

Orman ağacı ıslah çalışmalarında doğrusal karma model seçimi

Linear mixed model selection in forest tree breeding studies

Mehmet ACET¹, Zafer ÖLMEZ²

¹Orman Genel Müdürlüğü, Orman Ağaçları ve Tohumları Islah Araştırma Enstitüsü Müdürlüğü, Yenimahalle, Ankara

²Artvin Çoruh Üniversitesi, Orman Fakültesi, Orman Mühendisliği Bölümü, Merkez, Artvin

Eser Bilgisi / Article Info

Araştırma makalesi / Research article

DOI: 10.17474/artvinofd.1260542

Sorumlu yazar / Corresponding author

Mehmet ACET

e-mail: mehmetacet@ogm.gov.tr

Geliş tarihi / Received

05.03.2023

Düzeltilme tarihi / Received in revised form

05.05.2023

Kabul Tarihi / Accepted

08.05.2023

Elektronik erişim / Online available

15.05.2023

Anahtar kelimeler:

Model Seçimi

Doğrusal Karma Modeller

Mekansal Analiz

Akaike Bilgi Kriteri

Döl Denemeleri

Kızılçam

Keywords:

Model Selection

Linear Mixed Models

Spatial Analysis

Akaike Information Criteria

Progeny Trials

Turkish Red Pine

Özet

Orman ağacı ıslah çalışmalarında, uzun süre gözlemlenen genetik testler ile ıslah programları için genetik parametreler tahmin edilmektedir. Söz konusu parametrelerin tahminleri ıslah programını etkileyeceğinden tahmin için kullanılacak doğrusal karma modelin seçimi büyük önem taşımaktadır. Kullanılan doğrusal karma modellerde tahmin, genellikle artık (residual) veya kısıtlı maksimum olabilirlik (REML) yöntemi kullanılarak elde edilir. Farklı sabit etkileri olan modellerin olabilirliğe (likelihood) dayalı bilgi kriterleri ile kıyaslanabilmesi için, modellerin maksimum olabilirlik (maximum likelihood) kullanılarak tahmin edilmesi önerilmektedir. Orman ağaçları ıslah çalışmalarında doğrusal karma model seçiminde farklı modeller denenerek model uyumunu arttıran en kullanışlı model seçilmelidir. Orman ağaçları ıslah çalışmalarında modellerin uyumunu kıyaslamak için ise genellikle Akaike (AIC) bilgi kriterinin kullanılması önerilmektedir. Bu çalışmada, doğrusal karma model seçiminin gerekliliğini ve önemini ortaya koymak amaçlanmıştır. Bu amaç için, Muğla-Marmaris'te açık tozlaşma ürünü 168 aile (üvey kardeş) ile tesis edilmiş olan Kızılçam döl deneme sahasındaki ağaçların on ikinci yaş göğüs yüksekliği çap verileri kullanılarak modeller kıyaslanmıştır. Verilerin analizinde geleneksel (basit), mekânsal bileşen içeren, artığın bağımsız veya birinci dereceden iki boyutlu ayrılabilir otoregresif korelasyon hata yapısı olduğunu varsayan toplamda otuziki farklı model denenmiştir. Geleneksel modelin AIC değeri (Model-1=5594.1), mekânsal bileşen ve artığın otoregresif korelasyon yapısı içeren modellere kıyasla (Model-20=5447) daha yüksek bulunmuştur.

Abstract

In forest tree breeding studies, genetic parameters are estimated for breeding programs with long-term genetic tests. Since the estimations of these parameters will affect the breeding program, the selection of the linear mixed model to be used for the estimation is of great importance. In linear mixed models used for the estimation of these parameters, the estimation is usually obtained by using the residual or restricted maximum likelihood (REML) methods. In order to compare the models with different fixed effects with the information criteria based on likelihood, it is suggested that the models should be estimated using maximum likelihood. In forest tree breeding studies, different models should be tried in linear mixed model selection and the most useful model that increases model fit should be selected. It is generally recommended to use the Akaike (AIC) information criterion to compare the fit of models in forest tree breeding studies. In this study, it is aimed to reveal the necessity and importance of linear mixed model selection. For this purpose, the models were compared by using the twelfth year diameter at breast height data of the trees in the red pine progeny trial field established with 168 open pollinated families (half-sib) in Muğla-Marmaris. In the analysis of this data set, a total of 32 different models were tested, which include traditional (simple), spatial component, and assume that the residual is an independent or 1st order 2-dimensional separable autoregressive correlation error structure. The AIC value of the traditional model (Model-1=5594.1) was found to be higher than the models with the spatial component and the residual autoregressive correlation structure (Model-20=5447).

GİRİŞ

Orman ağaçları ıslah çalışmaları, diğer araştırmalardan farklı olarak, kullanılacak materyalin belli kurallara bağlı kalarak uygun bir deneme deseni ile araziye applike edilmesi ve elde edilen verilerin de bu desene ve uygun modele göre analiz edilmesini gerektirmektedir. Analizler

sonucu, tahmin edilen parametrelerin isabet derecesini etkileyen çeşitli çevre faktörleri bulunmaktadır. Bu faktörler, deneme sahasları arasında hatta aynı deneme sahasında bile farklılık/heterojenlik gösterebilmektedir. Çevresel heterojenlikten kaynaklı hatayı en aza indirmek için araştırma çalışmalarında deneme desenleri kullanılmaktadır.

