

Büyük Menderes Akımlarının Frekans Tabanlı Tahmini

Fatih DİKBAŞ¹

ÖZ

Bu çalışmada, mevsimsel salınımlar gösteren, içsel bağımlı ve kaotik zaman serilerinin veriye dayalı tahmini için yeni bir yöntem önerilmektedir. Yöntem, gözlenmiş serilerdeki değer aralıklarının frekanslarına bağlı olarak mevcut verilerin zamansal ve niceliksel ilişkilerine dayalı olarak eksik veriler için tahminler üretmektedir. Frekans tabanlı tahmin adı verilen yöntem, genel bir yaklaşıma sahiptir ve test/onay/düzenleme/ağırlık katsayısı belirleme gibi işlemler yapılması gerekmemektedir. Geliştirilen yöntem, Büyük Menderes havzasındaki 34 istasyonun 9050 adet aylık toplam akım gözleminin tahmini ve 1210 adet eksik verinin tamamlanmasında kullanılmıştır. Gözlemler ve tahminler arasında tüm istasyonlar için elde edilen yüksek korelasyon değerleri, önerilen yöntemin akım verilerinin tahmininde başarılı olduğunu göstermektedir.

Anahtar Kelimeler: Frekans tabanlı tahmin, veriye dayalı modelleme, aylık toplam akım verileri, Büyük Menderes Havzası, eksik verilerin tamamlanması.

ABSTRACT

Frequency Based Prediction of Büyük Menderes Flows

In this study, a new method for the data driven prediction of interrelated and chaotic time series data showing seasonal fluctuations is proposed. The method produces predictions based on the temporal and quantitative relationships among the available data related with the frequencies of the value ranges of observed data. The method, which is called frequency based prediction, has a general approach and requires no testing/validation/adjustment/weight determination steps. The developed method is used for predicting 9050 monthly total flow observations of 34 stations on Büyük Menderes River and for infilling 1210 missing data. High correlations obtained between the observations and predictions for all stations show that the proposed method is successful in the prediction of streamflow data.

Keywords: Frequency based prediction, data-driven modeling, monthly total streamflow data, Büyük Menderes Basin, estimation of missing data.

Not: Bu yazı

- Yayın Kurulu'na 07.08.2014 günü ulaşmıştır.
- 31 Mart 2016 gününe kadar tartışmaya açıktır.

1 Pamukkale Üniversitesi, İnşaat Mühendisliği Bölümü, Denizli - f_dikbas@pau.edu.tr

1. GİRİŞ

Hidrolojik çevrimin bileşenlerinin modellenmesi ve tahmini için mevcut verilerin doğru, hassas, güvenilir ve eksiksiz olması gerekmektedir. Bu verilerin, konuma ve zamana göre niceliksel değişimlerinin belirlenmesi, hidrolojik analizlerde ve su kaynakları sistemlerinin tasarımında önemli rol oynamaktadır. Nehir akımları, hidrolojik çevrimin önemli bir unsurudur ve nehir akımlarının bilimsel olarak incelenmesi için çok sayıda yöntem bulunmaktadır. Aylık toplam akım verileri hidrolojik araştırmalarda sıkça kullanılmaktadır ve akım miktarlarını etkileyen birçok unsur vardır. Akım serileri genellikle mevsimsellik göstermesine rağmen, etkileyen faktörlerin yüksek oranda değişkenlik göstermesi, kaotik ve nispeten rastgele bir davranışa sebep olmaktadır. Bu davranış, akım verilerinin modellenmesini ve önceden tahmin edilmesini zorlaştırmaktadır.

Son dönemlerde, yazılım teknolojilerindeki gelişmelere bağlı olarak, geleneksel hidrolik ve hidrolojik modeller, veriye dayalı yöntemlerle desteklenmeye/tamamlanmaya başlamıştır [1]. Veriye dayalı bir model, bir zaman serisi verisinin analizini içermektedir ancak fiziksel süreçleri ihmal eden bir bilgisayar egzersizi olarak düşünülmemelidir. Akım gözlemleri gibi bir zaman serisindeki konumsal ve zamansal iç ilişkilerin belirlenmesi, matematiksel olarak, akım serilerini oluşturan fonksiyonların arasındaki ilişkilerin belirlenmesi ile eşdeğerdir. Aslında, gözlenmiş akım değeri, akımı oluşturan tüm parametreler arasındaki ilişkileri belirleyen bir fonksiyonun sonucudur. Bu açıdan, doğrudan akım gözlem serileri üzerinde çalışılması, ilgili parametrelerin hiçbirini ihmal etmeyen bir çalışma haline gelir (parametrelerin birbirleriyle ilişkileri ve değişimleri değerlendirilmemiş olsa bile).

Veriye dayalı modelleme tekniklerinin gücü çoktan ispatlanmıştır ve araştırmacılar, veriye dayalı modelleri daha güçlü, anlaşılabilir ve yöneticiler için gerçekten faydalı hale getirmeye çalışmaktadırlar [1]. Hidrolojik süreçlerin modellenmesinde, veriye dayalı modelleme yöntemlerinin kullanıldığı çok sayıda çalışmadan birkaçı şöyle özetlenebilir:

Yapay Sinir Ağları (YSA) yaygın olarak kullanılan bir yöntemdir ve yağış-akış süreçlerinin modellenmesi [2-4]; nehir akımlarının tahmini [5-6], sediment konsantrasyonunun tahmini [7], evapotranspirasyon modellemesi [8] ve yağış şiddet-süre-frekans eğrilerinin belirlenmesi [9] gibi konularda uygulama alanı bulmuştur.

Bulanık kural tabanlı sistemler, kuraklık değerlendirmesi [10], yağış olaylarının tahmini [11], hidrolojik ekstremelerin modellenmesi [12], yağış-akış dinamiklerinin modellenmesi [13] ve taşkın tahmini [14] gibi alanlarda kullanılmıştır.

Destek vektör makineleri son yıllarda araştırmacılar arasında popülerlik kazanmıştır ve yağış akış modellemesi [15], yağış tahmini [16] ve akım tahmini [17-19] kullanıldığı alanlardan bazılarıdır.

Diğer dikkat çekici veriye dayalı modelleme çalışmaları arasında, örnek tabanlı öğrenme [20]; makine öğrenme yöntemleriyle akış tahmini [21] ve YSA, nöro-bulanık ve nöro-genetik algoritma modelleri ile taşkın tahmini [22] gösterilebilir. Veriye dayalı modelleme tekniklerinin hidrolojideki tahmin becerilerinin deneysel bir incelemesi Elshorbagy ve diğ. [23] tarafından sunulmuştur.

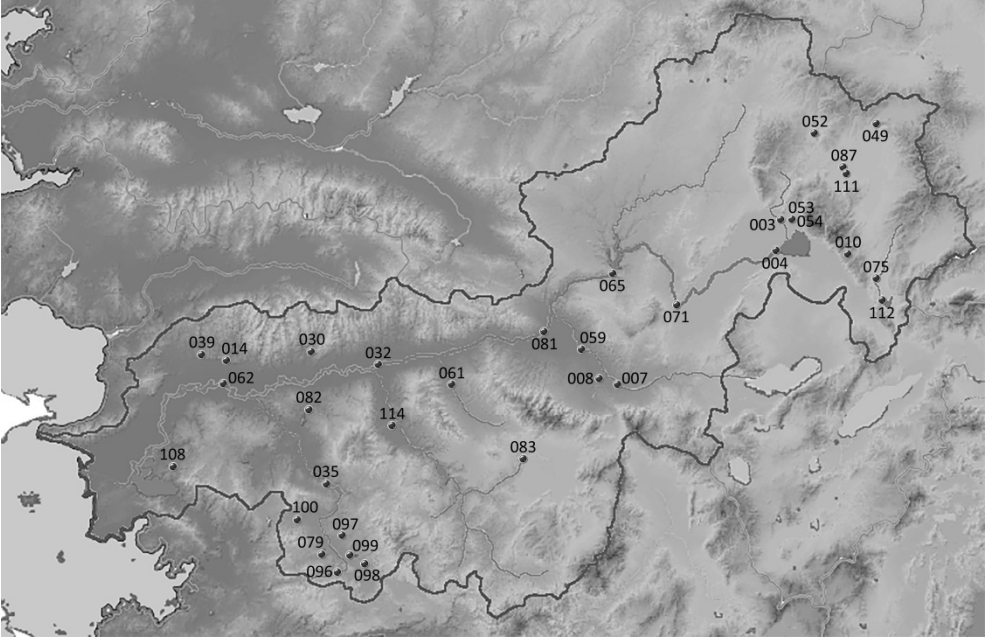
Mevcut çalışmaların birçoğunda, zaman serisi verileri tek boyutlu bir vektör olarak ele alınmaktadır. Genellikle hidrolojik zaman serileri bir yıllık bir mevsimsel çevrime sahiptir ve her satırda tam bir çevrimi içeren iki boyutlu bir matris, değişkenin zamana bağlı

davranışını bir boyutlu bir vektörden daha iyi temsil etmektedir. Örneğin, akım değerleri yıl içinde salınımlar göstermesine rağmen farklı yılların aynı aylarında genellikle benzer değerlere sahip olma eğilimi göstermektedir. Koşullu biçimlendirme ile renklendirilmiş iki boyutlu bir matris bu davranışı iyi bir şekilde görselleştirmektedir. Bu çalışmada, akım gözlem verileri, sütunlarda aylar, satırlarda yıllar olacak şekilde matrislere yerleştirilerek kullanılmıştır.

