

BULANIK ZAMAN SERİLERİ İLE ANKARA HAVA KALİTESİ VERİSİNİN ÇÖZÜMLENMESİ

Sibel ALADAĞ* Çağdaş Hakan ALADAĞ** Erol EĞRİOĞLU***

ÖZET

Son yıllarda, hava sıcaklığı ve borsa verileri gibi belirsizlik içeren zaman serilerinin çözümlenmesinde bulanık zaman serileri başarıyla kullanılmaktadır. Literatürde, daha güvenilir öngörüler elde etmek amacıyla, bulanık zaman serisi yaklaşımları üzerinde çalışılmaktadır. Güvenilir öngörüler elde edilmesinde anahtar noktalardan biri, aralık uzunluğunun doğru olarak belirlenebilmesidir. Bu problemi çözmek amacıyla literatürde önerilen bazı yaklaşımlar mevcuttur. Hava kalitesi verileri de yapılarından dolayı belirsizlik içeren zaman serileridir ve bu tür zaman serilerinin çözümlenmesinde de bulanık zaman serilerinin kullanılması gerekmektedir. Yapılan bu çalışmada, Ankara hava kalitesi verisi ilk defa bulanık zaman serileri yöntemiyle çözümlenmiştir. Çözümlemede aralık uzunluğu için farklı teknikler kullanan altı farklı bulanık zaman serisi öngörü yöntemi kullanılmış ve elde edilen sonuçlar karşılaştırılarak yorumlanmıştır.

Anahtar Kelimeler: Ankara hava kalitesi, Aralık uzunluğu, Bulanık zaman serileri, Optimizasyon, Öngörü.

1. GİRİŞ

Borsa verileri, hava sıcaklığı değerleri ya da hava kirliliği ölçümleri gibi zaman serileri belirsizlik içeren serilerdir. Bu çeşit verilerde gözlem değerleri aynı gün hatta aynı saat içinde bile değişiklik gösterebilmektedir. Bu tarz serilerin çözümlenmesinde klasik zaman serisi yaklaşımları yerine bulanık zaman serilerinin kullanılması gerekmektedir (Aladağ vd., 2009).

Bulanık zaman serisi yaklaşımı ilk olarak (Song vd., 1993a), (Song vd., 1993b), (Song vd., 1994) tarafından bulanık küme teorisine (Zadeh, 1965) dayalı olarak önerilmiştir. Çok karmaşık matris işlemleri içeren bu yöntem, (Chen, 1996) tarafından daha basit bir şekle getirilmiştir. Chen çalışmasında, Song ve Chissom tarafından geliştirilen yöntemdeki karmaşık matris işlemleri yerine, bulanık ilişki tablosundan yararlanma fikrini ortaya koymuştur (Chen, 1996).

Bulanık zaman serileri yaklaşımına olan ilgi giderek artmakta ve çeşitli alanlarda birçok uygulama alanı bulmaktadır (Aladağ vd., 2010). Daha güvenilir öngörülere ulaşmak amacıyla literatürde önerilen çeşitli bulanık zaman serisi yaklaşımları bulunmaktadır (Eğrioğlu vd., 2009). Uygulamalarda daha iyi öngörüler elde etmek amacıyla bulanık zaman serileri üzerinde çalışmalar devam etmektedir. Güvenilir öngörülere ulaşılmasında anahtar noktalardan biri aralık uzunluğunun doğru belirlenmesidir (Eğrioğlu vd., 2011). Seçilen aralık uzunluğu, kullanılacak bulanık zaman serisi modelinin öngörü performansını doğrudan etkileyen bir etkidir. Aralık uzunluğunun

*Sos. Güv. Uzm. Yrd., Sosyal Güvenlik Kurumu, HSGM, Elektronik Yayın ve Bilgi Dağıtım Daire Başkanlığı, Ankara, e-posta: saladag@sgk.gov.tr