Ormancılık ve tarımsal araştırmalardaki deneylerin en önemli unsurlarının içerisinde, deneme deseni ve analiz modelinin doğru seçimi de yer almaktadır (Casler 2015, Piepho ve ark. 2015). İyi tasarlanmış denemeler, Fisher (1936) tarafından önerilen rastgeleleştirme, tekrarlar ve yerel kontrol ilkelerine dayanmaktadır. Fisher'den sonra, rastgeleleştirme, tekrarlar ve bloklama ilkeleri değerlendirilmiştir (Yates 1939). Rastgeleleştirme, tüm deney birimlerinin herhangi bir işlem olasılığının eşit olmasını sağlayarak, deneyden kaynaklanan sistematik hataları en aza indirebileceği belirtilmiştir (Welham ve ark. 2014, Yates 1939). Tekrarlamaların, deneysel hata varyansının tahmin edilmesini sağlayacağı ve ne kadar çok tekrar olursa, elde edilen sonucun kesinliğinin (güvenirliliğinin) artacağı belirtilmiştir. Bunun yanında optimum tekrar sayısından fazla tekrar kullanılması, deneme alanının büyümesine sebep olacağından, çevresel varyansın artmasına, bunun sonucunda da kalıtım derecesi ve genetik kazanç tahminlerinde sapmalara sebep olabilmektedir (Matheson ve ark. 2008, Urhan ve ark. 2014). Bloklamanın ise, doğru şekilde uygulandığında, mekânsal varyasyonu kontrol edebileceği ve denemeden kaynaklı hataları azaltmaya yardımcı olabileceği belirtilmiştir (Welham ve ark. 2014, Yates 1939).

Ormancılıkta rastlantı blokları deneme deseni yaygın olarak kullanılmaktadır (RCBD) (Piepho ve ark. 2015). Açık arazi koşullarındaki bu denemelerde, ağaçların büyümesine etki eden çevresel faktörler birçok değişkenlik göstermekte ve genellikle mekânsal otokorelasyon içermektedir (Grondona ve ark. 1996, Legendre 1993). Bu durumun üstesinden gelebilmek için, mekânsal analizler geliştirilmiş ve rastlantı blokları deneme desenlerindeki verileri analiz etmede de kullanılmıştır (Chen ve ark. 2017, Isik ve ark. 2017).

Doğrusal karma modeller, varyans bileşenlerinin doğru tahminini sağlama potansiyelleri, ilişkili hataları modellemede esnek olduklarından işlem etkilerini yansız tahmin edebilmesi ve heterojen varyans yapılarının da modele dahil edilebilmesi gibi nedenlerle bitki ve hayvan ıslahı çalışmalarında yaygın olarak kullanılmaktadır (Brown ve Prescott 2015, Little ve ark. 2006, Mathew ve ark. 2016, Stroup 2012). Doğrusal karma modeller hem

sabit etkili hem de rastgele etkili değişkenleri içermektedir. Sabit etki, tekrarlanan örneklemede aynı kalan ve bilinen bir sabitken, rastgele etki ise, alt örneklemeden ve işlem düzeylerinin rastgele seçiminden kaynaklanan rastgele bir değişkendir (Isik ve ark. 2017). Diğer bir deyişle, seviyeleri rastgele olmayan bir süreçle seçilmişse veya çalışmaya dahil edilen seviyeler, olası seviyelerin tüm popülasyonunu kapsıyorsa, etkilerin sabit olduğu varsayılır. Rastgele etkiler ise, popülasyonda rastgele bir seviye (aile) örneklemeden oluşan faktörlerdir ve tüm popülasyon (aileler) için çıkarım yapılır (Mramba 2016).

Doğrusal karma modellerde, varyans bileşenlerini tahmin etmek için, genellikle kısıtlı maksimum olasılık (REML) yöntemi kullanılmaktadır. Bu yöntem, tahmin edilen hatanın varyansını en aza indirerek, gerçek ve tahmin edilen ıslah değerleri arasındaki korelasyonları en üst düzeye çıkarmaktadır. Raslantısal etkiyi tahmin etmede standart bir yöntem olan en iyi doğrusal yansız tahminlerini (BLUP'lar) elde etmek için, istatistiksel modellerin tüm olası varyasyon kaynaklarını içermesi önemlidir (Mramba 2016, Piepho ve ark. 2008).

Analizlerde, modellerin açıklayamadığı kontrol edilemeyen artıkların (hataların) rastlantısal ve aynı dağılıma sahip olduğu varsayılmaktadır. Diğer bir deyişle homojen yapıya sahip denemelerde modelin açıklayamadığı artık varyansın rastlantısal ve aynı dağılıma sahip olduğu kabul edilmektedir. Fakat bazen deneme sahasındaki heterojenlikten kaynaklı rastlantısal olmayan artıklar ortaya çıkabilmektedir. Bu durum ise deneme içindeki heterojenlik oluşturan mekânsal varyasyondan kaynaklanabilmektedir. Söz konusu bu durum deneme sahaslarında ölçümü yapılan örneklerin (ağaçlar vb.) gelişimlerinin aralarındaki mesafe ile ilişkili olabileceği belirtilmiştir (Gilmour ve ark. 2009, Stroup 2012).

Deney birimleri arasındaki fiziksel yakınlık ve dolayısıyla deneyin kesinliğini etkileyen mikro çevre değişkenliğinin paylaşılması nedeniyle mekânsal korelasyon ortaya çıkabilmektedir. Mekânsal olarak ilişkili hatalar genellikle 1. dereceden 2 boyutlu ayrılabilir otoregresif mekânsal hata yapısı kullanılarak modellenmektedir (Cullis ve ark.

2006, Gezan ve ark. 2010, Gilmour ve ark. 2009). Bunun yanında, diğer kullanılabilecek mekânsal hata yapıları da mevcuttur (Little ve ark. 2006).

