zaman serisi verileri kullanılarak yapılan çalışmalar, hava kirliliği konsantrasyonu ile kötü sağlık koşulları arasındaki günlük değişkenlik ilişkisini tahmin eder. Elde edilen sonuçlar, hava kirliliğinin mevcut seviyelerinin risklerini hesaplamak için kullanılan epidemiyolojik bilgilere katkı sağlar.

Hava kirliliği ve sağlık konusunda yapılan çalışmalarda yanıt değişkeni y_t , t . günde meydana gelen ölümlerin sayısı ya da bazı hastalık ölçümleridir. Genel olarak, ölümlerin incelendiği problemlerde, *kaza sonucu ölümler* ölüm sayısına dahil edilmemiştir. En basit yaklaşım yanıt değişkeni üzerinde logaritmik, karekök gibi dönüşümler uygulayarak sıradan en küçük kareler

modelini kullanmaktır (Draper, N.R. ve Smith, H. 1998; Montgomery vd. 2001). Fakat daha yaygın yaklaşım ise ölümlerin μ_t ortalama ile Poisson dağılımı göstermesi geleneksel varsayımından hareketle bir genelleştirilmiş doğrusal model (GLM) (McCullagh ve Nelder 1989) kullanılmasıdır (Schwartz vd. 1996; Kelsall vd. 1997; Fairley 1990). Fakat son yıllarda hava kirliliği ve sağlık ile ilgili zaman serisi verilerinin istatistiksel modellenmesinde epey gelişme sağlandı ve başlangıçta kullanılan standart regresyon modellerinin yerine, regresyon splayn ile GLM (He vd. 2006, 2007), splayn düzeltme (smoothing spline) ile GAM (Hastie ve Tibshirani 1990), regresyon splayn ile GAM ve cezalı splaynlar ile GAM (Marx ve Eiler 1998) gibi kısmi parametrik regresyon modelleri kullanılmaya başlandı (Hastie ve Tibshirani 1990, Green ve Silverman 1994).

Yapılan çalışmalarda regresyon modellerinde kullanılan açıklayıcı değişkenler temel olarak hava kirliliği, meteoroloji ve mevsimsellik ya da uzun dönem trendlerini ifade eden üç tiptedir. Modellerde dikkate alınan hava kirliliği değişkenleri PM_{10} , $PM_{2.5}$, SO_2 ve O_3 gibi değişkenlerdir. Yapılan çalışmalarda genel olarak “kirletici” olarak ifade edilen bir ana değişkenle ilgilenilmiş fakat bazı durumlarda diğer kirleticilerin de modele dahil edildiği durumlar olmaktadır. Meteoroloji değişkenleri ise nemlilik sıcaklık, atmosferik basınç gibi değişkenlerdir ve sıcaklık hemen hemen kurulan her modele dahil edilmiştir. Mevsimsellik ya da uzun dönem trend değişkenleri için geleneksel yaklaşım ise bunların, zamanın bazı doğrusal olmayan fonksiyonları olmasıdır. Bu doğrusal olmayan etki taban fonksiyonlar kullanılarak modellenebilir (Green ve Silverman 1994; Hastie ve Tibshirani 1990; Dominici vd. 2004, Wood ve Augustin 2002). Bu bilgi ışığında yapılan bir çok çalışmada, kirleticilerin genelleştirilmiş modelin doğrusal parçasını, meteoroloji ve zaman değişkenlerinin ise modelin parametrik olmayan parçasını oluşturduğu kısmi parametrik regresyon modelleri kurulmuştur (Touloumi vd. 2006; Dominici vd. 2004; Peng vd. 2006). Dominici vd. (2004) makalesinde kurulan kısmi parametrik modelin parametrik kısmındaki katsayılar, hava kirliliği seviyelerindeki her bir birimlik artışın ölüm ya da hastalığındaki yüzdelerdeki artışını, parametrik olmayan kısımdaki fonksiyonlar ise splayn düzeltme ya da doğal kübik splayn v.b. yaklaşımlar kullanılarak oluşturulan, takvim zamanı, sıcaklık ve nemliliğin pürüzsüz (smooth) fonksiyonlarını ifade etmektedir. Böyle bir modelin tercih edilmesinin temel nedeni zaman serisi çalışmalarının son zamanlardaki analizlerinde ortaya çıkan ve etki karışımları yanlılığı (etki karışımı yanlılığı) olarak bilinen önemli epidemiyolojik ve istatistiksel durumu kontrol altında tutabilmektir. Etki karışımı yanlılığını kontrol etmek için zaman ve sıcaklığın pürüzsüz (smooth) fonksiyonları kısmi parametrik Poisson regresyon modeline parametrik olmayan kısım olarak dahil edilmiştir. Poisson regresyon modeline kirlilik kestirimi için yeteri kadar esnek olan sıcaklık ve zamanın pürüzsüz fonksiyonlarını dahil edilmesi etki karışımı yanlılığı durumunu ortadan kaldırılabılır. Bahsedilen etki karışımı yanlılığı zaman ve sıcaklığın pürüzsüz (smooth) fonksiyonlarında kullanılan serbestlik derecesi ile kontrol edilmektedir. Serbestlik derecesinin seçimi uzman kararına ya da AIC (Akaike Information Criteria) gibi optimum kriterlerine dayandırılmaktadır. Dominici vd. zaman ve sıcaklığın pürüzsüz fonksiyonlarında kullanılan serbestlik derecesinin uygun bir değerini belirleyerek, hava kirliliği risk tahminlerinin varyansını ve asimptotik “bias”ını hesaplamışlardır. Daha sonra ise bu hesaplamalara dayanarak standart hatalara göre küçük etki karışımı yanlılığına (etki karışımı yanlılığı) sahip hava kirliliği riski tahminlerini sağlayan zaman ve sıcaklığın pürüzsüz fonksiyonları için bir bant (düzeltme parametresi) seçim stratejisi geliştirmişlerdir. Bu çalışmada aynı zamanda GAM’deki regresyon katsayıları (hava kirliliği risk tahmincileri) vektörünün kovaryans matrisini tahminlemek için bir algoritma geliştirmişlerdir ve bu algoritma “confounding” faktörler için düzeltmenin derecesini tam olarak hesaplamaktadır.

