

Coğrafi veri seçim işlemi sonuçlarının değerlendirilmesinde hata matrisinin kullanımı

Bülent ÇETİNKAYA*, **Gönül TOZ**

İTÜ Fen Bilimleri Enstitüsü, Jeodezi ve Fotogrametri Mühendisliği Programı, 34469, Ayazağa, İstanbul

Özet

Bilgi sistemlerinin yoğun olarak kullanıldığı günümüzde, Coğrafi Bilgi Sistemlerine (CBS) ve CBS'nin temel bileşenlerinden coğrafi verilere olan ihtiyaç, artarak devam etmektedir. Coğrafi verilerin üretimi, CBS bileşenleri arasında en çok zaman ve maliyet gerektiren işlemi oluşturmaktadır. Dolayısıyla coğrafi veri üreticileri için, üretim süreçlerinin otomasyonu son derece önem taşımaktadır. İstenilen nitelik ve içerikteki coğrafi veriler, daha yüksek mekânsal çözünürlüğe sahip coğrafi verilerin bulunması durumunda bu verilerden türetme yolu ile üretilebilir. Türetilen coğrafi verilerin doğruluğu, CBS içindeki veriyi ve dolayısıyla analiz sonuçlarını doğrudan etkilediğinden, basılı haritalardaki durumlarından daha da önemli hale gelmektedir. Bilgisayar destekli harita üretim tekniklerin kullanımı ile birlikte, genelleştirme işlemleri sonuçlarının değerlendirilmesinde kullanılacak yeni yöntemlere olan gereksinim, kendini belirgin olarak göstermektedir. Hata matrisi, raster verilerle yapılan sınıflandırma işlemlerinin doğruluk analizlerinde yoğun olarak kullanılmaktadır. Bu çalışmada; hata matrisinin, türetme coğrafi veri üretim amaçlı gerçekleştirilen coğrafi veri seçim işlemi sonuçlarının değerlendirmesinde kullanımı irdelenmiştir. Türetme coğrafi verilerin üretiminde gerçekleştirilen coğrafi veri seçimi, aslında bir tür sınıflandırma işlemidir. Coğrafi detaylar, "seçilen" ve "seçilmeyen" şeklinde iki ayrı sınıfa ayrıştırılmaktadır. Hata matrisinin, coğrafi veri otomatik seçim sonuçlarının değerlendirilmesinde kullanımı bir uygulama ile gerçekleştirilmiştir. Hata matrisinin, yapılan seçim işlemi sonucu ve doğruluğu hakkında yansız ve kapsamlı bir değerlendirme sunduğu ve bu amaçla kullanılabileceği görülmüştür. Hata matrisinin, seçim işlemi doğrudan etkileyen en iyi seçim ölçütlerinin belirlenmesinde de kullanılabilceği değerlendirilmektedir.

Anahtar Kelimeler: *Coğrafi veri, seçme ve eleme, doğruluk, hata matrisi.*

*Yazışmaların yapılacağı yazar: Bülent ÇETİNKAYA. bulent.cetinkaya@hgk.mil.tr; Tel: (312) 595 24 44.

Bu makale, birinci yazar tarafından İTÜ Fen Bilimleri Enstitüsü, Geomatik Mühendisliği Programında tamamlanmış olan "Topografik haritaların üretiminde eş yükseklik eğrileri, akan su ve su iletim hatları coğrafi verilerin otomasyon süreçleri" adlı doktora tezinden hazırlanmıştır. Makale, metni 19.10.2006 tarihinde dergiye ulaşmış, 01.11.2006 tarihinde basım kararı alınmıştır. Makale ile ilgili tartışmalar 30.11.2008 tarihine kadar dergiye gönderilmelidir.

Usage of error matrix in the accuracy assessment of geographic data selection results

Extended abstract

Nowadays, as being in the information age, the needs for Geographic Information Systems (GIS) and for its main component geographic data are incrementally increasing. Geographic data is the most expensive and time consuming component of GIS (Longley et al., 2001). Therefore construction of automation processes in geographic data production has high importance for data providers. In case of the availability of data having higher spatial resolution, derived geographic data with certain contents and specifications can be produced through generalization. Generalization is important to GIS, not only because of to drive towards automating the process to obtain derived databases, but also because generalization can seriously affect data within a GIS, even more so than within paper maps (Joao, 1998). A fully automated generalization cannot be anticipated at this moment (Spiess et al., 2005). Still lots of efforts are needed to automate the generalization processes (Stoter, 2005). And new automated methods need new quality control methods (Jaakola, 1994; Ruas, 2001). In this study, the usage of error (confusion) matrix is investigated in the accuracy assessment of selection process aiming to obtain derived geographic data.

Geographic data selection process constitutes an important step in the production of derived datasets through generalization. The results of the selection process build the main skeleton of the derived data. Selection process and criteria used are highly depended on the intended usage purpose of the generalized data. In the selection process, unwanted features are eliminated and important features are retained. It is the one of the most immediate differences between maps of the same region at different scales (Joao, 1998). Contextual relations among geographic features should be considered in the selection process. Although selection process of geographic features constitutes an important step in the production of derived datasets, it is not only enough by itself. It should be considered together with other generalization operators (Shea and McMaster, 1989) in map or geographic data production line.