Bitki ıslahı çalışmalarında yaygın olarak bulunan ve deneysel birimler arasındaki hesaba katılması gereken başka bir ilişki ise, analizler ve sonuçlar üzerinde arka plan gürültüsü (noise) yapabilen genetik akrabalıktır (Little ve ark. 2006). Bu nedenle, optimal bir deney oluşturmak, yukarıda bahsedilen varyasyon kaynaklarının tümünün veya çoğunun uygun bir şekilde kullanılmasını gerektirmektedir (Mramba 2016). Her bir deneme sahasında elde edilen verilerin analizindeki en verimli yöntem, tüm sabit ve rastgele etkilerin tek bir aşamada (tek aşamalı analiz) karma bir model temelinde tahmin edilmesidir (Caliński ve ark. 2005, Smith ve ark. 2001).

Ormancılıkta genetik testler olarak da bilinen döl denemeleri geniş (>1,5 ha.) alanlarda kurulduğu için mekândan kaynaklı varyasyon oluşabilmektedir. Analizlerde bu durumun üstesinden gelmek için, mekânsal analiz yöntemleri geliştirilmiştir. Ormancılıkta mekândan kaynaklı varyasyonu açıklayabilmek için, geleneksel doğrusal karma modellerin sabit ve rastlantısal etkileri kısmına mekânsal değişkenler (line, spline) ve hata kısmına da otoregresif korelasyon yapısı eklenerek yapılan analizlere mekânsal analizler denilmektedir (Gilmour ve ark. 2009, White ve ark. 2007). Mekânsal modellemenin amacı, ölçüm yapılan bireyler arasındaki bağımsızlığın mesafe arttıkça veya azaldıkça nasıl değiştiğini açıklayarak parametre tahminini güçlendirmek olduğu belirtilmiştir (VSNI 2021).

Döl denemelerinde, genel kombinasyon yeteneği (GKY) için varyans bileşenlerini ve BLUP'ları tahmin etmek için önerilen doğrusal karma modeller kullanılmaktadır (Isik ve ark. 2017, Piepho ve ark. 2008). Rastlantı blokları deneme deseniyle kurulmuş denemelerde, mekansal analiz ile blok varyasyonları içindeki veya arasındaki mikrosite varyasyonlar tespit edilerek modeli iyileştirdiği belirtilmiştir (Dutkowski ve ark. 2006, White ve ark. 2007, Ye ve Jayawickrama 2008).

Orman ağacı ıslahı çalışmalarında modellerin kıyaslanmasında model uyumunu ödüllendiren ayrıca parametre sayısı arttıkça cezalandıran Akaike Bilgi Kriteri

(AIC) değeri ve Bayesian Bilgi Kriterleri (BIC) yaygın olarak kullanılmaktadır (Bozdoğan 1987, Galwey 2014, Isik ve ark. 2017). Guerin ve Stroup (2000) tarafından tekrarlanan ölçümler için kovaryans modeli seçiminde AIC ve BIC'nin performansı karşılaştırılmıştır. Çalışmada, AIC'nin daha karmaşık bir model seçme eğiliminde olduğu, ancak BIC'den daha iyi tip-I hata kontrolüne sahip olduğu belirtilmiştir. AIC değerinin düşük olması, model uyumluluğunun (goodness of the fit) bir göstergesidir ve model seçiminde yaygın olarak kullanılmaktadır (Bozdoğan 1987, Galwey 2014, Isik ve ark. 2017, Verbyla 2019).

AIC veya BIC değerlerine göre modellerin kıyaslanabilmesi için sabit etkilerin aynı olması gerektiğini belirten çalışmalar literatürde bulunmaktadır (Isik ve ark. 2017, Welham ve ark. 2010). Doğrusal karma modellerde tahminler genellikle artık (residual) olabilirlik veya kısıtlı maksimum olabilirlik (REML) kullanılarak gerçekleştirilir. Farklı sabit etkileri olan modeller için, artık olabilirlik değerleri kullanılarak karşılaştırma yapılamadığı için AIC veya BIC gibi bilgi kriterlerinin kullanılamayacağı belirtilmiştir. Bunun yanında literatürde farklı sabit etkileri olan modellerin bilgi kriterleri kullanılarak kıyaslanabilmesi için modellerin maksimum olabilirliklerinin kullanılması önerilmektedir (Verbyla 2019). Bu durumda sabit etkileri aynı olmayan modellerde de kıyaslama yapılabilmektedir (Guerin ve Stroup 2000, Verbyla 2019).

Bu çalışmada, orman ağacı ıslahı çalışmalarında genetik parametre tahmininde kullanılan doğrusal karma model seçimi ele alınmaktadır. Bunun için farklı sabit ve rastlantısal etkilere sahip 32 farklı model denenmiş ve AIC değerlerine göre kıyaslanmıştır.

MATERYAL VE YÖNTEM

Materyal

Orman Ağaçları ve Tohumları Islah Araştırma Enstitüsü Müdürlüğü tarafından Türkiye Milli Ağaç Islahı ve Tohum Üretim Programı (1994-2003) kapsamında Ege Bölgesi Kızılçam Alt Islah Zonunda tesis edilen 3A numaralı deneme sahasındaki 168 açık tozlaşma ürünü ailenin on ikinci arazi yaşı çap ölçümleri bu çalışmanın materyalini

