

Uygun Stokastik Model Seçim Ölçütlerinin Değerlendirilmesi

Türkay BARAN*
Ülker G. BACANLI**

ÖZ

Stokastik hidrolojide karşılaşılan en önemli problemlerden biri, incelenen süreci tanımlayan modeller arasında en uygun olanın seçimidir. Sunulan çalışmada, gözlenmiş yıllık akımlar ve türetilmiş sentetik diziler kullanılarak, en uygun model seçiminde kullanılan yöntemler karşılaştırılmıştır. Sınamalarda, hidrolojide sık kullanılan AR(1), AR(2), AR(3), ARMA(1,1) ve ARMA(1,2) modelleri değerlendirilmiştir. Ceyhan havzası, Tanır/Gözlerüstü (2015/25) akım gözlem istasyonu yıllık ortalama akış değerleri ve sentetik serilerin kullanıldığı çalışmalarda elde edilen sonuçlar, uygunluk ölçütü performansının veri uzunluğuna paralel olarak arttığını göstermiştir. İkinci grup çalışmalarda uygunluk ölçütlerinin performansları, toplumu ve istatistiksel özellikleri bilinen otuz sentetik örnek üzerinde sınanmış ve elde edilen sonuçlar karşılaştırılmıştır.

ABSTRACT

Evaluation of Goodness of Fit Criterion in Time Series Analysis

The order determination of appropriate model type among various models is one of the most important problems in stochastic hydrology. In the presented study, various types of testing the goodness of fit criterion are compared by using observed annual mean flow and synthetic data sets. For this purpose, AR(1), AR(2), AR(3), ARMA(1,1) and ARMA(1,2) models which are mostly used in hydrology are investigated. In the presented study, Tanır/Gözlerüstü (2015/25) annual mean runoff series of Ceyhan basin and corresponding synthetic series are used. The results based on the performance analysis for criterions versus observation length showed that, testing the goodness of fit criterion performance improved with increasing data length. In the second part of the synthetic series evaluation, the performance of the criterion was tested on the thirty generated samples with well known statistical and population characteristics.

Not: Bu yazı

- Yayın Kurulu'na 13.07.2005 günü ulaşmıştır.
- 31 Aralık 2006 gününe kadar tartışmaya açıktır.

* Dokuz Eylül Üniversitesi, Mühendislik Fakültesi, İzmir – turkay.baran@deu.edu.tr

** Pamukkale Üniversitesi, Mühendislik Fakültesi, Denizli – ugbacanli@pamukkale.edu.tr

1. GİRİŞ

Hidrolojik zaman serileri niteliğindeki gözlem değerlerinin stokastik yapılarının belirlenmesi; tek değişkenli hidrolojik dizilerin sentetik olarak türetilmesi konularındaki çalışmalar 1960'lı yıllardan bu yana artarak sürmektedir.

Su kaynakları projelerinde, projelerin oldukça uzun sayılabilecek ekonomik ömürlerinin yanında, genellikle oldukça kısa gözlemlerden yararlanarak karar verme gerekliliği, karar verme sürecinde karşı karşıya kalınan belirsizliği önemli ölçüde arttırmaktadır. Belirsizliğin azaltılabilmesi için gereken uzun akış verileri bulunmadığında, aynı topluma ait olduğu varsayılan yeni örneklerin türetilmesi ancak modelleme yöntemleri kullanılarak mümkün olmaktadır. Bu nedenle, sentetik seri üretimi amacıyla stokastik modelleme yöntemlerine sıklıkla başvurulur. Modelleme sonucunda, incelenen konuya ve mevcut verinin güvenilirliğine bağlı olarak; gelecekte beklenen akımın (debi, seviye, akış hacmi) tahmini, çok amaçlı haznelerin işletilmesi, akarsuyun su potansiyelinin belirlenmesi, taşkın uyarılarının yapılması, kurak dönemlerde hidroelektrik üretiminin, şehir suyu ve sulama suyunun dağıtımı ve akarsularda ulaşımın planlanması gibi pek çok konuda tahmin yapmak mümkün olabilir. Bir başka deyişle, stokastik modeller planlama/tasarım için veri üretmek ya da süreçlerin gelecekteki değerlerini tahmin etmek/senaryolar üretmek amacıyla, karar verme sürecinin önemli bir bileşeni olarak da kullanılabilir.

Bir hidrolojik sürecin modellenmesi iki aşamada gerçekleştirilir. İlk aşamada, hidrolojik süreçlerin birçoğunda var olan periyodik unsurlar; bazı süreçlerde karşılaşılabilen eğilim (trend) ve sıçrama (jump) gibi ayıklanması gereken deterministik (gerekirci) bileşenler tanımlanır. İkinci aşamada ise, ancak olasılık yapılarıyla tanımlanabilen stokastik bileşenin modellenmesi gerçekleştirilir.

İncelenen sürecin olasılık yapılarının tanımlanmasında, sürecin niteliğine bağlı olarak farklı model tipleri uygulanabilmektedir. Modelleme aşamasında çoğunlukla izlenen yol, süreci tanımlamak için kullanılacak farklı tipte modellerin parametrelerini belirledikten sonra, en uygun model tipinin araştırılmasıdır.

Gerçekte, hidrolojik zaman serilerinin kesin matematiksel modelinin belirlenmesi mümkün değildir. Hidrolojik sürecin toplumu için tanımlanan matematiksel model yaklaşıktır. Model parametreleri sınırlı uzunlukta veriden tahmin edildiği için model parametrelerinin toplum değerleri bilinemez [1]. Farklı modeller arasından, incelenen hidrolojik sürece en uygun model tipinin seçimi, stokastik hidrolojide karşılaşılan en önemli problemlerden biridir. Pratikte, uygun model seçimi - özellikle model tipi ve mertebesinin belirlenmesi - kişisel deneyime bağlı olarak, sezgisel biçimde yapılmaktadır [2].

Gözlenmiş bir stokastik sürece uydurulmak istenen matematik modelin tipi seçilip parametreleri eldeki örnekten tahmin edildikten sonra seçilen modelin gözlenmiş seriye uygunluğunun kontrol edilmesi gerekir. Bu amaçla, minimum hata varyansı [$\min \text{Var}(e)$], korelogram [ACF - Autocorrelation Function], Akaike bilgi ölçütü [AIC - Akaike Information Criterion] gibi çeşitli yöntemler kullanılabilir. Ancak, her uygunluk sınaması yönteminde varolan farklı zayıflıklar nedeniyle, tek bir yöntemi kullanmak çoğu kez mümkün değildir. Bu nedenle, genellikle farklı sınama yöntemlerinin verdiği sonuçlar araştırılarak uygun modele karar vermek yoluna gidilmektedir. İncelenen sürecin toplumu hakkında verilecek kararlar seçilen modele bağlı olacağından, uygun model seçiminde kullanılan yöntemin güvenilirliği de önem taşımaktadır. Model uygunluğu için kullanılan ölçütlerin; incelenen hidrolojik sürecin yapısı, istatistiksel özellikleri, veri uzunluğu gibi çok sayıda özelliğe bağlı olarak oldukça farklı sonuçlar verdiği gösterilmiştir [3, 4, 5].

