

T.C.
ERCIYES ÜNİVERSİTESİ
BİLİMSEL ARAŞTIRMA PROJELERİ
KOORDİNASYON BİRİMİ

**VERİ MADENCİĞİ YAKLAŞIMLARI KULLANILARAK HİDROLOJİK VE
İKLİMSEL PARAMETRELER ARASINDAKİ İLİŞKİLERİN BELİRLENMESİ**

Proje No: FBA-09-866

Araştırma Projesi

SONUÇ RAPORU

Proje Yürütücüsü:

Yrd. Doç. Dr. Filiz DADAŞER ÇELİK
Mühendislik Fakültesi/Çevre Mühendisliği Bölümü

Yrd. Doç. Dr. Mete ÇELİK
Mühendislik Fakültesi/Bilgisayar Mühendisliği Bölümü

Ocak 2012

KAYSERİ

TEŞEKKÜR

Bu çalışma Erciyes Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimi tarafından FBA-09-866 no'lu araştırma projesi olarak desteklenmiştir. Bu destekten dolayı Erciyes Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimi'ne teşekkür ederiz. Proje kapsamında yürütülen çalışmalarındaki katkılarından dolayı yüksek lisans öğrencisi Eda CENGİZ ve yüksek lisans öğrencisi Ahmet Şakir DOKUZ'a teşekkür ederiz.

İÇİNDEKİLER

TEŞEKKÜR	İİİ
İÇİNDEKİLER	İV
TABLÖLAR LİSTESİ	Vİ
ŞEKİLLER LİSTESİ	Vİİİ
ÖZET	X
ABSTRACT	Xİ
1. GİRİŞ / AMAÇ VE KAPSAM	1
2. GENEL BİLGİLER	4
2.1. Türkiye Hidrolojik ve İklim Verileri Kullanılarak Daha Önce Yapılan Çalışmalar	4
2.1.1. Verilerin Karakterlerinin ve Mekansal Değişimlerinin Belirlenmesi	4
2.1.2. Veri Serilerin Homojenliğinin Değerlendirilmesi	5
2.1.3. Eğilim Analizleri	6
2.1.4. Veri Serilerinde Büyük Ölçekli Atmosferik Salınım Etkilerinin Araştırılması	8
2.1.5. Kümeleme Analizleri	10
2.2. Hidrolojik ve İklimsel Veri Analizinde Veri Madenciliği Yaklaşımı	11
2.3. Literatür Analizi	17
3. GEREÇ VE YÖNTEM	19
3.1. Veri Toplama ve Veri Kalitesinin Değerlendirilmesi	20
3.1.1. Homojenlik Analizleri	21
3.2. Veri Ön-İşleme	23
3.2.1. İstasyon Eşleştirme	23
3.2.2. Veri Birleştirme	24
3.2.3. Verilerin Mevsimselliğinin Giderilmesi ve Normalleştirme	24
3.2.4. Veri Ayırıklaştırma	24
3.3. Hidrolojik ve İklimsel İlişkilerin ve Örüntülerin Tanımlanması	27
3.4. Değerlendirme Ölçütü Tanımlama	27

3.5. Hızlı ve Özgün Algoritmaların Tasarlanması -----	30
3.6. İlişki ve Örüntü Madenciliği-----	32
3.7. Yöntemlerin ve Örüntülerin Değerlendirmesi-----	32
4. BULGULAR-----	33
4.1. Veri Toplama ve Veri Kalitesinin Değerlendirilmesi -----	33
4.1.1. Veritabanının Oluşturulması ve Verilerin Veritabanına Yerleştirilmesi -----	33
4.1.2. Verilerin Kalite Kontrolünün Yapılması-----	34
4.2. Veri Önileme-----	44
4.2.1. Veri Birleştirme-----	44
4.2.2. İstasyon Eşleştirme-----	44
4.2.3. Veri Ayırıştırma -----	54
4.2.4. Mevsimselliğin Giderilmesi -----	57
4.3. Hidrolojik ve İklimsel İlişkilerin ve Örüntülerin Tanımlanması -----	57
4.4. Değerlendirme Ölçütü Tanımlama-----	58
4.4.1. Destek Ölçütü-----	58
4.4.2. Hidrometeorolojik Yaygınlık Ölçütü -----	58
4.5. Hızlı ve Özgün Algoritmaların Tasarlanması -----	59
4.6. İlişki ve Örüntü Madenciliği-----	61
4.6.1. Bütün veriler kullanılarak elde edilen sonuçlar -----	62
4.6.2. Düşük ve Yüksek Akımlar için Sonuçlar -----	65
4.6.3. Yaz ve Kış Dönemleri için Elde Edilen Sonuçlar -----	71
4.7. Yöntemlerinin ve Örüntülerin Değerlendirilmesi-----	80
5. TARTIŞMA VE SONUÇ-----	81
KAYNAKLAR-----	84

TABLolar LİSTESİ

Tablo 2.1. Literatürde bulunan ve hidrolojik ve iklimsel parametreler arasındaki ilişkileri açıklayan örüntüler	16
Tablo 2.2. Literatürde bulunan ve hidrolojik ve iklimsel parametreler arasındaki ilişkileri keşfetmek için kullanılan algoritmalar	17
Tablo 3.1: Bir istasyona ait 10 günlük hidrolojik ve iklimsel örnek veriler.....	29
Tablo 4.1. 1975-2000 dönemi için homojenlik analizi sonuçları (Toplam istasyon sayısı rüzgar hızı için 208, diğer parametreler için 213'tür)	38
Tablo 4.2. 1975-2000 dönemi için homojenlik analizi sonuçlarının sınıflandırılması	38
Tablo 4.3. Eşleştirilen akım gözlem ve meteoroloji istasyonları	45
Tablo 4.4. Akım ve yağış, sıcaklık, bağıl nem, rüzgar hızı ve güneşlenme şiddeti arasındaki korelasyon katsayıları	47
Tablo 4.5. Akım ve geçmiş yıllara ait yağış değerleri arasındaki korelasyon katsayıları.....	49
Tablo 4.6. Akım ve geçmiş yıllara ait kümülatif yağış değerleri (1 ,2, 3, 4, 5 yıllık toplam) arasındaki korelasyon katsayıları.....	50
Tablo 4.7. Akım ve geçmiş yıllara ait sıcaklık değerleri arasındaki korelasyon katsayıları	52
Tablo 4.8. Yağış, ortalama sıcaklık, ort. nem, rüzgar hızı ve güneşlenme şiddeti verilerinin maksimum, minimum, ortalama ve standart sapma değerleri.....	53
Tablo 4.9. Kümülatif yağış değerleri ile akım arasındaki korelasyonların maksimum, minimum, ortalama ve standart sapma değerleri.....	53
Tablo 4.10. Hidrolojik ve iklimsel parametreler kullanılarak örüntü madenciliği algoritması	60
Tablo 4.11. Farklı destek ve hidrometeorolojik yaygınlık ölçütü değerleri için kural sayıları	62
Tablo 4.12. Tüm verilerin kullanılması durumunda elde edilen kurallar (destek: 0.15, hidrometeorolojik yaygınlık ölçütü :0.50). D: Düşük, O: Orta, Y: Yüksek	63

Tablo 4.13. Farklı destek ve hidrometeorolojik yaygınlık ölçütü değerleri için kural sayıları (Düşük Akım).....	66
Tablo 4.14. Farklı destek ve hidrometeorolojik yaygınlık ölçütü değerleri için kural sayıları (Yüksek Akım).....	67
Tablo 4.15. Tüm verilerin kullanılması durumunda elde edilen kurallar (Düşük Akım) (destek: 0.15, hidrometeorolojik yaygınlık ölçütü :0.50). D: Düşük, O: Orta, Y: Yüksek, AY: Aşırı Yüksek, OD: Orta Düşük	68
Tablo 4.16. Tüm verilerin kullanılması durumunda elde edilen kurallar (Yüksek Akım) (destek: 0.15, hidrometeorolojik yaygınlık ölçütü :0.50). D: Düşük, O: Orta, Y: Yüksek, AY: Aşırı Yüksek, AY: Aşırı Yüksek.....	70
Tablo 4.17. Farklı destek ve hidrometeorolojik yaygınlık ölçütü değerleri için kural sayıları (Yaz Mevsimi).....	72
Tablo 4.18. Farklı destek ve hidrometeorolojik yaygınlık ölçütü değerleri için kural sayıları (Kış Mevsimi)	73
Tablo 4.19. Yaz mevsimi için elde edilen kurallar (destek: 0.15, hidrometeorolojik yaygınlık ölçütü :0.50). D: Düşük, O: Orta, Y: Yüksek.....	74
Tablo 4.20. Kış mevsimi için elde edilen kurallar (destek: 0.15, hidrometeorolojik yaygınlık ölçütü :0.50). D: Düşük, O: Orta, Y: Yüksek.....	78

ŞEKİLLER LİSTESİ

Şekil 3.1. Projede uygulanan veri madenciliği yönteminin aşamaları.....	20
Şekil 3.2. DBSCAN algoritması ile kümeleme.....	26
Şekil 3.3. Apriori algoritması tarafından üretilen olası örüntüler (kesik çizgili bölge içerisindeki koyu renkli örüntüler Tablo 1’ de verilen veri kümesin den keşfedilen ve eşik değeri 4’ü sağlayan sık örüntü ve ilişkileri gösterir.	31
Şekil 4.1. Veritabanı için Varlık-İlişki (ER-Entity-Relationship) diyagramı.....	34
Şekil 4.2. Çalışmada kullanılmak üzere seçilen 218 meteoroloji istasyonunun konumları	35
Şekil 4.3. Çalışmada kullanılmak üzere seçilen akım gözlem istasyonlarının konumları	36
Şekil 4.4. Çalışmada kullanılmak üzere seçilen 213 meteoroloji istasyonunun konumları	37
Şekil 4.5. Yağış verilerinin homojenlik analizi sonuçları	39
Şekil 4.6. Ortalama sıcaklık verilerinin homojenlik analizi sonuçları	40
Şekil 4.7. Maksimum sıcaklık verilerinin homojenlik analizi sonuçları.....	40
Şekil 4.8. Minimum sıcaklık verilerinin homojenlik analizi sonuçları	41
Şekil 4.9. Bağıl nem verilerinin homojenlik analizi sonuçları	42
Şekil 4.10. Rüzgar hızı verilerinin homojenlik analizi sonuçları.....	42
Şekil 4.11. Eşleştirilen akım gözlem ve meteoroloji istasyonlarının konumları.....	45
Şekil 4.12. Akım ve yağış, sıcaklık, bağıl nem, rüzgar hızı ve güneşlenme şiddeti arasındaki korelasyon katsayılarının konuma bağlı aldıkları değerler.....	48
Şekil 4.13. Yüzdelik dilimler yöntemi ile ayrıklaştırılmış 17836 no’lu istasyona ait aylık ortalama sıcaklık verileri	55
Şekil 4.14. İstatistiksel yöntem ile ayrıklaştırılmış 17836 no’lu istasyona ait aylık ortalama sıcaklık verileri.....	55

Şekil 4.15. Kümeleme (DBSCAN) algortiması kullanılarak ile ayrıklaştırılmış 17836 no'lu istasyona ait aylık ortalama sıcaklık verileri	56
Şekil 4.16. Farklı destek ve hidrometeorolojik yaygınlık ölçütü değerleri için kural sayıları .	62
Şekil 4.17. Tüm verilerin kullanılması durumunda elde edilen kuralların coğrafi dağılımı (destek: 0.15, hidrometeorolojik yaygınlık ölçütü :0.50)	65
Şekil 4.18. Farklı destek ve hidrometeorolojik yaygınlık ölçütü değerleri için kural sayıları (Düşük Akım).....	66
Şekil 4.19. Farklı destek ve hidrometeorolojik yaygınlık ölçütü değerleri için kural sayıları (Yüksek Akım).....	67
Şekil 4.20. Düşük akımları için elde edilen kuralların coğrafi dağılımı (destek: 0.15, hidrometeorolojik yaygınlık ölçütü :0.50).....	69
Şekil 4.21. Yüksek akımları için elde edilen kuralların coğrafi dağılımı (destek: 0.15, hidrometeorolojik yaygınlık ölçütü :0.50).....	71
Şekil 4.22. Farklı destek ve hidrometeorolojik yaygınlık ölçütü değerleri için kural sayıları (Yaz Mevsimi).....	72
Şekil 4.23. Farklı destek ve hidrometeorolojik yaygınlık ölçütü değerleri için kural sayıları (Kış Mevsimi)	73
Şekil 4.25. Kış mevsimi akımları için elde edilen kuralların coğrafi dağılımı (destek: 0.15, hidrometeorolojik yaygınlık ölçütü :0.50).....	79

ÖZET

Bu projenin amacı Türkiye için hidrolojik ve iklimsel parametreler arasındaki ilişkilerin tespit edilmesi için veri madenciliği yöntemlerinin geliştirilmesidir. Hidrolojik ve iklimsel verilerin analizleri için bugüne kadar yapılan çalışmaların üç temel eksikliği bulunmaktadır: 1) Hidrolojik ve iklimsel parametreler arasındaki ilişkiler yeterli düzeyde incelenmemiştir; 2) Hidrolojik ve iklimsel verilerin analizinde büyük-ölçekli verilerin analizinde yetersiz olan istatistiksel yöntemler ağırlıklı olarak kullanılmıştır. 3) Birçok çalışmada verilerin zamansal boyutuna odaklanılmış ve verilerin mekansal ve mekan-zamansal boyutları ihmal edilmiştir. Bu projede hidrolojik ve iklimsel veriler beraber analiz edilerek, bu verilerin birlikte oluşturdukları ilişkilerinin ve hidrometeorolojik örüntülerin keşfi sağlanmıştır. Bu amaçla veri madenciliği yöntemleri arasında yer alan birliktelik analizi kullanılmıştır. Veri madenciliği, analizi manuel olarak mümkün olmayan veya analizi çok zaman alan büyük ve çok değişkenli veri kümelerinden daha önce keşfedilmemiş, ilginç ve anlamlı bilgileri çıkarmak için otomatik veya yarı-otomatik yöntemleri geliştirmeyi amaçlamaktadır. Birliktelik analizi bu kapsamda değişik parametreler arasındaki ilişki ve örüntülerin belirlenmesinde neden-sonuç ilişkilerini de ifade eden kuralların keşfinde kullanılır. Bu çalışmada kullanılacak hidrolojik ve meteorolojik veriler çeşitli devlet kurumlarından (Devlet Meteoroloji İşleri Genel Müdürlüğü-DMİ, Elektrik İşleri Etüd İdaresi Genel Müdürlüğü-EİE) elde edilmiştir. Veri kalitesinin değerlendirilmesi ve veri madenciliği için hazırlanmasından sonra, birliktelik örüntülerinin keşfi için hızlı ve özgün bir algoritma hazırlanmıştır. Geliştirilen yöntemler Türkiye akım ve iklimsel verileri üzerinde uygulanmıştır. Çalışma sonucunda veri madenciliği yöntemlerinden olan birliktelik analizinin hidrolojik ve iklimsel parametrelerin arasındaki ilişkilerin keşfinde kullanılabileceği tespit edilmiştir. Yapılan uygulamada Türkiye'nin büyük bir bölümü için yağış ve sıcaklığın akımı belirleyen temel değişkenler olduğu bulunmuştur.

Anahtar Kelimeler: hidrolojik ve iklimsel veri analizi, mekan-zamansal hidrometeorolojik örüntüler, veri madenciliği, birliktelik analizi

ABSTRACT

The purpose of this project was to determine the relationships between hydrologic and climatic parameters of Turkey using data mining techniques. The previous studies conducted using hydrologic and climatic data of Turkey has three major limitations. 1) the relationships between hydrologic and climatic parameters have not been sufficiently explored. 2) statistical methods, which are not sufficient for analyzing large datasets, were mostly used for analysis. 3) previous studies focused on the temporal aspects of the datasets and discovery of temporal patterns and ignored the spatial and spatio-temporal aspects of the data. In this project, hydrologic and climatic data were analyzed together to determine the relationships and spatio-temporal hydrometeorological patterns formed by them. For this purpose data mining techniques (association analysis) were used. Data mining aims to develop automatic or semi-automatic methods for discovering unforeseen, interesting, and meaningful patterns from heterogeneous and large datasets which cannot be analyzed manually. The dataset that were used in this project were obtained from various government institutions (Turkish State Meteorological Service – DMI, General Directorate of Electrical Power Resources Survey and Development Administration – EIE, etc.). After testing data quality and preparation of data for data mining, hydrologic and climatic datasets were analyzed by considering their spatial and temporal characteristics and a new algorithm for discovering association rules was proposed. The techniques developed in this Project was further used to discover the relationships between stream flow and climatic variables in Turkey. The analysis showed that data mining techniques, including association analysis, can successfully be used for discovering the relationships between hydrologic and climatic variables. The analysis also showed that stream flow in Turkey is mostly related with precipitation and air temperature.

Keywords: Hydrologic and climatic and data analysis, spatio-temporal hydrometeorological patterns, data mining, association analysis

1. GİRİŞ / AMAÇ VE KAPSAM

Bu projenin temel amacı Türkiye için hidrolojik ve iklimsel parametreler arasındaki ilişkilerin ve bu parametrelerin birlikte oluşturdıkları örüntülerin tespit edilmesi ve bu ilişki ve örüntülerin tespit edilmesi veri madenciliği yöntemlerinin geliştirilmesidir.

Küresel ısınmadan kaynaklanan iklim değişikliği tarım, hayvancılık ve enerji gibi ulusal ekonomi ve toplumsal refahı yakından etkileyen birçok sektörü tehdit etmektedir. Hükümetlerarası İklim Değişikliği Paneli (Intergovernmental Panel on Climate Change veya IPCC) tarafından hazırlanan rapora göre 1906 - 2005 döneminde küresel yüzey sıcaklığı 0.74 °C artmıştır [1, 2]. Küresel ısınmanın hidrolojik sistemde değişikliklere yol açtığı artık bilinen bir gerçektir. Örneğin, atmosferik nem oranı artmakta, yağış karakterinde ve şiddetinde değişimler gözlenmekte ve sel ve kuraklık gibi doğal afetlerde artışlar yaşanmaktadır. IPCC tarafından yayınlanan, “İklim Değişikliği ve Su (Climate Change and Water)” başlıklı raporda su kaynaklarının iklim değişikliğinden etkilenebilirliğinin belirlenmesinin önemi vurgulanmıştır [3]. Yine aynı raporda iklim değişikliğine uyum için karar verme mekanizmalarını güçlendirmek üzere hidrolojik ve iklimsel sistemler arasındaki ilişkilerin özellikle yerel ölçekte (mekana bağlı olarak) belirlenmesinin öneminden bahsedilmiş ve bu amaçla bu ilişkilerin belirlenmesi için yeni yöntemler geliştirilmesinin gerekliliğine işaret edilmiştir. Bu proje kapsamında hidrolojik parametreler (akarsu akımları) ve iklimsel parametreler (sıcaklık, yağış, nisbi nem, rüzgar şiddeti) arasındaki ilişkilerin ve bu parametrelerin birlikte oluşturdıkları örüntülerin belirlenmesi için yeni ve özgün yöntemler geliştirilmiştir. Geliştirilen yöntemler Türkiye’ye ait hidrolojik ve iklimsel veriler üzerinde uygulanmış ve Türkiye için hidrolojik parametreler ve iklimsel parametreler arasındaki birliktelik örüntüleri tespit edilmiştir.

Türkiye’de hidrolojik ve iklimsel veriler daha önce birçok çalışmada analiz edilmiştir. Bu çalışmalar genel olarak 1) verilerin genel karakterleri ve mekansal olarak değişimlerini ortaya koymaya, 2) veri kalitesini ve güvenilirliğini belirlemeye, 3) verilerde geçmişte meydana gelen değişimleri belirlemeye, 4) verilerde Güneyli Salınım veya Kuzey Atlantik Salınımı gibi büyük ölçekli atmosferik salınım olaylarının etkilerini açıklamaya, 5) verileri gruplandırmaya yani benzer karakter gösteren kümeler oluşturmaya yönelmişlerdir. Bu çalışmaların çoğunluğunda hidrolojik ve iklimsel sistemler birbirlerinden ayrı olarak incelenmiştir. Bu nedenle hidrolojik ve iklimsel parametreler arasındaki ilişkiler henüz yeterli düzeyde

belirlenenememiştir. Buna karşılık hidrolojik olaylar ve iklimsel faktörler arasındaki bağlantılar her zaman açık değildir. Bu projede Türkiye için hidrolojik ve iklimsel parametreler arasındaki bağlantılar belirlenerek bu parametrelerin birlikte oluşturdukları örüntüler mekansal ve mekan-zamansal olarak analiz edilmiştir.

Hidrolojik ve iklimsel parametrelerin ilişkilerinin analizinde daha önce korelasyon analizi gibi istatistiksel yöntemler kullanılmıştır. İstatistiksel yöntemler destekleyici (confirmatory) ve açıklayıcı (explanatory) veri analiz yöntemlerini içerir. Destekleyici yaklaşımlar analize başlamak için bir hipoteze ihtiyaç duyarlar. Örneğin korelasyon analizinde iki parametre arasındaki korelasyonun sıfırdan farklı olduğu hipotezi ile başlamak gereklidir. Yani analiz önceden belirlenen hipotezi desteklemek veya reddetmek için yapılmaktadır. Açıklayıcı yaklaşımlar ise verinin analiz edilerek veri karakteristiğine uygun hipotezlerin geliştirilmesi esasına dayanır. Ayrıca istatistiksel yöntemler küçük-ölçekli verilerin analizi için geliştirilmişlerdir ve dolayısı ile büyük-ölçekli verilerin analizinde yetersiz kalırlar. Büyük-ölçekli verilerin analizi için veri madenciliği yöntemleri kullanılabilir. Bu projede kullanılan yöntem olan veri madenciliği yöntemleri açıklayıcıdır ve veri analizi sürecinde hipotezler geliştirme esasına dayanır. Böylece açıkça tespit edilemeyen ilginç ve anlamlı ilişki ve örüntülerin modellenmesine ve dolayısıyla keşfedilmesine imkan verir. Ayrıca veri madenciliği yöntemleri büyük-ölçekli verileri analiz edebilme potansiyeline sahiptir. Bu projede veri madenciliği yaklaşımı kullanılarak hidrolojik ve iklimsel parametrelerin birlikte oluşturdukları örüntüler ortaya çıkarılmıştır.

Veri madenciliği temel olarak dört yöntem içerir [4, 5]. Bunlar kümeleme, sınıflandırma, anormallik tespiti ve birliktelik analizidir. Kümeleme danışmansız sınıflandırma olarak da tanımlanabilir ve sınıfları belli olmayan verileri ortak özelliklerine göre gruplamayı amaçlar. Sınıflandırma genel olarak danışmanlı gruplama olarak bilinir ve kümelemenin aksine sınıflar önceden verilir. Anormallik tespiti veri kümelerinde beklenen özellikleri göstermeyen örüntülerin tespit edilmesini amaçlar. Birliktelik analizi ise veri tabanlarında yer alan parametreler arasındaki açıkça görülemeyen ilginç ve anlamlı ilişki ve örüntülerin keşfedilmesini sağlayan bir yöntemdir. Bu çalışma kapsamında birliktelik analizi kullanılmıştır. Genel olarak tanımlanacak olursa birliktelik analizi “bir işlem kaydında (transaction) bir öğenin meydana gelme olasılığını diğer öğelerin meydana gelme olasılıklarından tahmin etmek için kurallar bulunmasıdır” [4]. Birliktelik analizi kullanıcı tarafından belirlenmiş frekans (sıklık) değerinden daha yüksek frekansa (daha sık) sahip olan

$\{A\} \rightarrow \{B\}$ şeklindeki kuralların bulunmasını sağlar. Bu kuralda A ve B olayları veya iki farklı parametreyi göstermektedir. Örneğin, A yağış, B ise pik akım biçimi olabilir. $\{A\} \rightarrow \{B\}$ kuralı ise “eğer A gerçekleşirse, B gerçekleşir” şeklinde okunur. Bu yönüyle birliktelik analizi korelasyon analizinde olduğu gibi sadece parametreler arasındaki ilişkiyi açıklamamakta aynı zamanda neden-sonuç ilişkilerini gösteren kurallar oluşturmaktadır. Birliktelik analizi aynı zamanda korelasyon analizinde olduğu gibi iki parametre ile sınırlı kalmayıp, birçok parametreyi aynı zamanda analiz etme kabiliyetine sahiptir. Yani birliktelik analiziyle $\{A, B, C\} \rightarrow \{D\}$ şeklinde veya $\{A, B, C\} \rightarrow \{D, E, F\}$ şeklinde kurallar bulmak mümkündür.