Error matrix is generally used in the accuracy assessments of supervised and unsupervised classifica-

tion of raster data (Congalton ve Green, 1999). An error matrix is a square array of numbers set out in rows and columns which expresses the number of sample units assigned to a particular class relative to actual class as verified by some reference data (Figure 3). The selection processes in the production of derived geographic data can be assessed as a kind of classification. Geographic data is classified in two different categories as "selected" or "eliminated". The usage of error matrix in the accuracy assessment of selection results has been realized with an application. In the application, the automatic selection result of running waters and canals for the production of derived topographic maps has been used.

Geographic data covered by Balıkesir J19 topographic map sheet was chosen as a test data. 1:25.000 scale content data was used for the automatic selection of running waters and canals for the production of 1:100.000 scale Balıkesir J19 topographic map sheet. Last edition of 1:100.000 scale Balıkesir J19 map sheet has been used as a reference data that is needed for verifying the selection result and constructing the error matrix. Due to the subjectivity of generalization, extra comments have been made in defining the accuracy of the selection results of individual features, in case of conflicts occurred between the selection results and the reference data. The lengths of geographic features can also be used in constructing error matrix. In this study, two types of error matrixes were constructed, using feature numbers and feature lengths, respectively. Then the accuracies of classification as a whole and as individual categories have been computed by using the error matrixes constructed.

As a result, it has been found that error matrix provides objective and thorough information for the accuracy assessment of the selection process results for obtaining derived geographic data. It can be used for this purpose. Selection criteria would directly affect the selection results. Determining optimum criteria that will be used in the production lines constitute a critical issue. The Kappa coefficient of agreement (Cohen, 1960), which takes into account all the elements of the error matrix, yields a robust assessment for the accuracy of whole classification process and can be used in determining the optimum selection criteria.

Keywords: Geographic features, selection process, accuracy assessment, error matrix.

Giriş

Özellikle 1990'lı yıllardan sonra büyük bir ivme kazanan bilgisayar teknolojisindeki baş döndürücü gelişmeler, diğer birçok bilim dalı gibi haritacılık bilimini de etkilemiştir. Klasik harita üretim teknikleri, yerini bilgisayar destekli harita üretim tekniklerine bırakmıştır. CBS hayatın her alanında vazgeçilmez bir unsur olmaya başlamıştır. Basılı harita ihtiyacı yanında, CBS'nin temel bileşenlerinden olan Coğrafi Veri Setlerine (CVS) ve sayısal haritalara olan ihtiyaç her geçen gün artmaktadır. Coğrafi verilerin üretimi, CBS bileşenleri arasında en çok zaman ve maliyet gerektiren işlemi oluşturmaktadır (Longley vd., 2001). Bu durum coğrafi veri üreticilerini hızlı, doğru ve ekonomik coğrafi veri üretim arayışlarına itmektedir. Türetme harita ve değişik nitelik ve içerikte coğrafi veri setlerinin üretimi; doğrudan fotogrametrik ya da yerel ölçüm yöntemleri ile yapılabilir (nesne genelleştirilmesi) olmasına rağmen, kendisinden daha yüksek mekânsal (geometrik ve semantik) çözünürlüğe sahip temel ölçek sayısal harita verisi ve coğrafi veri setlerinden (birincil modeller) model genelleştirme ve/veya kartografik genelleştirme yöntemleri kullanılarak üretilmesi daha akılcı ve tercih edilen yöntem olmaktadır (Hardy ve Meter, 2003).

Coğrafi verilerin doğruluğu, sayısal verilerin kullanımı ile birlikte daha da önem kazanmıştır. Coğrafi veriler, CBS'leri ile yapılacak analizleri ve sonuçlarını doğrudan etkilemektedirler. Bilgisayar destekli harita üretim tekniklerinin kullanımı ile birlikte, türetme coğrafi veri üretimi için ihtiyaç duyulan otomasyon süreçlerin geliştirilmesinde ve türetilen coğrafi verilerin doğruluklarının belirlenmesinde çeşitli değerlendirme yöntemlerine gereksinim doğmuştur (Jaakola, 1994; Ruas, 2001).

Bu çalışmada, türetme coğrafi veri üretim amaçlı yapılan seçim işlemleri sonuçlarının değerlendirilmesinde hata matrisinin kullanımı irdelenmiştir. Hata matrisinin, seçim sonucunun doğruluğu hakkında yansız ve kapsamlı bir bilgi sunduğu gerçekleştirilen bir uygulama ile görülmüştür. Uygulamada tez çalışmaları kapsamında gerçekleştirilen, türetme topografik harita üre-

tim amaçlı akan su ve su iletim hatları coğrafi detayların otomatik seçim sonuçları kullanılmış ve hata matrisi ile doğruluk değerlendirilmesi yapılmıştır.

Türetme coğrafi verilerin üretimi

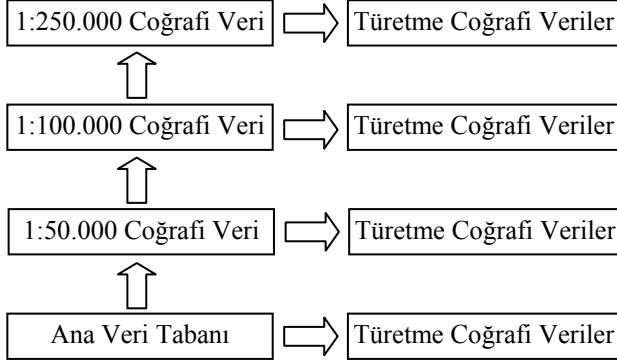
Coğrafi veri toplama, çoğunlukla çok maliyetli olan bir işlemdir. Fotogrametrik değerlendirme, coğrafi veri toplama yöntemlerinden biri olup, geniş alanlara ait coğrafi verilerin belirli ölçeklerde toplanmasında kullanılabilir. Fotogrametrik yöntem ile veri toplayan bilgisayar operatörünün yaptığı işlem, aslında nesne genelleştirilmesidir. Ana coğrafi veri tabanı oluşturulduktan sonra, türetme coğrafi veri setlerini genelleştirme yöntemleri kullanarak üretmek, çoğu zaman daha akılcı ve tercih edilen bir yöntem olmaktadır.