Önerilen frekans tabanlı tahmin yöntemi, mevsimsel salınımlar gösteren, içsel ilişkiye sahip kaotik zaman serilerinin tahmini için geliştirilmiş veriye dayalı bir metodolojiye sahiptir. Tahminler, gözlenmiş verilerin değer aralıklarının frekansları belirlenerek, mevcut veriler arasındaki zamansal ve niceliksel ilişkilere dayalı olarak belirlenmektedir. Yöntem, herhangi bir eksik veri için, hangi veri aralığının hangi olasılıkla mümkün olacağına dair tahminler üretmektedir. Böylece, hangi değerlerin daha olası, hangi değerlerin daha düşük olasılıklı olduğu değerlendirilebilmektedir. Bu yaklaşımla, bir eksik veri için birden fazla tahmin elde edilebilmektedir. Yöntem genel bir yaklaşıma sahiptir ve mevcut birçok veriye dayalı yöntemin aksine, öğrenme, test, onay, düzenleme, ağırlık katsayısı belirleme, verilerin yumuşatılması gibi aşamaların hiçbirine gerek duyulmamaktadır. Eldeki veriler doğrudan kullanılarak eksik veriler için tahminler bir tıklama ile belirlenmektedir. Önceden belirlenmesi gereken tek değer, veri aralığının bölüneceği maksimum küme sayısıdır ve bu değer, eldeki verinin niceliksel yapısına göre tespit edilmelidir.

Önerilen yöntem, Büyük Menderes Nehri üzerindeki 34 akım gözlem istasyonunun 9050 adet gözlenmiş aylık toplam akım değerinin tahmini ve 1210 adet eksik verinin belirlenmesi için kullanılmıştır. İstasyonlar, nehrin farklı kesimlerini temsil edecek şekilde seçilmiştir, böylece yöntemin, veri uzunluğu ve değerleri açısından farklılık gösteren istasyonların verilerinin tahminindeki başarısı test edilebilmiştir. Seçilen istasyonların konumları Şekil 1'deki harita üzerinde işaretlenmiştir. Adıgüzel Baraj Gölü çıkışındaki 065 Numaralı istasyona Uşak tarafından gelen kolda DSİ akım gözlem istasyonu bulunmadığı için o kolda değerlendirme yapılamamıştır. Zayıf yan kollar ve dereler şekilde gösterilmemiştir.

Değerlendirilen istasyonların gözlem süreleri 8 yıl (07-114 istasyonu) ile 41 yıl (07-003, 004 ve 010 istasyonları) arasında değişmektedir. Yedi istasyonun verileri tamdır ve diğer istasyonlarda eksik veri sayısı 12'den 122'ye kadar değişmektedir. İstasyon başına ortalama eksik veri sayısı 36'dır. İstasyonların büyük çoğunluğunda minimum akım gözlem değeri 0'dır ve bazı istasyonlarda gözlem süresi boyunca onlarca 0 gözlem değeri bulunmaktadır. Tüm istasyonlar için aylık toplam akım verilerinin ortalaması 18.9 hm^3 'tür. Tüm gözlem süresi boyunca en düşük ortalamaya sahip istasyon 0.3 hm^3 ile 07-097 istasyonu, en yüksek ortalamaya sahip istasyon ise 219.3 hm^3 ile 07-062 numaralı istasyondur. İstasyonların gözlem süreleri boyunca gözlenmiş maksimum aylık toplam akım değerleri 2.8 hm^3 (07-097) ile 1121 hm^3 (07-062) arasında değişmektedir. İstasyon kotları 17 m (07-062) ile 1145 m (07-111) aralığındadır. İstasyonların karakteristik bilgileri, tahmin ve gözlemlerin istatistiksel karşılaştırmaları ile birlikte Tablo 5'te sunulmuştur.



Şekil 1. Büyük Menderes Havzası'nda seçilen akım gözlem istasyonları

2. AKIM VERİLERİNİN FREKANS TABANLI TAHMİNİ

Matris formatındaki bir veri setindeki bitişik bir veri çifti, incelenen değişkenin zamansal ve niceliksel davranışını en küçük ölçekte temsil eder. İstatistiksel muhakeme ile veri setindeki bu ilişkilerden bilgi temin edilebilir ve eksik veriler için değer tahmini yapılabilir. Önerilen yöntemin ana fikri, gözlenmiş veri seti matrisindeki tüm bitişik veri çiftlerinin, akım değerlerinin zamansal ve niceliksel değişimi ile ilgili bilgi içerdikleri ve bu bilgiye dayalı olarak etraflarındaki gözlemlerin olası değer aralıklarının tahmin edilebileceğidir.

Yöntemin dayandığı temel fikir, periyodik davranış gösteren veri serilerindeki herhangi bir değer, kendisine yakın gözlemlerle güçlü, kendisinden uzak gözlemlerle zayıf niceliksel ilişki içinde olmasıdır. Önerilen frekans tabanlı tahmin yöntemi, iki boyutlu veri matrisindeki herhangi bir veri hücresinin etrafında bulunan 7 x 7 hücre boyutundaki komşu hücreler (Şekil 2) arasındaki niceliksel ilişkilerin veri setinin diğer kısımlarında tekrarlarına dayalı olarak tahminler üretmektedir. Komşuluk bölgesi 7 x 7 boyutlarında olmak zorunda değildir ancak bu çalışmada incelenen 34 veri setinin tümü için başarılı sonuçlar elde edilmesinde yeterli olmuştur. Daha geniş komşuluk alanı kullanılması işlem süresini gereksiz yere yükseltebilecektir.

Örneğin, bu çalışmada incelenen aylık toplam akım serileri mevsimsel salınım göstermektedir ve yaz aylarında düşük, kış aylarında yüksek değerler gözlenmektedir. Bu nedenle, Ocak ayındaki bir veri tahmin edilirken, yaz aylarında gözlenmiş akım değerlerinin yerine, Ocak ayı merkezde olacak şekilde 7 aylık bir zaman aralığında (Ekim-Nisan), incelenen ayın bulunduğu yıl ortada olacak şekilde, 7 yıllık bir aralıkta gözlenmiş

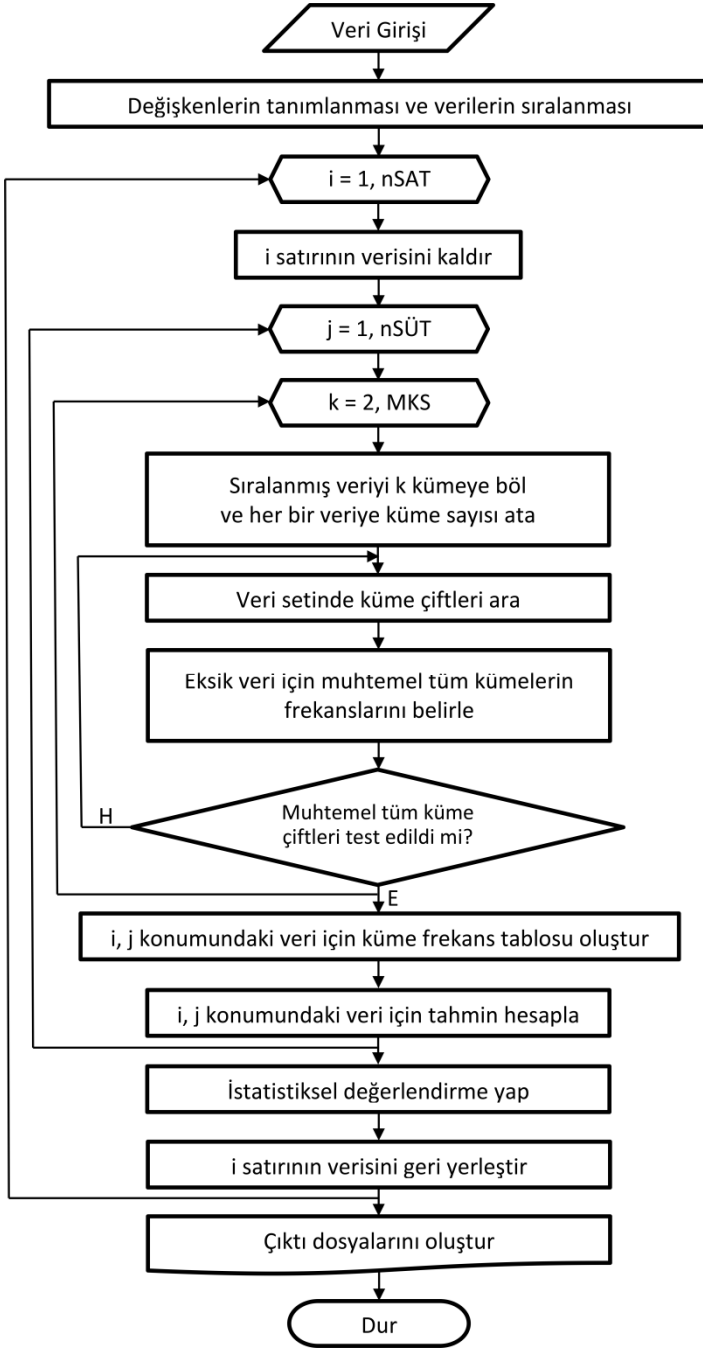
veri çiftlerinin tüm gözlem setindeki frekansları kullanılmaktadır. Benzer şekilde, Ağustos ayındaki bir eksik veri tahmin edilirken, Mayıs-Kasım aralığındaki 7 aylık periyod dikkate alınmaktadır. Böylece, tüm veri serisi için, birbiriyle niceliksel ve zamansal olarak en yakın ilişki içinde bulunan veri çiftlerinin frekanslarına göre tahminler elde edilebilmekte, birbiriyle çok sınırlı hatta negatif ilişki içinde olan veri çiftlerinin tahminlerde oluşturabileceği olumsuz etkiler önlenmektedir.