Bugüne kadar yapılan çalışmalar verilerin zamansal boyutuna odaklanarak (zaman serileri analizi gibi) zamansal örüntüleri tespit etmişler, ancak verilerin mekansal ve mekansal-zamansal boyutlarını ihmal etmişlerdir. Bu durum Tobler’in coğrafyanın ilk kuralı olarak bilinen “Herşey herşeyle ilişkilidir fakat uzayda birbirine yakın olanlar uzak olanlara göre daha çok ilişkilidir” kuralıyla çelişmektedir [6]. Aynı kuralı mekan-zamansal özellikler içinde genelleayebiliriz. Hidrolojik ve iklimsel veriler coğrafi (yani mekansal) özellik taşımaktadır ve dolayısıyla veriler mekansal ve mekan-zamansal ilişkiler içermektedirler. Bu projede hidrolojik ve iklimsel parametreler arasındaki mekan-zamansal ilişkiler belirlenmiştir.

Bu projede Türkiye’ye ait akarsu akımları örneğinde hidrolojik parametrelere ait veriler üzerindeki iklimsel etki incelenecektir. Akarsu akımları, yağış, sıcaklık, buharlaşma (ve buharlaşmayı etkileyen nisbi nem, rüzgar hızı) gibi pekçok iklimsel parametreden etkilenmektedir. Bu nedenle bu projede hidrolojik parametreler ile, yağış, sıcaklık, nisbi nem, rüzgar hızı gibi iklimsel parametreler arasındaki ilişkilerin ve bu parametrelerin beraber oluşturdukları birliktelik örüntüleri analiz edilmiştir.

2. GENEL BİLGİLER

Proje konusu ile ilgili kapsamlı bir literatür araştırması yapılmıştır. Literatür araştırması aşağıda üç bölüm halinde sunulmaktadır. Birinci bölümde Türkiye hidrolojik ve iklimsel verileri kullanılarak daha önce yapılmış olan çalışmalar tartışılmaktadır. İkinci bölümde hidrolojik ve iklimsel verilerin ilişkilerinin belirlenmesi için geçmişte kullanılan yöntemler özetlenmekte ve bu projede önerilen veri madenciliği yaklaşımı ile karşılaştırılmaktadır. Üçüncü ve son bölümde ise Türkiye’de yapılan çalışmaların kapsamı ve daha önceki çalışmalarda kullanılan yöntemlerin bir analizi sunulmaktadır.

2.1. Türkiye Hidrolojik ve İklim Verileri Kullanılarak Daha Önce Yapılan Çalışmalar

Türkiye’ye ait hidrolojik ve iklimsel veri serileri 1990’lı yıllardan itibaren birçok çalışmada analiz edilmiştir. Bugüne kadar yapılan çalışmalar 1) verilerin genel karakterlerinin ve mekansal olarak değişimlerinin belirlenmesi, 2) homojenlik analizleri, 3) eğilim analizleri, 4) veri serilerinde büyük ölçekli atmosferik salınım etkilerinin araştırılması, 5) kümeleme analizleri olarak beş grup altında sınıflandırılabilir.

Aşağıda Türkiye verisi üzerinde yapılan çalışmalar yukarıda belirtilen başlıklar altında ayrıntı olarak listelenmekte ve çalışma konuları ve analiz edilen verilerin özellikleri ve kullanılan yöntemler açıklanmaktadır.

2.1.1. Verilerin Karakterlerinin ve Mekansal Değişimlerinin Belirlenmesi

Bu grupta yer alan çalışmalar genel olarak sıcaklık (ortalama, minimum, maksimum) ve yağış gibi iklimsel parametrelerin genel karakterlerini (örneğin aylık, mevsimsel ve yıllık dalgalanmalar, ısrar ve dönemsellik) ve mekansal olarak dağılımlarını incelemişlerdir. Bu çalışmalar genel itibarıyla çeşitli parametreler için zaman serisi analizlerini kapsamaktadır ve Türkiye için genel durumu ortaya koymaktadır.

Türkeş v.d. [7] 85 istasyona ait 63 yıllık (1930-1992) yıllık ortalama sıcaklık verilerinin ve yıllık değişim katsayılarının Türkiye üzerindeki dağılımı incelemişlerdir.

Türkeş [8] 91 istasyona ait ortalama 60 yıllık (1930-1993) aylık toplam yağış verilerini kullanarak yıllık yağış verilerinin ve yıllık değişim katsayılarının Türkiye üzerindeki dağılımını incelemiştir.

Türkeş [9] 99 istasyona ait 54-65 yıllık (1929-1993) dönemi kapsayan aylık yağış verilerini kullanarak yıllık yağış ve mevsimsel yağış verilerinin Türkiye üzerindeki dağılımını, yıldan yıla ve mevsimsel olarak meydana gelen değişimleri incelemiştir. Ayrıca yağışlarda meydana gelen eğilimlerin, uzun dönem dalgalanmaların, yıldan yıla dalgalanmaların türünü açıklamış ve bunu jeopotansiyel yükseklik serisi ile ilişkilendirmiştir.

Kutiel v.d. [10] Türkiye'nin yağış bakımından farklılık gösteren 7 bölgesinden seçilmiş 7 istasyona ait aylık toplam yağış verilerini kullanarak yağışta gözlenen değişimlerin deniz seviyesi basıncı ile ilişkisini incelemişlerdir.

Türkeş v.d. [11] 91 istasyona ait aylık toplam yağış verilerini kullanarak yıllık ve mevsimsel yağış verilerindeki ısrar ve dönemselliği incelemişlerdir. Çalışmada serisel korelasyon katsayıları ve güç spektrası analizleri kullanılmıştır. Ayrıca yağış anormallikleri ile jeopotansiyel yükseklik serisindeki anormalliklerin ilişkisi incelenmiştir.

2.1.2. Veri Serilerin Homojenliğinin Değerlendirilmesi

Bu grupta yer alan çalışmalar hidrolojik ve iklimsel veri serilerinin homojenliğini sınamayı amaçlamıştır. Homojenlik analizi veri toplama yönteminde ve verinin toplandığı konumda yapılan değişikliklerden kaynaklanabilecek farklılaşmaları bulmayı amaçlamaktadır. Bu nedenle hidrolojik ve iklimsel verilerin analizlerini içeren birçok çalışmada bir ön analiz olarak yer almıştır.

Türkeş v.d. [7] 1930-1992 dönemini kapsayan yıllık ortalama sıcaklık verilerinin homojenliğini short-cut Barlett testini kullanarak analiz etmişlerdir.

Türkeş [8] 1930-1993 dönemini kapsayan aylık toplam yağış verilerinin homojenliğini Kruskal-Wallis testini kullanarak analiz etmiştir.

Türkeş v.d. [12] 1930-1993 dönemini kapsayan mevsimsel ortalama günlük maksimum ve minimum sıcaklık verilerindeki verilerinin homojenliğini Kruskal-Wallis testini kullanarak analiz etmişlerdir.

Tayanç v.d. [13] 81 istasyon için 1951-1990 dönemini kapsayan sıcaklık verilerinin homojenliğini grafiksel inceleme, Kruskall-Wallis testi ve Wald-Wolfowitz testini kullanarak analiz etmişlerdir.

Türkeş v.d. [11] 1929-1999 dönemini kapsayan aylık ve mevsimsel ortalama, maksimum ve minimum sıcaklık verilerinin homojenliğini Kruskall-Wallis testini kullanarak analiz etmişlerdir.

Kahya ve Karabörk [14] 300'den fazla akarsu gözlem istasyonuna ait akım verisini inceleyerek regülasyon ve su çevriminin mevcut olmadığı 83 istasyonu belirlemişlerdir. Bu istasyonlara ait veri zaman serilerinin homojenliği aynı zamanda insan-kaynaklı etkinin (örneğin, istasyon yeri değişimi, baraj veya savak inşaatı) olup olmadığını kontrol edilerek doğrulanmıştır.

Cıgızoğlu v.d. [15], bir önceki çalışmaya benzer şekilde regülasyon ve su çevriminin mevcut olmadığı 96 akım gözlem istasyonu belirlemişlerdir.

Karabörk v.d. [16] 1973-2002 dönemini kapsayan yıllık toplam yağış verilerinin homojenliğini Standart Normal Homojenlik Analizi ve Pettitt Testini kullanarak analiz etmişlerdir.

Göktürk v.d. [2] 267 istasyondan elde edilmiş aylık yağış serisinde yer alan aykırı verileri belirlemiş ve veri serilerin homojenliğini Standart Normal Homojenlik analizi ile kontrol etmişlerdir.

Türkeş v.d. [17] 1930-2002 dönemini kapsayan aylık, mevsimlik ve yıllık yağış verilerinin homojenliğini Kruskall-Wallis testini kullanarak analiz etmişlerdir.

Tayanç v.d. [18] 50 istasyona ait 54 yıllık (1950-2004) günlük maksimum, minimum ve ortalama sıcaklık ve günlük yağış verilerinin homojenliğini en yakın istasyonla farklılık serileri oluşturarak incelemişlerdir.

2.1.3. Eğilim Analizleri

Bu grupta yer alan çalışmalar yağış ve sıcaklık gibi iklimsel parametreler ve akarsu akımı ve göl seviyeleri gibi hidrolojik parametrelerde yer alan eğilimlerin belirlenmesine

odaklanmışlardır. Eğilim analizlerindeki amaç verilerde zamansal olarak meydana gelen değişimlerin belirlenmesidir.

Türkeş v.d. [7] 85 istasyona ait 63 yıllık (1930-1992) yıllık ortalama sıcaklık verilerinde gözlenen değişimleri ve eğilimleri Wald-Wolfowitz serisel korelasyon testi, Mann-Kendall Sıra Korelasyonu, Spearman Sıra Korelasyonu ve Cramer testini kullanılarak değerlendirmişlerdir.

Türkeş [8] 91 istasyona ait ortalama 60 yıllık (1930-1993) aylık toplam yağışta gözlenen eğilimleri Mann-Kendall eğilim testini kullanarak incelemiştir.

Türkeş v.d. [12] 59 istasyona ait ortalama 57 yıllık (1930-1993) yıllık günlük maksimum ve minimum sıcaklık verilerindeki eğilimleri Gauss filtresi, Mann-Kendall eğilim testi ve Wald-Wolfowitz serisel korelasyon testini kullanarak analiz etmişlerdir.

Kadıoğlu [19] 18 istasyon için 60 yıllık sıcaklık verisinde yer alan eğilimleri incelemiştir.

Tayanç ve Toros [20] 12 istasyona ait sıcaklık ve yağış verilerini analiz ederek kentleşmenin iklim değişikliği üzerindeki etkisini incelemişlerdir.

Türkeş v.d. [21] 70 istasyona ait 70 yıllık (1929-1999) aylık ortalama, maksimum ve minimum sıcaklık verilerinde gözlenen eğilimleri Mann-Kendall eğilim testi ve Wald-Wolfowitz serisel korelasyon testlerini kullanarak analiz etmişlerdir. Ayrıca bu parametrelerdeki lineer eğilimler de lineer regresyon analizi ile değerlendirilmiştir.

Kahya ve Kalaycı [22] Türkiye'nin 26 havzasına yer alan 83 akarsu gözlem istasyonundan elde edilen aylık akım verilerini kullanarak 1964-1994 dönemi için eğilimleri tespit etmişlerdir. Çalışmada dört adet parametrik olmayan test kullanılmıştır (Sen's T, Spearman's Rho, Mann-Kendall Testi ve Mevsimsel Kendall Testi). Ayrıca Van Belle ve Hughes havza bazında eğilim testleri kullanılmıştır.

Türkeş ve Sümer [23] 70 istasyona ait 70 yıllık (1929-1999) aylık maksimum ve minimum sıcaklık verilerini kullanarak günlük sıcaklık aralığında meydana gelen eğilimleri incelemişlerdir. Çalışmada Mann-Kendall sıra korelasyon testi kullanılmıştır. Ayrıca günlük sıcaklık aralığı ve maksimum ve minimum sıcaklık serileri arasındaki ilişkiler parametrik ve parametrik olmayan korelasyon testleri (Pearson ve Spearman) ve t-testi kullanılarak değerlendirilmiştir.

Yıldırım v.d. [24] Gediz Havzası'nda yer alan 4 akarsu gözlem istasyonuna ait akım verisi ve Akhisar İstasyonu'nda toplanmış yağış verilerindeki eğilimleri ve bu verilerin birbirleriyle ilişkilerini analiz etmişlerdir. Eğilim analizlerinde Mann-Kendall sıra korelasyon testi, ilişkilerin belirlenmesinde ise korelasyon katsayıları kullanılmıştır.

Cıgızoğlu v.d. [15] 25 ila 66 yıllık dönemi kapsayan 96 akım gözlem istasyonuna ait maksimum, ortalama ve düşük akımlardaki eğilimleri parametrik t-testi ve parametrik olmayan Mann-Kendall testini kullanarak analiz etmişlerdir.

Partal ve Kahya [25] 96 istasyona ait 1929-1993 dönemini kapsayan aylık toplam ve aylık toplam verilerin yıllık ortalaması parametrelerini ve aynı zamanda bölgesel indis serilerini Mann-Kendall eğilim testi ve Şen's T testini kullanarak analiz etmişlerdir.

Kahya ve Partal [26] 96 istasyona ait 1929-1003 dönemini kapsayan aylık toplam yağış verilerinden elde edilmiş mevsimsel ortalama yağış verilerindeki eğilimleri Mann-Kendall eğilim testi ile belirlemişlerdir.

Türkeş v.d. [17] 97 istasyona ait ve 72 yıllık (1930-2002) dönemi kapsayan aylık, mevsimlik ve yıllık yağış verilerinde meydana gelen uzun dönemli değişimleri ve eğilimleri incelemişlerdir. Eğilimleri tespit etmek için Mann-Kendall sıra korelasyon testi kullanılmıştır. Aynı zamanda 86 istasyona ait 49 yıllık (1953-2002) verileri kullanarak ana bileşenler analizi ile mekansal değişimler incelenmiştir.

Tayanç v.d. [18] 50 istasyona ait 54 yıllık (1950-2004) günlük maksimum, minimum ve ortalama sıcaklık ve günlük yağış verilerini kullanarak bu parametrelerdeki eğilimleri Mann-Kendall testi ile değerlendirmişlerdir.

2.1.4. Veri Serilerinde Büyük Ölçekli Atmosferik Salınım Etkilerinin Araştırılması

Bu grupta yer alan çalışmalar Güneyli Salınım (Southern Oscillation), Kuzey Atlantik Salınımı (North Atlantic Oscillation) gibi büyük ölçekli atmosferik salınım olaylarının Türkiye yağış, akım ve göl seviyelerindeki etkilerinin belirlenmesini incelemişlerdir.

Türkeş [9] 99 istasyona ait büyük çoğunluğu 54-65 yıllık (1929-1993) dönemi kapsayan aylık yağış verilerini kullanarak yıllık yağış ve mevsimsel yağış verilerinde gözlenen anormalliklerin extrem Güneyli salınım olayları ile ilişkisini t-testini kullanarak incelemiştir.

Kahya ve Karabörk [14] Türkiye'nin 22 havzasına dağılmış 76 istasyona ait aylık akım verilerinde (1964-1994) El Nino ve La Nina sinyallerinin bulunup bulunmadığını incelemişlerdir.

Kutiel v.d. [27] yüksek seviye atmosferik bağlantısı olan Kuzey Denizi-Karadeniz Örüntüsü'nün Yunanistan, Türkiye ve İsrail'e ait sıcaklık ve yağış verilerindeki etkilerini incelemişlerdir. Çalışmada Türkiye'den 7 istasyona ait aylık toplam yağış ve ortalama sıcaklık verileri kullanılmıştır.

Karabörk ve Kahya [28] Türkiye yağış verileri ve El Nino ve La Nina olayları arasındaki ilişkileri incelemişlerdir. Çalışmada 1940-1993 dönemine ait 94 meteoroloji istasyonundan aylık yağış verileri kullanılmıştır. Ayrıca yıllık döngü analizi, çapraz-korelasyon analizi ve spektra analizi ile El Nino ve La Nina ile yağışın ilişkisi sayısal olarak ifade edilmişlerdir.

Türkeş ve Erlat [29] 78 istasyona ait aylık toplam yağış verisini kullanarak Kuzey Atlantik Salınımı (NAO) indisleri ile Türkiye yağış verilerinde görülen değişimleri ilişkilendirmişlerdir. Analizlerde Pearson korelasyon katsayısı ve t-testi kullanılmıştır.

Tatlı v.d. [30] büyük ölçekli iklim süreçleri ile yerel iklim parametreleri arasındaki ilişkileri çok değişkenli istatistiksel teknikleri kullanarak incelemişlerdir.

Kutiel ve Türkeş [31] Kuzey Denizi-Karadeniz Örüntüsü'nün Kapadokya bölgesindeki yağış ve sıcaklık rejimlerine etkisini incelemişlerdir.

Karabörk v.d. [16] yağış, akarsu akımı, maksimum ve minimum sıcaklığı içeren Türkiye iklim değişkenleri ile Güneyli Salınım (SO) ve Kuzey Atlantik Salınımı (NAO) bağlantılarını incelemişlerdir.

Türkeş ve Erlat [32] 78 istasyon için kış aylarında düşen yağış verilerini ile Kuzey Atlantik Salınımı (NAO) indisleri arasındaki bağlantıları ilişkilendirmişlerdir. Analizlerde korelasyon ve kompozit analizleri kullanılmıştır.

Karabörk v.d. [33] 1973-2002 dönemini kapsayan aylık toplam yağış verilerini kullanarak yağış anormalliklerinde güneyli salınım etkisini gamma dağılımını kullanarak sayısal olarak belirlemişlerdir.

Kalaycı ve Kahya [22] El Nino olaylarından etkilenmiş ve etkilenmemiş akım serilerinin istatistiksel özelliklerinin farklı olup olmadığını incelemişlerdir.

Martı ve Kahya [34] ortalama yıllık sıcaklık verisinde El Nino Güneyli Salınım etkisinin bulunup bulunmadığını incelemişlerdir.

Kahya ve Cengiz [35] Beyşehir Gölü su seviyesi üzerinde Kuzey Atlantik Salınımının etkisini sürekli wavelet transform ve global spectra yöntemlerini kullanarak incelemişlerdir.

Türkeş ve Erhat [36] 70 istasyona ait 1950-2003 dönemine ait ortalama kış sıcaklıkları ve aylık ortalama sıcaklıkları ile Kutup Salınım İndisi (Arctic Oscillation Index) arasındaki ilişkileri incelemişlerdir. Analizlerde korelasyon ve kompozit analizleri kullanılmıştır.

Karabörk ve Kahya [37] katagorize güneyli salınım indisi ve çok değişkenli El Nino Güneyli Salınım indisi ile yağış, sıcaklık ve akarsu akımı parametreleri arasındaki ilişkiler incelenmiştir. Çalışmada parametrik olmayan bir korelasyon testi olan Spearman korelasyon testi kullanılmıştır.

Küçük v.d. [38] 7 göl için su seviyesi verilerini kullanarak Kuzey Atlantik Salınımının göl seviyeleri üzerindeki etkisini sürekli wavelet transform ve global spectra yöntemlerini kullanarak incelemişlerdir.

2.1.5. Kümeleme Analizleri

Bu grupta yer alan çalışmalar kümeleme algoritmalarını kullanarak benzer karakter gösteren alanları gruplamayı amaçlamıştır. Analizlerde yağış, sıcaklık ve akım verileri kullanılmıştır.

Ünal v.d. [39] 113 istasyona ait 47 yıllık (1951-1998) maksimum, ortalama ve minimum sıcaklık ve toplam yağış verilerini kullanarak Türkiye'nin iklim bölgelerini yeniden tanımlamışlardır. Analizlerde tek ağılamalı, tam bağlamalı, merkez (centroid) ve Ward's minimum değişim yöntemi ve ortalama uzaklık yöntemi kullanılmış ve sonuçta Ward algoritmasının en iyi sonucu verdiği bulunmuştur.

Kalaycı ve Kahya [40] ana bileşenler analizini kullanarak 78 istasyondan elde edilmiş akım verilerinin zamansal ve mekansal değişimlerini incelemişlerdir.

Kahya v.d. [41] 80 istasyon için 30 yıllık (1964-1994) yıllık ortalama akım verilerini K-ortalama algoritmasını kullanarak gruplandırmışlardır.

Demirel v.d. [42] 80 istasyon için 30 yıllık (1964-1994) yıllık minimum akım verisini K-ortalama ve ana bileşenler analizi ile incelemişlerdir.

Kahya ve Demirel [43] 80 istasyon için 30 yıllık (1964-1994) aylık minimum akım verilerini üç yığılmalı hiyerarşik kümeleme algoritmalarını kullanarak (tek bağlamalı, tam bağlamalı ve Ward's algoritmaları) gruplandırmışlardır.

Kalaycı ve Kahya [44] 78 istasyon için 30 yıllık (1964-1994) akım verisi kullanılarak elde edilen dönüştürülmüş ve dönüştürülmemiş ana bileşenlerini karşılaştırmışlardır.

Kahya v.d. [45] 78 istasyon için 30 yıllık (1964-1994) akım verisini homojen akım bölgeleri elde etmek için varimaks ortogonal dönüşümlü ana bileşenler analizini kullanarak gruplandırmıştır.

2.2. Hidrolojik ve İklimsel Veri Analizinde Veri Madenciliği Yaklaşımı

Dinamik bir yapıya sahip olan hidrolojik ve iklimsel sistemlerin birbirleriyle ilişkilerinin belirlenmesi ancak hidrolojik ve iklimsel veri kümelerinin etkin biçimde analizi ile mümkündür. Özellikle son yıllarda etkisini daha çok hissettiren iklim değişikliğine karşı hazırlıklı olabilmek için geçmişte kaydedilmiş verilerde gizlenmiş ve her geçen gün farklılaşan anlamlı ve önemli mekan-zamansal örüntülerin tespit edilmesi ve böylelikle hidrolojik sistemler üzerindeki iklimden kaynaklanan etkinin belirlenmesi gerekmektedir.

Hidrolojik ve iklimsel sistemler arasındaki ilişkiler her zaman belirgin değildir. İlişkilerin belirlenmesinde ortaya çıkan en temel problem hidrolojik ve iklimsel verilerin kendilerine has yapılarıdır [46]. Öncelikle hidrolojik ve iklimsel veriler coğrafi (mekansal) özellik taşımaktadır. Bu nedenle zamansal ve mekansal korelasyonlar (yani bağlantılar) içerirler. İkinci olarak hidrolojik ve iklimsel veriler birbirleriyle lineer olmayan biçimde ilişkilidir [47]. Bu durum ilişkilerinin lineer ilişkiler için uygun yöntemlerle (örneğin, lineer regresyon,

korelasyon) tanımlanmasını imkansız kılar. Üçüncü olarak hidrolojik ve iklimsel verilerde geçmişte meydana gelmiş olayların etkilerini görmek mümkündür yani uzun dönemli hafızaya sahiptir. Yağışın yeraltı suyu seviyesini gecikmeli olarak yükseltmesi buna bir örnektir. Son olarak hidrolojik ve iklimsel veriler mekansal boyutta bağımlılık gösterirler. Büyük-ölçekli atmosferik salınımların etkilerinin dünyanın çeşitli bölgelerinde hissedilmesi bu bağımlılığın göstergesidir.

Hidrolojik ve iklimsel parametreler arasındaki bağlantıları hidrolojik modelleme çalışmaları ile tanımlamak mümkündür [48]. Ancak hidrolojik ve iklimsel sistemlerin ve bunların birbirleriyle ilişkilerinin karmaşıklığı başarılı modeller geliştirmeyi zorlaştırır. Modelleme aynı zamanda uygulanacak alanın hidrojeolojik yapısı, toprak özellikleri, arazi örtüsü gibi elde edilmesi kolay olmayan verilere ihtiyaç duyar. Birden fazla çalışma alanı olduğu durumlarda veya bölgesel veya ülke çapında bir analiz yapılmak istendiğinde modelleme çalışmalarının yapılabilirliği ve başarısı daha da düşmektedir.

Hidrolojik ve iklimsel sistemlerin ilişkilerini belirlemek için istatistiksel yöntemler yoğun olarak kullanılmaktadır. Korelasyon analizleri bu alanda en sık olarak kullanılmış olan yöntemdir. Örneğin Groisman v.d. [49] Amerika üzerindeki akarsu akımı ve yağış ilişkisini korelasyon analizi ile incelemiştir. Burn ve Elnur [50] Kanada için akarsu akımı ve yağış, sıcaklık gibi iklimsel parametreler arasındaki ilişkileri bu yöntemle incelemiştir. Novotny ve Stefan [51] Amerika'nın Minnesota eyaleti için akarsu akımı ve yağış ilişkisini korelasyon katsayılarını hesaplayarak incelemiştir. Dadaser-Celik ve Stefan [52] Minnesota'da yer alan 25 göl için su seviyesi ve yağış, sıcaklık, çiy noktası sıcaklığı ve rüzgar hızı gibi iklimsel parametreler arasındaki ilişkileri korelasyon analizi ile incelemiştir. Dadaser-Celik ve Stefan [53] yine aynı 25 göl için su seviyesi ve buharlaşma ve yağış-buharlaşma farkının korelasyonunu incelemiştir. Geçmişte yapılan çalışmalar çoğunlukla yağış ve sıcaklık parametrelerini içermiştir. Bu iki iklimsel parametre dışındaki parametreleri içeren çalışmalar çok azdır. Aynı zamanda korelasyon analizi ancak iki parametrenin ilişkisinin araştırılmasını sağlamaktadır. İki'den fazla parametrenin dahil olduğu analizlerde bu yöntem yetersizdir. Ayrıca korelasyon analizleri ilişkilerin mevcudiyetini ve derecesini açıklamasına rağmen parametreler arasındaki neden-sonuç ilişkisini açıklayamamaktadır.

