

ZKÜ MÜHENDİSLİK FAKÜLTESİ

JEODEZİ VE FOTOGRAMETRİ

MÜHENDİSLİĞİ BÖLÜMÜ

UYDU GÖRÜNTÜLERİNİN SINIFLANDIRILMASI VE DOĞRULUK DEĞERLENDİRMESİ

UZM. MURAT ORUÇ

1 UZAKTAN ALGILAMANIN FİZİKSEL ESASLARI

Günümüzde güncelliği hızla artan Uzaktan Algılama Teknolojisi, arazi örtüsü ve kullanımı ile ilgili bilgileri hızla toplayarak, sorunları çabuk ortaya koyabilmesi, uydu verilerinin kullanıcıya hızla aktarılabilmesi ve uygulanan çözümün zamansal değişimin izlenebilmesi açısından en etkin yöntem olarak görülmektedir.

Uydu verileri ve görüntü işleme teknikleri;

- ✓ Doğal kaynaklara yönelik çok çeşitli alanlarda kullanılabilme
- ✓ Hızlı bilgi üretme özellikleriyle
- ✓ Hızla değişen, dinamik bir yapı sergileyen
- ✓ Doğal olmayan çevresel etkilerin alansal ve yapısal özelliklerinin kısa sürede belirlenmesine

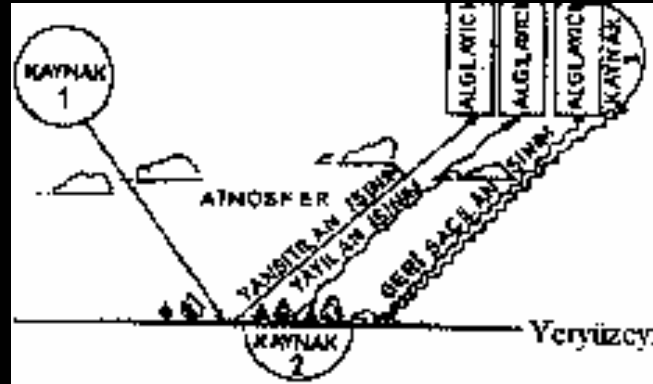
olanak sağlayabilecek özelliklere sahiptir.

Uzaktan algılama sistemlerinde en sık ölçülen nicelik, ilgilenilen nesneden yayılan elektromanyetik enerjidir.

Güneş ve çeşitli doğal ya da yapay kaynaklar, değişik dalga boylarında elektromanyetik enerji saçarlar.

Uzaktan algılama sisteminin dört ana bileşeni: **kaynak, yeryüzü ile karşılıklı etkileşim, atmosfer ile karşılıklı etkileşim** ve **algılayıcı** olmaktadır.

- **Kaynak** : Güneş ışığı veya yeryüzünün neşrettiği ısı gibi doğal ışınlamalar veya yapay elektromanyetik ışınlamalar elektromanyetik radyasyonun kaynağı olabilirler.
- **Yeryüzü ile karşılıklı etkileşim** : Yeryüzünden yansıyan veya neşredilen ışınlamaların miktarı ve ışınlamaların özellikleri, yeryüzündeki objelerin özelliklerine bağlıdır.
- **Atmosfer ile karşılıklı etkileşim** : Elektromanyetik enerji atmosferden geçerken çeşitli etkilerle saçılmaya ve bozulmaya uğrar.
- **Algılayıcı** : Atmosfer ve yeryüzü ile karşılıklı etkileşime uğrayan elektromanyetik ışınlamaların bir algılayıcı tarafından kaydedilir.



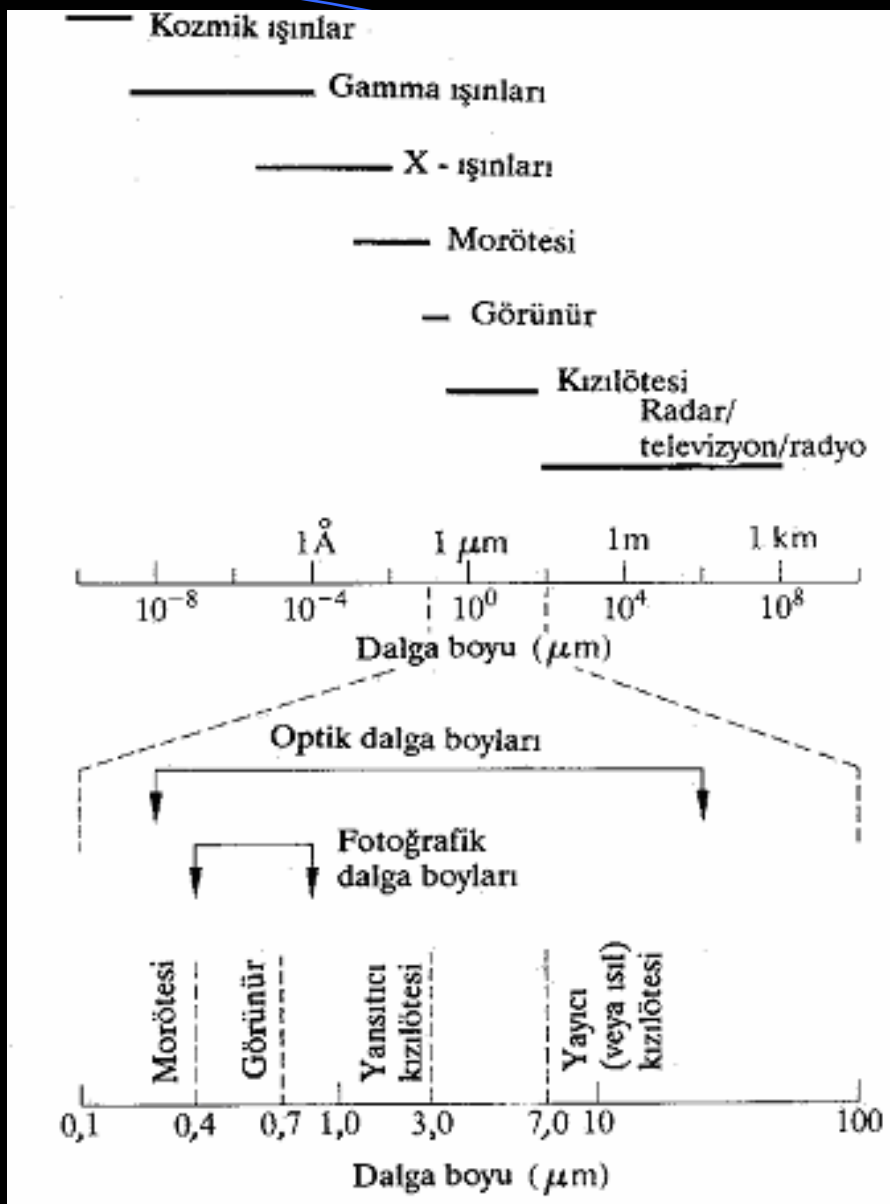
Şekil 1 Uzaktan algılama sistemi

2 ELEKTROMANYETİK SPEKTRUM

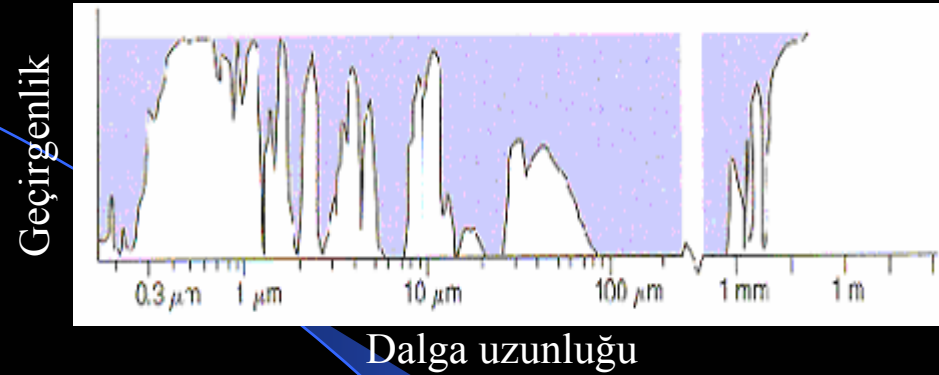
Işık hızı ile hareket eden ve dalga boyu nanometrelerden kilometrelere kadar uzanan sürekli enerji ortamıdır.

Elektromanyetik spektrumun farklı aralıklarında, farklı yeryüzü özellikleri kaydedilir.

Algılayıcıların tasarımında ve yapılacak bir çalışmada kullanılacak uydu görüntüsünün seçiminde, elektromanyetik spektrumun algılama aralıkları büyük önem taşımaktadır.



Şekil 2 Elektromanyetik Spektrum ve algılama aralıkları



Şekil 3 Atmosferik Pencere

Elektromanyetik enerjinin geçirildiği dalga uzunluğu bölgelerine, **atmosferik pencere** denir ve algılayıcı bantlar, atmosferik pencereye göre oluşturulurlar.