Sunulan çalışmada, en uygun model seçiminde kullanılan ölçütler, gözlenmiş yıllık akımlar ve türetilmiş sentetik seriler üzerinde yapılan modelleme çalışmalarıyla değerlendirilmiştir. Gözlenmiş veriler, Ceyhan havzasında Hurman suyu üzerinde bulunan Tanır - Gözlerüstü (2015/25) akım gözlem istasyonu (AGI) yıllık akım değerleridir. Tanır – Gözlerüstü AGI yıllık akım gözlemlerinin yanısıra, incelenen yıllık akış dizisinin örnek istatistiklerine uygun olarak türetilen farklı sentetik veri gruplarıyla da uygunluk ölçütlerinin değerlendirmesi yapılmıştır.

Bu değerlendirmeler için kullanılan ilk türetilmiş veri grubu 600 yıl uzunluğundadır. Veri uzunluğunun etkisinin belirlenmesine yönelik olarak incelenen bu veri grubu alt gruplara bölünerek değerlendirilmiş, veri uzunluğunun etkisi araştırılmıştır.

Model tipinin uygunluk ölçütleri üzerindeki etkisinin incelenmesine ve bir önceki adımda yapılan çalışmaların doğrulanmasına yönelik olarak yapılan ikinci grup çalışmada ise, Tanır – Gözlerüstü AGI gözlenmiş verileri ile aynı istatistik parametrelere sahip ve aynı uzunlukta, otuz farklı dizi üretilerek incelenen ölçütlerin öngörülen modelleri yansıtma sıklıkları (oranları) araştırılmıştır.

Stokastik modellerin uygunluk sınamalarında minimum hata varyansı [$\min \text{Var}(e)$], korelogram [ACF - Autocorrelation Function], Akaike bilgi ölçütü [AIC – Akaike Information Criterion] ölçütlerinin yanısıra, düzeltilmiş Akaike bilgi ölçütü [AICC – Modified Akaike Information Criterion] ve nihai hata tahmin ölçütü de[FPE - Final prediction error criterion] değerlendirilmiştir.

2. İNCELENEN MODELLER

2.1. Doğrusal İçsel Bağımlı Modeller (AR)

Doğrusal içsel bağımlı (Linear Autoregressive) veya Markov modelleri olarak da adlandırılan bu modeller

$$x_t = \mu + \sum_{j=1}^p \phi_j(x_{t-j} - \mu) + \varepsilon_t \quad (1)$$

biçiminde ifade edilmektedir. İfadede, p modelin mertebesini göstermekte olup, x_t herhangi bir yılda görülen akış, ε_t bağımsız bir süreç oluşturan, sıfır ortalamalı ve varyansı σ_ε^2 olan normal dağılmış değişken (noise, error term) olarak tanımlanmaktadır. ε_t normal dağılmış olduğundan, değişken de (x_t) normal dağılımlı olmaktadır. p-inci mertebe Markov modeli AR(p) biçiminde tanımlanmakta, model parametreleri (p+2) adet (μ , ϕ_1, \dots, ϕ_p ve σ_ε^2) olmaktadır [6, 7]. Hidrolojide birinci ve ikinci mertebe Markov modelleri [AR(1) ve AR(2)] geniş bir uygulama alanı bulmaktadır [8].

2.2. Doğrusal İçsel Bağımlı - Hareketli Ortalama Modeller (ARMA)

Doğrusal içsel bağımlı süreçlerle hareketli ortalamaların (Autoregressive - Moving Average) karışımı olan bu modellerin genel formları

$$x_t = \mu + \sum_{j=1}^p \phi_j(x_{t-j} - \mu) + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (2)$$

p sayıda içsel bağımlılık parametresi (ϕ_1, \dots, ϕ_p) ve q sayıda hareketli ortalama $(\theta_1, \dots, \theta_q)$ parametresi kullanarak tanımlanabilmektedir. Bu tür modeller kısaca ARMA veya ARMA(p,q) gösterimleriyle tanımlanmakta, ARMA(p,0) modeli AR(p) içsel bağımlı modeline indirgenmektedir. Benzer şekilde, ARMA(0,q) modeli, hareketli ortalama MA(q) modeli ile aynı anlamdadır.

Doğrusal içsel bağımlı AR(p) modellerde olduğu gibi, x_t herhangi bir yılda görülen akış, ε_t bağımsız bir süreç oluşturan, sıfır ortalama ve σ_ε^2 varyans değerlerine sahip, normal dağılmış değişken olarak tanımlanmaktadır. ARMA(p,q) mertebeye bir modelde parametre sayısı $(p + q + 2)$ adet $(\mu, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ ve $\sigma_\varepsilon^2)$ olmaktadır [6, 7, 8, 9].

Model parametrelerin tahmini için çeşitli yöntemler kullanılabilir. Momentler, maksimum olabilirlik ve en küçük kareler en çok kullanılan tahmin yöntemleridir. Sunulan çalışmada, model parametreleri momentler ve maksimum olabilirlik yöntemleriyle hesaplanmıştır.

3. YILLIK AKIŞLARIN MODELLENMESİ ve EN UYGUN MODELİN SEÇİMİ

3.1. Yıllık Akışların Modellenmesi

Hidrolojide, incelenen sürecin modeli seçilip, parametre tahminleri yapıldıktan ve modelin gözlenmiş seriye uygunluğu kontrol edildikten sonra, bu modeli kullanarak istenen uzunlukta sentetik seriler türetilebilir.

Yıllık akış serilerinin, varsa gidiş - eğilim (trend) ve sıçrama (jump) şeklindeki bileşenleri ayıklandıktan sonra, sürecin durağan (stasyon) olduğu kabul edilebilir. Yıllık akış serilerinin bazılarında iç bağımlılık çok küçük olduğundan, bunların bağımsız süreç oldukları da kabul edilebilir [8]. Bununla birlikte, özellikle yeraltı biriktirme haznelerinin akışlara önemli ölçüde katkılarının olabildiği karstik havzalarda, yıllık akış serilerinde de ihmal edilemeyecek mertebelerde stokastik bağımlılıklarla karşılaşmak mümkündür.