Bu çalışmada hidrolojik ve iklimsel verilerin lineer olmayan ilişkilerini açıklamak için verilerin yapısal özelliklerine uygun ve mekansal olarak hızlı ve etkin bir analiz geliştirilmesine imkan verecek veri madenciliği yaklaşımı kullanılacaktır.

Veri madenciliği büyük ölçekli ve çok değişkenli veri tabanlarından bilgi çıkarımı olarak tanımlanabilir [4, 5]. Veri madenciliği veri kümelerinden maksimum bilgiyi elde etmeyi, birbirleriyle ilişkili değişkenleri belirlemeyi ve geleceğe yönelik daha doğru tahminler yapılmasına olancak sağlayacak modeller geliştirmeyi hedefler [54, 55]. Bu hem değişkenler arasındaki sebep-sonuç ilişkilerinin belirlenmesini hem de alternatif stratejilerin sonuçlarının değerlendirilmesini kapsar. Veri madenciliği yöntemleri daha önce pazarlama, biyoenformatik ve madencilik gibi sistematik veri analizini gerektiren ve büyük veritabanlarına sahip olan pekçok alanda uygulanmasına rağmen [56], hidrolojik ve iklimsel veri analizlerindeki uygulamaları sınırlıdır.

Literatürde meteorolojik ve/veya hidrolojik veri ilişkilerinin nitelenmesi için çeşitli örüntü tanımlamaları yapılmıştır. Bu örüntüler genel veri madenciliği uygulamaları ile de uyumlu olarak 4 ana başlık altında toplanabilir: (1) Klasik birliktelik örüntüleri, (2) zamansal birliktelik örüntüleri, (3) mekansal birliktelik örüntüleri, (4) mekan-zamansal birliktelik örüntüleri. Aşağıda her bir örüntü türü kısaca tanımlanmakta ve bu örüntülerle ilgili olarak daha önce yapılan çalışmalar özetlenmektedir.

Klasik birliktelik örüntüleri verilerin zaman ve mekan boyutlarını ihmal eden ilişkileri tanımlamaktadır. Tan v.d. [57] çevresel ve iklimsel veritabanlarında klasik birliktelik örüntülerini incelemiştir. Bu çalışmada klasik birliktelik örüntüleri “bölge-içi sıralı olmayan birliktelik örüntüleri (intra-zone non-sequential association patterns)” olarak isimlendirilmiştir. Bu çalışmada kullanılan parametreler yağış, deniz yüzeyi sıcaklığı (sea surface temperature – SST) gibi gözlem verilerini, net birincil üretim (Net Primary Production – NPP) gibi model sonuçlarını ve SOI, NAO, deniz suyu sıcaklık anomalileri gibi iklim indislerini kapsamaktadır. Bu çalışmada veriler öncelikle önışlemden geçirilerek veri madenciliği uygulamaları için hazırlanmıştır. Önışlem aşamasında mevsimsel etki giderilmiş ve veriler ayrıklaştırılmıştır. Mevsimsellik giderimi için çeşitli yaklaşımların (Discrete Fourier Transform, Aylık Z-Score gibi) kullanılabilirliği değerlendirilmiştir. Çalışmada bölge-içi sıralı olmayan örüntülerin tespit edilmesinde Apriori [58] ve FP-tree [59] klasik veri madenciliği algoritmalarının kullanılabileceği belirtilmiştir. Ganguly ve Steinhäuser [46] veri madenciliği yaklaşımının iklim değişikliğini ve etkilerini incelemekte kullanılabileceğini göstermek için İspanya verileri üzerinde basit bir veri madenciliği uygulaması yapmışlardır. Bu uygulamada İspanya için extrem sıcaklıklar konu alınmıştır. Çalışmada Hükümetlerarası İklim Değişikliği Paneli (IPCC) tarafından geliştirilmiş iklim projeksiyonlarını kullanan iklim modelleri ve

gözlenen sıcaklık değerleri kullanılmıştır. Çalışma üç aşamada gerçekleştirilmiştir. Birinci aşamada 2000-2007 dönemi için model tahmin sonuçları gözlem verileri ile karşılaştırması ile belirsizlik sınırları çıkarılmıştır. İkinci aşamada 2045-2054 ve 2090-2099 dönemleri için modellerden elde edilen sonuçlar üzerinde tahminlerdeki belirsizliklerin etkisi değerlendirilmiştir. Üçüncü aşamada ise özel bir bölgeye yoğunlaşarak iklim değişikliğinin muhtemel etkileri tartışılmıştır. Shu v.d. [60] iklimsel veriler üzerinde bulanık birliktelik kurallarını (fuzzy association rules) araştırılmıştır. Bulanık birliktelikler incelendiği için sınıflandırmanın daha önemli olduğu üzerinde durulmuştur. C-means algoritması fuzzy olarak yeniden şekillendirilmiştir. Algoritmaya veriler giriş olarak verilmeden önce gürültülerden arındırılmış, eksik veriler giderilmiştir. Daha sonra girdiler bulanıklaştırılmıştır. Bu çalışmanın en önemli özelliği girdi değerlerinin bulanık olması ve böylece daha iyi çözümler elde edilebilmesidir. Algoritma sonucu çıkan kümelenmiş veriler, Apriori algoritması yardımıyla kurallar haline getirilmiştir.

Zamansal birliktelik örüntüleri verilerin zamansal özelliklerini dikkate alan ilişkileri ifade etmektedir. Tan vd. [61] tarafından yukarıda da bahsedilen çalışmada zamansal birliktelik örüntüleri bölge-içi sıralı örüntüler (intra-zone sequential patterns) olarak adlandırılmıştır. Bu çalışmada bu tür örüntülerin tespitinde Genelleştirilmiş Sıralı Örüntüler (Generalized Sequential Patterns – GSP) algoritması [62] kullanılmıştır. Çalışmada kullanılan veriler ve takip edilen yöntem klasik birliktelik örüntüleri başlığı altında verilmiştir. Tadesse vd. [63, 64] Amerika Birleşik Devletleri'nin Nebraska eyaletinde kuraklık indisleri ve büyük ölçekli atmosferik olayları arasındaki ilişkileri zamansal birliktelik örüntüleri şeklinde ifade etmişlerdir. Çalışmalarda kullanılan parametreler, yağış, sıcaklık, toprak nemi, standart yağış indisi (Standard Precipitation Index – SPI) ve Palmer Kuraklık Şiddet İndisi (Palmer Drought Severity Index – PSPİ), Kuzey Atlantik Salınım İndisi (North Atlantic Oscillation Index – NAO), Güneyli Salınım İndisi (South Oscillation Index – SOI), Pasifik Kuzey-Amerika İndisi (Pacific North-American Index – PNA) ve Pasifik On Yıllık Salınımı İndisi (Pacific Decadal Oscillation – PDO), Çok-Değişkenli ENSO İndisi (Multivariate ENSO Index – MEI)'dir. Çalışmada kullanılan veriler zaman serileri şeklinde düzenlenmiştir. Çalışmaların ilk aşamasında her parametre için veriler 7 grupta sınıflandırılmıştır. Ardından MOWCATL – Minimal Occurences With Constraints And Time Lags – algoritması veri seti üzerinde çalıştırılmış ve kurallar elde edilmiştir. Kuralları değerlendirmek için, destek (support) ve güven (confidence) değerleri yerine J-measure adında bir değerlendirme ölçütü kullanılmıştır. Nagesh Kumar v.d. [65] ve Dhanya ve Nagesh Kumar [66] Hindistan için kuraklık ve selleri

tahmin etmek için ekstrem yağışlar ve iklim indisleri arasındaki ilişkileri tespit etmek için zamansal birliktelik analizi yöntemini kullanmışlardır. Çalışmalarda 1960 – 2005 dönemi için Hindistan için yaz (Haziran-Eylül) muson yağmurları verileri, Darwin Deniz Seviyesi Basıncı – DSLP, Nino 3.4, Doğu orta tropik Pasifik Deniz yüzeyi sıcaklıkları, Kuzey Atlantik Salınım İndisi (NAO) ve 1x1 derecelik Deniz Seviyesi Sıcaklığı - SST verileri kullanılmıştır. Çalışmalarda daha önce Tadesse vd. (2005) tarafından da kullanılan MOWCATL algoritması kullanılmıştır. Veri seti 1960-1982 ve 1983-2005 olmak üzere iki parçaya bölünmüştür. İlk parça model için kontrol ölçüsü olarak kullanılmıştır. Makalede geniş bir örüntü çıkarma ve sıralı örüntüler için algoritmik araştırma yapılmıştır. Uygulama kısmında da önce bölgesel bazda daha sonra da tüm Hindistan için algoritma çalıştırılmış ve sonuçlar elde edilmiştir. Sonuç olarak da bazı birliktelik kuraklık veya sellere sebep olabilecek olan birliktelik kuralları bulunmuştur. Bu makalede destek (support) yerine J-measure ölçüğü kullanılmıştır. Lin v.d. [67] global yüzey yağışları ve deniz suyu sıcaklıkları arasındaki bölgeler-arası ilişkileri tespit etmek için birliktelik analizini kullanmışlardır. Çalışmada 100 yıllık global iklim verileri, Güneyli Salınım İndisi (South Oscillation Index – SOI) ve El Nino Bölgesi Deniz Suyu Sıcaklık verileri (Nino3.4) kullanılmıştır. Çalışmada öncelikle bir kümeleme algoritması kullanılarak ekstrem olaylar tespit edilmiş ve daha sonra ise olay birliktelik analizi (episode association mining) [68] mantığını kullanan bir minimum occurrence of episodes (MOE) algoritması ise birliktelik kuralları tespit edilmiştir.

Mekansal birliktelik örüntüleri verilerin mekansal özelliklerini dikkate alan ilişkileri ifade etmektedir. Tan v.d. [61] mekansal birliktelik örüntülerini “bölgeler-arası sıralı olmayan örüntüler (inter-zone non sequential patterns) olarak ifade etmişler ve iklimsel veritabanında keşfi konusunda çalışmışlardır. Çalışmada üzerinde analiz yapılan coğrafya ızgaralara (gridlere) bölünmüştür ve her bir ızgara için kaydedilen veriler kullanılmıştır. Farklı ızgaralardaki verilerin birbirleriyle ilişkilerinin ortaya konulmasıyla anlamlı mekansal birliktelikler ortaya çıkarılmıştır. Çalışmada mekansal birliktelik örüntülerinin tespit edilmesinde de Koperski ve Han [69] veya Shekhar ve Huang [70] tarafından geliştirilen mekansal veri madenciliği algoritmalarının kullanılabileceği belirtilmiştir. Çalışmada kullanılan veriler ve takip edilen yöntem klasik birliktelik örüntüleri başlığı altında verilmiştir.

Mekan-zamansal birliktelik örüntülerin verilerin zamansal ve mekansal özelliklerini dikkate alan ilişkileri ifade etmektedir. Tan v.d. [61] tarafından yapılan çalışmada mekan-zamansal birliktelik örüntüleri bölgeler-arası sıralı örüntüler (intra-zone non-sequential patterns) olarak

ifade edilmiştir. Bu tür örüntülerinin tespit edilmesinde de Koperski ve Han [69] veya Shekhar ve Huang [70] tarafından geliştirilen mekansal veri madenciliği algoritmalarının kullanılabileceği belirtilmiştir. Çalışmada kullanılan veriler ve takip edilen yöntem klasik birliktelik örüntüleri başlığı altında verilmiştir.

Literatürde yer aşan çalışmalar Tablo 2.1 ve Tablo 2.2’de özetlenmektedir.

Tablo 2.1. Literatürde bulunan ve hidrolojik ve iklimsel parametreler arasındaki ilişkileri açıklayan örüntüler

Makale/Örüntü Türü	Klasik	Zamansal	Mekansal	Mekan-Zamansal
Tadesse v.d. [63, 64]		✓		
Lin v.d. [67]		✓		✓
Ganguly ve Steinhäuser [46]	✓			
Tan v.d. [61]	✓	✓	✓	✓
Nagesh Kumar v.d. [65] Dhanya ve Nagesh Kumar [66]		✓		
Shu vd. [60]	✓			

Tablo 2.1 ve 2.2’de gösterildiği üzere hidrolojik ve iklimsel veritabanları üzerinde birliktelik analizinin yapıldığı çalışmalar çok az sayıdadır. Bugüne kadar klasik ve zamansal birliktelik örüntüleri ile ilgili bazı çalışmalar yapılmıştır. Mekansal ve zaman-mekansal birliktelik örüntüleri ile ilgili çalışmalar ise nadirdir. Ayrıca bu çalışmalar küresel ölçekte olup, iklimsel verilerin birbirleriyle ve ekolojik sistemle ilişkilerinin analizleri ile sınırlı kalmıştır. İklimsel ve hidrolojik verileri birlikte inceleyen daha kapsamlı çalışmalar gerekmektedir.

Tablo 2.2. Literatürde bulunan ve hidrolojik ve iklimsel parametreler arasındaki ilişkileri keşfetmek için kullanılan algoritmalar

Makale/Kullanılan Algortima	Apriori	GSP	MOWCOTL	MOE
Tadesse v.d. [63, 64]			✓	
Lin v.d. [67]				✓
Ganguly ve Steinhäuser [46]	✓			
Tan v.d. [61]	✓	✓		
Nagesh Kumar v.d. [65] Dhanya ve Nagesh Kumar [66]			✓	
Shu vd. [60]	✓			

2.3. Literatür Analizi

a) Geçmişte yapılan çalışmalardan birinci grup Türkiye üzerinde sıcaklık ve yağış verilerinin zamansal ve mekansal olarak dağılımını incelemiştir. İkinci grup çalışma yağış, sıcaklık ve akarsu akımı zaman serilerinin homojen olup olmadığının belirlenmesi üzerine yoğunlaşmıştır. Üçüncü grup çalışmada yağış, sıcaklık, akım ve göl seviyelerinde geçmişte meydana gelen değişimler incelenmiştir. Dördüncü grup çalışma Güneyli Salınım ve Kuzey Atlantik Salınımı gibi büyük ölçekli atmosferik salınım etkilerinin yağış, sıcaklık, akım ve göl seviyelerinde gözlenip gözlenmediğinin belirlenmesi üzerine yoğunlaşmıştır. Kümeleme analizleri olarak adlandırılan beşinci grup çalışmada ise yağış, sıcaklık ve akım verilerinin çeşitli yöntemlerle gruplandırılması yani birbirine benzer karakter gösteren verilerin biraraya getirilmesi yer almaktadır.

Bu proje kapsamında bugüne kadar yapılmış olan çalışmalardan farklı olarak hidrolojik ve iklimsel parametreler arasındaki ilişkilerin açıklanmasını diğer bir ifadeyle hidrolojik ve iklimsel parametrelerin birlikte oluşturdukları örüntülerin ortaya çıkarılması ile hidrolojik sistemin iklimsel faktörlerden etkilenebilirliğinin belirlenmesi amaçlanmıştır.

b) Türkiye hidrolojik ve iklimsel veri serilerinde yapılan çalışmalarda çoğunlukla kullanılan yöntemler Mann Kendall eğilim testi, korelasyon testi, lineer regresyon ve çeşitli homojenlik testleri gibi istatistiksel yöntemlerdir. Son yıllarda bazı veri madenciliği metotları (k-ortalama analizi gibi) kümeleme analizlerinde kullanılmıştır. Bu çalışmada kullanılması önerilen ve bir veri madenciliği yaklaşımı olan “Birliktelik Analizi” hidrolojik ve iklimsel bağlantıların belirlenmesi amacıyla kullanılmamıştır.

Bu projede parametreler arasındaki ilişki ve örüntülerin belirlenmesini amaçlayan birliktelik analizinin uygulanması veri serileri arasındaki ilişkilerin “birliktelik kuralları” şeklinde ortaya çıkarılmıştır. Birliktelik kuralları yaklaşımı basit korelasyona analizlerinden tamamen farklı bir yaklaşım olup parametrelerin lineer olmayan ilişkilerini kurallar ve neden-sonuç ilişkileri şeklinde açıklamayı amaçlamaktadır. Bu özelliğiyle geleceğe yönelik tahminlerde bulunabilmeye imkan sağlayan bilgiler üretmektedir.

c) Yapılan çalışmalar veri serilerinin eğilim analizleri ile zamansal özelliklerini analiz etmişlerdir. Çalışmalarda yapılan kümeleme analiz yöntemleri ise verilerin mekansal özelliklerini (komşuluk ilişkisi gibi) göz önüne almamıştır. Dolayısı ile yapılan çalışmalar verilerin mekansal ve mekan-zamansal özelliklerini analiz etmekte yetersiz kalmışlardır. Aynı zamanda daha önce yapılan çalışmalar hidrolojik ve iklimsel parametreler arasındaki ilişkileri mekansal olarak ifade etmemişlerdir.

Bu projede uygulanan olan veri madenciliği yaklaşımında mekansal özellikleri içeren yöntemler kullanılmıştır.

d) Bugüne kadar yapılan çalışmalarda kullanılan istatistiksel yöntemler büyük-ölçekli kümelerin hızlı biçimde analizi için yetersizdir.

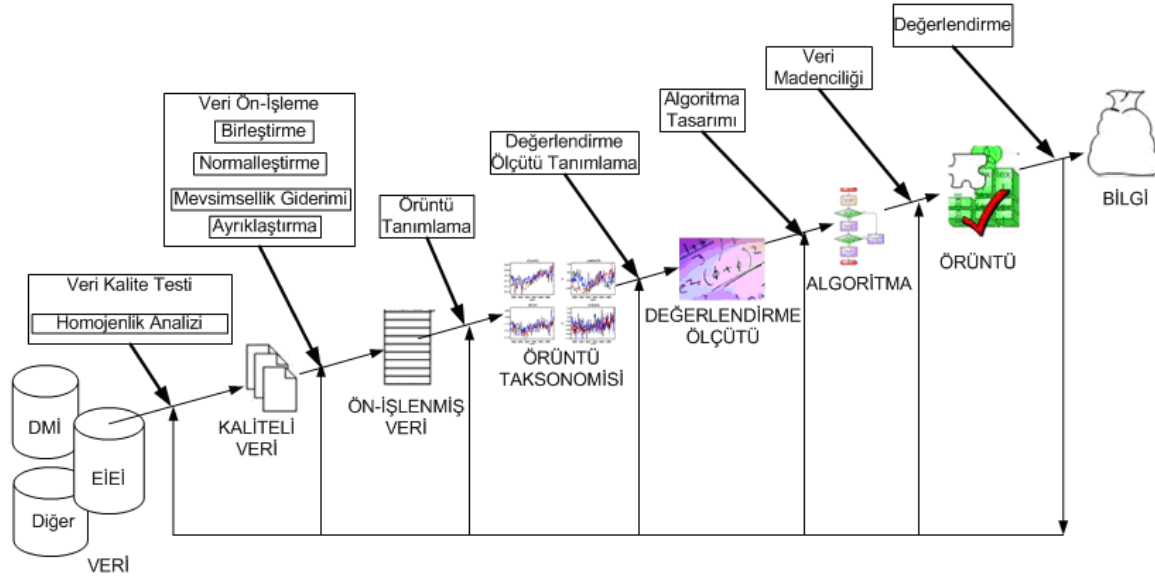
Bu projede hidrolojik ve iklimsel veri serileri arasındaki ilişkilerin belirlenmesi için hızlı ve etkin algoritmalar geliştirilmiştir.

3. GEREÇ VE YÖNTEM

Hidrolojik ve iklimsel veriler ülkemizde ve dünyada uzun süredir kaydedilmektedir ve kaydedilen verilerin çeşidi ve miktarı gelişen teknoloji ile paralel olarak artmaktadır. Bu verilerin etkili bir biçimde analiz edilmesi ve içerisinde saklı olan anlamlı mekan-zamansal ilişkilerin ve örüntülerin zamanında keşfi ise gün geçtikçe önem kazanmaktadır. Her geçen gün artan verilerin klasik yöntemler kullanılarak veya manuel olarak analiz edilmesi büyük miktarda iş gücü ve zaman kaybına neden olmakta ve veri kümelerinde gizlenmiş olan bilgilerin tamamının çıkarımında yetersiz kalmaktadırlar.

Bu projede hidrolojik ve iklimsel parametrelerin birbirleriyle ilişkileri ve beraber oluşturdukları mekansal-zamansal örüntüler veri madenciliği yaklaşımları kullanılarak keşfedilmiştir. Bu çalışmada uygulanan veri madenciliği süreci Şekil 3.1’de gösterilmektedir. İlk aşamada veriler çeşitli kurumlardan temin edilmiştir. Daha sonra temin edilen verilerin kalite analizleri ve ön-işlemesi çeşitli yöntemler kullanılarak gerçekleştirilmiştir. Hidrolojik ve iklimsel parametrelerin oluşturacakları birliktelik örüntüleri tanımlanmıştır. Belirlenen örüntü için bu ilişki ve örüntüleri niceleyen bir değerlendirme ölçütü tanımlanmıştır ve tanımlanan değerlendirme ölçütleri kullanılarak büyük veritabanlarından bu ilişkilerin ve örüntülerin hızlı bir şekilde keşfi için etkili ve özgün bir algoritma tasarlanmıştır. Son olarak geliştirilen yöntem Türkiye için hidrolojik ve iklimsel parametreler arasındaki ilişki ve örüntülerin keşfedilmesinde kullanılmıştır. Şekil 3.1’ den de görüleceği üzere uygulanan veri madenciliği süreci geri beslemeli bir yapıya sahiptir. Her bir aşamadan sonra elde edilen bulguların analizi sonucunda bir sonraki aşamaya geçilip geçilmeyeceğini veya hangi aşamaların tekrarlanması gerektiğini belirlenmiştir.

Aşağıda projede uygulanan yöntemin aşamaları ayrıntılı bir biçimde açıklanmaktadır.



Şekil 3.1. Projede uygulanan veri madenciliği yönteminin aşamaları

3.1. Veri Toplama ve Veri Kalitesinin Değerlendirilmesi

Bu çalışmada hidrolojik ve iklimsel verileri arasındaki ilişkilerin ve oluşturdukları örüntülerin tespiti amaçlanmaktadır. Hidrolojik parametre olarak uzun-sürekli ve düzenli verinin mevcut olduğu akarsu akımları verileri ve iklimsel parametreler olarak ise yağış, sıcaklık, nisbi nem ve rüzgar hızı verileri kullanılmıştır.

Hidrolojik ve iklimsel parametrelere ait veriler farklı kurumlardan temin edilmiştir. Türkiye’de hidrolojik ve iklimsel veriler 1920’li yılların sonundan itibaren kaydedilmeye başlanmıştır. Bazı parametreler için bugüne kadar yaklaşık 70 yıllık veri serisi elde edilmiş olsa da, mevcut verilerin üç temel problemi bulunmaktadır. Bunlardan birincisi çeşitli nedenlerle sürekli kayıt yapılamamasından kaynaklanan serilerdeki eksikliklerdir. İkinci problem ise istasyonların yer değiştirmesi ve ölçüm yönteminin veya aletlerinin değiştirilmesi gibi nedenlerle veri serilerinin homojenliğini yitirmesidir. Yine barajlar veya savaklar gibi yapılarla akarsuların doğal akışlarına ve göllerin doğal su seviyelerine müdahaleler olmuştur. Çalışmada kullanılacak istasyonlar belirlenirken bu tür problemleri taşımayan istasyonların seçimine ve elde edilen verilerin homojen özellik taşımasına dikkat edilmiştir.

İklim parametrelerine ait veriler Devlet Meteoroloji İşleri Genel Müdürlüğü’nden (DMİ) temin edilmiştir. DMİ büyük klima, küçük klima, yağış olmak üzere üç tip istasyonda meteorolojik veri toplamaktadır. Büyük klima istasyonları sıcaklık, yağış, nisbi nem parametreleri için istasyon kuruluşlarından itibaren ve rüzgar hızı, güneşlenme süresini de

kapsayan diğer 27 parametre için ise 1975 yılından itibaren günlük veri toplamaktadır. Küçük klima istasyonlarında ise toplanan veriler 1975 yılından başlamaktadır. Küçük klima istasyonlarında sıcaklık, yağış parametreleri için veri toplansa da genelde daha sınırlı sayıda parametre için olan veriler toplanmaktadır. Bu çalışma kapsamında büyük klima istasyonlarına ait veriler DMI'den talep edilmiştir.