Landsat 5'in taşıdığı **Tematik Mapper** (TM) algılayıcısının ölçüm yaptığı spektral bantlar ve bunların ana uygulama alanları aşağıda gösterilmiştir:

Bant - 1 (0.45 - 0.52 μm)

Mavi renge, su yüzeylerine, nemli alanlara, H_2O içeren minerallere hassastır. Su yüzeylerini tespit etmek, orman tiplerini belirlemek için kullanılır. Kıyı harita yapımında yararlanma ve toprağın bitkilerden ayırt edilmesinde rol oynar.

Bant - 2 (0.52 - 0.62 μm)

Bitkilerin canlılığını saptamada; görülen spektrumda ki yeşil bölgede oluşan pik yansımanın ölçülmesinde kullanılır.

Bant - 3 (0.63 - 0.69 μm)

Klorofil soğurma bandı olup, bitki türlerinin ayırt edilmesinde önemli rol oynar.

Bant - 4 (0.76 - 0.90 μm)

Yoğun bitki örtüsünün tespitinde, tarımsal bitkilerin ayırt edilmesinde ve su kütlelerinin ayırt edilmesinde yararlanır.

Bant - 5 (1.55 - 1.75 μm)

Bitki ve topraktaki nemi gösterir. Aynı zamanda karın buluttan ayırt edilmesinde yardımcı olur.

Bant - 6 (10.40 - 12.50 μm)

Termal infrared bant. Suda kirlenme tespitinde, yerleşim üretim alanlarının tespitinde ayrıca toprağın nemi ve volkan araştırmalarında faydalanılır.

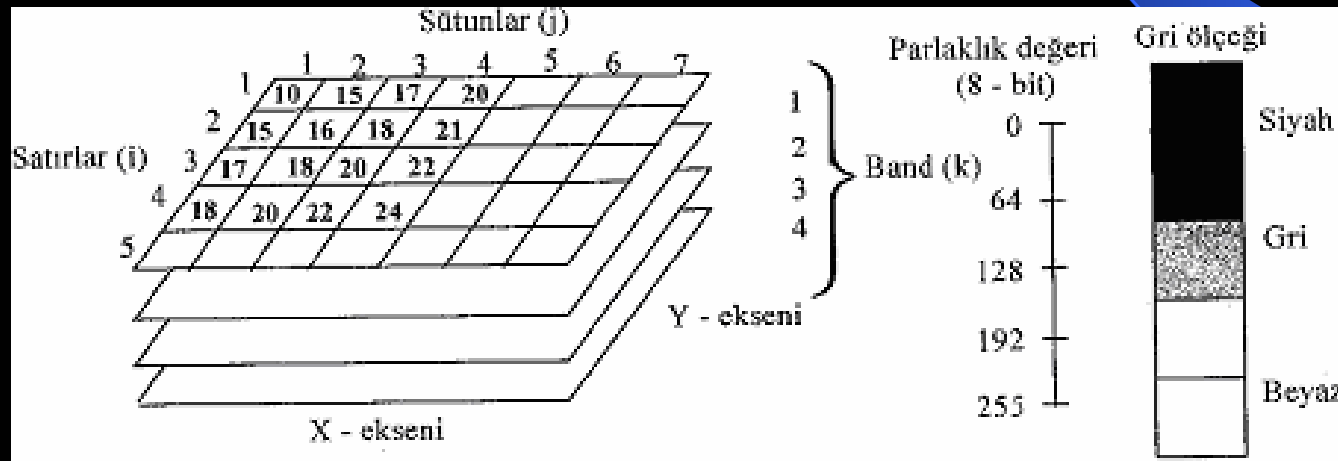
Bant - 7 (2.08 - 2.35 μm)

Kavac cinslerinin ayırt edilmesinde ve hidrotermal haritalamada kullanılabilir.

3 DİJİTAL GÖRÜNTÜ VE ÖZELLİKLERİ

Dijital görüntü, değişik dalga boyundaki ışınım vektörlerinin sayısal ifadesidir. Diğer bir deyişle, her elemanı vektör olan bir matristir. En küçük görüntü elemanına “**piksel**” denir.

Ham bir görüntüde her matris elemanı, karşı düştüğü coğrafi noktanın veya alanın elektromanyetik spektrumdaki yansıma veya yayılım değerlerinden oluşan bir vektördür.



Şekil 4 Dijital görüntünün tanımı

Yeryüzünde bir piksele karşı düşen alan ne kadar değişik görünüm ve değerlere sahip olursa olsun, pikselin bir spektral banttaki parlaklık değeri, 0 (siyah) ile 255 (beyaz) arasında değişen bir sayı ile ifade edilir.

4-GÖRÜNTÜ ZENGİNLEŞTİRME

• ANA BİLEŞENLER DÖNÜŞÜMÜ

Ana bileşenler dönüşümü, çok kanallı veri gruplarında benzer kanalları bastırmak ve daha çok bilgi veren yeni bir veri grubu oluşturmak için geliştirilmiş bir görüntü zenginleştirme yöntemidir. Yöntem, minimum korelasyonlu bilgileri sıkıştırarak veri grubu hakkında maksimum bilginin çıkarılmasını sağladığı için, çok kanallı verilere görsel yorumlama veya sınıflandırma öncesi uygulandığında, verilerden bilgi çıkarılmasını kolaylaştırmaktadır.

Ana bileşenler dönüşümü uygulandığında bantlar arasında spektral özellikler azaltılır veya başka bir deyişle benzer olan bantlar bastırılır. Amaç, veri kombinasyonunda görsel yorumlanabilirliği arttırmak için n kanallı orijinal veri dizisinde mevcut olan tüm bilgileri n 'den daha az sayıda yeni bantlara veya yeni bileşenlere sıkıştırmaktır.

Bu dönüşümde orijinal veriler içinde maksimum varyansa sahip doğrultular boyunca bileşen eksenleri adı verilen yeni eksenler oluşur. Pikseller yeni koordinat eksenine yerleştiklerinde 1. Ana bileşen eksenine 2. Ana bileşen eksenine göre daha büyük varyansa sahiptir. Bu nedenle 1. Ana bileşen eksenini veri grubunun uzun eksenini (ana eksen) belirtir ve daha fazla bilgiyi içerir. İkinci ana bileşen elipsin ana eksenini dik kesen eksendir ve birinci bileşende tanımlanmamış verilerin büyük bir kısmını tanımlar. Bu analiz sonucunda n tane çıkış kanalı olmasına rağmen verilerdeki bilginin büyük bir bölümü ilk iki veya üç bantta toplanır. Diğer bantlarda varyansın azalması nedeniyle gürültü etkisi ortaya çıkar ve yorumlayıcı için daha az bilgi içerdiğinden kullanılmaz.

●YOĞUNLUK-RENK TONU-DOYGUNLUK (IHS) RENK UZAY TRANSFORMASYONU

Dijital görüntüler genellikle üç ana renk olan kırmızı, yeşil ve mavi (RGB) kullanılarak toplamsal renk bileşikleri olarak görüntülenmektedir.

RGB bileşenlerine göre renkleri tanımlamanın bir seçeneği yoğunluk-renk tonu-doygunluk (IHS) sistemidir.

Yoğunluk (şiddet) bir rengin toplam parlaklığını,

Renk tonu, bir renge katkıda bulunan baskın veya ortalama dalga boyunu,

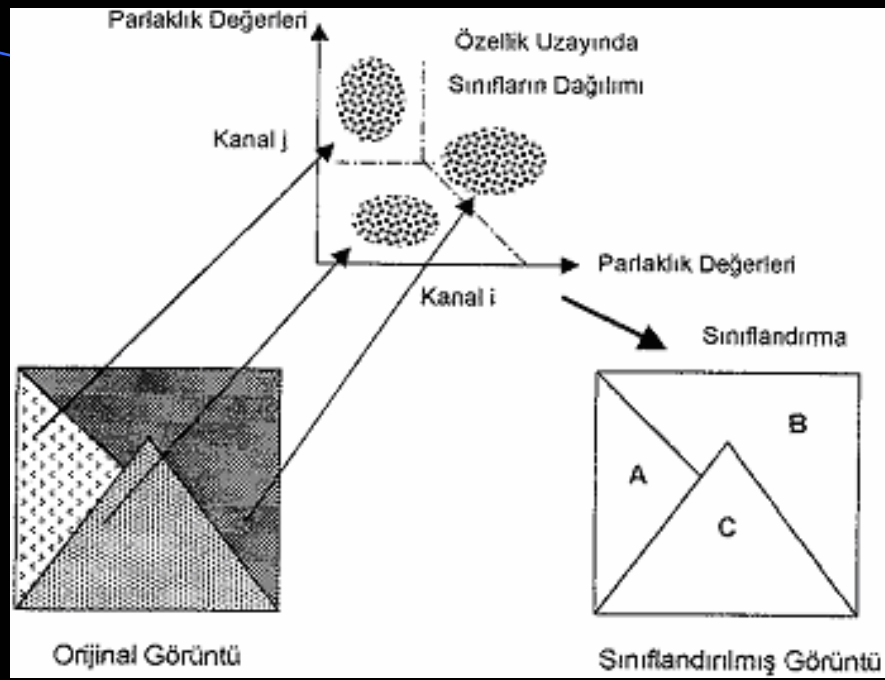
Doygunluk ise rengin griye göre saflığını belirtmektedir.