İstenilen nitelik ve içerikte ve geniş ölçek aralığında coğrafi veri setlerinin türetilebildiği dinamik genelleştirme kavramı, yakın gelecekte pek gerçekçi gözükmemektedir. Otomatik genelleştirme işlemlerinin yanı sıra operatör etkileşimli işlemlere olan gereksinim, devam etmektedir (Spiess vd., 2005; Stoter, 2005).

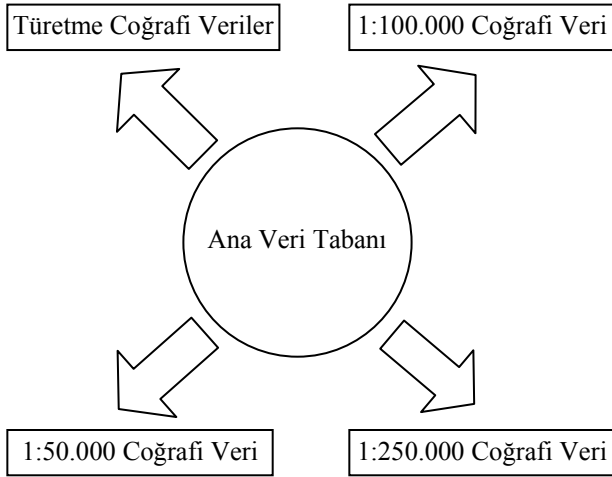
Coğrafi verilerin türetilmesinde alınması gereken önemli kararlardan biri de, basamak (ladder) ya da yıldız (star) yaklaşımlarından hangisinin uygulanacağıdır. Basamak yaklaşımında küçük ölçekli coğrafi veri setleri, kendisinden bir önceki büyük ölçekli coğrafi veri setlerinden türetilmektedirler (Şekil 1). Küçük ölçekli coğrafi setlerin yıldız yaklaşımı ile türetilmesinde ise, tüm türetme coğrafi veri setleri tek bir ana coğrafi veri setinden türetilmektedir (Şekil 2). Türetme coğrafi verilerin türetilmesinde bu iki yaklaşımın karışımını uygulamak da mümkündür (Stoter, 2005).

Seçme ve eleme işlemi, türetme coğrafi verilerin üretiminde çok önemli bir yere sahip olup türetilen coğrafi verilerin iskeletini oluşturmaktadır. Türetme coğrafi detayların üretiminde, sadece seçme ve eleme işlemi yeterli olmamakta, Shea ve McMaster (1989) tarafından sınıflandırılan diğer genelleştirme operatörlerine de ihtiyaç duyulmaktadır. Üretim sisteminde söz konusu genelleştirme operatörleri, geliştirilen diğer yöntemlerle birlikte uygun sırada ve karışımda uygulanmalıdır. Coğra-

fi veriler arasında var olan bağlamsal ilişkiler (contextual relation) seçme ve eleme işleminde dikkate alınmalıdır. Örneğin türetme harita üretim amaçlı akan su coğrafi detayların seçiminde hidrografik ağ yapısı dikkate alınmalıdır.



Şekil 1. Türetme coğrafi veri üretiminde basamak yaklaşımı (Eurogeographics, 2005)



Şekil 2. Türetme coğrafi veri üretiminde yıldız yaklaşımı (Eurogeographics, 2005)

Seçme ve eleme işleminde; hedef ölçekte istenmeyen coğrafi veriler elenmekte, istenen coğrafi detaylar ise seçilmektedir. Seçme ve eleme işlemi; aynı bölgeye ait farklı ölçeklerdeki haritalar arasında en belirgin farkı oluşturmaktadır (Joao, 1998). Türetme coğrafi verinin kullanım amacı, seçme ve eleme işleminde belirleyici rolü üstlenmektedir.

Hata matrisi

Hata matrisi, raster verilerin kontrollü (supervised) ve kontrolsüz (unsupervised) sınıf-

landırma işlemlerinin doğruluk analizlerinde yoğun olarak kullanılmaktadır. Sınıflandırma işleminin tamamlanmasının ardından, sonuç ürününün doğruluğunun belirlenmesi gerekir. Bunun için, çalışma bölgesine ait mevcut haritalar veya GPS ölçmeleri sonucunda elde edilen ve kesin olarak doğruluğuna güvenilen referans verileri kullanılır. Sınıflandırılmış görüntü verisindeki piksellerle bu referans verileri karşılaştırılır. Karşılaştırma işleminde piksellerin rastgele seçilmesi, doğruluk değerlendirmesi yapan kişinin olumlu veya olumsuz olarak hesaplamaya etkisini önler. Sınıflandırma doğruluğunun belirlenmesinde, sınıflandırma aşamasında kullanılmayan eğitim alanlarına ait piksellerin seçilmesi gerekir. Böylece yine sınıflandırma doğruluğunun eğitim alanları ile olumlu bir şekilde etkilenmesi önlenmiş olur. Hata matrisi yardımı ile sınıflandırılmış piksellerin doğruluk yüzdesi hesaplanır (Congalton ve Green, 1999). Şekil 3'te hata matrisinin genel şekli verilmektedir.