Çalışmada kullanılacak hidrolojik veriler olan akarsu akımlarına ait veriler ise Elektrik İşleri Etüt İdaresi Genel Müdürlüğü'nün (EİE) yayınlarından elde edilmiştir. EİE 26 akarsu havzasında yer alan 800'e yakın akım gözlem istasyonuna sahiptir. Bunlardan yaklaşık 300 tanesi halen faaliyettedir. Çalışmada kullanılacak akım gözlem istasyonları belirlenirken veri zaman serileri şu üç faktör açısından değerlendirilmiştir: 1) süreklilik 2) doğallık ve homojenlik 3) uzunluk. Süreklilik açısından değerlendirmede aylık bazda eksik veri bulunmamasına dikkat edilmiştir. Doğallık ve homojenlik açısından değerlendirilmede veri serilerinin alındığı akarsu kollarının doğal su rejimine sahip olup olmadıkları (baraj, savak gibi müdahalelerin mevcut olup olmadığı) ve homojenlikleri değerlendirilecektir. Daha önce Kahya ve Karabörk [14], 1964-1994 önemi 22 havzaya dağılmış 76 akım gözlem istasyonu verisinde regülasyon etkisinin mevcut olmadığı belirlemişlerdir. Bu projede aynı istasyonlara ait veriler seçilmiştir.

Elde edilen veriler Microsoft Access'de hazırlanan bir veritabanında depolanmıştır.. Aşağıda veri kalitesinin değerlendirilmesi için kullanılan homojenlik analizi yöntemleri açıklanmaktadır.

3.1.1. Homojenlik Analizleri

Meteorolojik veriler kullanılarak yapılan çalışmaların doğruluğu ve güvenilirliği verilerin kalitesi ile doğrudan ilgilidir. Verilerin kalitesini belirleyen faktörlerin başında ise veri serilerinin homojen olması gelmektedir. Serilerinin homojen olması, verilerin aynı yığından gelmesi ve rastgele karakterde (birbirinden bağımsız ve herhangi bir dış etki altında kalmamış) olması olarak tanımlanabilir. Homojen meteorolojik seriler, sadece iklimsel faktörlerden etkilenen seriler olarak da ifade edilebilir [71]. Meteorolojik serilerinin homojenliği, istasyonların yer değişimi, ölçüm aletinin veya yönteminin değişimi, istasyon çevresinin değişimi, ölçüm aletine bağlı hatalar ve gözlemsel değişimlerden etkilenir [72]. Veri serilerinin homejenliğini bozan bazı değişimler (ör., ölçüm yönteminin değişimi) keskin süreksizliklere neden olurken, istasyon çevresinin değişimi verilerde yavaş yavaş kademeli

sapmalara neden olabilir [73]. Bu çalışmada Türkiye meteorolojik verilerinin homojenliğinin değerlendirilmesi amaçlanmaktadır.

Bir veri serisinin homojen olup olmadığının anlaşılması için yapılan homojenlik analizleri doğrudan ve dolaylı olmak üzere iki yaklaşımla gerçekleştirilebilir. Doğrudan yaklaşımda istasyon geçmişinde veri serisinin sürekliliğinin yok olmasına neden olacak bir olayın var olup olmadığı istasyon geçmişi incelenerek belirlenmektedir [73]. Bu yöntemde sonuçların güvenilirliği istasyon geçmişi bilgilerinin doğruluğuna bağlıdır. Dolaylı yaklaşımda ise veri serisinin sürekliliği istatistiksel yöntemler aracılığıyla analiz edilmektedir. Dolaylı yaklaşımda kullanılan testler göreceli (relative) ve mutlak (absolute) testler olmak üzere iki kısımda incelenebilir [73]. Göreceli testler ile homojen olduğu düşünülen ve referans istasyon olarak adlandırılan komşu istasyonların verileri ve homojenlik analizine tabi tutulan istasyonun verileri karşılaştırılmaktadır [73]. Göreceli testler uygulandığında referans istasyon verilerinin aslında homojen olmaması durumunda yanlış sonuçlar ortaya çıkabilir [13]. Özellikle ölçüm aletlerin aynı zamanda değişimi gibi eş zamanlı değişikliklerin meydana gelmesi durumunda göreceli testlerin kullanılması mümkün değildir [33]. Mutlak testler ile ise, her bir istasyon verisi kendi içinde değerlendirilmektedir [74]. Homojenlik analizlerinde sıklıkla kullanılan mutlak testler arasında Pettitt testi, Standart normal homojenlik testi (SNHT), Buishand range testi ve Von Neumann ratio testi sayılabilir [75, 76]

Bu projede Türkiye meteorolojik verilerinin homojenliği mutlak testler kullanılarak değerlendirilmiştir. Çalışmada yağış, sıcaklık, bağıl nem dışında daha önce yapılan çalışmalardan incelenmeyen rüzgar hızı serilerinin de homojenliği analiz edilmiştir.

Homojenlik analizleri için Pettitt testi, Standart normal homojenlik testi (SNHT), Buishand range testi ve Von Neumann ratio testi olmak üzere dört farklı mutlak test kullanılmış ve veri serilerinin homojenliği %1 (0.01) anlamlılık düzeyinde araştırılmıştır. SNHT ve Buishand range testi verilerin normal dağılım gösterdiği varsayımına dayanır. SNHT tek kırılmalar için muhtemel kırılmanın hangi dönemde gerçekleştiğini saptayabilir ve zaman serilerinin başlangıcı ya da bitişine yakın meydana gelen kırılmalara karşı duyarlıdır. Buna karşılık Pettitt testi ve Buishand range testi zaman serilerinin ortalarına yakın kırılmalara karşı daha fazla duyarlıdır [76]. Von Neumann ratio testi ise kırılmanın zamanı hakkında bilgi vermemekle birlikte ani ve keskin değişimler dışında kalan sapmalara karşı da duyarlıdır [77].

Daha önce Wijngaard v.d. [76] tarafından yapılan çalışmada meteorolojik parametrelerin

homojenliđinin deęerlendirilmesinde bu alıřmada da kullanılan mutlak testler kullanılmıř ve testlerin homojenliđi reddetme durumlarına gre veri serilerinin kullanılabilirliđini ifade etmek iin bir sınıflandırma yapmıřlardır. Bu sınıflandırmaya gre; drt mutlak test iinde en az  test bir istasyon verisinin homojen olmadıđını gsteriyorsa veriler ‘řüpheli (suspect)’, iki test homojen olmadıđını gsteriyorsa veriler ‘belirsiz (doubtful)’, bir ya da sıfır test homojen olmadıđını gsteriyorsa veriler ‘kullanıřlı (useful)’ olarak nitelendirilmiřtir. Ayrıca ‘kullanıřlı’ ve ‘belirsiz’ olduđu bulunan serilerin kullanımının daha gvenilir olduđu ifade edilmiřtir. Bu projede de aynı sınıflandırma kullanılarak analiz edilen verilerin kullanılabilirliđini deęerlendirilmiřtir.

3.2. Veri n-İřleme

Veri n-İřleme verinin analiz veya veri madenciliđi iin daha uygun hale getirilmesini sađlamıřtır. Veri n-İřleme ok farklı ve karmařık teknikleri ierebilir (birleřtirme, rnekleme, boyut indirgemesi, ayırklařtırma, deđiřken dnřtrme ve mevsimselliđin giderilmesi gibi) [5]. Bu alıřmada genel olarak drt yntem kullanılmıřtır. Bunlar birbirleriyle iliřkilendirilecek akım gzlem istasyonları ve klima istasyonların belirlenmesi, veriler arasındaki period-zaman farklılıklarının giderilmesi (birleřtirme), mevsimselliđin giderilmesi ve normalleřtirme, ve verilerin sınıflandırılması (ayırklařtırma) olarak sıralanabilir. Bu iřlemler ařađıda aıklanmaktadır.

3.2.1. İstasyon Eřleřtirme

Hidrolojik ve iklim parametrelerine ait veriler birbirinden bađımsız olarak ve farklı kurumlarca toplanmaktadır. Bu nedenle hidrolojik ve iklim verisi sađlayan istasyonların eřleřtirilmesi gerekmektedir. Eřleřtirme yapmak iin Cıđızođlu [15] tarafından da uygulanmıř olan yntem uygulanmıřtır. Bu yntemde aynı havza iinde yer alan her akım lm istasyonuna yakın olan klima istasyonları belirlenmiř ve daha sonra korelasyon analizleri yapılarak akım verileri ile iklimsel veriler arasında en yksek korelasyon katsayılarına sahip klima istasyonu seilmiřtir. Ayrıca bu seimde havza sınırları gzetilerek, hidrolojik ve klima istasyonlarının aynı havza iinde yer almasına dikkat edilmiřtir.

İstasyon seimini takiben akım ile meteorolojik parametreler arasındaki iliřikler daha ayrıntılı olarak incelenmiřtir. Bu incelemede  tr korelasyon analizi yapılmıřtır: (1) Akım/gl

seviyeleri ile yağış, ortalama sıcaklık, rüzgar hızı, bağıl nem ve güneşlenme şiddeti parametreleri arasındaki ilişkiler, (2) Akım/göl seviyeleri ile 1, 2, 3, 4 ve 5 yıl önceki yağış verileri arasındaki ilişkiler, (3) Akım/göl seviyeleri ile 1, 2, 3, 4 ve 5 yıl önceki sıcaklık verileri arasındaki ilişkiler. Yağış ve sıcaklık akım/göl seviyelerini etkileyen en önemli parametreler oldukları için bu parametrelerin 5 yıllık dönemdeki değerler ile akım/göl seviyeleri arasındaki ilişkiler incelenmiştir.

3.2.2. Veri Birleştirme

Bu aşamada amaç farklı zaman dilimleri için veya farklı periodlarda kaydedilmiş olan verilerin aynı zaman dilimi veya period için ifade edilmesidir. Örneğin akım verileri aylık ortalama olarak mevcut olduğu için, yağış ise günlük olarak mevcut olduğu için her ikisinin de aylık olarak ifade edilmesi sağlanmıştır.

3.2.3. Verilerin Mevsimselliğinin Giderilmesi ve Normalleştirme

Mevsimselliğin kaldırılması, mevsimsel değişim maskesinin kaldırılarak verinin gerçek hareketinin, eğiliminin belirlenmesine imkan sağlamaktadır. Mevsimselliğin kaldırılması çeşitli uygulama alanlarında kullanılmıştır. Ekonomik analizde ekonomik dalgalanmaların mevsimsel etkilerin kaldırılarak gerçek ekonomik eğilimlerin tespitinde [78], halk sağlığında hastalıkların mevsimsel salınımları dışındaki ortaya çıkma eğilimlerinin tespitinde [79] kullanılmıştır. Mevsimsel olmayan ilginç hidrolojik ve iklimsel örüntülere örnek olarak El-Nino olayları, uzun dönemli salınımlar (decadal oscillations), eğilimler (küresel ısınma, sel taşkınlarının daha sık olması gibi) verilebilir. Veri üzerindeki mevsimsel değişimin kaldırılması için literatürde çeşitli yöntemler mevcuttur. Bunlardan bazıları ayırık fourier transformu, aylık z-skoru, tekil veri çözümlemesi, ve hareketli ortalama olarak listelenebilir [57]. Kumar v.d. [80] tarafından yapılan çalışmada ayırık fourier transformu, aylık z-skoru, tekil veri çözümlemesi, ve hareketli ortalama tekniklerinin mevsimsel değişimi kaldırmada etkilerini incelemişlerdir ve avantaj ve dezavantajlarını ortaya konulmuştur. Bu çalışmada mevsimselliğin giderilmesi ve normalleştirme için öncelikle z-skoru yönteminin kullanılabilirliği değerlendirilmiştir [57].

3.2.4. Veri Ayırıklaştırma

Kaydedilen hidrolojik ve iklimsel veriler sayısal zaman serileridir. Kaydedilen veriler farklı özelliklere sahip olabilir. Bu nedenle iki farklı parametre verilerinin birliktelik veya örüntü

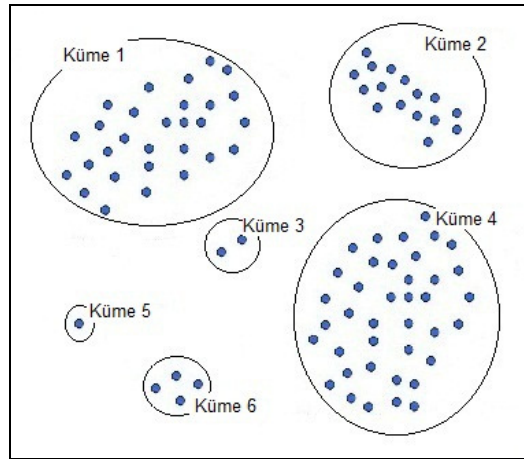
analizi yapıldığında bütün kombinasyonlar göz önüne alınacağından veri madenciliği karmaşıklığı eksponansiyel olarak artar. Bu nedenle bu aşamada amaç verileri ayrıklaştırılarak veya gruplara ayrılarak parametrelerin almış oldukları değerleri sınıflara ayırmaktır. Örnek olarak çeşitli eksi ve artı verileri alan sıcaklık verisi düşük, normal, yüksek veya çok soğuk, soğuk, normal, sıcak, çok sıcak gibi kavramsal bölmelere ayrılabilir veya akım verisi yüzdelik dilimler ile (%90'lık akım gibi) ifade edilebilir. Ayrıklaştıma için sınıflandırma algoritmaları (K-Means gibi) karmaşık veya eşit bölmelere ayırma gibi basit yöntemler kullanılabilir. Bu proje kapsamında veri ayrıklaştırma yüzdelik (persentil) dilimler yöntemi, istatistiksel yöntem ve kümeleme yöntemi olarak adlandırdığımız yöntemler üzerinde durulmuştur.

Yüzdelik yöntemde veriler yüzdelik dilimlere ayrılmaktadır. Bir yüzdelik değer yüz birimden oluşan bir ölçekte bir dağılımın bu değere eşit veya küçük yüzdesini ifade eder. Örneğin akım verilerini içeren bir seride %90'lık değer ölçülen akım değerlerinin %90'ından küçük olan değeri ifade etmektedir. Yüzdelik dilimler yöntemi özellikle hidrolojik çalışmalarda farklı düzeydeki akımları ifade etmek için sıklıkla kullanılmaktadır. Genellikle Q95 and Q90 olarak gösterilen %90'lık ve %95'lik akım değerleri düşük akım, Q10 ve Q5 olarak ifade edilen %5'lik ve %10'luk akım değerleri ise yüksek akım olarak nitelendirilmektedir [81]. Bir veri madenciliği uygulanmasında daha önce Tan vd. [61] tarafından iklim indislerini sınıflandırmak için %95 ve %5'lik dilimler sınır olarak kullanılmıştır.

İkinci yöntem olarak veri serisinin istatistiksel özelliklerini dikkate alan bir ayrıklaştırma yöntemidir. Bu yöntemde serinin ortalama ve standart sapma (SS) değerleri hesaplanmakta ve veriler bu değerler kullanılarak gruplara ayrılmaktadır. İstatistiksel yöntem daha önce bazı çalışmalarda veri ayrıklaştırma amacı ile kullanılmıştır. Novotny and Stefan [51] “ortalama+SS” (ortalama artı standart sapma) ve “ortalama+2SS” değerlerinden yüksek akımları çok yüksek akım olarak adlandırmışlardır. Bu yöntem bazı veri madenciliği uygulamalarında da ayrıklaştırma amacıyla kullanılmıştır. Yağış verilerinin sınıflandırılmasında Dhanya ve Kumar [66] “ortalama+1.5SS” değerinden büyük yağış düzeylerinin çok yüksek, “ortalama+1.5SS”den küçük değerleri ise çok düşük olarak sınıflandırmışlardır. Tan vd. [61] sıcaklık ve yağış gibi bazı iklim parametrelerine ait değerlerin ayrıklaştırılmasında bu yöntemi kullanmışlardır.

Son yöntem olan kümeleme yönteminde amaç birbirleri ile benzer özellik gösteren verilerin aynı grup altında toplanmasıdır. Proje kapsamında uzaklık tabanlı kümeleme yöntemlerinden biri olan DBSCAN (density-based clustering algorithm) algoritması [82] kullanılarak bir

çalışma yapılmıştır. DBSCAN veritabanında yer alan kümeleri tespit etmek yanında anormal verileri (yani hiçbir kümeye ait olmayan verileri) de tespit edebilmektedir. DBSCAN’de iki parametre bulunmaktadır: epsilon (uzaklık) ve minpts (minimum nokta sayısı). DBSCAN, verileri merkez nokta, sınır nokta ve anomali nokta olarak etiketlemektedir. Bir noktanın merkez nokta olabilmesi için epsilon uzaklığında en az minpts sayıda komşu noktayı bulundurması gerekmektedir. Sınır noktalar ise merkez nokta şartını sağlamayan ancak bir merkez noktanın komşusu bulunan noktalar olarak tanımlanabilir. Anomali noktalar da ne merkez ne de sınır nokta tanımına uymayan noktalardır. DBSCAN algoritmasının işleyiş biçimi Şekil 3.2’de gösterilmektedir. Şekil 3.2’deki kümeler sadece epsilon değerini kümeleme ölçütü olarak alan bir kümeleme yaklaşımının sonucunu göstermektedir. DBSCAN, bu uzaklık ölçütüne ek olarak, grubu bir küme olarak etiketleyebilmek için bir de minimum komşu noktaları, minpts, ölçütüne ihtiyaç duymaktadır. Örneğin; eğer minpts değeri 3 olarak seçilirse DBSCAN algoritması grup 1, grup 2, grup 4 ve grup 6’yı küme olarak etiketleyecektir. Buna karşılık bir küme oluşturabilmek için yeterli sayıda noktaya sahip olmadıkları için grup 3 ve grup 5 anomali olarak etiketlenmektedir. Benzer şekilde, eğer DBSCAN kullanılarak minpts değeri 5 olarak belirlenirse grup 3, grup 5 ve grup 6 anomali olarak etiketlenir. Çalışma kapsamında farklı minpts ve eps değerleri ile deneyler yapılmıştır.



Şekil 3.2. DBSCAN algoritması ile kümeleme

3.3. Hidrolojik ve İklimsel İlişkilerin ve Örüntülerin Tanımlanması

Projenin bu aşamasında hidrolojik ve iklimsel parametrelerin oluşturdukları birliktelik örüntüleri tanımlanmıştır.

Genel olarak tanımlanacak olursa birliktelik analizi bir işlem kaydında (transaction) bir ögenin meydana gelme olasılığını diğer ögelerin meydana gelme olasılıklarından tahmin etmek için kurallar bulunmasıdır. Birliktelik analizi kullanıcı tarafından belirlenmiş frekans (sıklık) değerinden daha yüksek frekansa (daha sık) sahip olan $\{A\} \rightarrow \{B\}$ şeklindeki kuralların bulunmasını sağlar. Bu kuralda A ve B olayları veya iki farklı parametreyi göstermektedir. Örneğin, A yağış, B ise pik akım biçimi olabilir. $\{A\} \rightarrow \{B\}$ kuralı ise “eğer A gerçekleşirse, B gerçekleşir” şeklinde okunur. Bu yönüyle birliktelik analizi sadece parametreler arasındaki ilişkiyi açıklamamakta aynı zamanda neden-sonuç ilişkilerini gösteren kurallar oluşturmaktadır. Birliktelik analizi aynı zamanda birçok parametreyi aynı zamanda analiz etme kabiliyetine sahiptir. Yani birliktelik analiziyle $\{A, B, C\} \rightarrow \{D\}$ şeklinde veya $\{A, B, C\} \rightarrow \{D, E, F\}$ şeklinde kurallar (örüntüler) bulmak mümkündür.

Bu çalışmada odaklanılan kurallar iklimsel parametrelerin kuralların sol tarafında, hidrolojik parametrelerin ise kurallarının sağ tarafından yer aldığı birlikteliklerdir. Yani iklimsel faktörlerin akım üzerinde ne tür bir etkiye neden olduğu araştırılmıştır.

3.4. Değerlendirme Ölçütü Tanımlama

Değerlendirme ölçütleri ilginç ve faydalı mekan-zamansal hidrometeorolojik ilişki ve örüntülerin nicelenmesi için tasarlanan formüldür. Değerlendirme ölçütleri ilişki ve örüntüleri belirlemek ve bunları değerlendirmek, yorumlamak ve kullanmak için tanımlanır. Örnek değerlendirme ölçütü olarak asit ve bazların tespiti için kullanılan turnusol kağıdı verilebilir. Ayrıca değerlendirme ölçütleri veri madenciliği yöntemlerinin işlemsel karmaşıklığını indirmek için kullanılabilirler . Bu ölçütlerin tasarlanmasındaki zorluklardan birisi, uzmanlar tarafından yöneltilen “ne zaman ve nerede” sorularının cevaplarının bulunmasında karşımıza çıkar.

Klasik veri madenciliğinde parametrelerin ve ilişkilerin anlamlılıklarının ölçülmesi için yaygın olarak destek (support) değerlendirme ölçütü kullanılır. Destek; bir parametre değerinin kaç kez tekrar ettiğini ifade eder. Eğer bir parametrenin veya ilişkinin destek değeri

belirli bir eşik değerinden (threshold) yüksek ise o parametrenin veya ilişkinin anlamlı olduğuna karar verilir ve daha karmaşık ilişkilerin bulunması için bir sonraki aşamalarda kullanılır. Destek değerlendirme ölçütünü açıklamak için Tablo 1 de verilen örnek veri kümesi kullanılabilir. Tablo 3.1’de bir istasyona ait 10 günlük hidrolojik ve iklimsel verilerin değerleri listelemiştir. Örnekte parametrelerin gerçek sayısal değerleri yerine verilerin ayrıklaştırılması sonucu kavramsal isimler kullanılmıştır. Verilen tabloda sıcaklık değeri 1.gün, 4. gün, 6. gün, 9. ve 10. günlerde yüksek olduğu görülmektedir. Dolayısı ile sıcaklık değeri 10 günün 5 tanesinde yüksek olmuştur ve sıcaklığın yüksek olmasının desteği 5 (veya ondalıklı ifade ile 5/10)’dır ve destek(sıcaklık=yüksek)=5 olarak ifade edilebilir. Aynı şekilde diğer sıcaklık verileri için de destekleri bulunabilir: destek(sıcaklık=normal)=3 ve destek(sıcaklık=düşük)=2 ‘dir. Benzer şekilde diğer parametre değerlerinin de destek değerleri bulunabilir. Nisbi nem için destek değerleri de; destek(nisbi_nem=yüksek)=4, destek(nisbi_nem=normal)=4, ve destek(nisbi_nem=düşük)=2 şeklinde olacaktır. Daha sonra bir eşik değeri üzerinde kalan yani çok tekrar eden parametre değerleri bir sonraki aşama için kullanılır. Eğer eşik değer 4 ise desteği 4 ten az olanlar elenecektir çünkü bu eşik değerine göre yeterince tekrar etmemişlerdir. Bu eleme işlemi yapıldığında sıcaklığın yüksek (destek(sıcaklık=yüksek)=5) ve nisbi nemin yüksek (destek(nisbi_nem=yüksek)=4) ve normal (destek(nisbi_nem=normal)=4) olduğu durumların destekleri eşik değerine eşit veya ondan yüksek olduğu için kalacaklardır. Eşik değerini geçen bu ilişkiler sık bağlantı veya örüntü olarak isimlendirilir. Daha sonra sık örüntülerin beraber olma olasılıklarının destek değerleri hesaplanır ve böylece karmaşık ilişkiler keşfedilmeye çalışılır. Sıcaklığın yüksek ve nisbi nemin normal olduğu yerler tabloda incelendiğinde 4. gün bu kriteri sağlamaktadır. Bu örüntünün destek değeri {destek(sıcaklık=yüksek) ve destek(nisbi_nem=normal)} 1’ dir ve eşik değeri olan 4’den küçük olduğu için elenir. Buna karşılık sıcaklığın yüksek ve nisbi nemin yüksek olduğu 4 gün (1., 6., 9. ve 10. günler) vardır ve dolayısı ile destek(sıcaklık=yüksek ve nisbi_nem=yüksek)=4 ‘dür. Bu destek değeri eşik değeri kriterini sağladığı için {sıcaklık=yüksek ve nisbi_nem=yüksek} örüntüsü veya ilişkisi sık bir örüntüdür ve daha sonraki kombinasyon üretimlerinde kullanılabilir. Bir sonraki aşamada aynı işlemler yapıldığı zaman {sıcaklık=yüksek, nisbi_nem=yüksek, akım=yüksek} ilişkisi veya örüntüsü tespit edilecektir. Çünkü bu örüntü 1., 6., 9., ve 10 günlerde olmaktadır ve dolayısı ile destek değeri 4’tür.