4 GÖRÜNTÜ SINIFLANDIRMA

Dijital görüntülerde farklı özellik tipleri, doğal spektral yansıtma ve yayma özelliklerine bağlı olarak farklı sayısal değerler içeren kombinasyonlar oluşturmaktadır. Bu farklılıktan yararlanılarak aynı spektral özellikleri taşıyan yer yüzündeki nesneler gruplandırılabilir. **Amaç uydu görüntülerindeki her pikseli spektral özelliklerine göre farklı gruplara ayırmak ve pikseli yansıtma değerlerine göre yer yüzünde karşılık geldiği kümeye atamaktır.**

Sınıflandırma işleminin gerçekleştirilmesinde dikkat edilecek hususlar şunlardır.

- Algılayıcı, algılama zamanı ve spektral bantların amaca uygun olarak seçimi.
- Yer yüzü özelliklerini ortaya koyabilecek kontrol alanlarının seçimi.
- Amaca yönelik sınıflandırma algoritmalarının seçimi.
- Belirlenen bu özelliklerin tüm görüntüye uygulanması ve görüntülenmesi.
- Sonuç görüntülerinde doğruluk değerlendirmelerinin yapılması.



Şekil 5 Sınıflandırmanın genel adımları

Sınıflandırılmış bir uzaktan algılama görüntüsüne bir bakıma bir tür **Dijital Tematik Harita** gözüyle bakılabilir.

Sınıflandırma yöntemleri;

1- PİXEL-TABANLI

- ❖ **Kontrollü** (Supervised) sınıflandırma

- ❖ **Kontrolsüz** (Unsupervised) sınıflandırma

2- OBJE-TABANLI

olarak ikiye ayrılır.

4.1.1 Kontrolsüz (Unsupervised) Sınıflandırma

Çalışma bölgesinde özellikleri bilinen **yeterince örnekleme bölgesi (test bölgesi)** veya **istatistiksel bilgi** bulunmadığı durumlarda, spektral olarak ayrılabilir sınıflar belirlenmekte ve bunlardan bilgi elde etme yoluna gidilmektedir.

Bu sınıflar görüntü dijital değerlerindeki doğal gruplaşmalara bağlı olup, oluşan spektral sınıfların ne olduğu önceden bilinmemektedir.

Oluşan sınıfların özellikleri bölgeye ait **hava fotoğrafları, topoğrafik haritalar** ve daha önce elde edilmiş var olan bilgilerle karşılaştırılarak belirlenir.

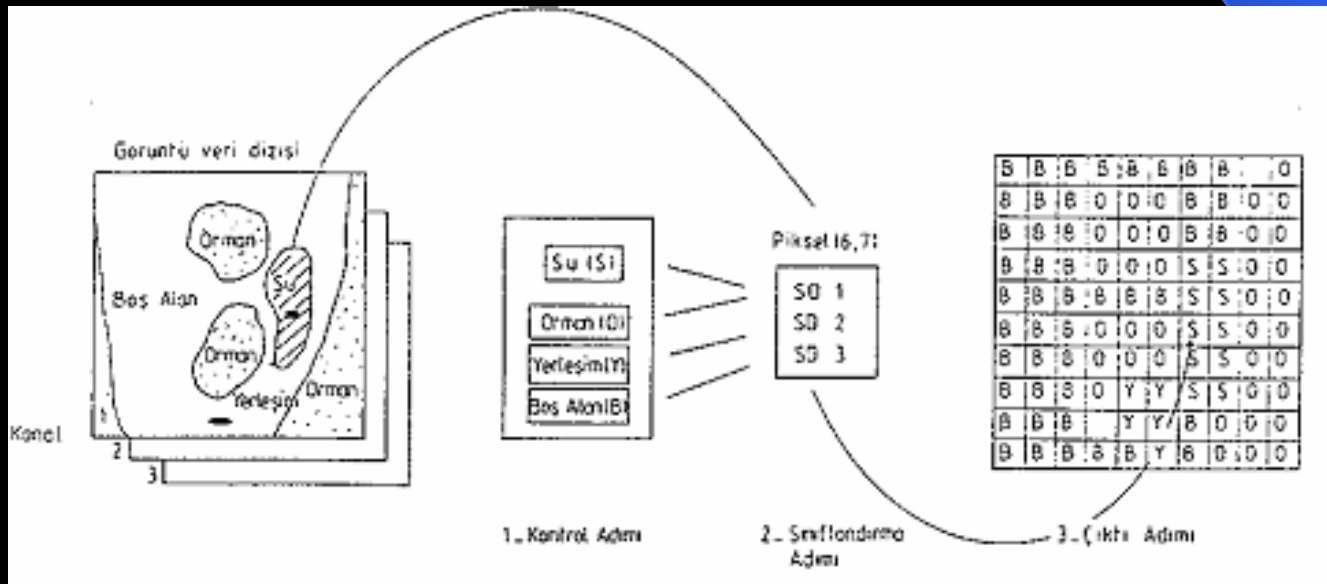
Kümeleme merkezi önce geçici olarak belirlenmekte olup, her bir pikselin açılan kümelerden hangisine dahil edileceği ise o görüntü elemanı ile açılan küme merkezi arasındaki **uzaklığın** hesaplanması ve **başlangıçta kabul edilen uzaklık değeri** ile karşılaştırılması sonucu belirlenir.

Bu belirlemede **en küçük uzaklık değeri** esas alınır.

Kümeye katılan her yeni piksel ile birlikte aritmetik ortalama hesaplanarak geçici küme merkezi n boyutlu uzayda ötelenmekte ve işlem geçici merkezin konumunda **değişim olmayıncaya kadar iteratif** olarak devam etmektedir.

4.1.2 Kontrollü (Supervised) Sınıflandırma

Bu yöntemde çalışma alanındaki yer yüzü özelliklerini tanımlayan yeteri sayıdaki örnek bölgeler (test alanlar) kullanılarak, sınıflandırılacak her bir cisim için spektral özellikleri tanımlı özellik dosyaları oluşturulur. Bu dosyaların görüntü verilerine uygulanması ile her bir görüntü elemanı (piksel), hesaplanan olasılık değerlerine göre en çok benzer olduğu sınıfa atanmaktadır.



Şekil 6 Kontrollü sınıflandırma yönteminin temel aşamaları.

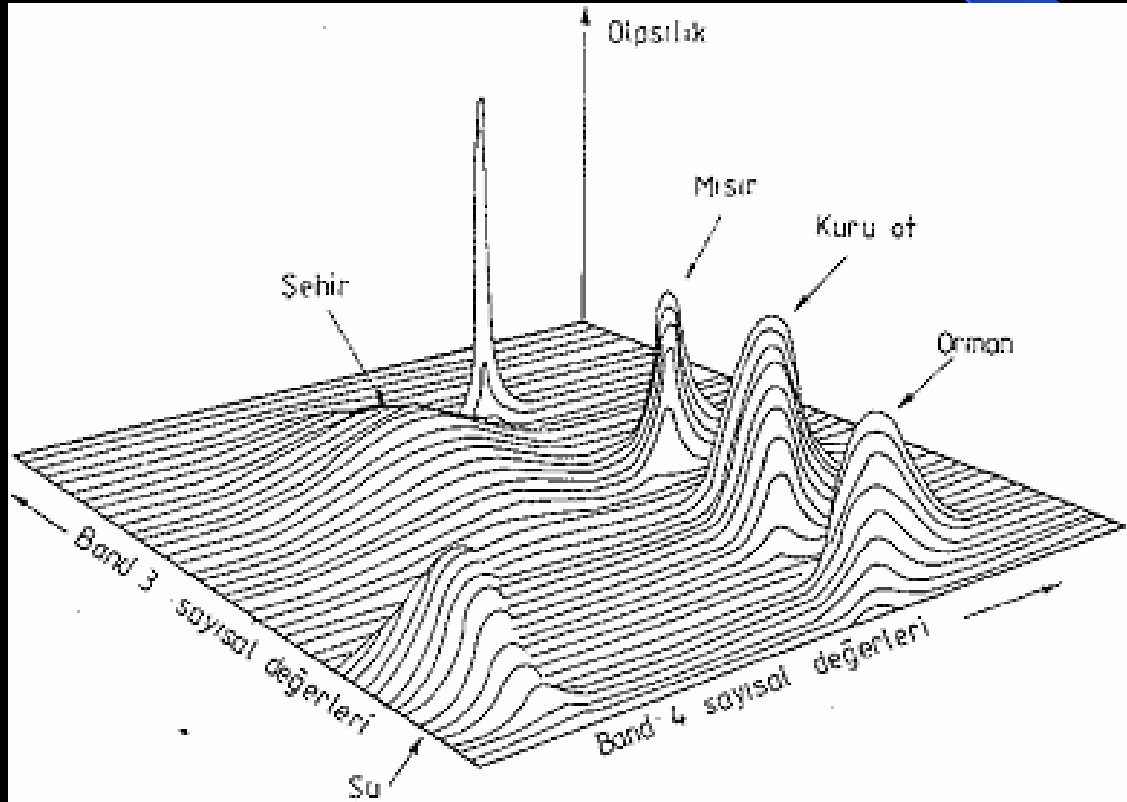
Kontrollü sınıflandırma işleminde;

- **En Büyük Benzerlik (Maximum Likelihood)**
- **Paralelyüz (Parallelepiped)**
- **En Küçük Uzaklık (Minimum Distance)**

adı verilen çeşitli sınıflandırma yöntemleri kullanılmaktadır.

