İstatistikçiler Dergisi: İstatistik & Aktüerya

Journal of Statisticians: Statistics and Actuarial Sciences

IDIA 11, 2018, 2, 81-92

Geliş/Received:24.05.2018, Kabul/Accepted: 04.11.2018

www.istatistikciler.org

Uygulama Makalesi / Application Article

Genetik-Simpleks hibrit algoritması ile doğrusal olmayan regresyon model parametrelerinin nokta tahmini

Özlem Türkşen

Ankara Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Ankara, Türkiye
turksen@ankara.edu.tr
0000-0002-5592-1830

Fikret Akgün

T.C. Enerji Piyasaları Düzenleme Kurumu Ankara, Türkiye
fikretakgun1990@hotmail.com
0000-0002-6689-9901

Öz

Bu çalışmada, Nelder-Mead Simpleks (NMS) algoritması ve Genetik Algoritma (GA) gibi türevden bağımsız optimizasyon algoritmalarının avantajlı yönlerinin birlikte kullanılması ile oluşturulan bir Genetik-Simpleks hibrit algoritması ile doğrusal olmayan model parametrelerinin nokta tahminlerinin elde edilmesine yer verilmiştir. Ayrıca, Taguchi deney tasarımı ile GA ayarlanabilir parametrelerinin optimal değerlerinin belirlenmesi konusunda çalışılmıştır. Çalışmada, önerilen optimizasyon yaklaşımları kullanılarak, literatürde tanımlı bir veri setine uygun olarak belirlenmiş negatif-üstel regresyon model parametrelerinin nokta tahminleri elde edilmiştir. Bulunan tahmin değerleri, literatürdeki tahmin sonuçları ile karşılaştırıldığında, Genetik-Simpleks hibrit algoritması ile model parametrelerinin tahminlerine kolaylıkla ulaşıldığı ve amaç fonksiyonunun minimum değerine tutarlı bir biçimde yaklaşıldığı gözlenmiştir.

Anahtar sözcükler: Doğrusal olmayan regresyon; Parametrelerin nokta tahmini; Genetik Algoritma (GA); Nelder-Mead Simpleks (NMS) Algoritması; Hibrit algoritma; Taguchi deney tasarımı.

Abstract

Point estimation of nonlinear regression model parameters with Genetic-Simplex hybrid algorithm

In this study, a Genetic-Simplex hybrid algorithm, which is composed of advantageous aspects of derivative-free optimization algorithms, such as Nelder-Mead Simplex (NMS) algorithm and Genetic Algorithm (GA), is used to obtain point estimates of nonlinear regression model parameters. In addition, it is studied to decide optimal values of GA tuning parameters by using Taguchi experimental design. In the study, point estimates of the parameters of a negative-exponential regression model, defined in the literature in accordance with a data set, are obtained by using the proposed optimization approaches. When the obtained results are compared with the results given in the literature, it is seen from the comperative results that estimates of model parameters are easily obtained and consistently approximated to the minimum value of the objective function by using Genetic-Simplex hybrid algorithm.

Keywords: *Nonlinear regression; Point estimations of the parameters; Genetic Algorithm (GA); Nelder-Mead Simplex (NMS) Algorithm; Hybrid algorithm; Taguchi experimental design.*

1. Giriş

Gerçek dünyada karşılaşılan bir probleme ilişkin gözlenen veri seti için, rastgelelik ve rastgele hata kavramlarını göz ardı etmeden, veri kümesindeki açıklayıcı ve açıklanan değişkenler arasındaki fonksiyonel ilişki yapısının minimum hata ile tanımlanması istenir. Bu amaçla en çok kullanılan istatistiksel modelleme yöntemi regresyon analizidir. Doğrusal regresyon modellemesine yönelik analizlerin kullanımı yaygın olmasına rağmen, mevcut veri kümesinde açıklanan ve açıklayıcı değişkenler arasındaki ilişki yapısının oluşturulmasında doğrusal olmayan regresyon analizinden de yararlanılır. Matematiksel olarak bir modelin parametrelerine göre doğrusal olmaması, o modelin parametrelerine göre birinci türevleri (kısmi türevleri) alındığında elde edilen sonuçların model parametrelerine bağlı olması durumu olarak tanımlanabilir [2].

Doğrusal olmayan regresyon model parametrelerinin nokta tahminlerinin elde edilmesine ilişkin yapılan hesaplamalarda güçlüklerle karşılaşılmasına rağmen Marquardt [13], Hartley ve Booker [9], Gallant [6], Seber ve Wild [16] doğrusal olmayan regresyon modellerinde istatistiksel sonuç çıkarımı konusunda çalışmalar yapmışlardır. Doğrusal olmayan regresyon modellerinde parametre tahmin edicilerinin analitik ifadesi kolaylıkla elde edilemediğinden, Jennrich [11], Malinvaud [12], Box [4], Gallant ve Fuller [5] çalışmalarında yer verdikleri tahmin ve hipotez testi konuları için asimptotik ifadelerden ve doğrusal yaklaşımlardan faydalanmışlardır. Bates ve Watts [3], çalışmalarında doğrusal olmayan regresyon modelleri konusunda açıklayıcı örneklere yer vermişlerdir.

Yapılan temel çalışmalar dikkatle incelendiğinde, doğrusal olmayan regresyon model parametrelerinin tahmini için en çok uygulanan yaklaşımın, türeve dayalı yinelemeli algoritmalar (örneğin, Gauss-Newton, En Hızlı İniş, Levenberg-Marquardt) kullanılarak, hata kareler toplamının en küçüklemesine dayalı olarak oluşturulan en küçük kareler (EKK) yönteminin uygulanması olduğu görülmüştür. Fakat, türeve dayalı algoritmalar ile yapılan eniyileme çalışmalarında problem için global değere yakınsayamama, yerel çözüm tuzaklarına takılma gibi sorunlar ortaya çıkabilmektedir. Ayrıca, bu yöntemler başlangıç çözüm değeri bilgisine ihtiyaç duyduğundan, parametrelerin başlangıç çözüm değerlerinin iyi belirlenememesi durumunda problem için optimal çözüme ulaşmanın oldukça zaman alıcı ve bazen mümkün olmadığı görülmüştür. Bu gibi dezavantajlarından dolayı türeve dayalı optimizasyon algoritmalarına alternatif olarak türevden bağımsız optimizasyon algoritmaları ile doğrusal olmayan regresyon model parametrelerinin nokta tahminlerinin elde edilmesi de mümkündür.