Tablo 3.1: Bir istasyona ait 10 günlük hidrolojik ve iklimsel örnek veriler

Gün	İklimsel veriler		Hidrolojik Veri
	Sıcaklık	Nisbi Nem	Akım
1. gün	Yüksek	Yüksek	Yüksek
2. gün	Düşük	Normal	Yüksek
3. gün	Normal	Normal	Yüksek
4. gün	Yüksek	Normal	Normal
5. gün	Düşük	Normal	Normal
6. gün	Yüksek	Yüksek	Yüksek
7. gün	Normal	Düşük	Yüksek
8. gün	Normal	Düşük	Yüksek
9. gün	Yüksek	Yüksek	Yüksek
10. gün	Yüksek	Yüksek	Yüksek

Klasik veri madenciliğinde güvenilirlik değerlendirme ölçütü sebep-sonuç ilişkisini gösterir ve hangi parametrenin hangisine sebep olduğunu belirlemede yani eğer- öyleyse (if-then) kurallarının çıkarımında kullanılır. Eğer $\{X,Y\}$ beraber iki parametre ise $Güvenilirlik(X)=destek (X \cap Y)/ destek (X)$ olarak hesaplanır. Yukarıdaki örnekte $\{sıcaklık=yüksek \text{ ve } nisbi_nem= yüksek\}$ birlikteliği tespit edilmiştir. Burada sorulması gereken soru acaba sıcaklığın yüksek olması mı nisbi_nemin yüksek olmasına neden olmuştur yoksa nisbi nemin yüksek olması mı sıcaklığın yüksek olmasına sebep olmuştur. Formüle göre hesaplanırsa eğer sıcaklık yüksek ise nisbi nemin yüksek olma güvenirligi = destek (sıcaklık=yüksek ve nisbi_nem= yüksek)/destek(sıcaklık=yüksek) = $4/5= 0.8$ 'dir. Benzer eğer nisbinem=yüksek ise sıcaklığın yüksek olma güvenirligi = destek (sıcaklık=yüksek ve

$\text{nisbi_nem}=\text{yüksek})/\text{destek}(\text{nisbi_nem}=\text{yüksek}) = 4/4= 1$ ’dir. Dolayısı ile bu istasyonda nisbi nemin yüksek olması durumunda sıcaklık ta yüksek oluyor denilebilir ve “Eğer nisbi nem yüksek ise sıcaklıkta yüksektir” bağlantısı çıkarılabilir.

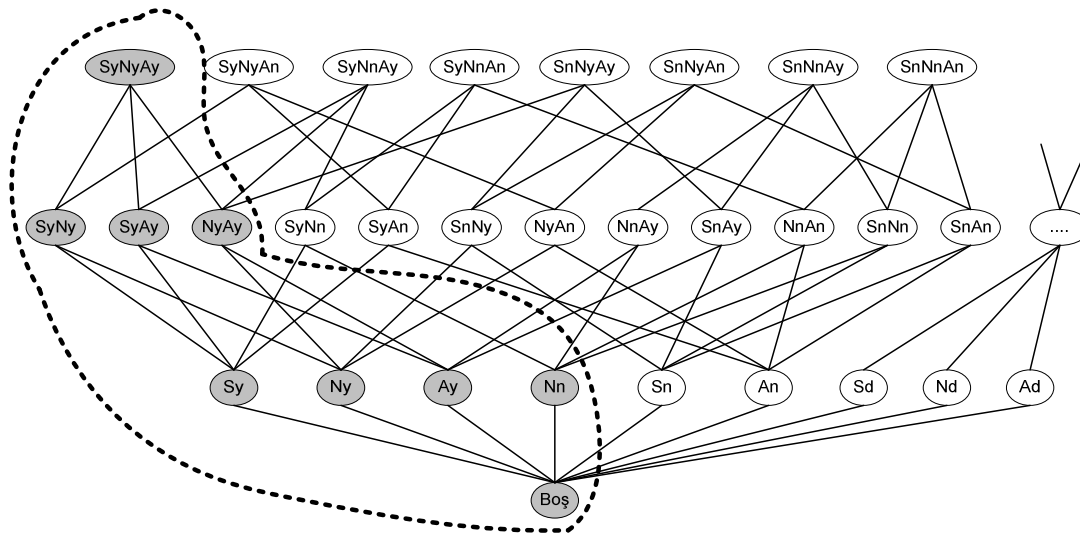
Bu araştırma projesinde öncelikle literatürde yer alan değerlendirme ölçütleri incelemiş ve gerekli olduğu durumda zaman, mekan ve diğer parametrelerin fonksiyonu olan değerlendirme ölçütleri formulize edilmiştir.

3.5. Hızlı ve Özgün Algoritmaların Tasarlanması

İklimsel ve hidrolojik verilerinin çeşitliliğinden ve büyük oluşundan dolayı mekansal-zamansal bağlantıların ve örüntülerin madenciliğinin işlemsel karmaşıklığı çok yüksek olacaktır. Yani ilişkilerin ve örüntülerin keşfi hem iş kaybına hem de zaman kaybına neden olacaktır. Bu işlemsel karmaşıklığı azaltmak için bir önceki aşamada (3.4) belirlenen veya tasarlanan değerlendirme ölçütlerinin avantajları kullanılarak yeni ve özgün algoritmalar tasarlanmıştır.

Verilerin mekansal ve mekan-zamansal özellikleri ihmal edildiğinde parametreler arasındaki örüntü ve bağlantıların keşfi için kullanılan en temel veri madenciliği algoritması apriori algoritmasıdır ve 1994 yılında Agrawal ve Srikant [58] tarafından geliştirilmiştir Apriori algoritması parametrelerin bütün olası ilişkilerini üretir. Daha sonra değerlendirme ölçütü (destek gibi) hesaplanır ve eşik değerini sağlayıp sağlamadığı araştırılır. Eğer eşik değerini sağlıyorsa bir sonraki ilişkiler üretilir ve destek değerleri eşik değerini sağlayıp sağlamadığı kontrol edilerek bütün sık örüntü ve ilişkilerin bulunması sağlanır. Apriori algoritmasında öncelikle parametrelerin destek değerleri bulunur ve sık olanlar ikili ilişkiler üretmek için kullanılır. Sonra ikili ilişkilerin destek değerleri bulunur ve sık olan ikili ilişkiler arasından üçlülere bularak madencilik işlemine devam eder. Özet olarak k boyutlu ilişkileri bulmak için sık k-1 boyutlu örüntüleri kullanır. Şekil 3.3’de Apriori algoritmasının Tablo 3.1’ de verilen veri kümesi için ürettiği kafes yapısı gösterilmektedir. Şekilde koyu renkle gösterilenler eşik değeri 4’ü sağlayan ilişkilerdir ve sık örüntülerdir. Sık örüntüler kesikli çizgi içerisinde gösterilmiştir. Algoritma öncelikle parametrelerin destek değerlerini bulmuştur ve Sy, Ny, Ay ve Nn olduğu durumlarının destek değerlerinin eşik değerini sağladığı ve sık olduğu tespit edilmiştir. Diğer durumlar (Sn, An, Sd, Nd, Ad) ise sık olmadıkları için elenirler ve bunların oluşturdukları kombinasyonlar üretilmezler. Sık olan Sy, Ny, Ay ve Nn durumları kullanılarak

SyNy, SyAy, NyAy, SyNn ve NnAy ikilileri bulunur. Tablo 3.1 incelendiğinde SyNn ve NnAy ikililerinin sık olmadıkları görülür ve sık olan SyNy, SyAy ve NyAy ikilileri kullanılarak SyNyAy üçlüsü bulunur. Tablo 1 incelendiğinde SyNyAy üçlüsünün de sık olduğu tespit edilir ve algoritma burada durur. Sonuç olarak Tablo 3.1’de verilen veri kümesi için apriori algoritması Sy, Ny, Ay ve Nn, SyNy, SyAy, NyAy ve SyNyAy sık ilişki ve örüntülerini tespit eder.



Şekil 3.3. Apriori algoritması tarafından üretilen olası örüntüler (kesik çizgili bölge içerisindeki koyu renkli örüntüler Tablo 1’ de verilen veri kümesin den keşfedilen ve eşik değeri 4’ü sağlayan sık örüntü ve ilişkileri gösterir.

Kısaltmaların açıklamaları : Sy=Sıcaklık yüksek, Sn=Sıcaklık normal, Sd=Sıcaklık düşük, Ny=Nisbi_nem yüksek, Nn=Nisbi_Nem normal, Nd=Nisbi_Nem düşük, Ay=Akım yüksek, An=Akım normal, Ad=Akım düşük,

Literatürde apriori algoritmasının geliştirilmiş çeşitli versiyonlarını bulmak mümkündür. Ayrıca çalışma prensibi farklı diğer algoritmalar da önerilmiştir. Han ve Kamber [5] tarafından sık örüntü ağacı ve algoritması (frequent pattern tree (FP-Tree)) önerilmiştir. Veri madenciliği algoritmaları örüntü ve ilişkilerin monotonik olması esasına dayanır. Yani eğer bir örüntünün destek değeri, bu örüntünün bütün alt kümelerinin destek değerinin en küçüğüne ya eşittir ya da alt kümelerin destek değerlerinden daha küçüktür. Örnek olarak {SyNyAy} sık örüntüsünün bütün alt kümeleri sık örüntüdür. Bu kurala monotoniklik prensibi

denir. Apriori algoritmasında esas amaç olabilecek bütün birliktelik ihtimallerini denemektir ve bu oldukça karmaşık ve zaman alan bir süreçtir (Şekil 3.1). Bu proje kapsamında apriori algoritması kullanılarak ve mekansal ve mekan-zamansal iklimsel ve hidrolojik veri madenciliği yöntemleri geliştirilecektir.

3.6. İlişki ve Örüntü Madenciliği

Bu aşamada öncelikle hazırlanan veriler klasik veri madenciliği yöntemleri (apriori gibi) kullanılarak ilk analizler yapılacaktır. Bu analizler verilerin düzenlenmesinde ve algoritmaların geliştirilmesinde yönlendirici olacaktır. Daha sonra geliştirilen algoritmalar ve yöntemler 3. aşamada sınıflandırmada verilen mekansal-zamansal hidrometeorolojik örüntülerin madenciliği için kullanılacaktır. Geliştirilen algoritmalar öncelikle küçük ve karakteristik veri kümeleri üzerinde çalıştırılacaktır. Böylece algoritmaların olası hataları ayıklanacak ve uygun veri yapıları kullanılarak algoritmaların optimizasyonları yapılacaktır. Bunun yanında özgün ve yeni algoritmaların ilk tasarımları ve denemeleri matris ve vektör işlemlerinde büyük kolaylık sağlayan ve kodların hatalarının ayıklanmasında ve değişken verilerinin takibinde arayüzleri bulunan Matlab programı kullanılarak yapılmıştır.

3.7. Yöntemlerin ve Örüntülerin Değerlendirmesi

Önerilen yöntemlerin ve değerlendirme ölçütlerinin analitik ve deneysel değerlendirmeleri gerçekleştirilecektir ve veritabanlarından keşfedilen bağlantıların ve örüntülerin geçerliliği (doğruluğu ve anlamlılığı) araştırılacaktır. Elde edilen algoritmalar Türkiye için hidrolojik ve iklimsel verilerin analizinde kullanılacaktır. Böylece Türkiye için hidrolojik ve iklimsel parametreler arasındaki ilişkiler ve örüntüler belirlenecektir.

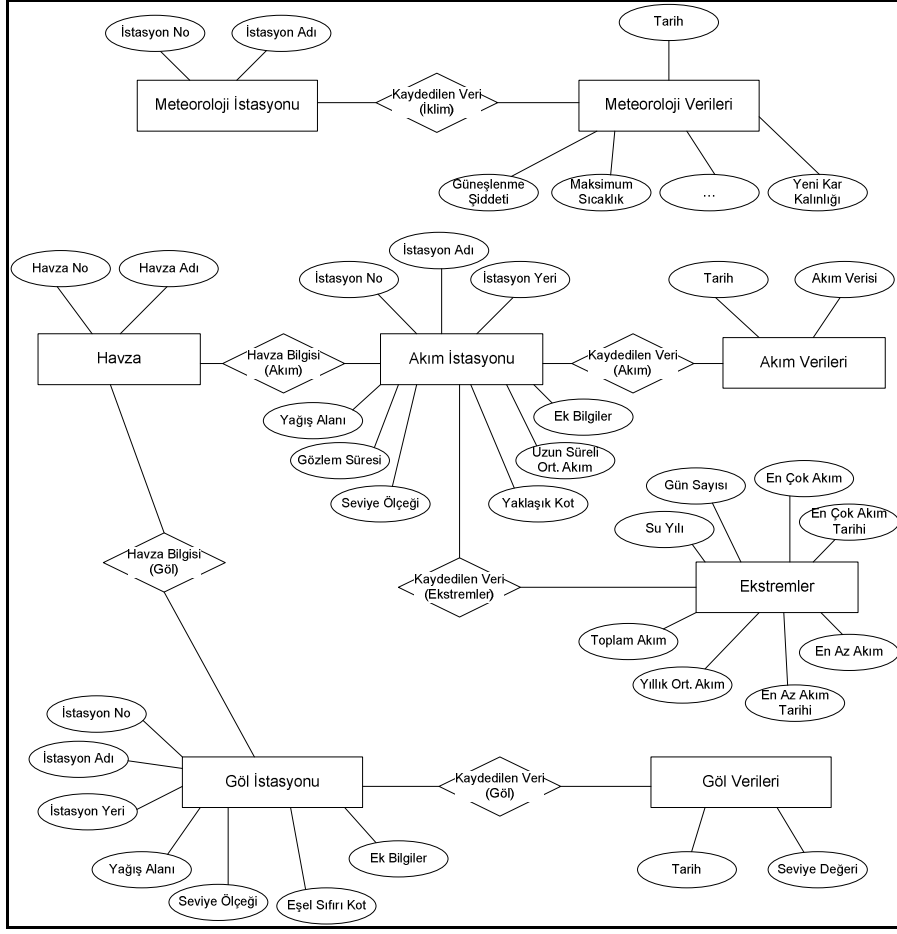
4. BULGULAR

4.1. Veri Toplama ve Veri Kalitesinin Değerlendirilmesi

Bu aşamada yapılan çalışmalar verilerin veritabanına yerleştirilmesini ve veri kalitesinin değerlendirilmesini içermektedir.

4.1.1. Veritabanının Oluşturulması ve Verilerin Veritabanına Yerleştirilmesi

Hidrolojik ve iklimsel veriler veritabanına aktarılmadan önce geliştirilecek veritabanının kavramsal modelini ifade eden Varlık-İlişki (ER-Entity-Relationship) diyagramı geliştirilmiştir (Şekil 4.1). ER diyagramında dikdörtgenler (□) varlıkları yani varolan ve diğer nesnelerden ayırt edilebilen nesneleri ifade etmektedir. Ayrıca varlıklar veritabanı tasarımında oluşturulacak olan tabloları göstermektedir. Örneğin Şekil 4.1’de “Meteoroloji İstasyonu”, “Akım İstasyonu” varlıkları ifade etmektedir ve veritabanında iki ayrı tabloya denk gelmektedir. ER diyagramlarında yer alan elipsler ise (○) varlıkların niteliklerini (özelliklerini) ifade etmektedir. Örneğin Şekil 4.1’de yer alan ve “Meteoroloji İstasyonu”na bağlı olan “İstasyon Adı” ve “İstasyon No” meteoroloji istasyonun tanımlamakta kullanılan niteliklerdir ve bu nitelikler “Meteoroloji İstasyonu” tablosu oluşturulurken tabloda her bir kaydın özelliklerini ifade eden iki sütun olarak tasarlanır. Son olarak ER diyagramlarında yer alan eşkenar dörtgen (◇) ise bir veya birden fazla tablo arasındaki ilişkiyi gösterir. Örneğin Şekil 4.1’de yer alan “Meteoroloji İstasyonu” ve “Meteoroloji Verileri” arasında yer alan “Kaydedilen Veri (İklim)” iki varlık arasındaki ilişkiyi (meteorolojik verilerin meteoroloji istasyonlarında kaydedildiğini) ifade eder ve hangi meteoroloji istasyonunda hangi bilgilerin tutulduğunu belirler. İlişkiler, veritabanı oluşturulurken genellikle ayrı bir tablo olarak tasarlanır.



Şekil 4.1. Veritabanı için Varlık-İlişki (ER-Entity-Relationship) diyagramı

ER diyagramı kullanılacak olan veritabanı yazılımından (Microsoft Access, Microsoft SQL Server, Oracle gibi) bağımsızdır. Diyagram geliştirilirken mümkün olduğunca birbirleriyle alakalı verilerin bir arada olmasını sağlayan tablolar ve ilişkiler tasarlanır. ER diyagramı tarafından tasarlanan tablolar herhangi bir veritabanı yazılımında gerçekleştirildikten sonra diğer bir veritabanı programına taşınması mümkündür.

Bu proje kapsamında hidrolojik ve iklimsel parametrelere ait veriler Microsoft Access’de hazırlanan veritabanına aktarılmıştır.

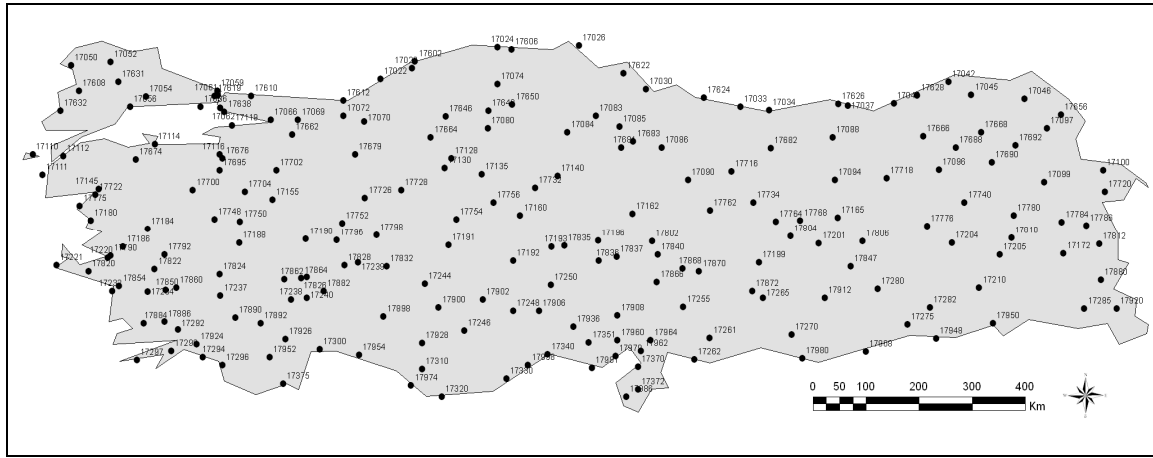
4.1.2. Verilerin Kalite Kontrolünün Yapılması

Verilerin kalite kontrolü iki şekilde yapılmıştır: (1) Veri serilerinin incelenmesi ve istasyon seçimi (2) Homojenlik analizleri. Bu iki aşamada elde edilen sonuçlar aşağıda açıklanmaktadır.

4.1.2.1. Veri serilerinin incelenmesi ve istasyon seçimi

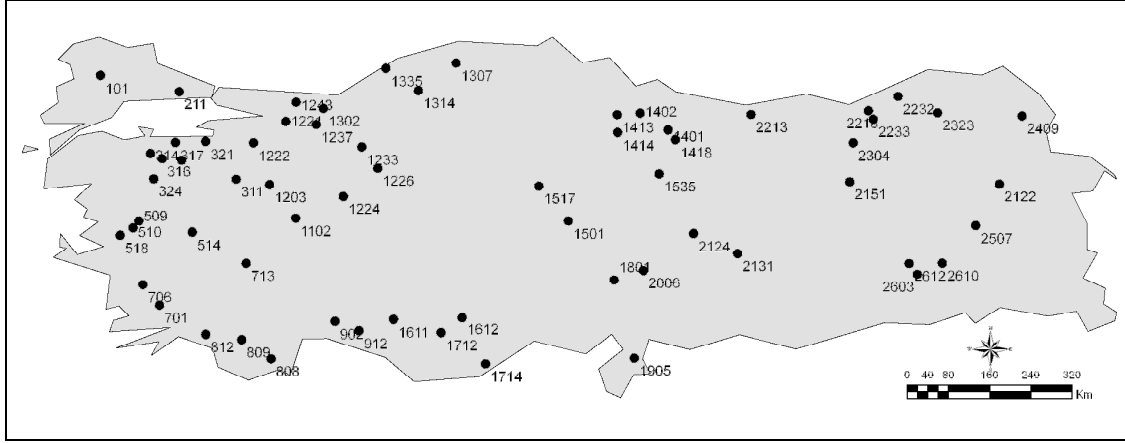
Analizler için 1975-2000 dönemi seçilmiştir.

Öncelikle meteorolojik parametrelere ait veriler incelenmiş ve 1975-2000 dönemi için yağış, sıcaklık, nisbi nem, rüzgar hızı ile ilgili verilerin büyük oranda eksiklik içermediği 218 istasyon belirlenmiştir. Çalışmada kullanılmak üzere seçilen 218 meteoroloji istasyonunun numaraları ve konumları Şekil 4.2’de gösterilmektedir.



Şekil 4.2. Çalışmada kullanılmak üzere seçilen 218 meteoroloji istasyonunun konumları

Çalışmada kullanılan akım gözlem istasyonları belirlenirken veri zaman serilerinin sürekli, doğal ve homojen özellikler taşımasına önem verilmiştir. Doğallık ve homojenlik özelliklerinin değerlendirilmesi için daha önce yapılan çalışmalardan yararlanılmıştır. Kahya ve Karabörk [14], 1964-1994 dönemi için 22 havzaya dağılmış 76 akım gözlem istasyonu verisinde regülasyon etkisinin mevcut olmadığı ve verilerin eksiksiz olduğunu belirlemişlerdir. Bu projede de aynı istasyonlara ait veriler öncelikli olarak alınmıştır. Daha sonra bu istasyonlara ait veriler 1975-2000 dönemi için yeniden gözden geçirilerek eksik veri içeren istasyonlar çıkarılmıştır. Sonuçta şu an için 23 akarsu havzasına dağılmış 64 istasyon belirlenmiştir. Çalışmada kullanılan akım gözlem istasyonlarının numaraları ve konumları Şekil 4.3’de gösterilmektedir.

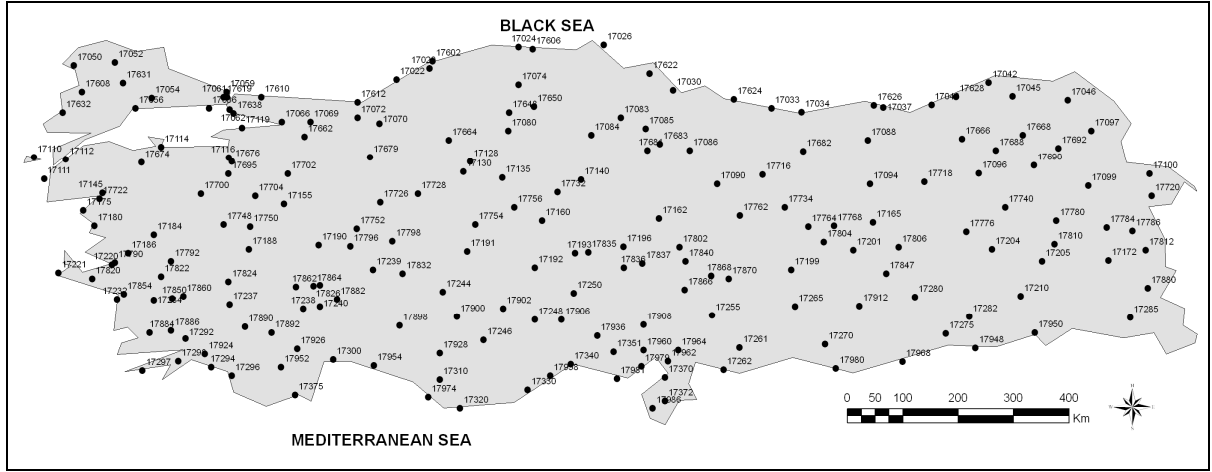


Şekil 4.3. Çalışmada kullanılmak üzere seçilen akım gözlem istasyonlarının konumları

4.1.2.2. Homojenlik analizleri

Çalışmada Devlet Meteoroloji İşleri (DMİ)'ne ait büyük klima istasyonlarında toplanmış günlük veriler kullanılmıştır. Çalışmada analiz edilmek üzere seçilen veriler 218 istasyondan (Şekil 4.2) 213 tanesi seçilmiştir (Şekil 4.4). Bu istasyonlar günlük olarak az miktarda eksik veri içeren ve yıllık olarak eksik veri içermeyen istasyonlardır. Günlük veri serilerinde yer alan eksik verilerin incelenen parametreler için önemli bir miktarda olmadığı için tamamlanmaları için herhangi bir işlem yapılmamıştır.

Günlük verileri yıllık ortalama veya toplam (yağış) değerlere çevrilmiştir. Diğer bir ifadeyle homojenlik analizine tabi tutulan veriler yıllık toplam yağış (mm), yıllık ortalama maksimum sıcaklık ($^{\circ}\text{C}$), yıllık ortalama sıcaklık ($^{\circ}\text{C}$), yıllık ortalama minimum sıcaklık ($^{\circ}\text{C}$), yıllık ortalama rüzgar hızı ($^{\circ}\text{C}$), yıllık ortalama bağıl nem ($^{\circ}\text{C}$) ve yıllık ortalama güneşlenme şiddeti (cal/cm^2)'dir. Veriler rüzgar hızı için 208 istasyonda mevcutken, diğer parametreler için 213 istasyonun tamamında mevcuttur.



