

**g-ROKY: MADEN KAYNAK KESTİRİMİNDE
HOMOJEN ORTAM MODELLENMESİNE YÖNELİK
YENİ BİR YÖNTEMİN GELİŞTİRİLMESİ**

**g-ROKY: DEVELOPING A NEW METHOD FOR DOMAIN
MODELING IN MINERAL RESOURCE ESTIMATION**

HAMZA KIVRAK

PROF. DR. A. ERHAN TERCAN
Tez Danışmanı

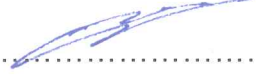
Hacettepe Üniversitesi
Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin
Maden Mühendisliği Anabilim Dalı için Öngördüğü
DOKTORA TEZİ olarak hazırlanmıştır.

2018

HAMZA KIVRAK'ın hazırladığı “g-ROKY: Maden Kaynak Kestiriminde Homojen Ortam Modellenmesine Yönelik Yeni Bir Yöntemin Geliştirilmesi” adlı bu çalışma aşağıdaki jüri tarafından **MADEN MÜHENDİSLİĞİ ANA BİLİM DALI**'nda **DOKTORA TEZİ** olarak kabul edilmiştir.

Prof. Dr. Neşe ÇELEBİ

Başkan



Prof. Dr. A. Erhan TERCAN

Danışman



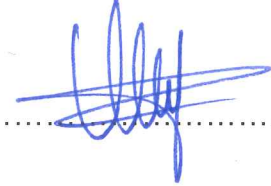
Prof. Dr. Sair KAHRAMAN

Üye



Doç. Dr. Nuray DEMİREL

Üye



Yrd. Doç. Dr. Güneş ERTUNÇ

Üye



Bu tez Hacettepe Üniversitesi Fen Bilimleri Enstitüsü tarafından **DOKTORA TEZİ** olarak onaylanmıştır.

Prof. Dr. Menemşe GÜMÜŞDERELİOĞLU

Fen Bilimleri Enstitüsü Müdürü

YAYINLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezimin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanması zorunlu metinlerin yazılı izin alarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

- Tezimin/Raporumun tamamı dünya çapında erişime açılabilir ve bir kısmı veya tamamının fotokopisi alınabilir.**

(Bu seçenikle teziniz arama motorlarında indekslenebilecek, daha sonra tezinizin erişim statüsünün değiştirilmesini talep etmeniz ve kütüphane bu talebinizi yerine getirirse bile, tezinin arama motorlarının önbelleklerinde kalmaya devam edebilecektir.)

- Tezimin/Raporumun tarihine kadar erişime açılmasını ve fotokopi alınmasını (İç Kapak, Özet, İçindekiler ve Kaynakça hariç) istemiyorum.**

(Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir, kaynak gösterilmek şartıyla bir kısmı ve ya tamamının fotokopisi alınabilir)

- Tezimin/Raporumun tarihine kadar erişime açılmasını istemiyorum, ancak kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisinin alınmasını onaylıyorum.**

- Serbest Seçenek/Yazarın Seçimi**

18 /01 / 2018


Hamza KIVRAK

ETİK

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada,

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

18.01.2018



HAMZA KIVRAK

ÖZET

g-ROKY: MADEN KAYNAK KESTİRİMİNDE HOMOJEN ORTAM MODELLENMESİNE YÖNELİK YENİ BİR YÖNTEMİN GELİŞTİRİLMESİ

Hamza KIVRAK

Doktora, Maden Mühendisliği Bölümü

Tez Danışmanı: Prof. Dr. A. Erhan TERCAN

Ocak 2018, 73 Sayfa

Maden kaynak kestiriminin önemli adımlarından biri maden yatağında birbirine benzer dağılım gösteren tenör, litolojik, mineralojik zonların varlığının sorgulanması ve tespit edilmesidir. Bu zonların homojenlik gösterdiği alanları belirlemenin kestirim kalitesini arttırdığı bilinmektedir. Literatürde maden yataklarında homojen ortamların tespit edilmesi amacıyla kullanılan yöntemler yalnızca örneklemdaki homojen zonlara ilişkin kategorik değişkenleri ve konumsal özellikleri dikkate alarak bir sınıflandırma yapmaktadır.

Tezin amacı, maden yataklarındaki homojen ortamların kestiriminde tenör değişkenlerini de hesaba katan yeni yöntemler geliştirmektir. Bu amaçla, Rastlantı Ormanı (RO) yöntemi ile Krigleme yönteminin birleştirilmesine dayanan iki yöntem geliştirilmiştir. Bunlardan Rastlantı Ormanı + Krigleme yöntemi (ROKY), örneklemdaki tenörlerle kategorik değişken arasındaki ilişkinin RO ile tanımlanması, örneklenmemiş lokasyonlardaki tenörlerin Ortalamasız Krigleme yöntemi ile kestirilmesi ve kestirim değerlerinin RO'ya beslenmesi adımlarını içerir. Gruplandırılmış Rastlantı Ormanı +

Krigleme yöntemi (g-ROKY) ise hatalı sınıflandırmalara yol açan Kriglemenin törpüleme (smoothing) etkisini azaltmak için geliştirilmiştir. ROKY'den farklı olarak orijinal ve kriglenmiş verilerin standartlaştırılarak gruplanmasına, gruplanmış verilerle RO'nun oluşturulması ve homojen ortamların tayinine dayanır.

İki yeni yöntemin uygulanabilmesi ve elde edilen sonuçların analiz edilebilmesi için bir MATLAB programı kodlanmıştır.

Geliştirilen yöntemler, kamuya açık Jura veri kümesi ile İğdekuzu linyit sahasında uygulanmış ve test edilmiştir. Yapılan sınıflandırmalar standart bir yaklaşım olan İndikatör Krigleme yöntemi ile karşılaştırılmıştır. Karşılaştırılan yöntemler içinde g-ROKY gerçek homojen ortamlara en yakın kestirim değerleri üreten yöntem olmuştur.

Anahtar Kelimeler: Homojen ortam, Jeostatistik, Variogram, Ortalamasız Krigleme, İndikatör Krigleme, Sınıflandırma, Rastlantı Ormanı, ROKY, g-ROKY

ABSTRACT

g-ROKY: DEVELOPING A NEW METHOD FOR DOMAIN MODELING IN MINERAL RESOURCE ESTIMATION

Hamza KIVRAK

Doctor of Philosophy, Department of Mining Engineering

Supervisor: Prof. Dr. A. Erhan TERCAN

January 2018, 73 Pages

One of the crucial steps of mineral resource estimation is to examine and determine the presence of domains which have similar grade, lithological and/or mineralogical distribution in mineral deposits. It is well known that defining the domains improves estimation quality. The methods traditionally used in modeling domains only take categorical variables of domains and their spatial relations into account.

The objective of this thesis is to develop new methods considering grade variable in domain modeling. For this purpose, two methods are developed, based on combination of Random Forest (RF) and Ordinary Kriging. The first method, called as ROKY (Random Forest + Kriging); includes the stages of defining the relationship between grades and categorical variable by RF, estimating the grade values at unsampled locations by Ordinary Kriging and feeding the estimated values into RF. The second method, g-ROKY (Grouped Random Forest + Kriging); is developed for avoiding the smoothing effect of kriging which causes misclassifications. g-ROKY, different from ROKY, is based on grouping the sample and kriged estimates in standardized form, constructing RO by grouped data and finally determination of the domains.

A MATLAB script is written in order to apply the two methods and to analyze the results obtained.

The methods are applied to public domain Jura data set and also İğdekuzu Lignite Field. In addition, the classifications made by these two methods are compared with traditional Indicator Kriging method. Within the compared methods, g-ROKY is the method that produces the closest estimation values to true domains.

Keywords: Domain, Geostatistics, Variogram, Ordinary Kriging, Indicator Kriging, Classification, Random Forest, ROKY, g-ROKY

TEŞEKKÜR

Deneyimlerinden yararlanmama olanak sağlayan, tezin olgunlaşmasında katkılarını esirgemeyen, teze ve tezin gelişim sürecine eleştirel gözle bakarak çalışmalarımı yoğunlaştırmam gereken alanlara ışık tutan tez danışmanım Prof. Dr. A. Erhan TERCAN'a teşekkür ederim.

Bu çalışmada geliştirilen yeni yaklaşımların esin kaynağı 215M999 no'lu TÜBİTAK projesinde yürütülen çalışmalar olmuştur. Bundan dolayı TÜBİTAK'a teşekkür ederim.

Yapıcı eleştirilerinden ve katkılarından dolayı Tez İzleme Komitesi ile Tez Savunma Jürisi üyeleri; Prof. Dr. Neşe ÇELEBİ, Prof. Dr. Sair KAHRAMAN, Doç. Dr. Nuray DEMİREL ve Yrd. Doç. Dr. Güneş ERTUNÇ'a teşekkür ederim.

Tüm desteklerinden dolayı Urmia Teknoloji Üniversitesinde öğretim görevlisi olarak görev yapan Babak SOHRABIAN'a ve Hacettepe Üniversitesi Maden Mühendisliği Bölümü'nde araştırma görevlisi olarak görev yapan Fırat ATALAY'a teşekkür ederim.

Yaşadığım olumsuzluklarda motivasyonumu yüksek tutmama vesile olan İş Başmüfettişi üstadım Rahmi İNAN'a, İş Müfettişi meslektaşlarım Mustafa Orkun ÇELİK, Mehmet Kamil GÜLSOY ve Ersan BAK'a çok teşekkür ederim.

Yoğun çalışma tempomda yanımda olduklarını hep hissettiren; Annem Meryem, babam Hasan KIVRAK'a, Ankara'da ve İstanbul'da yaşayan geniş ailemin bütün bireylerine teşekkür ederim.

Doktora eğitim sürecim boyunca sabrından, emeğinden ve bana kattıklarından dolayı hakkını ödeyemeyeceğim can yoldaşım, eşim Ebru'ya sonsuz teşekkür ederim.

İÇİNDEKİLER

	<u>Sayfa</u>
ÖZET	i
ABSTRACT	iii
TEŞEKKÜR	v
İÇİNDEKİLER	vi
SİMGELER VE KISALTMALAR	viii
1. GİRİŞ	1
1.1. Problemin Tanımı	2
1.2. Tezin Amacı ve Kapsamı	3
1.3. Araştırma Metodolojisi	3
1.4. Beklenen Bilimsel Etki.....	4
1.5. Tezin İçeriği	4
2. LİTERATÜR ÇALIŞMASI.....	5
2.1. Önceki Çalışmalar ve Tezde Geliştirilen Yöntemlerin Bunlar İçindeki Yeri	5
2.2. Krigleme yöntemi	7
2.2.1. Variogram Modelleme	8
2.2.2. Ortalamasız (Ordinary) Krigleme yöntemi	10
2.2.3. İndikatör Krigleme yöntemi	14
2.3. Rastlantı Ormanı Metodu İle Sınıflandırma	15
2.3.1. Sınıflandırma ve Regresyon Ağaçları (Srat), Gini İndeksi Ve Budama İşlemi ..	16
2.3.2. Bootstrap yöntemi	19
2.3.3. Bagging yöntemi.....	21
2.3.4. Rastlantısal Düğüm Değişkeni Seçimi.....	22
2.3.5. Rastlantı Ormanı Algoritması	23
2.3.6. Seçilmemiş Veri Sınıf Kestirimi Ve Kestirim Hatası	25
3. HOMOJEN ORTAM MODELLENMESİNE YÖNELİK YENİ GELİŞTİRİLEN YÖNTEMLER.....	27
3.1.1. Rastlantı Ormanı + Krigleme yöntemi (ROKY)	27
3.1.2. Gruplandırılmış Rastlantı Ormanı + Krigleme yöntemi (g-ROKY).....	28
3.2. Rastlantı Ormanı Algoritması, Kodlanması Ve Kodun Çalışma Prensipleri	29

4. YÖNTEMLERİN JURA VERİ KÜMESİ ÜZERİNDE TESTİ	32
4.1. Veri Özellikleri	32
4.2. Rastlantı Ormanı İle Jura Veri Kümesinin Sınıflandırılması	36
4.3. İndikatör Krigleme İle Kaya Türlerinin Kestirimi	38
4.4. Rastlantı Ormanı + Krigleme yöntemiyle (Roky) Jura Veri Kümesi Sınıflandırılması	41
4.5. Gruplandırılmış Rastlantı Ormanı + Krigleme yöntemiyle (G-Roky) Jura Veri Kümesinin Sınıflandırılması.....	47
5. GELİŞTİRİLEN YÖNTEMLERİN İĞDEKUZU LİNYİT SAHASINDA UYGULANMASI	49
5.1. İğdekuzu linyit sahası Verileri	49
5.2. İndikatör Krigleme İle Litolojilerin Kestirimi.....	51
5.3. Roky İle Litolojilerin Kestirimi	56
5.4. g-Roky İle Litolojilerin Kestirimi	63
6. SONUÇLAR VE ÖNERİLER.....	67
6.1. Sonuçlar.....	67
6.2. Öneriler	68
ÖZGEÇMİŞ	72
CURRICULUM VITAE	73

SİMGELER VE KISALTMALAR

Simgeler

$\gamma(h)$	Yarıvariogram (Yarı variogram)
$E []$	Beklenen değer (Expected Value)
$I []$	İndikatör fonksiyon
Prob []	Olasılık fonksiyonu

Kısaltmalar

RO	Rastlantı Ormanı
ROKY	Rastlantı Ormanı + Krigleme yöntemi
g-ROKY	Gruplandırılmış Rastlantı Ormanı+Krigleme yöntemi
TÜBİTAK	Türkiye Bilimsel ve Teknik Araştırma Kurumu
IK	İndikatör Krigleme
IBF	Işınsal Bazlı Fonksiyonlar
SRAT	Sınıflandırma ve Regresyon Ağacı Tekniği
OOB	Seçilmemiş Veri Kümesi (Out of bag)
İnd	İndikatör
Var	Varyans

1. GİRİŞ

Arama safhası madenciliğin ilk aşamalarından biri olup bu aşamada elde edilen verilerin değerlendirilmesi madencilik açısından oldukça büyük bir öneme sahiptir. Bu safhada elde edilen verilerin mühendislik süzgecinden geçirilmesi maden yataklarına ilişkin bilinmezliği en aza indirgeyecek ve madenin planlanmasına olumlu katkıda bulunacaktır.

Arama safhasında bir maden yatağının bütün noktalarından örnek alınması hem fiziksel olarak hem de maliyet açısından mümkün olmamaktadır. Bu sebeple örneklenmiş lokasyonlardan elde edilen bilgilerle maden yatağının tamamının tanımlanmasına yönelik çeşitli yaklaşımlar geliştirilmiştir. Poligon yöntemi, Uzaklığın Tersisi ile Ağırlıklandırma yöntemi, Krigleme (Kriging) yöntemi ve Jeostatistiksel Simülasyonlar bu işlevi gören farklı yöntemlerdir. Bu yöntemlerle bir maden yatağında cevherin kalitesine ve miktarına ilişkin kestirimler yapılabilmektedir.

Maden yatağında cevher kalitesini ve miktarını belirleyen değişkenlerin kestiriminden önce yatağın homojen ortamlara ayrılması kestirim kalitesini artırmaktadır. Homojen ortam, maden yatağı içerisinde benzer öznelik dağılımı ve jeolojik özellik gösteren zonlar olarak tanımlanabilir. Örneğin, bir maden yatağında düşük tenörlü cevherle yüksek tenörlü cevherin yatağın belirli bölgelerinde toplandığını varsayalım. Bu yatakta tenöre ilişkin homojen ortamlar belirlendiğinde, kestirim sırasında, yüksek tenörlerin düşük tenörlü zonu ya da düşük tenörlerin yüksek tenörlü zonu etkilemesinin önüne geçilmiş olur. Tenör değerleri yanında cevherin litolojik, mineralojik ve alterasyon özellikleri için de benzer nitelikte zonlar yani homojen ortamlar oluşabilmektedir.

Homojen ortamların tayin edilmesine yönelik literatürde bir çok yaklaşım mevcuttur. Bu yöntemler; kestirimi esas alan Kesit Yöntemi, Işınsal Bazlı Fonksiyonlar (Radial Basis Functions), Çoklu Gauss Krigleme (Multi Gaussian Krigging), İndikatör Krigleme ve İndikatör Eş Krigleme yöntemleri; sınıflandırmayı esas alan Log-Olasılık Grafiği yöntemi, K-Ortalamaları, Hiyerarşik Kümeleme (Hierarchical Clustering) yöntemleri olarak sayılabilir. Bütün bu sayılan yöntemlerde homojen ortamların belirlenmesi işlemi veri kümeleri içinde homojen ortamlara ilişkin kategorik veriler ile bunların konumsal özellikleri kullanılarak yapılmakta olup cevherin kalite değişkenlerinin homojen ortam

belirlenmesinde bir rolleri bulunmamaktadır. Örneğin, bir linyit sahasında yukarıda bahsedilen yöntemler kullanılarak linyitin litolojik homojenliğinin tespit edilmeye çalışıldığını varsayalım. Literatürde var olan yöntemlerle homojen litolojik birimlerin tespitinde sondaj ve karotlardan elde edilen litolojiye dair bilgiler ve her bir sondajın konumsal özellikleri kullanılmaktadır. Ancak bu yöntemlerle homojen ortamlar kestirilirken linyitin alt ısı değer, kül, nem vb. kalite değişkenleri dikkate alınmamaktadır.

1.1. Problemin Tanımı

Maden kaynak kestirimi, bir maden varlığının geometrisinin, miktarının, tenör ve dağılımının jeolojik bilgi ve sondaj verilerinden kestirimini içerir. Kaynak kestirim süreci temel olarak veri tabanı oluşturma, homojen jeolojik katı modelleme, blok modelleme-tenör kestirimi ve kaynak model üretimi adımlarından oluşur. Veri tabanı, cevher yatağı modelleme ve ocak tasarımına temel olan (jeolojik haritalar, sondaj logları vs.) verilerin yönetiminden sorumludur. Kavramsal jeolojik katı model, kaynak kestirim sınırlarını belirler. Bu sınırlar içinde benzer jeolojik özellikler ve öznitelik dağılımı gösteren ortamları (domains) tanımlamak gerekir. Sabit ya da değişken boyutlu bloklar bu ortamlar içinde oluşturulur ve blokların öznitelikleri bu ortamlar içinde kestirilir. Gerçekte homojen jeolojik ortam modellemesi ve bloklama işlemi bağımsız olmayıp birbirini tamamlayan süreçlerdir. Bir maden yatağında homojen ortamların belirlenmesi ve ortama göre bir kestirim süreci izlenmesi gerekir. Bu yapılmadığında standardın altında kestirimler elde edilir ve normalin çok altında ya da çok üstünde tenör ve tonaj değerleri üretilir [1], [2].

Homojen jeolojik ortam modellemesi ya da kısaca ortam modellemesi litoloji, mineraloji, alterasyon, yapısal unsurlar gibi jeolojik değişkenler ile sondaj karotlarının analizinden elde edilen öznitelik bilgisine dayanır. Problem, bu bilgiler verildiğinde maden kaynak kestiriminde kullanılan blokları ortamlara atamaktır.

Bir $D \in \mathbb{R}^3$ yatağının K sayıda ortama ($D \cup D_k$, $k=1, \dots, K$ ve $k \neq j$ ise $D_k \cap D_j = \emptyset$) ayrıldığını varsayalım. Ayrıca D yatağı içinde N adet lokasyonda [s_i , $i=1, \dots, N \in D$] çok değişkenli verilerin ölçüldüğünü ve bu ölçümlerin P sayıda sürekli bağımsız değişken $Z(s)=[Z_1(s), \dots, Z_P(s)]^t$ ve $Y(s)$ kategorik bağımlı değişkenle ilişkili olduğunu varsayalım.

Bu ifadede üstel t ; vektörün evriğini (transposition) göstermektedir. D yatağı, ölçümlerin yapıldığı malzeme (genelde sondaj karotları) hacmine göre çok daha büyük olan M sayıda $[b_j, j=1, \dots, M \in D, M \gg N]$ bloğa bölünsün. Bu durumda problem, $Z(s)$ ve $Y(s)$ bilgisini kullanarak $b_j, j=1, \dots, M$ blokların her birini $D_k, k=1, \dots, K$ ortamlarından birine atamak olarak tanımlanabilir.

1.2. Tezin Amacı ve Kapsamı

Tez, maden kaynak kestiriminde homojen ortam modellenmesine ilişkin daha objektif ve gerçek değerlere yakın sonuçlar üreten yeni yaklaşımlar geliştirmeyi amaçlamıştır. Yeni yöntemlerin, veri kümesindeki homojen ortamlar ve onların konumsal ilişkileri ile birlikte cevherin tenör değişkenlerini de dikkate alması amaçlanmaktadır. Bu amaçla iki yeni yöntem geliştirilmiştir.

Tez, temel olarak kaynak modeli oluşturan blokların homojen jeolojik ortamlara atanması ile ilgili olup öznelitliklerin ortam sınırlarındaki davranışı ve bunların modellenmesini içermemektedir.

1.3. Araştırma Metodolojisi

Tez kapsamında yeni geliştirilen yöntemler, Krigleme yöntemiyle Rastlantı Ormanı yönteminin birleşimine dayanmaktadır. Bu yöntemlerden ROKY (Rastlantı Ormanı + Krigleme yöntemi), örneklemdaki tenör ve kategorik değişkenleri ilişkilendiren Rastlantı Ormanlarının üretimi, tenör değişkenlerinin Ortalamasız Krigleme (OK) ile kestirimi ve kestirim değerlerinin Rastlantı Ormanına beslenmesi adımlarını içerir. g-ROKY (Gruplandırılmış Rastlantı Ormanı + Krigleme yöntemi) ise kriglemenin törpüleme (smoothing) etkisinin sebep olduğu sınıflandırma hatasını azaltmak amacıyla geliştirilmiştir. g-ROKY; gerçek ve kriglenmiş verilerin standartlaştırılması (z-skora dönüştürülmesi), z skorların gruplandırılması, gruplandırılmış gerçek verilerle rastlantı ormanının oluşturulması ve gruplandırılan kestirim değerlerinin Rastlantı Ormanına beslenerek sınıfların belirlenmesi adımlarından oluşur.

İndikatör Krigleme yöntemi (IK), ROKY ve g-ROKY'nin ortam modellemedeki performansları iki farklı veri kümesi için incelenmiştir. Yöntemlerin performansları, örneklemlerin gerçek homojen ortamlarına ne kadar yakın kestirimlerin yapıldığı dikkate alınarak karşılaştırılmıştır.

Araştırma metodolojisi aşağıdaki adımları içermektedir:

- Jeostatistiksel kestirimden elde edilen kestirim değerlerinin beslenebildiği, bu değerlere karşılık gelen homojen ortamları tayin eden ve yapılan sınıflandırma doğruluğuna ilişkin bilgiler üreten algoritmanın MATLAB programı aracılığıyla kodlanması,
- Geliştirilen yeni yaklaşımların ve IK'nın iki boyutlu Jura veri kümesi üzerinde testi,
- Yeni yaklaşımların ve IK yönteminin İğdekuzu linyit sahasına uygulanması,
- Uygulama sonuçlarının karşılaştırılması.

1.4. Beklenen Bilimsel Etki

Çalışma kapsamında yeni yöntemlerin geliştirilmiş olması maden kaynak kestirimi konusunda çalışanların ilgisini çekecektir. Özellikle gerçek homojen ortam değerlerine en yakın kestirim değerleri üreten g-ROKY'nin cevher modelleme ve ocak tasarımına ilişkin bütünleşik madencilik yazılımlarında kullanıma potansiyeli vardır. Yöntem, jeostatistik konusunda çalışan araştırmacılar, maden kaynak/rezerv kestirimi yapan (uygulayıcı) kişiler tarafından kullanılacaktır.

Bu konu ilk defa ele alınacağı için sonuçları ilgi çekici olacak ve pek çok tez çalışması ve yeni projelerin başlatılmasına yol açacaktır.

1.5. Tezin İçeriği

Tezin 2. bölümünde literatür çalışmalarına yer verilmiştir. 3. bölümde yeni geliştirilen yöntemlere yönelik bilgiler sunulmuştur. Yeni yöntemler 4. bölümde Jura veri kümesinde test edilmiştir. 5. bölümde İğdekuzu linyit sahası için yapılan uygulama verilmiştir. Tezin son bölümünde elde edilen sonuçlar analiz edilerek öneriler sunulmuştur.

2. LİTERATÜR ÇALIŞMASI

Bu bölümde homojen ortam modellenmesine yönelik literatürde kullanılan yöntemler ve tez çalışmasında geliştirilen yöntemlerin bunlar içindeki yerinden bahsedilmiştir. Ayrıca tez kapsamında kullanılan yöntemlere dair teorik bilgiler verilmiştir. Bu amaçla Krigleme (Variogram Modelleme, Ortalamasız Krigleme (Ordinary Kriging), İndikatör Krigleme), Sınıflandırma Ağaçları ve Rastlantı Ormanı yöntemleri açıklanmıştır.

2.1. Önceki Çalışmalar ve Tezde Geliştirilen Yöntemlerin Bunlar İçindeki Yeri

Ortam modellemesi için literatürde şimdiye kadar bir çok yöntem önerilmiştir. Bunları kestirime dayanan yöntemler ve sınıflandırma-kümeleme esaslı yöntemler şeklinde kabaca iki gruba ayırmak mümkündür. Kesit yöntemi, ışınsal bazlı fonksiyonlar (radial basis functions), çokluGauss krigleme (multiGaussian Krigleme yöntemi) ve indikatör eşkrigleme gibi yöntemler ilk grup içinde yer almaktadır. Log-olasılık (log-probability) grafiği yöntemi ve K-ortalamları (K-means), hiyerarşik kümeleme (hierarchical clustering) gibi yöntemler ise ikinci grup içinde yer almaktadır.

Kesit yöntemi, eski ve zaman alan geometrik bir yöntemidir. Cevherleşmeyi kontrol eden en etkili jeolojik değişkenleri ve tenör dağılımını göz önüne alarak yatak içinde iki boyutta kesitler oluşturulur ve bunlar üç boyutlu modeller yaratmak üzere birleştirilir. Bu yöntem, günümüz bilgisayar teknolojileri ile yarı otomatik hale getirilmesine karşın hala çok zaman almakta, kaynak veri tabanına yeni bilgiler eklendiğinde modelin güncellenmesi çok uzun sürmektedir. Bu dezavantajlarına rağmen yöntem, kaynak modellemede en çok kullanılan yöntemdir.

Modelleme işlemine hız kazandırmak amacıyla önerilen bir yöntem ışınsal bazlı fonksiyonlar (radial basis functions) yöntemidir [3]. Işınsal bazlı fonksiyonlar (IBF), D boyutlu (D= 1, 2 ve 3 olabilir) bir uzaydaki veri kümesinin iç kestirimi (interpolation) amacıyla türetilen sürekli fonksiyonlardır. Verilerin ölçüldüğü lokasyonlarda fonksiyonun değerleri aynı ancak diğer noktalarda törpülenmiştir (smoothing). IBF modelleme yaklaşımı literatürde örtülü (implicit) modelleme yaklaşımı olarak da bilinmektedir. IBF yönteminin çift krigleme (dual Krigleme yöntemi) yöntemine eşdeğer olduğu Matheron [4] tarafından gösterilmiştir. Yöntem, global (tüm verileri kullanan) bir iç kestirim yöntemi olduğu için son zamanlara kadar büyük boyutlu veri kümelerine

uygulama imkanı olmamıştır. Bilgisayar teknolojisi ve sayısal modellemedeki son gelişmeler IBF yönteminin ortam modellemedeki uygulama yeteneklerini artırmış ancak geniş bir kabul görmemiştir.

Ortamların modellenmesine yönelik olarak maden kaynak kestiriminde sıklıkla kullanılan bir yöntem log-olasılık grafiği yöntemidir [5]. Log olasılık grafikleri aynı zamanda jeokimyasal anomalilerin belirlenmesinde kullanılan fraktal yaklaşımların [6] temelini oluşturur. Yöntem, tenör değerlerinin birikimli olasılık grafiğinin oluşturulmasına dayanır. Hem tenörler hem de olasılıklar logaritmik ölçekte gösterilir. Yönteme göre grafiğin eğimindeki bir değişim, farklı tenör popülasyonlarının varlığına işaret eder. Değişimin gözlemlendiği tenör değeri, popülasyonları ayıran sınır değer olarak alınır. Log-olasılık grafiği yaklaşımı gerçekte tenör popülasyonlarını ayırmaktan çok, tenörün lognormallikten (ya da normallikten) sapmasını ölçen bir araçtır [7]. Bu yöntem jeolojik değişkenlerle tenör dağılımı arasındaki ilişkiyi dikkate almaz. Önce jeolojik değişkenlere göre bir modelleme yapılır daha sonra bu modeller içinde log-olasılık grafiği yaklaşımı ile belirlenen sınır değerlere dayanarak düşük, orta ve yüksek tenörlü zonlar belirlenir.

ÇokluGaussian krigleme [8], homojen tenör zonlarını modellemek amacıyla önerilen diğer bir yöntemdir. Örneklenmemiş her bir lokasyonda tenörün kestirimi yerine belirli bir sınır değerden yüksek olma olasılıkları hesaplanır. Tenöre ilişkin ortamlar bu olasılıklara göre sınıflara ayrılır ve tenör kabukları (grade shells) oluşturulur. Bu yaklaşımdaki temel problem, olasılık eşliğinin (probability threshold) seçimidir. Bunun seçimine ilişkin standart bir yöntem geliştirilmemiştir. Bu eşik yüksek seçildiğinde birbirinden ayrık çok sayıda tenör kabukları oluşur, eşik düşük olduğunda ise homojen olmayan çok büyük tenör ortamları yaratılır.

İndikatör eşKrigleme yöntemi standart bir yöntem olup bir değişkenin birden çok değişkenle kestirimini içerir. Değişken sayısı üç ve üçten büyük olduğunda çapraz variogramların modellenmesi ve eşkrigleme sistemlerinin çözümünde problemlerle karşılaşılır. Ayrıca bu yöntemde sadece litolojik bilgiler kullanılabilir. Bu kısıtlara rağmen bu yaklaşım maden kaynak kestiriminde kullanılmaktadır.

Kümeleme yöntemleri, ölçüm değerlerini aynı kümedeki veriler mümkün olduğunca birbirine benzeyecek şekilde kümelere ayırmaya çalışır. Kümeleme, danışmansız bir öğrenme (unsupervised learning) yöntemi olup makine öğrenmesi, veri madenciliği, örüntü tanıma, imge analizi ve uzaktan algılama gibi alanlarda geniş bir uygulama alanı bulmuştur. Buna karşılık maden kaynak kestirimindeki kullanımı oldukça sınırlıdır. İlk çalışma Romary vd. [9] tarafından gerçekleştirilmiştir. Bu araştırmacılar K-ortalamlar, Hiyerarşik Kümeleme ve model bazlı algoritmaları gözden geçirerek, bu yöntemleri bir uranyum yatağına uygulamışlardır. Ayrıca 2016 yılında, Mariz ve Wagner [10], Hiyerarşik Kümeleme ile Destek Vektör Makineleri yöntemlerinin bileşiminden oluşan bir yaklaşımla Avustralya'da bir demir yatağında düşük ve yüksek tenörlü demir cevherinin homojen ortamlarını belirlemişlerdir.

Kestirime ve sınıflandırmaya dayanan ve yukarıda bahsedilen yöntemlerin tümünün ortak özelliği, örneklenmemiş lokasyonlardaki homojen ortamların belirlenmesinde kategorik değişken ve onun konumsal özelliklerini kullanıyor olmalarıdır. Bu yöntemlerden hiçbiri örneklenmemiş lokasyonlardaki homojen ortamların tespitinde veri kümesindeki tenöre dair değişkenleri kullanmamaktadır. Tez kapsamında yeni geliştirilen ve Rastlantı Ormanı ile Krigleme yöntemini birleştiren ROKY ise konum ve homojen ortama ilişkin değişkenlerle birlikte cevherin özneteliğini de dikkate alarak bir sınıflandırma yapmaktadır. ROKY'nin Kriglemeden kaynaklı törpüleme etkisinden (smoothing) kaynaklanan sınıflandırma hatasını azaltabilmek için standartlaştırılmış ve gruplandırılmış verilerle sınıflandırma yapan g-ROKY yöntemi de tez kapsamında geliştirilen yeni bir yöntem olmuştur.

2.2. Krigleme yöntemi

Jeostatistik, örneklenmemiş lokasyonlardaki değişkenlerin kestirimi için kullanılan krigleme algoritmasının kullanıldığı yöntemleri kapsamaktadır. Krigleme yöntemi ise En Küçük Kareler Regresyon Algoritması ailesinin bir parçası olup bu yöneme yönelik ilk çalışmalar Krige [11] tarafından yapılmıştır [12].

Krigleme yöntemi, bir cevher bloğunda mevcut ve ilgili verileri en uygun şekilde ağırlıklandırarak en iyi doğrusal ve yansız kestirim sonuçları veren çoklu regresyon

yöntemidir. İki temel avantajı; sistematik yanlılıktan kaçınması ve kestirim hatalarını (krigleme hatası) asgariye indiriyor olmasıdır [13].