	EKİM	KASIM	ARALIK	OCAK	ŞUBAT	MART	NİSAN
1979	○	○	○	○	○	○	○
1980	○	○	○	○	○	○	○
1981	○	○	○	○	○	○	○
1982	○	○	○	○	○	○	○
1983	○	○	○	○	○	○	○
1984	○	○	○	○	○	○	○
1985	○	○	○	○	○	○	○

Şekil 2. Küme frekanslarını belirlemek amacıyla veri matrisinde aranacak olan veri çiftleri

Şekil 3'teki akış diyagramı, yöntemin genel uygulama adımlarını göstermektedir. Öncelikle, gözlem serisi dosyadan okunur ve veriler küçükten büyüğe sıralanarak, sıralanmış veriyi ve her bir verinin veri matrisindeki koordinatlarını içeren bir veri matrisi oluşturulur. Koordinat verisi önem taşımaktadır çünkü herhangi bir verinin gözlem zamanı, zaman serilerinin incelenmesinde oldukça önemlidir. Verilerin gözlem zamanları dikkate alınmadan sıralanması ve istatistiksel analizler yapılması, gözlemler arasındaki zamansal ilişkilerin ihmal edilmesi anlamına gelir. Bu çalışmada sıralama işlemi, değer aralığı kümelerinin elemanlarının belirlenmesi amacıyla yapılmaktadır.

Bir değer aralığı kümesi, gözlenmiş tüm verilerdeki en küçük ve en büyük değerler arasındaki tüm sıralanmış gözlemlerin mümkün olduğunca eşit sayıda elemana sahip kümeler oluşturacak şekilde gruplara bölünmesi ile elde edilir. Her bir istasyondaki gözlenmiş değer aralığı, sırasıyla, 2'den başlamak üzere, her tahmin adımında birer artacak şekilde maksimum küme sayısına kadar değer aralığı kümelerine bölünür. Bu çalışmada incelenen aylık toplam akım verileri için maksimum küme sayısı olarak 12 seçilmiştir ve elde edilen sonuçlar yeterli görülmektedir, ancak başka değişkenlerin incelenmesinde farklı sayıda küme sayısı gerekli olabilir. Maksimum küme sayısı, zaman serilerinin gösterdiği davranışa ve gözlenmiş veri miktarına göre belirlenmelidir. Yöntem, gözlenmiş verileri seriden çıkarıp tahminler ürettiği için, farklı küme sayıları denenerek en uygun çözümü veren küme sayısı kolay şekilde belirlenebilmektedir.



Not: nSAT: Satır sayısı, nSÜT: Sütun sayısı, MKS: Maksimum küme sayısı

Şekil 3. Frekans tabanlı tahmin yönteminin akış diyagramı

Kümelerin oluşturulmasında ve gözlem verilerinin hangi kümede bulunduğunu gösteren küme indislerinin belirlenmesinde iki farklı yaklaşım kullanılabilir. Birinci yaklaşımda, her küme mümkün olduğunca eşit sayıda elemana sahiptir ve kümelerin gözlem aralıkları değişkendir. Gözlemlerin kümelere atanması için Denklem 1 kullanılır.

İkinci yaklaşımda ise, her bir küme mümkün olduğunca eşit gözlem aralığına sahiptir ve kümelerdeki eleman sayısı değişkendir. Küme değer aralıkları, kümedeki gözlemlerin minimum ve maksimum değerlerine göre belirlenir. Gözlemlerin kümelere atanması için Denklem 2 kullanılır.

$$K_i = \text{int} \left(\frac{i * n_k}{n_v} \right) + 1 \quad (1)$$

$$K_i = \text{int} \left(\frac{(X_i - X_{\min}) * n_k}{X_{\max} - X_{\min}} \right) + 1 \quad (2)$$

Bu formüllerde:

n_v : Sıralanmış veri vektöründeki gözlem sayısı

i : 1 ile n_v arasında değişen gözlem indisi.

K_i : i indisli gözleme atanacak olan küme indisi (Bu çalışmadaki gözlemler için bu değer 1 ile 12 arasında değişmektedir).

$\text{int}()$: Ondalık sayıyı tamsayıya dönüştüren fonksiyon

n_k : Sıralanmış verinin bölüneceği küme sayısı (Bu çalışmadaki gözlemler için bu değer 12'dir)

X_i : Sıralanmış veri serisindeki i . eleman

X_{\min} , X_{\max} : Tüm veri serisindeki minimum ve maksimum gözlem değerleri.

Bu iki yaklaşımın birbirlerine göre avantajları ve dezavantajları vardır. Uygun kümeleme yönteminin seçimi, gözlenmiş veri serisinin değişkenliği ile bağlantılıdır. Örneğin, eşit aralıklı kümeler oluşturulduğunda, bazı kümelerdeki eleman sayısı diğer kümelere göre çok yüksek oluyorsa, eşit sayıda eleman içeren kümelerin oluşturulması daha doğru bir yaklaşım olabilir. Ancak bu durumda, aykırı değerleri içeren kümelerin değer aralığının artacağı ve yüksek değerlerin tahminlerinin daha küçük çıkabileceği unutulmamalıdır.

2.1 Önerilen Yöntemin Uygulama Örneği

Bu çalışmada önerilen yöntemin esas amacı, zaman serilerindeki eksik verilerin tahmin edilmesidir. Her bir hidrolojik zaman serisi, farklı değerlere sahip bir veri kümesi olduğu için, bir yöntemin tahmin başarısının istasyondan istasyona farklılık göstermesi doğaldır. Hatta, tahmin başarısı, herhangi bir veri setinin farklı kesimleri için değişkenlik gösterebilmektedir. Bu durum, bir istasyon için başarılı tahminler üreten bir yöntemin, başka bir istasyonun verilerinde kullanılmadan önce eldeki verilere göre test edilmesini gerektirir. Bu çalışmada önerilen yöntemin tahmin başarısının test edilmesi için, her bir veri setindeki her bir satır, veri setinden sırasıyla çıkarılmış ve geriye kalan gözlemler

arasındaki ilişkilerden faydalanılarak veri setinden çıkarılan satırlardaki değerler için tahminler hesaplanmıştır. Bu işlemler, geliştirilen yazılım tarafından otomatik olarak gerçekleştirilmekte ve elde edilen tahminler ile gözlenmiş değerler karşılaştırılarak yöntemin tahmin başarısı istatistiksel olarak değerlendirilmektedir. Bu işlemlerde, kasıtlı olarak çıkarılan gözlemlerle birlikte veri setinde bulunan eksik gözlemler için de tahminler hesaplanmaktadır.

Yöntemin uygulama adımlarının daha kolay kavranabilmesi amacıyla, Büyük Menderes havzasındaki 07-010 kodlu Dinar-Irgılı akım gözlem istasyonunun 1982 su yılı aylık toplam akım verilerinin tahmin aşamaları anlatılacaktır. Veri seti 1960-2000 yılları arasındaki 466 gözlemi kapsamaktadır. Gözlem süresince istasyonda aylık ortalama akım $7.7 \text{ m}^3/\text{s}$, aylık toplam akımların ortalaması ise 5.92 hm^3 'tür. En yüksek değer Mayıs 1970'te (18.2 hm^3), en düşük değer ise Haziran, Temmuz ve Ağustos 1995'te gözlenmiştir (0 hm^3). 1974 ve 1975 yıllarına ait tüm veriler ile 1960 yılının Ekim ve Kasım verileri eksiktir. Aylık toplam akım değerlerinin uzun yıllar ortalamaları hesaplandığında en yüksek değer Mart ayı (8.97 hm^3), en düşük değer ise Ağustos ayı (1.73 hm^3) için elde edilmiştir.

Tablo 1. a) Eksik verinin komşuları, b) Komşuların küme indisleri, c) İlk eşleşen bölgenin küme indisleri

	TEM	AĞS	EYL	EKM	KAS	ARL	OCK
1979	0.20	0.20	0.96	3.29	4.13	7.43	10.20
1980	0.24	0.24	0.23	3.33	7.62	9.50	9.13
1981	1.74	1.95	4.94	8.88	10.70	10.90	13.30
1982	0.60	0.75	1.75				
1983				7.05	8.21	8.34	8.74
1984	0.25	0.83	0.73	2.94	7.96	10.00	9.17
1985	2.49	2.89	5.28	6.87	10.00	10.30	10.90

a

	TEM	AĞS	EYL	EKM	KAS	ARL	OCK
1979	3	3	4	6	6	8	10
1980	3	3	3	6	9	10	10
1981	4	4	7	10	11	11	12
1982	4	4	4				
1983				8	9	9	10
1984	3	4	4	5	9	10	10
1985	5	5	7	8	10	11	11

b

	NİS	MAY	HAZ	TEM	AĞS	EYL	EKM
1961	9	7	7	6	5	5	6
1962	9	8	7	6	6	6	7
1963	10	12	12	10	9	10	11
1964	10	9	8	4	3	5	6
1965	11	11	11	9	9	10	12
1966	12	12	11	10	9	10	12
1967	12	12	9	7	6	10	11

c

Veri serisi kümelere ayrıldığında her bir veriye küme indisi atanması yapılmaktadır. Veri serisi 12 kümeye bölündüğünde, 1982 yılı Ekim ayının (tahmin edilecek olan ve Tablo 1a ve b'de mavi çerçeve ile gösterilen hücre) komşularına atanmış olan küme indisi değerleri Tablo 1b'de görülmektedir. Eksik verilere ve yöntemin tahmin performansını değerlendirmek amacıyla matristen çıkarılan verilere küme indisi atanması yapılmamaktadır.