Sınıf	1	...	n	Toplam (Sınıflandırılan)
1	X_{11}	...	X_{1n}	$\sum_{j=1}^n X_{1j}$
...
n	X_{n1}	...	X_{nn}	$\sum_{j=1}^n X_{nj}$
Toplam (Referans)	$\sum_{i=1}^n X_{i1}$...	$\sum_{i=1}^n X_{in}$	$\sum_{i=1}^n \sum_{j=1}^n X_{ij}$

Şekil 3. Hata matrisi

Birçok hata ölçüsü, bu hata matrisinden elde edilebilir. Bunlardan bazıları toplam doğruluk ile üretici ve kullanıcı doğruluklarıdır. Hata matrislerinden elde edilen sınıflar arası doğrulukların analiz edilmesinde K kappa katsayısı kullanılır. K kappa katsayısı, 0 ile 1 arasında değer alır ve hata matrisinin satır ve sütun toplamları ile köşegeni üzerindeki elemanlar kullanılarak hesaplanır. K'nın 1 olması istenen en ideal durumdur. Kappa değerinin hesaplanması aşağıda belirtilen biçimde gerçekleştirilebilir (Cohen, 1960);

Kappa katsayısının hesaplanması:

$$K = \frac{M \sum_{i=1}^r X_{ii} - \sum_{i=1}^r X_{i+} X_{+i}}{M^2 - \sum_{i=1}^r X_{i+} X_{+i}} \quad (1)$$

- r : Hata matrisindeki satır sayısı,
 X_{ii} : Hata matrisinin i nci satır ve i nci sütun elemanı,
 X_{i+} : Hata matrisinin i nci satır elemanlarının toplamı,
 X_{+i} : Hata matrisinin i nci sütun elemanlarının toplamı,
 M : Hata matrisi elemanlarının toplamı.

Uygulama ve sonuçların değerlendirilmesi

Hata matrisinin, türetme coğrafi veri seçim işlemi sonuçlarının değerlendirilmesinde kullanımı bir uygulama ile gerçekleştirilmiştir. Uygulamada tez çalışmaları kapsamında gerçekleştirilen, türetme topografik haritaların üretiminde akan su ve su iletim hatları detayların otomatik seçim sonuçları kullanılmış ve hata matrisi ile doğruluk değerlendirilmeleri yapılmıştır. Seçim işleminde, 1:25.000 ölçek içerikli coğrafi veri kullanılarak 1:100.000 ölçekli Standart Topografik Haritaların (STH) üretilmesi amaçlanmıştır.

Uygulamada, 1:100.000 ölçekli Balıkesir j19 STH paftasının kapsadığı bölge test verisi olarak kullanılmıştır. Otomatik seçim öncesi, 1:25.000 ölçek içerikli akan su ve su iletim hatları coğrafi detayların bir kısmı Şekil 4'te sunulmaktadır. 1:100.000 ölçekli topografik harita üretim amaçlı yapılan akan su ve su iletim hatları coğrafi verilerin otomatik seçim sonrası durumu ise Şekil 5'te gösterilmektedir. Otomatik seçim işleminin geliştirilmesinde ve sonucunun değerlendirilmesinde doğruluk hesaplama ve değerlendirme yöntemlerine gereksinim belirgin olarak kendini göstermektedir.

Akan su ve su iletim hatları otomatik seçimin doğruluğunu belirlemede, raster veri sınıflandırma işlemlerin doğruluk analizlerinde yoğun olarak kullanılan hata matrisinden yararlanılabilir. Yapılan coğrafi veri seçim işlemi, aslında bir sınıflan-

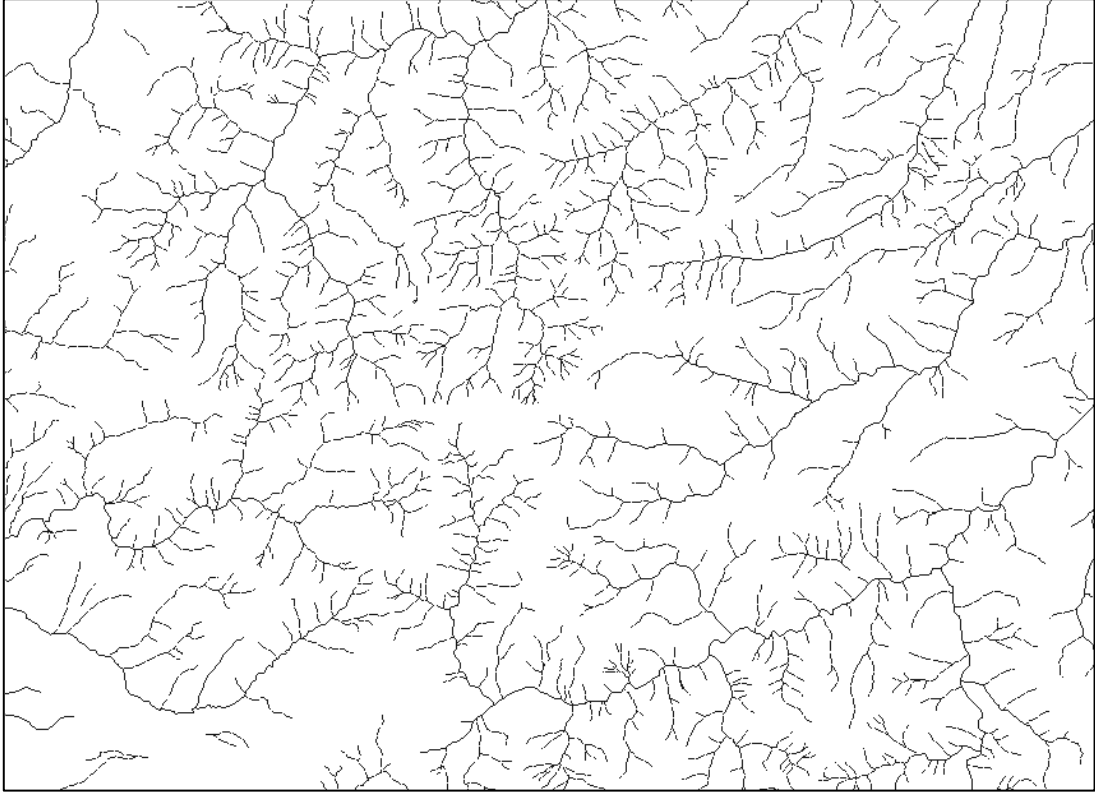
dırma işlemidir. Seçim işleminde, akan su ve su iletim hatları coğrafi detayları; “seçilen” ve “seçilmeyen” şeklinde iki ayrı sınıfa ayrıştırılmaktadır.