Krigleme yöntemi sırasıyla aşağıdaki adımları içermektedir:

- i. Konumsal ilişkiye sahip değişkenler içeren ve bu değişkenlere yönelik istatistiksel ortalama kavramını anlamlı kılacak homojenlikte bir alanın belirlenmesi,
- ii. Belirlenen bu alan içerisinde var olan bütün verilerin taranması suretiyle uzaklığa bağlı değişkenliğin tayinine yönelik mümkün olan her bir uzaklık ve yön için deneysel variogram değerlerinin hesaplanması,
- iii. Hesaplanan deneysel variogram değerlerinin uzaklığa bağlı değişimini tanımlayan bir teorik variogram modelinin oluşturulması,
- iv. Oluşturulan model kullanılarak Krigleme yöntemi ile örneklenmemiş lokasyonlarda kestirimlerin yapılması.

2.2.1. Variogram Modelleme

Variogram kavramı Matheron [14] tarafından ortaya atılmış olup ilk olarak 1963 yılında Güney Afrikada bir cevher yatağındaki tenör değerlerinin kestirimi için kullanılmıştır. Jeostatistiksel analizin en önemli araçlarından biri variogramlardır. Variogramlar, örneklerin uzaklığa bağlı ilişkisini veri çiftlerinin arasındaki uzaklığı ve açığı dikkate alarak hesaplar.

Deneysel yarı variogram değerleri Eş. 2.1 ile hesaplanmaktadır.

$$\gamma(h) = \frac{1}{2N(h)} \sum_{k=1}^{N(h)} (Z(x_k) - Z(x_k + h))^2 \quad (2.1)$$

Eşitlikte;

h : Veri çifti arasındaki uzaklığı,

$Z(x)$: x konumundaki Z değişkeninin değerini,

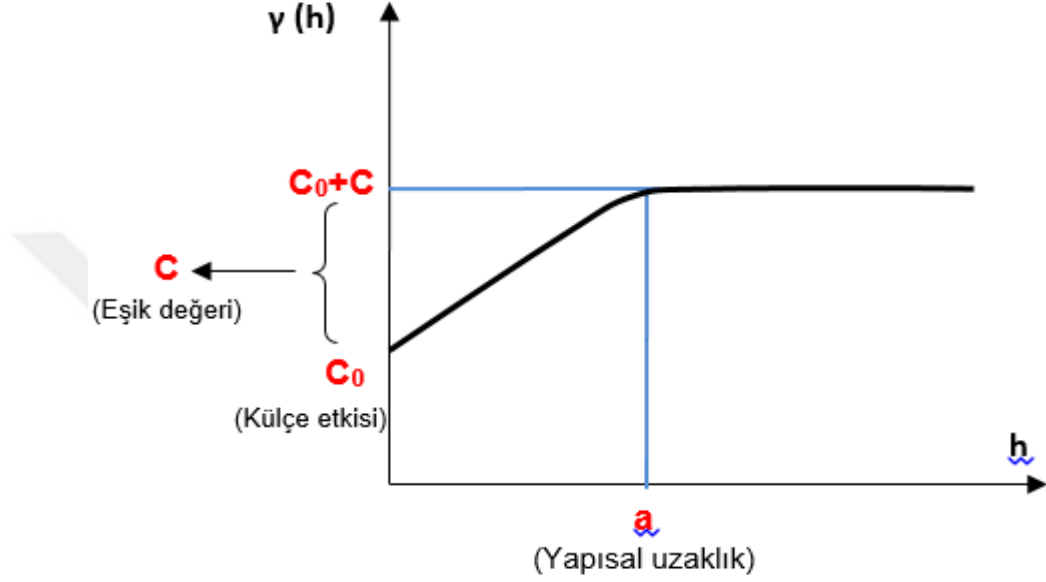
$Z(x+h)$: $x+h$ konumundaki Z değişkeninin değerini,

$\gamma(h)$: Yarı variogramı

$N(h)$: Veri çifti sayısını göstermektedir.

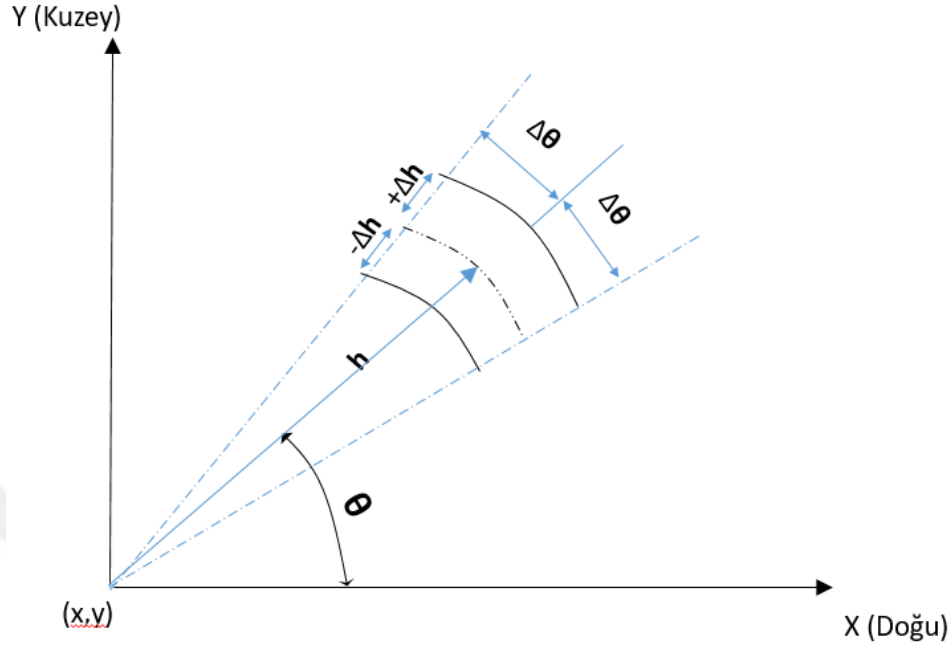
Tipik bir variogram fonksiyonu Şekil 1'de gösterilmiştir. Külçe etki değeri variogramın en önemli parametresidir. Bu parametre, h değeri 0'a yakınsarken değişkenin kısa mesafelerdeki değişkenliğinin bir ölçüsünü verir. Başka bir deneysel variogram

parametresi ise “*yapısal uzaklık (range)*” tır. Yapısal uzaklık, bir örneğin diğeri üzerinde etkili olduđu uzaklık yani örneğin etki mesafesidir. Eşik değeri ise variogramı hesaplamada kullanılan verilerin varyansına eşittir. Külçe etki değerin vanyansa oranı 1’e yakınsa veriler arasındaki deđişkenlik büyük ölçüde uzaklıktan bağımsızdır.



Şekil 1. Variogram Parametreleri (yapısal uzaklık, külçe etkisi ve eşik değeri)

Variogram yönün ve uzaklığın bir fonksiyonudur. Veriler arasındaki açıyı dikkate almadan yapılan variogram hesaplaması yön bağımsız (omnidirectional) olarak adlandırılır ve variogram değerlerinin hesaplamasında ilgili bölge içinde varolan ve bütün yönlerde birbirlerine h uzaklıkta yer alan tüm örnek çiftleri variogram değerinin hesaplamasında kullanılır. Variogramlar θ yönü ve h uzaklıkları birlikte dikkate alınarak hesaplanırsa yön bağımlı variogram olarak adlandırılır. θ , X ekseninden itibaren saat ibresinin tersi yönünde oluşan açıdır (Şekil 2). Örneğin $\theta = 0^\circ$, $\theta = 90^\circ$ yada $\theta = 45^\circ$ için variogramlar oluşturulabilir. $\Delta\theta$ açı toleransını göstermek üzere $(\theta - \Delta\theta < \theta < \theta + \Delta\theta)$ tanımlanabilir. Ancak pratikte deneysel variogramın hesaplanması aşamasında veri çiftlerinin her zaman bu şekilde birbirinden aynı uzaklıkta olması beklenemez. Bu durumda Eşitlik 2.1’in kullanımında yeni bazı kabuller yapılmalıdır. h artık kesin bir uzunluk değil bir aralık içerisinde $(h - \Delta h < h < h + \Delta h)$ olan değerler kümesinin ortalaması olarak düşünülür. Uzunluk toleransı olarak genellikle ilk seçilen h (lag mesafesi, gecikme uzaklığı) uzunluğunun yarısı düşünülür ($\Delta h = h/2$) [15].



Şekil 2. Variogramda yön ve uzaklık (θ adım uzaklığı yönünü, h adım uzaklığını, $\Delta\theta$ tolerans açısını, Δh tolerans mesafesini göstermektedir)

Deneysel yarı varyogramlarda varyogram değerleri belirli uzaklıklar için hesaplanır ve bunun dışındaki uzaklıklarda varyogram değerleri bilinmez. Bölgeselleşmiş değişkenin özelliklerinin belirlenmesinde ve özellikle örneklenmemiş noktadaki değerlerin kestiriminde, varyogramı bütün uzaklık değerleri ile birlikte bilmek gerekir. Bu durum varyogram modellemeyi yani deneysel varyogram değerlerine bir fonksiyon uyarlamayı gerektirir [16]. Çizelge 1'de sık kullanılan varyogram modelleri ve özellikleri gösterilmiştir.

2.2.2. Ortalamasız (Ordinary) Krigleme yöntemi

Teorik olarak bölgesel değişkenler sonsuz sayıdaki lokasyonda örneklenebilir. Ancak pratikte maliyete ilişkin sebeplerden dolayı bu mümkün değildir. Eğer veri olmayan lokasyonlardaki değerler bilinmek isteniyorsa, mevcut veriler kullanılarak bu değerler kestirilebilir [17].

Krigleme yöntemi, en iyi yansız doğrusal kestirim yapan bir yöntemdir. Burada en iyi kelimesi minimum kestirim varyansını ifade etmektedir.

Çizelge 1. Variogram Modelleri

Variogram Modeli Tipi	Variogram fonksiyonu	Variogram Şekli
Doğrusal Model	$h = 0 \text{ ise } \gamma(h) = 0$ $h > 0 \text{ ise } \gamma(h) = p \cdot h$	
Küresel Model	$h = 0 \text{ ise } \gamma(h) = 0$ $0 \leq h \leq a \text{ ise } \gamma(h) = C \left\{ \frac{3h}{2a} - \frac{h^3}{2a^3} \right\}$ $h > a \text{ ise } \gamma(h) = C$	
Üssel Model	$h = 0 \text{ ise } \gamma(h) = 0$ $\gamma(h) = C \{1 - \exp(-h/a)\}$	
Gauss Modeli	$h = 0 \text{ ise } \gamma(h) = 0$ $\gamma(h) = C \{1 - \exp(-h^2/a^2)\}$	
Yuvalı Yapı Modeli	<p>2 veya daha fazla modelin birleşimidir. Örnek: külçe etkisi modeliyle küresel modelin birleşimi</p> $h = 0 \text{ ise } \gamma(h) = 0$ $0 \leq h \leq a \text{ ise } \gamma(h) = C_0 + C \left\{ \frac{3h}{2a} - \frac{h^3}{2a^3} \right\}$ $h > a \text{ ise } \gamma(h) = C_0 + C$	

Krigleme yöntemi ile yapılacak bir kestirimin doğruluğu temel olarak aşağıdaki faktörlere bağlıdır:

1. Örnek sayısı ve her bir noktadaki veri kalitesi.
2. Örneklerin, maden yatağında sahip oldukları konum: Düzgün bir şekilde yayılmış örnekler kümeleşmiş verilere nazaran yatağı daha doğru temsil ederler.
3. Örnekler ve kestirim noktaları arasındaki mesafe: Kestirimde, komşu örneklerden yakın olanın uzak olana göre daha etkin olması doğaldır. Benzer şekilde, kestirimin örneklerin civarında daha doğru olması, örneklerden uzaklaşıldıkça kötüleşmesi beklenir.
4. Değişkenlerin uzaklığa bağlı süreklilikleri: Düzenli bir değişim gösteren değişkenlerin, ani bir şekilde değişen değişkenlere kıyasla kestirimleri daha kolaydır.

Krigleme yöntemi tüm bu faktörleri hesaba katan bir kestirim yöntemidir [18].

İki temel krigleme kestirim yöntemi; değişkenlere ilişkin ortalamının bilinmediğini varsayan Ortalamasız (ordinary) Krigleme yöntemi ve ortalamının bilindiğini varsayan Ortalamalı (simple) Krigleme yöntemidir. Tez kapsamında Ortalamasız Krigleme yöntemi kullanılmıştır.

Yöntemin amacı, bir ya da daha fazla örneklenmemiş nokta veya bloğa ilişkin rastlantısal bir değişkenin değerini, örnek verilerden yararlanarak kestirmektir. x_0 noktasındaki kestirim değeri,

$$Z^*(x_0) = \sum_{i=1}^N \lambda_i z(x_i) \quad (2.2)$$

ile hesaplanır. Eş. 2.2'de λ_i ağırlıkları kestirim hatalarının ortalaması sıfır ve varyansı minimum olacak şekilde belirlenir:

1. $\text{Ort}[Z^*(x_0) - Z(x_0)] = 0$
2. $\text{Var}[Z^*(x_0) - Z(x_0)]$ minimumdur [18].

Bu eşitliklerde;

$Z(x_0)$: x_0 noktasındaki Z rastlantısal değişkeninin gerçek değerini,

$Z(x_i)$: x_0 noktasındaki Z rastlantısal değişkenini kestirmek için kullanılacak örnek verileri.

$Z^*(x_0)$: x_0 noktasındaki Z rastlantısal değişkeninin kestirilen değerini,

λ_i : Kestirimde kullanılacak $Z(x_i)$ değerinin ağırlığını,

N : Kestirimde kullanılacak veri sayısını

göstermektedir.

Eş. 2.2'de, $Z^*(x_0)$ değerinin kestiriminin verilerin ağırlıklı ortalamaları kullanılarak hesap edildiği görülmektedir. Kestirimin yansız olması için λ_i ağırlıklarının toplamı 1'e eşit olmalıdır.

$$\sum_{i=1}^N \lambda_i = 1 \quad (2.3)$$

Kestirim hata varyansının hesaplanması için Eş. 2.4 kullanılır.

$$\begin{aligned} \text{Var}[Z^*(x_0) - Z(x_0)] &= E\{[Z^*(x_0) - Z(x_0)]^2\} \\ &= 2 \sum_{i=1}^N \lambda_i \gamma(x_i, x_0) - \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j \gamma(x_i, x_j) \end{aligned} \quad (2.4)$$

Eş. 2.4'te;

$\gamma(x_i, x_j)$: Z rastlantı değişkeninin x_i ve x_j veri noktaları arasındaki yarı variogramı,

$\gamma(x_i, x_0)$: x_i ve kestirim yapılacak nokta arasındaki yarı variogramı göstermektedir.

Noktasal veriler için Ortalamasız Krigleme yönteminin nokta Krigleme yöntemi eşitliği matris formunda Eş. 2.5'teki gibi gösterilebilir.

$$\begin{matrix} & & & & \mathbf{A} & \mathbf{x} & \boldsymbol{\lambda} & = & \mathbf{b} \\ \left[\begin{array}{cccc} \gamma(x_1, x_1) & \gamma(x_1, x_2) & \cdots & \gamma(x_1, x_N) \\ \gamma(x_2, x_1) & \gamma(x_2, x_2) & \cdots & \gamma(x_2, x_N) \\ \vdots & \vdots & \cdots & \vdots \\ \gamma(x_N, x_1) & \gamma(x_N, x_2) & \cdots & \gamma(x_N, x_N) \end{array} \right] & \left[\begin{array}{c} \mathbf{1} \\ \mathbf{1} \\ \vdots \\ \mathbf{1} \\ \mathbf{0} \end{array} \right] & \left[\begin{array}{c} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_i \end{array} \right] & = & \left[\begin{array}{c} \gamma(x_1, x_0) \\ \gamma(x_2, x_0) \\ \vdots \\ \gamma(x_N, x_0) \\ \mathbf{1} \end{array} \right] \end{matrix} \quad (2.5)$$

A matrisinin tersi ile b vektörü çarpımından ağırlıkları (λ) ve Lagrange Çarpanını (Ψ) bulmak mümkün olacaktır (Eş.2.6). Krigleme varyansı ise, Eş. 2.7 ile hesaplanır [17].

$$\boldsymbol{\lambda} = \mathbf{A}^{-1} * \mathbf{b} \quad (2.6)$$

$$\sigma^2(x_0) = \mathbf{b}^T * \boldsymbol{\lambda} + \mu(x_0) \quad (2.7)$$

Eş. 2.7'de $\sigma^2(x_0)$, x_0 noktası için yapılan krigleme varyansını; $\mu(x_0)$ ise global ortalamayı göstermektedir.

2.2.3. İndikatör Krigleme yöntemi

İndikatör Krigleme yöntemi ilk olarak Journel [19] tarafından ortaya atılmış olan bir tür parametrik olmayan koşullu Ortalamasız Krigleme yöntemidir ve genellikle bölgesel değişkenlerin örneklenmemiş lokasyonlarına ilişkin koşullu birikimli dağılım fonksiyonunun kestiriminde uygulanmaktadır. İndikatör Krigleme yönteminin uygulanabilmesi için bölgesel değişkenlere yönelik ikinci derece durağanlık veya normal dağılım gösterme gibi gereksinimler olmadığından bu yöntem geniş çaplı bir alanda kullanılabilir [20].

Çok metalli bir maden yatağında cevherin barit, diyorit bazalt gibi litolojilerle tanımlanmış olduğunu varsayalım. Eğer z , bu litolojik sınıfları gösterirse x lokasyonundaki indikatör fonksiyon Eş.2.8'deki gibi tanımlanabilir.

$$I(x, Y) = \begin{cases} 1 & y(x) \in Y \\ 0 & y(x) \notin Y \end{cases} \quad (2.8)$$

Burada $y(x)$ barit, diyorit, bazalt gibi litolojik sınıfları ve $y(x) \in Y$ ise ilgili litolojik sınıflardan birine ait olma durumunu ifade etmektedir. $I(x, Y)$ indikatör fonksiyonu bütün litolojik sınıflar için belirlenir. Y 'nin verilen koşullarında $y(x) \in Y$ olasılığı olan $F(y)$ ve indikatör fonksiyonun beklenen değeri olan $E\{I(x, Y)\}$ sırasıyla Eş.2.9 ve Eş.2.10 ile ifade edilir.

$$F(x, y) = Prob[I(x, Y) = 1] \quad (2.9)$$

$$E\{I(x, Y)\} = 1x F(x, y) + 0x [1 - F(x, y)] = F(x, y) \quad (2.10)$$

Beklenen değer eşitliği Eş. 2.10, bir D sahasında bağımsız değişkenlere ilişkin bir veri kümesinin litolojik sınıflardan birine ait olma olasılığının indikatör değişkenlerin ortalamasına eşit olduğunu göstermektedir. Eğer $I(x+h, Y)$ ve $I(x, Y)$ birbirlerinden h uzaklığıyla ayrılmış rastlantısal indikatör değişkenlerse, indikatör yarı variogramı $\gamma_I(h, Y)$, Eş. 2.11 ile tanımlanabilir:

$$\gamma_I(h, Y) = \frac{1}{2} E\{[I(x+h, Y) - I(x, Y)]^2\} \quad (2.11)$$

Eş. 2.11'e göre $\gamma_I(h, Y)$ 'nin h adım uzaklığına göre dağılım grafiği çizilebilir ve bu grafiğe küresel ve üssel teorik yarı-variogramlar modellenenir. Ortalamasız Krigleme yöntemi kullanılarak $I(x, Z)$ kestirilerek örneklenmemiş x noktasında bir litolojik sınıfın var olma olasılığı Eş. 2.12 ile hesaplanabilir:

$$F^*(x, Y) = \sum_{k=1}^N \lambda_k I(x_k, Y) \quad (2.12)$$

Burada, $F^*(x, Y)$ kestirilen belirli litolojik sınıfın var olma olasılığını, N ise kestirim yapılan D sahasında kestirim için kullanılan komşu örnek sayısını, x_k ise komşu örneklerden k 'inci örneği ve λ_k ise en iyi doğrusal yansızlık koşuluyla indikatör yarı variogram modeli kullanılarak o örneğin kestirim için kullanılacak ağırlığını ifade etmektedir [20].

2.3. Rastlantı Ormanı Metodu ile Sınıflandırma

Rastlantı Ormanları (RO) metodu, Bagging yönteminin geliştirilmesiyle oluşturulmuş olup temeli, bir makine öğrenme yöntemi olan Sınıflandırma ve Regresyon Ağaçları tekniğine (SRAT'a) dayanır. Yöntem, oluşturduğu ikili dallanmalara sahip ağaç yapılarıyla, veri kümesindeki bağımsız değişkenler ile bağımlı değişkenler arasındaki ilişkiyi ortaya koymayı amaçlar. RO'da, SRAT'den farklı olarak, sadece orijinal veri kümesinden ağaç yapısının oluşturulması yerine veri kümesinden Bootstrap yöntemi ile çekilen alt veri kümelerinin her biri için ağaç yapıları oluşturulur. Metodun adında orman kavramının geçiyor olması birden fazla ağaç yapısının oluşturulmasından ileri gelmektedir.

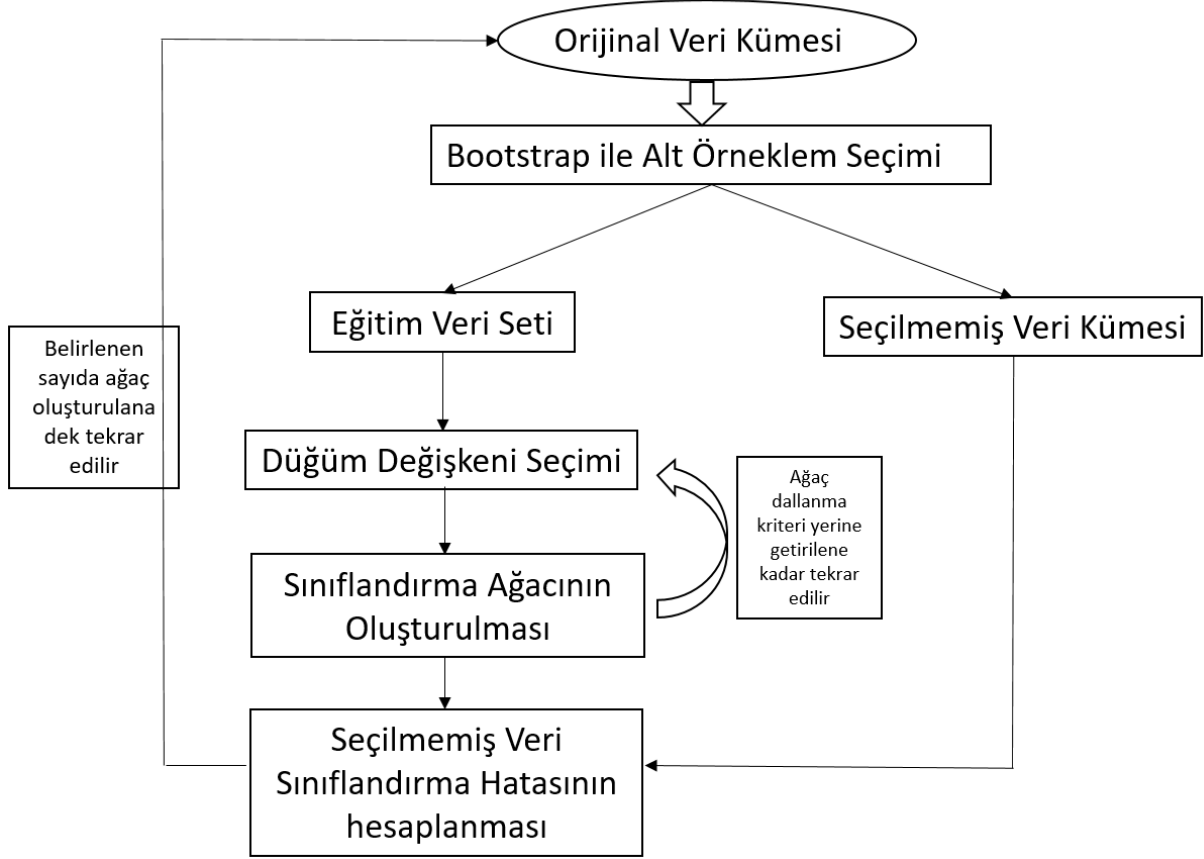
Bagging yöntemi Breiman [21] tarafından 1996 yılında geliştirilmiştir. Dietterich [22] ve Ho'nun [23] rastlantısal düğüm değişkeni seçimine yönelik çalışmalarının akabinde, 2001 yılında Breiman [24], Bagging yöntemi ile rastlantısal düğüm değişken seçiminin bileşiminden RO'yu geliştirmiştir.

Rastlantı Ormanı metodu aşağıdaki adımları içermektedir:

- 1) Orijinal veri kümesinden Bootstrap yöntemi ile rastlantısal alt veri kümesinin çekilmesi
- 2) Her bir rastlantısal alt veri kümesinin sınıflandırma ağaçlarının oluşturulması
- 3) Bootstrap yöntemi ile orijinal veri kümesi içinden seçilmemiş verilerin (out of bag samples) belirlenmesi, oluşturulan sınıflandırma ağaçları ile seçilmemiş verilerin sınıflandırılması ve sınıflandırma hatalarının hesaplanması
- 4) Belirlenen sayıda ağaç üretilerek orman yapısının oluşturulması

RO metodunun genel çalışma prensibi Şekil 3'te gösterilmiştir. Metodun kavranabilmesi için öncelikle Sınıflandırma ve Regresyon Ağaçları, Bootstrap, Bagging

yöntemi ve rastlantısal düğüm değişkeni seçimi konularının ele alınması gerekmektedir. Ayrıca bu yöntemin önemli bir argümanı olan seçilmemiş veri sınıflandırılması ve sınıflandırma hatasının incelenmesi gerekir.



Şekil 3. Rastlantı Ormanları Algoritması [25]

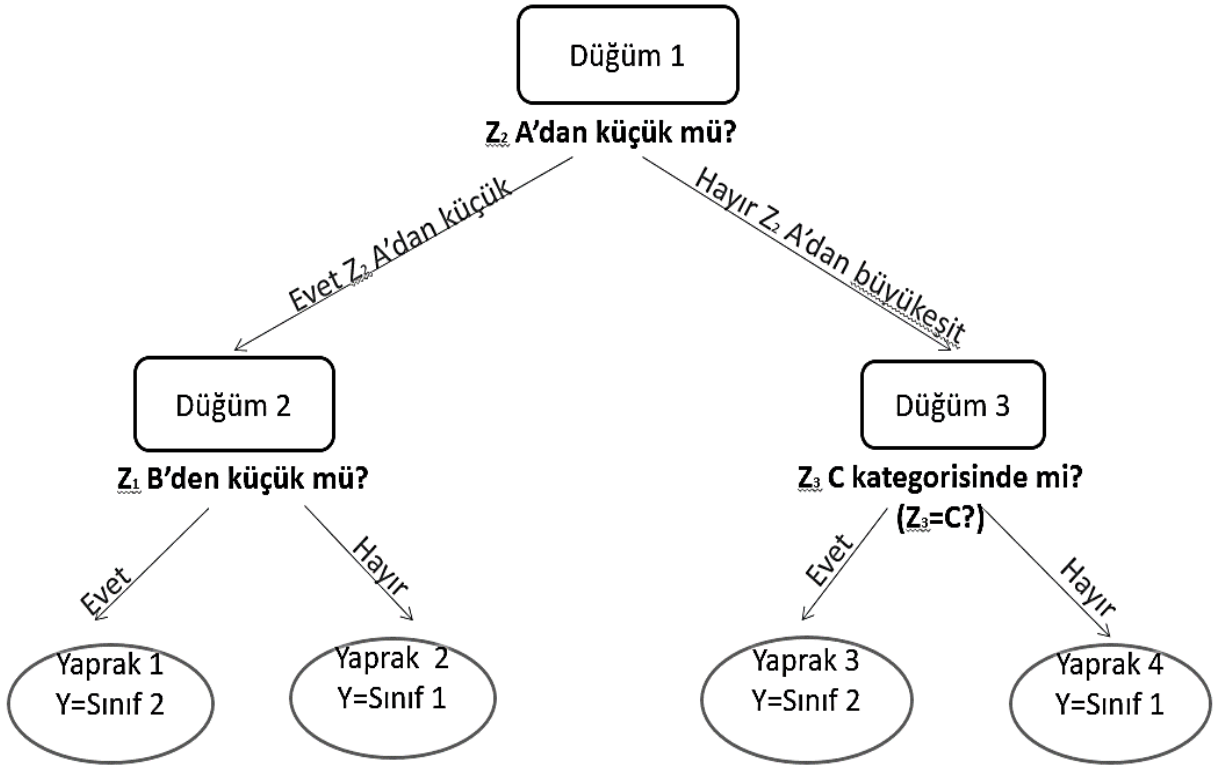
2.3.1. Sınıflandırma ve Regresyon Ağaçları (SRAT), Gini İndeksi ve Budama İşlemi

N adet gözlemden oluşan bir eğitim veri kümesinin (T) kategorik sınıf değişkeni Y'nin 1, 2, ..., K değerlerini aldığını ve bu sınıf değerleriyle ilişkili olan P adet X_1, X_2, \dots, X_P bağımsız değişken içerdiğini varsayalım. Sınıflandırmada amaç, örneklem dışında elde edilen $Z(x)$ vektörü değerlerine karşılık gelen sınıf değeri $Y(x)$ 'i kestirmektir. Teoride bu problemin çözümü Z veri kümesini k_1, k_2, \dots, k_K şeklinde K adet ayrık kümeye; $j=1, 2, \dots, K$ olmak üzere Z vektörü k_j kümesi içerisinde yer aldığı anda Y'nin j olarak tanımlanabileceği şekilde bölmektir [26].

SRAT ile veri kümesindeki bağımsız değişkenlere sorular sorularak ana düğümden her seferinde iki ayrı dala bölünmek suretiyle yapraklara kadar uzanan bir ağaç modeli

oluşturulur. Yapraklara ulaşıldığında veri kümesinde bağımsız değişkenler ile bağımlı değişkenin (sınıfların) ilişkisi ortaya konulmuş olur.

Bahsedilen sorular eğer değişken süreklirse ilgili değişkene ilişkin belirli bir sınır değerden küçük mü, büyük eşit mi olduğu şeklinde olabileceği gibi eğer değişken kategorikse ilgili kategoriye ait olup olmadığı şeklinde de olabilir. Optimal sınıflandırmayı sağlayan soruya evet cevabı veren veriler birinci dal, hayır cevabı veren veriler ikinci dal olarak ayrılır.



Şekil 4. Basitleştirilmiş bir sınıflandırma ağacı örneği.

Şekil 4; Z_1 , Z_2 , Z_3 'ün bağımsız değişkenleri, Y 'nin ise sınıf 1 ve sınıf 2 değerleri alabilen kategorik bağımlı değişkeni temsil ettiği örnek bir veri kümesiyle sınıflandırma ağacının çalışma prensibini basit bir biçimde açıklamaktadır. Z_1 ve Z_2 bağımsız değişkenleri sürekli değişken oldukları için sırasıyla A ve B değerlerinden küçük olup olmadıkları sorulmuş, Z_3 ise kategorik değişken olduğu için C kategorisine ait olup olmadığı sorulmuştur. Ağaçtaki her bir düğüm, veri kümesini en iyi şekilde sınıflandıracak sorunun ilgili bağımsız değişkene sorulmasına karşılık gelmektedir. Örneğin Şekil 4'ün ilk düğümünden (kök düğüm) anlaşılacağı üzere veri kümesinde Z_2 bağımsız

değişkeninin A' dan küçük mü büyük eşit mi olduğu sorusu ilk dallanmada dallar arasındaki sınıflandırma kontrastını en belirgin hale getiren soru olarak belirlenmiştir. Böylece Düğüm 1 'de veri kümesinde Z_2 'nin A değerinden küçük olduğu bütün veriler sol dala, büyük eşit olduğu veriler sağ dala ayrılmış ve daha sonra sınıflandırmayı optimize eden sorular sorulmaya devam edilerek yapraklara kadar ulaşılmıştır. Yapraklarda bağımlı değişkenin (Y) aldığı sınıf değerlerinin tespit edilmiş olduğu Şekil 4'te görülmektedir.

Bu noktada, sınıflandırma ağacı yapılandırılırken hangi bağımsız değişkene hangi soru sorulduğunda sınıflandırmanın en uygun şekilde yapılacağı belirlenmesinin büyük bir önem taşıdığı görülmektedir. SRAT'de bu problemin çözümü için farklı yöntemler uygulanabilmekte olup yaygın olarak başvuru Gini indeksinin hesaplanması tez kapsamında da kullanılan yöntem olmuştur.

Breiman vd. [27], düğüm değişkeni ve değişken eşik değeri seçimi için Gini indeksi olarak adlandırılan bir ölçüm kullanmışlardır. Gini fonksiyonu sınıflara ilişkin özniteliklerin safsızlığını ölçer.