oluşturmaktadır. Deneme alanı, 2000 yılında 1+0 yaşında fidanlar ile dört ağaçlı rastlantı blokları deneme deseni kullanılarak, Muğla Orman Bölge Müdürlüğü, Marmaris Orman İşletme Müdürlüğü Karaca Orman İşletme Şefliğinde kurulmuştur. Denemede yer alan aile sayısının fazla olması nedeni ile, çevresel varyansın etkilerini azaltan alt bloklama yapılmıştır (Schutz ve Cockerham 1966). Buna göre, her dört blok dört adet alt bloğa (set) ayrılmıştır. Ayrıca, deneme alanında ilk kuruluştan sonra fidanlarda %39 kuruma olduğu için, aynı *plus* ağaçlardan üretilen fidanlarla tamamlama yapıldığı, tamamlama etkisi dördüncü yaş-boy özelliği için istatistiksel olarak anlamlı farklılığa sahip olduğu için tamamlama etkisine göre veri düzeltilerek analizlerin yapıldığı ifade edilmiştir (Alan ve ark. 2005). Yaptığımız ön analizlere göre, bu etkinin zamanla azaldığı ve sekizinci yaştan itibaren istatistiksel olarak anlamlı düzeyde olmadığı görüldüğü için analizlere tamamlama etkisi eklenmemiştir.

Yöntem

Model seçiminde kullanılan AIC değeri hesaplanmasında Eşitlik 1 kullanılmıştır (Bozdoğan 1987, Gilmour ve ark. 2015).

$$AIC = -2(\log\text{-likelihood}) + 2k \quad (1)$$

AIC = Akaike Bilgi Kriteri değerini,
 k = model parametrelerinin sayısını,
 $\log\text{-likelihood}$ = log-olabilirlik değerini ifade etmektedir.

Uygun model seçilmesi için MrBean 2.0.6 programı ve AsReml R-4.0 paketleri kullanılmıştır (Aparicio ve ark. 2019, Gilmour ve ark. 2015). Analizlerde, 32 farklı doğrusal karma model (Çizelge 1) denenerek bu modellerden en düşük AIC değerine sahip olan model en uygun model olarak seçilmiştir (Bozdoğan 1987, Isik ve ark. 2017, Verbyla 2019). Kullanılan temel doğrusal karma model Eşitlik 2'deki gibidir.

$$y = Xb + Zu + e \quad (2)$$

Eşitlikte;

y : bireysel ağaç gözlemlerinin vektörünü,

b : sabit (fixed) etkilerin vektörünü (kesme noktası),

X : tasarım matrisini,

u : rastgele etkilerin vektörünü (deneme tasarım parametreleri ve aile etkileri),

Z : tasarım matrisini,

e : bağımsız artıkların (deneysel hata) vektörünü temsil eder.

Yukarıdaki modelde çevresel etkiler yalnızca deneysel tasarım özellikleri ve bağımsız bir hata ile modellenmiştir (Dutkowski ve ark. 2006). Ayrıca artıkların (hata) rastgele olmayan bir mekânsal (spatial) dağılım göstermesi durumunda, iki boyutlu (satır ve sütun) ayrılabilir otoregresif yapı (AR1) içeren modeller kurularak satır ve sütundan kaynaklı artık ortadan kaldırmaya çalışılmıştır (Dutkowski ve ark. 2006, Dutkowski ve ark. 2002). Modele satır ve sütun faktörleri sabit veya rastgele etkiler olarak eklenebilmektedir ve ayrıca artıklar arasındaki korelasyonlar modellenenilmektedir (Isik ve ark. 2017, White ve ark. 2007). Artık bileşenleri, otoregresif yapısı Eşitlik 3'deki gibi kurulmuştur (Dutkowski ve ark. 2002):

$$y = Xb + Zu + \xi + \eta \quad (3)$$

Eşitlikte;

y : bireysel ağaç gözlemlerinin vektörünü,

b : sabit (fixed) etkilerin vektörünü (kesme noktası),

X : tasarım matrisini,

u : rastgele etkilerin vektörünü (deneme tasarım parametreleri ve aile etkileri),

Z : tasarım matrisini,

ξ : mekânsal bağımlı artığı,

η : bağımsız artıkların (deneysel hata) vektörünü temsil eder.

Mekânsal bağımlı artığın formülasyonu Eşitlik 4'deki gibidir (Dutkowski ve ark. 2002).

$$R = \sigma_{\xi}^2 [AR1(\rho_{col}) \otimes AR1(\rho_{row})] + \sigma_{\eta}^2 I \quad (4)$$

σ_{ξ}^2 ve σ_{η}^2 : sırasıyla mekânsal bağımlı ve bağımsız artık varyanslarını,

⊗ : iki matrisin matris kronecker çarpımını,

$R1(\rho_{row})$: satır yönünde birinci derece otoregresif korelasyon matrisini,

I : ise birim matrisini temsil eder.

Sabit bir etki olarak kullanılan lin (sütun) ile birlikte rastgele bir etki olarak kullanılan spl (sütun), kübik düzleştirme eğrisini (cubic smooting spline) sütuna uyarlamaktadır (Burgueño ve ark. 2000).

Tüm modeller birlikte ve model grupları içinde karşılaştırılmıştır. Modellerin karşılaştırılması yapılırken sabit etkileri aynı olan modeller aynı grupta olmak üzere toplam 4 grup altında toplanmıştır. Kullanılan tam model (full model) Eşitlik 5'te ve kullanılan tüm modellerin bileşenleri Çizelge 1'de verilmiştir. Bileşenlerden a ile b sabit etkili ve diğer bileşenler raslantısal etkili olarak kullanılmıştır.