Stokastik süreçlerin matematik modellerinin genel ifadesi kapsamında tanımlanan, ε_t kalıntı değerlerinin bağımsız bir süreç oluşturduğu kabul edilmektedir. Seçilen modelin uygunluğunun sınanması için, öncelikle modelin kalıntıları (ε_t) gerçekten bağımsız bir sürece ait olup olmadığının belirlenmesi gereklidir.

Elde edilen korelogramın bağımsız bir sürece ait olduğu hipotezini kontrol etmek için yaygın olarak kullanılan test Anderson testidir. İçsel bağımlılık (Otokorelasyon) katsayısının (r_k) örnekleme dağılımının beklenen değer (ortalama) ve varyansı yaklaşık olarak;

$$E(r_k) = -\frac{1}{N-k} \quad \text{Var}(r_k) = \frac{N-k-1}{(N-k)^2} \quad (3)$$

ifadeleriyle belirlenebilir. İfadelerde N örnek uzunluğunu, k kaydırma (lag) sayısını tanımlamaktadır [6, 7].

Bu çalışmada, modelin kalıntı terimlerinin (ϵ_t) bağımsız seriler olup olmadığı Anderson testinin yanısıra, Box-Pierce portmanto testi uygulanarak da araştırılmıştır. Bu test,

$$Q = N \sum_{k=1}^L r_k^2(\epsilon) \quad (4)$$

istatistiğine dayanır. İfadede $r_k(\epsilon)$ kalıntıların korelogramı olup, L maksimum kaydırma (lag) olarak dikkate alınmıştır. Q istatistiğinin örnekleme dağılımı, serbestlik derecesi (L-p-q) olan Chi-kare (χ^2) dağılımıdır [6, 7, 8, 9].

3.2. En Uygun Modelin Belirlenmesi

Hidrolojik bir sürece uygun olabilecek AR(p), ARMA(p,q) benzeri model parametreleri belirlendikten sonra, modelin parametreleri durağanlık ve tersinirlik testlerinden geçirilir, kalıntıların bağımsız seriler olup olmadığı sınanır (model testing - model identification - diagnostic checking – order selection - order determination - testing the goodness of fit [6, 7, 9]).

Bu testlerin tümünden geçen stokastik modeller arasından, sürece en uygun modelin belirlenmesi aşaması da literatürde [6, 7, 9, 10, 11] benzer terimlerle karşılanmaktadır: model sınanması (model testing), modelin tanılanması (model identification), tanı kontrolü (diagnostic checking), merteye seçimi – belirlenmesi (order selection – order determination), uygunluk sınaması (testing the goodness of fit). İsimlendirme ne olursa olsun, bu aşamada sürecin fiziksel yapısına uygun olduğu düşünülerek korelogramı ve parametreleri belirlenen modeller arasından, incelenen sürece **en uygun olan modelin belirlenmesi** işlemi yapılmaktadır.

Seçilen modelin uygunluğunun sınanması aşamasında, korelogramın [ACF] en uygun modelin belirlenmesi amacıyla kullanılması mümkündür. Kalıntıların (ϵ_t) bağımsız bir sürece ait olduğu hipotezinin sınanması aşamasında kullanılan Anderson testi ifadelerinde, seçilen bir α anlamlılık düzeyinde korelogramın güven bölgesi belirlenebilmektedir. Hesaplanmış olan r_k değerlerinin güven bölgesi dışında kalanların oranı ile seçilen anlamlılık düzeyinin karşılaştırılmasına dayanan bağımsızlık hipotezi kontrolü – dolayısı ile modelin uygunluk sınaması - daha küçük anlamlılık düzeyleri (α) belirlenerek yapılabilir.

Model uygunluğunun sınanmasında kullanılan yöntemlerden biri, sürece ait gözlenmiş (X_t) veriler ile model tahminleri (\hat{X}_t) arasındaki farkların (e_t),

$$e_t = X_t - \hat{X}_t \quad (5)$$

varyans değerinin karşılaştırılmasıdır. Farklı modeller için hesaplanan varyans değerlerinden minimumu [minVar(e)], uygun modele karar vermeye yardımcı bir ölçüt olarak kullanılır.

Akaike tarafından tanımlanmış olan, esas olarak kalıntı varyansı cinsinden hesaplanan ölçütler, en uygun model seçiminde geniş bir kullanım alanı bulmuştur [10, 11]. Bilgi ölçütü olarak da tanımlanan ilk ölçüt, içsel bağımlı (AR) modellerin mertebesinin sınanması amacıyla önerilmiştir [12]. Nihai hata tahmin ölçütü [FPE - Final prediction error criterion] olarak isimlendirilen ölçüt;

Uygun Stokastik Model Seçim Ölçütlerinin Değerlendirilmesi

$$FPE(p) = \hat{\sigma}_\varepsilon^2 \frac{N+p}{N-p} \quad (6)$$

hata varyanslarının gözlenmiş veri uzunluğu ve modelin parametre sayısı dikkate alınarak düzeltilmesi esasına dayanmaktadır. Farklı modeller için hesaplanan nihai hata tahmin ölçütü değerlerinden minimumu [min FPE], uygun modele karar vermeye yardımcı bir ölçüt olarak kullanılmaktadır.

Akaike tarafından önerilen ve daha yaygın olarak kullanılan ölçüt, Akaike bilgi ölçütü [AIC- Akaike information criterion] olarak tanımlanır. Akaike bilgi ölçütü, örnek uzunluğu (N), dağılımın parametre sayısı (k) ve kalıntı varyansının ($\hat{\sigma}_\varepsilon^2$) maksimum olasılık tahmini kullanılarak

$$AIC = N * \ln(\hat{\sigma}_\varepsilon^2) + 2 * k \quad (7)$$

biçiminde ifade edilir [13, 14].

Özellikle ARMA tipi modellerin uygunluk sınamalarında yaygın olarak kullanılan bu ölçüt, parametre sayısı (k); ϕ ve θ katsayıları adedine (p, q) bağlı olarak hesaplanır:

$$k = p + q + 1 \quad (8)$$

Farklı modeller için hesaplanan nihai Akaike bilgi ölçütü değerlerinden minimumu [min AIC], uygun modele karar vermeye yardımcı bir ölçüt olarak kullanılır.

Hurvich ve Tsai [15] Akaike bilgi ölçütünde tarafsız tahminlerin kullanılabilmesi amacıyla Kullback-Leibler yaklaşımı kullanmış ve düzeltilmiş Akaike bilgi ölçütü için [AICC];

$$AICC(p, q) = N * \ln(\hat{\sigma}_\varepsilon^2) + \left(\frac{2(p+q+1) * N}{N-p-q-2} \right) \quad (9)$$

ifadesini önermiştir. Düzeltilmiş Akaike bilgi ölçütünde de, uygun model seçimi için minimum değer [min AICC] aranmaktadır.