Bu çalışmada, türevden bağımsız arama algoritmaları olan Nelder-Mead Simleks (NMS) algoritması ve Genetik Algoritma (GA)'nın avantajlı yönlerinin bir arada kullanılması ile oluşturulan, Türkşen [18], Türkşen ve Tez [19] çalışmalarında önerdikleri hibrit algoritma yaklaşımına benzer olarak tanımlanan bir Genetik-Simpleks hibrit algoritması ile doğrusal olmayan regresyon model parametrelerinin nokta tahminlerinin elde edilmesine yer verilmiştir. NMS algoritması, Nelder-Mead [14] tarafından geliştirilmiş simpleks esaslı bir arama algoritması olup, bölgesel olarak tanımlanmış bir alanda türevden bağımsız yaklaşımla tek nokta aramaya dayalı, pratik ve etkili sonuçlar veren bir optimizasyon algoritmasıdır. GA ise, temelleri Holland [10] tarafından evrim mekanizmasının bilgisayar sistemine aktarılması ile oluşturulmuş, popülasyon tabanlı aramaya dayalı başlangıç çözüm değeri bilgisine ihtiyaç duymayan sezgisel bir optimizasyon algoritmasıdır. GA'nın doğrusal olmayan regresyon model parametrelerinin tahmininde kullanımı Altunkaynak ve Esin [1] çalışmalarında görülebilir.

NMS algoritmasının ve GA'nın ayarlanabilir (tunable / tuning) parametrelerinin probleme uygun tanımlanması, en iyi çözümün elde edilmesinde oldukça önemlidir. Özellikle GA ile elde edilecek model parametre tahmin değerleri, GA'nın ayarlanabilir parametre değerlerinden oldukça etkilenebilmektedir. Bu nedenle, GA için etkili bir ayarlanabilir parametre değerlerini belirleme yaklaşımına ihtiyaç olduğu düşünülmüştür. Bu çalışmada, en az sayıda deney ile en uygun GA ayarlanabilir parametre kombinasyonuna ulaşmak için etkili bir deney tasarımı yöntemi olan Taguchi deney tasarımından yararlanılmıştır. GA ayarlanabilir parametrelerinin olabilecek çok sayıda farklı kombinasyonları arasından seçim yapmak, güvenilir ve detaylı bir hesaplama yaklaşımı olmasına rağmen zaman alıcı olabilmektedir. Bu nedenle, en uygun GA ayarlanabilir parametre kombinasyonunun seçimine ihtiyaç duymadan yürütülen Genetik-Simpleks hibrit algoritması ile de oldukça düşük hata değerine sahip parametre tahminlerine ulaşıldığı gözlenmiştir.

Çalışmanın 2. Bölümünde, doğrusal olmayan regresyon modellerinin genel yapısı tanımlanarak, model parametrelerinin nokta tahminlerini elde etmek amacıyla EKK yöntemi kullanılarak oluşturulan amaç fonksiyonlarının belirlenmesine yer verilmiştir. Çalışmanın 3. Bölümünde, doğrusal olmayan regresyon modellerinin nokta tahminlerinin bulunmasında kullanılan türevden bağımsız optimizasyon algoritmaları NMS ve GA hakkında bilgiler verilerek, bu iki algoritmanın hibriti ile oluşturulmuş olan Genetik-Simpleks hibrit algoritmasının adımları tanımlanmıştır. Ayrıca, GA'nın ayarlanabilir parametrelere düzeylerinin belirlenmesinde kullanılan Taguchi deney tasarım yöntemine yer verilmiştir. Çalışmanın 4. Bölümünde, literatürde tanımlı, Bates ve Watts [3] tarafından yapılan çalışmadan alınmış olan, "mikroorganizmaların biyokimyasal oksijen ihtiyacı (Biochemical Oxygen Demand-BOD)" olarak adlandırılmış veri setine ilişkin, negatif-üstel regresyon yapısına göre modellenmiş olan bir doğrusal olmayan regresyon modeli incelenmiştir. Analiz sonuçlarının değerlendirilmesine çalışmanın 5. Bölümünde yer verilmiştir.

2. Doğrusal olmayan regresyon modeli

2.1. Genel matematiksel yapısı

Parametrelerine göre doğrusal olmayan regresyon modelinin genel matematiksel yapısı vektörel formda

$$\boldsymbol{Y} = f(\boldsymbol{X},\boldsymbol{\theta}) + \boldsymbol{\varepsilon} \tag{1}$$

biçiminde ifade edilebilir. Burada, $\boldsymbol{Y}$, $n \times 1$ boyutlu açıklayıcı değişken vektörü; $\boldsymbol{X}$, $n \times q$ boyutlu tasarım matrisi; $\boldsymbol{\theta}$, $q \times 1$ boyutlu model parametre vektörü ve $\boldsymbol{\varepsilon}$, $n \times 1$ boyutlu rastgele hata vektörüdür. Modelin rastgele hata vektörünün, $E(\boldsymbol{\varepsilon}) = \boldsymbol{0}$, $C\mathrm{ov}(\boldsymbol{\varepsilon}) = \sigma^2 I$ ve $\boldsymbol{\varepsilon} \sim (\boldsymbol{0}, \sigma^2 I)$ varsayımlarını sağladığı kabul edilir. Ayrıca, f fonksiyonunun biçimsel olarak bilindiği, sürekli olduğu, parametre vektörüne göre en az iki defa türevinin alınabildiği de varsayımlar arasındadır.

2.2. Parametrelerin nokta tahmini için amaç fonksiyonu

Eşitlik (1) ile tanımlı modelde, f fonksiyonu biçimsel olarak bilinmesine rağmen oluşturduğu yüzey doğrusal değildir. Model parametrelerinin nokta tahminlerini elde etmek amacıyla f fonksiyonunun dönüşüm yapılarak doğrusallaştırılması durumunda, parametrelerin fiziksel gerçekliğine uygun olarak tahmin edilememesi söz konusu olabilir. Bu nedenle, doğrusal olmayan model için tanımlanan varsayımları dikkate alarak yapılan genel çözüm yaklaşımı

$$\phi(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = \left(\boldsymbol{Y}\text{-}f(\boldsymbol{X},\boldsymbol{\theta})\right)' \left(\boldsymbol{Y}\text{-}f(\boldsymbol{X},\boldsymbol{\theta})\right) \tag{2}$$

biçiminde tanımlı hata kareler toplamını amaç fonksiyonu olarak ele alıp, bu amaç fonksiyonunu minimum yapan parametre vektörünün elde edilmesi olarak tanımlanabilir. Bu yöntem kullanılarak elde edilen parametre değerleri, parametrelerin EKK tahmin değerleri olarak bilinmektedir. Eşitlik (2) ile tanımlı ϕ fonksiyonunun minimizasyonu, kısıtsız bir optimizasyon problemi olarak ele alınabilir. Türeve dayalı optimizasyon algoritmaları ile problemin en iyilenmesi için ϕ fonksiyonunun model parametrelerine göre türevlerinin alınıp, kısmi türev fonksiyonlarının elde edilmesi gerekir. f fonksiyonunun parametrelerine göre doğrusal olmayan bir fonksiyon olduğu göz önünde bulundurulduğunda, kısmi türev fonksiyonları da bilinmeyen parametrelerin fonksiyonları olacağından, analitik çözümlerin elde edilmesinin zorlayıcı olacağı açıktır. Ayrıca, türeve dayalı algoritmalar ile optimizasyonda, parametrelerin başlangıç değerlerinin belirlenmesi problemi de söz konusu olacaktır. Bu gibi nedenlerden dolayı, çalışmada, türevden bağımsız optimizasyon algoritmaları ile hibrit yaklaşım kullanılarak ϕ fonksiyonunu minimum yapacak model parametre tahminlerinin elde edilmesi hedeflenmiştir.