T veri kümesinin k sayıda sınıf değerinin Gini İndeksi; Gini (T), Eş. 2.13 ile tanımlanabilir:

$$\text{Gini (T)} = 1 - \sum_{i=1}^k f_i^2 \quad (2.13)$$

Eşitlikte f_i , T veri kümesindeki i sınıfının göreceli sıklığıdır. N adet veri içeren T veri kümesi, $N_{\text{sağ}}$ adet veri içeren $T_{\text{sağ}}$ ve N_{sol} adet veri içeren T_{sol} alt veri kümelerine ayrıldığında söz konusu ayırım için gini indeksi, $\text{Gini}_{\text{ayırım}}(T)$, Eş. 2.14 ile hesaplanır:

$$\text{Gini}_{\text{ayırım}}(T) = \frac{N_{\text{sağ}}}{N} \text{Gini}(T_{\text{sağ}}) + \frac{N_{\text{sol}}}{N} \text{Gini}(T_{\text{sol}}) \quad (2.14)$$

Eş. 2.14, veri kümesindeki bütün bağımsız değişkenlerin alabileceği bütün değerler için hesaplanır. Bu hesaplamalar sonucunda $\text{Gini}_{\text{ayırım}}$ değerini minimum yapan bağımsız değişken ve onun eşik değeri ayırım için belirlenen düğüm değişkeni ve eşik değeri olur. Bu değer baz alınarak ağaçta dallanma işlemi gerçekleştirir ve söz konusu işlemler sınıf değerlerinin tespit edildiği yapraklara ulaşıncaya kadar devam ettirilir. Bu aşamada büyük ağaç yapılarının oluşması ve aşırı dallanmanın sınıflandırma üzerinde olumsuz etkileri söz konusu olabilmektedir.

Sınıflandırma ağaçları sınıflar arasındaki ayrımı yüzde yüz oranda sağlayacak şekilde dallandırıldığında çoğu zaman aşırı dallanma yapısı sergileyen ağaç modelleri oluşmaktadır. Aşırı dallanmalar; özellikle veri sayısının yüksek olduğu örneklerde, verilerdeki muhtemel kirliliğin etkilerini, oluşturulan sınıflandırma ağacına taşımakta ya da ihmal edilebilir bilgi sağlamasına rağmen gereksiz dallanmalarla fazladan hesaplama yapılmasına neden olmaktadır. Bu durum optimal ağaç boyutunun belirlenmesini gerekli kılmaktadır. Budama (Pruning) adı verilen yöntem bu amaçla geliştirilmiştir. Budama işlemi iki farklı yaklaşımla gerçekleştirilebilmektedir. Bunlardan ilki; yüzde yüz saflıkta sınıflar oluşturulana dek ağaç yapılandırıldıktan sonra sınıflandırma üzerinde etkinliği düşük olan düğümlerin yaprak düğümünden başlayarak budanması şeklindedir. İkinci yaklaşımda ise ağaç yapısı henüz oluşturulurken bir alt dallanmaya geçilip geçilmeyeceğine yönelik karar, oluşturulacak yeni dallanmanın sınıflandırma üzerindeki etkisinin ihmal edilebilir olup olmadığı hususu gözetilerek verilir. Budama yapılıp yapılmayacağına ilişkin karar verilme sürecinde kullanılan farklı yöntemler mevcuttur. Ağaç kökünden başlanarak maksimum dallanma sınırının belirlenmesi, herhangi bir dalın yaprak olarak etiketlenmesini gerektirecek minimum veri sayısının belirlenmesi, yapraktan başlayarak her hangi bir düğümün kaldırılmasının kestirim doğruluğuna etkisinin belirlenmesi gibi yaklaşımlar budama kararı verilmesinde kullanılmaktadır.

Budama yöntemi uygulanmadan aşırı dallanmaların olumsuz etkilerinin bertaraf edildiği Bagging yöntemi (Bölüm 2.3.3'te açıklanmıştır) RastlantıOrmanlarının temel adımlarından biri olduğundan RO'da budama işlemine ihtiyaç duyulmaz.

2.3.2. Bootstrap yöntemi

Bootstrap yöntemi, var olan bir örneklemden alt örneklem kümeleri oluşturmayı amaçlar [28]. Kısaca Bootstrap, bağımlı ve bağımsız değişkenlerden oluşan örnekleme bir popülasyon olarak değerlendirir ve alt örneklemi gerçek verilerden seçer. Belirli bir dağılımdan (örneğin normal dağılım) rastlantısal şekilde veri çekmek yerine örneklem içerisinden yerine koymak suretiyle veri çeker. Bu sebeple deneysel dağılım fonksiyonunu gerçek dağılım fonksiyonu olarak ele alır [29].

Bootstrap yöntemi ile alt örneklem, $\{(Z_1, Y_1), \dots, (Z_N, Y_N)\}$ popülasyonundan N sayıda örneklem yerine koyma şartıyla çekilir. Popülasyonu temsil eden N sayıdaki orijinal örneklemden yine N sayıda örneklem çekilmesi işlemi parametrik olmayan Bootstrap yöntemi olarak adlandırılır ve bu yöntem ilk olarak 1979 yılında Efron [30] tarafından ortaya atılmıştır [28].

Y bağımlı değişkenine karşılık gelen P sayıda Z bağımsız değişkeni içeren N sayıda gözlem değerimiz olsun. $i=1, \dots, N$ olmak üzere, N adet $Z_{(1, \dots, P)i}$, Y_i değeri bağımsız olarak orijinal örneklemden yerine konulmak şartı ile çekildiğinde bir bootstrap örneklem kümesi oluşturulmuş olur. Her ne kadar bazı bağımlı-bağımsız değişken çiftleri tekrarlı şekilde kullanılsa ve bazıları hiç kullanılsa da, oluşturulan bootstrap örneği orijinal örneklem ile aynı sayıda veri içerir. Çoklu sayıda (B) bootstrap örneği oluşturulduğunda b'inci örneklem kümesindeki bağımlı ve bağımsız değişkenler Z_b, Y_b olarak ifade edilirse Z_b , N x P boyutlu matris ve Y_b ise N boyutlu bir sütun vektörü olacaktır [29].

Tercan [31], Bootstrap yöntemi ile, ortalaması kestirilmiş bir veri kümesinden örneklem dağılımının elde edilmesine yönelik basitleştirilmiş bir örnek vermiştir. Örnek kapsamında yapay olarak oluşturulan veri kümesi Çizelge 2'de verilmiş olup bu veri kümesinin bir maden yatağına ilişkin kalınlık değerlerine karşılık geldiği varsayılabilir. Söz konusu veri kümesi [2, 5, 8, 9, 7, 6, 3, 10, 4] değerlerini alan 9 adet veri içermektedir. Bootstrap işleminin ilk adımında veri kümesinden yerine koymak şartıyla ve rastlantısal bir şekilde orijinal veri sayısı kadar veri çekilir. Veri çekme, bir veri topluluğu içinden bir sayı (değer) çekme işlemidir. Çizelge 2'deki ikinci kolon, çekilen 9 veriyi göstermektedir. Bu veriler, birinci bootstrap örneğini oluşturmaktadır.

Çizelge 2'de orijinal veri kümesinden rastgele seçilen verilerle oluşturulan 3 adet örneklem incelendiğinde, orijinal veri kümesinde yer alan verilerden bazılarının hiç kullanılmadığı, bazılarının ise tekrar edecek şekilde kullanıldığı görülebilir. Bu durum, yukarıda da bahsedildiği üzere Bootstrap yönteminde veri çekme işleminin yerine koyarak ve rastgele yapılmasından ileri gelmektedir.

Çizelge 2. Bootstrap ile Örnekleme [31]

Mevcut Veriler	Bootstrap Örneklemleri		
	1.Örneklem Kümesi	2. Örneklem Kümesi	3. Örneklem Kümesi
2	9	3	8
5	8	5	7
8	6	5	10
9	9	7	7
7	3	8	5
6	7	2	6
3	7	5	4
10	4	6	4
4	2	6	9

2.3.3. Bagging yöntemi

Bagging yöntemi, Breiman [21] tarafından geliştirilen sınıf kestirimine ilişkin varyansı düşürerek kestirim kalitesini artıran ve bir çok sınıflandırma yöntemiyle birlikte kullanılabilen bir yöntemdir [32]. “Bagging”, kelime olarak “Boostrapping” ve “Aggregation” kelimelerinin bileşiminden oluşur. Dietterich [33] ile Dudoit ve Fridly [34] yaptıkları çalışmalarda Bagging yönteminin özellikle aşırı dallanmanın söz konusu olduğu örneklerde iyi bir örnekleme metodu olduğunu ortaya koymuşlardır.

Bagging yöntemi her bir ağaç yapısını farklı bir eğitim veri kümesini ($\mathcal{D}^{(B)}$) kullanarak oluşturur. Her bir eğitim veri kümesi orijinal eğitim veri kümesinden (\mathcal{D}) Bootstrap yöntemi ile tekrarlı bir şekilde çekilir. Başka bir deyişle orijinal veri kümesinden yeni eğitim veri kümesi çekilmesi işlemi çekilen verilerin yerine konulması şartı ile rastlantısal şekilde yapılır. Rastlantısal olarak çekilen her bir veri kümesindeki veri sayısı (N) orijinal veri kümesindeki veri sayısı ile aynı olacak şekilde belirlenir. Her ne kadar aynı veri sayısına sahip olsa da Bootstrap yöntemiyle rastlantısal şekilde çekilen her bir verinin yerine koyulmasından dolayı rastlantısal çekilen veri kümesinin orijinal veri kümesindeki örneklerin yaklaşık %63'ünü içereceği Breiman [21] tarafından ortaya konulmuştur. Seçilmemiş olan yaklaşık %37'lik örnek ise oluşturulan ağaç yapısının test edilmesi için seçilmemiş veri kümesi (out of bag sample) olarak ayrılır. N sayısının %37'si kadar verinin seçilmemiş veri kümesi olarak ayrılmış olması aynı zamanda

eđitim veri kümesinde N sayısının %37'si kadar kendisini tekrar eden örneđin olduđu anlamına gelir [35].

$\mathcal{E}(\vec{Z}, \mathcal{D}^{(B)})$ şeklinde tanımlanan bir sınıf kestirim algoritması ile bağımlı kategorik sınıf deđişkenine karşılık gelen $\mathcal{E}(\vec{Z})$ 'in, bağımsız deđişkenlere karşılık gelen \vec{Z} vektörü ile hesaplanması Eş. 2.15 ile yapılır.

$$\mathcal{E}(\vec{Z}) = \text{oylama } \mathcal{E}(\vec{Z}, \mathcal{D}^{(B)}) \quad (2.15)$$

Breiman [21], Bagging yöntemi ile yapılan sınıflandırmaların, orijinal örneklem kullanılarak yapılan kestirime göre daha düşük kestirim hatasına sahip olduğunu ortaya koymuştur. Dietterich [33], Bagging yönteminin özellikle sınıflandırma kirliliđinin olduđu örneklemlerde oldukça iyi bir örnekleme yöntemi olduğunu ortaya koymuştur.

Örnekleme, rastlantı ormanları yönteminin özünü oluşturan ve rastlantısallığın ilk olarak sağlandıđı aşamadır. Yapılan işlemin doğruluđuna ve genel hatasına yönelik bilgi sağlıyor olması Bagging yönteminin en önemli avantajıdır. [35]

2.3.4. Rastlantısal Düğüm Deđişkeni Seçimi

Bir ağaç topluluđunun kestirim hatasını etkileyen iki temel faktörden bahsedilebilir: Birincisi, her bir ağaç yapısının doğruluđu; ikincisi ise her bir ağaç yapısının bir birinden farklılıđıdır. Birinci faktör bütün orman yapısında ortaya çıkabilecek kestirim yanlılıđıyla, ikinci faktör ise ormanın kestirim varyansı ile ilişkilidir. Bagging yöntemi, birbirlerinden farklı ağaç yapılarının üretilmesini, orijinal örneklem popülasyonunu kullanarak yerine koyma metodu ile alt örneklem kümeleri oluşturarak gerçekleştirdiđinden düşük kestirim varyansı sağlar ve aşırı dallanmanın kestirim üzerindeki olumsuz etkisini yumuşatır (overfitting etkisini yumuşatır). Bagging yönteminden ayrı olarak farklı ağaç yapılarının oluşturulmasına yönelik yollar mevcuttur. Ho [23], Amit ve Geman [36] veri kümesi ile birlikte ağaç yapılandırılması esnasında da rastlantısallaştırmanın kestirim hataları açısından olumlu sonuç verdiđini göstermişlerdir. Ağaç yapılandırması esnasındaki rastlantısallaştırmalar; düğüm deđişkenlerinin her bir farklı ağaçta ya da aynı ağacın farklı düğüm noktalarında rastlantısal şekilde seçilmesi, eşik deđerlerinin rastlantısal şekilde seçilmesi gibi farklı yöntemlerle yapılabilmektedir. Daha önce yapılan çalışmalar, her bir ağaç yapısında rastlantısal şekilde seçilen düğüm

değişkenlerini kullanmak yerine aynı ağaç içindeki her bir düğümde düğüm değişkeninin rastlantısal şekilde seçilmesinin daha iyi sonuçlar verdiğini göstermektedir.

Rastgele seçilecek düğüm değişkeni sayısının belirlenmesinde farklı metotlar kullanılmaktadır. Sınıf kestiriminde kullanılacak toplam bağımsız değişken sayısı n olarak ifade edilirse, ağaç yapılandırılması sürecinde her bir düğümde \sqrt{n} sayıda düğüm değişkeni adayı rastlantısal şekilde belirlenerek bunlar arasından düğüm değişkeninin hesaplanması sıklıkla başvurulan metottur. Rastlantısal seçilen düğüm değişkeni adaylarının alabilecekleri bütün değerler için bölüm 2.3.1'de belirtildiği şekilde Gini indeksinin hesaplanması suretiyle optimum dallanmayı sağladığı tespit edilen düğüm değişkeni ve bu değişkene ilişkin eşik değer belirlenerek ağacın yapılandırılmasına devam edilir.

2.3.5. Rastlantı Ormanı Algoritması

Rastlantı Ormanı yöntemi ile, sınıf değişkeni bilinmeyen bir bağımsız değişken vektörü, rastlantısal şekilde çekilen alt örneklem kümeleri ile oluşturulan sınıflandırma ağaçlarına girdi olarak beslenerek bağımsız değişken vektörünün sınıfı belirlenir. Bu yöntemde, her bir ağaç modeli, sınıf değişkeni içermeyen bağımsız değişken vektörünün sınıfını tayin etmekte bir oy hakkına sahiptir [24]. Orman içerisindeki ağaç sayısı dikkate alınarak en çok oy alan sınıf, bağımsız değişken vektörünün sınıfı olarak kestirilmiş olur.

Rastlantı Ormanı Yöntemi; Sınıflandırma ve Regresyon Ağaçları, Bagging yöntemi ve Rastlantısal Düğüm Değişkeni Seçimi işleminin birleştirilmesi ile elde edilen bir sınıflandırma algoritması olarak da ifade edilebilir [24]. Bu yöntemde hem orijinal veri kümesinden Bagging yöntemiyle rastlantısal alt veri kümeleri oluşturularak hem de her bir ağacın yapılandırılması esnasında dallanma sürecinde rastlantısal düğüm değişkenleri seçilerek birbirlerinden farklı sınıflandırma ağacı yapıları oluşturulmuş olur. Rastgele ağaç yapılarında herhangi bir budama algoritması kullanılmaz [37]. Rastlantı Ormanı yönteminde sınıflandırma ağaçları oluşturulurken dallanma kriterlerinin belirlenmesinde başka bir ifadeyle dallanma için özniteliklerin (düğüm değişkenlerinin) seçiminde Gini İndeksi yöntemi kullanılır [38].

Rastlantı Ormanı yönteminde ağaç yapılarının oluşturulabilmesi için kullanıcı tarafından bazı parametrelerin belirlenmesine ihtiyaç duyulur. Bu parametreler; düğüm değişkeni seçiminin her bir dallanmada kaç adet bağımsız değişken arasından yapılacağı ve ormanı oluşturulacak ağaç sayısıdır. Her bir düğümde dallanmanın sınıflandırma kontrastını en iyi şekilde sağlamasına olanak sağlayan düğüm değişkeni, bütün bağımsız değişkenler kullanılarak değil kullanıcının sağladığı parametre göz önünde bulundurularak rastlantısal olarak belirlenen bağımsız değişkenler arasından seçilir. Ayrıca kullanıcı tarafından H adet ağaç oluşturulmasının istenildiği varsayılırsa, sınıflandırılacak veriler bu H adet ağacın her birinin sınıflandırma işlemine tabi tutulur. İlgili veri için H adet ağacın sınıflandırmasından en çok oyu alan sınıf verinin sınıfı olarak kestirilmiş olur [39].

E eğitim veri kümesinin; \vec{Z} bağımsız değişken vektörlerinden ve Y kategorik sınıf değerlerinden oluştuğunu varsayalım. Y kategorik sınıf değişkeni $k=1, 2, \dots, K$ olmak üzere K adet kategorik değişken değeri alabilsin. Ayrıca, bağımsız değişken vektörü olan \vec{Z}^F 'ye ait verileri içeren ancak sınıf değişkeni Y^F 'ye ilişkin verilerinin bilinmediği F ise kestirim veri kümesi olsun. Rastlantı Ormanı yöntemi ile F kestirim veri kümesinin sınıf değerlerinin kestirimi aşağıdaki adımlarla gerçekleştirilir:

1) Eğitim veri kümesi E'den B_1, \dots, B_M bootstrap örneklemi çekilir.

2) Her bir bootstrap örnekleme ($B_m, n=1, \dots, N$) için Sınıflandırma ve Regresyon Ağaçları (T_n) üretilerek \vec{Z} ve Y arasındaki ilişki her bir ağaç yapısıyla ortaya konulur. Bu aşamada 2.3.1 bölümünde anlatılan Sınıflandırma ve Regresyon Ağaçlarının oluşturulması sürecinden farklı olarak:

a. Her bir düğüm noktasında düğüm değişkeni ve değişken değeri seçiminin belirlenmesi için 2.3.4 bölümünde belirtildiği şekilde sınıf kestirimi için kullanılacak bağımsız değişkenler içinde rastgele seçilen düğüm değişken adayları kullanılır. P'nin E veri kümesi içindeki bağımsız değişken sayısını gösterdiği düşünülürse rastgele seçilecek düğüm değişkeni adayı sayısı \sqrt{P} olarak belirlenebilir.

b. Üretilen her ağaç yaprak düğüme erişilene kadar geliştirilir. Sınıflandırma ve Regresyon Ağaçları yönteminden farklı olarak her hangi bir budama işlemi yapılmaz.

3) Oluşturulan Rastlantı Ormanı ile, kestirim veri kümesi F 'nin \vec{Z}^F bağımsız değişken vektörü kullanılarak Y^F sınıflarının kestirimi her bir Sınıflandırma ve Regresyon Ağacı'ndan elde edilen sonucun değerlendirilmesi ile sağlanır. $\hat{f}_h(\vec{Z}^F)$ 'nin F veri kümesinin h 'inci ağaç yapısı ile sınıf değişkenlerinin kestirimini gösterdiği düşünülürse bootstrap örnekleme ile oluşturulan H adet ağaçtan F veri kümesinin Rastlantı Ormanı ile kestirilen Y^F sınıfları:

$$I[Oylama_{Y_k^F}]_{k=1}^K = \begin{cases} \mathbf{1} & \text{eğer } \hat{f}_h(\vec{Z}^F) = Y_k \\ \mathbf{0} & \text{eğer } \hat{f}_h(\vec{Z}^F) \neq Y_k \end{cases}$$

olmak üzere;

$$Y^F = \underset{Y_k^F}{\operatorname{argmax}} \left\{ \sum_{h=1}^H I[Oylama_{Y_k^F}] \right\} \quad (2.16)$$

şeklinde hesaplanabilir [40]. Eş. 2.16 kullanıldığında, H adet ağaç yapısından en fazla oyu alan sınıflar F kestirim veri kümesinin sınıfları olarak kestirilmiş olur.

2.3.6. Seçilmemiş Veri Sınıf Kestirimi ve Kestirim Hatası

Genelleştirilmiş hata tespiti, aykırı değerlerin (outlier) tespiti, değişken önemini belirlenmesi gibi olanaklar sağladığından seçilmemiş verilerin kullanılması Rastlantı Ormanları yöntemi ile geliştirilen en ilginç tekniklerden biridir. Daha önce belirtildiği gibi Bagging yöntemiyle seçilen örneklem orijinal verinin yaklaşık 2/3'ünü içerir. Seçilmemiş olanlar ise seçilmemiş veri kümesi (Out Of Bag sample) olarak anılır [35].

\vec{Z} bağımsız değişken vektörlerinden ve $k=1, 2, \dots, K$ olmak üzere K adet kategorik değişken değeri alabilen Y kategorik sınıf değerlerinden oluşan E eğitim veri kümesinden Bootstrap yöntemiyle çekilen E_n örnekleme E 'nin bir alt kümesi olacaktır. E_n örnekleme için oluşturulan Sınıflandırma ve Regresyon Ağacı'nın örneklem içindeki bağımsız değişkenlerle sınıf değişkeni arasındaki ilişkiyi $f_n(\vec{Z})$ modeli ile ortaya koyduğu düşünülürse, bu model E veri kümesinde yer alan ancak E_n örnekleme içinde bulunmayan verilerin (E/E_n) kestirimi için de kullanılabilir ve Bootstrap ile seçilen örneklem dışında kalan bütün örneklem kümeleri dikkate alınarak ortalama kestirim hatası tespit edilebilir [40].

Rastlantı Ormanı yönteminin eğitim veri kümesinden Bootstrap yöntemiyle alt örneklem kümeleri oluşturulması bootstrap'a dayalı hata kestirimini olanaklı kılmaktadır. $f_n(\vec{Z})$ modeli kullanılarak E/E_n 'in kestirilmesi seçilmemiş veri sınıf kestirimi olarak ve kestirimde yapılan hatalar seçilmemiş veri kestirim hatası (Out of bag kestirim hatası) olarak adlandırılır. Bu iki kavrama yönelik hesaplamalar aşağıda belirtilen şekilde yapılır [35]:

- 1) Eğitim veri kümesi olan N adet veri içeren E 'den Bootstrap yöntemi ile $n=1, \dots, N$ olmak üzere E_n alt örnekleme seçilir ve her bir E_n alt örnekleme için T_n ağaç yapısı oluşturulur.
- 2) $t=1, \dots, L$ olmak üzere E_t ; eğitim veri kümesi E 'nin bağımsız değişken vektörü \vec{Z}_t ve bu vektöre karşılık gelen bağımlı değişken (sınıf değeri) değeri Y_t yi içeren herhangi bir veri satırını göstermektedir. Ω_t ; E_t satırını içermeyen E_n alt örneklemlerinin indislerini, $[\Omega_t]$; toplam indis sayısını, $f_{\Omega_t}(\vec{Z}_t)$ ise E_t veri satırının sınıf değerinin t satırını içermeyen seçilmemiş veri örneklemleriyle kestirim modelini gösterebilir. Bu durumda E_t veri satırına ilişkin Y_t sınıfının seçilmemiş veri sınıf kestirimi Y_t^{OOB} Eş. 2.17 ile hesaplanır;

$$Y_t^{OOB} = \underset{Y_k^{OOB}}{\operatorname{argmax}} \left\{ \sum_{v=1}^{[\Omega_t]} I[f_v(\vec{Z}_t) = Y_k^{OOB}] \right\} \quad (2.17)$$

- 3) Bu durumda, seçilmemiş veri kestirim hatası $Error^{OOB}$;

$$Error^{OOB} = \frac{1}{n} \left\{ \sum_{t=1}^L I[Y_t \neq Y_t^{OOB}] \right\} \quad (2.18)$$

'dir.

Eş. 2.17 ve 2.18'de, $I[f_{\Omega_t}(\vec{x}_t) = \mathcal{E}(\vec{x}_t)^{OOB}]$ ve $I[\mathcal{E}(\vec{x}_t) = \mathcal{E}(\vec{x})^{OOB}]$, parantez içinde belirtilen koşul gerçekleştiğinde 1 gerçekleşmediğinde 0 değerini alan indikatör fonksiyonu göstermektedir.

3. HOMOJEN ORTAM MODELLENMESİNE YÖNELİK YENİ GELİŞTİRİLEN YÖNTEMLER

3.1.1. Rastlantı Ormanı + Krigleme yöntemi (ROKY)

ROKY, farklı homojen ortamların bulunduğu bir maden yatağında örneklenmemiş lokasyonların hangi ortama ait olduğunu kestirmek için kullanılabilecek yeni bir yöntemdir.

Bir $D \in \mathbb{R}^3$ yatağının K sayıda ortama ($D \cup D_k$, $k=1, \dots, K$ ve $k \neq j$ ise $D_k \cap D_j = \emptyset$) ayrıldığını varsayalım. Ayrıca D yatağı içinde N adet lokasyonda $[s_i, i=1, \dots, N \in D]$ çok değişkenli verilerin ölçüldüğünü ve bu ölçümlerin P sayıda sürekli bağımsız değişken $Z(s)=[Z_1(s), \dots, Z_P(s)]^t$ ve $Y(s)$ kategorik bağımlı değişkenle ilişkili olduğunu varsayalım. Bu ifadede üstel t ; vektörün evriğini (transposition) göstermektedir. D yatağı, ölçümlerin yapıldığı malzeme (genelde sondaj karotları) hacmine göre çok daha büyük olan M sayıda $[b_j, j=1, \dots, M \in D, M \gg N]$ bloğa bölünsün. $Z(s)$ ve $Y(s)$ bilgisini kullanarak $b_j, j=1, \dots, M$ blokların her birini $D_k, k=1, \dots, K$ ortamlarından birine atamak amacıyla geliştirilen yeni yaklaşımda;

1. $Z(s)$ ve $Y(s)$ arasındaki ilişki $f(Z(s))=Y(s)$, Rastlantı Ormanı Yöntemi ile tanımlanır.
2. M sayıda örneklenmemiş blok lokasyonuna $[x_i, i=1, \dots, M \in D]$ ait P sayıdaki bağımsız değişken $[Z_1(x), \dots, Z_P(x)]$ Ortalamasız Krigleme yöntemi ile kestirilir.
3. M adet bloğun hangi homojen ortama atanacağı, birinci adımda, örneklenmiş lokasyonlar için kurulan $Z(s)$ ve $Y(s)$ ilişkisinden faydalanılarak oluşturulan Rastlantı Ormanına, Ortalamasız Krigleme yöntemiyle kestirilmiş olan $Z(x)^*$ bağımsız değişkenleri beslenerek $f(Z(x)^*)= Y(x)$ ile örneklenmemiş x lokasyonundaki $Y(x)$ sınıf değerleri hesaplanır.

3.1.2. Gruplandırılmış Rastlantı Ormanı + Krigleme yöntemi (g-ROKY)

Krigleme yönteminin düşük değişken değerlerini yüksek, yüksek değerleri ise düşük kestirerek bir törpüleme etkisi (smoothing) gösterdiği bilinmektedir. ROKY yönteminde Rastlantı Ormanı gerçek verilerin bağımsız değişkenleri ile bağımlı sınıf değişkeni arasındaki ilişkiyi ortaya koymakta olup sınıf değerleri hesap edilecek verilerin bağımsız değişkenleri krigleme ile kestirilmektedir. Kriglenmiş bağımsız değişkenler, orijinal verilerle oluşturulan Rastlantı Ormanına beslendiğinde törpüleme etkisi sınıflandırma hatasına sebep olmaktadır. Kriglemeyi takiben gerçek ve kriglenmiş bağımsız değişkenlerin standartlaştırılarak (z-score transformation) gruplandırılmaları, kestirim verilerindeki törpülenmelerin gruplar içinde gizlenmesine olanak sağlayarak sınıflandırma hatasını azaltmaktadır. Bu amaçla geliştirilen g-ROKY'nin çalışma prensibi aşağıdaki adımlarla anlatılmıştır;

1. Örneklemden P sayıdaki $[Z_1(s), \dots, Z_P(s)]$ bağımsız sürekli değişkenleri kullanılarak M sayıda örneklenmemiş blok lokasyonuna $[x_i, i=1, \dots, M \in D]$ ait $[Z_1(x), \dots, Z_P(x)]$ bağımsız sürekli değişken değerleri Ortalamasız Krigleme yöntemi ile kestirilir.
2. Kriglemeyi takiben P sayıdaki örneklem bağımsız değişkeni ($Z(s)$) ve kriglenmiş bağımsız değişken ($Z(x)^*$) aşağıdaki eşitliklerle standartlaştırılır.

$$Z(s)_{st} = \frac{Z(s) - \mu(Z(s))}{\sigma(Z(s))} \quad Z(x)_{st}^* = \frac{Z(x) - \mu(Z(x)^*)}{\sigma(Z(x)^*)} \quad (3.1)$$

Eşitlik 3.1'de $Z(s)_{st}$ standartlaştırılmış örneklem bağımsız değişkenlerini, $Z(x)_{st}^*$ standartlaştırılmış kriglenmiş bağımsız değişkenleri, μ ortalamayı, σ ise standart sapmayı göstermektedir.

3. Standartlaştırılan bağımsız değişkenlerin gruplandırılması işleminin ilk adımında Eşitlik 3.2 ile Grup Aralığı hesaplanır:

$$\text{Grup Aralığı} = \{\text{Maks.}(Z(x)_{st}^*) - \text{Min}(Z(x)_{st}^*)\} / G \quad (3.2)$$

Eşitlikte G, oluşturulmak istenilen grup sayısına karşılık gelmektedir.

Bu durumda, gruplandırma işlemi aşağıdaki fonksiyonlarla yapılır:

$$f(Z(x)_{st}^*) = \begin{cases} 1 & \text{eğer } Z(x)_{st}^* < \{\min(Z(x)_{st}^*) + 1 \times \text{Grup Aralığı}\} \\ 2 & \text{eğer } \{\min(Z(x)_{st}^*) + (2 - 1) \times \text{Grup Aralığı}\} < Z(x)_{st}^* < \{\min(Z(x)_{st}^*) + 2 \times \text{Grup Aralığı}\} \\ \vdots & \vdots \\ G - 1 & \text{eğer } \{\min(Z(x)_{st}^*) + (G - 2) \times \text{Grup Aralığı}\} < Z(x)_{st}^* < \{\min(Z(x)_{st}^*) + (G - 1) \times \text{Grup Aralığı}\} \\ G & \text{eğer } \{\min(Z(x)_{st}^*) + (G - 1) \times \text{Grup Aralığı}\} < Z(x)_{st}^* \end{cases}$$

$$f(Z(s)_{st}) = \begin{cases} 1 & \text{eğer } Z(s)_{st} < \{\min(Z(x)_{st}^*) + 1 \times \text{Grup Aralığı}\} \\ 2 & \text{eğer } \{\min(Z(x)_{st}^*) + (2 - 1) \times \text{Grup Aralığı}\} < Z(s)_{st} < \{\min(Z(x)_{st}^*) + 2 \times \text{Grup Aralığı}\} \\ \vdots & \vdots \\ G - 1 & \text{eğer } \{\min(Z(x)_{st}^*) + (G - 2) \times \text{Grup Aralığı}\} < Z(s)_{st} < \{\min(Z(x)_{st}^*) + (G - 1) \times \text{Grup Aralığı}\} \\ G & \text{eğer } \{\min(Z(x)_{st}^*) + (G - 1) \times \text{Grup Aralığı}\} < Z(s)_{st} \end{cases}$$

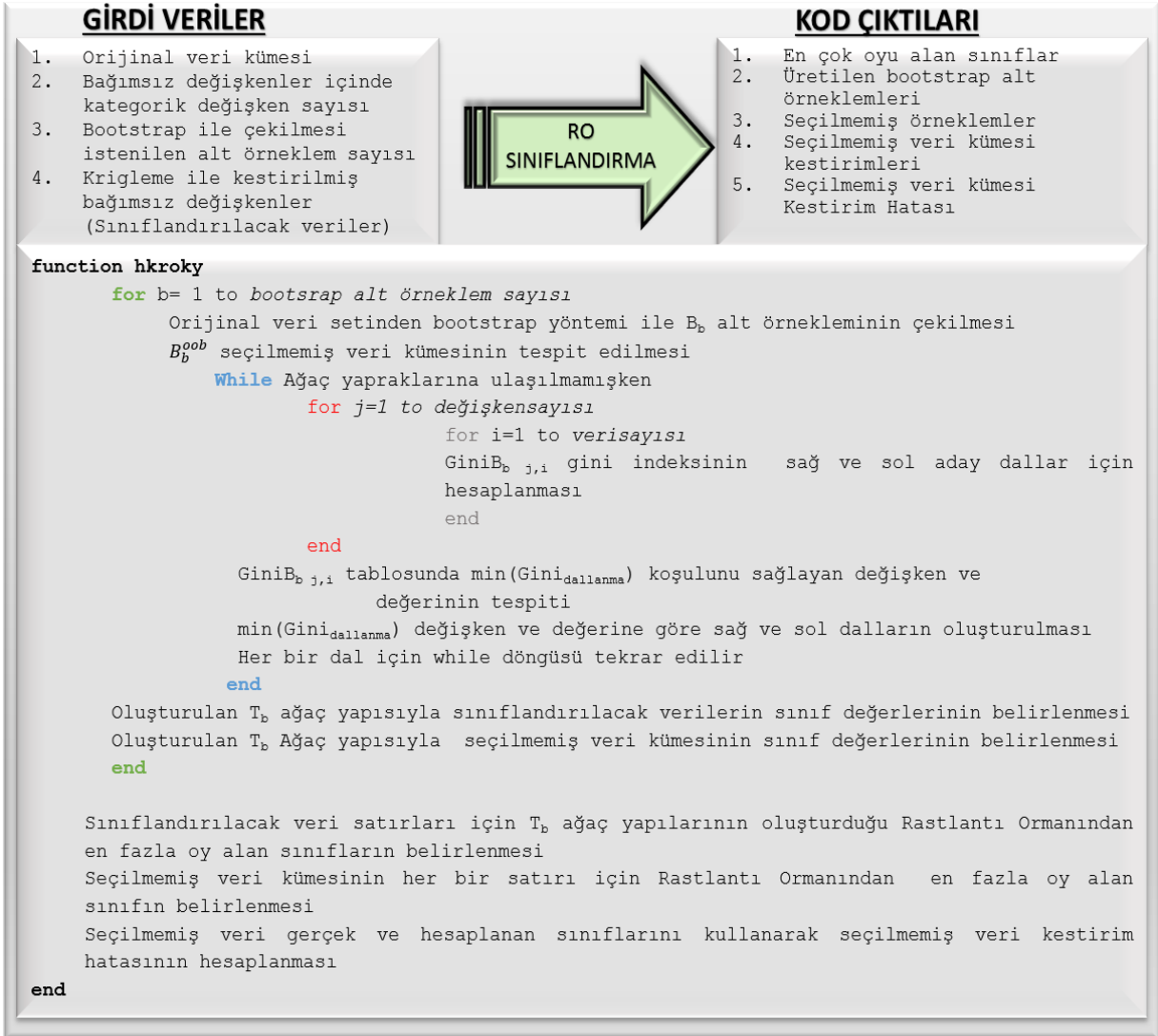
$f(Z(x)_{st}^*)$ ve $f(Z(s)_{st})$ fonksiyonları ile standartlaştırılmış örneklem grupları $Z^G(s)_{st}$ ve kriglenmiş veri grupları $Z^G(x)_{st}^*$ hesaplanır.