Bu hesaplamaların yapılmasında kolaylık olması amacıyla S-Plus'da yer alan gam paketinin genişletilmiş bir versiyonu olan gam.exact paketini geliştirmişlerdir.

Hava kirliliği ve sağlık zaman serileri konusunda yapılan kısmi parametrik regresyon modellerinde bahsedilen etki karışımı yanlılığı ile ilgili önemli çalışmalardan biri de Touloumi vd. (2006)'nin yapmış oldukları çalışmadır. Bu çalışmada da kirletici, modelin parametrik kısmında yer almakta, ortalama günlük sıcaklık ve ortalama günlük oransal nemlilik değişkenleri modelin parametrik olmayan kısmını oluşturmaktadır. Ek olarak haftanın günleri ile ilgili değişken de modelde yapay değişken olarak yer almaktadır. Bu çalışmada modelde kullanılan pürüzsüz fonksiyonlarda kullanılan serbestlik derecesinin seçimi için çeşitli kriterler önerilmiş ve karşılaştırılmıştır. Bu amaç kısmi otokorelasyon fonksiyonu (PACF-partial autocorrelation function), Akaike bilgi kriteri (AIC), Bayesian bilgi kriteri (BIC-Bayesian Information Criteria) ve NMMAPS çalışmalarında sabitlenmiş olarak kullanılan zaman için 7, sıcaklık için 6 ve nemlilik için 3 değerinin kullanılması (Samet vd. 2000c; Dominici vd. 2000) dikkate alınmıştır.

Yukarıda bahsedilen modellerden farklı olarak, doğrusal olmayan etkilerin sadece mevsimsellik ya da uzun dönem trendleri ve meteoroloji ile sınırlı olmadığı; diğer değişkenlerin de (kirleticilerin de) modelin parametrik olmayan kısmına dahil olduğu bir çok çalışma mevcuttur (Wood, 2006; Xia ve Tong, 2006; Memmedli ve Omay, 2007a,b; Omay, 2007). Üstelik bu değişkenlerin etkileşimlerini de ortaya koyan ince tabakalı splayn (TPS-) içeren GAM modelleri tercih edilmektedir.