3.Türevden bağımsız optimizasyon algoritmaları ve hibrit algoritma

Doğrusal olmayan regresyon model parametrelerinin nokta tahminlerinin elde edilmesinde en çok izlenen yaklaşım, türeve dayalı yinelemeli algoritmalar kullanılarak modele en küçük kareler (EKK) yönteminin uygulanmasıdır. Fakat, bu algoritmalar kullanılarak yapılan aramalarda, model parametrelerinin başlangıç değerlerinin iyi tanımlanamadığı durumlarda yerel çözüm tuzaklarına takılma, global çözüme yakınsayamama gibi sorunlar ortaya çıkabilmektedir. Türev bilgisinin elde edilemediği, elde edilse bile türev bilgisi kullanımının pratik olmadığı problemlerin çözümünde kullanılan türevden bağımsız optimizasyon algoritmaları, problemin çözümü için belirlenen kurallar çerçevesinde yinelemeli olarak ilerleyen, her yinelemede elde edilen sonuçları kullanarak optimal sonuca ulaşmaya çalışan algoritmalardır [15].

3.1. Nelder-Mead Simpleks algoritması

Nelder-Mead Simpleks (NMS) algoritması, doğrusal olmayan modellerin optimizasyonunda, tek nokta aramalara dayalı, türev bilgisine ihtiyaç duymayan bir stokastik optimizasyon algoritmasıdır. NMS algoritması, Spendley vd. [17] tarafından tanıtılan simpleks esaslı arama algoritmasındaki yansıma ve küçülme operasyonlarına, Nelder-Mead [14] tarafından eklenen genişleme ve büzülme operasyonları ile geliştirilerek, doğrusal olmayan modellerin optimizasyonu için kullanılır biçime getirilmiştir. q sayıda parametreye sahip bir doğrusal olmayan regresyon modelinde, öncelikli olarak seçilen q+1 tane parametre tahmin vektörünün amaç fonksiyon değerlerinin bulunması ile işleme başlanır. Parametre tahminleri, ϕ amaç fonksiyonundaki değerlerine göre sıralanır. Sıralama işleminde, en düşük amaç fonksiyonu değerine sahip parametre tahmin vektörü (*L*), en düşük ikinci amaç fonksiyonu değerine sahip parametre tahmin vektörü (*S*), en yüksek amaç fonksiyonu değerine sahip parametre tahmin vektörü (*H*) ve en yüksek ikinci amaç fonksiyonu değerine sahip parametre tahmin vektörü (*N*) belirlenir. Sıralama işlemi sonrasında NMS algoritmasında sırasıyla yansıma, genişleme, büzülme ve küçülme operasyonları uygulanır. Bu operasyonlara ilişkin NMS ayarlanabilir parametreleri α (>0), β (>1), γ $(0<\gamma<1)$ ve σ $(0<\sigma<1)$ ile tanımlıdır. Yapılan çalışmalarda genellikle bu parametre değerleri, $\alpha=1$, $\beta=2$, $\gamma=1/2$ ve $\sigma=1/2$ olacak biçimde seçilir [8]. Her yineleme sonucunda algoritmada bulunan q+1 parametre tahmin vektörü ϕ fonksiyonu değerine göre yeniden sıralanır ve *L*, *H*, *S*, *N* parametre tahmin vektörleri güncellenir. Böylece, NMS algoritması, bu operasyonlar ile adım adım ilerleyerek belirlenmiş iyi bir başlangıç çözüm değeri ile amaç fonksiyonu değerini en küçük yapacak parametre tahminine yakınsar. Algoritma, amaç fonksiyon değerlerinin birbirine çok yaklaşması veya belirli bir yineleme sayısına ulaşılması durumlarından birinin gerçekleşmesi yoluyla sonlandırılır.

3.2. Genetik Algoritma

Genetik algoritma (GA), evrim sürecini taklit ederek nesiller sonucunda "en iyi" çözümlerin topluluk içinde çoğunluğu oluşturması ve belli bir kuşak sayısı sonunda bu çözümler içinden de en iyisinin seçilmesine dayalı oluşturulmuş popülasyon tabanlı, türevden bağımsız, bir stokastik optimizasyon algoritmasıdır. Evrimin bilgisayar bilimlerine entegre edilmesinde 1950'li yıllardan itibaren pek çok bilim insanının katkısı olmasına rağmen GA kavramı John Holland ve öğrencileri tarafından geliştirilmiştir. Holland [10], GA konulu çalışmalarını "Adaptation in Natural and Artificial Systems" isimli kitabında toparlayıp, yayınlamıştır. Holland'ın öğrencilerinden Goldberg [7] ise yaptığı çalışmaları ile GA'nın tanınmasına büyük katkı sağlamıştır. GA, evrim ilkelerinin bilgisayar ortamına uyarlanması fikrine dayandığından genetik biliminde kullanılan pek çok terim (gen, kromozom, popülasyon, kuşak sayısı) GA için uyarlanmıştır.

GA, başlangıç popülasyonunun oluşturulması, uygunluk değerinin hesaplanması, genetik operatör işlemlerinin (seçim, çaprazlama, mutasyon) uygulanması, yeni kuşak popülasyonunun oluşturulması ve durdurma koşulunun kontrolü gibi adımlardan oluşmaktadır. Burada, genetik operatörler, GA'da evrim sürecinin taklit edildiği bölümdür. Önceki çözümlere göre yeni çözümlerin daha başarılı olması genetik operatörler aracılığı ile sağlanır. Bu nedenle genetik operatörlerin seçimi önemlidir. Literatürde tanımlı mevcut seçim yöntemleri (örneğin, rulet tekerleği, stokastik uniform, remainder, rank, turnuva), çaprazlama yöntemleri (tek nokta, iki nokta, scattered, aritmetik, heuristic) ve mutasyon yöntemleri (uniform, adaptive feasible) arasından probleme uygun kombinasyon seçimi önemlidir. Genetik operatörlerin yanı sıra, popülasyon büyüklüğü ve kuşak sayısının seçimi de model parametrelerinin tahminini etkileyen diğer GA ayarlanabilir parametreleridir.