Eksik verinin komşuluk bölgesi içinde şu şekilde sorular üretmek ve bu sorularla bağlantılı ilişkileri matrisin diğer kısımlarında bulmak mümkündür:

Eğer Ekim 1980'deki toplam akım 3.33 hm^3 ve Ekim 1981'deki toplam akım 8.88 hm^3 ise (Tablo 1a'daki kırmızı çerçeve) Ekim 1982'deki toplam akım kaç hm^3 olur?

Eğer Ağustos 1980'deki toplam akım 0.75 hm^3 ve Eylül 1980'deki toplam akım 1.75 hm^3 ise (Tablo 1a'daki sarı çerçeve) Ekim 1982'deki (yani 1 ay sonraki) toplam akım kaç hm^3 olur?

Bu iki soru, küme indisleri kullanılarak şöyle ifade edilebilir:

Eğer Ekim 1980 verisi 6. kümede ve Ekim 1981 verisi 10. kümede ise Ekim 1982 verisi hangi kümede yer alır?

Eğer Ağustos 1980 verisi 4. kümede ve Eylül 1980 verisi de 4. kümede ise Ekim 1982 verisi hangi kümede yer alır?

Yukarıdaki iki soru, tahmin edilecek olan verinin komşuluk alanı içerisinde gözlemlenmiş değerlerin niceliksel ilişkilerine dayanarak ilgili eksik verinin değerini sorgulamaktadır. Mavi çerçeve ile gösterilmiş eksik verinin komşuluk alanı içerisindeki yatay, düşey ve çapraz yönde bitişik veri çiftleri kullanılarak, aranan değer ile ilgili, benzer fakat tekrarlanmayan 158 soru üretilebilir. 1982 yılı Ekim ayının komşuluk bölgesindeki bitişik veri çiftleri Tablo 1.a'da gösterilmiştir. Her bir hücre, veri matrisi üzerinde farklı bir hücrede yer aldığı için, her bir hücrenin 7×7 hücrelik komşuluk bölgesi de o veriye özeldir ve farklıdır.

Örnek olarak verilen bu iki soru ve diğer veri çiftleri kullanılarak oluşturulan soruların cevapları veri matrisi içindeki küme çiftleri ile eşleştirmeler yapılarak aranır. Örneğin birinci sorunun cevabını bulmak için alt alta bulunan 6 ve 10 değerleri tüm veri setinde aranır. İlk eşleşme, Tablo 1.c'de gösterilen veri bölgesinde yer alan Temmuz 1962 ve 1963 kümeleri için tespit edilmiştir. Aynı tabloda, Eylül 1962 ve Eylül 1963 küme indislerinin de sırasıyla 6 ve 10 olduğu görülmektedir. Bu eşleşmelerin altında, değeri sorgulanan Ekim 1982 verisinin konumuna karşı gelen hücrelerdeki küme değerleri sırasıyla 4 ve 5'tir. Aramalar tüm veri setinde sürdürülerek, alt alta bulunan 6 ve 10 değerlerinin altındaki küme değerlerinin frekansları belirlenir. Aramalar, aranan değerlerin komşuluk bölgesindeki toplam 158 çapraz ve yatay yönde bitişik küme çifti için de gerçekleştirilir ve Ekim 1982 verisinin konumuna karşı gelen konumdaki kümelerin frekansları her eşleşmede birer artırılır. Tüm küme çiftlerinin aranması sona erdiğinde, 12 kümeye ait toplam frekans değerleri belirlenmiş olur ve Tablo 2'de Ekim ayı için verilen frekans tablosunun 12 numaralı sütunundaki frekans değerleri elde edilir. En yüksek frekansa sahip küme en büyük olasılığa sahip küme olarak, en düşük frekansa sahip küme ise en düşük olasılığa sahip küme olarak değerlendirilir.

12 küme için yapılmış olan bu işlemler 2'den 11'e kadar tüm kümeleme işlemleri için tekrarlanarak 1982 Ekim ayı verisi için küme frekans tablosu elde edilir. Bunun için tahmin sürecinin başlangıcında, sıralanmış veri seti, bir kümede düşük değerler, diğer kümede yüksek değerler olacak şekilde iki kümeye bölünür, ardından, ilk kümedeki verilere (düşük değerliler) küme indisi olarak 1 değeri, ikinci kümedekilere (yüksek değerliler) ise 2 değeri atanır. Küme indislerinin atanmasından sonra küme frekansı belirleme ve tahmin işlemleri yapılır. İki küme için işlemler tamamlandıktan sonra sıralanmış veri seti, mümkün olduğunca eşit sayıda eleman içerecek şekilde üç kümeye bölünür ve her bir veriye 1 ile 3 arasında değişen küme indisleri atanarak yeni kümeler için frekanslar hesaplanır ve tahminler yapılır. Bu işlemler, her adımda küme sayısı birer artırılarak tekrarlanır ve maksimum küme sayısı için de tahminler oluşturulduktan sonra işlemler sonlandırılır.

Büyük Menderes Akımlarının Frekans Tabanlı Tahmini

Kümeleme işlemleri sonucunda her bir gözlem ve eksik veri için tahminlerin altyapısını oluşturan küme frekans tabloları oluşturulmuş olur (Tablo 2).

Frekans tablolarındaki koyu yeşil bölgeler yüksek olasılıklı veri aralıklarını, kırmızı renkli bölgeler ise düşük olasılıklı veri aralıklarını göstermektedir. Tablo 2'nin sağ tarafındaki "Min" ve "Maks" başlıklı sütunlar, küme sayısının 12 olması durumunda her bir kümenin değer aralıklarını göstermektedir. Örneğin, Ekim 1982 için 12 küme içinde en yüksek frekansa sahip 4. kümenin değer aralığı 0.5-2.0 hm³tür.

Tablo 2'de 07-010 istasyonu 1982 su yılının ayları için oluşturulmuş küme frekans tabloları görülmektedir. Tabloların sütun başlıklarındaki sayılar veri setinin bölündüğü küme sayılarını, satır başlıklarındaki sayılar ise küme indislerini göstermektedir. Örneğin, Ekim ayı için oluşturulan tabloda veri serisi iki kümeye bölündüğünde, ilk kümenin frekansı 7491, ikinci kümenin frekansı ise 12345 olmuştur. Bu değerler, 1982 Ekim ayının aylık toplam akım verisinin daha büyük olasılıkla ikinci kümenin değer aralığında olabileceğini belirtmektedir. Veri seti 12 kümeye bölündüğünde ise en yüksek frekans olan 112 değeri 4. küme için elde edilmiştir. Gözlemleri yüksek oranda değişkenlik gösteren Ekim ve Kasım gibi aylar için elde edilen küme frekans tabloları bulanık iken, değişkenliğin daha az olduğu Mart, Nisan ve Ağustos ayları için elde edilen frekans tablolarında kırmızı ve yeşil bölgeler daha belirgin bir hat izlemektedir.

Tablo 2. 07-010 istasyonunun 1982 verileri için oluşturulan küme frekans tabloları