Hata matrisini oluşturmak ve seçim sonuçlarını değerlendirmek için referans veri gereklidir. Referans veri, her bir detayın seçim doğruluğunu belirlemede kullanılır. Bu çalışmada referans veri olarak 1:100.000 ölçekli Balıkesir-j19 paftasının klasik genelleştirme yöntemiyle üretilen son basımına ait mavi kalıp taranarak kullanılmıştır. Mavi kalıp, harita basımına esas olan ve haritadaki mavi renk ile gösterilen detayları içeren bir katmandır. Otomatik seçim sonuçlarının değerlendirilmesinde büyük ölçüde faydalanılan mavi kalıp, her durumda mutlak doğru şeklinde kabul edilmemiştir. Seçim sonucunun mavi kalıp ile uyuşmadığı durumlarda, detayın sınıflandırma sonucu, klasik genelleştirmede çalışmış uzman kartograf ile birlikte sayısal verinin sağladığı olanaklar da göz önünde tutularak yapılan yorum ile kesinlik kazandırılmıştır. Akan sular, arazide pek değişmediği kabul edilen coğrafi verilerdir. Fakat üretimde, söz konusu verilerin kontrolü mutlaka yapılmalıdır.

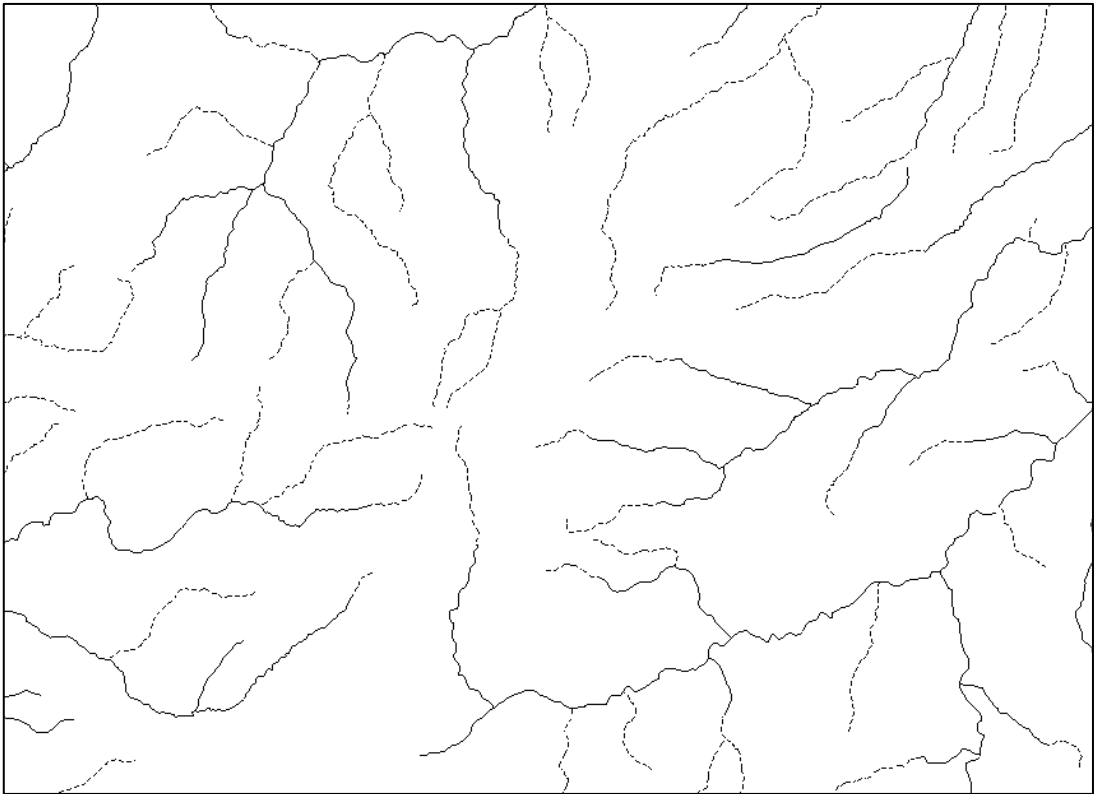
Seçim işlemine girdi verilerin analizi, seçim sonuçlarının değerlendirilmesine olumlu katkılar sağlayacaktır. Girdi veri, 1:25.000 ölçekli STH'ların üretiminde doğrudan kullanılan akan su ve su iletim hatları vektör verileridir. Akan su ve su iletim hatları girdi verisi, farklı özneliklere ve çok farklı uzunluklara sahip coğrafi detayları içermektedir (Şekil 4). Küçük uzunluklara sahip detaylar, genelde detayların birleşim yerlerindeki bağlantı çizgileridir ve hidrografik ağ yapısını oluşturmada önemli görevler üstlenmektedirler. Girdi verileri ile ilgili istatistik Tablo 1'de verilmektedir.