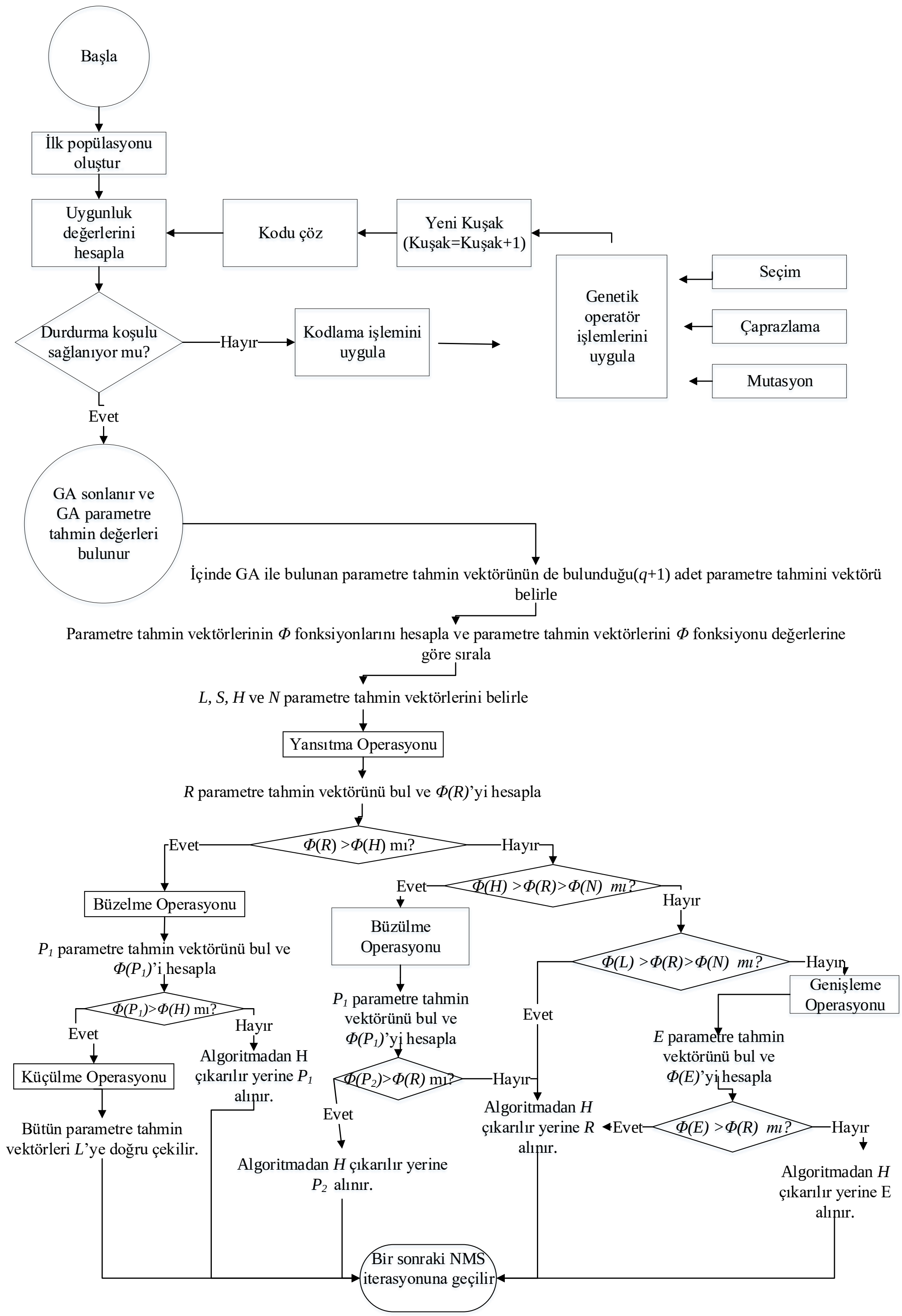
GA ayarlanabilir parametrelerinin kombinasyonları arasından seçim yapmak ciddi emek ve zaman kayıplarına yol açacağından, etkili ve güçlü bir GA araması için ayarlanabilir parametre değerlerini belirleme yöntemine ihtiyaç vardır. Bu amaçla çalışmada, etkin ayarlanabilir parametre kombinasyonuna ulaşabilmek için Taguchi deney tasarımı kullanılmıştır. Taguchi deney dizilimlerinin hazırlanmasında dikey dizilim kullanılmıştır. Dikey dizinlerden üretilen Taguchi tablolarının gösterimi $T_n(s^k)$ biçimindedir. Burada, n, dizindeki deneme sayısı (satır sayısı), s, düzey sayısı (eleman sayısı) ve k etken sayısıdır (sütun sayısıdır). Akgün [2] çalışmasında, oluşturulma biçimlerine göre Taguchi tasarım sınıflandırmalarına ve Taguchi dizininde kullanılan üreteç matrislerinin oluşturulmasına yer verilmiştir.

3.3. Genetik-Simpleks hibrit algoritması

NMS algoritması ve GA'nın avantajlı yönlerinin bir arada kullanılması ile oluşturulan Genetik-Simpleks hibrit algoritmasında, öncelikle GA ile global arama yapılarak, geniş arama uzayı daraltılır. Daha sonra, GA ile bulunan model parametre tahminleri NMS için model parametrelerinin başlangıç değerleri olarak kabul edilerek, en iyi çözümün bulunduğu bölgeye indirgenir. Şekil 1'de, çalışmada kullanılan Genetik-Simpleks hibrit algoritmasının akış diyagramı verilmiştir.

4. Uygulama

Çalışmanın uygulama bölümünde, mikroorganizmaların biyokimyasal oksijen ihtiyacını belirlemek amacıyla, Bates ve Watts [3] tarafından yapılan çalışmadaki veri seti ele alınmıştır. Verilerin elde edilmesinde öncelikli olarak, nehir suyundan örnek alınmıştır. Suya çözünmüş organik madde, inorganik besinler ve çözünmüş oksijen ilave edilmiştir. Su, şişelere bölüştürülmüştür. Her bir şişeye sabit sıcaklıkta ve eşit miktarda mikroorganizma kültürü enjekte edilmiş ve şişelerin ağzı kapatılmıştır. Daha sonra, şişeler günlere bağlı olarak sırasıyla açılmış ve şişelerdeki suyun içinde çözünmüş oksijen miktarları ölçülmüştür. Çizelge 1'de, mikroorganizmaların zamana göre biyokimyasal oksijen ihtiyacına ilişkin veriler yer almaktadır.