4. Rastlantı Ormanı ile $Z^G(s)_{st}$ ve örneklem bağımlı sınıf değişkeni olan $Y(s)$ arasındaki ilişki ortaya konulur.
5. $Z^G(x)_{st}^*$, dördüncü adımda oluşturulan Rastlantı Ormanına beslenerek örneklenmemiş x lokasyonundaki $Y(x)$ sınıfları hesaplanır.

g-ROKY'de ROKY'ye ek olarak grup sayısı G 'nin belirlenmesi gerekir. G değeri seçilmemiş veri kestirim hatasını en aza indiren minimum sayı olarak belirlenir.

3.2. Rastlantı Ormanı Algoritması, Kodlanması ve Kodun Çalışma Prensibi

Tez kapsamında Ortalamasız Krigleme yöntemiyle kestirilen öznitelik değerlerine göre bir sınıflandırma yapabilecek ve yapılan sınıflandırma sonuçlarının analiz edilebileceği nitelikte rastlantı orman algoritması oluşturma ihtiyacı doğmuştur. Bu amaçla MATLAB yazılımı kullanılarak bu işlemi gerçekleştirecek bir kod geliştirilmiştir. Oluşturulan rastlantı orman algoritması kodunun kullanıcıdan girdi olarak istediği veriler ve çıktı olarak ürettiği değerler Şekil 5'te verilmiştir.

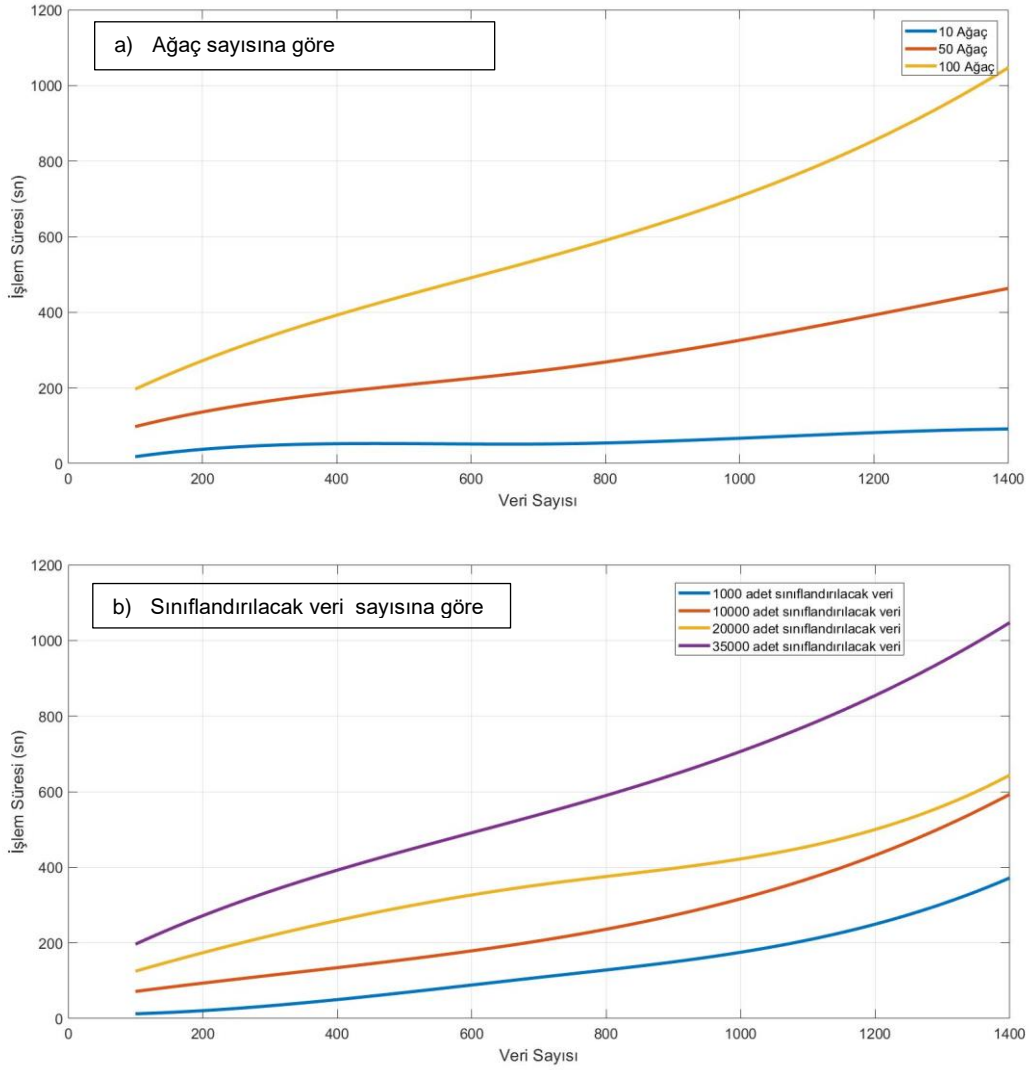


Şekil 5. Tez kapsamında MATLAB programında kodlanan RO algoritmasının girdi ve çıktıları ile basitleştirilmiş kod

Kodlanan Rastlantı Ormanı Algoritması Şekil 5'te basitleştirilerek gösterilmiştir.

Geliştirilen kodun; Rastlantı Ormanı oluşturulacak veri sayısı, sınıflandırılacak veri sayısı ve üretilmesi istenilen ağaç sayısına göre işlem süreleri Şekil 6'daki grafiklerde verilmiştir. Şekil 6'da verilen süreler 8 GB bellek (ram) ve 1.3 GHz, 4 çekirdekli işlemciye sahip bir bilgisayar için hesaplanmıştır. Kullanılan veriler 6 adet bağımsız sürekli değişken ve 5 farklı değer alabilen kategorik sınıf değişkenini içermektedir.

Şekil 6-a'da yapılan işlemler, RO'nun oluşturulması ile birlikte 35000 adet verinin sınıflandırılmasını kapsamaktadır. Şekil 6-b'de ise 100 adet sınıflandırma ağacı içeren RO ile farklı sayıda sınıflandırılacak veri için işlem süreleri verilmiştir.



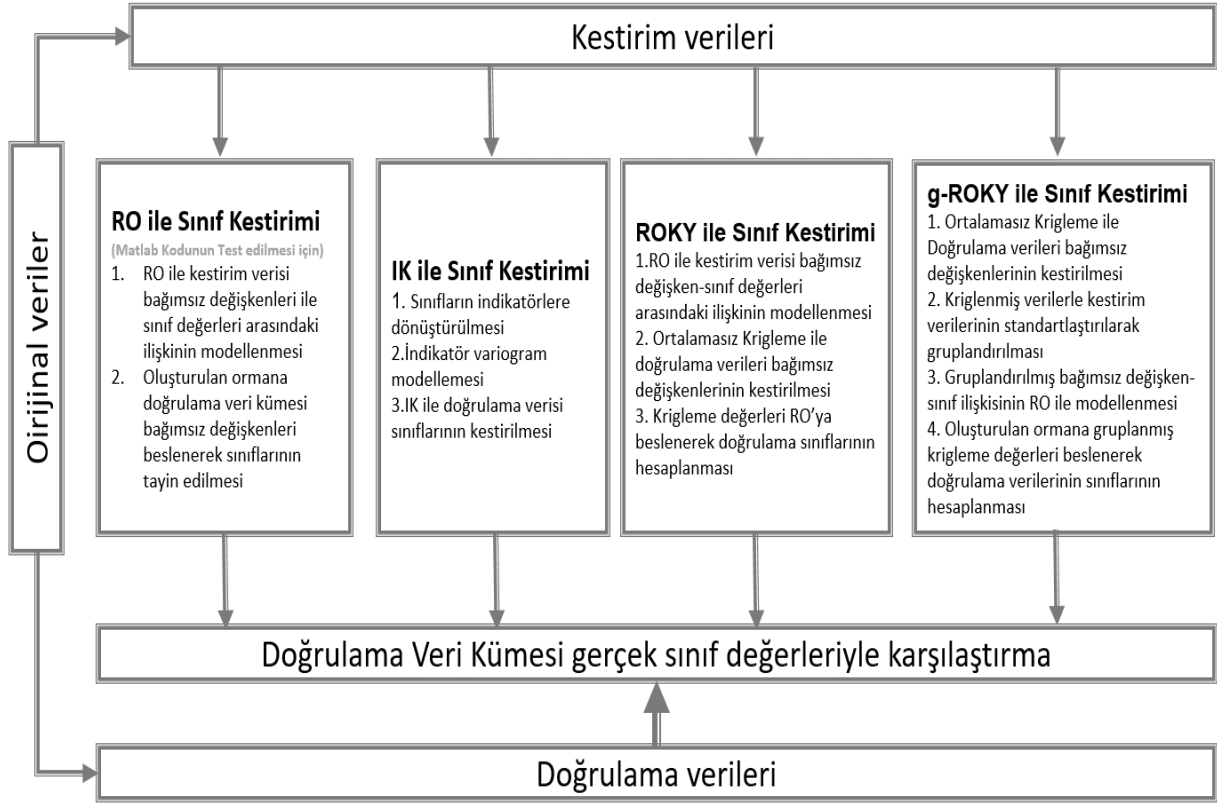
Şekil 6 Tez kapsamında geliştirilen RO kodunun sınıflandırma ağacı (a) ve sınıflandırılacak veri sayısına göre (b) işlem süreleri

Şekil 6, 400 adet veri içeren bir veri kümesi ile 100 adet ağaçtan oluşan bir RO oluşturma ve oluşturulan orman ile 35000 adet verinin sınıflandırma işlemlerinin yaklaşık 7 dakikalık bir süre içinde yapılabileceğini göstermektedir.

4. YÖNTEMLERİN JURA VERİ KÜMESİ ÜZERİNDE TESTİ

Bu bölümde tez kapsamında MATLAB programı aracılığıyla geliştirilen Rastlantı Ormanı kodu Jura veri kümesinde test edilmiştir. IK, ROKY ve g-ROKY kullanılan yöntemlerdir.

Yeni geliştirilen yöntemleri test etmek amacıyla kullanılan ve özellikleri 4.1 bölümünde verilen Jura veri kümesi için yapılan çalışmaların genel adımları Şekil 7’de gösterilmiştir.

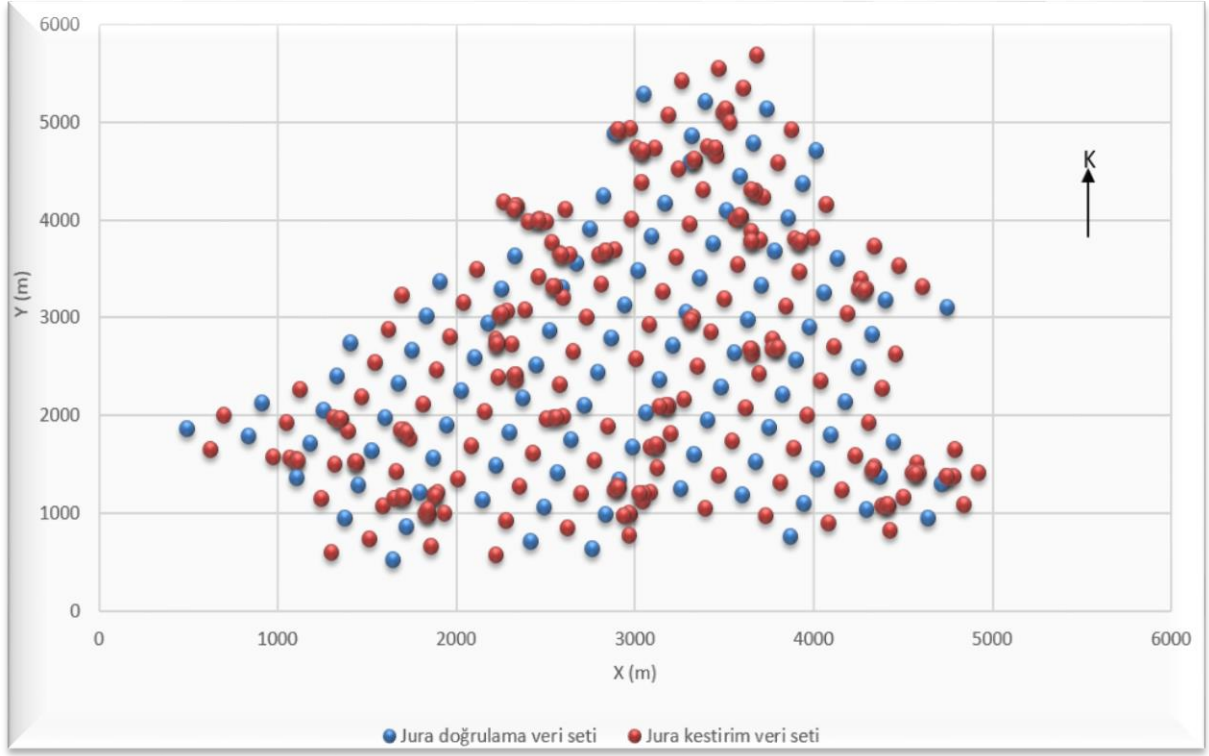


Şekil 7 Jura veri kümesi için uygulama adımları

4.1. Veri Özellikleri

Jura veri kümesi, eser miktarda ağır metaller içeren toprak yapısını analiz etmek için İsviçre Federal Teknoloji Enstitüsü'nce sağlanan kamuya açık bir veri kümesidir [41]. Veri kümesi; X ve Y koordinatları ile Cd, Co, Cr, Cu, Ni, Pb, Zn kalite değerleri sürekli değişkenler; üst örtü tipi ve kaya türü kategorik değişken olmak üzere 12 adet değişken değerini içermektedir. Bu veri kümesinin 259 adedi (Jura kestirim veri kümesi) kestirim için, 100 adedi (Jura doğrulama veri kümesi) ise doğrulama için ayrılmıştır. Jura

kestirim veri kümesi ile Jura doğrulama veri kümesinin konumları Şekil 8’de gösterilmiştir. Tez kapsamında yapılan uygulama çalışmalarının daha önce aynı veri kümesiyle Bel, Allard vd.’nin [42] Sınıflandırma Ağaçları ile yaptıkları sınıflandırma doğruluğuyla kıyaslanabilmesi için kestirim ve doğrulama veri kümeleri ile sınıf değerleri bahsedilen çalışmada kullanıldığı şekilde belirlenmiştir.



Şekil 8, Jura verilerin konumları

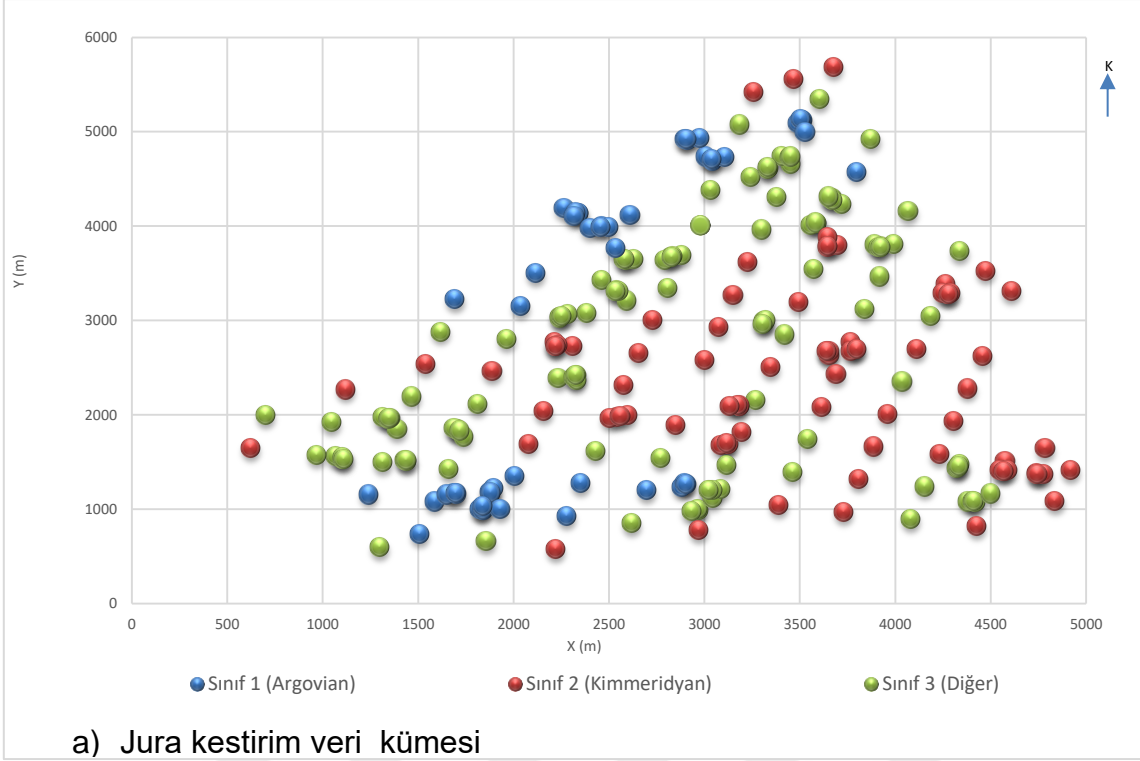
Veri kümesinde sürekli değişkenler dışında kategorik bağımsız değişken olan üst örtü tipi değişkeni de bulunmaktadır. Bu değişken; 1: Orman, 2: Mera, 3: Çimenlik, 4: Tarım Alanı değerlerini alabilmektedir. Üst örtü tipi değişkeni kategorik nitelikte olduğu için ROKY ile yapılan kestirimde kullanılmamış MATLAB programı ile kodlanan RO'nun çalışma doğruluğunun test edilmesi aşamasında dikkate alınmıştır.

Jura kestirim ve doğrulama veri kümelerinin sürekli bağımsız değişkenlerinin özet istatistikleri Çizelge 3'te ait oldukları kaya türü sınıfları da dikkate alınarak verilmiştir.

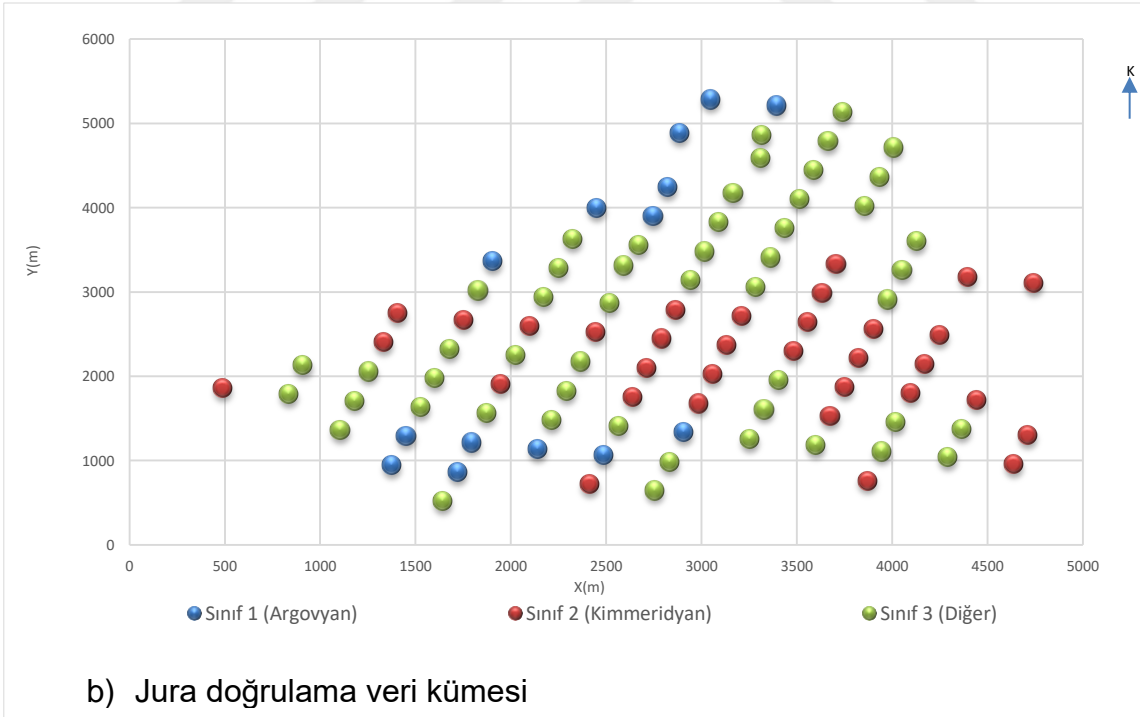
Çizelge 3. Jura kestirim ve doğrulama veri kümesine ilişkin sürekli değişkenlerin özet istatistikleri

İstatistiksel Bilgiler		Cd mg/kg	Co mg/kg	Cr mg/kg	Cu mg/kg	Ni mg/kg	Pb mg/kg	Zn mg/kg	
JURA KESTİRİM VERİLERİ	Tüm Sınıflar	Ortalama	1.31	9.3	35.07	23.73	19.73	53.92	75.08
		En yüksek	5.13	17.72	67.6	166.4	53.2	229.56	219.32
		En Düşük	0.13	1.55	8.72	3.96	4.2	18.96	25.2
		<i>Stand.Sapma</i>	0.92	3.58	10.96	20.71	8.23	29.79	29.02
	Sınıf 1	Ortalama	1.14	5.39	28.58	16.38	11.66	41.26	60.10
		En yüksek	5.13	10.32	52.40	66.12	22.92	141.00	145.60
		En Düşük	0.14	3.12	15.08	4.52	4.20	21.60	27.20
		<i>Stand.Sapma</i>	1.27	2.07	9.51	9.92	5.41	17.75	31.40
	Sınıf 2	Ortalama	1.35	11.05	35.41	22.23	24.99	56.37	77.08
		En yüksek	4.18	15.28	54.00	79.20	53.20	172.12	166.28
		En Düşük	0.22	1.55	8.72	4.20	5.36	27.20	28.88
		<i>Stand.Sapma</i>	0.70	2.81	8.84	17.77	7.49	26.41	22.93
Sınıf 3	Ortalama	1.36	9.79	37.67	28.00	19.57	57.73	80.23	
	En yüksek	4.02	17.72	67.60	166.40	35.28	229.56	219.32	
	En Düşük	0.28	2.12	14.52	3.96	4.64	18.96	25.20	
	<i>Stand.Sapma</i>	0.86	3.31	11.78	24.79	6.73	34.55	29.78	
JURA DOĞRULAMA VERİLERİ	Tüm Sınıflar	Ortalama	1.23	9.79	34.88	23.22	20.76	56.48	77.96
		En yüksek	3.78	20.6	70	154.6	43.68	300	259.84
		En Düşük	0.33	1.65	3.32	3.55	1.98	18.68	25
		<i>Stand.Sapma</i>	0.69	3.54	9.91	25.98	7.71	40.54	35.13
	Sınıf 1	Ortalama	0.94	6.49	29.84	20.98	13.71	49.82	66.97
		En yüksek	3.78	12.04	53.20	117.60	24.24	152.80	175.20
		En Düşük	0.38	2.36	18.12	3.55	5.60	21.12	25.00
		<i>Stand.Sapma</i>	0.78	2.96	9.25	23.14	5.77	31.86	36.79
	Sınıf 2	Ortalama	1.31	10.80	36.44	17.51	24.89	54.49	78.70
		En yüksek	2.65	15.32	60.40	127.00	43.68	300.00	192.00
		En Düşük	0.42	1.65	3.32	4.44	1.98	26.32	48.80
		<i>Stand.Sapma</i>	0.57	2.77	9.45	19.24	7.26	42.18	23.98
Sınıf 3	Ortalama	1.34	10.76	36.34	30.43	20.80	62.56	83.86	
	En yüksek	3.02	20.60	70.00	154.60	30.20	239.96	259.84	
	En Düşük	0.33	1.92	14.92	3.98	4.68	18.68	26.80	
	<i>Stand.Sapma</i>	0.72	3.42	9.98	31.88	6.04	43.55	42.37	

Veri kümesindeki değişkenlerden kategorik olan kaya türü değişkeni, sınıfları (homojen ortamları) temsil etmektedir. Bu sınıflar: 1: Argovian, 2: Kimmeridgian, 3: Sequanian, 4: Portlandian, 5: Quaternary'dir. Sınıflardan 3, 4 ve 5. sınıf kaya türleri 3. sınıfta birleştirilmiş ve bu çalışma toplam 3 sınıf üzerinden yapılmıştır (1. Argovian, 2. Kimmeridgian ve 3. diğerleri). Jura kestirim ve doğrulama verileri için bu 3 farklı sınıfın lokasyonları Şekil 9'da gösterilmiştir.



a) Jura kestirim veri kümesi

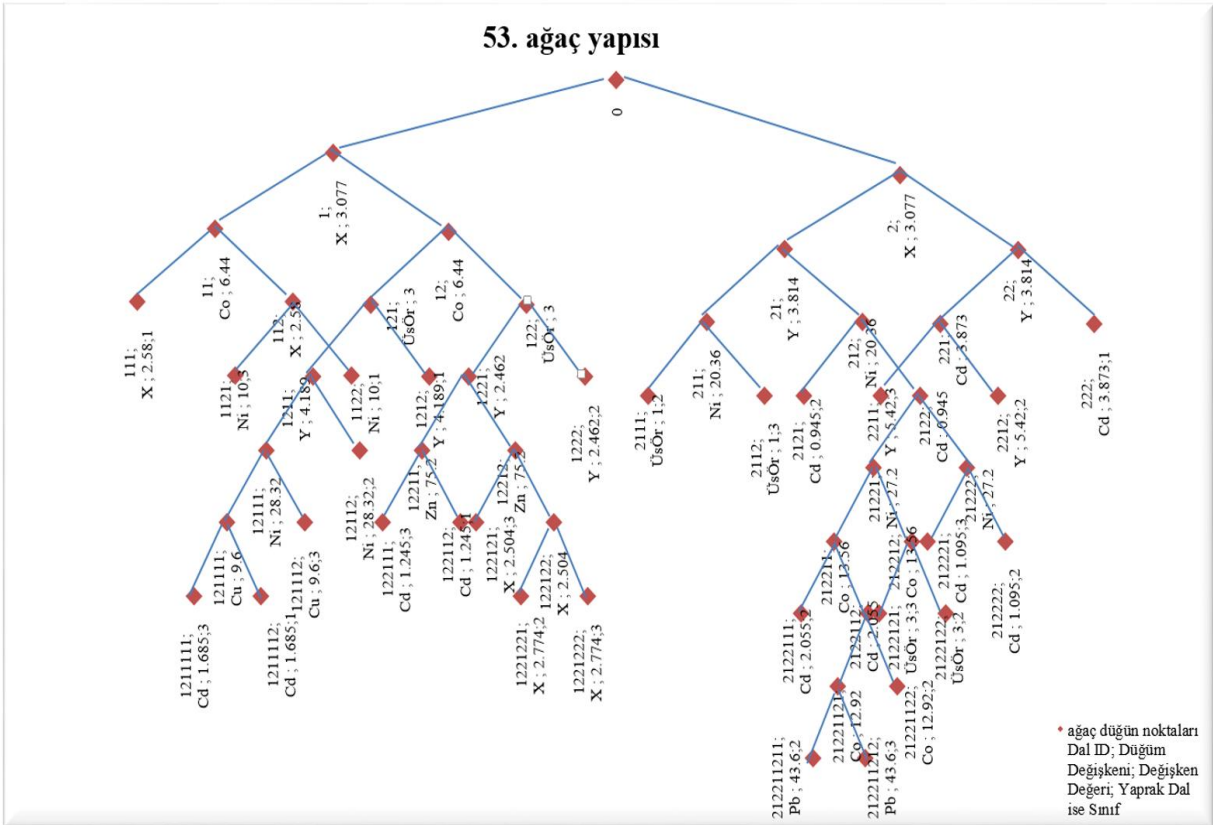


b) Jura doğrulama veri kümesi

Şekil 9. Jura verilerinin kaya türü sınıflarına göre lokasyon haritaları

4.2. Rastlantı Ormanı ile Jura Veri Kümesinin Sınıflandırılması

MATLAB programı aracılığıyla tez kapsamında kodlanan Rastlantı Ormanı algoritmasının test edilmesi amacıyla Jura doğrulama veri kümesinin bağımsız değişkenlerini (Cd, Co, Cr, Cu, Ni, Pb, Zn, Üstörtü Tipi) bildiğimizi ancak bağımsız değişken olan sınıf (kaya türü) değerlerini bilmediğimizi varsayalım. Jura kestirim veri kümesi ile Rastlantı Ormanı oluşturulmuş ve bağımlı ve bağımsız değişkenler arasındaki kural tabanlı ilişkiler belirlenmiştir. Bu ilişki aynı zamanda bağımsız değişken değerlerini bildiğimiz doğrulama veri kümesinin sınıf değerlerini tespit etmemize olanak sağlayacaktır. Bu amaçla oluşturulan ve 100 adet ağaçtan oluşan Rastlantı Ormanının seçilmemiş veri kestirim hatası %18.53 olarak hesaplanmıştır. Dolayısıyla, RO seçilmemiş verilerin %81.47'sinin sınıf değerlerini gerçek değerleriyle aynı şekilde kestirmiştir. RO'yu oluşturan 100 adet ağaç yapısından 53. ağaç Şekil 10'da verilmiştir.



Şekil 10. Jura kestirim veri kümesi için oluşturulan Rastlantı Ormanındaki 100 adet ağaçtan 53. Ağaç yapısı.

Şekil 10'da dal ya da düğüm etiketleri (id) 1 ve 2 rakamları kullanılarak verilmiştir. Veri kümesinin ayırım yapılmamış ilk hali 0 etiket numarası ile ağaç yapısının en tepesinde

yer almaktadır. Veri kümesinde düğüm için belirlenen sürekli bağımsız değişken değerinden küçük olanlar 1 dalına büyük eşit olanlar 2 dalına ayrılmaktadır. Örneğin 1 no'lu düğümden 11 ve 12 no'lu dallara ayrılma şartı incelenecek olursa, düğüm değişkeni Co değeri %6.44'ten küçük olan veri satırları 11 no'lu dala, büyük eşit olan veri satırları ise 12 no'lu dala ayrılacaktır.

İşlem sonucunda RO ile oluşturulan 100 adet sınıflandırma ağacı ile kestirilen Doğrulama Veri kümesinin kaya türleri ve bunların doğruluk oranları Çizelge 4'te verilmiştir. Aynı Çizelgede Bel, Allard vd.(2009) 'nin daha önce Sınıflandırma ve Regresyon Ağacı yöntemi (SRAT) ile aynı veri kümesini kullanarak yaptıkları kaya türü kestiriminin doğruluk oranları da verilmiştir. Hesaplanan kaya türleri Şekil 11'de gösterilmiştir.