Veri sayısı arttıkça Akaike tarafından önerilen ölçütlerin [FPE, AIC, AICC] sayısal değerleri birbirine yaklaşmakta, asimptotik halde ise birbirine eşit olmaktadır.

4. UYGULAMA

4.1. Gözlenmiş Akımlar

Sunulan çalışmada, Ceyhan havzasında bulunan, Hurman suyu Tanır (2015) akım gözlem istasyonunda (AGI) gözlenmiş (Şekil 1) yıllık akım verileri kullanılmıştır. Tanır AGI 1180 m yükseltide olup, 915,2 km² yağış alanına sahiptir. Tanır AGI 1957 – 1995 yılları süresinde yıllık ortalama akımları ortalaması 8,21 m³/s olarak hesaplanmıştır [16]. İstasyon 1996 su yılından itibaren 175 m akış yukarı (menbaa) taşınmış olup Gözlerüstü (2025) adı altında ölçümlere devam edilmektedir. Yeni AGI yükseltisi 1222 m, yağış alanı 914,7 km² olarak verilmiş olup, 1996 – 2000 yılları akış ortalaması 8,26 m³/s'dir [17, 18]. Tanır (2015) ve Gözlerüstü (2025) yıllık ortalama akımları aynı toplumdaki veriler olarak kabul edilebilir niteliktedir. Bu durumda, 1957-2000 yılları (44 yıl) ortalama akış değeri, 8,212 m³/s olmaktadır.

Hurman suyu, Yukarı Ceyhan havza alt bölümünde yer alan Afşin projesi kapsamında olup, karstik formasyonlar gözlenmektedir. Türkiye yüzölçümünün önemli bir kısmında - Güneyde Dalaman havzasından başlayarak, Eşençay, Manavgat, Göksu, Seyhan, Ceyhan, Fırat ve Dicle gibi büyük akarsu havzalarına kadar ulaşan bu kuşak boyunca - akarsu debilerinin önemli bir bölümü karst formasyonlardan kaynaklı pınar katkılarında oluşmaktadır. Türkiye akarsularında pınar katkılarının araştırıldığı çeşitli çalışmalarda, Tanır AGI yıllık akışlarının ortalama 3,50 m³/s'lik kısmının karst pınar katkılarında kaynaklandığı belirlenmiştir [19, 20, 21]. Bu değer göz önüne alındığında, yıllara göre değişimle birlikte, yıllık ortalama akımların % 40 ~ 45 mertebesinde bir kısmının pınar katkılarında olduğu söylenebilir. Dolayısıyla, yıllık ortalama akımların AR(p) tipi modellerle tanımlanması beklenmektedir.

Çalışmanın ilk aşamasında, Tanır/Gözlerüstü (2015/25) AGI 1957-2000 arasında gözlenmiş yıllık ortalama akışlarının, AR(1), AR(2), AR(3), ARMA(1,1) ve ARMA(1,2) modelleri ile tanımlanabilirliği araştırılmıştır. İncelenen modellerin parametreleri belirlendikten sonra, durağanlık (stasyonellik) ve tersinirlik (invertibility) testlerinden geçirilmiştir. Tanır/Gözlerüstü (2015/25) AGI 1956-2000 döneminde gözlenmiş yıllık ortalama akışlarının temel istatistikleri Çizelge 1'de, gidişleri Şekil 2'de sunulmuştur.

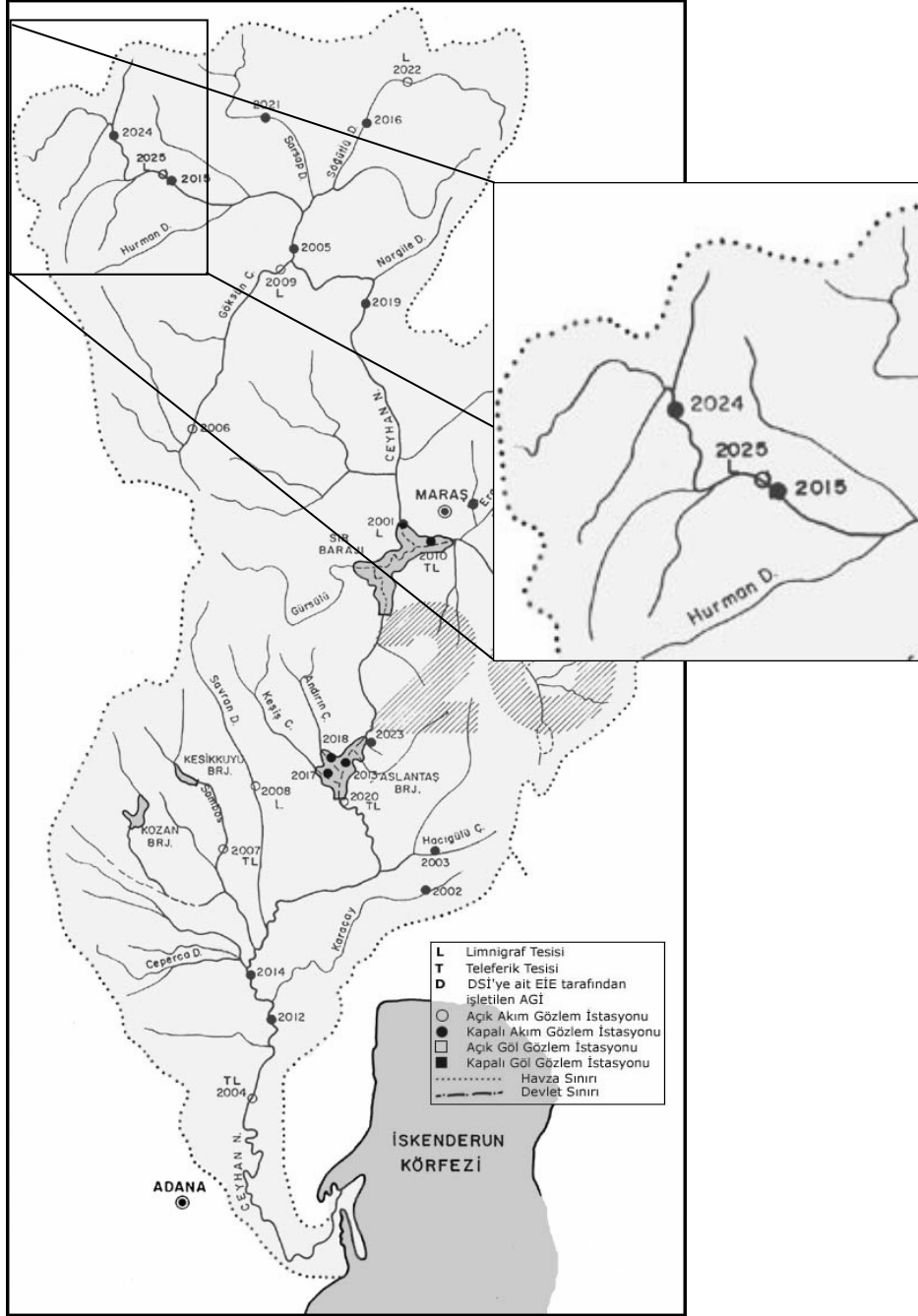
Çizelge 1: Tanır/Gözlerüstü (2015/25) AGI 1957-2000 yıllık ortalama akışlarının temel istatistiksel özellikleri