Şekil 4.4. Çalışmada kullanılmak üzere seçilen 213 meteoroloji istasyonunun konumları

Analiz edilen 213 istasyon içerisinde Pettitt testi, Standart Normal Homojenlik Testi (SNHT), Buishand range testi ve Von Neumann ratio testi ile homojen olarak bulunan istasyonların sayısı her bir parametre için Tablo 4.1’de sunulmuştur. Tablo 4.2 ise şüpheli, belirsiz ve kullanışlı olarak nitelendirilebilecek istasyon sayılarını her bir parametre için göstermektedir. Yağış parametresi için Pettitt testi ve Buishand range testi; ortalama sıcaklık, maksimum sıcaklık, minimum sıcaklık, bağıl nem ve rüzgar hızı parametreleri için ise Pettitt testi en fazla sayıda istasyona ait verilerin homojen özellik gösterdiği belirlenmiştir. Serilerin ortalarındaki kırılmalara karşı daha duyarlı olan iki test Pettitt testi ve Buishand range testi beklendiği şekilde birbirlerine benzer sonuçlar vermişlerdir. Von Neumann ratio testi, yağış, rüzgar hızı, bağıl nem ve güneşlenme şiddeti parametreleri için ve SNHT ise maksimum, ortalama ve minimum sıcaklık parametreleri için en fazla sayıda homojen olmayan istasyon ortaya çıkarmışlardır.

Tablo 4.1. 1975-2000 dönemi için homojenlik analizi sonuçları (Toplam istasyon sayısı rüzgar hızı için 208, diğer parametreler için 213'tür)

Parametre	TESTLER							
	Pettitt testi		SNHT		Buishand Range testi		Von Neumann ratio testi	
	Homojen İstasyon Sayısı	Homojen İstasyon Oranı (%)	Homojen İstasyon Sayısı	Homojen İstasyon Oranı (%)	Homojen İstasyon Sayısı	Homojen İstasyon Oranı (%)	Homojen İstasyon Sayısı	Homojen İstasyon Oranı (%)
Yağış	208	98	206	97	208	98	204	96
Ortalama Sıcaklık	202	95	192	90	197	93	194	91
Maksimum Sıcaklık	184	87	166	78	182	86	173	81
Minimum Sıcaklık	176	83	164	77	174	82	174	82
Bağıl Nem	94	44	82	39	90	42	40	19
Rüzgar Hızı	39	19	33	16	36	17	9	4

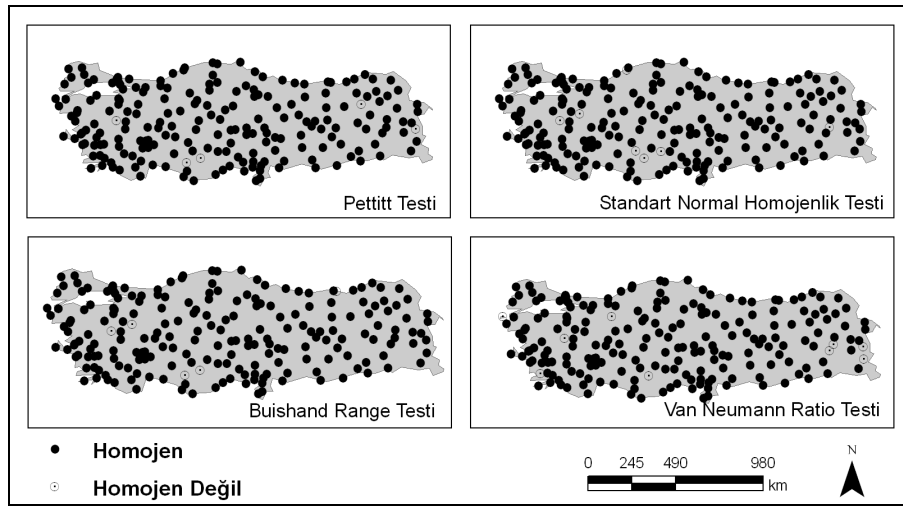
Tablo 4.2'ye göre ise yağış, ortalama sıcaklık, maksimum sıcaklık, minimum sıcaklık parametrelerine ait veriler çoğunlukla 'kullanışlı' olarak nitelendirilebilirken, bağıl nem, rüzgar hızı ve güneşlenme şiddeti parametrelerine ait veriler çoğunlukla 'şüpheli' olarak nitelendirilmektedir. Aşağıda her bir parametre için sonuçlar ayrıntılı olarak açıklanmaktadır.

Tablo 4.2. 1975-2000 dönemi için homojenlik analizi sonuçlarının sınıflandırılması

Parametreler	Şüpheli	Kuşkulu	Kullanışlı	Toplam
Yağış	2	4	207	213
Ortalama Sıcaklık	12	3	198	213
Maksimum Sıcaklık	27	13	173	213
Minimum Sıcaklık	31	12	170	213
Bağıl Nem	120	20	73	213
Rüzgar Hızı	170	12	26	208

Yağış verilerine ait homojenlik analizi sonuçları

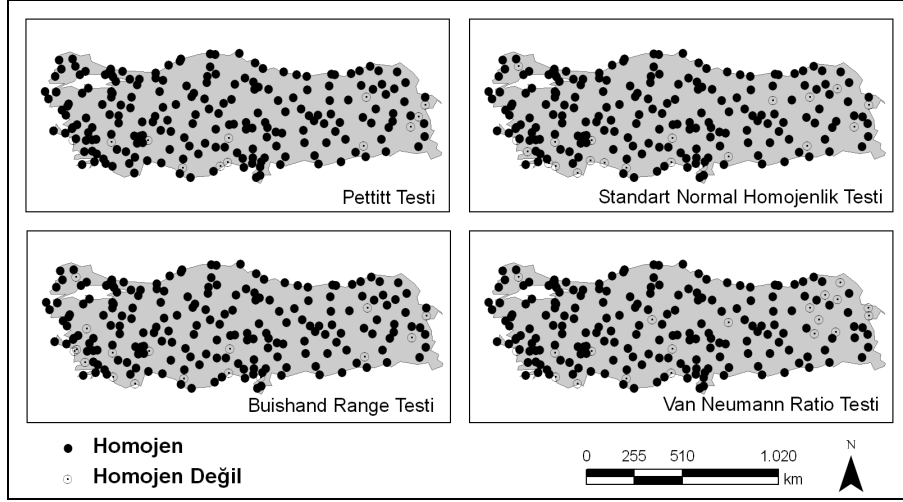
Türkiye yağış verilerinin homojenlik analizlerinde farklı testler ile bulunmuş sonuçlar Şekil 4.5’de gösterilmektedir. Toplam 213 istasyon arasından sadece 2 istasyonun her dört test ile homojen yağış verilerine sahip olmadığı ortaya çıkmıştır. Bu 2 istasyona ait yağış verileri ‘şüpheli’ olarak nitelendirilebilir. 197 istasyonun yağış verilerinin bütün testlerle homojen olduğu belirlenirken, 207 istasyon verilerinin üç veya dört testle ve 211 testin ise iki veya daha fazla testle homojen olduğu ortaya çıkmıştır. Diğer bir ifadeyle ‘kullanışlı’ veya ‘belirsiz’ olarak nitelendirilebilecek 211 istasyon yağış verileri homojenlik testinden olumlu sonuç almış olup, veriler kullanılabilir durumdadır (Tablo 4.2).



Şekil 4.5. Yağış verilerinin homojenlik analizi sonuçları

Ortalama sıcaklık verilerine ait homojenlik analizi sonuçları

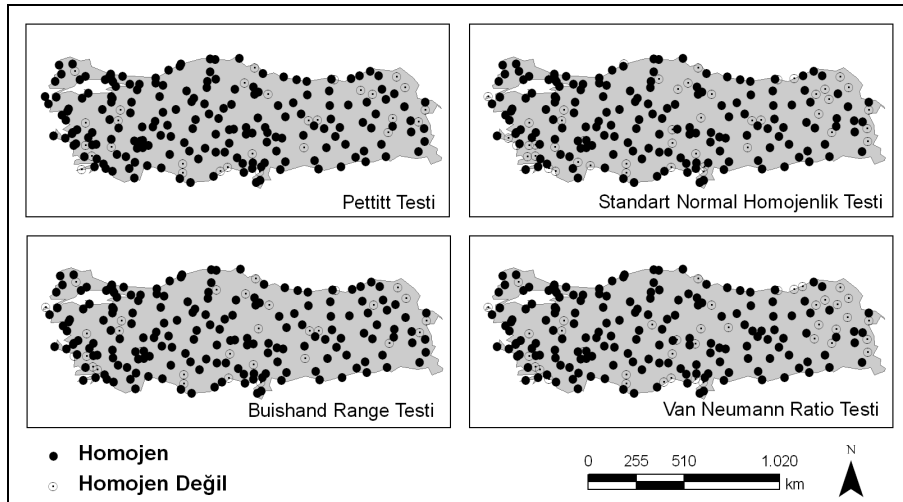
Ortalama sıcaklık verilerinin homojenlik analizi sonuçları Şekil 4.6’da gösterilmiştir. Toplam 213 istasyon arasından 7 istasyona ait verilerin bütün testlerle homojen olmadığı bulunurken, 180 istasyona ait verilerin ise bütün testlerde homojen olduğu ortaya çıkmıştır. 213 istasyon arasında 12 istasyon verisi ‘şüpheli’, 3 istasyon verisi ‘belirsiz’, 198 istasyon ise ‘kullanışlı’ grubuna dahil edilebilir. Böylece 201 istasyon verisinin kullanımının uygun olduğu söylenebilir (Tablo 4.2).



Şekil 4.6. Ortalama sıcaklık verilerinin homojenlik analizi sonuçları

Maksimum sıcaklık verilerine ait homojenlik analizi sonuçları

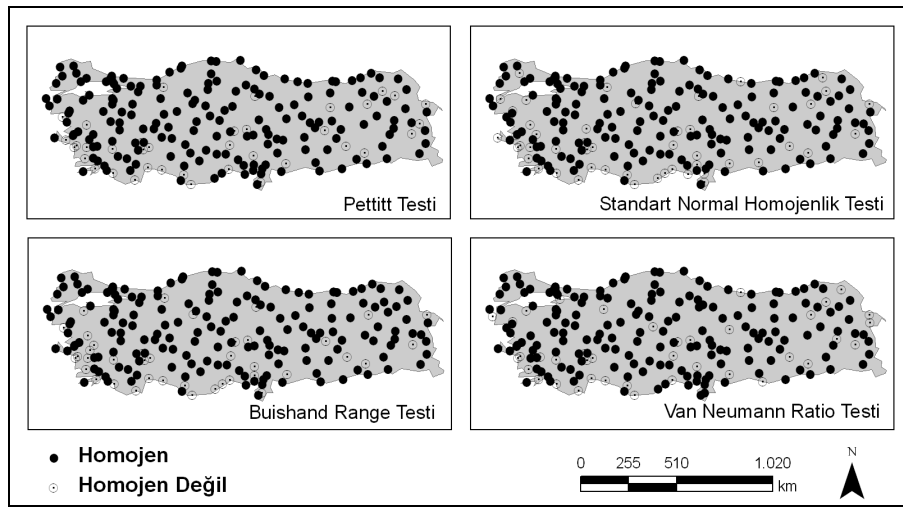
Maksimum sıcaklık verilerinin homojenlik analizi sonuçları Şekil 4.7’de gösterilmiştir. 17 istasyona ait maksimum sıcaklık verilerinin bütün testlerde homojen olmadığı, 150 istasyona ait verilerin ise bütün testlerle homojen olduğu belirlenmiştir. Tablo 4.2’de de görüldüğü üzere 213 istasyona ait maksimum sıcaklık verileri arasından 27 istasyon verileri ‘şüpheli’, 13 istasyon verileri ‘belirsiz’, 173 istasyon verileri ise ‘kullanışlı’ olarak nitelendirilebilir. Bu durumda 186 istasyon verisinin kullanılabilir olduğu söylenebilir (Tablo 4.2).



Şekil 4.7. Maksimum sıcaklık verilerinin homojenlik analizi sonuçları

Minimum sıcaklık verilerine ait homojenlik analizi sonuçları

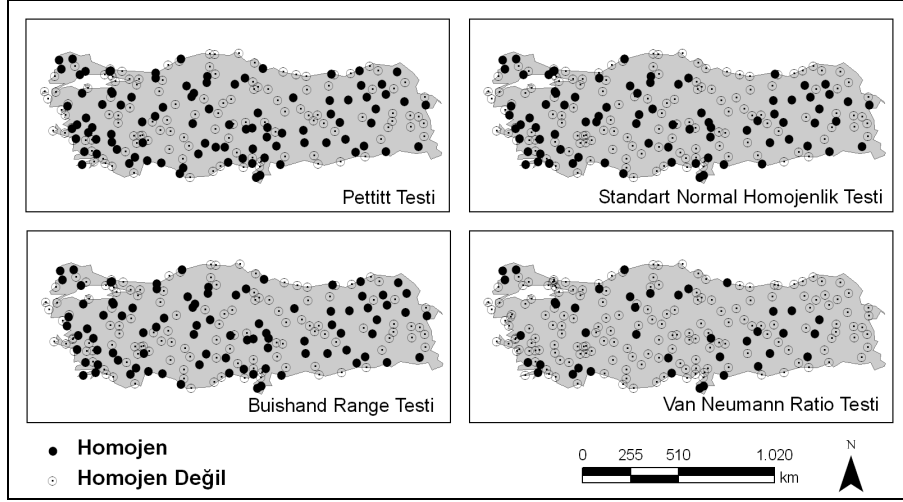
Minimum sıcaklık verilerinin homojenlik analizi sonuçları Şekil 4.8’de gösterilmiştir. Toplam 213 istasyon arasından 24 istasyona ait minimum sıcaklık verilerinin bütün testlerle homojen olmadığı, 147 istasyona ait verilerin ise bütün testlerde homojen olduğu bulunmuştur. 31 istasyona ait minimum sıcaklık verileri ‘şüpheli’, 12 istasyona ait veriler ‘belirsiz’, 170 istasyona ait veriler ise ‘kullanışlı’ grubuna dahil edilebilir. Böylece 182 istasyon verisi kullanılabilir durumda olmaktadır (Tablo 4.2).



Şekil 4.8. Minimum sıcaklık verilerinin homojenlik analizi sonuçları

Bağıl nem verilerine ait homojenlik analizi sonuçları

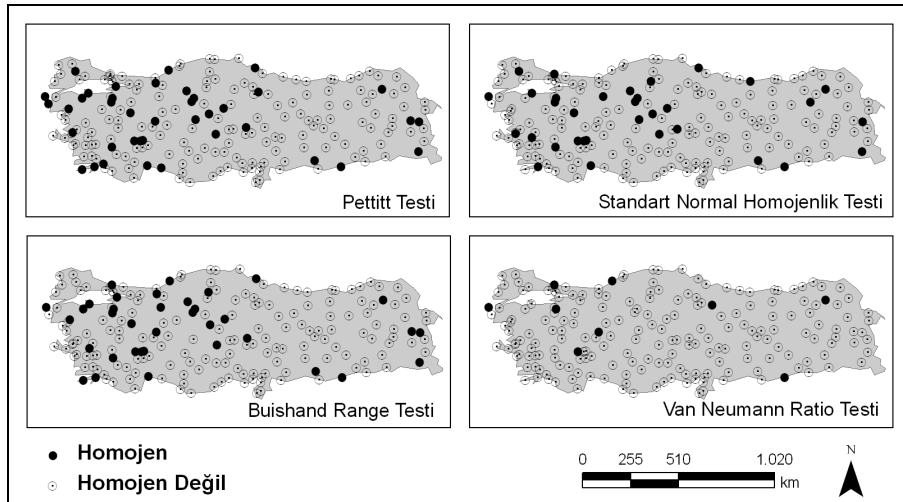
Bağıl nem verilerinin homojenlik analizi sonuçları Şekil 4.9’da gösterilmiştir. Bağıl nem verilerinin 106 istasyonda bütün testlerde homojen olmadığı, 33 istasyonda ise bütün testlerde homojen olduğu belirlenmiştir. 120 istasyona ait veriler ‘şüpheli’, 20 istasyona ait veriler ‘belirsiz’, 73 istasyona ait veriler ise ‘kullanışlı’ olarak nitelendirilebilir. Böylece 93 istasyona ait veriler kullanılabilir durumdadır (Tablo 4.2).



Şekil 4.9. Bağıl nem verilerinin homojenlik analizi sonuçları

Rüzgar hızı verilerine ait homojenlik analizi sonuçları

Rüzgar hızı verilerinin homojenlik analizi sonuçları Şekil 4.10’da gösterilmiştir. 1975-2000 dönemi için eksiksiz veri 208 istasyonda mevcuttur. Rüzgar hızı verilerinin 208 istasyondan 160’ında bütün testlerde homojen olmadığı, 5 istasyonda ise bütün testlerde homojen olduğu belirlenmiştir. 170 istasyona ait veriler ‘şüpheli’, 12 istasyona ait veriler ‘belirsiz’, 26 istasyona ait veriler ise ‘kullanışlı’ olarak nitelendirilebilir. Bu bilgilerden yola çıkarak 38 istasyon verisinin kullanılabilir durumda olduğu söylenebilir. (Tablo 4.2).



Şekil 4.10. Rüzgar hızı verilerinin homojenlik analizi sonuçları

Özet

Bu çalışmada Devlet Meteoroloji İşleri (DMİ)'ye ait 280 klima istasyonu arasından 1975-2000 dönemi için 213 istasyona ait yağış, ortalama sıcaklık, maksimum sıcaklık, minimum sıcaklık ve bağıl nem; 208 istasyona ait rüzgar hızı ve 43 istasyona ait güneşlenme şiddeti serilerinin homojenliğinin tespiti için %1 (0.01) anlamlılık düzeyinde Pettitt testi, Standart normal homojenlik testi, Buishand range testi ve Von Neumann ratio testi uygulanmıştır. Testlerin sonucuna göre seriler homojenlik açısından 'şüpheli', 'belirsiz' ve 'kullanışlı' olarak sınıflandırılmıştır.

Analizler sonucu yağış serileri için 207 adet, ortalama sıcaklık serileri için 198 adet, maksimum sıcaklık serileri için 173 adet, minimum sıcaklık serileri için 170 adet, bağıl nem serileri için 73 adet, rüzgar hızı serileri için 26 adet, güneşlenme şiddeti serileri için 13 adet 'kullanışlı' olarak nitelendirilebilen seri tespit edilmiştir. Bunun yanında yağış serileri için 4 adet, ortalama sıcaklık serileri için 3 adet, maksimum sıcaklık serileri için 13 adet, minimum sıcaklık serileri için 12 adet, bağıl nem serileri için 20 adet, rüzgar hızı serileri için 12 adet, güneşlenme şiddeti serileri için 2 adet seri 'belirsiz' olarak bulunmuştur. Ayrıca yağış serileri için 2 adet, ortalama sıcaklık serileri için 12 adet, maksimum sıcaklık serileri için 27 adet, minimum sıcaklık serileri için 31 adet, bağıl nem serileri için 120 adet, rüzgar hızı serileri için 170 adet ve güneşlenme şiddeti serileri için 28 adet 'şüpheli' nitelikte seri tespit edilmiştir.

Yapılan analizler bu çalışmada kullanılması planlanan ve hidrolojik parametreler (akım) üzerinde en önemli etkiye sahip yağış ve sıcaklık parametrelerine ait verilerin büyük oranda homojen olduğunu göstermiştir. Buna karşılık bağıl nem ve rüzgar hızı parametreleri için homojen seriye sahip istasyon sayısı son derece azdır. Veri serilerinin homojenliği istasyonların yer değişimi, ölçüm yönteminin değişimi veya istasyon çevresinin değişimi gibi insan-kaynaklı faktörlerden etkilenebilir. Bunun yanında veri serisinde yer alan uzun-dönemli değişimler homojen olmama durumunu ortaya çıkarabilir. Konu ile ilgili olarak DMİ ile iletişime geçilmiş ve rüzgar hızı ve bağıl nem açısından homojen nitelik göstermeyen bazı istasyonların metaverilerine ulaşılarak homojen olmama durumunun nedenleri araştırılmıştır. Bu araştırmada incelenen istasyonların metaverilerinde insan-kaynaklı değişikliklerden biri tespit edilmemiştir. Bütün istasyonlar için metaverilerin ulaşılabilir olmaması bu tür analizin bütün istasyon verileri için yapılmasını engellemiştir.

Homojenlik analizleri sonucunda, projenin ilerleyen aşamaları için yağış ve sıcaklık verilerinin homojen özellik gösterdiği istasyonların kullanılması mümkün olacaktır. Bu iki parametre hidrolojik sistem üzerinde doğrudan etkilidir ve homojen istasyon sayısı bir sonraki aşamada yapılan istasyon eşleştirme uygulaması için yeterli düzeydedir. Ancak rüzgar hızı ve bağıl nem gibi parametrelerde homojen seri sayısı çok azdır ve sadece homojen veriye sahip istasyonlara ait verilerin kullanılması birbirleriyle ilişkili olmayan istasyonların eşleştirilmelerine neden olacaktır. Bu sebeplerden dolayı rüzgar hızı ve bağıl nem verilerinin kullanılması durumunda, homojenlik aranmamıştır.

4.2. Veri Önışleme

Projenin II. aşaması veri önışleme aşamasıdır. Bu aşama veri birleşirme, istasyon eşleşirme, veri ayrıklaştırma, mevsimselliğin giderilmesi iş paketlerini kapsamaktadır.

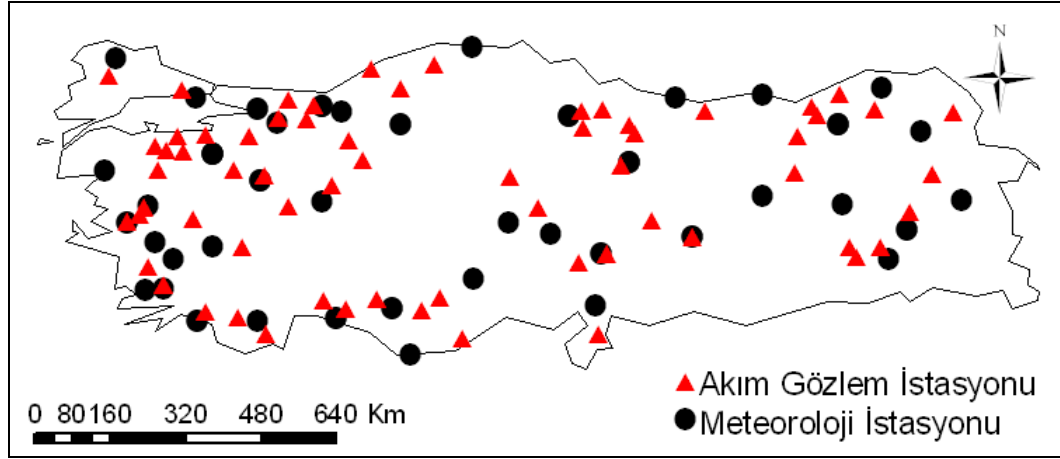
4.2.1. Veri Birleşirme

Veri birleşirme çeşitli parametrelere ait veriler arasındaki period-zaman farklılıklarının giderilmesidir. Verilerin tamamı aylık ortalama veya aylık toplam olarak ifade edilmiştir.

4.2.2. İstasyon Eşleşirme

Hidrolojik ve iklim parametrelerine ait veriler birbirinden bağımsız olarak ve farklı kurumlarca toplanmaktadır. Bu nedenle hidrolojik ve iklim verisi sağlayan istasyonların eşleştirilmesi gerekmektedir. Bu konuda yapılan çalışmalar ve elde edilen sonuçlar aşağıda açıklanmaktadır.

Akım ve meteoroloji istasyonlarının Tablo 4.3’de gösterildiği biçimde eşleştirilmiştir. Akım istasyonların ve eşleştirildikleri meteoroloji istasyonlarının konumları Şekil 4.11’de gösterilmektedir.