Şekil 1. Genetik-Simpleks hibrit algoritması akış diyagramı

Çizelge 1. Mikroorganizmaların zamana göre biyokimyasal oksijen ihtiyacı.

t (gün)	1	2	3	4	5	7
Biyokimyasal oksijen ihtiyacı (mg/l)	8.3	10.3	19	16	15.6	19.8

Çizelge 1'de verilen veri setine uygun olarak kabul edilen doğrusal olmayan regresyon modeli

$$\mathbf{Y} = \theta_1\left(1 - e^{-\theta_2 \mathbf{t}}\right) + \boldsymbol{\varepsilon} \tag{3}$$

biçimindedir. Burada, $\boldsymbol{\theta} = [\theta_1\ \theta_2]$ bilinmeyen model parametre vektörüdür. EKK kriterine göre oluşturulan amaç fonksiyonu

$$\phi(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = \left(\boldsymbol{Y}\text{-}\theta_1\left(1 - e^{-\theta_2 \mathbf{t}}\right)\right)'\left(\boldsymbol{Y}\text{-}\theta_1\left(1 - e^{-\theta_2 \mathbf{t}}\right)\right) = \sum_{i=1}^{6}\left(Y_i - \theta_1(1 - e^{-\theta_2 t_i})\right)^2 \tag{4}$$

ifadesi ile tanımlanır. Buna göre kısıtsız optimizasyon problemi

$$\min_{\boldsymbol{\theta}}\left\{\phi([\theta_1\ \theta_2]) = \sum_{i=1}^{6}\left(Y_i - \theta_1(1 - e^{-\theta_2 t_i})\right)^2\right\} \tag{5}$$

biçiminde olur. Eşitlik (5) ile tanımlı minimizasyon problemini Bates ve Watts [3] çalışmalarında, türeve dayalı bir optimizasyon algoritması ile en iyileyerek, $\hat{\boldsymbol{\theta}} = \left[\hat{\theta}_1\ \hat{\theta}_2\right] = \left[19.1426\ \ 0.5311\right]$ ve $\phi(\hat{\boldsymbol{\theta}}) = 25.9903$ sonuçlarına ulaşmışlardır.

Bu çalışmada, Eşitlik (5) ile tanımlı kısıtsız problemin GA ile optimizasyonu için öncelikli olarak, GA ayarlanabilir parametrelerinin probleme uygun tanımlanması gerekir. Bu amaçla çalışmada, GA ayarlanabilir parametreleri için (popülasyon büyüklüğü, kuşak sayısı, seçim yöntemi, çaprazlama yöntemi, mutasyon yöntemi) beş seviye belirlenmiştir. Çizelge 2'de GA ayarlanabilir parametrelerin uygun seviyeleri tanımlanmıştır.

Çizelge 2. GA ayarlanabilir parametreleri ve seviyeleri

GA Ayarlanabilir Parametreleri	**Seviyeler**				
	1	**2**	**3**	**4**	**5**
Popülasyon Büyüklüğü - A	30	50	100	150	200
Kuşak Sayısı - B	50	75	100	200	500
Seçim Yöntemi - C	Stokastik uniform	Remainder	Uniform	Rulet tekerleği	Turnuva
Çaprazlama Yöntemi - D	Scattered	İki nokta	Tek nokta	Heuristic	Aritmetik
Mutasyon Yöntemi - E	Uniform 0.2	Uniform 0.01	Uniform 0.05	Uniform 0.1	Adaptive feasible

Çizelge 3. Taguchi T_{25} tablosu

Deneme Sayısı	**A**	**B**	**C**	**D**	**E**
1	1	1	1	1	1
2	1	2	2	2	2
3	1	3	3	3	3
4	1	4	4	4	4
5	1	5	5	5	5
6	2	1	2	3	4

7	2	2	3	4	5
8	2	3	4	5	1
9	2	4	5	1	2
10	2	5	1	2	3
11	3	1	3	5	2
12	3	2	4	1	3
13	3	3	5	2	4
14	3	4	1	3	5
15	3	5	2	4	1
16	4	1	4	2	5
17	4	2	5	3	1
18	4	3	1	4	2
19	4	4	2	5	3
20	4	5	3	1	4
21	5	1	5	4	3
22	5	2	1	5	4
23	5	3	2	1	5
24	5	4	3	2	1
25	5	5	4	3	2

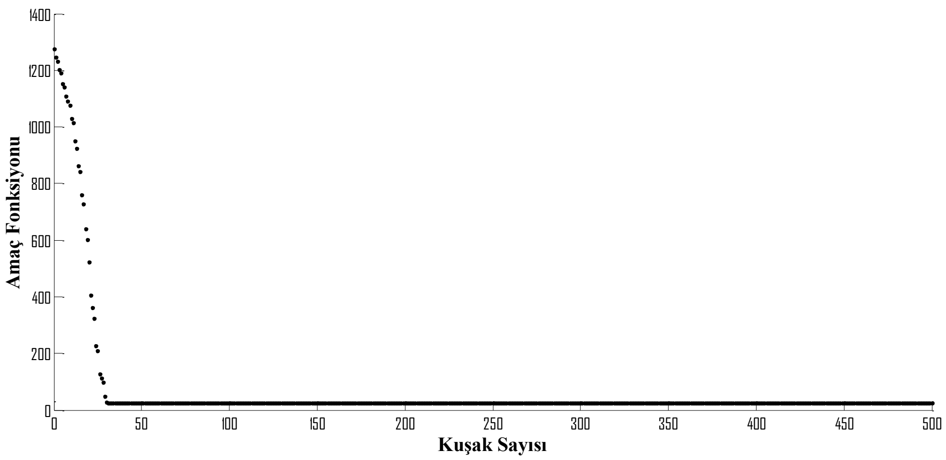
Beş parametreli ve beş seviyeli duruma göre Çizelge 3'te tanımlı Taguchi' nin T_{25} tablosu kullanılarak, 25 farklı denemenin her biri için GA ile 200 defa hesaplama yapılmıştır. Hesaplamalarda MATLAB programı kullanılmıştır. Her bir denemede 200 hesaplama için hata kareler toplamının en küçük, ortalama, medyan ve en büyük değerleri ile standart sapma istatistikleri hesaplanmıştır. Elde edilen sonuçlara göre hata kareler toplamının en küçük olduğu durumdaki parametre tahmini değerleri model parametrelerinin nokta tahmini olarak seçilmiştir. 25 deneme için elde edilen amaç fonksiyonu (hata kare toplamı) sonuçları ve bu sonuçlara ilişkin istatistikler Çizelge 4'te verilmiştir.