Tanır/Gözlerüstü(2015/25) yıllık ortalama akışları (1957-2000)	
Ortalama (\bar{x}) (m ³ /sn)	8,212
Standart Sapma (S_x) (m ³ /sn)	2,643
Maksimum Değer (x_{max}) (m ³ /sn)	16,743
Minimum Değer (x_{min}) (m ³ /sn)	4,224
Değişkenlik Katsayısı (C_v)	0,322
Çarpıklık Katsayısı (C_s)	0,917
Sivrilik Katsayısı (E)	1,555

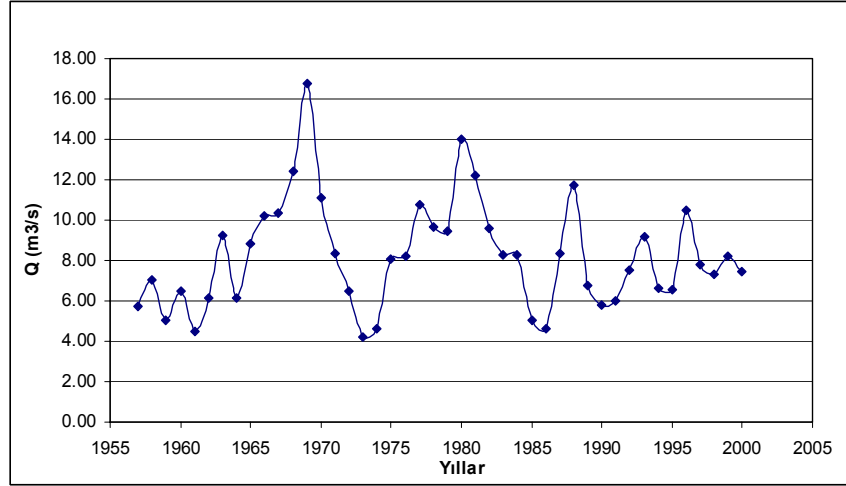
Hurman suyu, Tanır/Gözlerüstü AGI yıllık ortalama akışlarının AR(p) ve ARMA(p,q) tipi modellerle tanımlanabilirliğinin araştırılması aşamasında hesaplanan parametre değerleri Çizelge 2'de sunulmuştur.

Gözlenen akışlara en uygun modeli belirlemede yardımcı olmak üzere, gözlenen akışların içsel bağımlılık (ACF-korelogram) ve kısmi içsel bağımlılık fonksiyonunun (PACF-Partial Autocorrelation Function) ordinatları hesaplanmıştır (Şekil 3 ve 4). Korelogramlar incelendiğinde; içsel bağımlılık katsayılarının belli bir k değerinden sonra birden sıfır olmadığı; büyük k değerinde bile sıfıra yavaşça yaklaştıkları görülmektedir. Bu durum, modelde içsel bağımlılık (AR) terimlerinin bulunduğu göstergesidir. İncelenen modeller için, uygunluk ölçütlerinin sonuçları Çizelge 3'de sunulmuş olup, Tanır/Gözlerüstü yıllık ortalama akışları için en uygun model AR(2) olarak belirlenmiştir.

Uygun Stokastik Model Seçim Ölçütlerinin Değerlendirilmesi



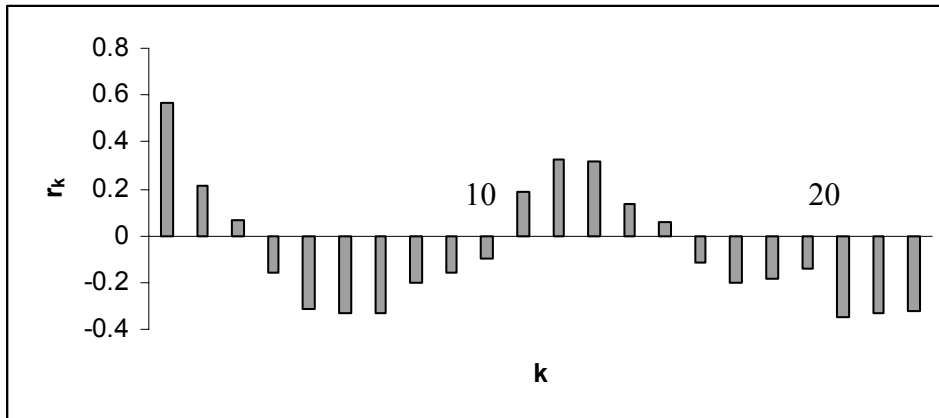
Şekil 1: Ceyhan havzasında kullanılan akım gözlem istasyonu yeri.[18]



Şekil 2: Hurman Suyu Tanır/Gözlerüstü AGI yıllık akımlarının gidişi.

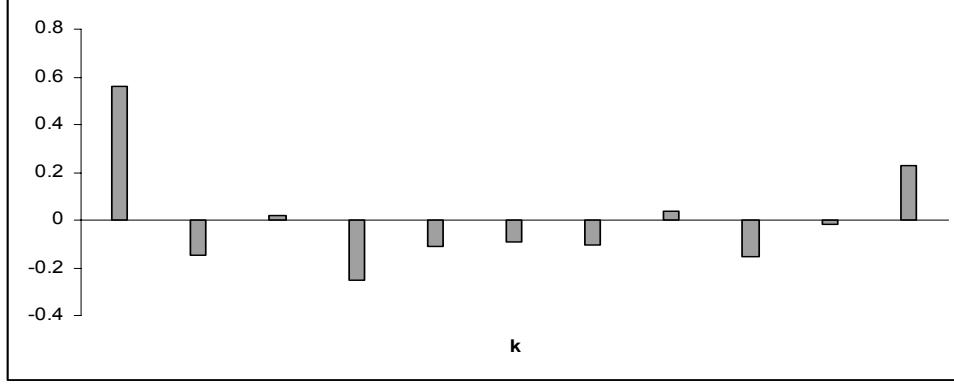
Çizelge 2: Tanır/Gözlerüstü AGI yıllık ortalama akışları doğrusal içsel bağımlı(AR) ve içsel bağımlı – hareketli ortalama (ARMA) model parametreleri.