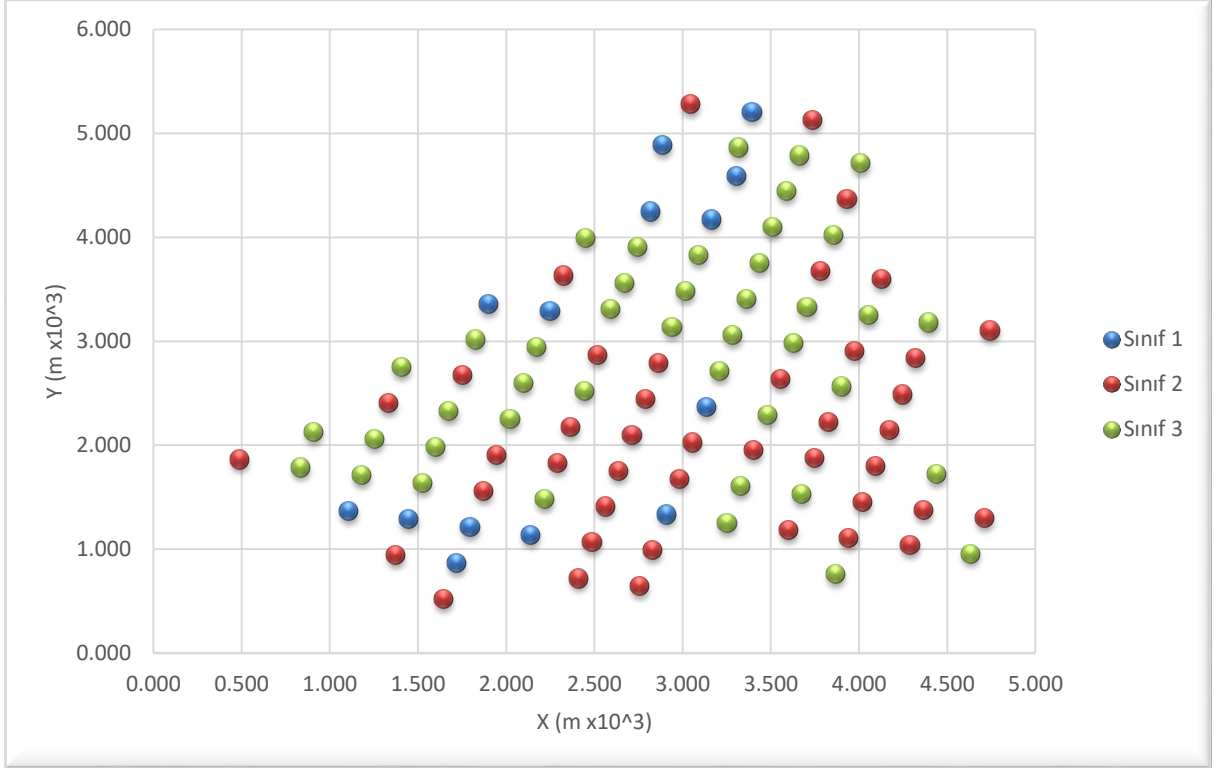
Çizelge 4. Matlab programıyla oluşturulan Rastlantı Ormanı koduyla yapılan sınıflandırmanın önceki çalışmalarla kıyaslanması

Açıklama	Kaya Türleri*			Toplam
	Sınıf 1	Sınıf 2	Sınıf 3	
Jura doğrulama veri kümesinin gerçek kaya türü sayıları	23	39	38	100
Bel, Allard vd.'nin [42] Jura kestirim kümesiyle oluşturdukları SRAT ile Jura doğrulama kümesine ilişkin doğru olarak tahmin ettikleri kaya türü sayıları	14	27	18	59
Jura kestirim veri kümesiyle oluşturulan Rastlantı Ormanı ile Jura doğrulama verilerinin tahmin edilen kaya türü sayıları	15	39	46	100
Jura kestirim veri kümesiyle oluşturulan Rastlantı Ormanı ile Jura doğrulama verilerinin doğru tahmin edilmiş kaya türü sayıları	12	30	29	71

*Sınıf 1: Argovian, 2: Kimmeridgian, 3: Diğer (Sequanian,+ Portlandian, + Quaternary)

Çizelge 4'ten görüleceği üzere Matlab Programı aracılığıyla kodlanan Rastlantı Ormanı algoritması Jura doğrulama veri kümesinin sınıf değerlerini, Bel, Allard vd.'nin [42] sınıflandırma ağacıyla yaptıkları kestirime kıyasla çok daha doğru şekilde kestirmiştir. RO ile Jura doğrulama veri kümesi için sınıf 3 olarak kestirilen toplam 46 adet sınıf

değerinden yaklaşık %63'ü doğrulama veri kümesinin gerçek değerleriyle aynı olmuştur. Sınıf 3 için kestirim sayısı ile doğru kestirim sayısı arasındaki bu verimsiz oran eğitim verileri olarak kullanılan Jura kestirim veri kümesinde sınıf 3 olarak tasnif edilmiş veri sayısının doğrulama veri kümesine göre daha yüksek olmasından ileri gelmektedir. RO, Jura kestirim veri kümesi ile oluşturulduğundan sınıflandırmada bu şekilde bir yanlılığa sebep olmuştur.



Şekil 11. Jura Doğrulama Veri kümesinin Rastlantı Ormanı Yöntemi ile hesaplanan homojen ortamları (Sınıf 1: Argovian, Sınıf 2: Kimmeridian, Sınıf 3: Diğer)

4.3. İndikatör Krigleme ile kaya türlerinin kestirimi

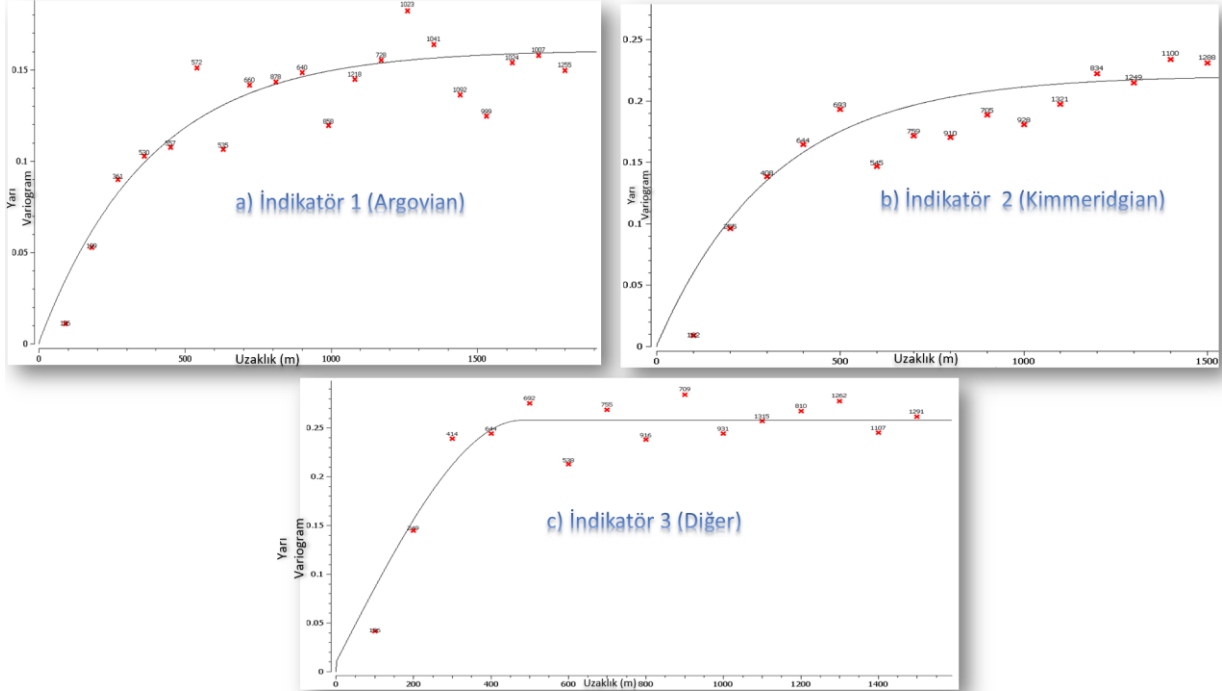
İndikatör Krigleme yöntemiyle kaya türlerinin tayini için ilk olarak Jura kestirim veri kümesindeki kaya türleri indikatörlere dönüştürülmüştür. Bu amaçla öncelikle sınıf 1'de yer alan (kaya türü Argovian olan) verilere 1 değeri, sınıf 2 ve 3 olan verilere ise 0 değeri atanarak birinci indikatör veri kümesi; daha sonra sınıf 2'de yer alan (kaya türü Kimmeridgian olan) verilere 1 değeri, sınıf 1 ve 3 olan verilere 0 değeri atanarak ikinci indikatör veri kümesi; son olarak sınıf 3'te yer alan (kaya türü diğer olan) verilere 1

değeri sınıf 1 ve 2 verilere ise 0 değeri atanarak üçüncü indikatör veri kümesi oluşturulmuştur.

Bu üç indikatör için oluşturulan variogramlara ilişkin parametreler Çizelge 5'te gösterilmiştir.

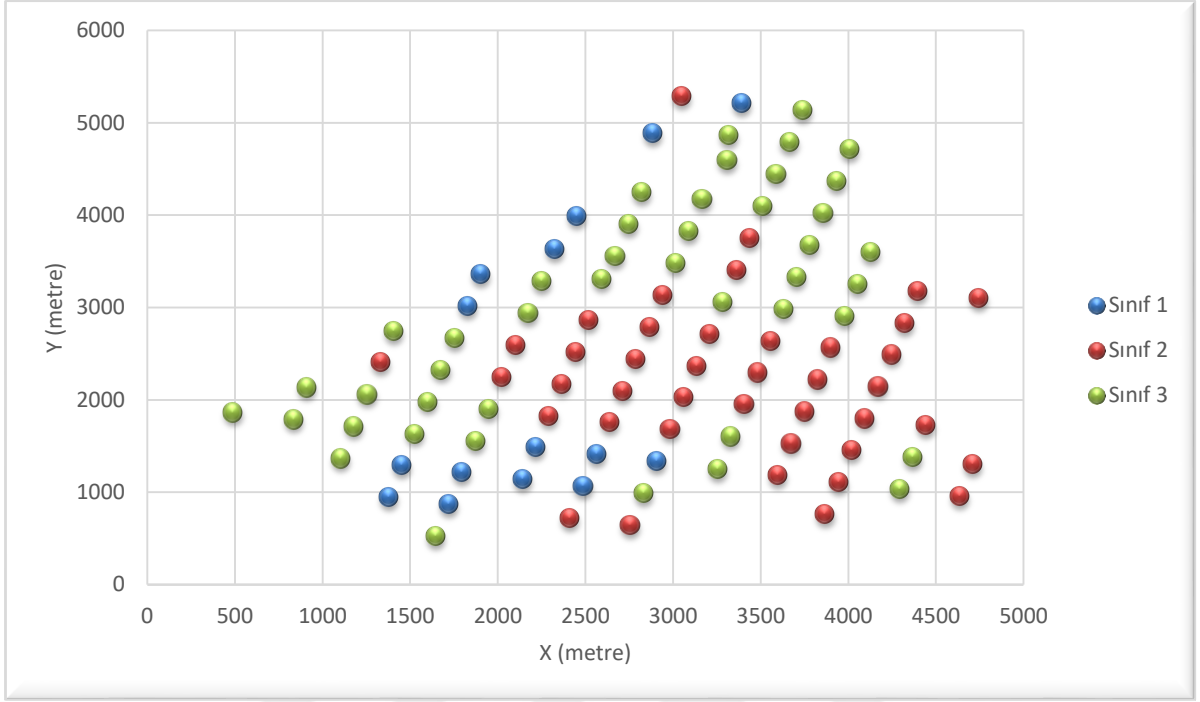
Çizelge 5. Jura kestirim veri kümesi kaya türü indikatörlerinin

İndikatör	C ₀	C	a	Variogram Türü
İnd1 (Argovian)	0	0.16	1134	Üssel
İnd2 (Kimmeridgian)	0	0.22	940	Üssel
İnd3 (Diğer)	0	0.248	450	Kürese



Şekil 12. Jura kestirim veri kümesi indikatörlerinin yön bağımsız deneysel variogramları ve variogram modelleri

Şekil 12 'deki Jura kestirim verileri için oluşturulan variogram modelleri kullanılarak Jura doğrulama verilerinin kaya türleri İndikatör Krigleme yöntemi ile kestirilmiştir. Her üç indikatör için yapılan kestirimlerden (sınıf 1, sınıf 2 ve sınıf 3) maksimum olanı ilgili veri satırının kaya türü olarak belirlenmiştir. Kestirilen sınıf değerlerine ilişkin lokasyon haritası Şekil 13 'te verilmiştir.



Şekil 13 Jura Doğrulama Veri kümesinin İndikatör Krigleme yöntemi ile kestirilmiş kaya türü sınıflarına ilişkin konum haritası (Sınıf 1: Argovian, Sınıf 2: Kimmeridian, Sınıf 3: Diğer)

Çizelge 6'da yapılan sınıf kestirimlerinin doğruluk oranları verilmiştir.

Çizelge 6. İndikatör Krigleme ile kestirilen Doğrulama Veri kümesi sınıflarının doğruluk oranları ve karşılaştırma çizelgesi

Açıklama	Kaya Türleri*			Toplam
	Sınıf 1	Sınıf 2	Sınıf 3	
Jura doğrulama veri kümesinin gerçek kaya türü sayıları	23	39	38	100
Bel, Allard vd'nin [42] Jura kestirim veri kümesiyle oluşturdukları SRAT ile Jura doğrulama veri kümesine ilişkin doğru olarak tahmin ettikleri kaya türü sayıları	14	27	18	59
İndikatör Krigleme ile tahmin edilen kaya türü sayıları	15	41	44	100
İndikatör Krigleme ile doğru tahmin edilen kaya türü sayıları	13	28	24	65

*Sınıf 1: Argovian, 2: Kimmeridgian, 3: Diğer (Sequanian,+ Portlandian, + Quaternary)

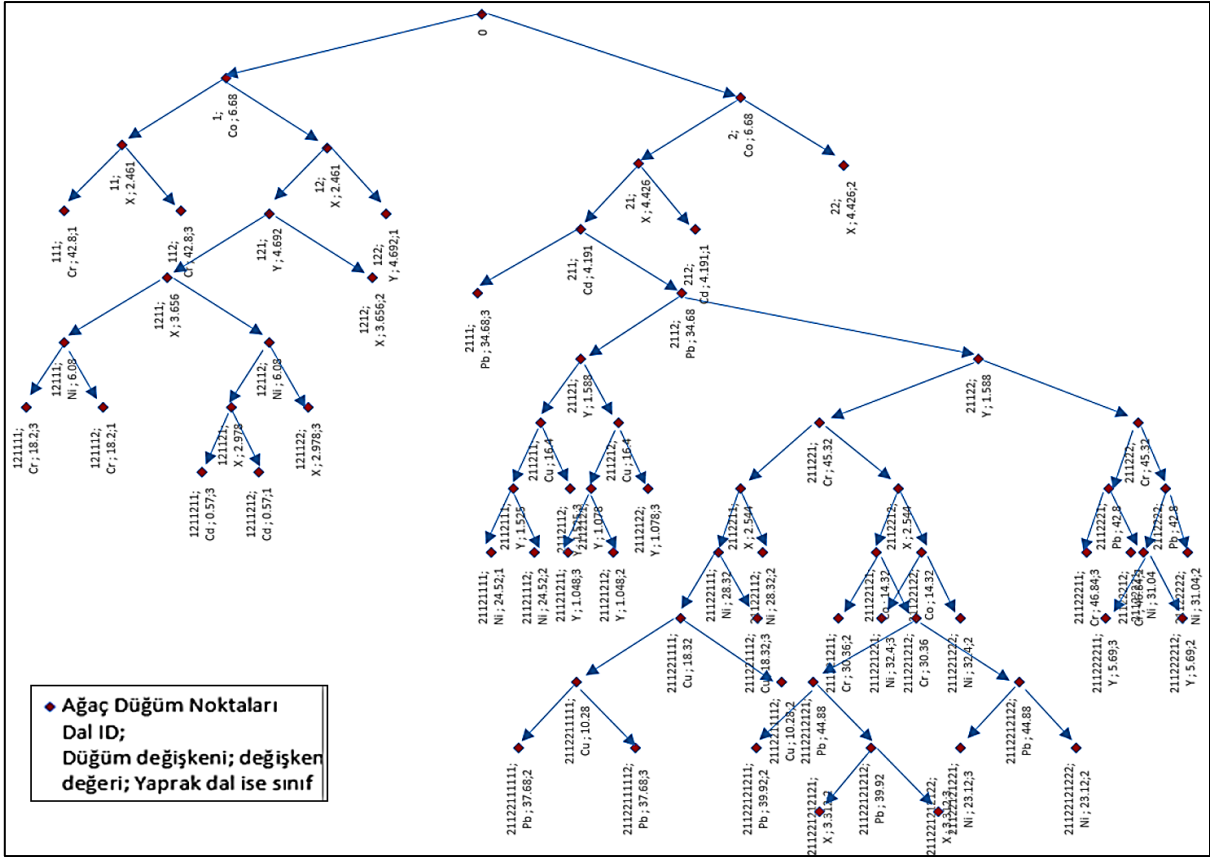
İndikatör Krigleme ile yapılan Jura doğrulama veri kümesi sınıf kestiriminde doğrulama verilerinin %65'inin gerçek sınıf değerleriyle aynı kestirildiği ve İndikatör Krigleme ile

yapılan sınıflandırmanın SRAT ile yapılmış olan sınıflandırmaya kıyasla daha doğru kestirim sonuçları ürettiği görülmektedir.

4.4. Rastlantı Ormanı + Krigleme yöntemiyle (ROKY) Jura Veri kümesi Sınıflandırılması

Jura veri kümesinden oluşan örnekte kaya türü değerleri (homojen sınıflar) bağımlı değişken, öznitelik ile koordinat değerleri bağımsız değişkenlerdir. Bu değişkenler arasındaki ilişkiyi ortaya koymak için Rastlantı Orman algoritması kullanılmıştır. Amaç, örnek olmayan lokasyonlardaki kaya türlerinin belirlenmesidir. Jura doğrulama veri kümesinin bulunduğu lokasyonlara ilişkin değişken değerlerinin ve kaya türlerinin bilinmediğini varsayalım. Bu değerleri elimizde var olan örnekleri, yani Jura kestirim veri kümesini kullanarak tayin etmeye çalışacağız. Bu amaçla ilk etapta Jura kestirim veri kümesindeki bağımlı değişken (kaya türü) ile bağımsız değişkenler (X,Y, Cd, Co, Cr, Cu, Ni, Pb, Zn) arasındaki ilişkiyi tespit etmek için 100 adet sınıflandırma ağacından oluşan Rastlantı Ormanı oluşturulmuştur (üst örtü tipi kategorik değişken olduğu için Ortalamasız Krigleme ile kestirilmemiş dolayısıyla Rastlantı Ormanı + Krigleme yönteminde dikkate alınmamıştır). Bu ağaçlardan birinin örneği Şekil. 14'te verilmiştir. Oluşturulan Rastlantı Ormanının seçilmemiş veri kestirim hatası %22'dir.

Şekil 14'te düğüm etiketleri 1 ve 2 rakamları kullanılarak verilmiştir. Veri kümesinin ayırım yapılmamış ilk hali 0 numarası ile ağaç yapısının en tepesinde yer almaktadır. Ağaç yapısı oluşturulurken veri kümesinde düğüm için belirlenen sürekli bağımsız değişken değerinden küçük olanlar 1 dalına büyükeşit olanlar ise 2 dalına ayrılmaktadır. Örneğin sınıflandırma işlemi 12 no'lu düğüm için gerçekleştirildiğinde Y koordinat değerinin 4692'den küçük olan veri satırları 121 no'lu dala, büyükeşit olduğu veri satırları ise 122 no'lu dala ayrılmaktadır.



Şekil 14. Jura kestirim veri kümesi için oluşturulan ve 100 adet sınıflandırma ağacından oluşan Rastlantı Ormanından 40. ağaç yapısı.

İkinci aşamada Jura kestirim veri kümesi kullanılarak Ortalamasız Krigleme ile Jura doğrulama veri kümesinin bulunduğu lokasyonlara ilişkin değişkenlerin (Cd, Co, Cr, Cu, Ni, Pb, Zn) değerleri kestirilmiştir.

Son olarak Jura doğrulama veri kümesinin bulunduğu lokasyonlar için kestirilmiş bağımsız değişkenlerin sınıfları (kaya türleri) Jura kestirim veri kümesi kullanılarak oluşturulan Rastlantı Ormanı yardımıyla hesaplanmıştır.

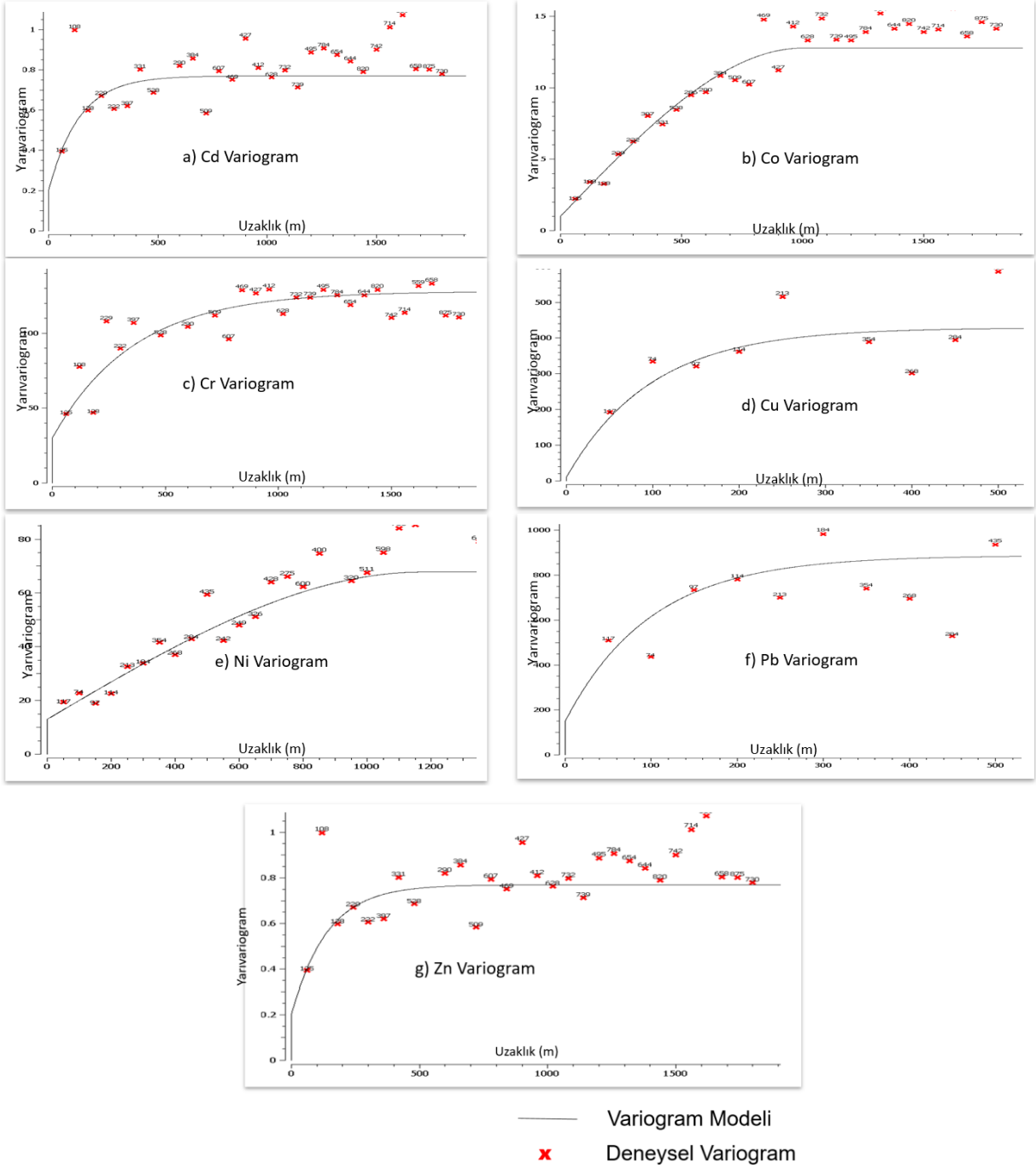
Jura kestirim veri kümesindeki değişkenler için (Cd, Co, Cr, Cu, Ni, Pb, Zn) belirlenen variogram parametreleri Çizelge 7’de belirtilmiştir:

Çizelge 7. Jura kestirim veri kümesi değişkenlerinin variogram parametreleri

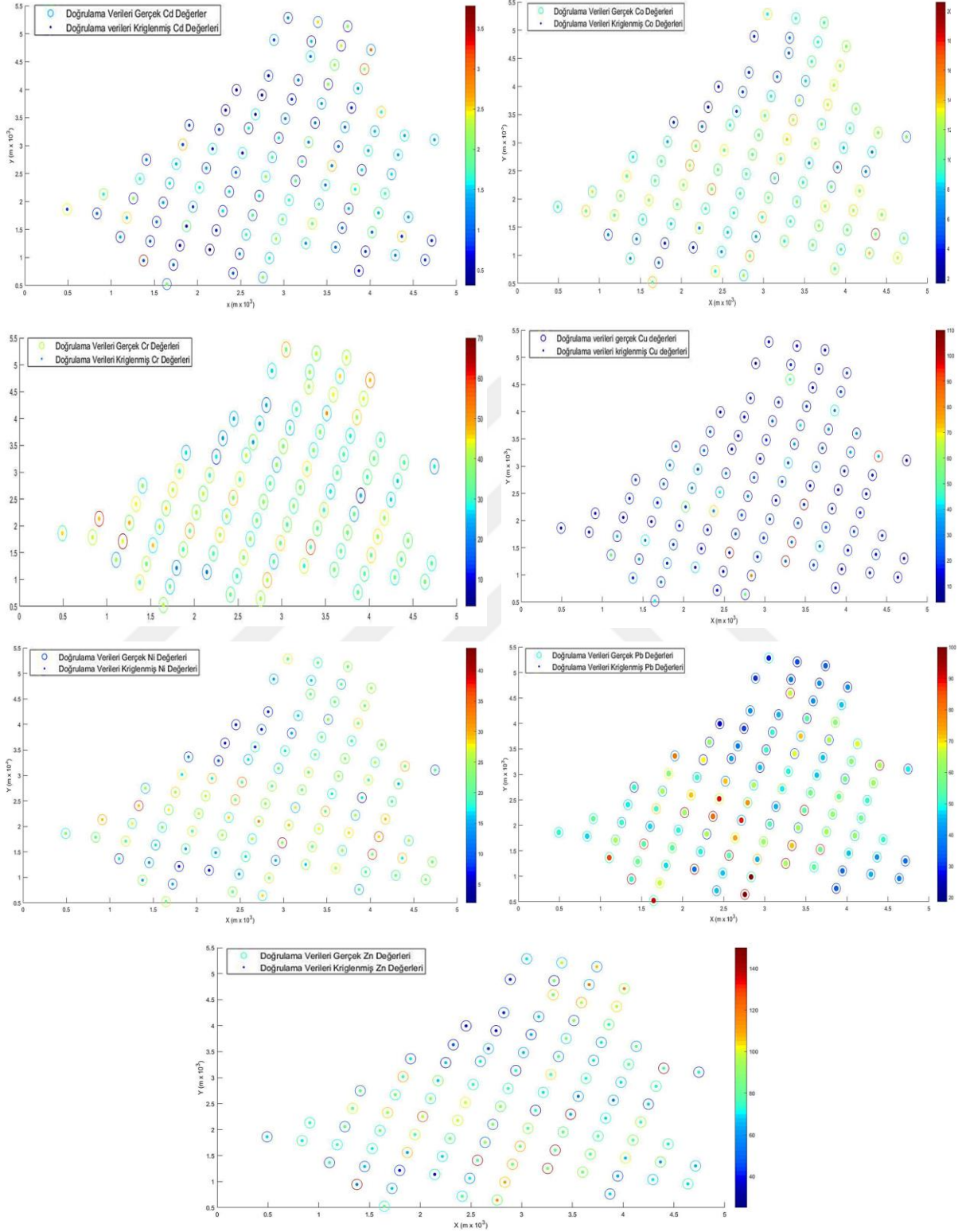
Değişken	C0	C	a	Variogram modeli
Cd	0.2	0.57	450	Üssel
Co	1	11.78	1000	Küresel
Cr	30	98	1080	Üssel
Cu	10	419	295	Üssel
Nı	13	54.78	1170	Küresel
Pb	100	787.57	300	Üssel
Zn	110	732	585	Üssel

Jura kestirim veri kümesi için oluşturulan variogram modelleri Şekil 15'te gösterilmiştir. Söz konusu variogram modelleri yön-bağımsız (omni-directional) olarak tespit edilmiştir.

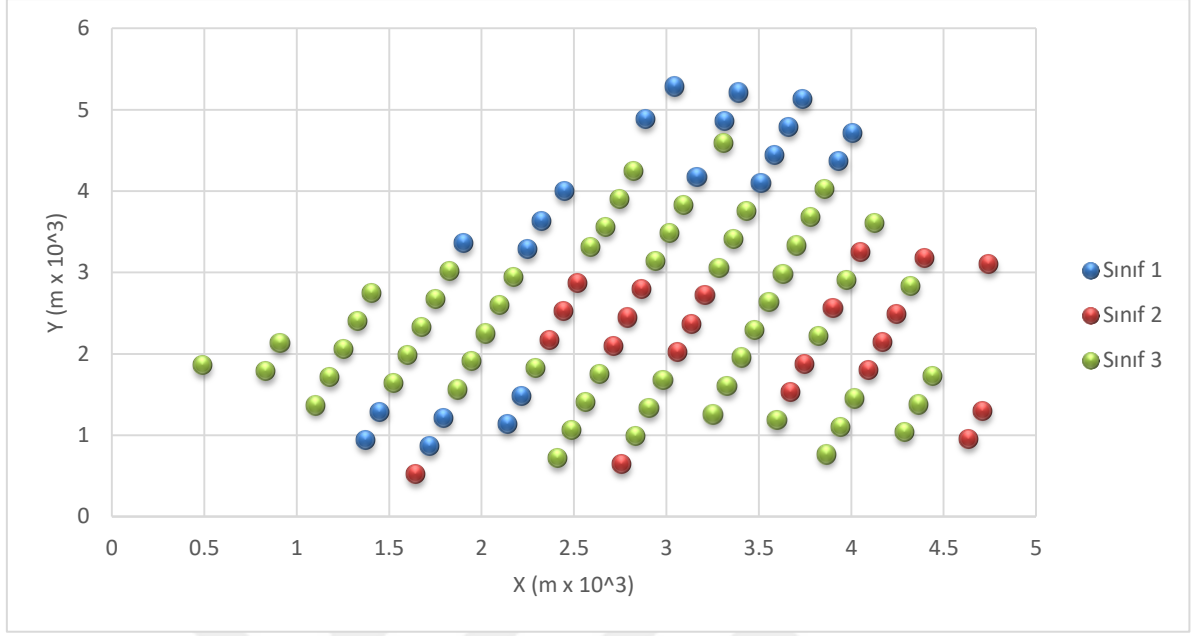
Şekil 15 'te gösterilen variogram modelleri kullanılarak Jura doğrulama verilerinin bağımsız değişkenleri Ortalamasız Krigleme yöntemi ile kestirilmiştir. Kestirilen ve gerçek değerlerin lokasyon haritaları Şekil 16'da gösterilmiştir. Şekil 17 ise ROKY ile kestirilen homojen ortamları göstermektedir. Yapılan homojen ortam kestirimlerinin Jura doğrulama veri kümesinin geçek kaya türleri ile kıyaslaması Çizelge 8'de verilmiştir.



Şekil 15. Jura kestirim veri kümesi değişkenlerinin yön bağımsız (omni-directional) deneysel variogramları ve variogram modelleri.



Şekil 16. Jura Doğrulama Veri kümesinin sürekli bağımsız değişkenlerinin gerçek değerleri ve Ortalamasız Krigleme ile kestirilmiş değerleri



Şekil 17. Jura doğrulama veri kümesinin ROKY ile hesaplanan homojen sınıfları (Sınıf 1: Argovian, Sınıf 2: Kimmeridian, Sınıf 3: Diğer)

Çizelge 8. ROKY ile kestirilen Jura doğrulama veri kümesi sınıflarının doğruluk oranları

Açıklama	Kaya Türleri			Toplam
	Sınıf 1	Sınıf 2	Sınıf 3	
Jura doğrulama veri kümesinin gerçek kaya türü sayıları	23	39	38	100
Bel, Allard vd.'nin [42] Jura kestirim veri kümesiyle oluşturdukları SRAT ile Jura doğrulama veri kümesine ilişkin doğru olarak tahmin ettikleri kaya türü sayıları	14	27	18	59
İndikatör Krigleme ile doğru tahmin edilen kaya türü sayıları	13	28	24	65
ROKY ile tahmin edilen kaya türü sayıları	13	27	60	100
ROKY ile doğru tahmin edilen kaya türü sayıları	11	20	30	61

Çizelge 8, ROKY'nin tek bir sınıflandırma ağacıyla yapılan sınıf kestirimine göre daha doğru sonuçlar ürettiğini göstermektedir. Burada; Bel, Allard vd.'nin [42] yaptıkları çalışmada doğrulama veri kümesinin gerçek öznelik değerleri kullanarak sınıflandırma ağacı oluşturdukları ancak ROKY ile doğrulama veri kümesinin özneliklerinin bilinmediği varsayılarak krigleme ile kestirildikten sonra kestirim değerlerinin Rastlantı Ormanına beslendiği hususu göz ardı edilmemelidir. Yani, kestirilmiş öznelik değerleri kullanılmış olmasına rağmen; ROKY, sınıflandırma ağacıyla yapılan sınıflandırmaya

kıyasla daha doğru sınıf değerleri hesaplamıştır. Diğer taraftan, İndikatör Krigleme yönteminin sınıf kestirimi ROKY'ye kıyasla gerçek değerlere daha yakın sonuçlar vermiştir. ROKY yalnızca sınıf 3 türünün kestirimini İndikatör Krigleme yöntemine göre daha doğru yapmış olsa da Çizelge 8'de ROKY ile tahmin edilen kaya türü sayıları satırında da belirtildiği gibi ROKY 100 adet doğrulama verisininin 60 adedini sınıf 3 olarak kestirmiş ve 60 adet sınıf 3 kestirimini ancak %50'si gerçek değerlerle aynı olabilmıştır. Jura kestirim veri kümesinde Jura doğrulama veri kümesine nazaran sınıf 3 türünün diğer türlerden sayıca daha fazla olması (sınıf 1, 53 adet; sınıf 2, 85 adet ve sınıf 3, 121 adet), Rastlantı Ormanında sınıflandırma açısından bu yanlılığa sebep olmuştur. İndikatör Krigleme yöntemi sonuçları ise Jura kestirim veri kümesindeki sınıf 3 türü fazlalığını kestirim sonuçlarına çok daha az yansıtmıştır.