Kirleticilerin de modelde pürüzsüz fonksiyonlar yardımıyla nonparametrik olarak değerlendirildiği çalışmalardan biri Wood (2006) tarafından yapılmıştır. Wood çalışmasında Chicago'da meydana gelen hava kirliliğinin insan sağlığı üzerindeki etkisini kısmi parametrik regresyon modellerini kullanarak incelemiştir. Fakat şimdiye kadar değinmiş olduğumuz çalışmalardan farkı, kirleticileri de modele nonparametrik olarak dahil etmesidir. Çalışmalarını gerçekleştirirken, splayn fonksiyonlarının taban fonksiyonlar yardımıyla ifade edilmesine dayanan ve kendisinin geliştirmiş olduğu mgcv paketini kullanmıştır. Uygulamada öncelikle o ana kadar geleneksel olarak süre gelen kısmi parametrik regresyon modelini kurmuş fakat bu modele ilişkin kontrol grafiklerinden yola çıkarak, modelin esneklikten çok uzak olduğuna karar verip kirleticileri de pürüzsüz fonksiyonlar kullanarak modele nonparametrik olarak dahil etmiştir. Sonuç olarak ozon ve sıcaklığın birlikte parametrik olmayan etkisini ve PM₁₀ ve zamanın ayrı ayrı parametrik olmayan etkilerini içeren bir modelin en uygun model olduğuna karar vermiştir. Bu model Chicago'daki hava kirliliğinin insan sağlığı üzerindeki etkisini ifade eden bir modeldir. Fakat bu çalışmada elde edilen modele çok fazla güvenmeden önce, aynı modelin diğer bölgelerde ya da diğer zaman periyotlarında da iyi sonuç verip vermediğinin incelenmesi gerekmektedir.

Benzer çalışmalar Memmedli ve Omay (2007a,b) ve Omay (2007)'de de gerçekleştirilmiştir. Memmedli ve Omay (2007a)'da, Wood'un çalışmasına benzer bir çalışma, Los Angeles'daki hava kirliliğinin insan sağlığı üzerindeki etkisini modellemek amacıyla yapılmıştır. Bu çalışmada, mgcv ve NMMAPSdata paketleri kullanılmıştır ve dikkate alınan veriler, 01/01/1987-31/12/2000 yılları arasında, Los Angeles'da kaydedilen günlük ölüm sayısı (kazalar hariç tutularak) ve hava kalitesi değerleridir. İlgilenilen yanıt değişkeni yıllar içerisinde Los Angeles'da meydana gelen günlük ölüm sayılarıdır. Açıklayıcı değişkenler ise ozon, karbon monoksit, PM₁₀, PM_{2.5}, sıcaklık ve zamandır. Ele alınan regresyon modelleri ise doğrusal model, GLM, toplamsal model, GAM ve ince tabakalı splayn ile GAM'dır.

Toplamsal model, GAM ve ince tabakalı splayn kullanılarak oluşturulan GAM’da *regresyon splayn* kullanılarak model uyumu yapılmıştır. Bu modeller AIC ve sapma (deviance) değerleri göz önünde bulundurularak karşılaştırılmış ve bu kriterlere göre Los Angeles’deki hava kirliliğinin insan sağlığı üzerindeki etkisini en iyi açıklayan modelin ince tabakalı splayn içeren GAM olduğuna karar verilmiştir. Bu uygulamaya ait sonuçlar Tablo1’de yer almaktadır.

Tablo 1. Los Angeles verileri için kurulan modellerin AIC ve sapma değerleri

Modeller	AIC	Sapma
Doğrusal Model	43601.46	1506674
Genelleştirilmiş Doğ. Model	5157.500	1081.8
Toplamsal Model	5015.326	149266.1
Genelleştirilmiş Top. Model	5047.006	910.0135
TPS ile elde edilen GAM	4929.725	673.4297

Birinci ve ikinci terimi sırasıyla zaman, PM₁₀ ve PM_{2,5}’in etkileşimleri ve sıcaklık ve ozonun etkileşimini içeren ince tabakalı splayn olan GAM aşağıda verilmiştir.