	AR(1)	AR(2)	AR(3)	ARMA(1,1)	ARMA(1,2)
ϕ_1	0,5652	0,6516	0,6548	0,3812	0,3141
ϕ_2	-	-0,1529	-0,1663	-	-
ϕ_3	-	-	0,0206	-	-
θ_1	-	-	-	-0,2775	-0,3406
θ_2	-	-	-	-	-0,0570



Şekil 3: Hurman Suyu Tanır AGI yıllık ortalama akışlarının korelogramı

Uygun Stokastik Model Seçim Ölçütlerinin Değerlendirilmesi



Şekil 4: Hurman Suyu Tanır AGI yıllık ortalama akışlarının kısmi otokorelasyon katsayıları

Çizelge 3: Tanır/Gözlerüstü AGI yıllık ortalama akışları için en uygun model önerileri

Uygunluk Ölçütü	ACF	Var(e)	AIC	AICC	FPE
Tüm Modeller İçin Değerlendirme	Tüm Modeller	AR(2)-AR(3)	AR(2)-AR(3)	AR(2)	---
İçsel Bağımlı Modeller İçinde Değerlendirme	Tüm Modeller	AR(2)-AR(3)	AR(2)-AR(3)	AR(2)	AR(1)

4.2. Sentetik Seriler

Stokastik modellerde kullanılan uygunluk testlerinin değerlendirilmesi amacıyla, modellenen gözlenmiş dizinin yanısıra sentetik seriler de değerlendirilmiştir. Yıllık akımlarda bu uzunlukta gözlemlere ulaşmak mümkün olmamakla birlikte, uygunluk ölçütlerinin performanslarının veri sayısı ile değişiminin incelenmesi açısından uzun süreli türetilmiş (600 yıl) bir veri grubu kullanılmıştır. Sentetik dizilerin türetilmesinde, Tanır/Gözlerüstü AGI yıllık ortalama akımlarının modellenmesi aşamasında belirlenen AR(2) model parametreleri (Çizelge 2) kullanılmıştır.

Öngörülen AR(2) model parametreleri kullanılarak türetilen, dolayısıyla stokastik yapısı bilinen bu veri grubu, 100, 200, ..., 600 yıl süreli veri alt guruplarına ayrılarak incelenmiştir. Her veri alt grubu, gözlenmiş akış seri kabul edilerek tekrar modellenmiş, en uygun model seçimi için, uygunluk ölçütleri kullanılarak sonuçlar karşılaştırılmış, veri uzunluğunun uygunluk ölçütleri üzerindeki etkisi araştırılmıştır. Değerlendirilen AR ve ARMA modelleri için, çalışma kapsamında incelenen uygunluk ölçütlerinin verdiği en uygun model sonuçları Çizelge 4'de; sadece AR tipi modeller için Çizelge 5'de sunulmuştur.

Çizelge 4: Toplamı AR(2) olarak türetilen sentetik seri alt gurupları için, incelenen AR ve ARMA tipi modellerin uygunluk değerlendirmeleri

Uygunluk Ölçütü Veri Sayısı (N)	ACF	Var(e)	AIC	AICC	FPE
44	Tüm modeller	ARMA(1,2)	ARMA(1,2)	ARMA(1,2)	AR(1)
100	Tüm modeller	AR(2) AR(3)	AR(2)	AR(2)	AR(2)
200	Tüm modeller	AR(2)	AR(2)	AR(2)	AR(2)
300	Tüm modeller	AR(2)	AR(2)	AR(2)	AR(2)
400	Tüm modeller	AR(2)	AR(2)	AR(2)	AR(2)
500	Tüm modeller	AR(2)	AR(2)	AR(2)	AR(2)
600	Tüm modeller	AR(2)	AR(2)	AR(2)	AR(2)

Çizelge 4 ve 5’de sunulmuş olan sonuçlardan da görülebileceği gibi, beklenildiği gibi veri sayısının artmasıyla birlikte tüm uygunluk ölçütlerinin performansı artmaktadır. Uygunluk ölçütlerinin hesaplanmasından önce, incelenen her bir model için Anderson ve Box-Pierce portmanto testleri uygulanmıştır. Korelogram sonuçları, seçilen güven sınırı için - % 95 anlamlılık düzeyinde - tüm modellerin uygun kabul edilebileceğini göstermektedir.

Sentetik seriler ile yapılan ikinci gurup çalışmada, Tanır/Gözlerüstü 1957-2000 yılları arasında gözlenmiş yıllık ortalama akım verileriyle aynı parametre değerlerini koruyan 44 adet veriye sahip otuz farklı sentetik seri türetilmiştir. Bu seriler, aynı toplumdaki gelen farklı gözlem dizileri gibi değerlendirilerek, yıllık ortalama akışların modellenmesi bölümünde açıklanan esaslar çerçevesinde modellenmiştir.

İncelenen doğrusal içsel bağımlı (AR) ve doğrusal içsel bağımlı – hareketli ortalama (ARMA) tipi model parametreleri belirlendikten sonra, her model için Anderson ve Box-Pierce portmanto testleri uygulanmıştır. Korelogram sonuçları, seçilen güven sınırı için -% 95 anlamlılık düzeyinde - tüm modellerin uygun kabul edilebileceğini göstermiş olduğundan, bu bölümdeki karşılaştırmalara dahil edilmemiştir.

İkinci mertebeden, doğrusal içsel bağımlı bir toplumdaki gelen otuz alt serinin her biri üzerinde yapılan modelleme sonucunda, uygunluk ölçütlerinin başarı oranları (mutlak ve nisbi frekanslar) belirlenmiş; incelenen model tipleri ve değerlendirilen uygunluk ölçütleri için elde edilen sonuçlar Çizelge 6’da özetlenmiştir.