4.5. Gruplandırılmış Rastlantı Ormanı + Krigleme yöntemiyle (g-ROKY) Jura Veri kümesinin Sınıflandırılması

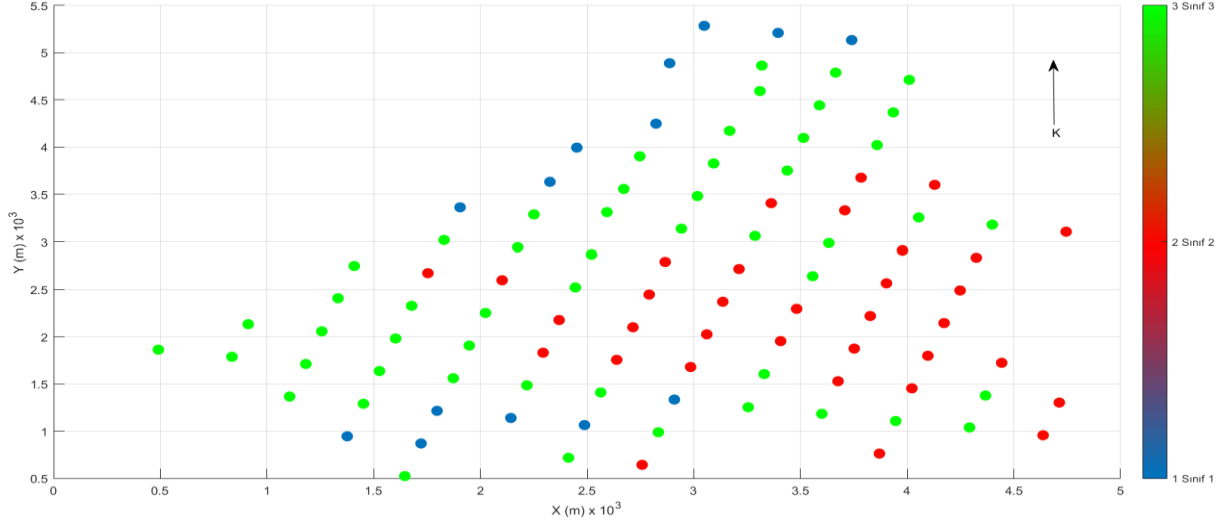
g-ROKY ile Jura doğrulama veri kümesinin sınıflarının hesaplanmasının ilk adımında 4.4. bölümde ROKY ile yapılan sınıflandırmada kullanılan kriglenmiş veriler ve Jura kestirim verilerinin öznitelikleri standartlaştırılarak gruplandırılmıştır. Çizelge 9'da her bir gruba karşılık gelen standartlaştırılmış bağımsız değişken değer aralıkları verilmiştir.

Çizelge 9 Jura kestirim veri kümesi bağımsız değişkenleri ve kriglenmiş Jura doğrulama veri kümesi bağımsız değişken grupları

Grup	Cd (mg/kg)		Co (mg/kg)		Cr (mg/kg)		Cu (mg/kg)		Ni (mg/kg)		Pb (mg/kg)		Zn (mg/kg)	
1	<	-1.61	<	-1.99	<	-1.95	<	-0.94	<	-2.17	<	-1.33	<	-2.01
2	-1.61	-1.10	-1.99	-1.56	-1.95	-1.49	-0.94	-0.29	-2.17	-1.71	-1.33	-0.79	-2.01	-1.48
3	-1.10	-0.58	-1.56	-1.12	-1.49	-1.03	-0.29	0.36	-1.71	-1.25	-0.79	-0.26	-1.48	-0.96
4	-0.58	-0.07	-1.12	-0.69	-1.03	-0.57	0.36	1.01	-1.25	-0.80	-0.26	0.27	-0.96	-0.44
5	-0.07	0.45	-0.69	-0.26	-0.57	-0.11	1.01	1.66	-0.80	-0.34	0.27	0.80	-0.44	0.08
6	0.45	0.96	-0.26	0.18	-0.11	0.34	1.66	2.31	-0.34	0.12	0.80	1.34	0.08	0.60
7	0.96	1.48	0.18	0.61	0.34	0.80	2.31	2.96	0.12	0.57	1.34	1.87	0.60	1.13
8	1.48	1.99	0.61	1.05	0.80	1.26	2.96	3.61	0.57	1.03	1.87	2.40	1.13	1.65
9	1.99	2.51	1.05	1.48	1.26	1.72	3.61	4.26	1.03	1.49	2.40	2.94	1.65	2.17
10	2.51	<	1.48	<	1.72	<	4.26	<	1.49	<	2.94	<	2.17	<

Gruplandırılan Jura kestirim veri kümesi ile 100 adet sınıflandırma ağacı içeren Rastlantı Ormanı oluşturulmuştur. RO'nun seçilmemiş veri sınıf kestirimi hatası

%.17.8'dir. g-ROKY ile hesaplanan Jura doğrulama veri kümesi sınıfları Şekil 17'de gösterilmiştir.



Şekil 18. Jura doğrulama veri kümesinin g-ROKY ile hesaplanan sınıf haritası

Oluşturulan Rastlantı Ormanı kullanılarak yapılan Jura doğrulama veri kümesi sınıflandırmaları ve yapılan sınıflandırmaların diğer çalışmalarla kıyaslamaları Çizelge 10'da verilmiştir.

Çizelge 10. Jura doğrulama veri kümesinin g-ROKY ile hesaplanan sınıfları ve karşılaştırmalar

Açıklama	Kaya Türleri			Toplam
	Sınıf 1	Sınıf 2	Sınıf 3	
Jura doğrulama veri kümesinin gerçek kaya türü sayıları	23	39	38	100
İndikatör Krigleme ile doğru tahmin edilen kaya türü sayıları	13	28	24	65
ROKY ile doğru tahmin edilen kaya türü sayıları	11	20	30	61
g-ROKY ile tahmin edilen kaya türü sayıları	18	34	48	100
g-ROKY ile doğru tahmin edilen kaya türü sayıları	15	25	28	68

g-ROKY ile yapılan Jura doğrulama veri kümesi sınıflandırmaların SRAT'la daha önceki çalışmada [42] yapılan sınıflandırmaya ve İndikatör Krigleme ve ROKY ile yapılan sınıflandırmaya göre daha doğru sonuç verdiği Çizelge 10'da görülmektedir.

5. GELİŞTİRİLEN YÖNTEMLERİN İĞDEKUZU LİNYİT SAHASINDA UYGULANMASI

Bu bölümde yeni geliştirilen yöntemler İğdekuzu linyit sahasına uygulanmıştır.

İğdekuzu linyit sahasında; sondaj verilerinden elde edilen bilgiler kullanılarak, herhangi bir örnekleme yapılmamış olan ve linyit sahasının tamamını kapsayacak şekilde oluşturulan 20x20x1 m³'lük blokların homojen ortamları (litolojik sınıfları) İK, ROKY ve g-ROKY ile belirlenmiştir.

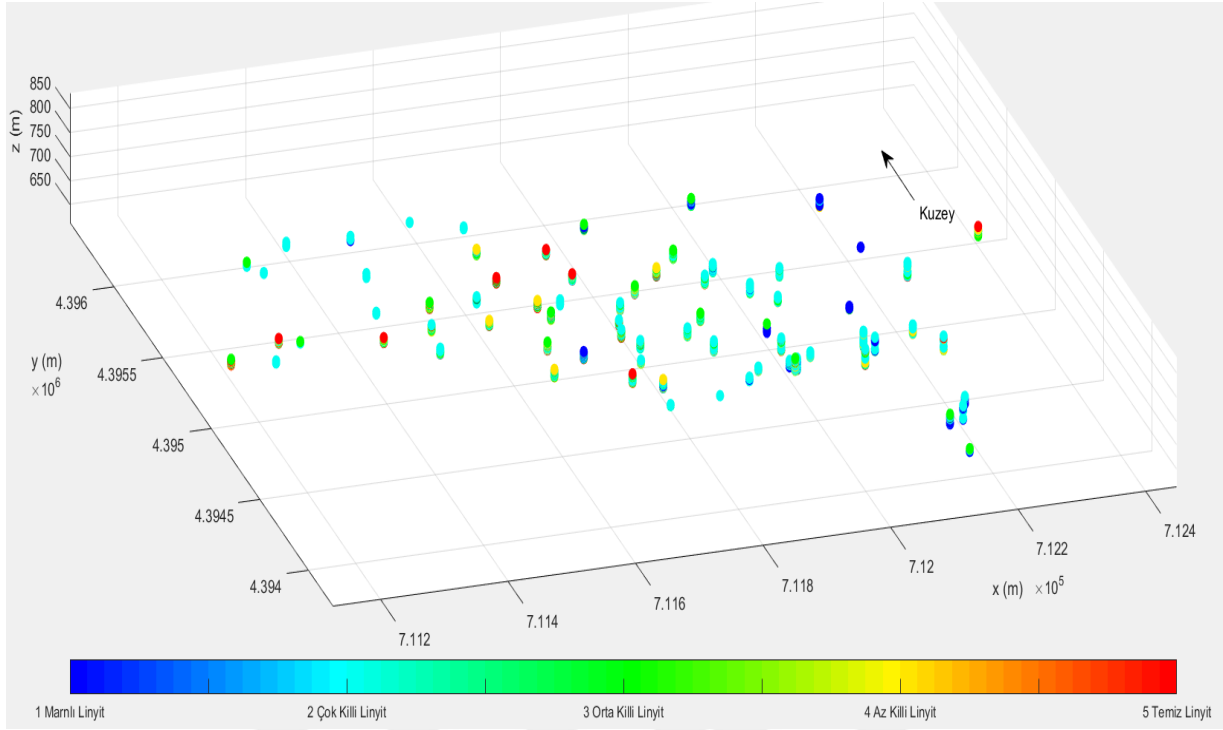
5.1. İğdekuzu linyit sahası Verileri

İğdekuzu linyit sahası; Kütahya ilinin Tavşanlı ilçesinde, Tunçbilek bölgesinde yer almaktadır. Türkiye Kömür İşletmeleri Kurumu'na ait olan saha özel bir firma tarafından işletilmektedir.

İğdekuzu linyit sahası veri kümesi; koordinat değerleri (X, Y, Z), alt ısı değer, kül ve nem bağımsız değişkenleri ile bu değerlere karşılık gelen ve bağımlı değişken olan litolojik birimleri içermektedir. Sahadaki veriler toplam 72 adet sondajdan elde edilen 1461 adet karot analizinden oluşmaktadır. Linyit yatağına ilişkin 1461 adet veri bulunmaktadır. Sondaj lokasyonları litolojiler ile birlikte Şekil 19'da verilmiştir.

Sondajlardan alınan karot örnekleri alt ısı değer, kül ve nem için analiz edilmiş ve litolojileri belirlenmiştir. Veri kümesi 5 adet litolojik sınıf verisi içermektedir. Bu litojik sınıflar; 1: Marnlı Linyit (ML), 2: Çok Killi Linyit (ÇKL), 3: Orta Killi Linyit (OKL), 4:Az Killi Linyit (AKL), 5: Temiz Linyit (TL) olarak ayrılmıştır. En fazla sayıda veri Çok Killi Linyit litolojisinde (675 adet) mevcuttur. En az sayıda veriyi ise Marnlı Linyit (84 adet) ve Temiz Linyit (133 adet) litolojik sınıfları içermektedir.

Linyitin kalite değişkenlerine ilişkin özet istatistikler Çizelge 11'de verilmiştir. Çizelgeden Temiz Linyit olarak sınıflandırılmış verilerin alt ısı değerlerinin en yüksek, kül değerlerinin ise en düşük ortalamaya sahip oldukları görülmektedir. Ayrıca Marnlı Linyit lotolojik sınıfındaki verilerin en düşük alt ısı değer, en yüksek kül ortalamasına sahip oldukları anlaşılmaktadır. Nem değişkenine ilişkin en düşük ortalamaya sahip sınıf Marnlı Linyit sınıfı olmuştur.



Şekil 19. İğdekuzu linyit sahası sondaj ve litolojilerin perspektif görüntüsü

Çizelge 11. İğdekuzu linyit sahası kalite değişkenlerinin özet istatistikleri

Litolojik Sınıf	Değişken	Ortalama	En		Standart Sapma
			Düşük	Yüksek	
Tüm Sınıflar	Nem (%)	6.51	0.70	41.00	2.69
	Kül (%)	40.59	6.60	87.30	14.02
	Alt Isıl Değer (kcal/kg)	3645	705	7009	1252
Sınıf 1	Nem (%)	5.34	1.50	12.00	2.35
	Kül (%)	63.48	60.06	70.24	2.18
	Alt Isıl Değer (kcal/kg)	1436	705	2067	214
Sınıf 2	Nem (%)	6.69	1.00	41.00	2.98
	Kül (%)	50.20	8.00	87.30	7.25
	Alt Isıl Değer (kcal/kg)	2831	797	4484	636
Sınıf 3	Nem (%)	6.43	0.70	15.70	2.36
	Kül (%)	35.48	23.05	50.33	3.19
	Alt Isıl Değer (kcal/kg)	4113	2927	5405	415
Sınıf 4	Nem (%)	6.58	1.00	19.70	2.31
	Kül (%)	25.11	16.20	39.07	3.17
	Alt Isıl Değer (kcal/kg)	4969	29	5904	387
Sınıf 5	Nem (%)	6.35	0.90	24.60	2.62
	Kül (%)	18.07	6.60	42.45	6.42
	Alt Isıl Değer (kcal/kg)	5619	3456	7009	638

5.2. İndikatör Krigleme ile Litolojilerin Kestirimi

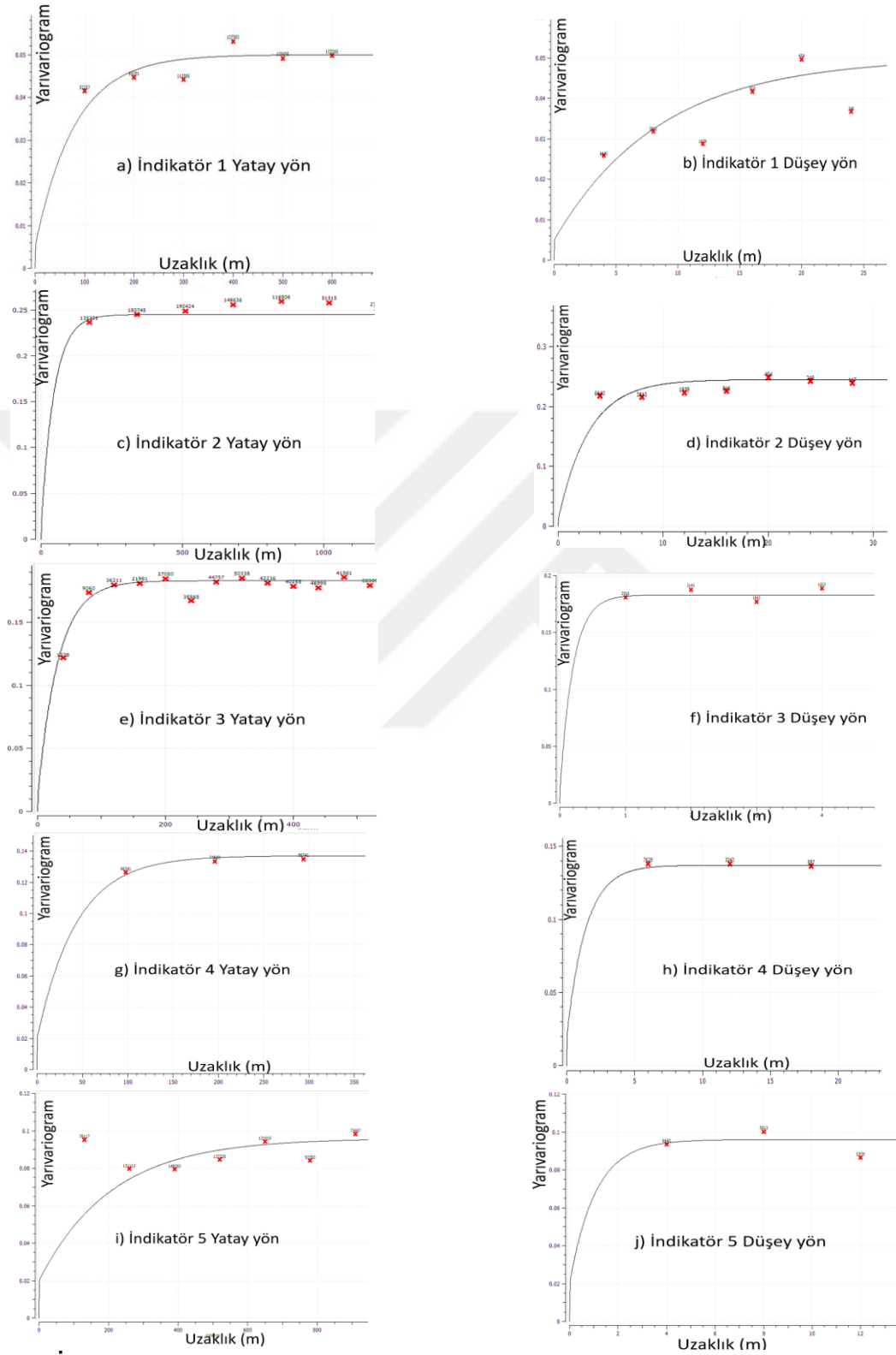
İndikatör Krigleme yöntemiyle litolojilerin tayini amacıyla İğdekuzu veri kümesindeki litolojik birimler indikatörlere dönüştürülmüştür. İlk olarak, sınıf 1’de yer alan (Marnlı Linyit) verilere 1 değeri, sınıf 2, 3, 4 ve 5 olan verilere ise 0 değeri atanarak birinci indikatör; daha sonra sınıf 2’de yer alan (Çok Killi Linyit)) verilere 1 değeri, sınıf 1, 3, 4 ve 5 olan verilere 0 değeri atanarak ikinci indikatör belirlenmiş ve bütün litolojik sınıflar için aynı işleme devam edilerek her bir sınıfa karşılık gelen 5 farklı indikatör veri kümesi oluşturulmuştur.

Litolojik birimlere karşılık gelen 5 indikatör için hesaplanan ve modellenen variogramlara ilişkin parametreler Çizelge 12’de gösterilmiştir.

Çizelge 12. İğdekuzu linyit sınıfı indikatörlerinin variogram modeli parametreleri

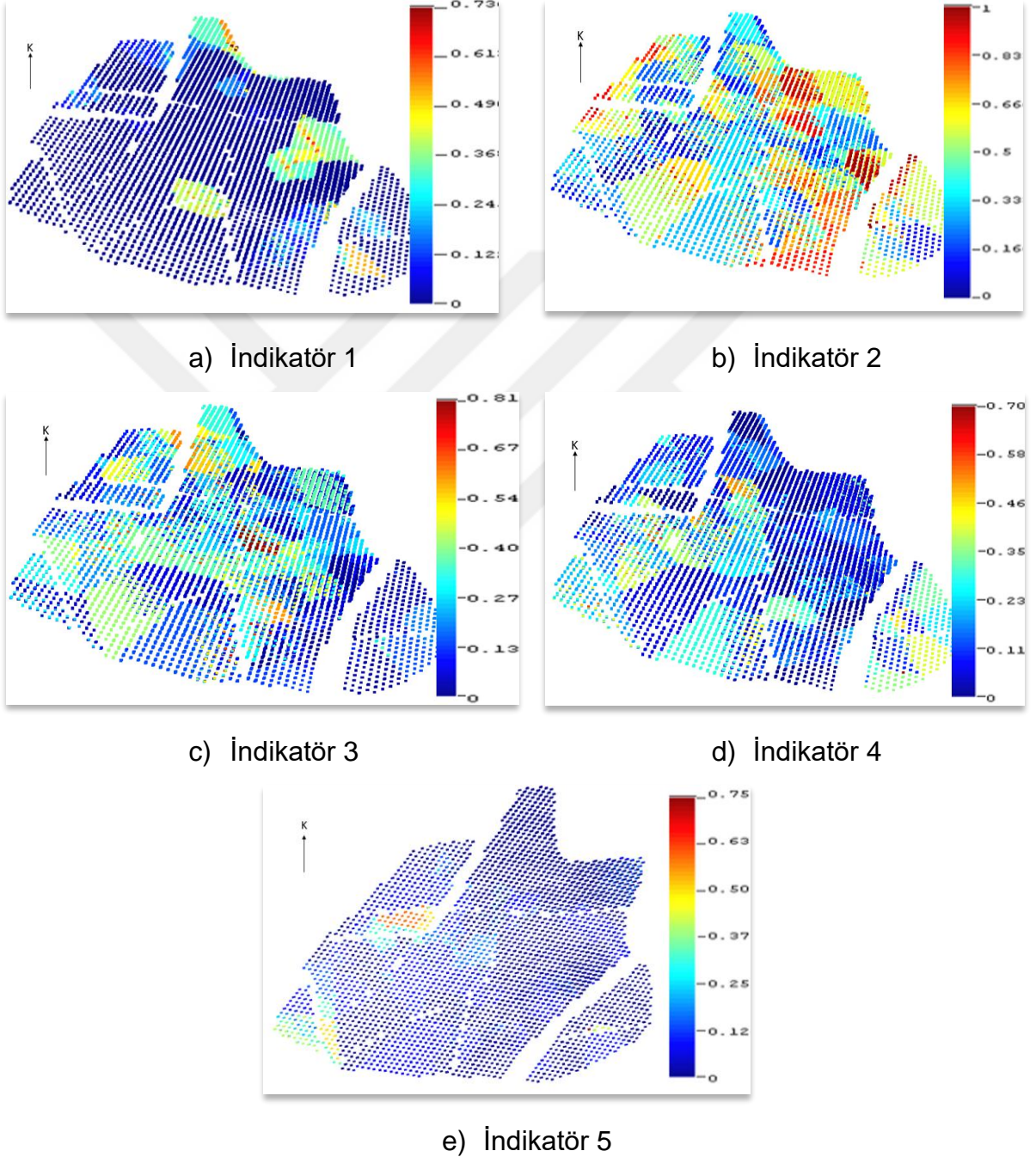
Sınıf	Yön	Külçe etkisi (C ₀)	Eşik değer (C)	Etki mesafesi (a)	Model
Marnlı Linyit	Yatay	0.005	0.045	240	Üssel
	Düşey	0.005	0.045	25.6	Üssel
Çok Killi Linyit	Yatay	0.01	0.235	136	Üssel
	Düşey	0.01	0.235	8.8	Üssel
Orta Killi Linyit	Yatay	0.01	0.173	96	Üssel
	Düşey	0.01	0.173	0.6	Üssel
Az Killi Linyit	Yatay	0.02	0.117	110	Üssel
	Düşey	0.02	0.117	3	Üssel
Temiz Linyit	Yatay	0.02	0.076	624	Üssel
	Düşey	0.02	0.076	3.2	Üssel

Şekil 20’de İğdekuzu veri kümesi litolojik sınıf indikatörleri için hesaplanan ve modellenen variogramlar gösterilmiştir.



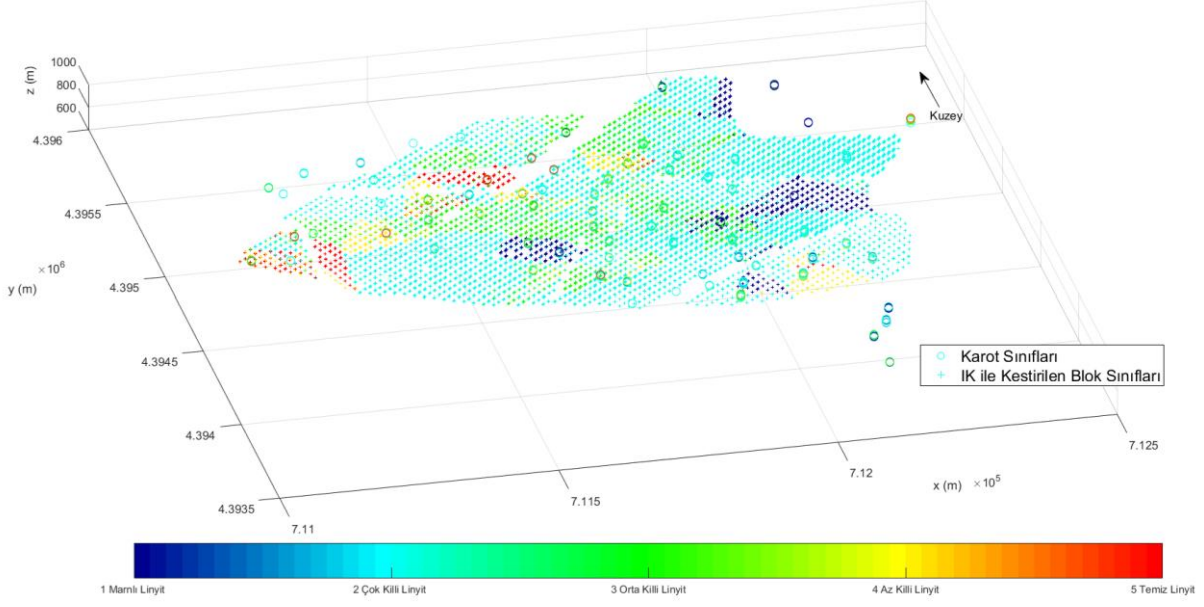
Şekil 20. İğdekuzu veri kümesi indikatörlerinin deneysel variogramları ve variogram modelleri

Şekil 20'de verilen variogramlar kullanılarak herhangi bir örneklemin bulunmadığı linyit zonu içindeki $20 \times 20 \times 1 \text{ m}^3$ 'lük blokların litolojik birimleri İndikatör Krigleme ile kestirilmiştir. Kestirim haritaları Şekil 21'de verilmiştir.



Şekil 21. İğdekuzu linyit sahası $20 \times 20 \times 1 \text{ m}^3$ 'lük bloklara ilişkin İndikatör Krigleme ile kestirilen linyit sınıfları haritası

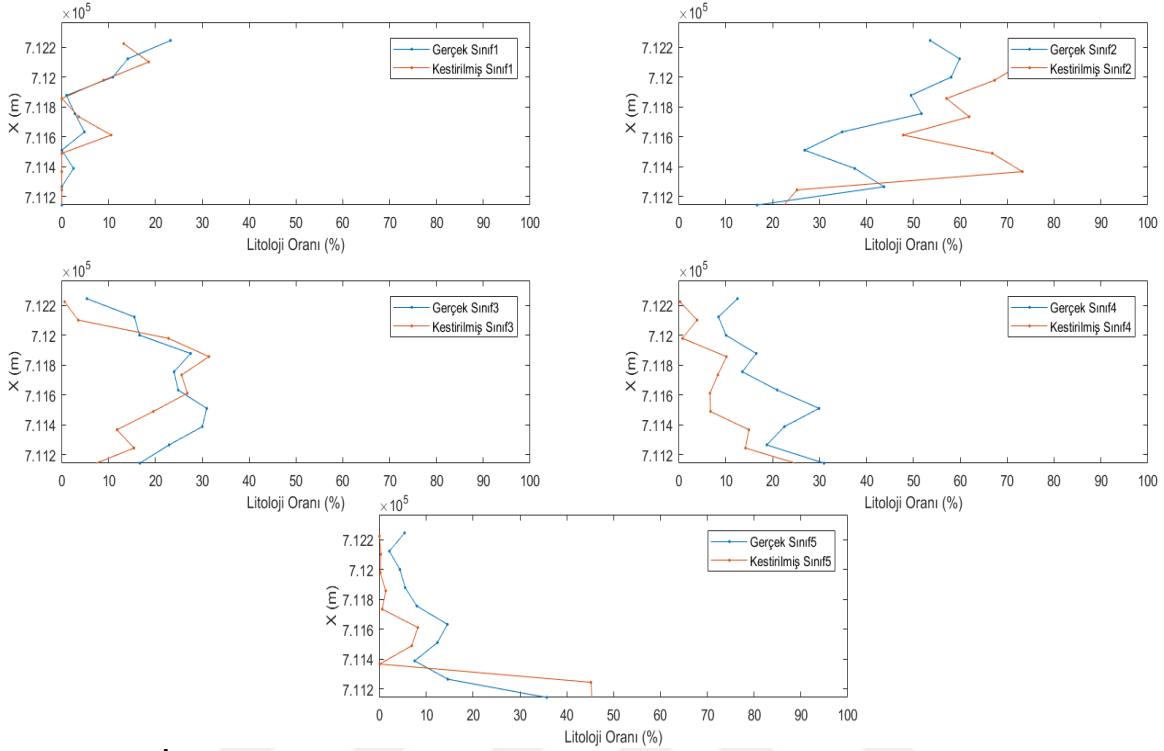
Her bir lokasyonda İndikatör Krigleme yöntemi ile yapılan beş indikatör kestiriminden maksimum olanı ilgili lokasyonun litolojisi olarak belirlenmiştir. Kestirilen sınıf değerlerine ilişkin lokasyon haritası Şekil 22’de verilmiştir.



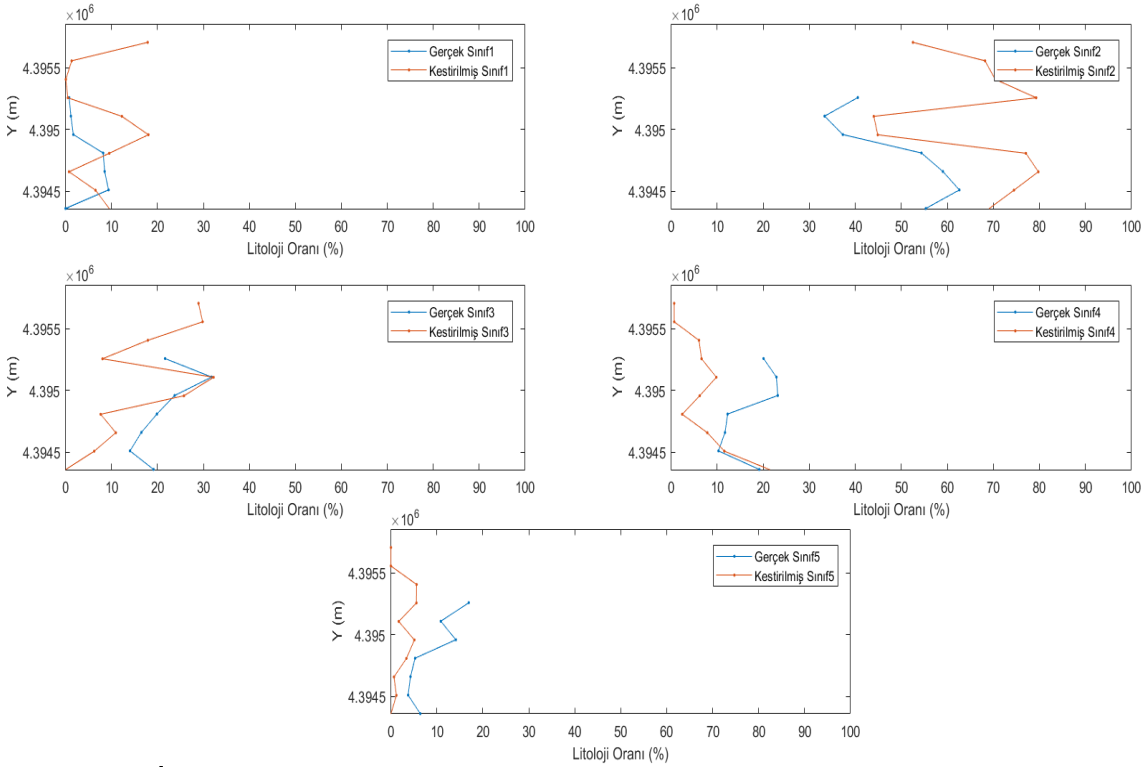
Şekil 22. İğdekuzu linyit sahası için oluşturulan 20 x 20 x 1 m³’lük blokların İndikatör Krigleme yöntemiyle belirlenen litolojik sınıfları

İK ile yapılan sınıflandırmada, sahanın batısında, orta kısımlarında ve güney doğusunda daha az kil içeren litolojik birimlerin bulunduğu Şekil 22’de görülmektedir.