$$\log(E[death_i]) = f_1(time_i, pm10median_i, pm25median_i) + f_2(tmpd_i, o3median_i) + f_3(comedian_i) + \varepsilon_i$$

$$death_i \sim Poi(E[death_i])$$

Yukarıda bahsetmiş olduğumuz çalışmalar alternatif olarak dikkate alınabilecek diğer çalışmalar ise prospektif çalışmalar (prospective study) (Dockery vd., 1993; Pope vd., 1995; Krewski vd. 2000) ve “case-crossover” çalışmalarıdır (Levy vd., 2001; Janes vd. 2005). Prospektif çalışmalar, zaman serisi çalışmalarını yapmayan araştırmacılara avantaj sağlamaktadır. Zaman serisi çalışmalarında regresyon parametreleri her bir şehir için hesaplanır ve daha fazla doğruluk sağlamak amacıyla ancak daha sonra şehirler birleştirilir. Prospektif çalışmalarda ise, zaman serisi çalışmalarının aksine, regresyonların kendisi şehirler arasındaki karşılaştırmalara dayanır. “Case-crossover” çalışmalarda ise temel düşünce ölüm, kalp krizi gibi bazı kötü olayların hemen öncesinde bireyin bir kirleniciye maruz kalmasıyla, diğer bir kontrol ya da referans zamanda aynı bireyin aynı kirleniciye maruz kalmasının karşılaştırılmasıdır (Smith, 2007).

HAVA KİRLİLİĞİNİN SAĞLIK ÜZERİNDEKİ ETKİSİNİ İNCELEMELİK İÇİN KULLANILAN REGRESYON MODELLERİ

Hava kirliliğinin insan sağlığı üzerindeki etkisini ortaya koyacak sıradan en küçük kareler modeline göre daha yaygın bir yaklaşımın, özellikle ölüm olaylarında, y_i bağımsız değişkeninin $E(y_i) \equiv \mu_i$ ortalama ile bir Poisson dağılıma sahip olması varsayımından yola çıkarak kullanılan genelleştirilmiş doğrusal model olduğundan daha önce bahsetmiştik. Bu model (1) denkleminde ifade edilmiştir.

$$\eta_i = g(\mu_i) = \log \mu_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i \quad (1)$$

Bu model bir genelleştirilmiş doğrusal modeldir (GLM) ve modelde yer alan g , pürüzsüz bir link fonksiyonudur. η_i ise doğrusal kestirici olarak adlandırılmaktadır. Yaygın olarak kullanılan iteratif olarak yeniden ağırlıklandırılmış en küçük kareler (IRLS- *Iteratively Re-*

weighted Least Squares) algoritması kullanılarak (1) modelinin uyumu yapılır (Nelder, Wedderburn 1972, McCullagh, Nelder 1989).

Yukarıda ifade edilen doğrusal regresyon ve GLM parametrik regresyon problemleri sınıfına aittir. Bu problemlerde regresyon fonksiyonlarının biçimi bilinmemektedir ve bu fonksiyonların tahmini için sadece sonlu sayıda parametreyi tahmin etmek gerekmektedir. Bu parametre tahminleri ise normal denklemler veya skor denklemleri sisteminin çözümü ile belirlenmektedir. Hava kirliliğinin insan sağlığı üzerindeki etkisi ile ilgili uygulama çalışmalarında, açıklayıcı değişkenlerin yanıt değişkenine etkisinin sonlu parametre içeren fonksiyonel bir ilişkiden daha karmaşık bir ilişkiye sahip olduğu gözlenmiştir. Bu durumda, sonlu sayıda parametre tahmini problemi sonsuz boyutlu uzayda bir eleman tahmini problemine, diğer bir ifadeyle parametrik olmayan regresyon problemine dönüşmüş olur. Parametrik olmayan regresyon modellerinin en basit formu, bir açıklayıcı değişken içeren yani tek bir düzeltici fonksiyonu olan modeldir ve (2)'de gösterilmiştir.

$$y_t = f(x_t) + \varepsilon_t \quad (2)$$

Bu modelin daha gelişmiş formları ise kısmi parametrik, toplamsal, genelleştirilmiş toplamsal ve ince tabakalı splayn modelleridir.