EKM	2	3	4	5	6	7	8	9	10	11	12	KAS	2	3	4	5	6	7	8	9	10	11	12	Min	Maks	
1	7491	1601	525	234	161	78	27	22	0	0	0	1	6078	1085	283	148	78	48	14	15	0	0	0.0	0.0		
2	12345	4106	1673	741	459	207	151	124	78	63	38	2	16304	4236	1567	541	260	131	96	72	38	35	11	0.0	0.1	
3		4552	2030	1074	679	464	268	173	69	61	46	3		7219	2332	943	621	267	162	93	32	43	21	0.1	0.5	
4			1963	1004	691	453	317	231	198	188	112	4			3296	1323	711	381	349	210	161	141	69	0.5	2.0	
5				1042	521	395	314	167	167	127	87	5				1622	842	470	340	170	138	130	117	2.0	3.3	
6					541	339	228	307	164	81	101	6					927	503	292	305	147	95	77	3.3	4.9	
7						237	244	105	135	134	72	7						468	353	188	186	167	70	4.9	6.3	
8							187	180	95	83	106	8							314	270	157	98	110	6.3	7.4	
9								145	112	89	56	9									258	168	127	116	7.4	8.6
10									68	118	76	10										145	142	98	8.6	10.2
11										66	82	11											140	117	10.2	12.3
12											35	12												107	12.3	18.2
ARL	2	3	4	5	6	7	8	9	10	11	12	OCK	2	3	4	5	6	7	8	9	10	11	12	Min	Maks	
1	5756	773	244	87	54	23	13	2	0	0	0	1	5660	749	201	57	26	6	1	0	0	0	0.0	0.0		
2	20701	4217	1354	450	138	77	42	41	15	18	4	2	23838	4555	1262	472	139	76	63	32	24	1	4	0.0	0.1	
3		9706	2660	928	579	237	126	70	25	23	15	3		11845	2909	1076	531	209	130	56	27	24	16	0.1	0.5	
4			5105	1564	799	355	286	141	139	127	39	4			7002	1809	836	324	257	153	134	110	36	0.5	2.0	
5				2577	1131	463	335	191	125	95	91	5				3274	1276	439	338	253	139	103	118	2.0	3.3	
6					1533	694	372	252	161	122	100	6					2202	782	381	293	203	178	99	3.3	4.9	
7						810	457	300	219	143	73	7						1253	624	342	212	102	102	4.9	6.3	
8							544	366	185	129	136	8							883	431	230	156	103	6.3	7.4	
9								376	282	135	133	9									643	374	213	153	7.4	8.6
10									246	223	162	10										312	355	187	8.6	10.1
11										192	167	11											290	279	10.2	12.3
12											173	12												252	12.3	18.2
ŞUB	2	3	4	5	6	7	8	9	10	11	12	MAR	2	3	4	5	6	7	8	9	10	11	12	Min	Maks	
1	5579	802	226	59	13	7	6	0	0	0	0	1	5547	957	264	90	31	5	7	0	0	0	0	0.0	0.0	
2	24891	4517	1111	444	146	65	51	11	9	4	2	2	22971	3905	957	518	233	75	44	25	11	4	4	0.0	0.1	
3		12965	2842	987	401	194	115	52	36	26	17	3		11721	2541	858	369	113	98	96	43	18	26	0.1	0.5	
4			7090	1874	780	322	193	137	94	74	36	4			7064	1723	729	325	151	94	56	48	42	0.5	2.0	
5				3877	1393	362	335	234	133	91	67	5				3978	1366	319	260	170	114	59	53	2.1	3.3	
6					2765	863	367	263	196	224	102	6					2881	915	342	211	168	180	75	3.3	4.9	
7						1660	587	398	225	85	135	7						1664	605	302	169	103	115	4.9	6.3	
8							1114	531	234	169	87	8							1178	565	258	134	98	6.3	7.4	
9								800	501	274	171	9								716	512	257	116	7.4	8.6	
10									377	453	225	10									419	450	233	8.6	10.2	
11										338	344	11										334	373	10.2	12.3	
12											264	12												251	12.3	18.2

Tablo 2. 07-010 istasyonunun 1982 verileri için oluşturulan küme frekans tabloları (devam)

HAZ	2	3	4	5	6	7	8	9	10	11	12	TEM	2	3	4	5	6	7	8	9	10	11	12	Min	Maks
1	11197	2978	1288	571	291	133	81	34	0	0	0	1	12988	4131	1653	822	464	209	133	54	0	0	0	0.0	0.0
2	11025	3099	1786	1339	1013	629	514	290	155	174	103	2	7405	2158	1845	1369	1441	1009	738	448	310	319	172	0.0	0.1
3		4578	1506	780	574	456	352	255	189	255	199	3		2180	1090	704	520	406	389	417	222	305	274	0.1	0.5
4			2842	773	525	374	190	200	166	161	82	4			1118	446	375	312	189	148	129	186	84	0.5	2.1
5				1756	585	220	235	246	79	78	124	5			687	349	176	169	157	65	63	109	2.1	3.3	
6					1289	435	221	149	85	105	52	6				475	221	163	80	86	74	43	3.4	5.1	
7						788	331	158	102	63	91	7					352	127	107	42	34	71	5.1	6.4	
8							659	293	145	56	39	8						269	111	66	52	15	6.4	7.5	
9								303	283	150	73	9							129	101	59	66	7.5	8.7	
10									216	214	101	10								78	90	26	8.7	10.3	
11										142	200	11									59	73	10.3	12.2	
12											122	12										54	12.3	18.2	

AĞS	2	3	4	5	6	7	8	9	10	11	12	EYL	2	3	4	5	6	7	8	9	10	11	12	Min	Maks
1	13633	4712	1660	905	471	225	151	51	0	0	0	1	14059	4061	1359	729	405	183	107	19	0	0	0	0.0	0.0
2	5891	2333	2311	1552	1671	1184	679	519	359	289	152	2	9620	3465	2816	1451	1430	997	460	403	280	210	132	0.0	0.1
3		1502	920	812	615	429	432	438	248	289	256	3		2826	1581	1124	861	570	493	457	258	289	166	0.1	0.5
4			610	408	357	315	341	140	151	147	170	4			1182	674	545	449	511	230	189	147	231	0.5	2.1
5				253	230	157	126	151	104	113	84	5			544	352	370	258	192	217	188	120	2.1	3.3	
6					121	131	91	70	76	112	75	6				259	229	172	171	92	155	102	3.4	5.1	
7						97	99	73	52	56	45	7					183	137	116	103	146	92	5.1	6.4	
8							90	40	37	31	31	8						179	63	88	69	87	6.4	7.5	
9								39	32	26	31	9							79	68	48	68	7.5	8.7	
10									19	25	24	10								42	47	38	8.7	10.3	
11										16	26	11									29	48	10.3	12.2	
12											7	12										18	12.3	18.2	

2.2. Eksik Verilerin Frekans Tablolarına Göre Hesaplanması

Küme frekans tablolarından faydalanılarak veri setindeki eksik gözlemlerin tahmini için farklı yaklaşımlar kullanılabilir. Örneğin belirli bir değer tahmini yapmak yerine en yüksek frekansa sahip kümelerin değer aralıkları olası tahmin aralıkları olarak değerlendirilebilir. Bunun yanında, frekans tablolarında belirginleşen yeşil tonlu bölgelerin doğrultusunun değer aralık tablosunu kestiği konum, eksik verinin tahmini değeri olarak belirlenebilir. Bu çalışmada, eksik verilerin ve kasıtlı olarak veri setinden çıkarılan tahmin değerlerinin elde edilmesi için izlenen yöntem şu şekildedir: Frekans tablosu oluşturulurken her bir kümenin frekansını oluşturan değerlerin toplamı hesaplanır. Frekans tablosu oluşturma işlemi sona erdiğinde, frekans değerlerini oluşturan gözlemlerin toplam değerleri de belirlenmiş olur. 12 Küme için elde edilen frekanslardan en yüksek 5 tanesi belirlenir ve büyükten küçüğe sıralanır. Ekim 1982 verisi için bu 5 küme, Tablo 2'de de görüldüğü gibi, sırasıyla 4., 8., 6., 5. ve 11. kümelerdir. Bu kümelerin her biri için elde edilmiş toplam gözlem değerleri, küme frekans değerlerine ayrı ayrı bölünerek Ekim 1982 için en olası ilk 5 tahmin belirlenmiş olur. Ekim 1982 ayında gerçek gözlem değeri 4.42 hm³tür ve bu değer 6. kümenin değer aralığındadır ve bu küme için elde edilen tahmin değeri 3.77 hm³tür.

Tahmin işlemleri yılın diğer ayları için de aynı şekilde tekrarlanarak her ay için en olası ilk beş tahmin değeri elde edilir. Tablo 3.a, 1982 yılı gözlemleri ile birlikte, her ay için elde edilen ilk 5 olası tahmin değerlerini ve bu ilk 5 tahmin değeri içinde gözlemlenen gerçek değerlere en yakın tahmin serisini göstermektedir. Tablonun sağ tarafında, tahminler ile gözlem serisi arasındaki korelasyon değerleri verilmiştir. En yakın tahmin değerleri koyu renkle belirginleştirilmiştir. Tahminleri yapılan 12 değer için en yaklaşık değerlerden 6 tanesi ilk tahminde, 2 tanesi ikinci tahminde ve geriye kalan 4 tanesi de üçüncü tahminlerde elde edilmiştir. İlk tahminler ile gözlem serisi arasındaki korelasyon 0.859, ilk üç tahmindeki en yaklaşık tahminler ile gözlem serisi arasındaki korelasyon ise 0.980 olmuştur. Açıkça görüldüğü gibi bu senenin verileri için en iyi yaklaşımın elde edilmesi için ilk üç tahmin yeterli olmuş ve 4. ve 5. tahminlerin gözlemler ile tahminler arasındaki korelasyonun iyileşmesi doğrultusunda bir katkısı olmamıştır. Tekrar belirtilmelidir ki,

yöntemin uygulanmasından önce 1982 yılına ait veriler veri setinden çıkarılmıştır ve işlemlerin hiçbir aşamasında bu değerler yazılım tarafından bilinmemektedir. Önerilen yöntem, veri setinden çıkarılan veya önceden eksik olan her akım verisi için çoklu tahminler üretmektedir.

Önerilen yöntemin tahminlerdeki başarısının karşılaştırılması amacıyla, 07-010 istasyonunun 1982 yılı gözlemleri için, çoklu doğrusal regresyonla 5 farklı tahmin serisi belirlenmiştir. Hesaplamalarda, önerilen yöntemde olduğu gibi, 1982 verileri veri setinden çıkarılmış ve geriye kalan değerlerden faydalanılarak tahminler elde edilmiştir. Gözlem değerleri ve tahminler ile gözlenmiş değerler arasındaki korelasyonlar, en iyi tahmin değerleri ile birlikte Tablo 3.b’de sunulmuştur. Önerilen yöntemin elde ettiği tahminlerin, çoklu doğrusal regresyonla elde edilen sonuçlara göre gerçek değerlere genel olarak daha yakın olduğu görülmüştür.