Şekil 4.11. Eşleştirilen akım gözlem ve meteoroloji istasyonlarının konumları

Tablo 4.3. Eşleştirilen akım gözlem ve meteoroloji istasyonları

Akım İst. No	Meteoroloji İst. No	Akım İst. No	Meteoroloji İst. No	Akım İst. No	Meteoroloji İst. No
101	17050	1203	17155	1712	17928
211	17636	1221	17066	1714	17320
302	17695	1222	17119	1801	17866
311	17155	1224	17726	1905	17964
314	17674	1226	17664	2006	17866
316	17695	1233	17072	2122	17784
317	17695	1237	17072	2124	17199
321	17695	1243	17066	2131	17764
324	17184	1302	17072	2151	17806
509	17186	1307	17606	2213	17033
510	17186	1314	17602	2218	17626
514	17792	1335	17070	2232	17045
518	17822	1401	17716	2233	17628
701	17884	1402	17085	2304	17666
706	17884	1413	17085	2323	17666
713	17824	1414	17085	2409	17692
808	17952	1418	17716	2507	17205
809	17294	1501	17196	2603	17776
812	17294	1517	17193	2610	17205
902	17300	1535	17716	2612	17776
912	17954	1611	17928		
1102	17155	1612	17248		

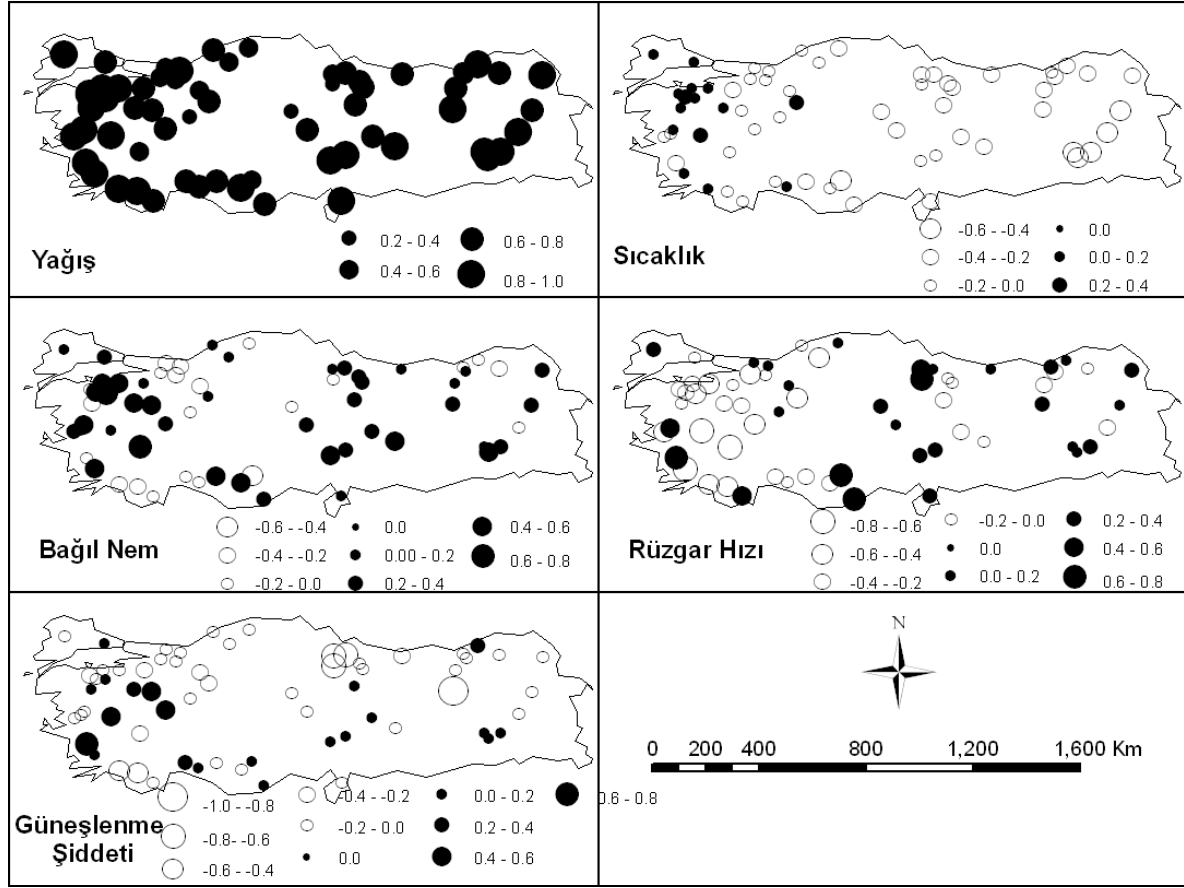
Akım ve meteorolojik parametreler arasındaki ilişkileri belirlemek için Türkiye bazında 1976-2000 (su yılı) dönemlerindeki daha önce belirlenmiş olan 64 akım istasyonu ile bu istasyonlarla eşleştirilmiş meteoroloji istasyonlarına ait 1976-2000 dönemleri için meteorolojik parametre değerleri kullanılmıştır.

Korelasyon analizinin ilk aşamasında akım ile aynı yıla ait toplam yağış, ortalama sıcaklık, ortalama bağıl nem, ortalama rüzgar hızı ve ortalama güneşlenme şiddeti verileri arasındaki korelasyon katsayıları hesaplanmıştır (Tablo 4.4). Analizler sonucunda en yüksek ilişki beklendiği gibi akım ile yağış arasında tespit edilmiştir. Akım ve yağış arasındaki ilişki pozitifdir ve korelasyon katsayıları 0.26 ile 0.94 arasında değişmiş ve 64 istasyon için ortalama 0.71 olarak gerçekleşmiştir. Yağıştan sonra akım ile en güçlü ilişki sıcaklık ve bağıl nem parametreleri arasında bulunmuştur. Akım ile ortalama sıcaklık arasındaki ilişki ortalama olarak negatif (-0.14), akım ile bağıl nem arasındaki ilişki ise ortalama olarak pozitifdir (0.15). Her iki parametre için de değerler pozitif ve negatif olmak üzere geniş bir aralıkta değişmektedir. Özellikle bağıl nem için 64 istasyon için değişkenlik daha yüksektir. Buna rağmen sıcaklıkla akım arasındaki ilişki istasyonların çoğunluğunda negatif, akım ile bağıl nem arasındaki ilişki ise istasyonların çoğunluğunda pozitif olarak bulunmuştur. Akım ile rüzgar hızı arasındaki ilişkiler diğer parametrelere göre daha zayıftır. Bu sonuçlar akımı etkileyen en önemli parametrenin yağış, yağıştan sonra ise sıcaklık ve bağıl nem olduğunu göstermiştir. Daha önce yapılan çalışmalarda da [ör., 48] akım ve meteorolojik parametreler arasında benzer ilişkiler elde edilmiştir.

Akım ve meteorolojik parametreler arasındaki ilişkiler mekansal anlamda da farklılıklar göstermektedir (Şekil 4.12). Yağış ve akım arasındaki ilişkiler genel olarak Ege Denizi ve Akdeniz kıyıları, Trakya ve Güneydoğu Anadolu Bölgeleri'nde daha güçlü, ancak iç kesimlerde zayıftır. Akım ve sıcaklık arasındaki ilişkiler Türkiye'nin doğu bölgelerinde daha güçlüdür. Ayrıca beklenen bir durum olmasa da, akım ve sıcaklık arasındaki ilişkiler batı bölgelerinde zayıf derecede pozitif çıkmıştır. Bağıl nem ve rüzgar hızı ile akım arasındaki mekansal ilişkiler daha az belirgindir.

Tablo 4.4. Akım ve yağış, sıcaklık, bağıl nem, rüzgar hızı ve güneşlenme şiddeti arasındaki korelasyon katsayıları

Akım İstasyonu	İstasyon No	Yağış	Sıcaklık	Bağıl Nem	Rüzgar Hızı	Akım İstasyonu	İstasyon No	Yağış	Sıcaklık	Bağıl Nem	Rüzgar Hızı
101	17052	0.82	0.04	0.08	0.25	1314	17664	0.56	-0.12	0.05	-0.54
211	17636	0.76	0.15	0.33	-0.11	1335	17070	0.69	-0.13	0.15	-0.08
302	17695	0.83	0.07	0.61	-0.41	1401	17090	0.62	-0.23	0.21	-0.11
311	17155	0.80	0.01	0.46	-0.26	1402	17085	0.66	-0.26	0.26	0.07
314	17145	0.82	0.03	-0.08	-0.34	1413	17085	0.55	-0.24	0.09	0.48
316	17695	0.85	0.22	0.51	-0.39	1414	17085	0.33	-0.18	-0.02	0.67
317	17695	0.87	0.16	0.55	-0.39	1418	17090	0.64	-0.25	0.21	-0.15
321	17695	0.84	0.07	0.58	-0.40	1501	17836	0.67	-0.25	0.32	0.02
324	17184	0.84	0.08	-0.21	-0.11	1517	17193	0.37	-0.29	-0.18	0.36
509	17186	0.88	0.01	0.46	0.17	1535	17090	0.68	-0.28	0.28	-0.28
510	17186	0.79	-0.03	0.33	0.42	1611	17928	0.80	-0.24	0.49	-0.29
514	17860	0.81	0.38	0.11	-0.75	1612	17248	0.48	-0.53	-0.45	0.64
518	17822	0.81	0.00	0.30	-0.48	1712	17928	0.81	-0.19	0.58	-0.39
701	17886	0.94	0.07	0.53	-0.72	1714	17320	0.72	-0.26	0.32	0.76
706	17884	0.81	-0.29	-0.02	0.71	1801	17866	0.87	-0.09	0.46	0.32
713	17824	0.55	-0.07	0.78	-0.77	1905	17964	0.81	-0.29	0.10	0.22
808	17952	0.62	-0.13	-0.04	0.60	2006	17866	0.90	-0.06	0.40	0.31
809	17294	0.83	-0.01	-0.23	-0.57	2122	17784	0.79	-0.40	0.24	0.10
812	17294	0.87	0.04	-0.20	-0.50	2124	17199	0.69	-0.22	0.25	-0.22
902	17954	0.76	-0.01	-0.10	-0.26	2131	17199	0.83	-0.32	0.47	-0.13
912	17954	0.79	0.20	-0.08	-0.12	2151	17165	0.81	-0.32	0.30	0.35
1102	17155	0.67	0.00	0.37	-0.49	2213	17033	0.70	-0.29	0.09	0.02
1203	17155	0.73	-0.06	0.45	-0.24	2218	17626	0.44	-0.18	-0.17	0.24
1221	17662	0.44	-0.06	-0.03	-0.55	2232	17045	0.81	-0.22	-0.09	0.10
1222	17066	0.74	-0.29	0.04	-0.02	2233	17666	0.35	-0.20	0.03	-0.20
1224	17752	0.26	-0.08	-0.12	0.05	2304	17666	0.73	-0.34	0.09	-0.25
1226	17664	0.71	0.26	0.02	-0.55	2323	17666	0.68	-0.36	-0.20	-0.02
1233	17072	0.54	-0.03	-0.33	0.09	2409	17692	0.81	-0.33	0.32	0.28
1237	17072	0.58	-0.17	-0.28	-0.01	2507	17205	0.84	-0.55	0.00	-0.35
1243	17066	0.50	-0.18	-0.29	0.19	2603	17210	0.88	-0.48	0.41	0.05
1302	17072	0.83	-0.16	-0.23	0.02	2610	17205	0.88	-0.53	0.39	0.21
1307	17606	0.51	-0.23	-0.02	0.17	2612	17776	0.82	-0.42	0.16	0.03



Şekil 4.12. Akım ve yağış, sıcaklık, bağıl nem, rüzgar hızı ve güneşlenme şiddeti arasındaki korelasyon katsayılarının konuma bağlı aldıkları değerler

Korelasyon analizinin ikinci aşamasında herhangi bir yıldaki akım değerleri ile geçmiş yıllara (0-5 yıl) ait yağış değerleri arasındaki ilişkiler incelenmiştir. Öncelikle her yılın yağış değeri ile akım arasındaki (Tablo 4.5), daha sonra ise son 1,2,3,4,5 yıllık kümülatif yağışlar ve akım arasındaki korelasyonlar (Tablo 4.6) incelenmiştir. Tablo 4.5'e göre akım ile aynı yılda gerçekleşen yağış arasındaki ilişkinin daha güçlüdür (ortalama 0.71). 64 istasyonun 59 tanesinde en güçlü ilişkinin aynı yıl gerçekleşen yağışla olduğu tespit edilmiştir. Akım ve yağış arasındaki ilişki geçmişe doğru ilerledikçe genel olarak azalmaktadır. Örneğin akım ile bir önceki yıldaki yağış arasındaki ortalama korelasyon katsayısı 0.15'e gerilemektedir.

Tablo 4.6'da verilen kümülatif yağış değerleri (1, 2, 3, 4, 5 yıllık toplam) ile akım arasındaki korelasyon katsayıları da benzer bir sonuç göstermektedir. Yağış ile akım arasındaki en güçlü ilişki bütün istasyonlar için akımın gerçekleştiği yıldaki yağış ile gerçekleşmiştir. Bu da akımın geçmiş yıllardaki toplam yağışlardan çok son 12 aylık yağışlardan en fazla etkilendiğini göstermektedir.

Tablo 4.5. Akım ve geçmiş yıllara ait yağış değerleri arasındaki korelasyon katsayıları

Akım İst. No	Meteoroloji İst. No	Yağış (t)	Yağış (t-1)	Yağış (t-2)	Yağış (t-3)	Yağış (t-4)	Yağış (t-5)	Akım İst. No	Meteoroloji İst. No	Yağış (t)	Yağış (t-1)	Yağış (t-2)	Yağış (t-3)	Yağış (t-4)	Yağış (t-5)
101	17052	0.82	0.31	0.26	0.39	0.33	-0.13	1314	17664	0.56	0.08	0.32	0.11	-0.34	0.10
211	17636	0.76	0.37	0.06	0.04	-0.25	-0.24	1335	17070	0.69	0.15	0.29	0.03	-0.44	0.15
302	17695	0.83	0.07	0.22	0.30	0.14	-0.04	1401	17090	0.62	-0.28	0.05	-0.21	-0.36	0.12
311	17155	0.80	0.35	0.39	0.60	0.28	0.16	1402	17085	0.66	-0.29	0.00	0.15	-0.10	-0.03
314	17145	0.82	-0.14	-0.04	-0.23	0.16	-0.06	1413	17085	0.55	-0.25	-0.13	0.18	-0.14	0.12
316	17695	0.85	0.11	0.16	0.24	0.08	-0.10	1414	17085	0.33	-0.18	-0.11	-0.01	0.00	0.17
317	17695	0.87	0.11	0.16	0.23	0.04	-0.11	1418	17090	0.64	-0.44	0.04	-0.07	-0.46	0.19
321	17695	0.84	0.23	0.29	0.17	0.18	-0.15	1501	17836	0.67	-0.03	-0.08	-0.01	-0.30	0.34
324	17184	0.84	0.22	0.21	0.34	0.41	0.04	1517	17193	0.37	0.10	0.00	0.16	-0.18	-0.13
509	17186	0.88	0.33	0.11	0.38	0.45	0.04	1535	17090	0.68	-0.37	-0.06	-0.03	-0.47	0.32
510	17186	0.79	0.39	0.19	0.47	0.56	0.13	1611	17928	0.80	0.19	0.06	-0.04	0.15	0.19
514	17860	0.81	0.34	0.50	0.48	0.46	0.19	1612	17248	0.48	0.41	0.34	-0.17	0.35	0.37
518	17822	0.81	0.50	0.45	0.69	0.49	0.29	1712	17928	0.81	0.36	0.09	0.13	0.21	0.23
701	17886	0.94	0.58	0.54	0.46	0.26	0.00	1714	17320	0.72	0.45	0.51	0.74	0.47	0.26
706	17884	0.81	0.61	0.58	0.56	0.44	0.24	1801	17866	0.87	0.33	-0.17	-0.02	0.10	-0.02
713	17824	0.55	0.41	0.51	0.62	0.52	0.35	1905	17964	0.81	0.35	-0.01	-0.36	-0.22	-0.09
808	17952	0.62	0.51	0.47	0.55	0.46	0.30	2006	17866	0.90	0.37	-0.10	-0.02	0.19	-0.06
809	17294	0.83	0.52	0.57	0.53	0.58	0.18	2122	17784	0.79	0.03	-0.04	0.21	-0.02	0.02
812	17294	0.87	0.59	0.65	0.54	0.58	0.18	2124	17199	0.69	0.27	-0.27	-0.16	0.14	0.21
902	17954	0.76	0.07	-0.03	0.30	-0.11	-0.58	2131	17199	0.83	0.14	-0.32	-0.12	-0.03	0.34
912	17954	0.79	0.13	0.00	0.34	0.02	-0.53	2151	17165	0.81	-0.05	-0.02	0.04	-0.11	-0.07
1102	17155	0.67	0.09	0.21	0.62	0.22	0.09	2213	17033	0.70	0.03	-0.16	-0.22	-0.09	0.09
1203	17155	0.73	0.41	0.27	0.64	0.40	0.24	2218	17626	0.44	-0.05	0.02	0.46	0.22	-0.05
1221	17662	0.44	0.22	0.09	0.52	-0.02	0.15	2232	17045	0.81	0.08	0.05	0.10	0.12	-0.03
1222	17066	0.74	0.01	-0.20	-0.05	-0.29	-0.04	2233	17666	0.35	0.29	0.21	0.00	-0.01	0.27
1224	17752	0.26	0.11	0.27	0.26	0.28	0.40	2304	17666	0.73	0.03	0.07	-0.21	0.30	0.09
1226	17664	0.71	0.37	0.11	0.06	-0.16	-0.33	2323	17666	0.68	0.07	0.09	-0.05	0.20	-0.02
1233	17072	0.54	0.20	-0.15	0.23	-0.32	-0.12	2409	17692	0.81	0.16	0.02	-0.03	0.30	-0.11
1237	17072	0.58	0.25	0.07	0.21	-0.39	-0.07	2507	17205	0.84	0.30	0.14	0.07	-0.12	0.40
1243	17066	0.50	0.10	-0.18	-0.02	-0.36	-0.02	2603	17210	0.88	-0.23	0.11	-0.01	-0.38	0.06
1302	17072	0.83	0.02	0.14	0.15	-0.57	0.22	2610	17205	0.88	0.04	0.01	-0.04	-0.22	0.12
1307	17606	0.51	-0.58	0.20	-0.34	0.11	0.18	2612	17776	0.82	-0.40	0.03	-0.04	-0.08	0.08

Tablo 4.6. Akım ve geçmiş yıllara ait kümülatif yağış değerleri (1 ,2, 3, 4, 5 yıllık toplam) arasındaki korelasyon katsayıları

Akım İst. No	Meteoroloji İst. No	Yağış (1 yıl)	Yağış (2 yıl)	Yağış (3 yıl)	Yağış (4 yıl)	Yağış (5 yıl)	Akım İst. No	Meteoroloji İst. No	Yağış (1 yıl)	Yağış (2 yıl)	Yağış (3 yıl)	Yağış (4 yıl)	Yağış (5 yıl)
101	17052	0.82	0.65	0.50	0.41	0.32	1314	17664	0.56	0.40	0.46	0.41	0.31
211	17636	0.76	0.61	0.43	0.34	0.27	1335	17070	0.69	0.55	0.54	0.49	0.41
302	17695	0.83	0.44	0.32	0.18	0.09	1401	17090	0.62	0.17	0.12	0.01	-0.02
311	17155	0.80	0.57	0.49	0.38	0.23	1402	17085	0.66	0.29	0.19	0.19	0.18
314	17145	0.82	0.53	0.39	0.18	0.16	1413	17085	0.55	0.17	-0.01	-0.02	-0.04
316	17695	0.85	0.52	0.39	0.24	0.14	1414	17085	0.33	0.01	-0.18	-0.23	-0.26
317	17695	0.87	0.53	0.39	0.23	0.12	1418	17090	0.64	0.13	0.00	-0.05	-0.04
321	17695	0.84	0.67	0.60	0.50	0.45	1501	17836	0.67	0.31	0.11	0.06	0.03
324	17184	0.84	0.65	0.55	0.43	0.37	1517	17193	0.37	0.24	0.09	0.06	0.03
509	17186	0.88	0.71	0.54	0.42	0.34	1535	17090	0.68	0.14	0.04	0.04	0.04
510	17186	0.79	0.67	0.49	0.37	0.29	1611	17928	0.80	0.53	0.21	0.01	-0.03
514	17860	0.81	0.56	0.49	0.33	0.21	1612	17248	0.48	0.43	0.21	-0.13	-0.20
518	17822	0.81	0.74	0.61	0.49	0.34	1712	17928	0.81	0.52	0.13	-0.09	-0.17
701	17886	0.94	0.78	0.68	0.57	0.47	1714	17320	0.72	0.51	0.37	0.28	0.18
706	17884	0.81	0.73	0.64	0.54	0.43	1801	17866	0.87	0.66	0.29	0.11	0.03
713	17824	0.55	0.38	0.29	0.23	0.14	1905	17964	0.81	0.61	0.28	-0.07	-0.22
808	17952	0.62	0.59	0.55	0.49	0.41	2006	17866	0.90	0.65	0.22	0.01	-0.08
809	17294	0.83	0.71	0.62	0.53	0.44	2122	17784	0.79	0.40	0.18	0.05	-0.04
812	17294	0.87	0.81	0.75	0.66	0.57	2124	17199	0.69	0.64	0.28	0.11	0.09
902	17954	0.76	0.54	0.36	0.28	0.15	2131	17199	0.83	0.51	0.07	-0.08	-0.11
912	17954	0.79	0.64	0.44	0.37	0.27	2151	17165	0.81	0.40	0.17	0.04	-0.06
1102	17155	0.67	0.34	0.28	0.27	0.18	2213	17033	0.70	0.40	0.16	0.03	0.01
1203	17155	0.73	0.53	0.40	0.32	0.19	2218	17626	0.44	0.22	0.17	0.19	0.17
1221	17662	0.44	0.32	0.21	0.22	0.09	2232	17045	0.81	0.40	0.28	0.22	0.19
1222	17066	0.74	0.47	0.32	0.24	0.17	2233	17666	0.35	0.34	0.27	0.14	0.06
1224	17752	0.26	0.13	0.13	0.04	-0.04	2304	17666	0.73	0.27	0.11	-0.02	-0.01
1226	17664	0.71	0.60	0.50	0.45	0.37	2323	17666	0.68	0.26	0.10	0.00	-0.01
1233	17072	0.54	0.49	0.35	0.31	0.18	2409	17692	0.81	0.41	0.14	-0.02	-0.05
1237	17072	0.58	0.53	0.42	0.40	0.31	2507	17205	0.84	0.72	0.62	0.55	0.49
1243	17066	0.50	0.33	0.19	0.14	0.03	2603	17210	0.88	0.35	0.20	0.06	-0.06
1302	17072	0.83	0.61	0.53	0.50	0.40	2610	17205	0.88	0.45	0.24	0.08	-0.03
1307	17606	0.51	-0.05	0.11	0.10	0.15	2612	17776	0.82	0.16	0.03	-0.09	-0.14

Akım-yağış ilişkisine benzer şekilde akım-sıcaklık ilişkisi geçmişe (0-5 yıl) yönelik olarak incelenmiştir. Sıcaklık ve akım arasında ters orantı beklendiği için en güçlü negatif korelasyonlar dikkate alınmıştır (Tablo 4.7). Tablo 4.7’de görüldüğü üzere akım değerlerinin en güçlü ilişkiyi beş yıl önceki sıcaklık değerleriyle gösterdiği 22 istasyon, dört yıl önceki sıcaklıkla gösterdiği 6 istasyon, üç yıl önceki sıcaklıkla gösterdiği 4 istasyon, iki yıl önceki sıcaklıkla gösterdiği 15 istasyon, bir yıl önceki sıcaklıkla gösterdiği 5 istasyon, aynı yıldaki sıcaklıkla gösterdiği 17 istasyon bulunmaktadır.