Parametrik olmayan regresyon modellerinin analizinde yer alan temel yaklaşım, pürüzlülük ceza yaklaşımıdır (Green ve Silverman, 1994; Hastie ve Tibshirani, 1999). Bu yaklaşımda, modeldeki bilinmeyen fonksiyonları tahmin etmek için, hata kareler toplamına veya log-olabilirlik fonksiyonuna bir ceza terimi eklenir. Ceza teriminin eklenmesi mevcut problemi, geleneksel enterpolasyon probleminden gerçek regresyon problemine dönüştürmüştür.

Geleneksel olarak ceza terimi $\int_a^b \{g''(x)\}^2 dx$ integralinin bir α düzeltme parametresine

çarpımı olarak kullanılır ve \hat{g} tahmin fonksiyonu ikinci mertebeden sürekli türeve sahip $g(\cdot)$ fonksiyonlar uzayından seçilir. Schonberg'in 1964 yılında yayınladığı polinomial splaynların optimumluk özelliği, bu sonsuz boyutlu problemi sonlu boyutlu parametrik bir probleme dönüştürmek için imkan sağlamaktadır. Yani, cezalı hataya optimum değer veren fonksiyonun bir kübik splayn olması bulgusu parametrik olmayan regresyonda tahmin problemini parametrik hale getirir. Dolayısıyla, parametrik olmayan regresyon probleminin polinomial splayn fonksiyonlarıyla incelenmesi, modelin analizini çok kolaylaştırmaktadır.

Pürüzlülük ceza yaklaşımında kullanılan yöntemlerden biri regresyon splaynı yöntemidir (Wood, 2002). Regresyon splaynı ile pürüzlülük ceza yaklaşımında splayn fonksiyonu belirli taban splaynların toplamı şeklinde ifade edilir ve burada bilinmeyenler, kullanılan taban fonksiyonların katsayılarıdır. Bu katsayılar cezalı hatanın minimizasyonu probleminden bulunur. Regresyon splaynında düğüm noktalarının sayısı (konumu) veya taban fonksiyonların sayısı özel seçilerek, tüm gözlemlerin sayısından çok daha küçük alınabilir. Bu splayn düzeltme yöntemindeki hesaplama karmaşıklığını epey hafifletmiş olur. Wood taban fonksiyonlar olarak kübik splayn tabanlarını (Gu, 2002) ve sürekli ceza terimini kullanarak regresyon splaynını uygulamaktadır (Wood, 2000; Wood vd., 2002, Wood, 2003; Wood, 2004).

Cezalı regresyon splaynını daha ayrıntılı incelemek için aşağıda ifade edilen kısmi parametrik toplamsal regresyon modeli dikkate alınsın (Green ve Silverman, 1994; Hastie ve Tibshirani, 1999).

$$y_t = \mathbf{z}_t^T \boldsymbol{\beta} + f_1(x_{1t}) + \dots + f_p(x_{pt}) + \varepsilon_t, t=1,2,\dots,n \quad (3)$$

(3) toplamsal modeli *genelleştirilmiş toplamsal modele* (GAM) şöyle genişletilebilir:

$$g(\mu_t) = \mathbf{z}_t^T \boldsymbol{\beta} + f_1(x_{1t}) + \dots + f_p(x_{pt}), Y_t \sim \text{üstelaile} \quad (4)$$

(4) denkleminde yer alan $f_k, k=1,2,\dots,p$ fonksiyonları cezalı kübik regresyon splaynı ile aranan düzeltme fonksiyonlarıdır ve aşağıdaki taban fonksiyonlar yardımıyla verilir.

$$f_k(t) = \sum_{j=1}^{q_k} \beta_{kj} b_{kj}(t) \quad (5)$$

(5) ifadesindeki $b_{kj}(\cdot)$ taban fonksiyonlarının, sırasıyla $f_k, k=1,2,\dots,p$ fonksiyonları için *kübik splayn taban fonksiyonları* olduğu varsayılır (Wood, 2002). Düzeltici f_k fonksiyonlarının kestirimi için *cezalı log-olabilirlik* (log-likelihood) yöntemi kullanılabilir.

f_k fonksiyonlarından bazıları birden çok x_j değişkenin fonksiyonu olduğunda GAM modeli *ince tabakalı splayn* bileşenine sahip olur. Hava kirliliği riski ile ilgili çalışmalarda özellikle sıcaklığın bazı değişkenlerle birlikte değişme eğiliminde olduğu gözlenmiş ve bu yönde yapılan çalışmalarda ince tabakalı splayn içeren genelleştirilmiş toplamsal modellerin daha iyi sonuçlar verdiği ortaya konmuştur (Memmedli ve Omay, 2007a,b; Omay, 2007; Wood, 2006.).