Tablo 3.a. 07-010 istasyonu 1982 gözlemleri; önerilen frekans tabanlı tahmin yöntemiyle belirlenen en olası ilk 5 tahmin; en yakın tahminler ve seriler arası korelasyonlar.

Ay	EKM	KAS	ARL	OÇK	ŞUB	MAR	NİS	MAY	HAZ	TEM	AĞS	EYL	
Gözlem	4.42	7.98	10.30	10.60	9.55	11.70	11.30	10.40	5.65	0.56	0.93	1.73	Korel.
Tahmin 1	1.15	2.70	14.47	11.13	11.13	11.16	11.21	11.22	11.31	0.26	0.25	0.86	0.859
Tahmin 2	6.66	11.36	11.12	14.36	14.32	14.34	9.32	14.37	0.28	0.04	0.97	0.25	0.898
Tahmin 3	3.77	7.95	9.31	9.27	9.22	9.39	14.42	9.40	2.72	2.68	0.04	0.05	0.926
Tahmin 4	2.63	6.80	6.77	7.98	8.02	8.08	5.56	5.66	14.40	1.27	2.76	2.84	0.522
Tahmin 5	11.19	14.53	7.90	2.61	5.59	5.66	8.15	2.62	0.03	11.31	4.17	4.03	-0.106
En Yakın Thmn	3.77	7.95	11.12	11.13	9.22	11.16	11.21	11.22	2.72	0.26	0.97	0.86	0.980

Tablo 3.b. 07-010 istasyonu 1982 gözlemleri; çoklu doğrusal regresyon yöntemiyle belirlenen 5 tahmin; en yakın tahminler ve seriler arası korelasyonlar.

Ay	EKM	KAS	ARL	OÇK	ŞUB	MAR	NİS	MAY	HAZ	TEM	AĞS	EYL	
Gözlem	4.42	7.98	10.30	10.60	9.55	11.70	11.30	10.40	5.65	0.56	0.93	1.73	Korel.
Tahmin 1	1.07	4.13	17.50	6.13	2.32	8.74	9.90	14.70	2.68	0.21	0.31	5.28	0.681
Tahmin 2	0.50	16.30	8.40	7.11	13.60	15.10	6.91	11.90	0.24	10.20	0.31	0.79	0.581
Tahmin 3	2.97	10.70	7.49	5.20	7.23	7.61	12.90	6.19	0.28	0.06	0.00	2.53	0.798
Tahmin 4	0.04	11.60	3.10	15.00	2.44	8.01	17.70	6.19	14.20	0.42	0.57	0.04	0.603
Tahmin 5	3.21	2.48	17.50	3.27	6.29	15.10	6.91	6.25	0.24	9.82	0.18	10.20	0.294
En Yakın Thmn	3.21	10.70	8.40	7.11	7.23	8.74	9.90	11.90	2.68	0.42	0.57	2.53	0.892

İncelenen veri setinin değişkenlik gösterme durumuna bağlı olarak, elde edilecek tahminlerin sayısı azaltılabilir veya artırılabilir. Bilindiği gibi, nehir akım verileri nispeten kaotik bir davranış göstermektedir ve en olası akım değeri, gözlemlenen akım değeri olarak ortaya çıkmayabilmektedir. Bu nedenle, bir eksik veri için birden fazla tahmin değerinin elde bulunması, akım verileri ile çalışan bilim adamları, uygulayıcılar ve yöneticiler için çok faydalı olacaktır. Bu çalışmada incelenen istasyonların akım verisi serileri için 5 tahmin üretilmesi, başarılı sonuçlar elde edilmesi için yeterli olmuştur.

Bir eksik veri için birden fazla tahmin hesaplanmasının avantajını test etmek amacıyla, artan tahmin sayısına bağlı olarak, gözlem serisi ile en iyi tahmin serileri arasındaki korelasyonlardaki artış incelenmiştir. Tablo 4, 07-010 istasyonunun gözlem serileri ile ilk 2, 3, 4 ve 5 tahmin içindeki en yaklaşık tahminler arasındaki korelasyonları hem her yıl için hem de tüm seri için göstermektedir.

İlk iki tahmin serisi içindeki en iyi tahminler ile gözlenmiş seriler arasındaki yıllık korelasyonların %77'si (30/39) 0.70'in üzerindedir. Bu oran, ilk üç ve dört tahmin için %90'a (35/39) ve ilk beş tahmin için %97'ye (38/39) yükselmektedir. 0.90'ın üzerindeki korelasyonların oranı ise, ilk 2, 3, 4 ve 5 tahmin için sırasıyla %38 (15/39), %56 (22/39), %77 (30/39) ve %85 (33/39) olmuştur.

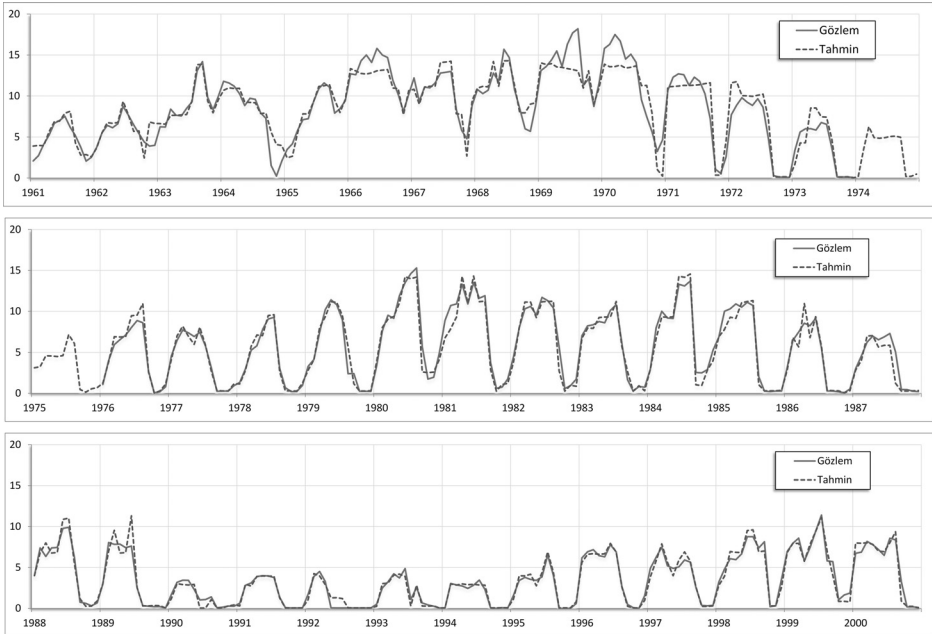
Tablo 4'ün son sütununda, 466 gözlemden oluşan tüm seri ile ilk 2, 3, 4 ve 5 tahmin içinde gözlemlere en yaklaşık değerlerin oluşturduğu seri arasındaki korelasyon değerleri verilmiştir. İlk iki tahmin için elde edilen 0.84 korelasyon değeri uygulamada yeterli kabul edilebilir ancak tahminlerin sayısı artırıldığında korelasyon değerleri 0.9'un üzerine çıkmış ve her ay için 5 tahmin hesaplandığında korelasyon değeri tüm seri için 0.97 olmuştur. Aylık toplam akım serilerinin tahmini için bu korelasyon değeri oldukça yüksek ve güven veren bir değer olarak kabul edilebilir.

Tablo 4. 07-010 İstasyonunun gözlenmiş serileri ile ilk 5 tahmin serileri içindeki en yaklaşık tahminler arasındaki korelasyonlar

Tahminler	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971
1-2	0.46	0.70	0.22	0.64	0.63	0.22	0.93	0.58	0.74	-0.35	0.83	0.95
1-3	0.36	0.72	0.41	0.94	0.85	0.72	0.89	0.83	0.69	0.04	0.86	0.85
1-4	0.34	0.61	0.55	0.97	0.90	0.87	0.88	0.82	0.93	0.68	0.86	0.88
1-5	0.89	0.89	0.70	0.98	0.97	0.97	0.88	0.94	0.92	0.68	0.88	0.93
Tahminler	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983
1-2	0.95	0.90			0.88	0.62	0.91	0.83	0.94	0.81	0.91	0.84
1-3	0.96	0.92			0.96	0.89	0.89	0.96	0.97	0.93	0.98	0.87
1-4	0.97	0.94			0.98	0.91	0.97	0.97	0.97	0.95	0.98	0.96
1-5	0.97	0.91			0.99	0.99	0.99	0.97	0.97	0.98	0.98	0.99
Tahminler	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
1-2	0.94	0.76	0.79	0.78	0.73	0.79	0.59	0.97	0.72	0.95	0.95	0.99
1-3	0.96	0.95	0.93	0.92	0.84	0.91	0.85	0.98	0.78	0.96	0.96	0.99
1-4	0.98	0.97	0.95	0.91	0.94	0.94	0.92	1.00	0.93	0.97	0.98	0.99
1-5	0.99	0.98	0.97	0.92	0.98	0.95	0.93	1.00	0.93	0.97	0.98	0.99
Tahminler	1996	1997	1998	1999	2000	TÜM						
1-2	0.87	0.87	0.90	0.90	0.92	0.84						
1-3	0.94	0.93	0.95	0.90	0.93	0.91						
1-4	0.96	0.98	0.97	0.94	0.97	0.94						
1-5	1.00	0.97	0.97	0.98	0.97	0.97						

Bu sonuçlar, artan tahmin sayısının daha fazla güvenilirlik ve hassasiyet sağladığını göstermektedir ancak bazı durumlarda daha yakın sonuçların daha yüksek korelasyon sağlamadığı unutulmamalıdır. Örneğin, Tablo 4'te, 1966 yılı korelasyon değerlerinin, artan tahmin sayısı ile birlikte azaldığı görülmektedir. Bu durum korelasyon değerinin hesabında kullanılan fonksiyondan kaynaklanmaktadır fakat karşılaştırılan serilerin değerlerinin birbirine yaklaşmasına rağmen korelasyon değerinin azalması durumuyla nadiren karşılaşılmakta ve daha yaklaşık tahminler genellikle daha yüksek korelasyon sağlamaktadır. Hidrolojik değişkenlerin modellenmesinde amaç çoğunlukla daha yaklaşık tahminler elde edebilmek olduğu için istatistiksel değerlendirmede korelasyon katsayısı tek başına yeterli olmamaktadır. Bu nedenle birkaç istatistiksel parametreye birden bakılması daha sağlıklı yorum yapılabilmesini sağlamaktadır.