**Doç. Dr., Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Beytepe, Ankara, e-posta: aladag@hacettepe.edu.tr

***Doç. Dr., Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, Samsun, e-posta: erole@omu.edu.tr

belirlenmesi problemini çözmek amacıyla literatürde önerilen çeşitli yaklaşımlar mevcuttur. Aralık uzunluğunun belirlenmesi için, Huarng (2001) ortalamaya ve dağılıma dayalı yaklaşımlar önermiştir. Huarng ve Yu (2006) aralık uzunluğunun belirlenmesi için orana dayalı yeni bir yaklaşım geliştirmiştir. Literatürde aralık uzunluğunun belirlenmesi için bir başka yaklaşım ise Eğrioğlu vd. (2011) tarafından önerilmiştir. Eğrioğlu vd. (2011) çalışmalarında, en iyi aralık uzunluğunu belirlemek için kısıtlı optimizasyona dayalı bir yaklaşım kullanarak, Chen (1996) tarafından önerilen yöntemi geliştirmişlerdir.

Özellikle büyük şehirlerde hava kalitesi verilerinin değerlendirilmesi, hem insan sağlığı, hem çevre koşulları, hem de yapılacak yatırımlar açısından büyük önem taşımaktadır. Hava kalitesi verileri yapılarından ötürü belirsizlik içermektedirler ve klasik zaman serisi yöntemleri bu tür serilerin içerdiği belirsizliği modellemede yetersiz kalmaktadır. Bu nedenle, şehirlerin hava kalitesi verilerinin çözümlenmesinde bulanık zaman serileri yaklaşımlarından yararlanmak çok daha akıllıca olacaktır. Yapılan çalışmada, Ankara hava kalitesi verisi ilk kez bulanık zaman serileri yöntemiyle çözümlenmiştir. Çözümlemede kullanılan bulanık zaman serisi yöntemleri Song ve Chissom (1993a), Chen (1996), Huarng (2001), Huarng ve Yu (2006) ve Eğrioğlu vd. (2011) tarafından önerilen yaklaşımlardır. Huarng (2001) çalışmasında hem ortalamaya hem dağılıma dayalı olmak üzere iki farklı yöntem önermiştir ve bu iki yöntemde yapılan çalışmada kullanılmıştır. Buna göre, Ankara hava kalitesi verisi toplamda altı farklı bulanık zaman serisi yaklaşımıyla öngörülmüştür. Uygulama sonucunda elde edilen sonuçlar incelenmiş ve yorumlanmıştır.

Makalenin ikinci bölümünde, bulanık zaman serileri ve temel tanımları kısaca verilmiştir. 2.1’de Chen (1996) tarafından önerilen bulanık zaman serisi öngörü yöntemi özetlenmiştir. 2.2’de Eğrioğlu vd. (2011) tarafından önerilen optimizasyona dayalı aralık uzunluğu belirleme yaklaşımını tanıtılmıştır. Yapılan uygulama ve elde edilen sonuçlar üçüncü bölümde verilmiştir. Son bölümde ise elde edilen bulgular ışığında ulaşılan sonuçlar değerlendirilmiştir.

2. BULANIK ZAMAN SERİLERİ

Aladağ vd. (2010) tarafından bulanık zaman serileri ve temel bazı tanımları aşağıdaki gibi verilmiştir.

$U = \{u_1, \dots, u_b\}$ evrensel küme olsun. U ’nun elemanları aralıklardır. Bu aralıklar zaman serisinin tüm değerlerini kapsayan evrensel kümenin parçalanması ile elde edilir.

U ’nun elemanlarına bağlı olarak A_i bulanık kümeleri aşağıdaki gibi tanımlanabilir.

$$A_i = f_{A_i}(u_1)/u_1 + \dots + f_{A_i}(u_b)/u_b$$

Burada f_{A_i} , A_i bulanık kümesinin üyelik fonksiyonudur ve $f_{A_i} : U \rightarrow [0,1]$ olmaktadır. $f_{A_i}(u_a)$, u_a ’nın A_i bulanık kümesine ait olmasının üyelik derecesidir.

Tanım 1. $Y(t)$ ($t = \dots, 0, 1, 2, \dots$) reel sayıların bir alt kümesi olsun ve A_i bulanık kümelerinden oluşan evrensel kümeyi temsil etsin. A_j ($j = 1, 2, \dots$) bulanık kümelerinden

oluşan $F(t)$, $Y(t)$ ($t = \dots, 0, 1, 2, \dots$) evrensel küme üzerindeki bulanık zaman serisi olarak adlandırılır.