4.1.2.1 En Büyük Benzerlik (Maximum Likelihood) Yöntemi

Bu yöntem, sınıflar için eş olasılık eğrilerinin tanımlanmasına ve sınıflandırılacak piksellerin üyelik olasılığı en yüksek olan sınıfa atanması ilkesine dayanır.



Şekil 7 Olasılık Yoğunluk Fonksiyonları

Yöntemin etkinliği, her spektral sınıf için **ortalama vektör** ile **kovaryans matrisin** doğru biçimde tahmin edilmesine bağlı olmaktadır.

Bu koşul ise, sınıfların her biri için yeterli miktarda örnekleme verisinin (pikseller) bulunabilmesine bağlıdır.

Örnekleme bölgesine ait yeterli miktarda veri bulunmadığı zaman (sınıfların olasılık dağılımlarını doğru bir şekilde tahmin edebilecek özellikte veri olmadığı zaman) arzulanan sınıflandırma doğruluklarına ulaşamaz.

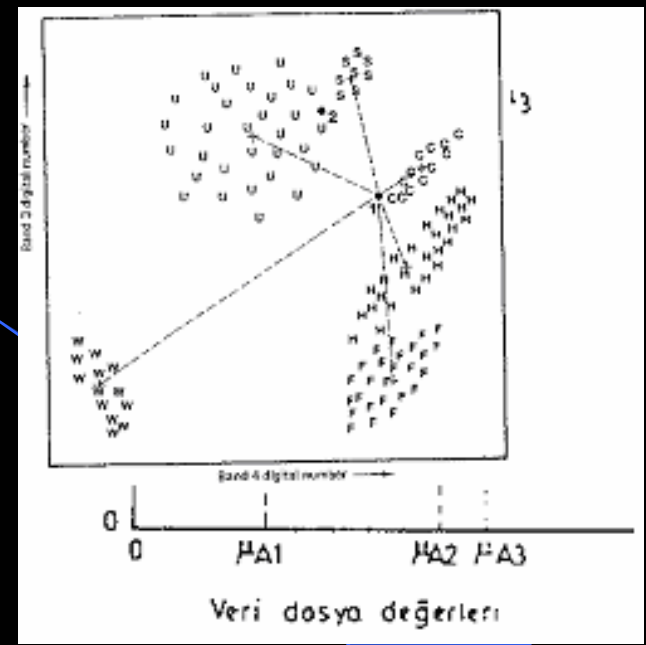
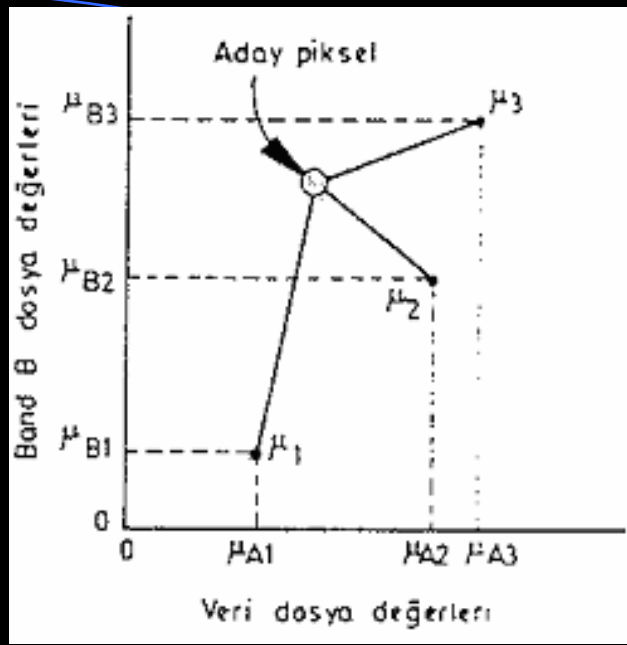
Böyle bir durumda kovaryans bilgilerini kullanmayan başka sınıflandırma yöntemlerine başvurmak gerekir.

4.1.2.2 En Küçük Uzaklık (Minimum Distance) Yöntemi

Örnekleme bölgelerine ait örnekleme verileri sınırlı olduğu zaman, kovaryans bilgilerini kullanmayan yalnızca **spektral sınıfların ortalama vektörlerinin** kullanılmasına dayanan sınıflandırma yöntemlerine başvurmak gerekir.

Böyle bir durumda kısıtlı örnekten ortalama değerler kovaryanslara nazaran daha doğru tahmin edilebilir.

Örnekleme verilerinden sınıf ortalamaları belirlenir, sınıflandırmada bir piksel en yakın ortalamaya sahip sınıfa atanır.



Şekil 8 En Küçük Uzaklık Yöntemi

En küçük uzaklık yöntemi, en büyük benzerlik yönteminden daha **hızlı** olduğu için ilgi çekicidir. **Ancak kovaryans bilgilerini kullanmadığından en büyük benzerlik yöntemi kadar esnek değildir.**

Yöntemde kovaryans verilerinin kullanılmaması nedeniyle sınıf modelleri spektral anlamda **simetrik**dir. Bu nedenle bazı sınıflar iyi modellenmeyebilir.

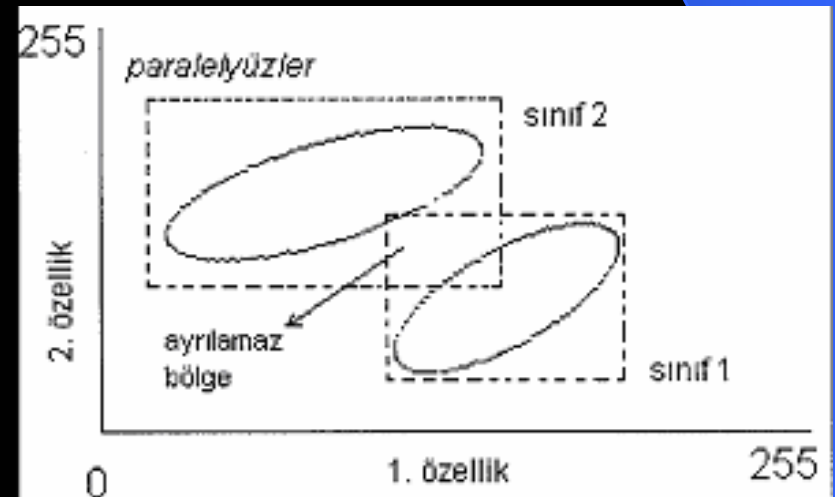
Ancak örnekleme verileri sınırlı olduğu zaman en küçük uzaklık yöntemi, en büyük benzerlik yönteminden daha doğru sonuçlar verir.

4.1.2.3 Paralelyüz (Parallelepiped) Yöntemi

Bu yöntem, kontrol (örnekleme) verilerinin spektral bileşenlerinin **histogramlarının incelenmesine** dayalı çok basit bir kontrollü sınıflandırmadır. Aynı zamanda bir sınıflandırma analistinden en az bilgiyi gerektiren yöntemdir.

Tanımlanan her sınıf için, kullanılan her bandın **minimum** ve **maksimum piksel değerleri** kullanılmaktadır. Uygulanan karar kuralı, her bilinmeyen piksel, özellik değeri ile ilgili olduğu paralelyüz içine atanır.

Şekil 9 İki Boyutta Paralelyüz Sınıflandırma



Paralelyüz sınıflandırma yöntemi, uygulaması hızlı ve kolay olmasına rağmen; her bir örnekleme seti için özelliklerin yalnızca **minimum** ve **maksimum** değerleri kullanılması nedeniyle, gerçek spektral sınıfların iyi temsil edilememesine yol açmaktadır.