Hava kirliliği riski ile ilgili çalışmalarda yaş ve ölüm nedeni gibi değişkenler modele yapay (dummy) değişken olarak dahil edilebilir. Bu durumda değişken katsayılı modellerin (varying-coefficient model) kullanılması uygun olabilir (Hastie, 1993). Dağılımı bir η parametresine bağlı olan bir Y rassal değişkenine ve aynı zamanda X_1, X_2, \dots, X_p ve R_1, R_2, \dots, R_p kestiricilerine sahip olduğumuzu varsayalım. Bu durumda bir değişken katsayılı model aşağıdaki formda olacaktır:

$$\eta = \beta_0 + X_1 \beta_1(R_1) + \dots + X_p \beta_p(R_p) \quad (6)$$

(1) modelinde, $R_1, R_2, \dots, R_p, \beta_1(\cdot), \dots, \beta_p(\cdot)$ fonksiyonları vasıtasıyla X_1, X_2, \dots, X_p 'nin katsayılarını değiştirmektedir. $\beta_j(\cdot)$ 'nin R_j üzerindeki bağımlılığı her bir R_j ve X_j arasındaki özel bir çeşit etkileşimi ifade eder (Hastie, 1993). Değişken katsayılı model oldukça spesifik görünmesine karşın daha geneldir; (6) modelinin bazı özel durumları yukarıda ifade edilen doğrusal, genelleştirilmiş doğrusal, genelleştirilmiş toplamsal ya da yapay değişkenli model gibi modelleri ifade eder.

SONUÇ VE ÖNERİLER

Bu çalışmada hava kirliliğinin insan sağlığı üzerindeki etkisini ortaya koymak için kullanılan GAM'in önemini vurguladık. İncelemiş olduğumuz bir çok çalışmada temel olarak iki yaklaşım gördük, birincisi kirleticilerin doğrusal, diğer değişkenlerin ise nonparametrik olarak dikkate alındığı semiparametrik genelleştirilmiş toplamsal modeller, diğeri ise kirletici değişkenlerin de modele nonparametrik olarak dahil edilebildiği genelleştirilmiş toplamsal modeller. Bu modeller özellikle son on yılda hava kirliliği ve sağlık çalışmalarında doğrusal ve genelleştirilmiş doğrusal modellerin yerini almıştır. İnsan sağlığı üzerinde etkisi olması muhtemel olan değişkenler incelenirken, onların sadece bireysel katkıları değil birlikte etkileşimleri de göz önünde bulundurulmalıdır. Bu durum ince tabakalı splayn ile GAM kullanılarak gerçekleştirilebilmektedir.

Hava kirliliği ve sağlık ile ilgili yapılan parametrik olmayan regresyon çalışmalarında, yaş, haftanın günleri vb. değişkenler de birer yapay değişken olarak modele dahil edilebilir. Bu çalışmaların devamı olarak, yapay değişkenlerin genelleştirilmiş toplamsal modellerde dikkate alınmasının yanı sıra bu değişkenin diğer değişkenlerle doğrusal ve doğrusal olmayan etkileşimlerinin varlığının incelenmesi, incelediğimiz probleme farklı bakış açıları getirebilecek bir yaklaşımdır (Hastie, 1993). Genel olarak, genelleştirilmiş doğrusal model için yapay değişkenlerin kullanılması literatürde henüz çok incelenmemiştir.

Şu ana kadar, hava kirliliği ve meteoroloji ile ilgili daha ayrıntılı verilere ulaşabilmemize karşın, ölüm ile ilgili istediğimiz ayrıntıları taşıyan verilere ulaşamamız nedeniyle, Türkiye'deki herhangi bir şehir için bu uygulamaları gerçekleştiremedik. Veriler ile ilgili bu eksikliği giderebildiğimiz taktirde, önerilen regresyon modellerinin Türkiye'deki hava kirliliğinin insan sağlığına etkisi gibi önemli ve güncel bir problemin çözümüne, katkı sağlayabileceği görüşündeyiz.