Tanım 2. * herhangi bir operatörü göstermek üzere, eğer $F(t) = F(t-1) * R(t, t-1)$ olacak şekilde bir $R(t, t-1)$ bulanık ilişkisi varsa, $F(t)$ sadece $F(t-1)$ 'den etkilenmektedir. Bu durumda $F(t)$ birinci dereceden bulanık zaman serisi olarak da adlandırılabilir. $F(t-1) = A_i$ ve $F(t) = A_j$ olduğunda, $F(t-1)$ ile $F(t)$ arasındaki bulanık ilişki $A_i \rightarrow A_j$ şeklinde de gösterilebilir. A_i bulanık ilişkinin sol yanı ve A_j bulanık ilişkinin sağ yanı olarak isimlendirilir.

2.1 Chen'in Bulanık Zaman Serisi Öngörü Yöntemi

Tanım 1 ve Tanım 2'ye bağlı olarak, Chen (1996) tarafından önerilen birinci dereceden bulanık zaman serisi yaklaşımı aşağıda verildiği gibi altı adımda özetlenebilir.

Adım 1. $U = \{u_1, u_2, \dots, u_b\}$ evrensel küme ve evrensel kümenin elemanları olan alt aralıklar ($u_i, i = 1, 2, \dots, b$) tanımlanır. Evrensel kümenin başlangıç ve bitiş noktaları belirlenir. Buna göre evrensel küme

$$U = [\text{Başlangıç}, \text{Bitiş}]$$

şeklinde ifade edilebilir. Daha sonra, belirlenen uygun aralık uzunluğuna göre evrensel küme eşit alt aralıklara bölünür.

Adım 2. Evrensel küme ve belirlenen alt aralıklara bağlı olarak A_i bulanık kümeler

$$A_i = f_{A_i}(u_1)/u_1 + \dots + f_{A_i}(u_b)/u_b$$

şeklinde tanımlanır.

Adım 3. Gözlemler bulanıklaştırılır. Örneğin, eğer bir gözlemin en büyük üyelik derecesi A_j bulanık kümesinde ise, bu gözlemin bulanık değeri A_j olarak belirlenir.

Adım 4. Bulanık ilişki ve grup ilişki tablosu oluşturulur. Gruplandırma için örnek verilirse, örneğin $A_1 \rightarrow A_2$, $A_1 \rightarrow A_1$ ve $A_1 \rightarrow A_3$ ilişkileri bulunsun, bu ilişkiler

$$A_1 \rightarrow A_1, A_2, A_3$$

biçiminde gruplandırılabilir.

Adım 5. Öngörüler elde edilir. Öngörü elde edilirken birkaç durum söz konusudur. $F(t-1) = A_j$ olsun. Buna göre, öngörü elde ederken oluşabilecek durumlar aşağıda verilmiştir.

Durum 1. Bulanık mantık ilişki tablosunda sadece bir ilişki olabilir. Örneğin, sadece $A_j \rightarrow A_i$ ilişkisi varsa, $F(t)$ öngörü değeri A_i 'dir.

Durum 2. Bulanık grup ilişki tablosunda $A_j \rightarrow A_i, A_j, \dots, A_k$ ilişkisi varsa, $F(t)$ öngörü değeri A_i, A_j, \dots, A_k 'dir.

Durum 3. Bulanık grup ilişki tablosunda $A_j \rightarrow Boş$ ise $F(t)$ öngörü değeri A_j 'dir.

Adım 6. Berraklaştırma işlemi uygulanır. Bir önceki adım sonunda elde edilen bulanık sayılar, berraklaştırma işlemi kullanılarak gerçel sayılara dönüştürülür. Berraklaştırma işlemi için merkezileştirme yöntemi kullanılır.