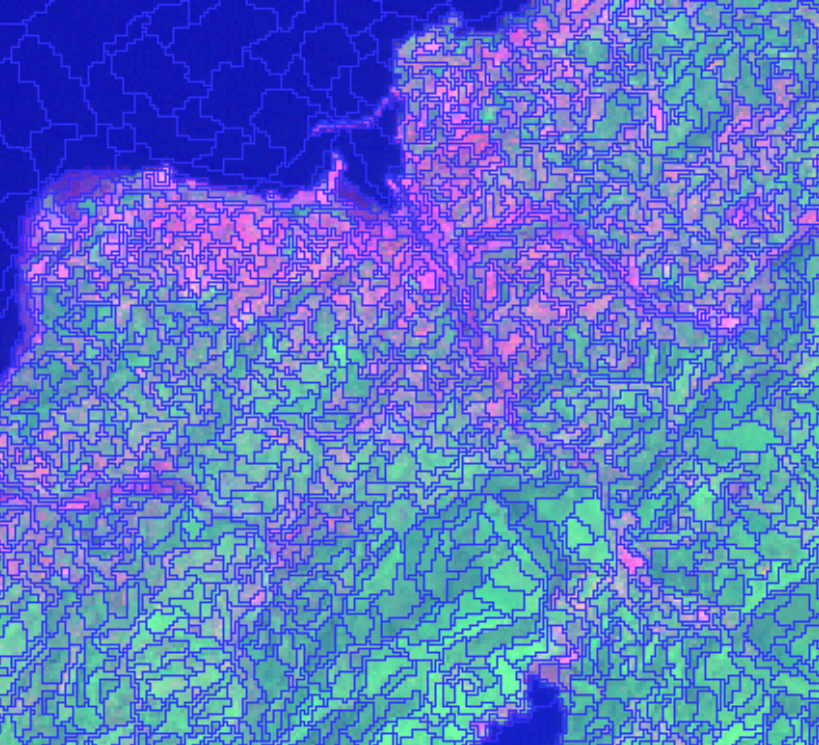
Diğer bir sakınca ise, sınıf üyeliklerinin öncül olasılıklarının dikkate alınmamasıdır.

4.2 OBJE-TABANLI SINIFLANDIRMA YÖNTEMİ

Bu sınıflandırma yönteminde pixel boyutunda gruplanmış veriler yerine oluşturulan segmentler kullanılır. Segment, belli bir gri değerine veya aralığa sahip ve bu gri değerleri grubunu temsil eden alanlar olarak nitelendirilebilir. Bu alanlar (segmentler) seçilen segment parametrelerinin değerlerine göre görüntüde farklılık gösterir.

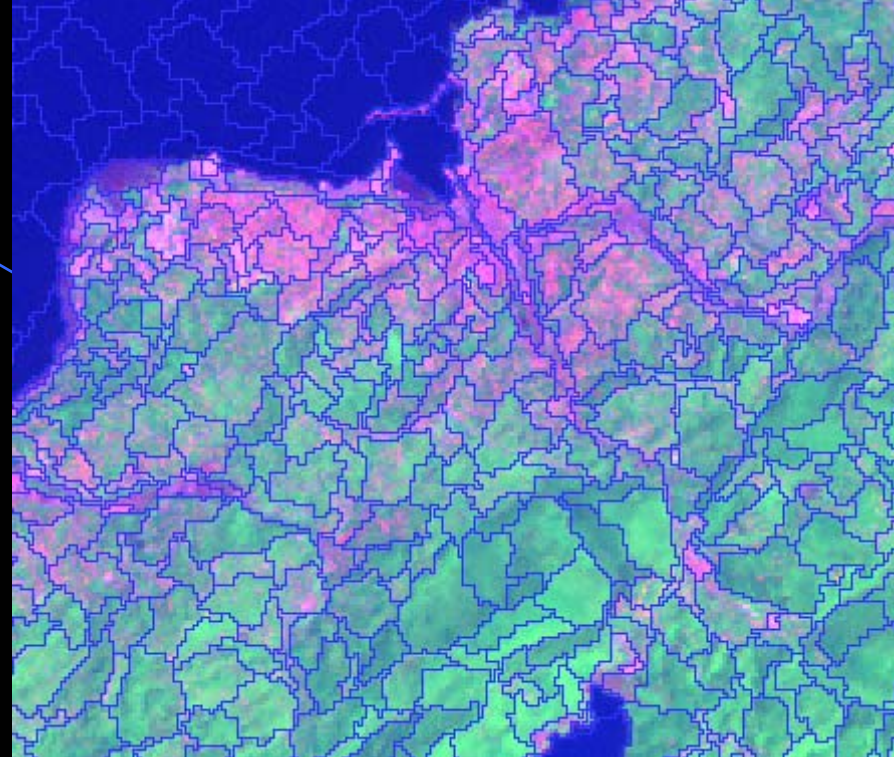
Segment parametreleri, görüntünün özelliklerine göre ve yapılacak sınıflandırmaya göre birden fazla Level değerleri oluşturularak belirlenir ve amaca en uygun parametreler (Level) sınıflandırma işlemi için kullanılır.

Level	Scale Parameter	Color	Shape	Smoothness	Compactness
Level 1	5	0.7	0.3	0.9	0.1
Level 2	10	0.5	0.5	0.5	0.5
Level 3	25	1.0	0	0.5	0.5