KAYNAKLAR

Bayram, H. Türkiye'de hava kirliliği sorunu: Nedenleri, alınan önlemler ve mevcut durum, *Toraks dergisi*, 6(2), 159-165, 2005.

Dockery, D.W., Pope, C.A., Xu, X., Spengler, J.D., Ware, J.H., Fay, M.E., Ferris, B.G. ve Speizer, F.E. An association between air pollution and mortality in six U.S. cities, *New England Journal of Medicine*, 329, 1753-1759, 1993.

Dominici, F., McDermott, A. ve Hastie, T.J. Improved semiparametric time series models of air pollution and mortality, *Journal of the American Statistical Association*, 99, 938-948, 2004.

Dominici, F., Samet, J.M. ve Zeger, S.L. Combining evidence on air pollution and daily mortality from the 20 largest US cities: a hierarchical modelling strategy, *Journal of the Royal Statistical Society, Series A*, 163, 263-302, 2000.

Draper, N.R. ve Smith, H. *Applied Regression Analysis*, Wiley-Interscience, 1998.

Fairley, D., the relationship of daily mortality to suspended particulates in Santa Clara County, 1980-1986, *Environmental Health Perspectives*, 89, 159-168, 1990.

Green, P.J. ve Silverman, B.J. Nonparametric Regression and Generalized Linear Models, Chapman Hall, 1994.

Gu, C. Smoothing Spline ANOVA Models, Springer-Verlag, New York, 2002.

Hastie, T.J. ve Tibshirani, R.J. Generalized Additive Models, Chapman Hall, 1990.

Hastie, T.J. and Tibshirani, R.J. Varying-coefficient Models, *Journal of the Royal Statistical Society*, Series B, 55(4), 1993.

He, S., Mazumdar, S. ve Arena, V.C. A comparative study of the use of GAM and GLM in air pollution research, *Environmetrics*, 17, 81-93, 2006.

He, S., Mazumdar, S., Arena, V.C. ve Tang, G. Partial regression method to fit a generalized additive model, *Environmetrics*, 18, 599-606, 2007.

Janes, H., Sheppard, L. ve Lumley, T. Overlap bias in the case-crossover design, with application to air pollution exposures, *Statistics in Medicine*, 24, 285–300, 2005.

Katsouyanni, K., Toulomi, G., Samoli, E., Gryparis, A., LeTertre, A., Monopolis, Y., Rossi, G., Zmirou, D., Ballester, F., Boumghar, A. ve Anderson, H.R. Confounding and effect modification in the short term effects of ambient particles on total mortality: result from 29 European cities within the APHEA2 project, *Epidemiology*, 12, 521-531, 2001.

Kelsall, J.E., Samet, J.M., Zeger, S.L. ve Xu, J. Air pollution and mortality in Philadelphia, 1974-1988, *American Journal of Epidemiology*, 146, 750-762, 1997.

Krewski, D., Burnett, R.T., Goldberg, M.S., Hoover, K., Siemiatycki, J., Jerrett, M., Abrahamowicz, M. ve White, W.H. Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality. A Special Report of the Institute's Particulate Epidemiology Reanalysis Project. Health Effects Institute, 2000.

Levy, D., Lumley, T., Sheppard, L., Kaufman, J. ve Checkoway, H. Referent selection in case-crossover analyses of health effects of air pollution, *Epidemiology*, 12, 186–192, 2001.

Marx, B.D. ve Eilers, P.H.C. Direct generalized additive modeling with penalized likelihood, *Computational Statistics and Data Analysis*, 28, 193-209, 1998.

McCullagh, P. ve Nelder, J.A. Generalized Linear Models, Chapman Hall, 1989.

Memmedli M., Omay R.E. Regression Models with Thin Plate Spline, *First International Conference on Soft Computing Technologies in Economy (ICSCTE-2007)*, Baku, Azerbaijan, November 19-20, 2007a.

Memmedli, M. ve Omay, R.E. Hava Kirliliğinin Ölüm Oranı Üzerindeki Etkisinin Genelleştirilmiş Toplamsal Regresyon Modelleri ile İncelenmesi, 5. İstatistik Kongresi, Bildiri Özetleri, 40, 2007b.