Çizelge 4. GA ile 25 deneme için elde edilen amaç fonksiyonu değerleri ve fonksiyon değerlerine ilişkin istatistikler

	ϕ				
Deneme	**En küçük**	**Ortalama**	**Medyan**	**En büyük**	**Std. Sapma**
1	1266.1069	1268.1609	1267.7243	1275.3778	1.6948
2	1266.1809	1271.8305	1270.6387	1293.6239	4.8862
3	1266.1666	1270.3261	1269.2471	1288.8628	3.6711
4	***25.9903***	***26.9687***	***26.0649***	***44.1037***	***2.6238***
5	***26.0154***	***28.4509***	***27.3296***	***45.0495***	***2.9837***
6	1266.0197	1267.5013	1267.1452	1273.4304	1.3149
7	***25.9903***	***26.0451***	***25.9907***	***36.6875***	***0.7563***
8	1266.1611	1267.5561	1267.2924	1272.1778	1.0619
9	1266.2625	1270.1013	1269.7053	1281.3135	2.6714
10	1266.1006	1268.4849	1267.8787	1280.2719	2.2139
11	1266.5785	1273.5719	1272.8794	1286.0550	4.2411
12	1266.0438	1267.3516	1267.1156	1273.5329	1.1553
13	1266.0015	1266.8039	1266.5714	1270.5032	0.7305
14	***25.9903***	***57.6098***	***26.0004***	***359.7778***	***86.2753***
15	***25.9903***	***25.9905***	***25.9903***	***25.9983***	***0.0010***
16	***25.9903***	***63.1432***	***26.0036731***	***425.3242***	***108.9083***

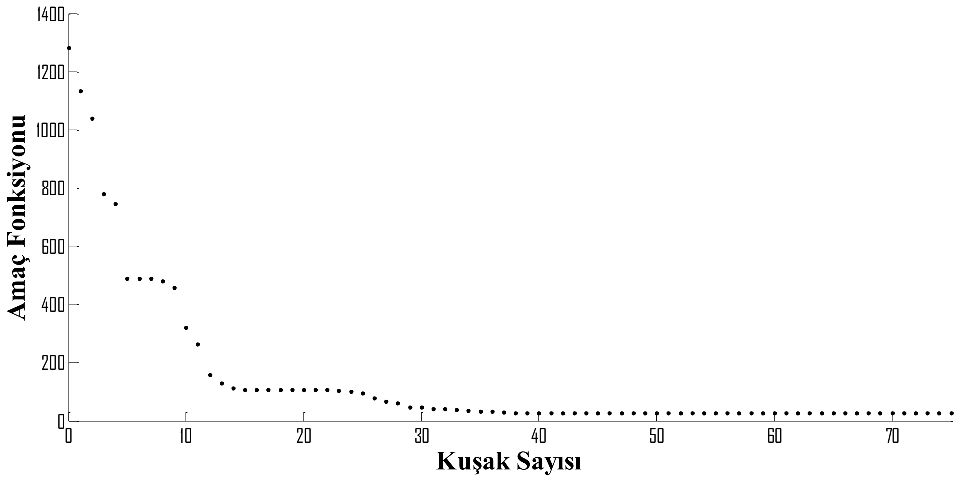
17	1266.0071	1266.3736	1266.2989	1267.7037	0.2942
18	***25.9903***	***27.4609***	***26.1367***	***37.1909***	***2.3987***
19	1266.0686	1267.4308	1267.0441	1272.2281	1.2543
20	1266.0461	1266.8803	1266.6852	1268.9877	0.6845
21	1211.4667	1232.8045	1233.3594	1252.2715	8.36306
22	1266.0349	1266.7299	1266.5748	1269.9987	0.6045
23	***25.9903***	***58.6576***	***25.9915***	***310.2354***	***86.7679***
24	1266.0026	1266.4471	1266.3649	1268.2569	0.3673
25	1266.0279	1267.0354	1266.7937	1270.8184	0.8839

Çizelge 4'teki sonuçlar incelendiğinde, en küçük amaç fonksiyon değerinin yaklaşık olarak 26 olduğu ve bu sonucun 4., 5., 7., 14., 15., 16., 18. ve 23. denemelerde elde edildiği görülmektedir. Bu denemelere ilişkin Çizelge 4'teki istatistiklere göre, 14., 16. ve 23. denemelerde elde edilen 200 farklı hesaplamaya ilişkin standart sapma değerlerinin oldukça büyük olduğu söylenebilir. Bu durum, stokastik aramalara dayalı olan GA'nın ayarlanabilir parametrelerinin model parametrelerinin nokta tahmini üzerindeki etkisini gösterir. GA'da popülasyon büyüklüğü ve kuşak sayısının gereğinden büyük seçilmesi, hesaplama süresinin uzamasına neden olacaktır. GA'nın bir yinelemede kaçıncı kuşaktan itibaren optimum değere yaklaştığı bilgisi, GA'nın optimal sonuca ne kadar çabuk ulaşabildiğinin bir yanıtıdır. En iyi çözüme yakınsama sağlandıktan sonraki kuşaklar, denemelerin büyük çoğunluğunda, fazladan işlem yükü olabilmektedir. Çizelge 4'teki düşük standart sapmalı 4., 5., 7., 15. ve 18. denemeler arasında en küçük standart sapma değerine sahip 15. denemeye ilişkin GA ayarlanabilir parametreleri incelendiğinde, popülasyon büyüklüğünün 100 ve kuşak sayısının 500 olduğu görülür. Deneme 15'te kaçıncı kuşaktan itibaren optimum değere yakınsama sağlandığının belirlenmesi amacıyla 500 kuşakta hesaplanan amaç fonksiyonu değerleri incelenmiştir. Şekil 2'de verilen grafikte, deneme 15 için bir yinelemede 500 kuşak boyunca elde edilen amaç fonksiyon değerleri görülmektedir.



Şekil 2. Deneme-15' te bir yineleme için 500 kuşakta hesaplanan amaç fonksiyonu değerleri

Şekil 2'de görüldüğü üzere kuşak sayısı yaklaşık olarak 30'a ulaştığında amaç fonksiyonu en iyi değerine yakınsamaktadır. Buna göre, 500 kuşak sayısına ulaşıncaya kadar yapılan yinelemelerde, yaklaşık olarak 30. kuşaktan sonra yapılan hesaplamalarda benzer fonksiyon değeri civarında sonuçlara ulaşıldığı ve bu hesaplama yükünün hesaplama süresini uzattığı söylenebilir. Düşük standart sapmalı denemeler arasında en az kuşak sayısına sahip 7. denemeye ilişkin Şekil 3'te verilen amaç fonksiyon değerleri grafiği incelendiğinde, oldukça makul kuşak sayısında problem için en iyi fonksiyon değerine tutarlı olarak ulaşıldığı gözlenmektedir.