2.2 Eğrioğlu vd. Yöntemi

Eğrioğlu vd. (2011) yaptıkları çalışmada, aralık uzunluğunu belirlemek için kısıtlı optimizasyona dayalı bir yöntem önermişlerdir. Aralık uzunluğu optimizasyona dayalı yaklaşımla bulunduktan sonra, Chen (1996) tarafından önerilen yöntemi kullanarak öngörülerini hesaplamışlardır. Eğrioğlu vd. (2011) tarafından kullanılan optimizasyon yöntemi altın arama yöntemine dayalı bir süreçtir. Belirtilen süreçte,

$\text{Min } f(x)$

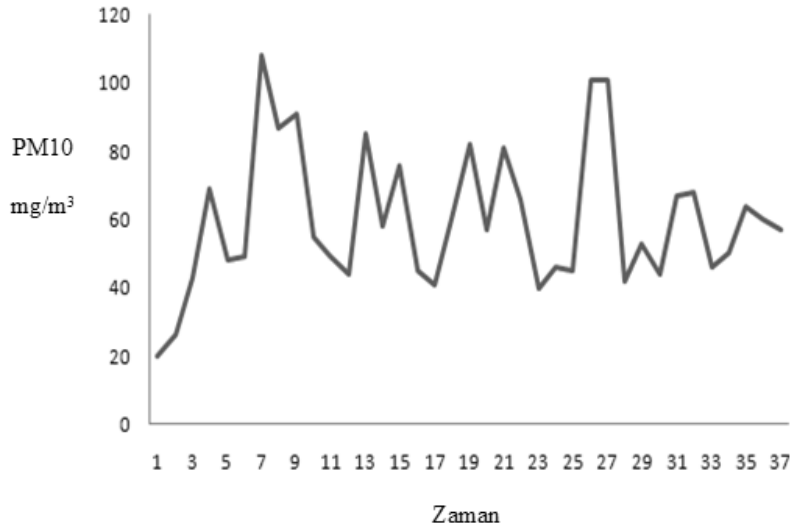
Kısıtlar:

$$x_1 < x < x_2$$

şeklinde verilen optimizasyon problemi çözümlenir. Verilen optimizasyon probleminin çözümü sonucunda, $f(x)$ fonksiyonunu minimum yapan ve (x_1, x_2) aralığında yer alan x değeri bulunur. Verilen minimum problemi, aralık uzunluğunun en iyi değerinin bulunması problemi olarak düşünüldüğünde, $f(x)$ fonksiyonu, x değeri için Chen (1996) yöntemiyle bulunan öngörülerden hesaplanan hata kareler ortalaması karekök (HKOK) değerini vermektedir. Buna göre, belirtilen optimizasyon problemi çözümlendiğinde, (x_1, x_2) aralığında yer alan ve en küçük hata değerini veren aralık uzunluğu bulunmaktadır. Eğrioğlu vd. (2011) belirtilen optimizasyon probleminin çözümünde MATLAB paket programından yararlanmışlardır. Yöntem hakkında detaylı bilgi Eğrioğlu vd. (2011) çalışmasında bulunabilir.

3. UYGULAMA

Uygulamada, 2007 Kasım ile 2010 Ağustos ayları arasındaki, aylık gözlemlerden (PM10 maddesi, mg/m^3) oluşan Ankara hava kalitesi zaman serisi çözümlenmiştir. Çözümlenen serinin grafiği Şekil 1'de verilmiştir.



Şekil 1. Ankara hava kalitesi zaman serisi

Şekil 1'de grafiği görülen zaman serisi 6 farklı bulanık zaman serisi yaklaşımıyla çözümlenmiştir. Kullanılan bulanık zaman serisi yaklaşımlarının uygulanış şekilleri aşağıda özetlenmiştir:

Song ve Chissom (1993b) ve Chen (1996) yaklaşımları kullanırken, aralık uzunluğu olarak 5, 10, 15 ve 20 değerleri denenmiştir. İki yaklaşım içinde 10 aralık uzunluğunun en iyi sonucu ürettiği görülmüştür.