Montgomery, D.C., Peck, E.A. ve Vining, G.G. Introduction to Linear Regression Analysis, Wiley-Interscience, 2001.

Nelder, J.A. ve Wedderburn, R.W.M. Generalized linear models, *Journal of the Royal Statistical Society, Series A*, 135, 370-384, 1972.

Omay, R.E., Regresyonda Pürüzlülük Ceza Yaklaşımı, Tez Danışmanı: Prof. Dr. Memmedağa MEMMEDLİ, 2007.

Öztürk, M., Şehir İçi Bölgelerde Hava Kirliliğinin Sağlık Üzerine Etkileri, 2005. <http://www.cevreorman.gov.tr/belgeler1/hk.doc>

Peng, R.D., Dominici, F. ve Louis, T.A. Model choice in time series of air pollution and mortality, *Journal of the Royal Statistical Society, Series A*, 169, 179-203, 2006.

Pope, C.A., Thun, M.J., Namboodiri, M.M., Dockery, D.W., Evans, J.S., Speizer, F.E. ve Heath, C.W. Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults, *American Journal of Respiratory and Critical Care Medicine*, 151, 669-674, 1995.

Samet, J.M., Dominici, F., Zeger, S.L., Schwartz, J., ve Dockery, D.W. National Morbidity, Mortality, and Air Pollution Study. Part I: Methods and Methodologic Issues, Cambridge: Health Effect Institute, 2000a.

Samet, J.M., Zeger, S.L., Dominici, F., Curriero, F., Coursac, I., Dockery, D., Schwartz, J., ve Zanobetti, A. National Morbidity, Mortality, and Air Pollution Study. Part II: Morbidity and Mortality from Air Pollution in the United State, Cambridge: Health Effect Institute, 2000b.

Samet, J.M., Dominici, F., Curriero, F.C., Coursac, I. ve Zeger, S.L., Fine particulate air pollution and mortality in 20 U.S. cities, 1987-994, *New England Journal of Medicine*, 343, 1742-1749, 2000c.

Samoli, E., Schwartz, J., Wojtyniak, B., Toulomi, G., Spix, C., Balducci, F. ve Medina, S. Investigating regional differences in short-term effects of air pollution on daily mortality in the APHEA Project: A sensitivity analysis for controlling long-term trends and seasonality, *Environmental Health Perspectives*, 109, 349-353, 2002.

Schwartz, J. ve Marcus, A. Mortality and air pollution in London: A time series analysis, *American Journal of Epidemiology*, 131, 185-194, 1990.

Schwartz, J. ve Dockery, D.W. Increased mortality in Philadelphia associated with daily air pollution concentrations, *The American Review of Respiratory Disease*, 145, 600-604, 1992.

Schwartz, J., Spix, C., Touloumi, G., Bacharova, L., Barumamdzadeh, T., Le Tertre, A., Piekarksi, T., Ponce de Leon, A., Ponka, A., Rossi, G., Saez, M., Schouten, J.P. Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions, *Journal of Epidemiology and Community Health*, 50, 3-11, 1996.

Smith, R.L. Air Pollution Risk, to appear in the *Encyclopedia of Quantitative Risk Assessment*, edited by Brian Everitt and Ed. Melnick, to be published by John Wiley and Sons, Ltd., 2007.

Touloumi, G., Samoli, E., Pipikou, M., Tertre, A.L., Atkinson, R. ve Katsouyanni, K. Seasonal confounding in air pollution and health time-series studies: Effect on air pollution effect estimates, *Statistics in Medicine*, 25, 4164-4178, 2006.

Wood, S.N. Modeling and Smoothing Parameter Estimation With Multiple Quadratic Penalties, *Journal of Royal Statistical Society, Ser. B*, 62, 413-428, 2000.

Wood, S.N. ve Augustin, N.H. GAMs with integrated model selection using penalized regression splines and applications to environmental modelling, *Ecological Modelling*, 157, 157-177, 2002.

Wood, S.N. Thin Plate Regression Splines, *Journal of Royal Statistical Society, Ser. B*, 65, 95-114, 2003.

Wood, S.N. Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models, *Journal of the American Statistical Association*, 99, 673-686, 2004.

Wood, S.N. Generalized Additive Models: An Introduction with R, Chapman Hall, 2006.