$AR1(\rho_{col})$: sütun yönünde birinci derece otoregresif korelasyon matrisini,

$$y = a + b + c + d + e(d) + f(d) + g(d) + h + i + j \quad (5)$$

a = lin (satır), satırın doğrusal (tek serbestlik derecesi) etkisini,

b = lin (sütun), sütunun doğrusal (tek serbestlik derecesi) etkisini,

c = aile etkisini,

d = blok etkisini,

$e(d)$ = alt bloklama (set) etkisini, (yuvalanmış, nested)

$f(d)$ = satırın blok içindeki etkisini, (yuvalanmış, nested)

$g(d)$ = sütunun blok içindeki etkisini, (yuvalanmış, nested)

h = spline (satır), ortak değişken olarak satırın kübik spline'in rastgele etkisini,

i = spline (sütun): ortak değişken olarak sütunun kübik spline'in rastgele etkisini,

j = artığı (hatanın, residual) temsil etmektedir.

Çizelge 1. Model seçimi için denenen modeller ve bileşenler

Model	Sabit Etkiler					Rastlantısal Etkiler				Artık (j)
	lin (satır) (a)	lin (sütun) (b)	Aile (c)	Blok (d)	Set(Blok) ($e(d)$)	Satır(Blok) ($f(d)$)	Sütun(Blok) ($g(d)$)	spline (satır) (h)	spline (sütun) (i)	
1/17			X	X	X					Indep/ar1
2/18			X	X	X	X				Indep/ar1
3/19			X	X	X	X	X			Indep/ar1
4/20	X		X	X	X	X		X		Indep/ar1
5/21		X	X	X	X	X			X	Indep/ar1
6/22	X		X	X	X	X	X	X		Indep/ar1
7/23	X	X	X	X	X	X		X	X	Indep/ar1
8/24		X	X	X	X	X	X		X	Indep/ar1
9/25	X	X	X	X	X	X	X	X	X	Indep/ar1
10/26			X	X	X		X			Indep/ar1
11/27	X		X	X	X			X		Indep/ar1
12/28		X	X	X	X				X	Indep/ar1
13/29	X		X	X	X		X	X		Indep/ar1
14/30	X	X	X	X	X			X	X	Indep/ar1
15/31		X	X	X	X		X		X	Indep/ar1
16/32	X	X	X	X	X		X	X	X	Indep/ar1

Lin (satır): Satırın doğrusal (tek serbestlik derecesi) etkisini, lin (sütun): sütunun doğrusal (tek serbestlik derecesi) etkisini, Aile: aile etkisini, Blok: blok etkisini, Set: alt bloklama (set) etkisini, () yuvalanmış, nested, spline(satır): ortak değişken olarak satırın kübik spline'in rastgele etkisini, spline (sütun): ortak değişken olarak sütunun kübik spline'in rastgele etkisini, indep: artığın (hatanın, residual) bağımsız olduğunu (model 1-16), ar1: birinci derece otoregresif korelasyonu (model 17-32) ifade etmektedir.

BULGULAR

Farklı bileşenlerden oluşan 32 farklı model ve model gruplarına ait elde edilen bilgiler Çizelge 2’de verilmiştir. Artığın (hatanın) birinci derece otoregresif korelasyonunu içeren modeller, bağımsız olduğunu varsayan modellere nispeten daha düşük AIC değerlerine sahip olduğu bulunmuştur (Çizelge 2). AIC değeri hatanın (artığı) bağımsız olduğunu varsayan Model-1 de 5594.1’den, aynı bileşenlere sahip ama hatanın otoregresif korelasyon içerdiğini varsayan Model17 de 5472.04’e düşmüştür. Sabit etki içermeyen birinci gruptaki, geleneksel (basit) modelin (Model-1) AIC değeri 5594.1’den mekânsal bileşen ve otoregresif korelasyon içeren Model-18’de 5448.7’ye düşmüştür. Modeller içinde en düşük AIC değerine (5447) sahip olan Eşitlik 6’daki Model-20 bulunmuştur. AIC değeri yüksek olan modeller ya satır etkisini içermeyen (Model-1) ya da satır etkisini içermemekle beraber sütun etkisi içeren (Model-10, Model-12 ve Model-15) modellerdir (Çizelge 1-2).

$$y = a + c + d + e(d) + f(d) + h + j \quad (6)$$

a = lin (satır), satırın doğrusal (tek serbestlik derecesi) etkisini,

c = aile etkisini,

d = blok etkisini,

$e(d)$ = alt bloklama (set) etkisini, (yuvalanmış, nested)

$f(d)$ = satırın blok içindeki etkisini, (yuvalanmış, nested)

h = spline (satır), ortak değişken olarak satırın kübik spline’in rastgele etkisini,

j = artığı (hatanın, residual) temsil etmektedir.