Tablo 1. Akan su ve su iletim hatları girdi veri istatistikleri

Detay sayısı	12 253
Toplam uzunluk	5 339 902 m
Ortalama uzunluk	436 m
Standart sapma	409 m
Minimum uzunluk	1 m
Maksimum uzunluk	6 356 m



Şekil 4. Akan su ve su iletim hatları (otomatik seçim öncesi)



Şekil 5. Akan su ve su iletim hatları (otomatik seçim sonrası)

Her bir seçimin doğruluğunun değerlendirilmesi sonucu oluşturulan hata matrisi Tablo 2’de gösterilmektedir. Tablo 2’de sunulan hata matrisinin oluşturulmasında, pafta içerisindeki tüm coğrafi veri kullanılmıştır. Hata matrisi, tüm kütleliyi temsil eden örnek bir küme kullanılarak da oluşturulabilir.

Tablo 2. Akan su ve su iletim hatları otomatik seçim hata matrisi (detay sayıları)

DETAY SAYILARI	Referans			
	Seçilen	Seçilmeyen	Toplam	
Otomatik Seçim	Seçilen	3691	239	3930
	Seçilmeyen	190	8133	8323
	Toplam	3881	8372	12253
Tespit Doğruluğu	95.10	97.15		
İsabet Doğruluğu	93.92	97.72		
Ortalama Tespit Doğruluğu				96.12
Ortalama İsabet Doğruluğu				95.82
Genel Doğruluk				96.50
Bileşik Tespit Doğruluğu				96.31
Bileşik İsabet Doğruluğu				96.16
Kappa katsayısı				91.94

Hata matrisinde, akan su ve su iletim hatları coğrafi verisine ait seçim sonuçları sunulmaktadır. Girdi verisindeki hatadan dolayı, 3 çizgi parçasında bilgisayar operatörü tarafından düzeltme (editleme) yapılmış olup bu çizgilerden ikisi program tarafından otomatik seçilen diğeri ise operatör etkileşimli olarak sonradan seçilen çizgiler arasındadır. Ayrıca hidrografya alandan çizgiye geometrik ve semantik dönüşüm işlemi sonucu 98 çizgi detayı oluşmuş olup, söz konusu çizgi detayları ayrı bir işlemin sonucu olduğundan, bu değerlendirmeye dahil edilmemiştir. Söz konusu alandan çizgiye dönüşüm işlemi, seçim işlemi için gerekli olan hidrografya ağ yapısını oluşturmada kullanılmıştır.

Tablo 2’de sunulan hata matrisi kullanılarak aşağıda tanımlanan doğruluk hesaplamaları yapılmıştır:

• **Tespit doğruluğu :**

Üretici doğruluğu şeklinde de tanımlanabilir. Tanımlanan sınıf için doğru olarak sınıflandırılanların sayısının, referans veride söz konusu sınıfa ait olanların sayısına bölümüdür. Bu doğruluk, ıskalamanın (omission) da bir ölçüsüdür (Story ve Congalton, 1986). Seçilen detaylar için bu doğruluk, 3691/3881 veya %95.10’dır (Tablo 2).

• **İsabet doğruluğu :**

Kullanıcı doğruluğu şeklinde de tanımlanabilir. Tanımlanan sınıf için doğru olarak sınıflandırılanların sayısının, söz konusu sınıfa sınıflandırılanların toplam sayısına bölümüdür. Bu doğruluk, yanlış sınıflandırmanın (commission) da göstergesidir. Seçilen detayların isabet doğruluğu 3691/3930 veya %93.92’dir (Tablo 2).

• **Genel doğruluk :**

Doğru olarak sınıflandırılanların toplam sayısının (hata matrisindeki diyagonal elemanlar toplamının) toplam örnek sayısına bölümüdür. Seçim işleminin genel doğruluğu (3691+8133)/12253 veya %96.50’dir (Tablo 2). Bu doğruluk, herhangi bir sınıfın seçim doğruluğunu göstermek için tüm seçim işleminin doğruluğunu göstermektedir.

• **Ortalama tespit doğruluğu :**

Ortalama üretici doğruluğu şeklinde de tanımlanabilir. Her bir sınıfa ait tespit doğruluklarının ortalamasıdır. Ortalama tespit doğruluğu (95.10+97.15)/2 veya %96.12’dir (Tablo 2). Genel doğruluğun aksine bu doğruluk, küçük sayıda bulunan sınıfa doğru kayma (bias) eğilimindedir.

• **Ortalama isabet doğruluğu :**

Ortalama kullanıcı doğruluğu şeklinde de tanımlanabilir. Her bir sınıfa ait isabet doğruluklarının ortalamasıdır. Ortalama isabet doğruluğu (93.92+97.72)/2 veya %95.82’dir (Tablo 2). Genel doğruluğun aksine bu doğruluk, küçük sayıda bulunan sınıfa doğru kayma eğilimindedir.