Şekil 3. Deneme-7'de bir yineleme için 75 kuşakta hesaplanan amaç fonksiyonu değerleri

Taguchi deney tasarımı ile elde edilen deneme sonuçlarına göre, hesaplama süresi de dikkate alındığında, probleme en uygun GA ayarlanabilir parametre değerlerinin 7. denemeye ilişkin olduğu görülmüştür. Deneme-7'ye ilişkin GA ayarlanabilir parametre değerleri, 50 birimlik popülasyon, 75 kuşak sayısı, Uniform seçim, Heuristic çaprazlama ve Adaptive feasible mutasyon olarak tanımlıdır. Bu değerler kullanılarak, GA ile elde edilen model parametre tahminleri $\hat{\boldsymbol{\theta}} = \left[\hat{\theta}_1 \ \hat{\theta}_2\right] = \left[19.1426 \ \ 0.5311\right]$ dir.

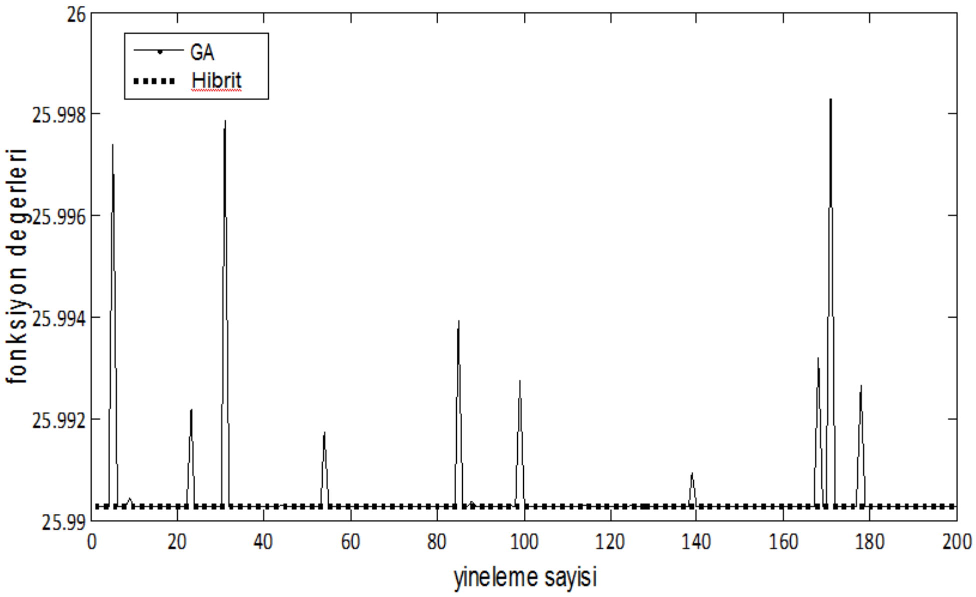
GA ile ulaşılan 25 farklı model parametre tahmin vektörü, NMS algoritması için parametre başlangıç başlangıç değerleri olarak kabul edilip, Genetik-Simpleks hibrit algoritması ile elde edilen model parametrelerinin tahmin değerlerine Çizelge 5'te yer verilmiştir.

Çizelge 5. Model parametrelerinin başlangıç değerleri, hibrit algoritma ile 25 deneme için bulunan amaç fonksiyonu değerlerinin istatistikleri ve model parametre tahminleri

	Parametre başlangıç değerleri		ϕ					Parametre tahmin değerleri	
Deneme	$\hat{\boldsymbol{\theta}}_1^0$	$\hat{\boldsymbol{\theta}}_2^0$	**En küçük**	**Ortalama**	**Medyan**	**En büyük**	**Std. Sapma**	$\hat{\boldsymbol{\theta}}_1$	$\hat{\boldsymbol{\theta}}_2$
1	0.9996	0.9978	25.9903	25.9903	25.9903	25.9903	$3*10^{-9}$	19.1426	0.5311
2	0.9989	0.9988	25.9903	25.9903	25.9903	25.9903	$2*10^{-9}$	19.1426	0.5311
3	0.9994	0.9959	25.9903	25.9903	25.9903	25.9903	$2*10^{-9}$	19.1426	0.5311
4	19.1426	0.5311	25.9903	25.9903	25.9903	25.9903	$4*10^{-9}$	19.1457	0.5311
5	18.9972	0.54370	25.9903	25.9903	25.9903	25.9903	$2*10^{-9}$	19.1426	0.53111
6	0.9999	0.9991	25.9903	25.9903	25.9903	25.9903	$2*10^{-9}$	19.1426	0.5311
7	19.1429	0.5311	25.9903	25.9903	25.9903	25.9903	$2*10^{-9}$	19.1426	0.5311
8	0.9997	0.9938	25.9903	25.9903	25.9903	25.9903	$2*10^{-9}$	19.1426	0.5311
9	0.9995	0.9915	25.9903	25.9903	25.9903	25.9903	$2*10^{-9}$	19.1426	0.5311
10	0.9996	0.9986	25.9903	25.9903	25.9903	25.9903	$2*10^{-9}$	19.1426	0.5311
11	0.9965	0.9988	25.9903	25.9903	25.9903	25.9903	$2*10^{-9}$	19.1426	0.5311

12	0.9998	0.9989	25.9903	25.9903	25.9903	25.9903	$3*10^{-9}$	19.1426	0.5311
13	0.9999	0.9999	25.9903	25.9903	25.9903	25.9903	$2*10^{-9}$	19.1426	0.5311
14	19.1443	0.5311	25.9903	39.25663	25.9903	136.7617	36.02	19.1426	0.5311
15	19.1426	0.5311	25.9903	25.9903	25.9903	25.9903	$2*10^{-9}$	19.1426	0.5311
16	19.1414	0.5311	25.9903	37.60793	25.9903	136.9752	34	19.1426	0.5311
17	0.9999	0.9999	25.9903	25.9903	25.9903	25.9903	$2*10^{-9}$	19.1426	0.5311
18	19.1426	0.5311	25.9903	25.9903	25.9903	25.9903	$2*10^{-9}$	19.1426	0.5311
19	0.9996	0.9999	25.9903	25.9903	25.9903	25.9903	$2*10^{-9}$	19.1426	0.5311
20	0.9999	0.9981	25.9903	25.9903	25.9903	25.9903	$2*10^{-9}$	19.1426	0.5311
21	1.3298	1.1666	25.9903	25.9903	25.9903	25.9903	$2*10^{-9}$	19.1426	0.5311
22	0.9999	0.9982	25.9903	25.9903	25.9903	25.9903	$2*10^{-9}$	19.1426	0.5311
23	19.1428	0.5311	25.9903	39.80883	25.9903	136.8726	36.65	19.1426	0.5311
24	0.9999	0.9999	25.9903	25.9903	25.9903	25.9903	$4*10^{-9}$	19.1426	0.5311
25	0.9999	0.9987	25.9903	25.9903	25.9903	25.9903	$3*10^{-9}$	19.1426	0.5311