SP=5

SP: Scale
Parameter










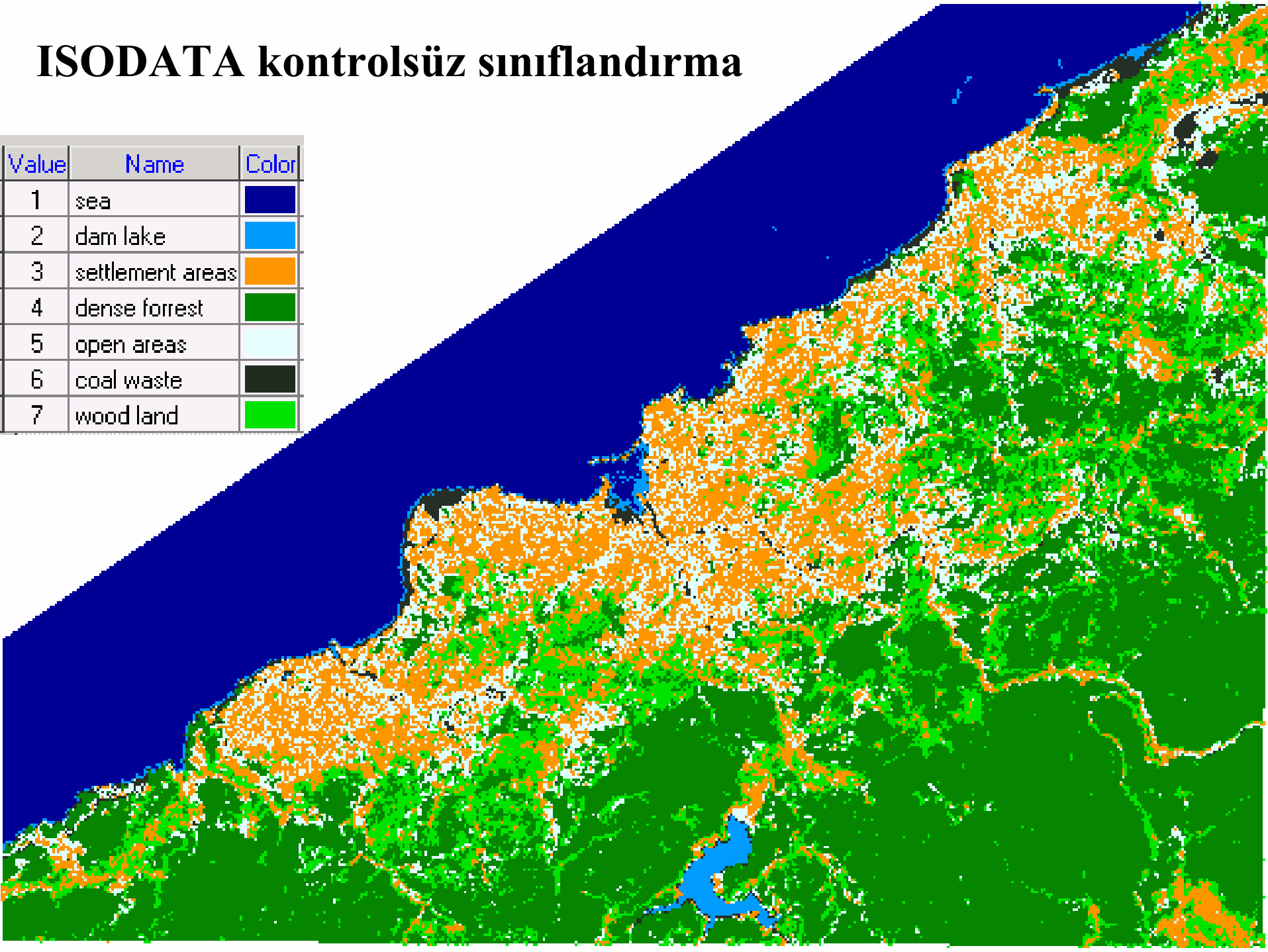
SP=10



SP=25

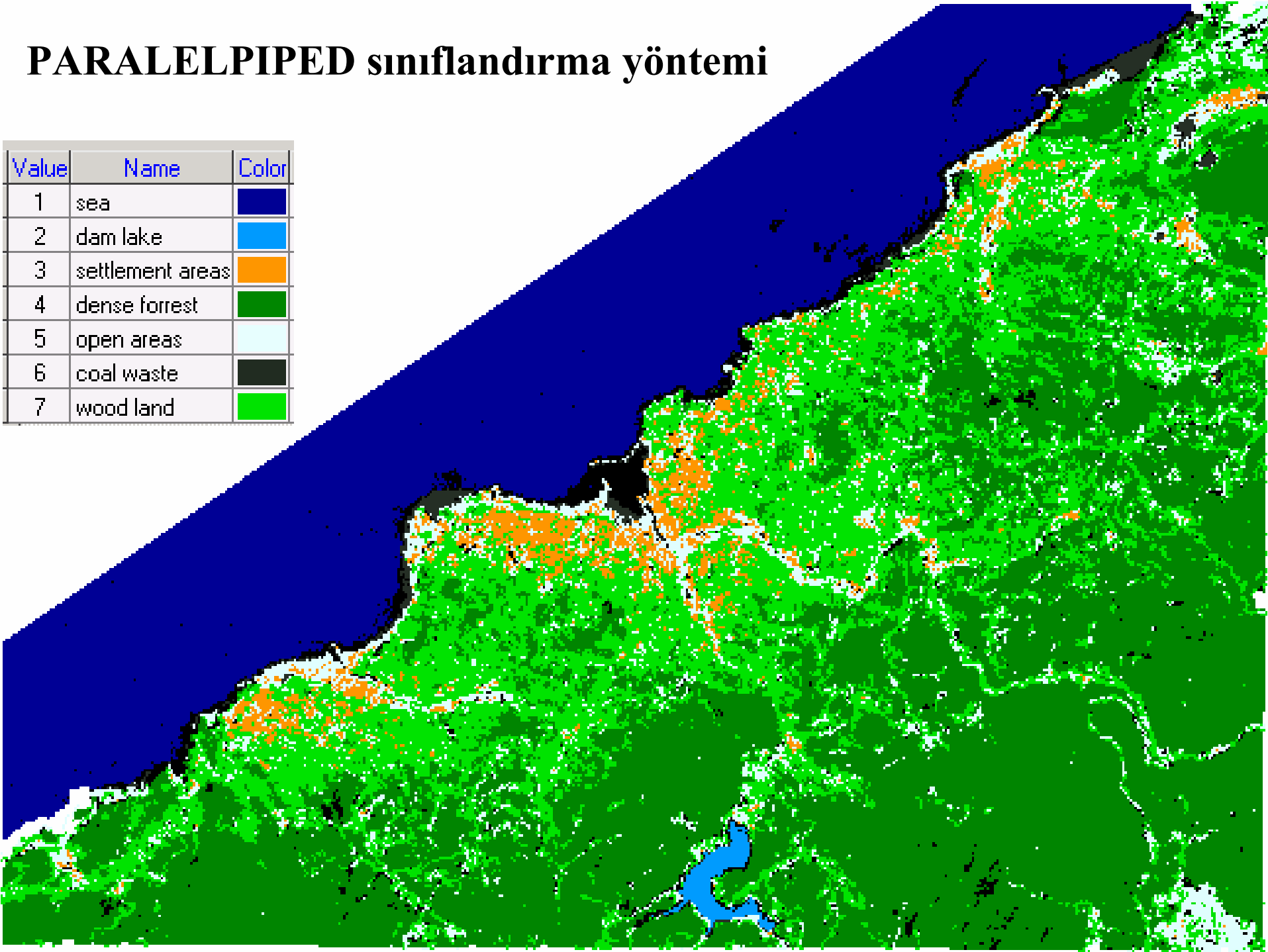
ISODATA kontrolsüz sınıflandırma

Value	Name	Color
1	sea	
2	dam lake	
3	settlement areas	
4	dense forrest	
5	open areas	
6	coal waste	
7	wood land	



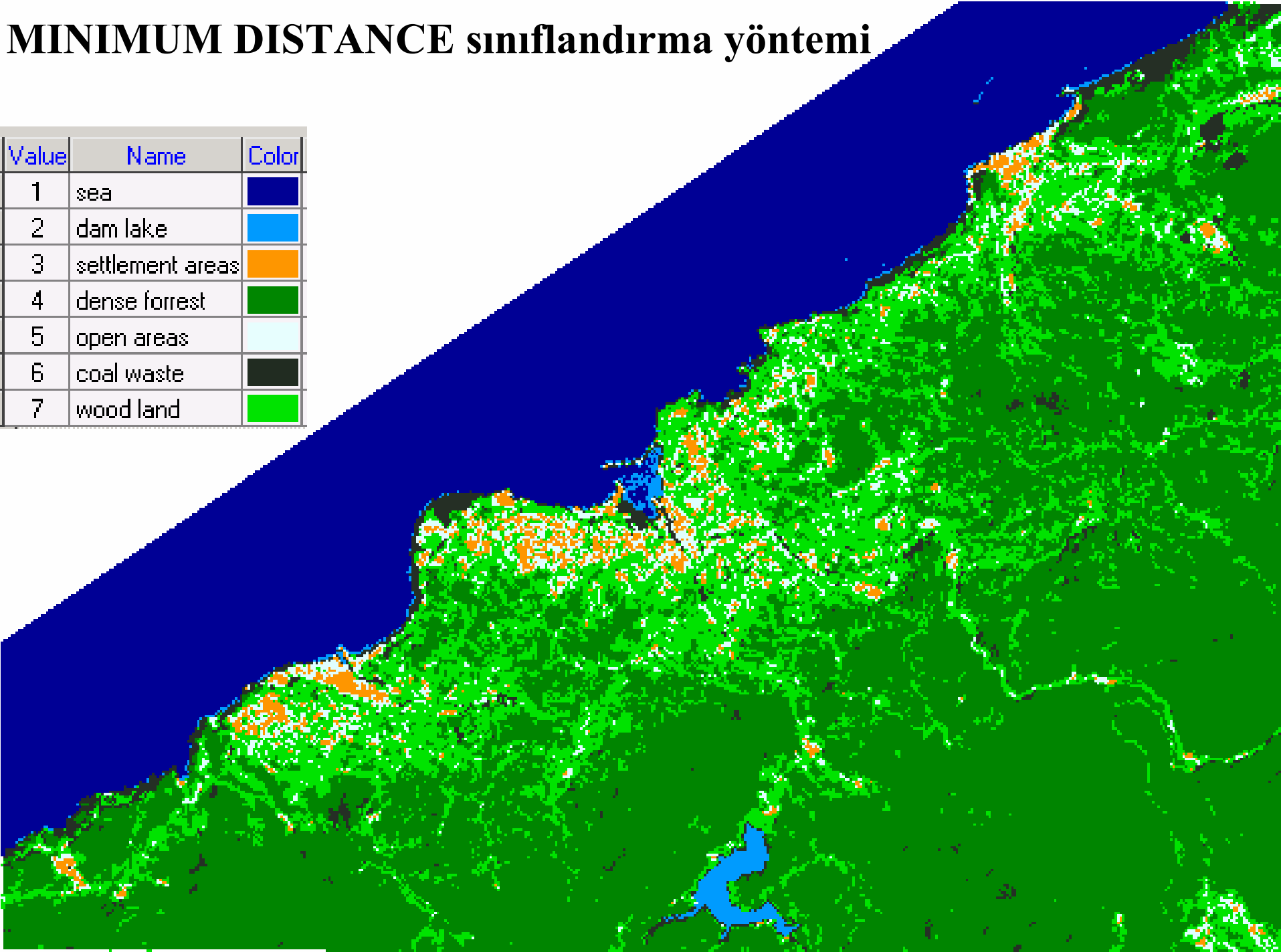
PARALELPIPED sınıflandırma yöntemi

Value	Name	Color
1	sea	dark blue
2	dam lake	light blue
3	settlement areas	orange
4	dense forrest	dark green
5	open areas	light green
6	coal waste	black
7	wood land	medium green