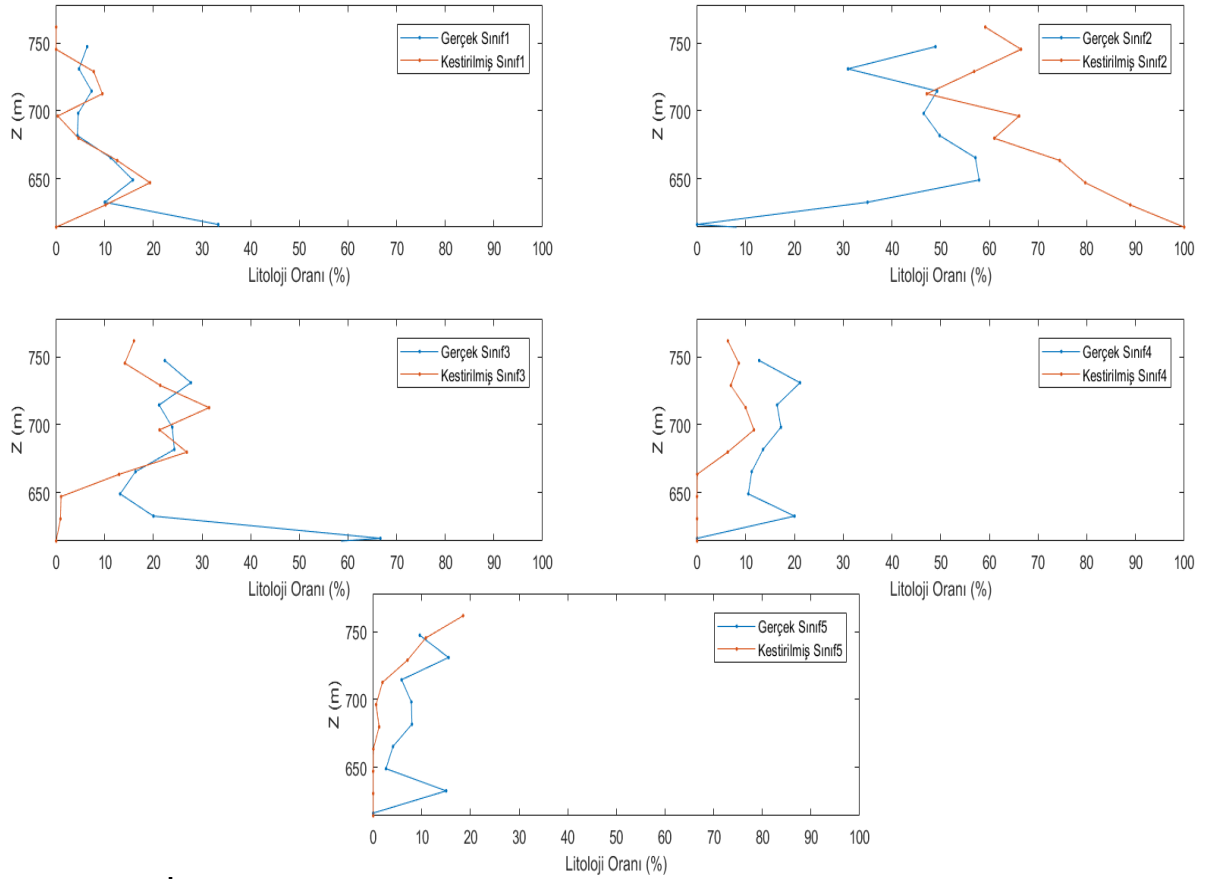
Litolojik sınıfların X, Y ve Z yönlerindeki değişimi Şekil 23’te gösterilmiştir. İK ile yapılan sınıflandırmalarda bütün yönlerde; Çok Killi Linyit ve Az Killi Linyit litolojik sınıflarının orijinal verilerdeki sınıflara kıyasla daha fazla, Temiz Linyit sınıfının ise daha az oranla kestirildiği görülmektedir. Gerçek verilerin litolojik sınıf oranlarına en yakın kestirimler Marnlı Linyit ve Orta Killi Linyit sınıflarında yapılmıştır.



Şekil 23 a. İğdekuzu linyit sahasının orijinal ve IK ile kestirilen sınıflarının doğu - batı (X) yönündeki değişim grafikleri (Sınıf 1: Marnlı Linyit, Sınıf 2: Çok Killi Linyit, Sınıf 3: Orta Killi Linyit, Sınıf 4: Az Killi Linyit, Sınıf 5: Temiz Linyit)



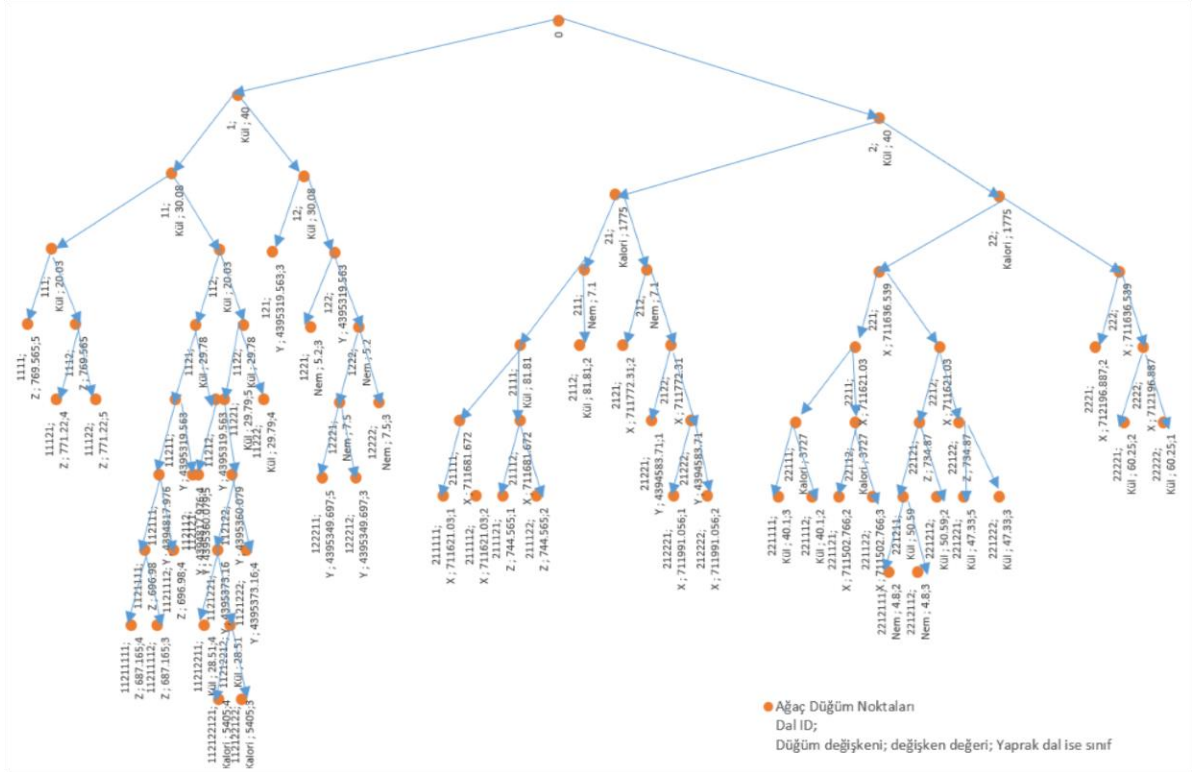
Şekil 23 b. İğdekuzu linyit sahasının orijinal ve IK ile kestirilen sınıflarının kuzey-güney (Y) yönündeki değişim grafikleri



Şekil 23 c. İğdekuzu linyit sahasının orijinal ve IK ile kestirilen sınıflarının düşey (Z) yönündeki değişim grafikleri

5.3. ROKY ile Litolojilerin Kestirimi

İğdekuzu veri kümesinde alt ısı değer, kül ve nem içerikleri ile X, Y, Z koordinat değerleri bağımsız değişkenlere ve litolojik sınıflar ise bağımlı değişkene karşılık gelmektedir. Değişkenler arasındaki ilişkiyi tanımlamak için 100 adet ağaç ile yapılandırılan Rastlantı Ormanı oluşturulmuştur. Ormanın 65'inci ağaç yapısı Şekil 24'te verilmiştir. Rastlantı Ormanının seçilmemiş veri kestirim hatası %3.01'dir.

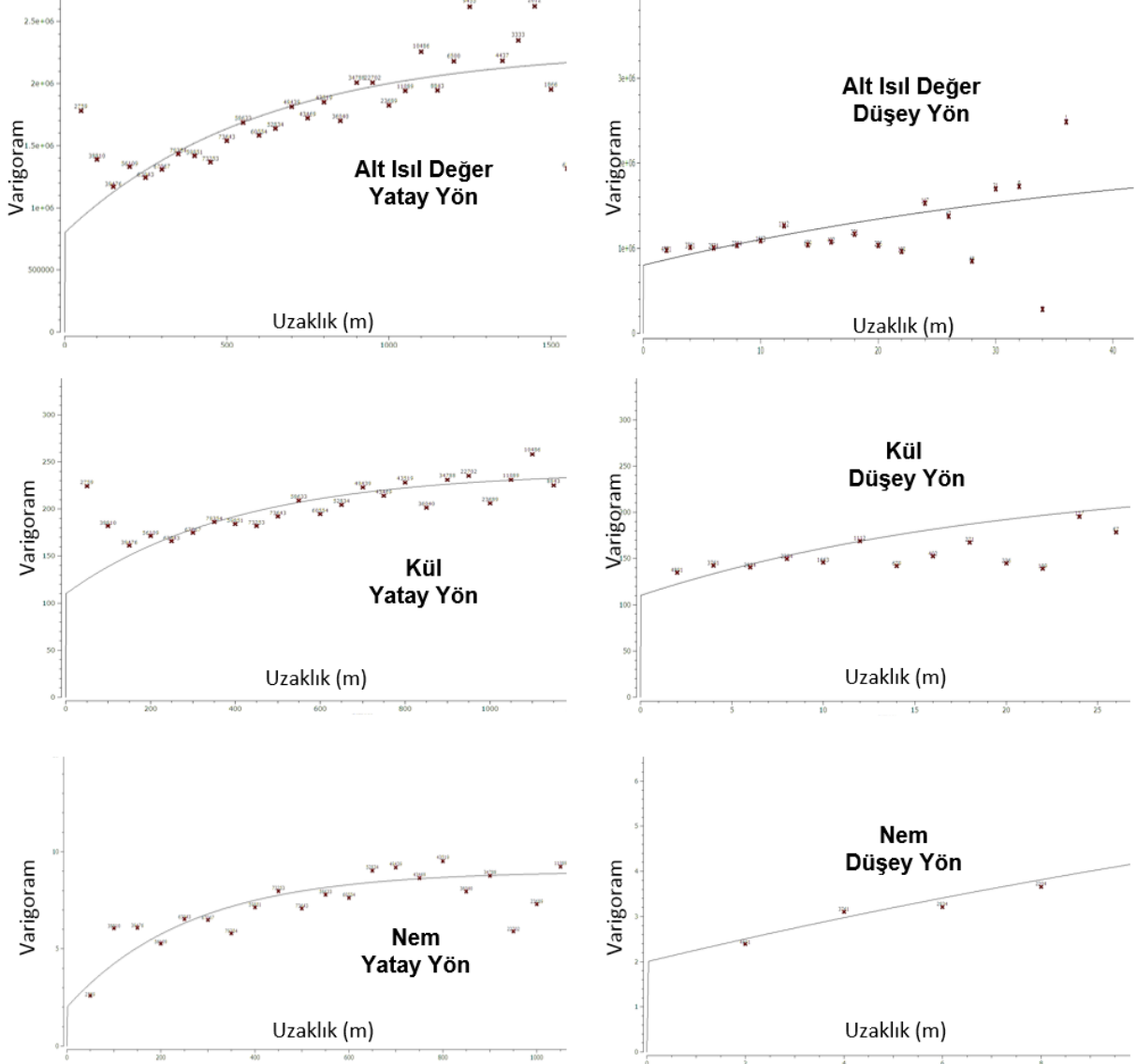


Şekil 24. İğdekuzu linyit sahası ham verileriyle oluşturulan Rastlantı Ormanında 65. ağaç yapısı

İğdekuzu veri kümesinin alt ısıl değer, kül ve nem değişkenleri için oluşturulan variogram modelleri Şekil 25'te bu modelin belirlenmesinde kullanılan parametreler ise Çizelge 13'te verilmiştir.

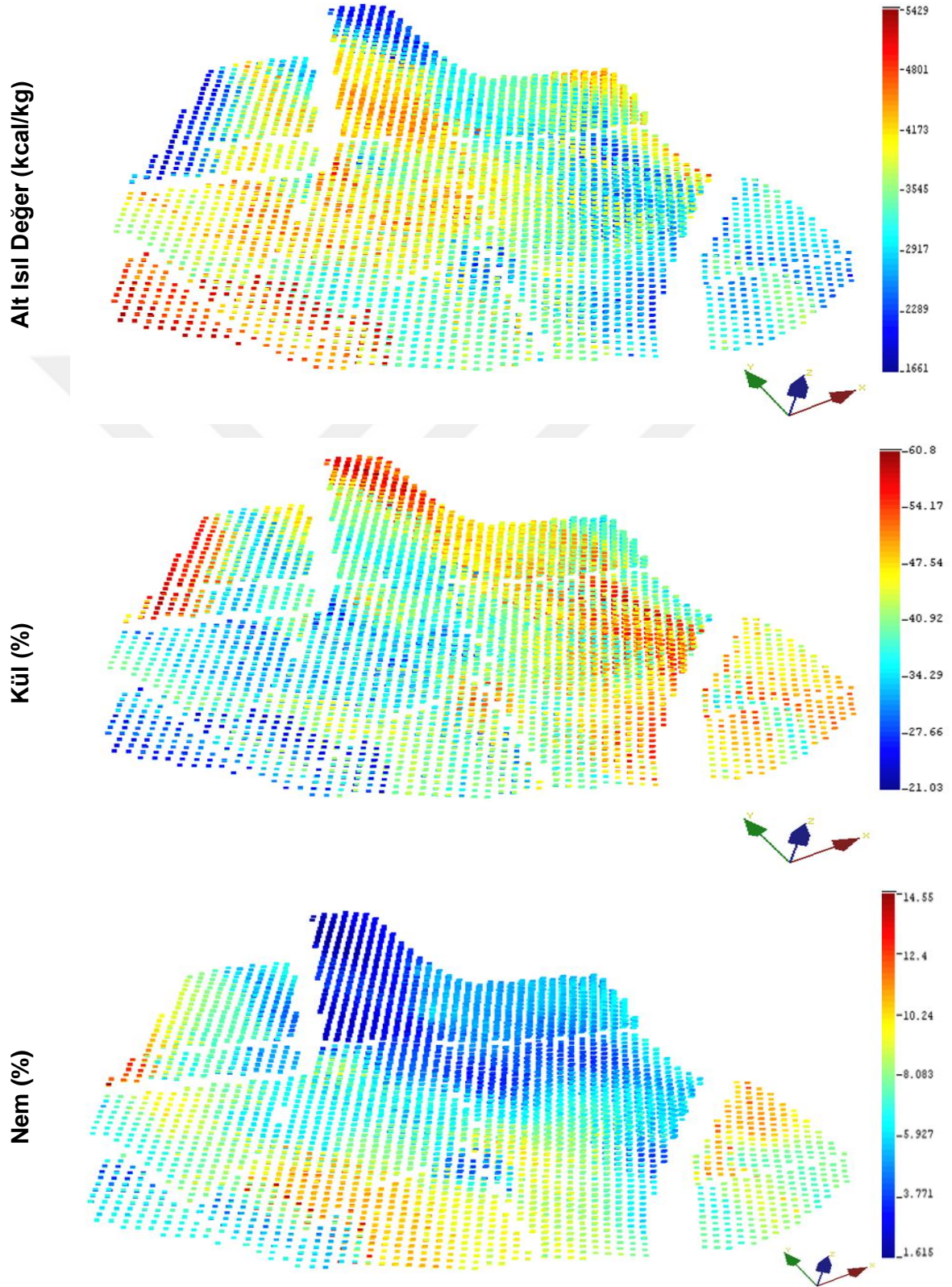
Çizelge 13 İğdekuzu veri kümesi variogram parametreleri

Değişken	Yön	Külçe Etkisi	Eşik Değeri	Etki Mesafesi (m)	Model
Alt Isıl Değer	Yatay	800000	1480000	1800	Üssel
	Düşey			132	
Kül	Yatay	110	130	1200	Üssel
	Düşey			60	
Nem	Yatay	2	7	780	Üssel
	Düşey			80	



Şekil 25. İğdekuzu veri kümesi alt ısı değer, kül ve nem değişkenleri için hesaplanan yarıvariogramlar ve variogram modelleri

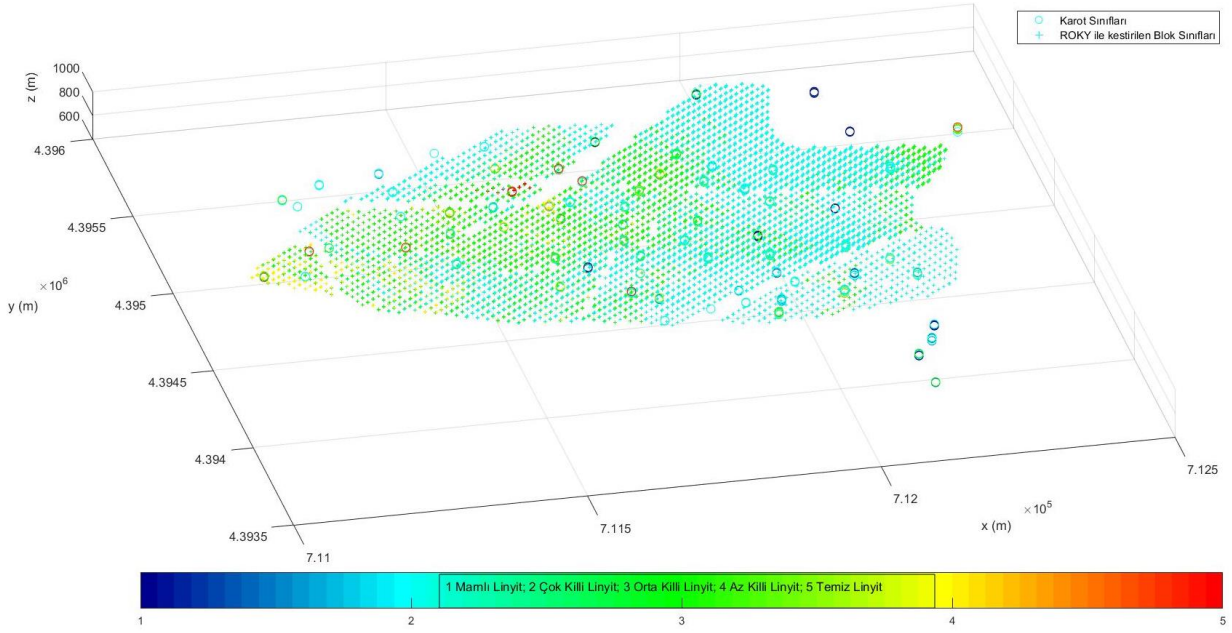
İğdekuzu linyit sahasında linyit horizonunu kapsayan 20 x 20 x 1 m³lük blokların alt ısı değer, kül ve nem içerikleri Ortalamasız Krigleme yöntemi ile kestirilmiştir. Yapılan kestirimler şekil 26'da gösterilmiştir.



Şekil 26. İgdekuzu linyit sahası için oluşturulan 20x20x1 m³lük blokların Ortalamasız Krigleme ile kestirilen alt ısıl değer, kül ve nem değişkenlerinin konumsal dağılımı

Sahada alt ısı değer değişkenine ilişkin en yüksek kestirim değerleri ile kül değişkenine ilişkin en düşük kestirim değerlerinin sahanın batısı ile orta-kuzey kısımlarında toplandığı; nem değişkenine ilişkin ise en yüksek kestirim değerlerinin sahanın güneyinde ve kuzey batısında toplandığı Şekil 26'da görülmektedir.

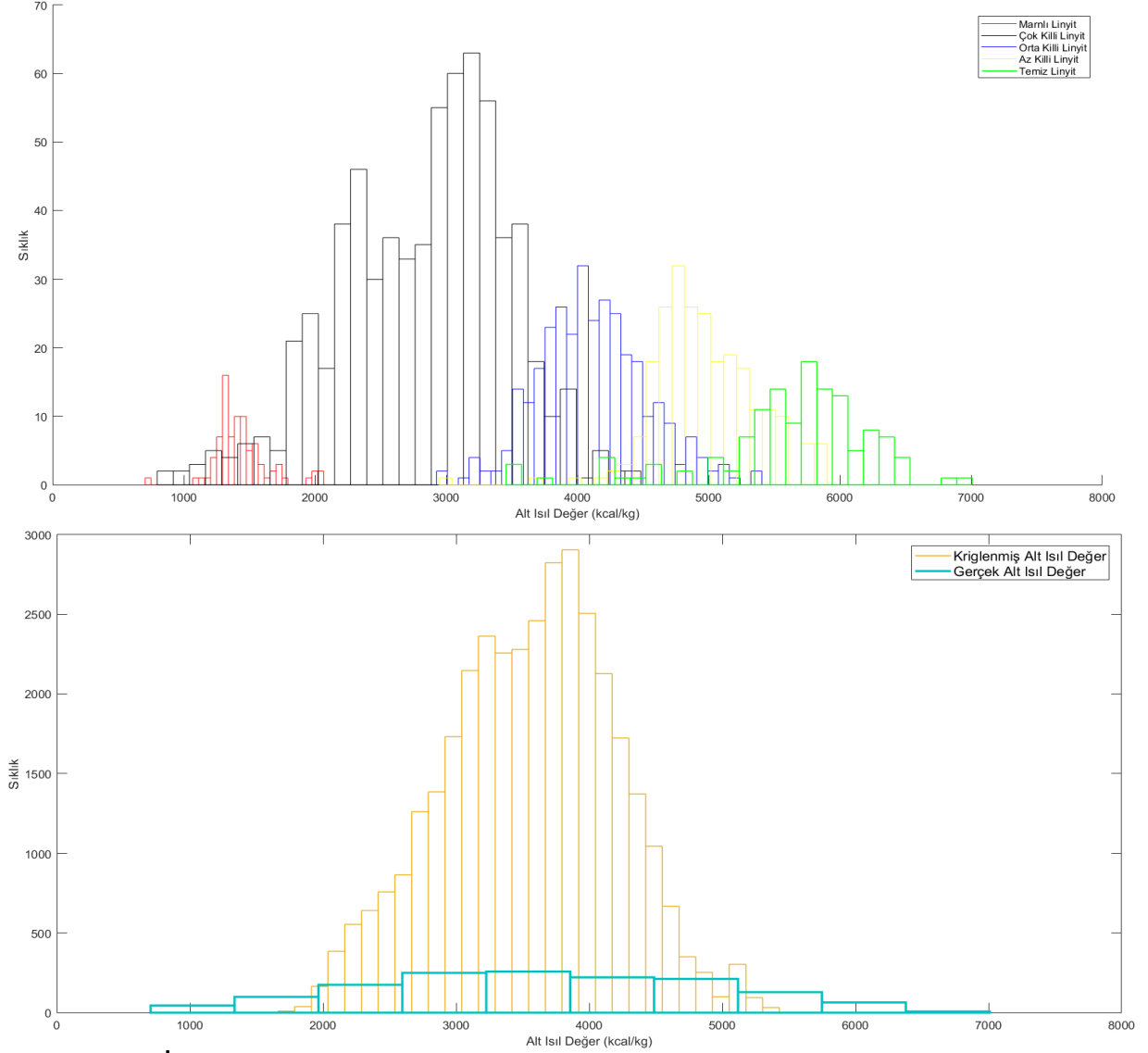
Şekil 26 'da gösterilen bloklara atanmış kestirim değerleri ham verilerle oluşturulan Rastlantı Ormanına beslenerek blokların litolojik sınıfları tayin edilmiştir. Hesaplanan litolojik sınıflar Şekil 27'de gösterilmiştir.



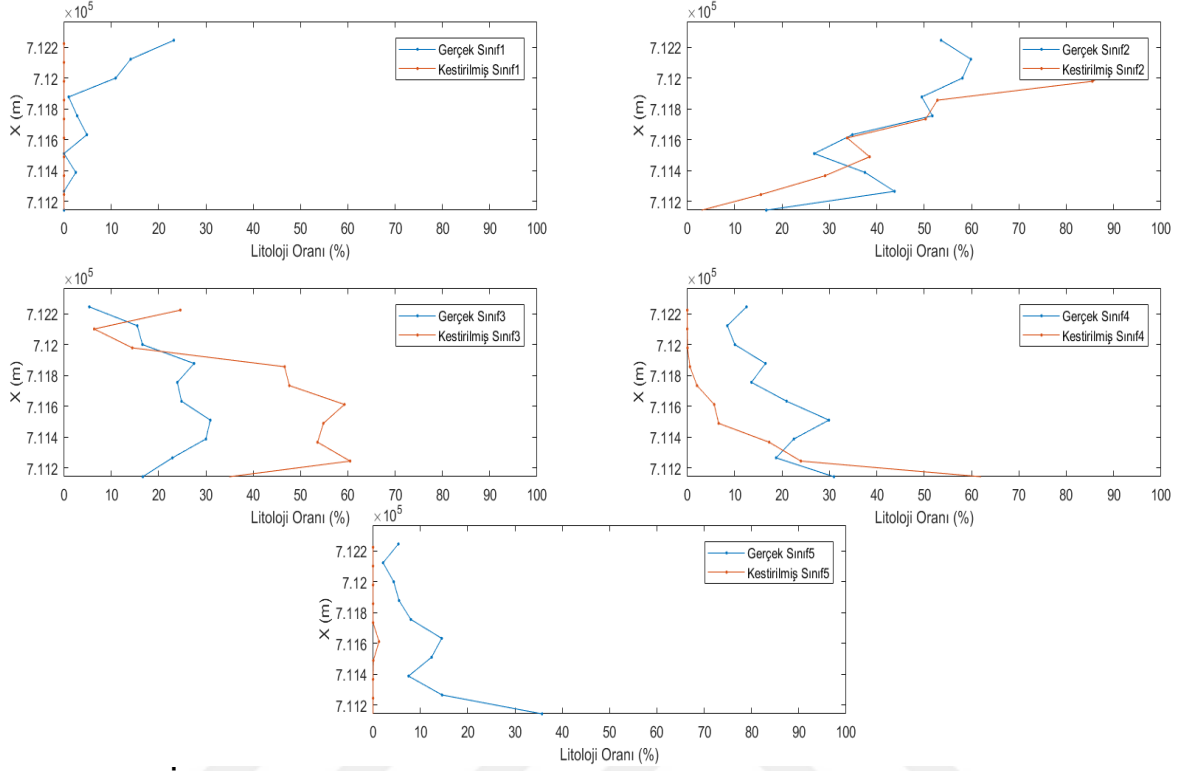
Şekil 27. İğdekuzu linyit sahası için oluşturulan 20 x 20 x 1 m³'lük blokların ROKY ile belirlenen litolojik sınıfları.

ROKY ile alt ısı değer değişkeninin yüksek, kül değişkeninin düşük kestirildiği bölgelerde daha az kil içeriği olan litolojik sınıfların hesaplandığı görülmektedir. Ayrıca ROKY'nin en yüksek alt ısı değer değişken değeri ve en düşük kül değişken değerlerinin toplandığı Temiz Linyit litolojisinde ve aynı değişkenler için sırasıyla en düşük ve en yüksek değerlerin toplandığı Marnlı Linyit litolojisinde neredeyse hiç sınıf hesaplamadığı görülmektedir. Bunun sebebinin anlaşılması için Şekil 28'de verilen gerçek İğdekuzu veri kümesi alt ısı değer histogramı ile Kriglenmiş alt ısı değer histogramını incelemek yerinde olacaktır. Şekilde sınıflara göre orijinal alt ısı değer histogramları incelendiğinde Marnlı Linyit litolojisine sahip verilerin en düşük; Temiz Linyit litolojisine sahip verilerin en yüksek alt ısı değerlere sahip olduğu görülebilir.

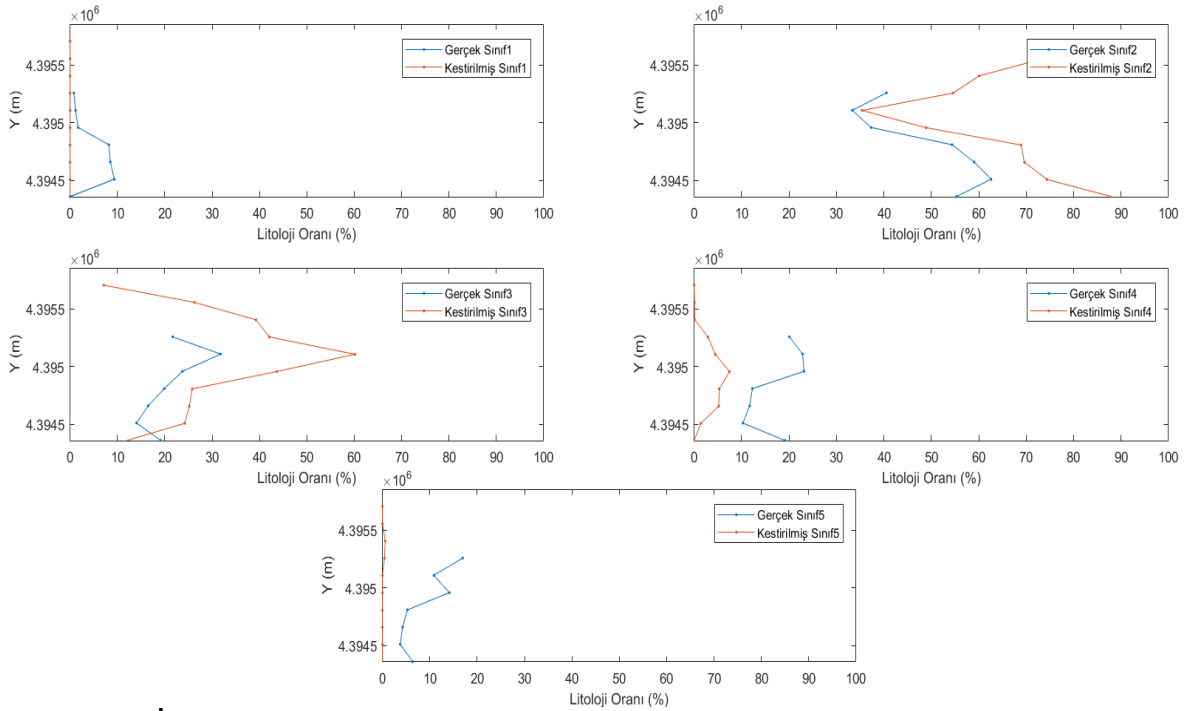
Ancak kestirim verilerinin histogramından anlaşılacağı üzere Krigleme yöntemi verileri törpülemiş (smoothing), en yüksek ve en düşük değerleri ortalamalara yaklaştırmıştır. ROKY ile Marnlı Linyit ve Temiz Linyit litolojisinde sınıf tayin edilememesi, kriglemenin törpüleme etkisinden kaynaklanmaktadır.



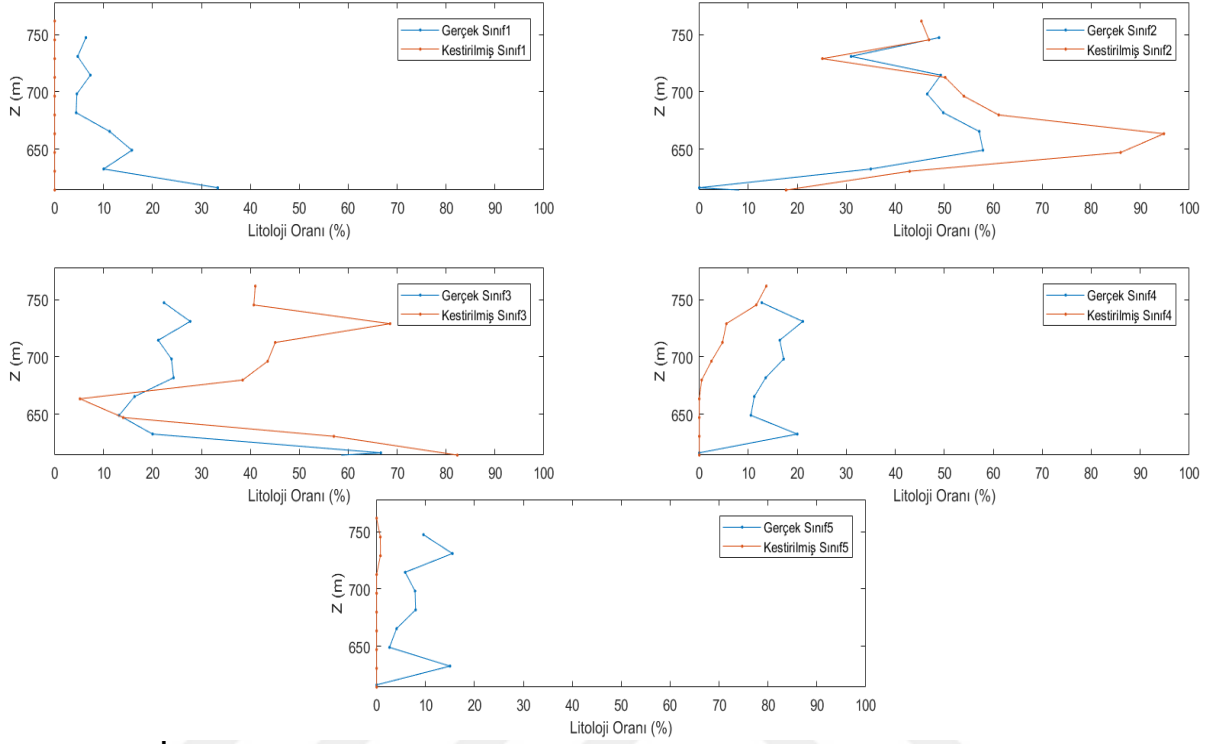
Şekil 28. İğdekuzu linyit sahası gerçek ve kriglenmiş verilerin alt ısıl değer histogramları. (Üstteki grafikte, gerçek değerlerin sınıflara göre histogramı, alttaki grafikte ise gerçek ve kriglenmiş alt ısıl değerlerin histogramı gösterilmiştir.)