Şekil 4'teki grafikler, 07-010 istasyonunun gözlemleri ile frekans tabanlı tahmin yönteminin tahminlerini karşılaştırmaktadır. Gözlem süresi içinde akım değerleri niceliksel olarak belirgin değişiklikler göstermiş olmasına rağmen tahminler ile gözlemler arasında genel olarak oldukça iyi bir uyum elde edilmiştir. 1966, 1969 ve 1970 yıllarında gözlemlenen ekstrem değerlerin tahmin hassasiyeti diğer yıllara göre daha düşük kalmıştır. Bu durum, yöntemin kullandığı yaklaşımdan kaynaklanmaktadır. Yöntem herhangi bir veri için en olası değerleri tahmin etmeye çalıştığı için tüm gözlem süresince sadece birkaç defa karşılaşılan ekstrem değerlerin düşük olasılıkları, tahminlerin de düşük değerli kalmasına neden olmaktadır. Yöntemin ekstrem değerleri tahmin becerisi, akım değerleriyle doğrudan bağlantılı olan yağış gibi diğer hidrolojik değişkenlerin zamansal ve konumsal değişimi dikkate alınarak geliştirilebilir.



Şekil 4. 07-010 İstasyonunun gözlemleri ile Frekans Tabanlı Tahmin yönteminin tahminlerinin karşılaştırması

3. FREKANS TABANLI TAHMİN YÖNTEMİNİN DİĞER 33 İSTASYONA UYGULANMASI

Yukarıda sunulan değerlendirmeler sadece 07-010 istasyonunun değerlerinin tahmin aşamalarını kapsamaktadır. Bir yöntemin sadece bir istasyonun değerlerinin tahmininde başarılı olması, diğer istasyonların tahmininde de başarılı olacağı anlamına gelmez. Bu durumu test etmek için, geliştirilen yöntem, Büyük Menderes havzasında bulunan diğer 33 istasyonun verilerini tahmin etmek için de kullanılmıştır. Şekil 1’de de görüldüğü gibi istasyonlar, nehrin farklı kesimlerinin farklı akım özelliklerini iyi bir şekilde temsil edecek şekilde seçilmiştir.

Yöntem, istasyonların 9050 adet gözlenmiş ve 1210 adet eksik aylık toplam akım verilerinin tahmini için kullanılmıştır. Gözlem serileri 2. bölümde anlatıldığı gibi matrislere yerleştirilmiş ve her satır sırasıyla matrislerden çıkarılarak tahminler elde edilmiştir. Tablo 5’te, elde edilen sonuçların istatistiksel değerlendirmeleri ve istasyonlar ile ilgili bazı özellikler sunulmuştur. Yöntemin gözlenmiş değerleri tahmin başarısını sınamak amacıyla, hidrolojik değişkenlerin istatistiksel değerlendirmesinde sıklıkla kullanılan korelasyon katsayısı, Nash-Sutcliffe randıman katsayısı, normalize kök ortalama hata karesi (NRMSE) ve ortalama mutlak ölçekli hata (MASE) istatistikleri tüm istasyonlar için hesaplanmıştır. İstatistiksel değerlendirmelerin yanında, istasyonların gözlem süresi yıl olarak, mevcut aylık gözlemlerin toplam sayısı (n), eksik değerlerin sayısı, minimum, ortalama ve maksimum aylık toplam akım değerleri ve istasyonun kot değeri sunulmuştur.

Tüm istasyonlarda korelasyon değeri 0.75’in üzerinde iken, %88’inde 0.85’in ve %68’inde 0.9’un üzerindedir. En yüksek korelasyon değerleri 0.99 olmuştur ve 07-007 ile 07-054 istasyonları için elde edilmiştir. Ayrıca, 07-039 ve 07-061 istasyonları hariç, gözlem süresi 21 yıldan fazla olan tüm istasyonlarda korelasyon değeri 0.9’un üzerinde çıkmıştır. Bu durum, gözlem süresi artışının tahmin performansına olumlu katkı sağladığını göstermektedir. Nash-Sutcliffe randıman katsayıları 0.47 (07-083 istasyonu) ile 0.98 (07-007 istasyonu) aralığında değişmektedir ve %86’sı 0.70’in, %65’i %80’in ve %21’i de 0.90’ın üzerindedir. NRMSE değerleri 0.05 (istasyonlar:07-007, 008, 032, 039, 054 ve 059) ile 0.15 (istasyon 07-112) arasında değişmektedir ve %88’i 0.1’in altındadır. MASE değerleri ise 0.2 (07-007 istasyonu) ile 0.94 (07-083) arasında değişmektedir ve %71’i 0.4’ün altındadır.

İstatistiksel değerlendirmeler, yöntemin tahmin performansının genel olarak oldukça iyi düzeyde olduğunu göstermektedir. Gözlenmiş seriler ile tahmin serilerinin karşılaştırılması ile elde edilen bu sonuçlar, yöntemin, incelenen 34 istasyona ait aylık toplam akım serilerinin tahminindeki başarısının ispatı için yeterli kanıt sunmaktadır. Daha önce de belirtildiği gibi herhangi bir yöntemin bir zaman serisinin tahmininde başarılı olup olmayacağı söyleyebilmek için ilgili serinin ilgili yöntemle çözümünün yapılması zorunludur. Yine de elde edilen sonuçlar, önerilen yöntemin, diğer akım serilerinin tahmininde de başarılı olacağı doğrultusunda oldukça ümit vericidir. Geliştirilen yöntem genel bir yaklaşıma sahip olduğu için, diğer hidrolojik değişkenlerin ve biyoistatistik, ekonomi ve sosyal bilimler gibi diğer bilim dallarına ait farklı zaman serilerinin ve veri setlerinin tahmininde de kullanılabilir kapasitesi vardır.

Tablo 5. Gözlenen ve tahmin edilen akım serilerinin istatistiksel karşılaştırması ve bazı istasyon parametreleri

İSTASYON	Korel	Na-Su.	NRMSE	MASE	Göz. Yılı	n	Eksik	Min.	Ort.	Maks.	Kot
07-003	0.93	0.83	0.06	0.35	41	480	12	0.0	7.1	95.5	837
07-004	0.93	0.86	0.06	0.36	41	456	36	0.0	33.3	162.0	814
07-007	0.99	0.98	0.05	0.20	16	177	15	0.0	8.0	23.6	260
07-008	0.97	0.93	0.05	0.40	22	264	0	0.0	5.5	12.9	300
07-010	0.97	0.93	0.06	0.48	41	466	26	0.0	5.9	18.2	841
07-014	0.90	0.79	0.06	0.35	39	419	49	0.0	5.7	65.3	70
07-030	0.91	0.82	0.08	0.37	39	408	60	0.0	2.7	20.6	177
07-032	0.97	0.94	0.05	0.37	38	396	60	4.6	112.1	447.0	68
07-035	0.94	0.87	0.06	0.32	36	402	30	0.1	24.1	216.0	112
07-039	0.85	0.68	0.05	0.35	36	381	51	0.0	1.8	36.9	73
07-049	0.93	0.86	0.07	0.38	29	240	108	0.0	2.4	17.7	1025
07-052	0.91	0.81	0.07	0.39	33	381	15	0.0	1.2	11.6	980
07-053	0.94	0.89	0.07	0.61	37	322	122	0.7	9.9	29.8	829
07-054	0.99	0.97	0.05	0.25	30	288	72	0.0	1.6	6.9	829
07-059	0.98	0.96	0.05	0.27	33	321	75	0.0	22.6	61.2	160
07-061	0.88	0.71	0.08	0.53	31	252	120	0.1	5.4	31.3	197
07-062	0.95	0.89	0.06	0.33	33	355	41	4.7	219.3	1121.0	17
07-065	0.93	0.86	0.08	0.39	31	324	48	0.0	64.5	206.0	307
07-071	0.96	0.90	0.08	0.33	31	372	0	1.0	27.5	86.4	758
07-075	0.94	0.88	0.07	0.36	24	239	49	0.0	0.7	4.9	1010
07-079	0.91	0.77	0.08	0.43	21	171	81	0.0	2.0	15.0	355
07-081	0.89	0.78	0.08	0.58	20	216	24	10.0	64.3	249.0	150
07-082	0.89	0.75	0.10	0.45	18	202	14	0.0	2.5	21.7	111
07-083	0.75	0.47	0.10	0.94	18	216	0	0.2	2.8	29.8	855
07-087	0.95	0.89	0.08	0.30	15	180	0	0.0	1.0	5.5	1067
07-096	0.89	0.77	0.08	0.35	13	143	13	0.0	0.7	6.2	450
07-097	0.95	0.85	0.08	0.29	11	120	12	0.0	0.3	2.8	425
07-098	0.78	0.58	0.09	0.38	13	156	0	0.0	0.4	5.7	500
07-099	0.88	0.75	0.09	0.40	13	139	17	0.0	0.9	7.2	395
07-100	0.91	0.79	0.07	0.32	13	156	0	0.0	0.8	8.8	325
07-108	0.90	0.77	0.11	0.39	11	120	12	0.0	0.8	5.2	160
07-111	0.85	0.70	0.07	0.34	10	120	0	0.0	1.0	11.4	1145
07-112	0.82	0.64	0.15	0.93	10	84	36	0.5	4.1	8.1	1005
07-114	0.84	0.66	0.09	0.38	8	84	12	0.0	1.5	16.0	140
Min:	0.75	0.47	0.05	0.20	8	84	0	0.0	0.3	2.8	17
Ort:	0.91	0.81	0.08	0.41	25.1	266.2	36	0.65	18.9	90.2	492.6
Maks:	0.99	0.98	0.15	0.94	41	480	122	10.0	219.3	1121.0	1145