Uygun Stokastik Model Seçim Ölçütlerinin Değerlendirilmesi

Çizelge 5: Toplamu AR(2) olarak türetilen sentetik seri alt gurupları için, incelenen AR tipi modellerin uygunluk değerlendirmeleri

Veri Sayısı (N) \ Uygunluk Ölçütü	ACF	Var(e)	AIC	AICC	FPE
44	AR(1)- AR(2)-AR(3)	AR(1)	AR(1)	AR(1)	AR(1)
100	AR(1)- AR(2)-AR(3)	AR(2) AR(3)	AR(2)	AR(2)	AR(2)
200	AR(1)- AR(2)-AR(3)	AR(2)	AR(2)	AR(2)	AR(2)
300	AR(1)- AR(2)-AR(3)	AR(2)	AR(2)	AR(2)	AR(2)
400	AR(1)- AR(2)-AR(3)	AR(2)	AR(2)	AR(2)	AR(2)
500	AR(1)- AR(2)-AR(3)	AR(2)	AR(2)	AR(2)	AR(2)
600	AR(1)- AR(2)-AR(3)	AR(2)	AR(2)	AR(2)	AR(2)

5. SONUÇ ve ÖNERİLER

Hidrolojik zaman serilerinin matematiksel modelinin belirlenmesi aşamasında, model parametreleri sınırlı uzunlukta veriden tahmin edildiği için model parametrelerinin toplum değerleri bilinemez. Bu nedenle, süreci tanımlayabilecek farklı modeller arasından, incelenen hidrolojik sürece en uygun model tipinin seçimi problemi, stokastik hidrolojide karşılaşılan en önemli problemlerden biri niteliğindedir.

Hidrolojik zaman serilerini tanımlayabilecek çeşitli matematik modeller arasından en uygun olanın belirlenmesi amacıyla kullanılan ölçütler; incelenen hidrolojik sürecin yapısı, istatistiksel özellikleri, veri uzunluğu gibi çok sayıda özelliğe bağlı olarak oldukça farklı sonuçlar verebilmektedir. Stokastik hidrolojide yaygın olarak kullanılan minimum hata varyansı [min Var(e)], korelogram [ACF - Autocorrelation Function], Akaike bilgi ölçütü [AIC – Akaike Information Criterion] gibi uygunluk ölçütlerinin, en uygun modele karar verme sürecinde güvenilirliklerinin belirlenmesi amacıyla yapılan bu çalışmanın sonuçları aşağıda maddeler halinde özetlenmiştir.

- Tanır/Gözlerüstü yıllık ortalama akışları için, korelogramlar (ACF ve PACF) birlikte değerlendirildiğinde; modelde içsel bağımlılık (AR) terimlerinin bulunduğu, kısmi içsel bağımlılık fonksiyonunun (PACF) ordinatlarının ikinci adımdan sonra sifıra çok yaklaştığı görülmüştür. İncelenen modeller için, uygunluk ölçütlerinin sonuçları da değerlendirilerek, Tanır/Gözlerüstü yıllık ortalama akışları için en uygun model AR(2) olarak belirlenmiştir.

Çizelge 6: Toplamı AR(2) olarak türetilen sentetik seriler için, incelenen modellerin uygunluk değerlendirmeleri (m: mutlak; f: nisbi frekans).

VAR(e)	İNCELENEN TÜM MODELLER ARASINDA DEĞERLENDİRME		İNCELENEN AR TİPİ MODELLER ARASINDA DEĞERLENDİRME	
	m	f	m	f
AR(1)	0	0	2	0,067
AR(2)	10	0,333	12	0,400
AR(3)	12	0,400	16	0,533
ARMA(1,1)	1	0,033	---	---
ARMA(1,2)	7	0,234	---	---
TOPLAM	30	1,000	30	1,000
AIC	m	f	m	f
AR(1)	1	0,033	6	0,200
AR(2)	12	0,400	15	0,500
AR(3)	9	0,300	9	0,300
ARMA(1,1)	1	0,033	---	---
ARMA(1,2)	7	0,234	---	---
TOPLAM	30	1,000	30	1,000
AICC	m	f	m	f
AR(1)	0	0,000	2	0,066
AR(2)	10	0,333	11	0,367
AR(3)	12	0,400	17	0,567
ARMA(1,1)	1	0,033	---	---
ARMA(1,2)	7	0,234	---	---
TOPLAM	30	1,000	30	1,000
FPE	m	F	m	f
AR(1)	---	---	7	0,233
AR(2)	---	---	16	0,534
AR(3)	---	---	7	0,233
ARMA(1,1)	---	---	---	---
ARMA(1,2)	---	---	---	---
TOPLAM	---	---	30	1,000

- Sürece ait korelogramlar (ACF-PACF), en uygun modelin belirlenmesinde ölçüt olarak kullanılamamakla birlikte, hidrolojik süreci tanımlayan modelin belirlenmesinde yol gösterici olmaktadır. Gerekğinde, güven sınırları daraltılarak en uygun model seçiminde kullanılabilir.
- Uzun süreli sentetik serilerle yapılan çalışmalar, veri sayısındaki artışın tüm uygunluk ölçütlerinin performanslarını olumlu etkilediğini göstermektedir. Seçilen anlamlılık düzeyi için korelogram [ACF] incelenen tüm modelleri kabul ettiğinden, veri uzunluğundan en az etkilenen ölçüt konumundadır.

Uygun Stokastik Model Seçim Ölçütlerinin Değerlendirilmesi

- Kısa süreli gözlemlerde, hiçbir model uygunluk ölçütü güvenilir görünmemektedir (Çizelge 4 ve 5). Ancak, bu çizelgelerde sunulan sonuçların; AR(2) olarak tanımlanmış 600 verili tek bir diziye dayandığı dikkate alınmalıdır.
- Kısa süreli, AR(2) olarak tanımlanmış otuz farklı örnek üzerinde yapılan çalışmaların bulguları (Çizelge 6), bir önceki adımda ulaşılan bulguları doğrulamaktadır. Kısa örnekler için, hiç bir uygunluk ölçütü güvenilir olmayıp, en uygun modeli belirleme oranı en yüksek olan nihai hata tahmin ölçütü [FPE] için dahi, başarı oranı % 53,4 değerine ulaşmaktadır.
- Sentetik seriler kullanılarak yapılmış çalışmalar sonucunda, güvenilirliği en düşük olan ölçüt, minimum hata varyansı [min Var(e)] olarak belirlenmiştir (Çizelge 4 ila 6).
- Akaike bilgi ölçütü [AIC] ve düzeltilmiş Akaike bilgi ölçütü [AICC] ölçütlerinin verdiği sonuçlar incelendiğinde; uzun örnekte belirgin bir performans farklılığı görülmezken, kısa örneklerde Akaike bilgi ölçütünün performansının daha iyi olduğu izlenmektedir (AIC: % 40; AICC: % 33).
- Sadece AR tipi modeller değerlendirildiğinde, kısa örneklerde Akaike bilgi ölçütleri arasında nihai hata tahmin ölçütü [FPE] en yüksek başarı oranına sahiptir (%53,4). Ancak, ikinci sırada bulunan Akaike bilgi ölçütü [AIC] % 50 başarı oranına sahip olup, aradaki fark anlamlı değildir.
- Kısa örnekler için yapılan çalışma sonuçlarında, düzeltilmiş Akaike bilgi ölçütünün [AICC] model mertebesini genellikle yüksek tanımladığı (tüm modeller için % 63,4; AR tipi modeller için % 56,7) izlenmektedir. AR ve ARMA tipi modeller birlikte değerlendirildiğinde, model mertebesinin daha düşük tanımlanma sıklığının sıfıra indiği görülmektedir.
- Akaike bilgi ölçütünün [AIC] performansı daha iyi olmakla birlikte, model mertebesini yüksek belirleme sıklığının oldukça yüksek (% 53,4) olduğu görülmektedir.