• **Bileşik tespit doğruluğu :**

Bileşik üretici doğruluğu şeklinde de tanımlanabilir. Ortalama tespit doğruluğu ile genel doğruluğun ortalamasıdır. Bileşik tespit doğruluğu $(96.12+96.50)/2$ veya %96.31'dir (Tablo 2). Bu doğruluk, ortalama ve genel doğrulukların kayma (bias) eğilimlerini azaltmak için kullanılır.

• **Bileşik isabet doğruluğu :**

Bileşik kullanıcı doğruluğu şeklinde de tanımlanabilir. Ortalama isabet doğruluğu ile genel doğruluğun ortalamasıdır. Bileşik tespit doğruluğu $(95.82+96.50)/2$ veya %96.16'dir (Tablo 2). Bu doğruluk, ortalama ve genel doğrulukların kayma eğilimlerini azaltmak için kullanılır.

• **Kappa katsayısı :**

Cohen (1960) tarafından geliştirilen Kappa katsayısı, hesaplanmasında hata matrisinin tüm elemanları kullandığından, ıskalama (ommission) ve isabet (commission) hatalarının her ikisini de dikkate almaktadır (Lillesand ve Kiefer, 1994). Her bir sınıflandırma işlemi için Kappa katsayısı hesaplanır. Bu katsayısı, sınıflandırma işlemi referans veri ile karşılaştırarak sınıflandırmanın doğruluğu ve güvenilirliği hakkında bilgi vermektedir. Kappa katsayısı, coğrafi veri işlemleri için gerekli olan eşik değer (threshold) belirleme hesaplamalarında da yoğun olarak kullanılmaktadır (Çetinkaya ve Türker, 2005).

Bir başka deyişle, Kappa değeri; sınıflandırma işlemi sonucu oluşan hatanın, tamamen rastgele sınıflandırmadaki hataya göre, oransal azalma miktarını ifade etmektedir. Hata matrisinden hesaplanan 0.9194 Kappa değeri; sınıflandırma işleminin, tamamen rastgele sınıflandırmada oluşacak hataların %91.94'ünden arındığını göstermektedir (Tablo 2).

Verideki akan su ve su iletim hatları detayların uzunlukları çok farklılık göstermektedir. Her ne kadar detay sayısı, otomasyon olmaksızın yapılacak seçim veya silme işleminde bilgisayar operatörünün yapacağı işler hakkında bilgi verse de, değerlendirmede sadece detay sayısına bakmak yanıltıcı olabilir. Dolayısıyla, seçim işlemi daha iyi değerlendirmek amacıyla, detayların uzunluklarından oluşan hata matrisi de elde edilmiştir (Tablo 3).

Tablo 3. Akan su ve su iletim hatları otomatik seçim hata matrisi (detay uzunlukları)

UZUNLUK (km)	Referans			
	Seçilen	Seçilmeyen	Toplam	
Otomatik Seçim	Seçilen	1918.1	156.9	2075
	Seçilmeyen	75.1	3189.8	3264.9
	Toplam	1993.2	3346.7	5339.9
Tespit Doğruluğu	96.23	95.31		
İsabet Doğruluğu	92.44	97.70		
Ortalama Tespit Doğruluğu				95.77
Ortalama İsabet Doğruluğu				95.07
Genel Doğruluk				95.66
Bileşik Tespit Doğruluğu				95.71
Bileşik İsabet Doğruluğu				95.36
Kappa katsayısı				90.79

Yapılan sınıflandırma işlemi değerlendirildiğinde, seçim işleminin yaklaşık %96.50 ve %95.66 genel doğruluklarla gerçekleştirildiği ve otomatik seçim sonrası toplam detay sayısı ve uzunluğun Töpfer yasasına (Töpfer, 1974) uygun olduğu görülmektedir. Ancak doğruluk ne kadar yüksek olursa olsun, coğrafi veri üretimi söz konusu olduğunda, sonuç ürün mutlaka kontrol sürecinden geçmelidir. Burada kontrolcü kartografa yaklaşık %96 doğrulukla seçimi gerçekleştirmiş coğrafi veri sunulmaktadır. Kartograf kontrol ve düzeltmede, alınmaması gereken detayları silmekte ve ilave alınması gereken detayları altlık veriden kopyalamaktadır. Akan sular seçim işlemi gerçekleşen Balıkesir-J19 paftası için seçilen 239 detay operatör etkileşimli olarak silinmiş ve seçilmeyen 190 detay ilave olarak alınmıştır (Tablo 2). Operatör etkileşimli işlemlere bakıldığında silme işleminin daha fazla olduğu gözükmemektedir. Otomatik seçim işlemi sonucu, arazi karakteristiğini yansıtan ve büyük olasılıkla alınması gereken veri kartografa sunulmakta ve kartografin yoğun veri ile uğraşma zorunluluğu ortadan kaldırılmaktadır. Otomatik seçim işlemi ile genelleştirmenin öznel yapısından büyük ölçüde kurtulmakta ve üretimde kalite, otomasyon ve standardizasyona önemli katkılar yapılmaktadır.

Hata matrisi, seçim işlemi sonuçlarını değerlendirmede daha kapsamlı ve yansız bir bilgi sunmaktadır. Yapılan seçim işlemi sonucunda tüm detayların seçilmesi durumunda, seçilen detayların tespit doğruluğu %100 olurken, değerlendirmelerde olabilecek bu tür yanlışlardan, seçilen detaylar için hesaplanan seçim isabet doğruluğu kurtarmaktadır. Ayrıca seçilmeyen detaylar için hesaplanan tespit ve isabet doğrulukları ile hata matrisinden türetilen diğer doğruluklar, seçim sonuçlarının daha sistematik ve kapsamlı olarak değerlendirilmesine olanak sağlamaktadır.