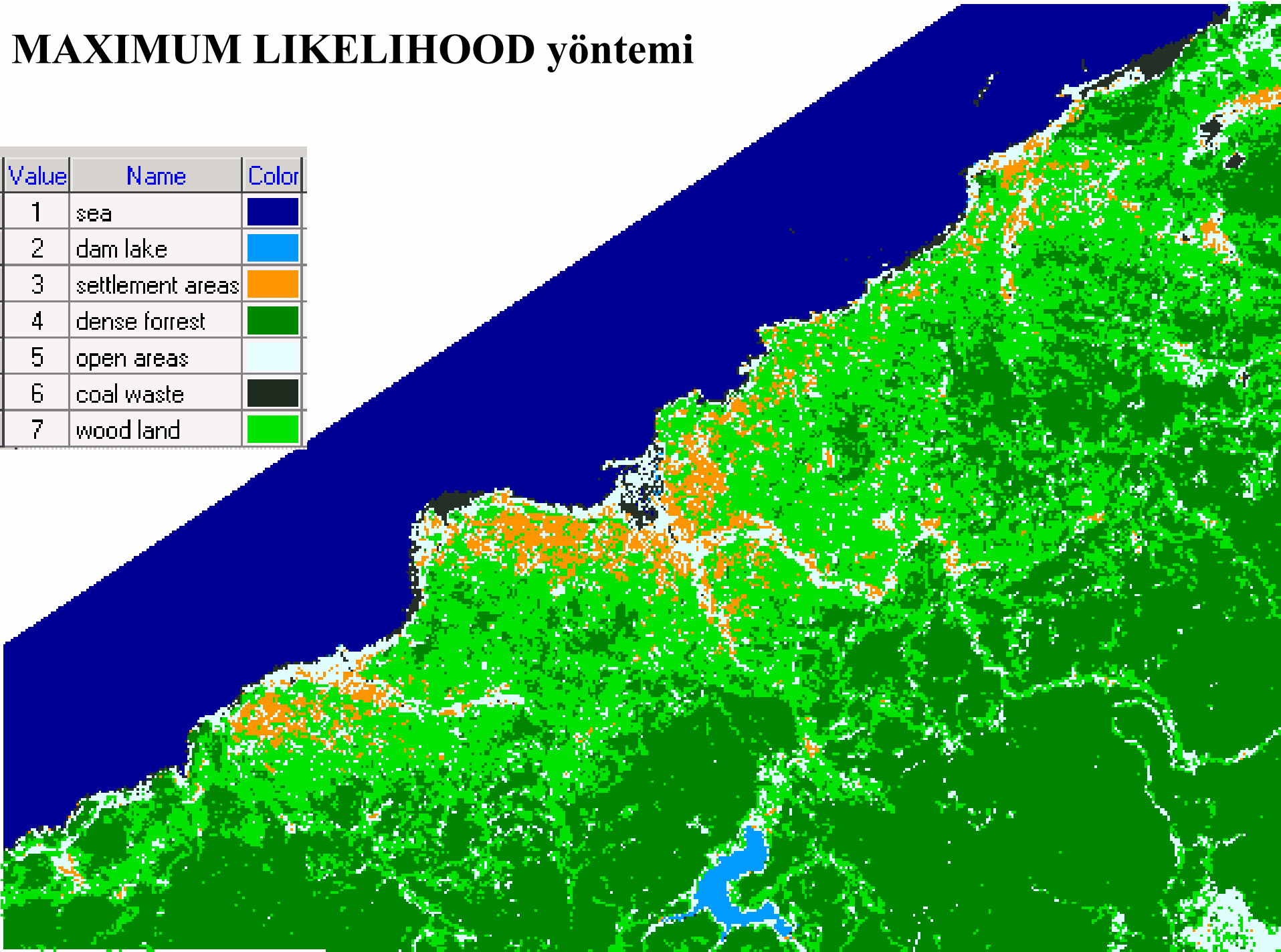
MINIMUM DISTANCE sınıflandırma yöntemi

Value	Name	Color
1	sea	dark blue
2	dam lake	light blue
3	settlement areas	orange
4	dense forrest	dark green
5	open areas	light green
6	coal waste	black
7	wood land	medium green










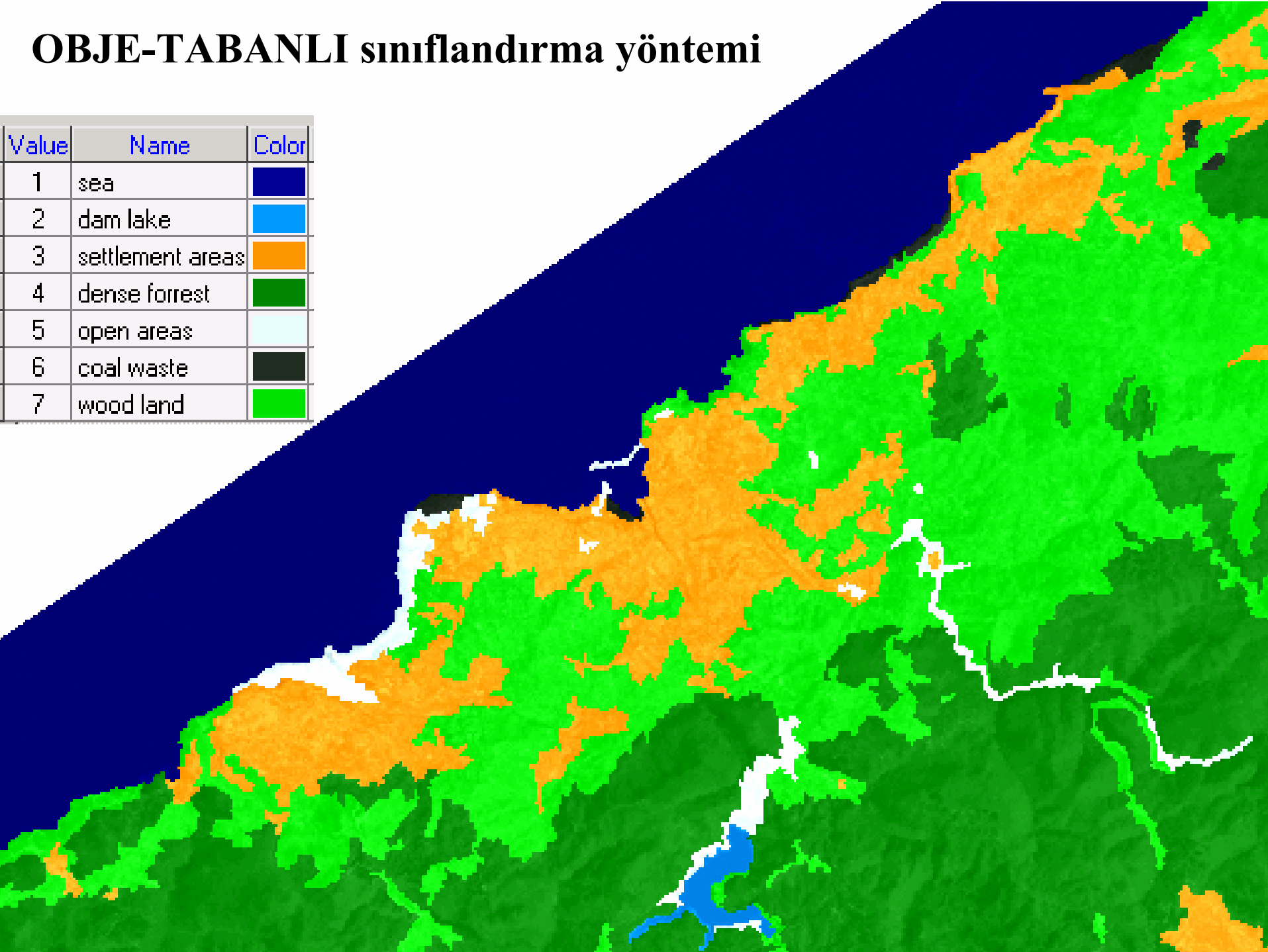
MAXIMUM LIKELIHOOD yöntemi

Value	Name	Color
1	sea	
2	dam lake	
3	settlement areas	
4	dense forrest	
5	open areas	
6	coal waste	
7	wood land	



OBJE-TABANLI sınıflandırma yöntemi

Value	Name	Color
1	sea	
2	dam lake	
3	settlement areas	
4	dense forrest	
5	open areas	
6	coal waste	
7	wood land	