Yıllık ortalama akışlar gibi, genellikle çok uzun süreli gözlem dizilerinin bulunamayacağı zaman serilerinin modellenmesinde; en uygun modelin belirlenmesi aşamasında kullanılan ölçütlerin, farklı zayıflıklarının bilinerek değerlendirilmesi gereklidir.

Sadece stokastik hidrolojide değil, istatistiksel ve fiziksel modelleme çalışmalarında da yaygın olarak kullanılan uygunluk ölçütlerinden biri olan minimum hata varyansı [min Var(e)] güvenilirliği en düşük olan ölçüt olarak belirlenmiştir. Bununla birlikte, model mertebesinin daha düşük belirlenmemesi açısından, bu ölçütün düzeltilmiş Akaike bilgi kriteri ile benzer olumlu performans sergilemesi dikkat çekicidir.

Zaman serilerinin modellenerek karar verme sürecinde kullanılacağı dikkate alındığında,, uygunluk ölçütlerinin tümünün değerlendirilmesinin; incelenen hidrolojik sürecin özellikleri de dikkate alınarak en uygun modelin seçiminin yerinde olacağı görülmektedir.

Kaynaklar

- [1]. Salas, J. D., Smith, R. A. Physical basis of stochastic models of annual flows. *Water Resources Research*, 17, 2, 428-430, 1981.
- [2]. Salas, J. D., Obeysekera, J. T. B., Smith, R. A. Identification of streamflow stochastic models. *ASCE Journal of Hydraulics Division*, 107, HY7, 853-868, 1985.
- [3]. Bacanlı, Ü. G., Baran, T. Stokastik modellerde yıllık akım verilerinde uygunluk kriterlerinin değerlendirilmesi. IV Ulusal Hidroloji Kongresi, 23-25 Haziran, İstanbul, Türkiye, 215-225, 2004.
- [4]. Baran, T., Bacanlı, Ü. G. Evaluation of suitability criteria in stochastic modeling. *EWRA Symposium on Water Resources Management: Risks and Challenges for the 21st Century, Proceedings V.II., Theme XI: Modeling as a Tool in Decision Making*, 937 – 948, 2004.
- [5]. Bacanlı, Ü. G. Uncertainty analysis in selection of models best fit to stochastic processes. Ph.D. thesis in Hydraulic-Hydrology and Water Resources Program (Sup.:T.Baran) Izmir, Dokuz Eylül University, Graduate School of Natural and Applied Sciences, Civil Engineering Department, 232 pp., 2004.
- [6]. Salas, J.D., Delleur, J.R., Yevjevich, V., Lane, W.L. Applied modeling of hydrologic time series. *Water Resources Publication*, Littleton-Colorado, 1980.
- [7]. Salas, J.D., Analysis and modeling of hydrologic time series. Mc Graw Hill, in *Handbook of Hydrology* V II, 19.1 – 19.72, New York, 1992.
- [8]. Bayazit, M. Hidrolojide istatistiksel yöntemler. İstanbul, İTÜ, 1981.
- [9]. Box, G.E.P., Jenkins, G. M., Time series analysis forecasting and control. San Francisco, Holden Day, 575 pp., 1976.
- [10]. Fan, J., Yao, Q., Nonlinear time series. Springer, New York, , 2003.
- [11]. Brockwell, P.J., Davis, R.A. Introduction to time series and forecasting. Springer, New York, 2002
- [12]. Akaike, H. Fitting autoregressive model for prediction. *Annals of the Institute of Statistical Mathematics*, 21, 243 - 247, 1969.
- [13]. Akaike, H. Information theory and an extension of the maximum likelihood principle. 2nd International Symposium on Information Theory, (eds. B.N.Petrov and F. Csaki), Akademiai Kiado, Budapest, 267 - 281, 1973.
- [14]. Akaike, H. “A New Look at the Statistical Model Identification”. *IEEE Transactions on Automatic Control*, 19, 716-723, 1974.
- [15]. Hurvich, C. M., Tsai, C. L. Regression and time series model selection in small samples. *Biometrika*, 76, 297-307, 1989.
- [16]. Elektrik İşleri Etüd İdaresi (EIE), Su Akım Aylık Ortalama (1935-1995). Ankara, 2000.
- [17]. Elektrik İşleri Etüd İdaresi (EIE), Su Akımları Yıllığı (1996-2000). Ankara, 2003.
- [18]. Elektrik İşleri Etüd İdaresi (EIE), Hidrolojik çalışmalar, Akarsu havzaları, <http://www.eie.gov.tr/turkce/hidroloji/20ceyhan.html>
- [19]. Baran, T. Türkiye'nin güneyindeki akarsu havzalarının brüt su kuvveti potansiyeli. İzmir, Dokuz Eylül Üniversitesi, Fen Bilimleri Enstitüsü, İnşaat Mühendisliği Böl., Hidroloji ve Su Yapıları Yüksek Lisans Tezi No.15 (Yön.Ü.Öziş), 255 s., 1987.

Uygun Stokastik Model Seçim Ölçütlerinin Değerlendirilmesi

- [20]. Baran, T., Harmancıođlu, N., Öziş, Ü. Türkiye'nin akarsu havzalarında karst pınar katkıları. Ankara, İnşaat Mühendisleri Odası, Türkiye İnşaat Mühendisliği IX. Teknik Kongresi, Bildiriler C II: Su Kaynakları Mühendisliği, s. 299 - 311. 1987.
- [21]. Baran, T., Harmancıođlu, N., Öziş, Ü. Average Base Flow Rates of Karst Spring Effluents in Turkey. Antalya, International Symposium and Field Seminar on Karst Waters and Environmental Impacts, 1995.