Huang'ın (2001) ortalamaya dayalı ve dağılıma dayalı yaklaşımlarıyla en iyi aralık uzunluğu sırasıyla 20 ve 2 olarak bulunmuştur. Huang ve Yu (2006) önerdikleri yaklaşım için α parametresinin 0.50 değerine yakın alınması gerektiğini belirtmişlerdir. Buna göre, Huang ve Yu (2006) tarafından önerilen yaklaşım kullanılırken α değeri için 0.40, 0.45, 0.50, 0.55 ve 0.60 değerleri denenmiş ve en iyi sonuçlar 0.45 değeri için elde edilmiştir.

Eğrioğlu vd. (2011) tarafından önerilen yaklaşımda, optimizasyon sürecinde en iyi aralık değerinin arandığı aralığın sınırları, veri yapısına uygun olarak 1 ve 20 olarak alınmıştır. Optimizasyon süreci sonunda, en iyi aralık uzunluğu 1.4057 olarak belirlenmiştir. Yapılan tüm işlemlerde MATLAB paket programı kullanılmıştır.

Elde edilen öngörü sonuçlarını daha iyi değerlendirebilmek amacıyla, HKOK, ortalama mutlak hata yüzdesi (OMHY) ve yön sağlamlığı (YS) olmak üzere üç farklı performans ölçütü kullanılmıştır. Kullanılan altı farklı bulanık zaman serisi yönteminden elde edilen tüm sonuçlar Tablo 1'de özetlenmiştir. Tablo 1'de, Ort. ve Dağ. sırasıyla, Huang (2001) tarafından önerilen ortalamaya ve dağılıma dayalı yaklaşımları ve Oran ise Huang ve Yu (2006) tarafından önerilen orana dayalı yaklaşımı göstermektedir. Tablo 1 incelendiğinde, hesaplanan tüm performans ölçütlerine göre, en iyi öngörüler Eğrioğlu vd. (2011) tarafından önerilen bulanık zaman serisi yaklaşımının verdiği görülmüştür. Bir başka ifadeyle, Eğrioğlu vd. (2011) tarafından önerilen yöntem kullanıldığında, en küçük HKOK ve OMHY değerine sahip ve en büyük YS değerine sahip, en güvenilir öngörüler elde edilmiştir.

Tablo1. Tüm yöntemlerde test kümesi üzerinden elde edilen performans ölçütleri değerleri

Ölçütler	Song ve Chissom	Chen	Huarng	Huarng ve Yu	Eğrioğlu vd.	
	(1993b)	(1996)	Ort.(2001) Dağ(2001)	Oran (2006)	(2011)	
HKOK	18.168	18.442	22.665	14.164	18.361	13.127
OMHY	0.272	0.283	0.380	0.149	0.274	0.117
YS	0.743	0.743	0.657	0.829	0.743	0.857

4. SONUÇLAR

Belirsizlik içeren zaman serilerinin çözülmesinde klasik zaman serisi yaklaşımları yetersiz kalmaktadır. Son yıllarda, bu tür zaman serilerinin çözümlenebilmesi için bulanık zaman serisi yaklaşımları geliştirilmekte ve kullanılmaktadır. Bulanık zaman serileri yaklaşımlarında, güvenilir öngörüler elde edilmesinde, en önemli kararlardan biri aralık uzunluğunun doğru olarak belirlenebilmesidir. Bu nedenle, literatürde bu problemin çözümü için önerilen çeşitli yaklaşımlar bulunmaktadır. Yapılan çalışmada, yapısı gereği belirsizlik içeren, 2007 Kasım ile 2010 Ağustos ayları arasındaki, aylık gözlemlerden oluşan Ankara hava kalitesi zaman serisi farklı bulanık zaman serisi öngörü modelleri ile çözümlenmiştir. Çözümlemede, literatürde iyi bilinen Song ve Chissom (1993b) ve Chen (1996) tarafından önerilen bulanık zaman serileri yaklaşımları kullanılmıştır. Ek olarak, aralık uzunluğunun belirlenmesinde, Huarng (2001) tarafından önerilen ortalama ve dağılıma dayalı yaklaşımlar ve Huarng ve Yu (2006) tarafından önerilen orana dayalı yaklaşımın uygulandığı yöntemler de kullanılmıştır. Son olarak, Eğrioğlu vd. (2011) tarafından önerilen, aralık uzunluğunun belirlenmesinde kısıtlı optimizasyona dayalı bir yaklaşımın kullanıldığı, bulanık zaman serisi yöntemiyle de çözümleme yapılmıştır. Ankara hava kalitesi verilerinin ilk kez bulanık zaman serisiyle çözümlendiği bu çalışmada, hem literatürde en çok bilinen iki bulanık öngörü modeli hem de farklı aralık uzunluğu belirleme tekniklerinin kullanıldığı dört farklı bulanık öngörü modeli uygulanarak çözümleme yapılmıştır. Tüm yöntemlerden hesaplanan öngörüler, HKOK, OMHY ve YS olmak üzere üç farklı performans ölçütüne göre değerlendirilmiştir. Elde edilen sonuçlara göre, belirsizlik içeren Ankara hava kalitesi zaman serisi için en güvenilir öngörülerin, kullanılan üç performans ölçütüne göre, Eğrioğlu vd. (2011) tarafından önerilen yöntemden elde edildiği açıkça görülmüştür.