Sonuç ve öneriler

Bilgisayar destekli harita üretim tekniklerinin kullanımı ile birlikte türetme coğrafi verilerin üretimi ve otomasyonu daha da önem kazanmıştır. Türetme coğrafi verilerin üretimi ve otomasyonu için birçok ilave yeni genelleştirme yöntem, yaklaşım ve algoritmalara ihtiyaç vardır (Hardy ve Meyer, 2003; Stoter, 2005). Oluşturulan otomasyon süreçlerinin ve sonuçlarının yansız olarak değerlendirilmesinde de birçok ilave doğruluk değerlendirme yöntemlerine gereksinim duyulmaktadır (Ruas, 2001).

Bu çalışmada, raster veri sınıflandırma işlemlerin doğruluk analizlerinde yoğun olarak kullanılan hata matrisinin, türetme coğrafi verilerin seçim işlemi sonuçlarının değerlendirilmesinde de kullanılabilmesi yapılan uygulama ile görülmüştür. Hata matrisi, seçim sonuçlarının değerlendirilmesinde, yansız ve kapsamlı bir bilgi sunmaktadır. Üretimde otomasyon oranı hakkında daha detaylı bir bilgi vermektedir. Doğru olarak seçilen ve seçilmeyen detaylar ile yanlış olarak seçilen ve seçilmeyen detayların seçim doğrulukları hakkında daha bütün bir bilgi sunmakta ve seçim sonuçların değerlendirilmesinde çeşitli yanlışlardan kurtarmaktadır. Coğrafi veri seçim sonuçlarının değerlendirilmesinde kullanılan hata matrisinde, detay sayılarının kullanıldığı gibi, detay uzunlukları da kullanılabilir.

Sonuç olarak hata matrisinin, türetme coğrafi verilerin üretiminde gerçekleştirilen seçim işlemi sonuçlarının değerlendirilmesinde kullanıla-

bileceği ve değerlendirme için kapsamlı ve yansız bir bilgi sunduğu görülmüştür. Otomatik seçim işleminde kullanılan seçim ölçütleri, seçim sonuçlarını doğrudan etkilemektedir. Eşik değer belirleme hesaplarında kullanılan Kappa katsayısı değerinin, coğrafi veri üretiminde oldukça önemli olan en iyi seçim ölçütlerinin seçim işlemi için belirlenmesinde de kullanılabilmesi değerlendirilmektedir.

Kaynaklar

- Cohen, J., (1960). A Coefficient of Agreement for Nominal Scales, *Educational and Psychological Measurement*, **20**, 37-46.
- Congalton R. G. ve Green K., (1999). *Assessing the Accuracy of Remotely Sensed Data: Principles and practices*. Boca Raton, Florida: Lewis Publications.
- Çetinkaya, B. ve Türker, M., (2005). Automatic detection of earthquake-damaged buildings using DEMs created from pre- and post-earthquake stereo aerial photographs, *International Journal of Remote Sensing*, **26**, 823-832.
- EuroGeographics, (2005). Generalisation Processes: a benchmark study of the expert group on quality. http://www.eurogeographics.org/eng/05_quality_reports.asp, (01.02.2005)
- Hardy, P. ve Meyer, M., (2003). Efficient Map Production by Re-Engineering and Generalising Your Data Assets, *Proceedings*, Cambridge Conference for National Mapping Organisations, July 2003, Cambridge, UK.
- Jaakola, O., (1994). Finnish Corine Land Cover - A Feasibility Study of Automatic Generalization and Data Quality Assessment, Reports of the Finnish Geodetic Institute, Helsinki.
- Joao, E. M., (1998). *Causes and Consequences of Map Generalization*, Taylor and Francis: London, England.
- Lillesand, T. M. ve Kiefer, R. W., (1994). *Remote Sensing and Image Interpretation*, 3rd edition, John Wiley and Sons, Inc.
- Longley, P. A., Goodchild M. F., Maguire D. J. ve Rhind D. W., (2001). *Geographic Information Systems and Science*, ESRI Pres.
- Ruas, A., (2001). Automatic Generalisation Project: Learning process from interactive generalisation, Reported by Anne Ruas, OEEPE, March 2001, Official Publication No 39.
- Shea, K. S. ve McMaster, R. B., (1989). Cartographic Generalization in a Digital Environment: When and How to Generalize, *Proceedings*,

- AUTO-CARTO 9, 9th International Symposium on Computer Assisted Cartography, Baltimore, 56-67.
- Spiess, E., Baumgartner, U., Arn, S. ve Vez, C., (2005). *Topographic Maps - Map Graphics and Generalisation*, Swiss Society of Cartography, Cartographic Publication Series No. 17, Zurich.
- Story, M. ve Congalton, R. G., (1986). Accuracy Assessment: A User's Perspective, *Photogrammetric Engineering and Remote Sensing*, **52**, 397-399.
- Stoter, J. E., (2005). Generalisation: The Gap Between Research and Practice, *Proceedings*, 8th ICA Workshop on Generalisation and Multiple Representation, A Coruna, Spain.
- Töpfer, F., (1974). *Kartographische Generalisierung*, Haack-Gotha, Leipzig.