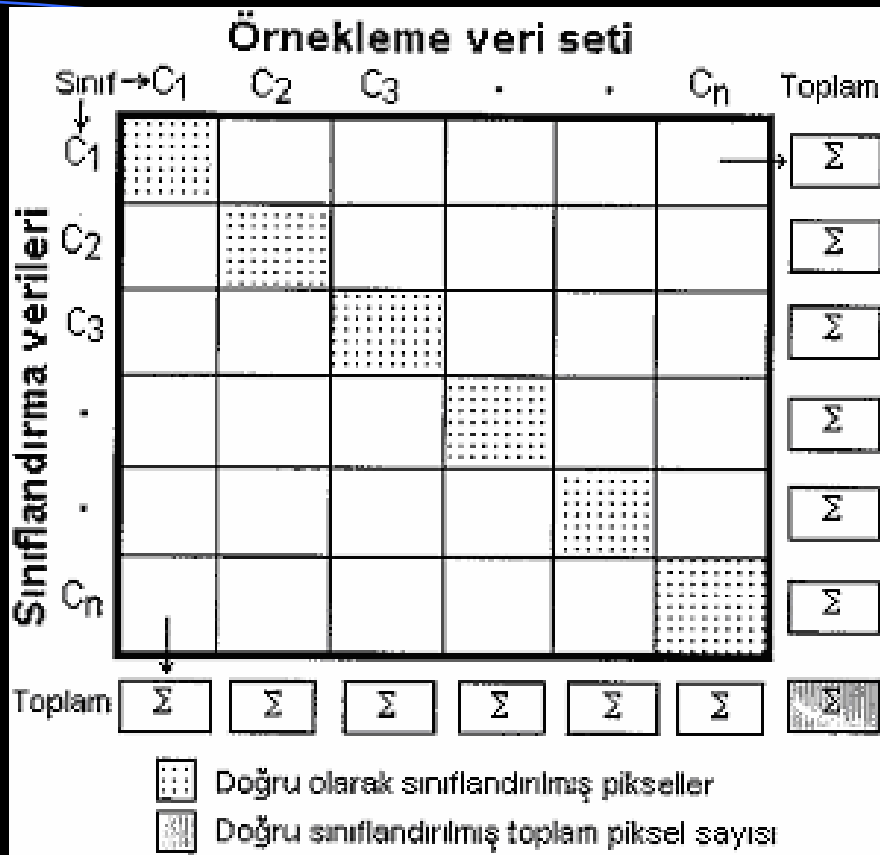
5 GÖRÜNTÜ SINIFLANDIRMASININ DOĞRULUK DEĞERLENDİRİLMESİ

Uzaktan algılamada görüntü sınıflandırmalarında doğruluk, bir piksele tahsis edilen **sınıf etiketi** ile “**gerçek**” sınıf arasındaki uygunluğu gösterir.

Gerçek sınıf, gerek hava fotoğraflarından gerekse de mevcut harita ve planlardan doğrudan veya dolaylı olarak gözlenebilir.

Doğruluk değerlendirmesinde çok kullanılan yöntem hata matrislerinin (confusion matrix, contingency table) hazırlanmasıdır.

Hata matrisleri sınıf bazında **referans verileri** (gerçek-yersel veriler) ile **otomatik sınıflandırma sonuçları** arasındaki ilişkiyi karşılaştırır.



Şekil 10 Hata matrisinin şematik gösterimi

Matrisin sütun elemanları örnekleme veri setine ilişkin değerleri gösterirken, satır elemanları sınıflandırılmış piksel verilerini göstermektedir.

Doğru arazi örtüsüne sınıflandırılmış örnekleme veri seti pikselleri **ana köşegen** (diagonal) üzerinde bulunmaktadır.

User Class \ Sample	sea	damlake	settlementareas	denseforest	openareas	coalwaste	woodland	Sum
Confusion Matrix								
sea	1	0	0	0	0	0	0	1
damlake	0	1	0	0	0	0	0	1
settlementareas	0	0	6	0	1	4	1	12
denseforest	0	0	0	12	0	0	0	12
openareas	0	0	1	0	4	0	0	5
coalwaste	0	0	0	0	1	4	0	5
woodland	0	0	1	0	0	0	11	12
unclassified	0	0	0	0	0	0	0	0
Sum	1	1	8	12	6	8	12	

Accuracy

Producer	1	1	0.5	1	0.8	0.8	0.917
User	1	1	0.75	1	0.667	0.5	0.917
Hellden	1	1	0.6	1	0.727	0.615	0.917
Short	1	1	0.429	1	0.571	0.444	0.846
KIA Per Class	1	1	0.667	1	0.628	0.442	0.889

Totals

Overall Accuracy 0.813
KIA 0.766

☐ reduce
 ☒ expand

Close

Hata matrislerinden çeşitli sınıflandırma doğruluk kriterleri türetilir.

“Toplam doğruluk” (overall accuracy), doğru biçimde sınıflandırılmış piksellerin toplam sayısının (köşegen toplamı) referans piksellerin toplam sayısına bölünmesiyle elde edilir.

Matrisin köşegeni üzerinde bulunmayan elemanları **“ihmal hatası”** veya **“dahil etme hatası”** (comission error)’ nı temsil eder.

“Üretici doğruluğu” (producer’s accuracy), her sınıf içinde doğru olarak sınıflandırılmış piksellerin sayısını, bu sınıf için kullanılan örnekleme veri seti pikselleri sayısına bölerek bulunur ve verilen bir arazi örtü türünün **örnekleme seti piksellerinin ne kadar iyi sınıflandırılabilirdiğini** gösterir.

“Kullanıcı doğruluğu” (user’s accuracy), her sınıf içinde doğru sınıflandırılmış piksel sayısını, bu kategori içinde sınıflandırılan piksellerin toplam sayısına bölünmesiyle bulunur ve **“dahil etme hatası”** nı gösteren bir ölçüdür.

Bu doğruluk değeri, **herhangi bir sınıfa atanan bir pikselin bu sınıfı gerçekte temsil etme olasılığını** gösterir.

Kappa (κ) istatistik ölçüsü, referans verilerle otomatik sınıflandırılmış veriler arasındaki gerçek uyum ile, yine referans verilerle rastgele sınıflandırılmış veriler arasındaki şans uyumu arasındaki farkı göstermek için kullanılmaktadır. Kavramsal olarak κ aşağıdaki şekilde tanımlanır:

$$\kappa = \frac{\text{Ö.D.} - \text{Ş.U.}}{1 - \text{Ş.U.}}$$

Burada Ö.D., ölçülen doğruluğu, Ş.U. ise şans uyumunu ifade etmektedir.

Bu κ istatistiki değeri, hata matrisinin doğru değerlerine erişimin doğru uyuşum sonucu mu yoksa “şans” uyuşumu nedeniyle mi olduğunun bir işaretini verir.

Doğru uyuşum (ölçülen) 1’e ve şans uyuşumu 0’a yaklaştıkça κ değeri 1’e yaklaşır. Buda ideal durum olmaktadır.

Kappa katsayısı, hata matrisinin satır ve sütun toplamaları ve köşegeni üzerindeki elemanlar kullanılarak hesaplanır.

Buradan anlaşılacağı üzere, Kappa istatistiği, köşegen üzerinde bulunmayan elemanları da doğruluk değerlendirmesine katmaktadır.

Kappa istatistiğinin temel avantajlarından biri de; verilen bir hata matrisinin veya matrisler arasındaki farkların istatistiksel anlamlılığını (signifikans) belirlemek için kullanılabilmesidir.

Sözgelimi farklı tarihlerdeki uydu görüntülerinin sınıflandırma sonuçlarının karşılaştırılmasında veya farklı sınıflandırma teknikleri ile gerçekleştirilen sınıflandırma doğruluklarının karşılaştırılması için kullanılabilir.

Confusion Matrix

Class on map	Class on ground				Total sampled sites
	Forest	Pasture	Arable	Bushland	
Forest	93	8	15	-	116
Pasture	6	65	23	1	95
Arable	11	34	503	32	580
Bushland	5	-	21	72	98
Total sampled sites	115	107	562	105	889

Total accuracy = $733/889 = 82\%$; 68% (65/95) of pasture was correctly classified; 32% (34/107) of pasture was incorrectly classed as arable; while 24% (23/95) of pasture on the map was actually arable on the ground. For pasture, the **commission error** = $30/95 = 32\%$, the **omission error** = $42/107 = 39\%$.

Cohen's Kappa

A measure considers significantly unequal sample sizes and likely probabilities of expected values for each class:

$$\kappa = \frac{d - q}{N - q}$$

where N = total number of samples;
 d = total number of cases in diagonal cells;

$$q = \frac{\sum_{i=1}^n \left(\sum_{j=1}^n a_{i,j} \sum_{j=1}^n a_{j,i} \right)}{N}$$

Kappa Computation

Class on map	Class on ground				Total sampled sites
	Forest	Pasture	Arable	Bushland	
Forest	93	8	15	-	116
Pasture	6	65	23	1	95
Arable	11	34	503	32	580
Bushland	5	-	21	72	98
Total sampled sites	115	107	562	105	889

In our case: $d = 93 + 65 + 503 + 72 = 733$

$$q = \frac{\sum_{i=1}^n \left(\sum_{j=1}^n a_{i,j} \sum_{j=1}^n a_{j,i} \right)}{N} = \frac{116 \times 115 + 95 \times 107 + 580 \times 562 + 98 \times 105}{889} = 404.67$$

$$\kappa = \frac{d - q}{N - q} = \frac{733 - 404.67}{889 - 404.67} = 0.678$$

The optimal score is 1.0 (i.e. 100% correct).