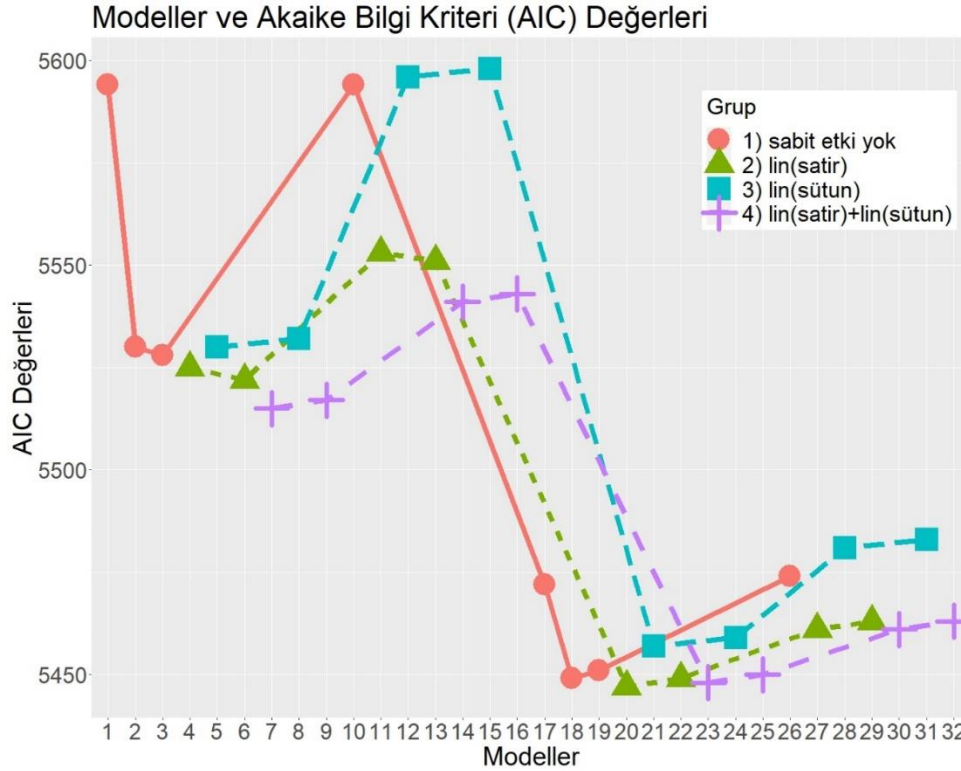
Genel olarak sabit etkinin olmadığı birinci grup ve sabit etkili (lin (sütun)) üçüncü grup modeller diğer iki model grubuna göre AIC değerleri yüksek bulunmuştur (Şekil 1). Hatanın bağımsız olduğunu varsayan, sabit etkinin olmadığı 1 ve 10 nolu modeller ile sabit etkili (lin (sütun)) 12 ve 15 nolu modeller en yüksek AIC değerine sahiptirler. Genel olarak sabit etkili (lin (sütun))+ lin (satır)) dördüncü grup modellerin AIC değerleri düşük bulunmuştur. Artığın

(hatanın) birinci derece otoregresif korelasyonunu içeren birinci, ikinci, üçüncü ve dördüncü model gruplarından sırasıyla Model-18 (5448.7), Model-20 (5447), Model-21 (5457) ve Model-23 (5448.4) en düşük AIC değerine sahiptirler. Sabit etki içeren üç grup modelde benzer trendler göstermişlerdir. Gruplar içerisinde en düşük AIC değerine sahip modeller grup içinde denenen önceki modellere kıyasla istatistiksel olarak anlamlı ($Pr < .00001$) farklılık göstermişlerdir.

Çizelge 2. Denenen modellerin varyans bileşen sayıları, maksimum olabilirlik, AIC ve modeller arası ki-kare olasılık değerleri

Model No	Grup	n.VB	logL	AIC	LR	Pr(>Chisq)
1	1	5	-2793	5594.1		
2	1	6	-2760.2	5530.4	-65.698	< .00001
3	1	7	-2757.8	5527.68	-4.716	< .00001
10	1	6	-2791.8	5593.54	67.862	< .00001
17	1	7	-2730	5472.04	-123.51	< .00001
18	1	8	-2717.3	5448.69	-25.354	< .00001
19	1	9	-2717.3	5450.68	0	1
26	1	8	-2730	5474.04	25.352	< .00001
4	2	7	-2756.4	5524.74		
6	2	8	-2754.1	5522.13	-4.608	< .00001
11	2	6	-2771.3	5552.59	34.466	< .00001
13	2	7	-2769.7	5551.32	-3.278	< .00001
20	2	9	-2715.5	5447.02	-108.3	< .00001
22	2	10	-2715.5	5449.02	0	1
27	2	8	-2723.7	5461.32	16.3	0.000054
29	2	9	-2723.7	5463.32	0	1
5	3	7	-2759.2	5530.32		
8	3	8	-2758.8	5531.64	-0.678	< .00001
12	3	6	-2792.8	5595.69	68.046	< .00001
15	3	7	-2792.8	5597.67	-0.016	< .00001
21	3	9	-2720.5	5457	-144.67	< .00001
24	3	10	-2720.5	5459	-0.004	< .00001
28	3	8	-2733.7	5481.46	26.466	< .00001
31	3	9	-2733.7	5483.46	-0.004	< .00001
7	4	8	-2750.4	5514.83		
9	4	9	-2750.3	5516.62	-0.218	< .00001
14	4	7	-2764.3	5540.55	27.932	< .00001
16	4	8	-2764.3	5542.54	-0.006	< .00001
23	4	10	-2715.2	5448.36	-98.186	< .00001
25	4	11	-2715.2	5450.35	-0.002	< .00001
30	4	9	-2722.7	5461.36	15.006	0.000107
32	4	10	-2722.7	5463.36	0	1

Grubu: sabit etkilere göre model grubu, Model no: Model sıra numarası, n. VB: Varyans bileşen sayısını, logL: maksimum olabilirlik değeri, AIC: Akaike Bilgi Kriteri değerini, LR: Önceki modele göre olasılık oranı, Pr(>Chisq): log olabilirlik oranı testinin ki-kare olasılık değerini, koyu olarak vurgulananlar grup içinde en düşük AIC değerine sahip modelleri ifade eder. İlk 16 model (Model-1-Model-16) hatanın (artığı) bağımsız, diğerleri ise otoregresif korelasyon içerdiğini varsayan modellerdir.



Şekil 1. Denenen modellerin AIC değerleri

TARTIŞMA

Orman ağacı ıslah programları kapsamında tesis edilen döl denemeleri, tarımsal bitkilerdeki denemelerden çok daha büyük alanlarda (yaklaşık 1-10 ha.) kurulmakta ve homojene yakın deneme sahası bulmak kolay olamamaktadır. Bundan dolayı her deneme sahasından elde edilen verilerin yapısı farklılık göstermekte ve her veriye uygun model seçimini gerektirmektedir.