Şekil 29 a. İğdekuzu linyit sahasının orijinal ve ROKY ile hesaplanan sınıflarının doğu-batı (X) yönündeki değişim grafikleri (Sınıf 1: Marnlı Linyit, Sınıf 2: Çok Killi Linyit, Sınıf 3: Orta Killi Linyit, Sınıf 4: Az Killi Linyit, Sınıf 5: Temiz Linyit)



Şekil 29 b. İğdekuzu linyit sahasının orijinal ve ROKY ile hesaplanan sınıflarının kuzey-güney (Y) yönündeki değişim grafikleri



Şekil 29 c. İğdekuzu linyit sahasının orijinal ve ROKY ile hesaplanan sınıflarının düşey (Z) yöndeki değişim grafikleri

Orijinal veri ve ROKY ile hesaplanan sınıf oranlarının X, Y ve Z yönlerinde değişimi Şekil 29 a, 29 b ve 29 c'de gösterilmiştir. ROKY ile Çok Killi Linyit litolojisi dışındaki bütün litolojik sınıflarda gerçek sınıf litolojilerinden farklı oranlarda sınıf değerleri hesaplanmıştır. Özellikle Marnlı Linyit ve Temiz Linyit sınıfları neredeyse hiç kestirilmemiştir

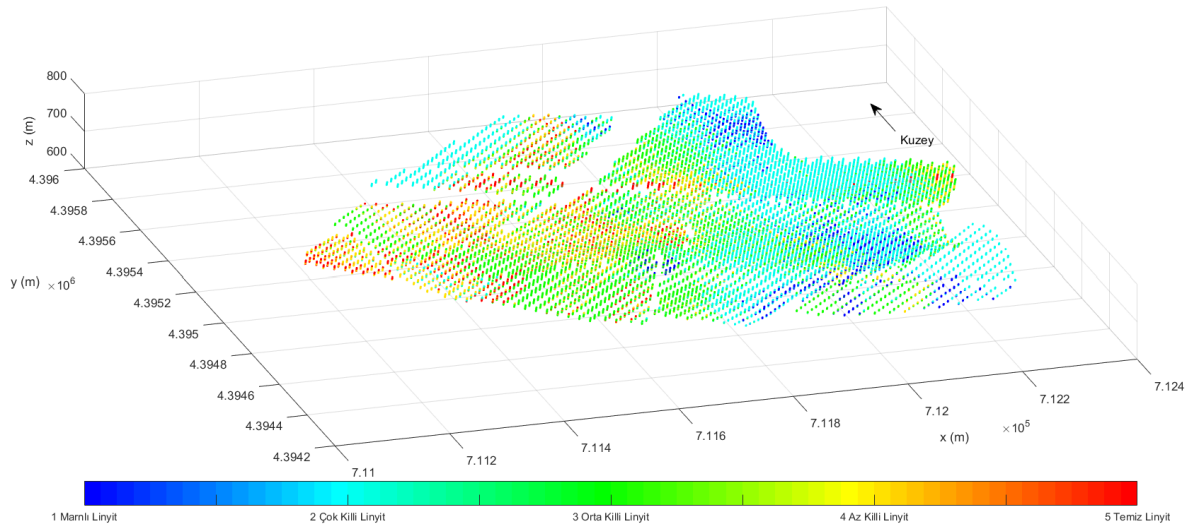
5.4. g-ROKY ile Litolojilerin Kestirimi

Üç kalite değişkenine ilişkin 20 x 20 x 1 m boyutundaki blokların kestirim değerleri ve karot analiz değerleri standartlaştırılarak gruplandırılmıştır. Çizelge 14'te standartlaştırılmış kalite değişkenlerine ilişkin her bir gruba karşılık gelen aralıklar verilmiştir.

Çizelge 14 İğdekuzu veri kümesi ve kriglenmiş blokların standartlaştırılmış bağımsız değişken grupları

Grup	Nem (%)		Kül (%)		Alt Isıl Değer (kcal/kg)	
1	<	-1.60	<	-2.46	<	-2.34
2	-1.60	-1.00	-2.46	-1.89	-2.34	-1.76
3	-1.00	-0.39	-1.89	-1.31	-1.76	-1.18
4	-0.39	0.21	-1.31	-0.74	-1.18	-0.59
5	0.21	0.82	-0.74	-0.16	-0.59	-0.01
6	0.82	1.42	-0.16	0.41	-0.01	0.57
7	1.42	2.03	0.41	0.98	0.57	1.16
8	2.03	2.63	0.98	1.56	1.16	1.74
9	2.63	3.24	1.56	2.13	1.74	2.32
10	3.24	>	2.13	>	2.32	>

Bağımsız değişkenleri Çizelge 14'te gösterilen şekilde gruplandırılan İğdekuzu linyit sahası sondaj verileri ile 100 adet ağaç yapısından oluşan Rastlantı Ormanı oluşturulmuştur. Oluşturulan RO'nun seçilmemiş veri sınıf kestirimi hatası %11.5'tir. g-ROKY ile hesaplanan Jura doğrulama veri kümesi sınıfları Şekil 30'da verilmiştir. Şekil 30'dan görüleceği üzere g-ROKY ile yapılan sınıf hesaplamasında sahanın orta ve batı kısımlarında kil içeriği az olan sınıflar; kuzey batı, kuzey doğu, doğu ve güney doğu kısımlarında ise kil içeriği yüksek olan litolojik sınıflar hesaplanmıştır.



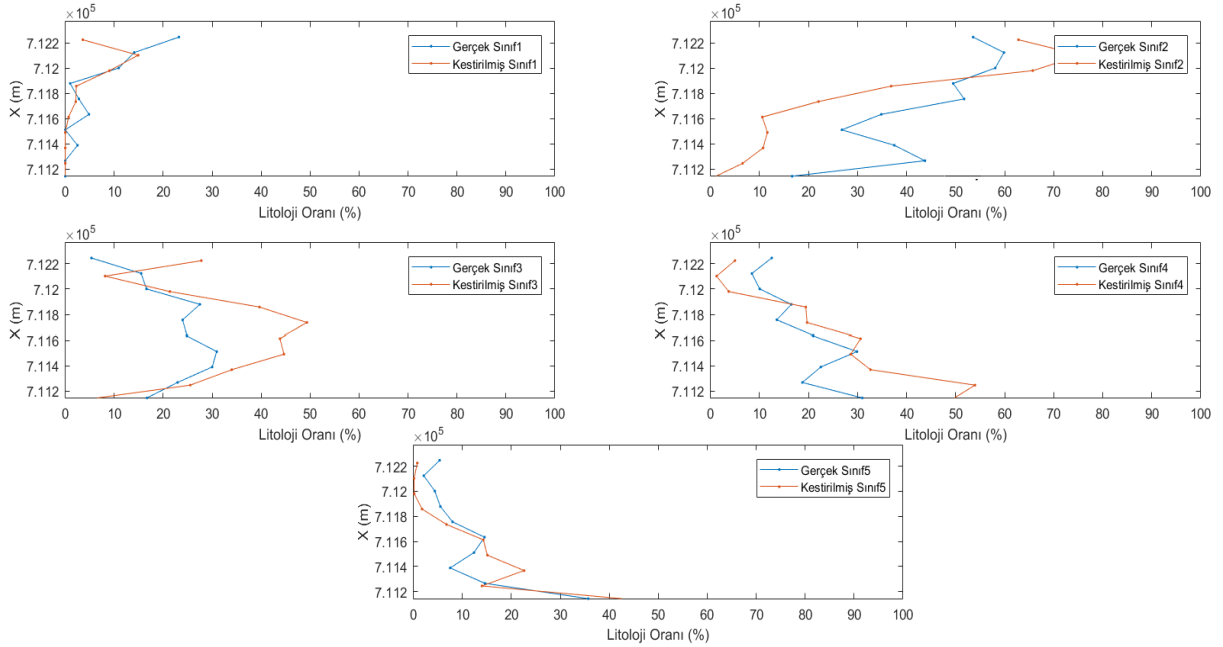
Şekil 30. İğdekuzu linyit sahası için oluşturulan 20 x 20 x 1 m³'lük blokların g-ROKY ile belirlenen litolojik birimleri

Sahanın tümü için yönden bağımsız olarak kestirilen sınıf oranları orijinal verilerin sınıf oranlarıyla karşılaştırıldığında g-ROKY'nin gerçek sınıf oranlarına en yakın kestirimi yaptığı Çizelge 15'te görülmektedir.

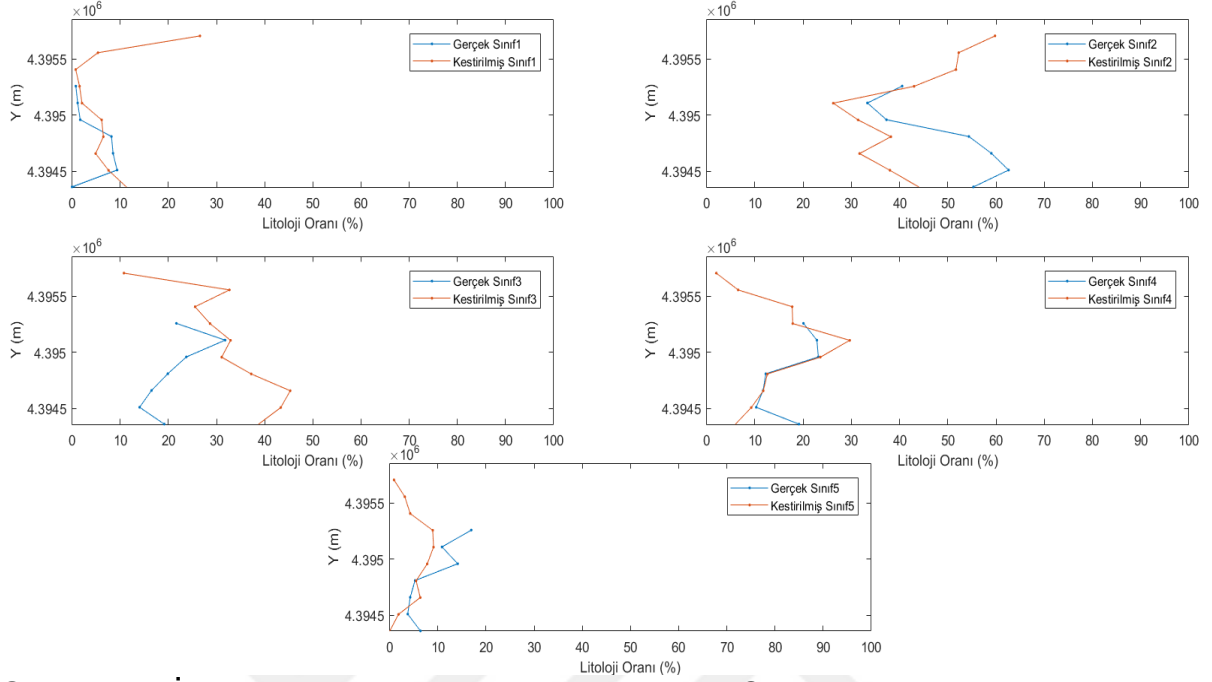
Çizelge 15. IK, ROKY ve g-ROKY ile hesaplanan sınıf oranlarının karşılaştırması

Kestirim Türü	Sınıf 1 Oranı (%)	Sınıf 2 Oranı (%)	Sınıf 3 Oranı (%)	Sınıf 4 Oranı (%)	Sınıf 5 Oranı (%)
Orijinal Veri	5.75	46.20	22.45	16.50	9.10
IK	7.42	65.69	17.70	5.87	3.32
ROKY	0.00	60.06	36.20	3.58	0.15
g-ROKY	5.77	38.79	33.55	13.98	7.91

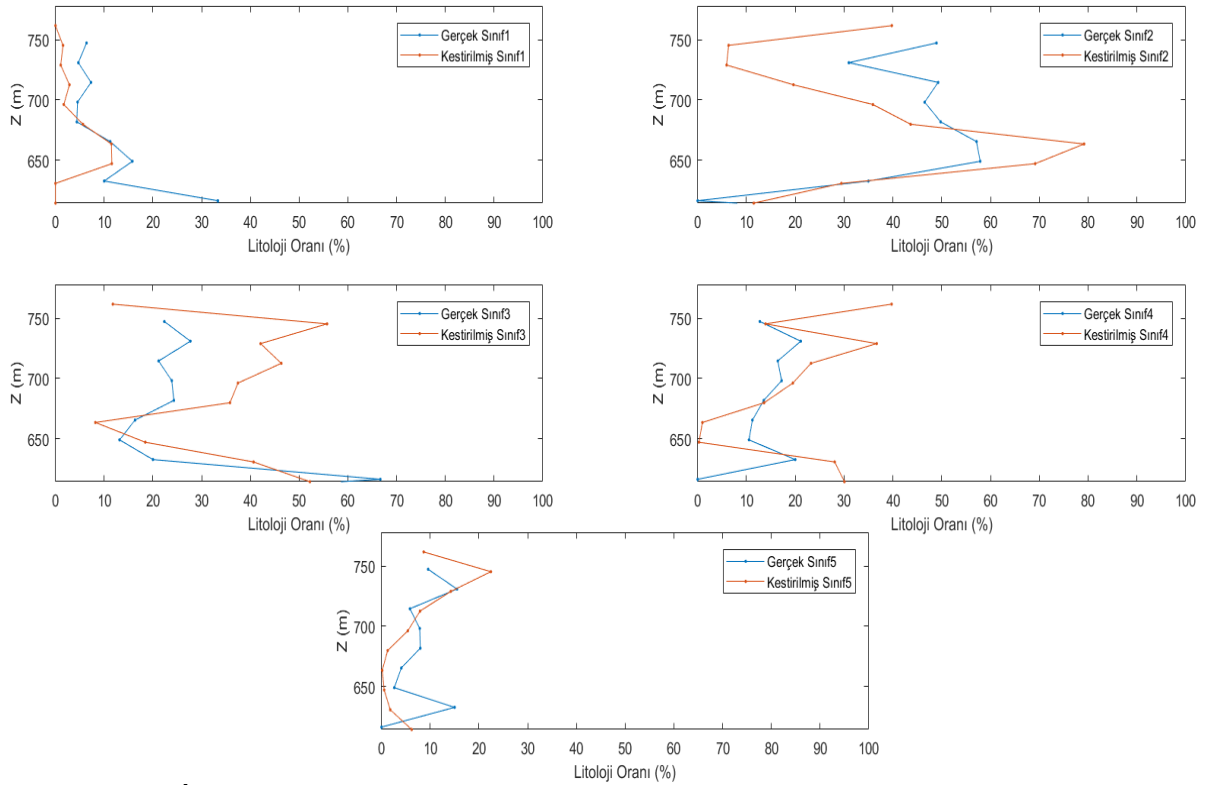
Orijinal veri ve g-ROKY ile hesaplanan sınıf oranlarının X, Y ve Z yönlerinde değişimi Şekil 31'de gösterilmiştir. g-ROKY ile bütün yönlerde, Orta Killi Linyit litolojisindeki sınıfların gerçek değerlerden daha fazla oranda, Çok Killi Linyit litolojisindeki sınıfların ise daha az oranda hesaplandığı görülmektedir. ROKY ile Temiz Linyit ve Marnlı Linyit litolojilerine ilişkin neredeyse hiçbir sınıflandırma yapılmamış olmasına rağmen; g-ROKY, kriglemenin törpüleme etkisini gizlemiş ve bahsedilen sınıfları tespit edebilmiştir. IK ve g-ROKY'nin sınıflandırmaları karşılaştırıldığında IK'nin yalnız Orta Killi Linyit litolojisinde gerçek sınıf oranlarına daha yakın sonuçlar ürettiği, g-ROKY'nin ise bütün yönlerde Az Killi Linyit ve Temiz Linyit sınıflarını gerçek sınıf değerlerine daha benzer oranlarda tespit ettiği görülmektedir.



Şekil 31 a. İğdekuzu linyit sahasının orijinal ve g-ROKY ile hesaplanan sınıflarının doğu-batı (X), yönündeki değişim grafikleri



Şekil 31 b. İğdekuzu linyit sahasının orijinal ve g-ROKY ile hesaplanan kuzey-güney (Y) yönündeki değişim grafikleri



Şekil 31 c. İğdekuzu linyit sahasının orijinal ve g-ROKY ile hesaplanan sınıflarının düşey (Z) yöndeki değişim grafikleri

6. SONUÇLAR ve ÖNERİLER

6.1. Sonuçlar

Tez kapsamında elde edilen sonuçlar:

1. İğdekuzu linyit sahası için ROKY ile yapılan sınıflandırmalarda kriglemenin törpüleme etkisinden dolayı Marnlı Linyit ve Temiz Linyit sınıflarının neredeyse hiçbir bloğa atanmadığı görülmektedir. Ancak g-ROKY, törpülemeden kaynaklı hatalı sınıflandırmalar açısından ROKY'den çok daha olumlu sonuçlar üretmiştir.
2. İğdekuzu linyit sahasında ve Jura veri kümesi üzerine yapılan çalışmalarda gerçek verilerin sınıflarına en yakın kestirim değerlerini g-ROKY üretmiştir.
3. Yeni geliştirilen yöntemlerden ROKY homojen ortamların belirlenmesinde İndikatör Krigleme yöntemine göre daha olumlu bir sonuç sağlayamasa da, g-ROKY hem İndikatör Kriglemeden hem de ROKY'den daha doğru sınıf hesaplamaları yapabilmektedir.
4. İndikatör Krigleme yöntemi ile ROKY ve g-ROKY'nin uygulanma zorlukları kıyaslanırken verideki bağımlı ve bağımsız değişkenlerin sayısı ve nitelikleri değerlendirilmelidir. Örneğin, 3 adet bağımsız değişken ve 5 adet farklı sınıf değeri alabilen bağımlı kategorik değişken içeren İğdekuzu veri kümesini ele alalım. İndikatör Krigleme ile yapılacak sınıf kestiriminde öncelikle 5 adet sınıf için indikatörler belirlenmelidir. Daha sonra bu indikatörler için 5 farklı variogram modelinin oluşturulması ve kriglemenin yapılması gerekmektedir. ROKY veya g-ROKY'de ise 3 adet bağımsız değişkenin variogram modelleri oluşturulması, bağımsız değişkenlerin krigleme ile kestirimlerinin yapılması ve kestirim değerlerinin Rastlantı Ormanına beslenmesi süreci izlenecektir. Dolayısıyla böyle bir örnek veri kümesinde variogram modelleme ve krigleme işlemleri açısından ROKY veya g-ROKY'nin uygulanması daha kolay olacaktır. Ancak Jura veri kümesinde görüldüğü üzere bağımsız değişken sayısı bağımlı sınıf değişkeninin alabileceği sınıf değerlerinden fazla ise durum tersine dönmektedir ve İndikatör Krigleme yönteminin uygulanması daha az zaman alan bir yöntem olmaktadır.
5. IK, ROKY ve g-ROKY'nin uygulanma süreleri karşılaştırıldığında veri sayısı, sınıflandırılacak veri sayısı ve RO'nun içereceği sınıflandırma ağacı sayısı etkili

olmaktadır. Üretilecek sınıflandırma ağacı sayısı ve veri sayıları arttıkça ROKY ve g-ROKY'deki işlem süreleri artmaktadır.

6.2. Öneriler

1. Geliştirilen g-ROKY yönteminde; grup sayısının, oluşturulan Rastlantı Ormanının seçilmemiş veri kestirim hatasını en aza indirgeyecek minimum değer olarak belirlenmektedir. Bu sebeple grup sayısı deneme yanılma yoluyla tespit edilmektedir. En uygun grup sayısının belirlenmesine yönelik geliştirilecek yeni yaklaşımlar yöntemin uygulanmasını kolaylaştıracaktır.
2. Sınıflandırmaya yönelik olumlu sonuçlar üreten g-ROKY ile bir maden yatağının homojen ortamlarının belirlenmesi ve ortam sınırlarında özneliklerin davranışlarının analiz edilmesi ilginç bir çalışma olacaktır.

KAYNAKLAR

- [1] Tercan, A. E., Ünver, B., Hindistan, M. A., Üç Boyutlu Cevher Yatağı Modelleme ve Ocak Tasarlama, *4. Ulusal Savunma Uygulamaları Modelleme ve Simülasyon Konferansı*, Haziran, Ankara, **2011**.
- [2] Tercan, A. E., Ünver, B., Hindistan, M. A., Erunç, G., Atalay, F., Ünal, S., Killioğlu, Y., Seam Modelling and Resource Estimation in Coalfields of Western Anatolia, *International Journal of Coal Geology*, 112, 94-106, **2013**.
- [3] Cowan, E. J., Spragg, K. J., Everitt, M. R., Wireframe-Free Geological Modelling – An Oxymoron or a Value Proposition, *Eighth International Mining Geology Conference*, Queenstown, New Zeland, 22-24 August 2011, **2011**.
- [4] Matheron, G., *Splines and Kriging - Their Formal Equivalence, Down-to-Earth Statistics – Solutions looking for geological problems*, pp. 77-95, Syracuse University Geology Contributions, **1980**.
- [5] Sinclair A. J., Blackwell, G. H., *Applied Mineral Inventory Estimation*, Cambridge University Press, **2002**.
- [6] Cheng, R. C. H., Traylor, L., Non-regular maximum likelihood problems, *Journal of the Royal Statistical Society, Series B, Statistical Methodology* 57, 3–4, **1995**
- [7] Emery, X., Oritz, J. M., Estimation of Mineral Resources Using Grade Domains: Critical Analysis and A Suggested Methodology, *The Journal of The South African Institute of Mining and Metallurgy*, Volume 105, April 2005, **2005**.
- [8] Leuangthong, O., Srivastava, R. M., On the use of multigaussian kriging for grade domaining in mineral resource characterization, *Proceedings of the 9th International Geostatistics Congress*, Oslo, Norway, 10–15 June 2012. 14 pp., **2012**.
- [9] Romary, T., Rivoirard, J., Deaisme, J., Quinones, C., Freulon, X., Domaining by Clustering Multivariate Geostatistical Data, *Quantitative Geology and Geostatistics* 17, Oslo, **2012**.
- [10] Mariz, C., Wagner, L., Data classification using geostatistical hierarchical clustering for robust and dynamic domaining, *Geomet 2016*, 3rd International Seminar on Geometallurgy, pp.80-81, **2016**.
- [11] Krige, D. G., A Statistical Approaches to Some Basic Mine Valuation Problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52, 119-139, **1951**.
- [12] Li, J., Heap, A. D., *A Review of Spatial Interpolation Methods for Environmental Scientist*, Australia Department of Resources, Energy and Tourism Minister for Resources and Energy, Sayfa 11, **2008**.
- [13] Krige, D. G., *Lognormal-de Wijisian Geostatistics for Ore Evaluation*, South African Institue of Mining and Metallurgy, Sayfa 25, **1981**.
- [14] Matheron, G., *Traite degeostatistique appliquee*, tome 1.MemoiresdeBureaude RecherchesGeologiqueset Minieres, no. 14, EditionsTechnip, Paris, Englishtranslation:

Treatise on Applied Geostatistics, Kennecott Copper Corporation, Salt Lake City, Utah, **1960**.

[15] Isaacs, E. H. and Srivastava, M. R., *An Introduction to Applied Geostatistics*, Oxford University press, 561p, **1989**.

[16] Tercan, A. E, Saraç, C., *Maden Yataklarının Değerlendirilmesinde Jeostatistiksel Yöntemler*, Jeoloji Mühendisleri Odası Yayınları, **1998**.

[17] Webster, R. and Oliver, M. A., *Geostatistics For Environmental Scientists*, John Wiley&Sons, Chichester, England, **2001**.

[18] Armstrong, M., *Basic linear geostatistics*, Springer, 153p, **1998**.

[19] Journel, A. G., Nonparametric Estimation of Spatial Distributions, *Journal of the International Association for Mathematical Geology*, June 1983, Volume 15, pp 445-468, **1982**.

[20] Liu R. M., Wang X. J., Zhang W., *Regional Environment Risk Assessment and Probability Distribution of Topsoil PAHs in Tianjin Area*, **2008**.

[21] Breiman, L., *Out Of Bag Estimation*, Statistics Department, University of California, Berkeley, **1996**.

[22] Dietterich, T. G., Approximate statistical test for comparing supervised classification learning algorithms, *Neural Computation*, (7):1895–1924, **1998**.

[23] Ho, T. K., The random subspace method for constructing decision forests, *Transactions on Pattern Analysis and Machine Intelligence*, 20(8):832–844, **1998**.

[24] Breiman, L., *Random Forests*, Statistics Department, University of California, January 2001, Berkeley, **2001**.

[25] Boulesteix, A. L., Janitza, S., Kruppa, J., König, I.R., *Overview of Random Forest Methodology and Practical Guidance with Emphasis on Computational Biology and Bioinformatics*, Department of Statistics University of Munich, Technical Report Number 129, <http://www.stat.uni-muenchen.de>, **2012**.

[26] Loh, W., *Classification and regression trees*, John Wiley & Sons, **2011**

[27] Breiman, L., Friedman, J., Stone, C. J., Olshen, R.A., *Classification and Regression Trees*, Taylor & Francis, **1984**.

[28] Shi, X., *Bootstrap I*, The University of Wisconsin Madison, Econometric Methods Lecture Notes, <http://www.ssc.wisc.edu/~xshi/econ715.html>, **2016**.

[29] Kurt, S., *The Bootstrap, Short Guides to Microeconometrics*, University of Basel,s.1., **2016**.

[30] Efron, B., Bootstrap Methods: Another Look at Jackknife, *The Annals of Statistics*, Volume 7, Number 1, 1-26, **1979**.

[31] Tercan, A. E., Global Rezerv Kestirimine İlişkin Güven Aralıklarının Uzaklık Bağımlı Bootstrap yöntemi İle Belirlenmesi, *Madencilik Dergisi*, Eylül 2002, Cilt 41, Sayı 3, Sayfa 35-42, **2002**.

- [32] Sutton, C. D., Classification and Regression Trees, Bagging, and Boosting, *Handbook of Statistics*, Vol. 24, sayfa 305, **2004**.
- [33] Dietterich, T. G., An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees, Bagging, Boosting and Randomization, *Machine Learning*, August 2000, Volume 40, pp 139-157, **2000**.
- [34] Dudoit, S., Fridlyand, J., Bagging to Improve the Accuracy of a Clustering Procedure, *Bioinformatics* 19(9), 1090-1099, **2003**.
- [35] Xu, B., *An Empirical Study of Classification and Regression Tree and Random Forests*, Doktora Tezi, Stony Brook Üniversitesi, Sayfa 33, **2006**.
- [36] Amit, Y., Geman, D., Shape Quantization and Recognition with Randomized Trees, *Neural Computation* 9, 1545-1588, **1997**.
- [37] Melville, P., Mooney, R., Creating Diversity in Ensembles Using Artificial Data, *Information Fusion: Special Issue on Diversity in Multiclassifier Systems*, **2004**.
- [38] Pal, M., Random forest classifier for remote sensing classification, *International Journal of Remote Sensing*, 26, 217-222, **2005**.
- [39] Çölkesen, İ., *Uzaktan Algılamada İlerisınıflandırma Tekniklerinin Karşılaştırılması Ve Analizi*, Yüksek Lisans Tezi, Gebze Yüksek Teknoloji Enstitüsü Mühendislik ve Fen Bilimleri Enstitüsü, Sayfa 74, **2009**.
- [40] Efron, B., Tibshirani, R. J., Improvement on Cross-Validation: The 632+ Bootstrap Method, *Journal of the American Statistical Association*, Volume 92, **1997**.
- [41] Goovaerts, P., *Geostatistics for Natural Resources Evaluation*, Oxford Univ. Press, New-York, 483 p. Appendix C describes (and gives) the Jura data set, **1997**.
- [42] Bel, L., Allard, D., Laurent, J.M., Cheddadi, R. and Bar-Hen, A., CART algorithm for spatial data: Application to environmental and ecological data, *Computational Statistics and Data Analysis*, 53, 3082-3093, **2009**.

ÖZGEÇMİŞ

Kimlik Bilgileri

Adı Soyadı : Hamza KIVRAK
Doğum Yeri : Ankara
Medeni Hali : Evli
E-posta : hamzakivrak@gmail.com
Adresi : Şahintepe Mah. 642. Sok. No:35/7 Mamak/ANKARA

Eğitim

Lisans : Hacettepe Üniversitesi Maden Mühendisliği Bölümü
Yüksek Lisans : Hacettepe Üniversitesi Maden Mühendisliği Bölümü
Doktora : Hacettepe Üniversitesi Maden Mühendisliği Bölümü

Yabancı Dil ve Düzeyi: İngilizce-Advanced

İş Deneyimi

2013-... : İş Müfettişi Yrd.- Çalışma ve Sosyal Güvenlik Bakanlığı
2007 - 2013 :Maden Mühendisi - TKİ Göynük Kontrol Müdürlüğü,
Göynük/BOLU
2006 - 2007 : Maden Mühendisi - Eti Bakır A.Ş. Küre/KASTAMONU

Deneyim Alanları

Maden Mühendisliği, Jeostatistik, Kriging, Sınıflandırma ve Regresyon Ağaçları,
Rastlantı Ormanı, Madencilikte İş Sağlığı ve Güvenliği,

Tezden Üretilmiş Projeler ve Bütçesi

-

Tezden Üretilmiş Yayınları

-

Tezden Üretilmiş Tebliğ ve/veya Poster Sunumu ile Katıldığı Toplantılar

-

CURRICULUM VITAE

Credentials

Name Surname : Hamza KIVRAK
Place of Birth : Ankara
Marital Status : Married
E-mail : hamzakivrak@gmail.com
Address : Şahintepe Mah. 642. Sok. No:35/7 Mamak/ANKARA

Education

BSc : Hacettepe University Department of Mining Engineering
MSc : Hacettepe University Department of Mining Engineering
PhD : Hacettepe University Department of Mining Engineering

Yabancı Dil ve Düzeyi: English-Advanced

Work Experience

2013-... : Labour Inspector – The Ministry of Labour and Social Security
2007 - 2013 : Mining Engineer - TKİ Göynük Kontrol Müdürlüğü, Göynük/BOLU
2006 - 2007 : Mining Engineer - Eti Bakır A.Ş. Küre/KASTAMONU

Areas of Experiences

Mining Engineering, Geostatistics, Kriging, Classification and Regression Trees, Random Forest, Occupational Health and Safety in Mining

Project and Budgets

-

Publications

-

Oral and Poster Presentations

-



HACETTEPE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
YÜKSEK LİSANS/DOKTORA TEZ ÇALIŞMASI ORJİNALLİK RAPORU

HACETTEPE ÜNİVERSİTESİ
FEN BİLİMLER ENSTİTÜSÜ
MADEN MÜHENDİSLİĞİ ANABİLİM DALI BAŞKANLIĞI'NA

Tarih: 18/01/2018

Tez Başlığı / Konusu: g-ROKY: Maden Kaynak Kestiriminde Homojen Ortam Modellenmesine Yönelik Yeni Bir Yaklaşımın Geliştirilmesi


Yukarıda başlığı/konusu gösterilen tez çalışmamın a) Kapak sayfası, b) Giriş, c) Ana bölümler d) Sonuç kısımlarından oluşan toplam 73 sayfalık kısmına ilişkin, 18/01/2018 tarihinde şahsım/tez danışmanım tarafından Turnitin adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı % 5 'tir.

Uygulanan filtrelemeler:

- 1- Kaynakça hariç
- 2- Alıntılar hariç/dâhil
- 3- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve bu Uygulama Esasları'nda belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.



Tarih ve İmza 18.01.2018

Adı Soyadı: Hamza KIVRAK
Öğrenci No: N11145231
Anabilim Dalı: MADEN MÜHENDİSLİĞİ
Programı:
Statüsü: Y.Lisans Doktora Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.



Prof. Dr. A. Erhan TERCAN

(Unvan, Ad Soyad, İmza)