5. KAYNAKLAR

Aladağ, C. H., Başaran, M. A., Eğrioglu, E., Yolcu, U., Uslu, V. R., 2009. Forecasting in High Order Fuzzy Times Series by Using Neural Networks to Define Fuzzy Relations. *Expert Systems with Applications*, 36 (3), 4228-4231.

Aladağ, C. H., Yolcu, U., Eğrioglu, E., 2010. A High Order Fuzzy Time Series Forecasting Model Based on Adaptive Expectation and Artificial Neural Networks. *Mathematics and Computers in Simulation*, 81, 875-882.

Chen, S. M., 1996. Forecasting Enrollments Based on Fuzzy Time-series. *Fuzzy Sets and Systems*, 81, 311-319.

Eğrioglu, E., Aladağ, C. H., Uslu, V. R., Başaran, M. A., Yolcu, U., 2009. A New Hybrid Approach Based on SARIMA and Partial High Order Bivariate Fuzzy Time Series Forecasting Model. *Expert Systems with Applications*, 36 (4), 7424-7434.

Eğrioglu, E., Aladağ, C. H., Başaran, M. A., Yolcu, U., Uslu, V. R., 2011. A New Approach Based on the Optimization of the Length of Intervals in Fuzzy Time Series. *Journal of Intelligent and Fuzzy Systems*, 22, 15-19.

Huang, K., 2001. Effective Length of Intervals to Improve Forecasting in Fuzzy Time-Series. *Fuzzy Sets and Systems*, 123, 387-394.

Huang, K., Yu, H. K., 2006. The Application of Neural Networks to Forecast Fuzzy Time Series. *Physica A*, 363, 481-491.

Song, Q., Chissom, B. S., 1993a. Fuzzy Time Series and its Models. *Fuzzy Sets and Systems*, 54, 269-277.

Song, Q., Chissom, B. S., 1993b. Forecasting Enrollments with Fuzzy Time Series - Part I. *Fuzzy Sets and Systems*, 54, 1-10.

Song, Q., Chissom, B. S., 1994. Forecasting Enrollments with Fuzzy Time Series - Part II. *Fuzzy Sets and Systems*, 62 (1), 1-8.

Zadeh, L. A., 1965. Fuzzy Sets. *Inform and Control*, 8, 338-353.

ANALYSIS OF ANKARA AIR QUALITY DATA WITH FUZZY TIME SERIES

ABSTRACT

Fuzzy time series have been successfully used in recent years to analyze time series which include uncertainty such as air temperature and stock market data. In the literature, fuzzy time series approaches have been studied in order to reach more accurate forecasts. Determining the length of interval correctly is one of the crucial points in obtaining accurate forecasts. Some methods have been proposed in the literature to solve this problem. Air quality data also include uncertainty due to its nature, and such time series should be analyzed by using fuzzy time series. In this study, Ankara air quality time series are analyzed by using the fuzzy time series method for the first time. Six different fuzzy time series approaches, which employ different techniques for determining the length of interval, are used in the implementation, and the results are compared and discussed.

Keywords: Ankara air quality, Length of interval, Fuzzy time series, Optimization, Forecasting.