Model seçimi, ıslah programlarında sonuçları etkilediğinden literatürde büyük ilgi görmekte ve yaygın olarak kullanılmaktadır (Isik ve ark. 2017, Kehel ve ark. 2010, Shalizi ve ark. 2020, Spilke ve ark. 2010, Verbyla 2019). Çalışmamızda aynı sabit ve raslantısal etkilere sahip modellerin artığının birinci derecede otoregresif korelasyon içermesi durumunda AIC değerinin düşmesinin sebebi, deneme sahasında çap büyümesi üzerine etki eden mekânsal varyasyondan kaynaklandığı değerlendirilmiştir. Çoğu durumda, mekânsal korelasyon içeren modellerin, mekânsal korelasyon içermeyen modellere kıyasla model uyumu daha iyi olduğu, ancak mekânsal heterojenliğin az olduğu uygun deneme

desenine sahip denemelerde iyi performans göstermediği belirtilmiştir. Bu nedenle, mekânsal modellemenin iyi deneme deseninin yerini tutamayacağı belirtilmiştir (Borges ve ark. 2019).

Basit rastlantısal doğrusal modele (rastlantısal etkili aile, blok, set (blok)) nispeten, mekânsal analizde kullanılan modellerin AIC değerlerinde ciddi bir düşüş olduğu görülmektedir (Çizelge 2). Bu denemede mekânsal korelasyondan kaynaklı hatanın mevcut olduğu görülmektedir. Mekândan kaynaklı trendin olduğu denemelerde model uyumundaki iyileştirmelerin genellikle yüksek olduğu ve mekândan kaynaklı trendin olmadığı denemelerde ise standart bir model kullanılması önerilmiştir (Dutkowski ve ark. 2006). Çalışmamızda AIC değerleri dikkate alındığında deneme sahasında satırdan kaynaklı mekânsal varyasyonun olduğu söylenebilir. Satırdan kaynaklı mekânsal bileşen ve otoregresif hata yapısı içeren Model-20 en düşük AIC değerine sahiptir. Aralarında istatistiksel olarak fark bulunmayan Model-20 ve Model-22 incelendiğinde, AIC değeri düşük olan Model-20'nin Model-22'ye kıyasla daha basit bir model olduğu ortaya çıkmaktadır. AIC değeri hesaplamasında

parametre sayısına göre ceza verildiği için bu bilgi kriteri basit modeli öncelemektedir. Verbyla (2019) tarafında yapılan bir çalışmada sabit ve raslantısal etkileri farklı 10 farklı doğrusal modelde en düşük AIC değerine sahip model mekansal bileşen ve otoregresif hata yapısı içermektedir. Özellikle büyük deneme sahalarında model seçerken mekânsal korelasyon içeren modeller de denenmelidir.

Sabit etkileri aynı olan model grupları içinde en düşük değere sahip modeller özellikle üç grupta birbirine yakındır. Farklı sabit etkilere sahip modellerin en düşük AIC değerlerinin birbirine yakın olması model seçiminde sabit etkilerin aynı olmamasına rağmen kıyaslama yapılan çalışmaları (Guerin ve Stroup 2000, Verbyla 2019) destekleyebileceği değerlendirilmektedir.

SONUÇ VE ÖNERİLER

Doğrusal model (sabit veya raslantısal, karma) kullanarak yapılan analizlerden önce model seçimi büyük önem taşımaktadır. Özellikle mekânsal varyasyon içerebilen denemelerde mekânsal bileşen içeren modeller denenerek model uyumunu arttıran en uygun model seçilerek analizlere devam edilmelidir. Aksine, uygun olmayan model kullanılarak yapılacak genetik parametre tahminleri olduğundan fazla (overestimate) veya az (underestimate) tahmin edilebilir ve ıslah programlarının ileri aşamalarını etkileyebilir.

TEŞEKKÜRLER

Bu çalışma Artvin Çoruh Üniversitesi Lisansüstü Eğitim Enstitüsünde yürütülen “Kızılçamda Genetik ve Fenotipik Parametrelerin Tahmini” adlı doktora çalışmasının bir bölümüdür. Ayrıca Orman Genel Müdürlüğü Orman Ağaçları ve Tohumları Islah Araştırma Enstitüsü Müdürlüğü tarafından “Diğer Çalışmalar” kapsamında desteklenmektedir. Bu çalışmaya katkısı olan başta Oğuz Şerif URHAN olmak üzere bireysel ve kurumsal katkılar için teşekkür ederiz.