Çizelge 5'ten görüldüğü gibi, Genetik-Simpleks hibrit algoritması ile her denemede elde edilen parametre tahmin değerlerinin tamamı ve hemen hemen bütün en büyük fonksiyon değerleri, GA ile 7. denemede elde edilen en iyi parametre tahmin değerlerine ($\hat{\boldsymbol{\theta}} = \left[\hat{\theta}_1\ \hat{\theta}_2\right] \cong [19.1426\ \ 0.5311]$) ve fonksiyon değerine ($\phi^* \cong 26$) sırasıyla eşittir. Buna göre, hibrit algoritmanın her deneme kombinasyonunda GA'ya göre daha tutarlı sonuçlar verdiği söylenebilir. Şekil 4'te verilen, deneme 7 için 200 yinelemede elde edilen GA ve Genetik-Simpleks hibrit algoritması ile elde edilen ϕ amaç fonksiyonu değerleri karşılaştırılması, hibrit algoritma ile GA'dan daha tutarlı sonuçlar elde edildiği desteklemektedir.



Şekil 4. Deneme 7 için uygulanan 200 hesaplamaya göre GA ve Genetik-Simpleks hibrit algoritması ile elde edilen hata kareler toplamlarının karşılaştırılması

5. Sonuç

Bu çalışmada, türevden bağımsız optimizasyon algoritmaları olan GA ve NMS algoritmasının avantajlı yönlerinin birleştirilmesi ile oluşturulan Genetik-Simpleks hibrit algoritması kullanılarak, doğrusal olmayan regresyon model parametreleri için EKK tahminlerinin elde edilmesi hedeflenmiştir. Ayrıca çalışmada, GA ayarlanabilir parametrelerinin belirlenmesinde Taguchi deney tasarımından yararlanılarak, GA ile model parametrelerinin EKK tahminlerinin elde edilmesi istenmiştir. Bu amaçla, GA ve Genetik-Simpleks hibrit algoritması, literatürde tanımlı bir negatif-üstel regresyon model için optimizasyon aracı olarak kullanılmıştır. Elde edilen sonuçlara göre, GA ayarlanabilir parametre değerlerinin seçiminin GA ile yapılan optimizasyon sonuçlarını etkilediği görülmüştür. Genetik-Simpleks hibrit algoritmasının, negatif-üstel biçimde tanımlı doğrusal olmayan model parametrelerinin nokta tahminlerinin elde edilmesinde model parametrelerinin başlangıç değerlerine ihtiyaç duymadan, GA ayarlanabilir parametrelerinin probleme uygun belirlenemediği durumlarda bile tutarlı tahmin sonuçları veren bir optimizasyon algoritması olarak kullanılabileceği sonucuna ulaşılmıştır.

Kaynaklar

[1] B. Altunkaynak, A. Esin, 2004, Doğrusal Olmayan Regresyonda Parametre Tahmini İçin Genetik Algoritma Yöntemi, Gazi Üniversitesi Fen Bilimleri Dergisi, 17(2), 43-51.

[2] F. Akgün, 2018, Doğrusal Olmayan Regresyon Model Parametrelerinin Nokta ve Aralık Tahmini İçin Bir Yaklaşım, Ankara Üniversitesi, Fen Bilimleri Enstitüsü Yüksek Lisans Tezi.

[3] D. M. Bates, D. G. Watts, 1988, Nonlinear Regression Analysis and Its Applications, John Willey & Sons, Inc., New York.

[4] M. J. Box, 1971, Bias in Nonlinear Estimation, Journal of Royal Statisical Society: Series B, 33(2), 171-201.

[5] A. R. Gallant, W. A. Fuller, 1973, Fitting Segmented Polynomial Regression Models Whose Join Points Have to Be Estimated, Journal of the American Statistical Association, 68(1), 144-147.

[6] A. R. Gallant, 1975, Nonlinear Regression, Journal of the American Statistical Association, 29(2), 73-81.

[7] D. E. Goldberg, 1989, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison - Wesley, Boston.

[8] A. P. Gurson, 2000, Simplex Search Behaviour in Nonlinear Optimization, Underground Honors Thesis, College of William & Mary.

[9] H. O. Hartley, A. Booker, 1963, Non-Linear Least Squares Estimation, Unpublished Report, Iowa State University, Ames.

[10] J. H. Holland, 1975, Adaptation in Natural andArtificial Systems, The University of Michigan Press, Ann Arbor.

[11] R. I. Jennrich, 1969, Asymptotic Properties of Non-linear Least Squares Estimators, The Annals of Mathematical Statistics, 40(2), 633-643.

[12] E. Malinvaud, 1970, Consistency of Nonlinear Regressions, The Annals of Mathematical Statistics, 41(3), 956-969.

[13] D. W. Marquardt, 1963, An Algorithm for Least-Squares Estimation of Nonlinear Parameters, Journal of the Society of Industrial and Applied Mathematics, 11(2), 431-441.

[14] J. A. Nelder, R. Mead, 1965, A Simplex Method for Function Minimization, The Computer Journal, 7(4), 308-313.

[15] L. M. Rios, N. V. Sahinidis, 2013, Derivative-Free Optimization: a Review of Algorithms and Comparison of Software Implementations, Journal of Global Optimization, 56(3), 1247-1293.

[16] G. A. F. Seber, C. J. Wild, 1989, Nonlinear Regression, John Willey&Sons, Inc., New York.

[17] W. Spendley, G. R. Hext, F. R. Himsworth, 1962, Sequential Application of Simplex Designs in Optimization and Evolutionary Operation, Technometrics, 4(4), 441-461.

[18] Ö. Türkşen, 2014, Estimation of Fault Plane Parameters by Using Stochastic Optimization Methods, International Journal of Earthquake Engineering and Hazard Mitigation, 2(2), 61-66.

[19] Ö. Türkşen, M. Tez, 2016, An Application of Nelder-Mead Heuristic-based Hybrid Algorithms: Estimation of Compartment Model Parameters, International Journal of Artificial Intelligence, 14(1), 112-129.