4. TARTIŞMA VE SONUÇLAR

Bu çalışmada, frekans tabanlı tahmin adı verilen veriye dayalı bir metodoloji önerilmiş ve Büyük Menderes Havzası'nda yer alan 34 istasyona ait 9050 adet aylık toplam akım verisinin tahmini ve 1210 adet eksik verinin tamamlanması için kullanılmıştır. Gözlenmiş veri serilerinin tamamı 12'li gruplar halinde yıllık olarak veri matrislerinden çıkarılmış, geriye kalan gözlemler kullanılarak tahminler yapılmıştır. Geliştirilen yöntemin ana hedefi veri serilerindeki eksik verilerin gözlenmiş serilerden faydalanılarak tahmin edilmesidir. Gözlemlerin ve tahminlerin istatistiksel karşılaştırmaları, yöntemin, kasıtlı olarak veri setlerinden çıkarılan tüm veriler için 34 istasyonun tamamında başarılı tahminler hesapladığını göstermektedir. Yöntemin uygulanışı sırasında, veri setlerindeki eksik veriler de tahmin edilmiştir. Önerilen yöntemin avantajları şu şekilde özetlenebilir:

- Yöntemin genel bir yaklaşımı vardır ve iki boyutlu herhangi bir veri setine, tek adımda, herhangi bir kalibrasyon, yumuşatma veya ağırlık katsayısı belirleme işlemi gerektirmeksizin uygulanabilmektedir.
- Tüm eksik veriler için önceden belirlenen sayıda çoklu tahminler belirlenmektedir. Belirlenen tahminler, önerilen yaklaşıma göre en yüksek olasılıklı tahminlerdir ve gözlenmiş serilerle yapılan karşılaştırmalar, hesaplanan tahminlerin tüm istasyonların tüm serileri için başarılı olduğunu göstermektedir. Bu özellik, akım gibi kısmen kaotik davranış gösteren verilerin değerlendirilmesine katkı sağlayacaktır.
- Büyük Menderes Havzası'ndaki 34 akım gözlem istasyonu için elde edilen sonuçlar, yöntemin, aylık toplam akım verilerinin tahmininde güvenilir şekilde kullanılabileceğini göstermektedir.

Teşekkür

Yazar, bu çalışmada kullanılan verileri sağlayan Devlet Su İşleri Genel Müdürlüğü'ne ve makalenin geliştirilmesine katkılarından dolayı danışmanlara ve dergi yayın kuruluna teşekkürlerini sunar.

Kaynaklar

- [1] Solomatine, D. P., Abrahart, R. J., See, L. M., Data-Driven Modelling: Concepts, Approaches and Experiences. Practical Hydroinformatics: Computational Intelligence and Technological Developments in Water Applications, R. J. Abrahart, L. M. See and D. P. Solomatine, (editörler), Springer, Berlin, 17-30, 2008.
- [2] Dawson, C. W., Wilby, R., An Artificial Neural Network Approach to Rainfall-Runoff Modelling. Hydrol. Sci. J., 43(1), 47-66, 1998.
- [3] Govindaraju, R. S., Ramachandra, R.A., Artificial Neural Networks in Hydrology, Kluwer, Dordrecht, 2001.
- [4] Tayfur, G., Singh, V. P., ANN and Fuzzy Logic Models for Simulating Event-Based Rainfall-Runoff. J. Hydraul. Eng., 132(12), 1321-1330, 2006.

- [5] Abrahart, R. J., Anctil, F., Coulibaly, P., Dawson, C. W., Mount, N. J., See, L. M., Shamseldin, A. Y., Solomatine, D. P., Toth, E., Wilby, R. L., Two Decades of Anarchy? Emerging Themes and Outstanding Challenges for Neural Network River Forecasting. *Prog. Phys. Geogr.*, 36(4), 480-513, 2012.
- [6] Huo, Z., Feng, S., Kang, S., Huang, G., Wang, F., Guo, P., Integrated Neural Networks for Monthly River Flow Estimation in Arid Inland Basin of Northwest China. *J. Hydrol.*, 420-421, 159-170, 2012.
- [7] Bhattacharya, B., Van Kessel, T., Solomatine, D. P., Spatio-Temporal Prediction of Suspended Sediment Concentration in the Coastal Zone Using Artificial Neural Network and a Numerical Model. *J. Hydroinform.*, 14(3), 574-594, 2012.
- [8] Adamala, S., Raghuwanshi, N. S., Mishra, A., Tiwari, M. K. Evapotranspiration Modeling Using Second-Order Neural Networks. *J. Hydrol. Eng.*, 19(6), 1131-1140, 2014.
- [9] Mirhosseini, G., Srivastava, P., Fang, X., Developing Rainfall Intensity-Duration-Frequency Curves for Alabama Under Future Climate Scenarios Using Artificial Neural Networks. *J. Hydrol. Eng.*, 10.1061/(ASCE)HE.1943-5584.0000962 , 04014022, 2013.
- [10] Pesti, G., Shrestha, B. P., Duckstein, P., Bogárdi, I. A., Fuzzy Rule-Based Approach to Drought Assessment. *Water Resour. Res.*, 32(6), 1741-1747, 1996.
- [11] Abebe, A. J., Solomatine, D. P., Venneker, R., Application of Adaptive Fuzzy Rule-Based Models for Reconstruction of Missing Precipitation Events. *Hydrol. Sci. J.*, 45 (3), 425-436, 2000.
- [12] Bogardi, I., Duckstein, L., Pongracz, R., Galambosi, A., Experience With Fuzzy-Rule-Based Modeling of Hydrological Extremes. *Proc., Risk-Based Decisionmaking in Water Resources IX*, ASCE, Reston, VA, 44-60, 2001.
- [13] Vernieuwe, H., Georgieva, O., De Baets, B., Pauwels, V. R. N., Verhoest, N. E. C., De Troch, F. P., Comparison of Data-Driven Takagi–Sugeno Models of Rainfall–Discharge Dynamics. *J. Hydrol.*, 302(1-4), 173-186, 2005.
- [14] Nayak, P. C., Explaining Internal Behavior in a Fuzzy If-Then Rule-Based Flood-Forecasting Model. *J. Hydrol. Eng.*, 15(1), 20-28, 2010.
- [15] Bray, M., Han, D., Identification of Support Vector Machines for Runoff Modelling. *J. Hydroinform.*, 6, 265-280, 2004.
- [16] Chen, X., Zhu, S., Improved Hybrid Model Based on Support Vector Regression Machine for Monthly Precipitation Forecasting. *J. Computers*, 8(1), 232-239, 2013.
- [17] Lin, J. Y., Cheng, C. T., Chau, K. W., Using Support Vector Machines for Long-Term Discharge Prediction. *Hydrol. Sci. J.*, 51(4), 599-612, 2006.
- [18] Kişi, Ö., Çimen, M., A Wavelet-Support Vector Machine Conjunction Model for Monthly Streamflow Forecasting. *J. Hydrol.*, 399(1-2), 132-140, 2011.
- [19] Karahan, H., İplikçi, S., Gürarşlan, G., River Flow Estimation From Upstream Flow Records Using Support Vector Machines, *J. Appl. Math.*, 2014

- [20] Solomatine, D. P., Shrestha, D. L., Maskey, M., Instance Based Learning Compared to Other Data-Driven Methods in Hydrological Forecasting. *Hydrol. Process.*, 22, 275-287, 2007.
- [21] Yılmaz, A. G., Muttil, N., Runoff Estimation by Machine Learning Methods and Application to the Euphrates Basin in Turkey. *J. Hydrol. Eng.*, 19(5), 1015-1025, 2014.
- [22] Mukerji, A., Chatterjee, C., Raghuwanshi, N. S., Flood Forecasting Using ANN, Neuro-Fuzzy, and Neuro-GA Models. *J. Hydrol. Eng.*, 14(6), 647-652, 2009.
- [23] Elshorbagy, A., Corzo, G., Srinivasulu, S., Solomatine, D. P., Experimental Investigation of the Predictive Capabilities of Data Driven Modeling Techniques in Hydrology - Part 1: Concepts and methodology. *Hydrol. Earth Syst. Sci.*, 14, 1931-1941, 2010.