KAYNAKLAR

- Alan M, Öztürk H, Şıklar S, Ezen T, Korkmaz B, Doğan B, Keskin S, Tulukçu M, Derilgen SI, Çalışkan B (2005) Ege bölgesi alt yükselti kuşağı ıslah zonunda (0-400 m) Kızılçam (*Pinus brutia* ten.) döl denemeleri 4. yaş sonuçları. (Teknik Bülten:13), Orman Ağaçları ve Tohumları Islah Araştırma Enstitüsü Müdürlüğü, Ankara.
- Aparicio J, Ariza-Suarez D, Raatz B (2019) Web application for spatial modelling of field trials. Paper presented at the XXIX Simposio Internacional de Estadística, Barranquilla,
- Borges A, González-Reymundez A, Ernst O, Cadenazzi M, Terra J, Gutiérrez L (2019) Can spatial modeling substitute for experimental design in agricultural experiments? *Crop Science*, 59(1):44-53
- Bozdoğan H (1987) Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3):345-370
- Brown H, Prescott R (2015) Applied mixed models in medicine. John Wiley & Sons, New York
- Burgueño J, Cadena A, Crossa J, Banziger M, Gilmour A, Cullis B (2000) User's guide for spatial analysis of field variety trials using ASREML. International Maize and Wheat Improvement Center (CIMMYT), Mexico
- Caliński T, Czajka S, Kaczmarek Z, Krajewski P, Pilarczyk W (2005) Analyzing multi-environment variety trials using randomization-derived mixed models. *Biometrics*, 61(2):448-455
- Casler MD (2015) Fundamentals of experimental design: Guidelines for designing successful experiments. *Agronomy Journal*, 107(2):692-705
- Chen Z, Karlsson B, Wu H (2017) Patterns of additive genotype-by-environment interaction in tree height of Norway spruce in southern and central Sweden. *Tree Genetics & Genomes*, 13(1):1-14
- Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. *Journal of Agricultural, Biological, Environmental Statistics*, 11(4):381-393
- Dutkowski GW, Costa e Silva J, Gilmour AR, Wellendorf H, Aguiar A (2006) Spatial analysis enhances modelling of a wide variety of traits in forest genetic trials. *Canadian Journal of Forest Research*, 36(7):1851-1870
- Dutkowski GW, Silva JCe, Gilmour AR, Lopez GA (2002) Spatial analysis methods for forest genetic trials. *Canadian Journal of Forest Research*, 32(12):2201-2214
- Fisher RA (1936) Design of experiments. *British Medical Journal* 1(3923):554
- Galwey NW (2014) Introduction to mixed modelling: beyond regression and analysis of variance. John Wiley & Sons, New York
- Gezan SA, White TL, Huber DA (2010) Accounting for spatial variability in breeding trials: a simulation study. *Agronomy Journal*, 102(6):1562-1571
- Gilmour A, Gogel B, Cullis B, Thompson R (2009) ASReml user guide release 3.0 VSN International Ltd, Hemel Hempstead
- Gilmour A, Gogel B, Cullis B, Welham S, Thompson R (2015) ASReml user guide release 4.1 functional specification. VSN International Ltd, Hemel Hempstead
- Grondona M, Crossa J, Fox P, Pfeiffer W (1996) Analysis of variety yield trials using two-dimensional separable ARIMA processes. *Biometrics*, 763-770
- Guerin L, Stroup WW (2000) A simulation study to evaluate PROC MIXED analysis of repeated measures data

- Isik F, Holland J, Maltecca C (2017) Genetic data analysis for plant and animal breeding. Springer International Publishing, Switzerland
- Kehel Z, Habash D, Gezan S, Welham S, Nachit MJAJ (2010) Estimation of spatial trend and automatic model selection in augmented designs. *Agronomy Journal*, 102(6):1542-1552
- Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology*, 74(6):1659-1673
- Little R, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O (2006) SAS for mixed models. SAS Institute Inc, North Carolina
- Matheson A, Gapare W, Ilic J, Wu H (2008) Inheritance and genetic gain in wood stiffness in radiata pine assessed acoustically in young standing trees. *Silvae Genetica*, 57(2):56-64
- Mathew B, Holand AM, Koistinen P, Léon J, Sillanpää MJ (2016) Reparametrization-based estimation of genetic parameters in multi-trait animal model using Integrated Nested Laplace Approximation. *Theoretical Applied Genetics*, 129(2):215-225
- Mramba LK (2016) Optimal experimental designs for spatially and genetically correlated data using linear mixed models. University of Florida
- Piepho HP, Möhring J, Melchinger AE, Büchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica*, 161(1-2):209-228
- Piepho HP, Williams ER, Michel V (2015) Beyond latin squares: a brief tour of row-column designs. *Agronomy Journal*, 107(6):2263-2270
- Schutz W, Cockerham CC (1966) The effect of field blocking on gain from selection. *Biometrics*, 22(4):843-863
- Shalizi MN, Gezan SA, McKeand SE, Sherrill JR, Cumbie WP, Whetten RW, Isik F (2020) Correspondence between breeding values of the same *Pinus taeda* L. genotypes from clonal trials and half-sib seedling progeny trials. *Forest Science*, 66(5):600-611
- Smith A, Cullis B, Thompson R (2001) Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics*, 57(4):1138-1147
- Spilke J, Richter C, Piepho HJPb (2010) Model selection and its consequences for different split-plot designs with spatial covariance and trend. *Plant Breeding*, 129(6):590-598
- Stroup WW (2012) Generalized linear mixed models: modern concepts, methods and applications. CRC Press, Boca Raton.
- Urhan OS, Kolpak SE, Jayawickrama KJS, Howe GT (2014) Early genetic selection for wood stiffness in juvenile Douglas-fir and western hemlock. *Forest Ecology and Management*, 320:104-117
- Verbyla AP (2019) A note on model selection using information criteria for general linear models estimated using REML. *Australian & New Zealand Journal of Statistics*, 61(1):39-50
- VSNI T (2021) A brief look at spatial modelling. In. /, (01.10.2022).
- Welham SJ, Gezan SA, Clark SJ, Mead A (2014) Statistical methods in biology: design and analysis of experiments and regression. CRC Press, London.
- Welham SJ, Gogel BJ, Smith AB, Thompson R, Cullis BR (2010) A comparison of analysis methods for late-stage variety evaluation trials. *Australian and New Zealand Journal of Statistics*, 52(2):125-149
- White TL, Adams WT, Neale DB (2007) Forest genetics. Cabi, United Kingdom.
- Yates F (1939) The recovery of inter-block information in variety trials arranged in three-dimensional lattices. *Annals of Eugenics*, 9(2):136-156
- Ye TZ, Jayawickrama KJ (2008) Efficiency of using spatial analysis in first-generation coastal Douglas-fir progeny tests in the US Pacific Northwest. *Tree Genetics & Genomes*, 4(4):677-692