Tablo 4.7. Akım ve geçmiş yıllara ait sıcaklık değerleri arasındaki korelasyon katsayıları

Akım	Meteoroloji	Sıcaklık	Sıcaklık	Sıcaklık	Sıcaklık	Sıcaklık	Sıcaklık	Akım	Meteorol	Sıcaklık	Sıcaklık	Sıcaklık	Sıcaklık	Sıcaklık	Sıcaklık
101	17052	0.04	0.15	-0.32	0.00	0.23	0.30	1314	17664	-0.12	0.21	0.42	0.18	0.22	-0.25
211	17636	0.15	0.23	-0.29	0.29	0.29	0.05	1335	17070	-0.13	0.16	0.25	0.31	0.42	-0.40
302	17695	0.07	0.27	-0.01	0.11	-0.07	0.19	1401	17090	-0.23	-0.02	0.23	0.07	0.04	-0.35
311	17155	0.01	0.23	-0.10	0.03	-0.10	0.09	1402	17085	-0.26	-0.16	0.35	0.03	0.02	-0.13
314	17145	0.03	0.07	-0.31	0.27	0.15	-0.04	1413	17085	-0.24	-0.25	0.31	0.00	0.03	-0.03
316	17695	0.22	0.26	-0.11	0.06	0.03	0.22	1414	17085	-0.18	-0.31	0.27	-0.02	0.09	-0.01
317	17695	0.16	0.27	-0.10	0.11	0.02	0.15	1418	17090	-0.25	-0.22	0.31	0.12	0.19	-0.21
321	17695	0.07	0.31	0.07	0.12	-0.01	0.12	1501	17836	-0.25	-0.21	0.41	-0.01	0.20	-0.22
324	17184	0.08	0.00	-0.32	0.06	0.04	0.01	1517	17193	-0.29	-0.33	0.16	-0.16	-0.01	0.05
509	17186	0.01	0.08	-0.23	-0.05	-0.10	0.15	1535	17090	-0.28	-0.12	0.36	0.05	0.18	-0.26
510	17186	-0.03	-0.04	-0.30	-0.06	-0.09	-0.04	1611	17928	-0.24	0.18	0.14	-0.05	-0.01	-0.14
514	17860	0.38	0.37	0.17	0.30	0.20	0.19	1612	17248	-0.53	-0.31	0.06	0.12	-0.27	-0.34
518	17822	0.00	-0.01	-0.18	-0.11	-0.05	-0.03	1712	17928	-0.19	0.13	0.16	-0.02	0.02	-0.15
701	17886	0.07	-0.21	-0.25	0.00	-0.01	0.02	1714	17320	-0.26	-0.12	-0.14	-0.29	-0.14	-0.19
706	17884	-0.29	-0.35	-0.35	-0.27	-0.33	-0.21	1801	17866	-0.09	0.15	0.35	0.02	-0.41	-0.40
713	17824	-0.07	-0.07	-0.04	0.00	-0.02	0.00	1905	17964	-0.29	0.06	0.08	0.25	-0.42	-0.48
808	17952	-0.13	-0.13	-0.13	0.07	0.01	-0.10	2006	17866	-0.06	0.03	0.39	-0.07	-0.42	-0.45
809	17294	-0.01	-0.02	-0.18	-0.17	0.07	-0.07	2122	17784	-0.40	-0.15	-0.09	-0.04	-0.01	-0.45
812	17294	0.04	-0.01	-0.19	-0.09	0.10	0.00	2124	17199	-0.22	-0.07	0.30	0.17	-0.11	-0.38
902	17954	-0.01	0.06	0.03	0.21	-0.04	-0.19	2131	17199	-0.32	-0.11	0.27	0.04	-0.17	-0.59
912	17954	0.20	0.08	0.10	-0.09	0.06	0.09	2151	17165	-0.32	0.10	0.15	0.06	-0.12	-0.43
1102	17155	0.00	0.07	-0.01	-0.08	0.14	0.18	2213	17033	-0.29	0.12	0.09	-0.17	-0.10	-0.09
1203	17155	-0.06	0.15	-0.05	-0.04	-0.06	0.01	2218	17626	-0.18	-0.24	-0.16	-0.37	0.11	-0.11
1221	17662	-0.06	0.21	0.32	0.21	0.42	-0.09	2232	17045	-0.22	-0.14	0.12	0.06	-0.07	-0.11
1222	17066	-0.29	0.35	0.09	0.23	0.01	-0.20	2233	17666	-0.20	-0.20	0.30	-0.11	0.06	-0.37
1224	17752	-0.08	-0.07	-0.02	-0.01	0.00	0.05	2304	17666	-0.34	0.01	0.05	0.01	-0.28	-0.38
1226	17664	0.26	0.20	0.47	0.21	0.15	-0.06	2323	17666	-0.36	-0.30	0.02	-0.05	-0.07	-0.08
1233	17072	-0.03	0.27	0.15	0.17	0.23	-0.23	2409	17692	-0.33	-0.03	-0.07	-0.12	-0.37	-0.11
1237	17072	-0.17	0.16	0.19	0.45	0.13	-0.32	2507	17205	-0.55	-0.28	0.01	-0.04	-0.03	-0.58
1243	17066	-0.18	0.15	0.17	0.14	0.23	-0.23	2603	17210	-0.48	0.02	-0.07	-0.13	0.03	-0.42
1302	17072	-0.16	0.24	0.34	0.34	0.03	-0.54	2610	17205	-0.53	-0.06	-0.06	-0.01	-0.04	-0.53
1307	17606	-0.23	0.23	0.03	0.45	-0.05	-0.24	2612	17776	-0.42	0.11	-0.01	-0.02	0.04	-0.63

Tablo 4.8’de akım ve meteorolojik parametreler arasındaki ilişkiler 64 istasyon için elde edilen sonuçların maksimum, minimum, ortalama ve standart sapma değerlerine göre özetlenmektedir. Tablo 4.9’da ise kümülatif yağış verilerinin akım ile korelasyonlarının maksimum, minimum, ortalama ve standart sapma değerleri özetlenmektedir.

Tablo 4.8. Yağış, ortalama sıcaklık, ort. nem, rüzgar hızı ve güneşlenme şiddeti verilerinin maksimum, minimum, ortalama ve standart sapma değerleri

Parametreler	Zaman	Minimu	Maksimu	Standart Sapma	Ortalama
Yağış	t	0.26	0.94	0.16	0.71
	t-1	-0.58	0.61	0.26	0.15
	t-2	-0.32	0.65	0.22	0.13
	t-3	-0.36	0.74	0.28	0.17
	t-4	-0.57	0.58	0.31	0.06
	t-5	-0.58	0.40	0.20	0.07
Sıcaklık	t	-0.55	0.38	0.20	-0.14
	t-1	-0.35	0.37	0.19	0.02
	t-2	-0.35	0.47	0.22	0.06
	t-3	-0.37	0.45	0.16	0.04
	t-4	-0.42	0.42	0.18	0.00
	t-5	-0.63	0.30	0.23	-0.16
Bağıl Nem	t	-0.45	0.78	0.28	0.15
Rüzgar Hızı	t	-0.77	0.76	0.37	-0.06
Güneşlenme Şiddeti	t	-0.84	0.70	0.32	-0.05

Tablo 4.9. Kümülatif yağış değerleri ile akım arasındaki korelasyonların maksimum, minimum, ortalama ve standart sapma değerleri

Kümülatif Yağış Süresi	Minimu	Maksimu	Ortalama	Standart Sapma
1	0.26	0.94	0.16	0.71
2	-0.05	0.81	0.20	0.46
3	-0.18	0.75	0.20	0.31
4	-0.23	0.66	0.21	0.21
5	-0.26	0.57	0.20	0.14

Özet

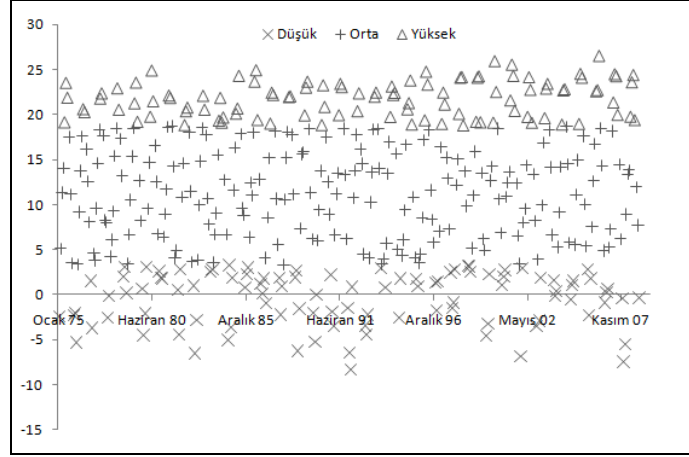
Korelasyon analizleri sonucunda her akım istasyonu ile bir meteoroloji istasyonu eşleştirilmiştir. Analizlerde hem akım hem de göl seviyesi verileri üzerinde en çok etkili meteorolojik parametrenin yağış olduğu tespit edilmiştir. Akım verileri üzerinde sıcaklık ve bağıl nem yağıştan sonra daha etkiliyken, göl seviyeleri üzerinde bağıl nem yağıştan sonra en etkili parametre olarak tespit edilmiştir. Geçmiş yıllara ait meteorolojik durumun akım ve göl seviyeleri üzerindeki etkilerini araştırmak için yapılan analizlerde, akım verileri üzerinde geçmişe dönük etkinin daha düşük olduğu, buna karşılık göl seviyelerinin geçmişteki meteorolojik koşullardan (özellikle yağıştan) büyük oranda etkilendiği bulunmuştur.

4.2.3. Veri Ayırıklaştırma

Veri madenciliği yaklaşım ve algoritmaları hem işlem maliyetini arttırdığı için hem de olası ihtimal sayısının çok yüksek olabileceği nedeniyle çoğunlukla gerçek sayılar üzerinde çalışamazlar. Bu nedenle verilerin sınıflandırılması yanı ayrılaştırılması yoluna gidilir. Verilerin ayrılaştırılması verilerin gruplara ayrılmasıyla parametrelerin almış oldukları değerleri sınıflara ayırmaktır. Bu proje kapsamında veri ayrılaştırmak için kullanılabilecek üç yöntem (yüzdelik dilimler yöntemi, istatistiksel yöntemi, kümeleme yöntemi) belirlenmiştir. Bu üç yöntemin uygulanabilirliği bir meteoroloji istasyonundan (17836 no'lu) elde edilen aylık ortalama sıcaklık verileri üzerinde deneyler yapılarak araştırılmıştır ve sonuçlar karşılaştırılmıştır. Aşağıda elde edilen sonuçlar aşağıda açıklanmaktadır.

Yüzdelik (Persentil) Dilimler Yöntemi

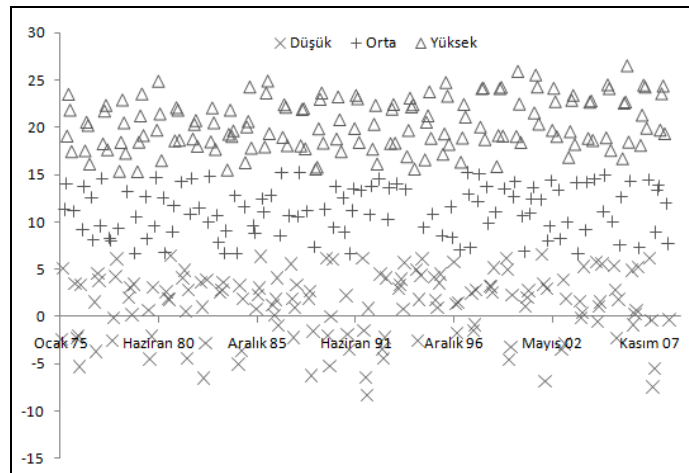
Yüzdelik (persentil) dilimler olarak adlandırılan ayrılaştırma yönteminde veriler yüzdelik dilimlere ayrılmaktadır. Yüzdelik dilimler yöntemi 17836 no'lu meteoroloji istasyonundan elde edilmiş aylık ortalama sıcaklık verileri üzerinde uygulanmıştır. Çok farklı sınıflandırma biçimleri mümkün olsa da, bu örnek uygulama da %75'den büyük olan dilimler "yüksek", %25 ile %75 arasında olan değerler "orta" ve %25'den küçük değerler "düşük" olarak adlandırılmıştır. Sonuçlar Şekil 4.13'de sunulmaktadır.



Şekil 4.13. Yüzdelik dilimler yöntemi ile ayrılaştırılmış 17836 no'lu istasyona ait aylık ortalama sıcaklık verileri

İstatistiksel Yöntem

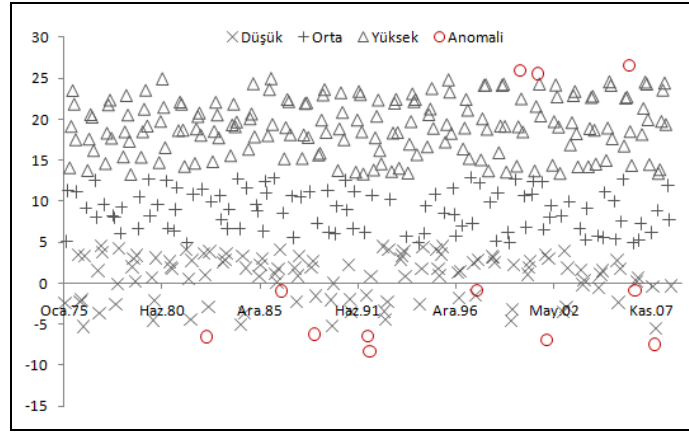
İkinci yöntem olarak veri serisinin istatistiksel özelliklerini dikkate alan bir ayrıklaştırma yöntemidir. Bu yöntemde serinin ortalama ve standart sapma (SS) değerleri hesaplanmakta ve veriler bu değerler kullanılarak gruplara ayrılmaktadır. İstatistiksel yöntem 17836 no'lu meteoroloji istasyonundan elde edilmiş aylık ortalama sıcaklık verileri üzerinde uygulanmıştır. Bu uygulamada “ortalama+0.5SS” değerinden büyük değerler “yüksek”, ortalama-0.5SS” değerinden küçük değerler ise “düşük”, “ortalama+0.5SS” ve “ortalama-0.5SS” arasındaki değerler “orta” olarak adlandırılmıştır. Sonuçlar Şekil 4.14’de sunulmaktadır.



Şekil 4.14. İstatistiksel yöntem ile ayrılaştırılmış 17836 no'lu istasyona ait aylık ortalama sıcaklık verileri

Kümeleme Yöntemi

Verilerin ayrıklaştırılmasında kullanılabilecek diğer bir yöntem kümeleme yöntemleridir. Kümelemede amaç birbirleri ile benzer özellik gösteren verilerin aynı grup altında toplanmasıdır. Proje kapsamında uzaklık tabanlı kümeleme yöntemlerinden biri olan DBSCAN (density-based clustering algorithm) algoritması [82] kullanılarak bir çalışma yapılmıştır. DBSCAN veritabanında yer alan kümeleri tespit etmek yanında anormal verileri (yani hiçbir kümeye ait olmayan verileri) de tespit edebilmektedir. DBSCAN algoritması 17836 nolu meteoroloji istasyonundan elde edilmiş aylık ortalama sıcaklık verileri üzerinde uygulanmıştır. Sonuçlar Şekil 15’de gösterilmektedir.



Şekil 4.15. Kümeleme (DBSCAN) algoritması kullanılarak ayrıklaştırılmış 17836 no’lu istasyona ait aylık ortalama sıcaklık verileri

Özet

Veri ayrıklaştırma amacıyla yüzdelik dilimler yöntemi, istatistiksel yöntemi ve kümeleme yöntemi olmak üzere üç yöntem belirlenmiş ve bu yöntemler 17836 nolu meteoroloji istasyonundan elde edilmiş aylık ortalama sıcaklık verileri üzerinde uygulanmıştır. Yapılan uygulamada birinci ve ikinci yöntemin birbirlerine nispeten benzer sonuçlar verdiği bulunmuştur. Bu iki yöntemde önemli olan hangi düzeylerden yüksek ve düşük değerlerin yüksek ve düşük olarak sınıflandırılacağıdır. Bu konuda karar grafiksel gösterim veya frekans analizi gibi yöntemlerde verilerin incelenmesi ile verilebilir. Kümeleme yöntemi veri kümesinde mevcut doğal grupları ortaya koyması bakımından avantajlıdır. Ancak özellikle DBSCAN yöntemi ile anomali olarak adlandırılan ve hiçbir gruba dahil olmayan grupların

belirlenmesi ve ortaya çıkan kümelerin etiketlenmesi gerekliliği bu yöntemin kullanımı zorlaştırabilir. Bu proje kapsamında yapılan çalışmalarda, analizlerin özelliklerine uygun olarak veri ayrıklaştırma için istatistiksel yöntemin kullanılmasına karar verilmiştir.

4.2.4. Mevsimselliğin Giderilmesi

Mevsimselliğin giderilmesi için çeşitli yöntemler (ayrık fourier transformu, aylık z-skoru, tekil veri çözümlemesi, ve hareketli ortalama gibi) literatürde mevcuttur [4]. Projenin ilk 6 aylık döneminde z-skoru yöntemi örnek bir veri kümesi üzerinde uygulanmış ve sonuçlarına 1. Gelişme Raporu'nda yer verilmiştir. Aylık z-skoru yöntemi Kumar v.d. [83] tarafından da belirtildiği gibi normal durumdan sapmayı ifade edebilmekte ve kolay uygulanabilmektedir. Bu tür avantajlarından dolayı aylık z-skoru yönteminin mevsimsellik giderimi gerekli olduğu durumlarda kullanılmasına karar verilmiştir.

4.3. Hidrolojik ve İklimsel İlişkilerin ve Örüntülerin Tanımlanması

Çalışmanın bu bölümünde hidrolojik ve iklimsel parametrelerin oluşturdukları birliktelik örüntülerine odaklanılmaktadır.

Genel olarak tanımlanacak olursa birliktelik analizi bir işlem kaydında (transaction) bir öğenin meydana gelme olasılığını diğer öğelerin meydana gelme olasılıklarından tahmin etmek için kurallar bulunmasıdır. Birliktelik analizi kullanıcı tarafından belirlenmiş frekans (sıklık) değerinden daha yüksek frekansa (daha sık) sahip olan $\{A\} \rightarrow \{B\}$ şeklindeki kuralların bulunmasını sağlar. Bu kuralda A ve B olayları veya iki farklı parametreyi göstermektedir. Örneğin, A yağış, B ise pik akım biçimi olabilir. $\{A\} \rightarrow \{B\}$ kuralı ise “eğer A gerçekleşirse, B gerçekleşir” şeklinde okunur. Bu yönüyle birliktelik analizi sadece parametreler arasındaki ilişkiyi açıklamamakta aynı zamanda neden-sonuç ilişkilerini gösteren kurallar oluşturmaktadır. Birliktelik analizi aynı zamanda birçok parametreyi aynı zamanda analiz etme kabiliyetine sahiptir. Yani birliktelik analiziyle $\{A, B, C\} \rightarrow \{D\}$ şeklinde veya $\{A, B, C\} \rightarrow \{D, E, F\}$ şeklinde kurallar (örüntüler) bulmak mümkündür.

Bu çalışmada odaklanılan kurallar iklimsel parametrelerin kuralların sol tarafında, hidrolojik parametrelerin ise kurallarının sağ tarafından yer aldığı birlikteliklerdir. Yani iklimsel

faktörlerin akım üzerinde ne tür bir etkiye neden olduğu araştırılmaktadır

4.4. Değerlendirme Ölçütü Tanımlama

Bu çalışmada bulunan kuralların anlamlılığını değerlendirmek için iki değerlendirme ölçütü kullanılmıştır. Bu ölçütlerden birincisi veri madenciliği uygulamalarında sıklıkla kullanılan “Destek” ölçütüdür. Çalışmada kullanılan diğer ölçüt ise bu çalışma kapsamında yeni tanımlanan “Hidro-Meteorolojik Yaygınlık Ölçütü”dür. Aşağıda bu iki ölçüt ve özellikleri açıklanmaktadır.

4.4.1. Destek Ölçütü

Klasik veri madenciliğinde parametrelerin ve ilişkilerin anlamlılıklarının ölçülmesi için yaygın olarak destek (support) değerlendirme ölçütü kullanılır. Destek; bir parametre değerinin veritabanında kaç kez tekrar ettiğini ifade eder. Eğer bir parametrenin veya ilişkinin destek değeri belirli bir eşik değerinden (threshold) yüksek ise o parametrenin veya ilişkinin anlamlı olduğuna karar verilir ve daha karmaşık ilişkilerin bulunması için bir sonraki aşamalarda kullanılır. Bu çalışmada destek değerlendirme ölçütü kullanılmıştır.

4.4.2. Hidrometeorolojik Yaygınlık Ölçütü

Bu çalışmada tanımlanan hidrometeorolojik yaygınlık ölçütü destek kriterini sağlayan kuralların mekânsal olarak yaygınlığını belirlemek için kullanılmıştır. Hidrometeorolojik yaygınlık ölçütü belirli bir destek değerine sahip kuralın analiz edilen veri gruplarında hangi oranda var olduğunu ifade etmektedir.

Bu çalışmada istasyonlar farklı veri gruplarını ifade etmektedir. Birliktelik kuralları her bir istasyon verisi üzerinde tespit edilmiş ve daha sonra varolan kuralların istasyonların kaç tanesinde olduğu, diğer bir ifadeyle çeşitli istasyonlarda manalı olan bir kuralın bütün istasyonlarda kaç oranında bulunduğu tespit edilerek istasyon yaygınlığı hesaplanmıştır. Eğer bulunan değer bir eşik değerinden yüksek ise kuralın hidrometeorolojik yaygın olduğu ifade edilmiştir. Hidro-meteorolojik yaygınlık ölçütü aşağıdaki eşitlikle de ifade edilebilir.

Hidrometeorolojik Yaygınlık Ölçütü = (Destek Kriterini Sağlayan İstasyon Sayısı)/(Toplam İstasyon Sayısı)

4.5. Hızlı ve Özgün Algoritmaların Tasarlanması

Hidrometeorolojik kuralların keşfedilmesindeki temel mantık öncelikle her istasyon için bütün manalı kuralların üretilmesi ve daha sonra bir son işlem adımında üretilen bütün kuralların hidrometeorolojik yaygınlıklarının belirlenmesidir. Ancak bu yaklaşımda gereksiz yere hidrometeorolojik yaygın olmayan kuralların üretilmesi ihtimali vardır. Bu nedenle bu yaklaşımın karmaşıklığı yüksektir.

Bu algoritmaya karşılık proje kapsamında yeni bir algoritma önerilmiştir. Algoritma öncelikle her bir istasyon için *min_sup* eşik değerini sağlayan manalı ikili kuralları üretir. Daha sonra bu ikili kuralların istasyonların kaçta kaçında olduğu hesaplanarak bu değerin *min_ist* değerini karşılayıp karşılamadığı belirlenir. Bu değeri karşılıyorsa kurallar hidrometeorolojik yaygındırlar ve her bir istasyonda 3'lü kuralların üretilmesi için kullanılırlar. 3'lü manalı kurallar üretildikten sonra yine hidrometeorolojik yaygınlığı kontrol edilir. Yaygın olanlar 4'lü kuralların üretilmesi için kullanılır ve bu işlem bütün hidrometeorolojik yaygın kurallar üretilene kadar devam eder. Bu algoritma Tablo 4.10'da sunulmaktadır.

Tablo 4.10. Hidrolojik ve iklimsel parametreler kullanılarak örüntü madenciliği algoritması

Girdiler:

min_sup: Algoritmaya verilecek olan destek ölçütü değeri

min_ist: Algoritmaya verilecek olan hidrometeorolojik yaygınlık ölçütü değeri

Çıktı:

Kurallar: Gerekli şartları sağlayan n’li kurallar

Algoritma:

```

1. okunan_veri=dosyadan_veri_oku
2. [Tum_Istasyon_Verileri,Istasyon_No]=istasyon_bilgilerini_al [Okunan_veri]
3. kural_ikili = ikili_kurallari_olustur (Tum_istasyon_verileri)
4. Kural_Ist = matris(Istasyon_Sayisi, length(kural_ikili))
5. for i = 1:Istasyon_Sayisi
6.     for j = 1: length(kural_ikili)
7.         sup = a_priori(Ist_Veri, kural_ikili(j, :))
8.         if sup >= min_sup
9.             Kural_Ist(i, j) = Kural_Ist(i, j) + 1
10.        end
11.    end
12. end
13. dosya_yaz(kurallar)
14. kurallar=istasyonlarda_yayginlik_hesapla (kurallar)
15. if kurallar>= min_ist then
16.     secilen_kurallar = kurallar
17. Kural_Coklu = secilen_kurallar
18. while(Kural_Coklumatrisi bos değilse)
19.     Yeni_Aday = bir_ust_aday_olustur(Kural_Coklu)
20.     if satir_Yeni_Aday< 2
21.         Dongudencik
22.     end
23.     Kural_Ist_2 = matris(Istasyon_Sayisi,length(Yeni_Aday))
24.     for i = 1:Istasyon_Sayisi
25.         for j = 1:satir_Yeni_Aday
26.             sup = a_priori(Ist_Veri, Yeni_Aday(j, :))
27.             if sup >= min_sup
28.                 Kural_Ist(i, j) = Kural_Ist(i, j) + 1
29.             end
30.         end
31.     end
32.     dosya_yaz(kurallar)
33.     istasyonlardayayginlik hesapla
34.     Kural_Coklu = kurallar(yayginlik>= min_ist)
35. end

```

Tablo 4.10’da verilen algoritmanın çalışma prensibi adım adım aşağıda açıklanmıştır.

Adım 1 – *dosyadan_veri_oku*: Bu fonksiyon, üzerinde çalışacağımız veriyi dosyadan okunur.

Adım 2 – *istasyon_bilgilerini_al*: Bu adımda istasyon numaraları tespit edilir ve bir diziye atılır. Ayrıca veriler ayrıklaştırılarak tüm_istasyon_verileri değişkenine aktarılır.

Adım 3 – *ikili_kurallari_olustur*: Bu adımda ayrıklaştırılmış verilerden elde edilen tüm değerler kullanılarak aday ikili kurallar üretilir. Bu aşamada modifiye edilmiş olan kısım kuralın en sağında her zaman akım parametresinin olmasıdır. Çünkü bu çalışmada diğer parametrelerin akım üzerindeki etkisini incelenmiştir.

Adım 4 – *matris*: Bu adımda, verilen boyutlarda, içi sıfırla doldurulmuş bir matris oluşturuluyor.

Adım 5 ile 12 arasındaki döngüde her bir istasyon için işlem yapılmaktadır.

Adım 6 ile 11 arasındaki döngüde her istasyon içerisinde bulunan her bir ikili kural için işlem yapılacaktır.

Adım 7’de *a-priori* fonksiyonu kullanılarak kuralların destek değerleri hesaplanmaktadır ve sup değişkenine atanmaktadır. Bu değer belli bir eşik değerini sağlıyor ise bu kural anlamlıdır denilir. Bu işlemin kontrolü ise **Adım 8’de** yapılmaktadır ve anlamlı kuralların kaydı adım 9’ da tutulmaktadır.

Adım 13 – *dosya_yaz*: Bu fonksiyon ile adım 5-12 arasında her bir istasyon için üretilen kurallar bir dosyaya yazdırılır.

Adım 14’de kuralların hidr-meteorolojik yaygınlıkları hesaplanır.

Adım 15 ve 16’da ise hidrometeorolojik yaygınlığı min_ist’i sağlayan kurallar belirlenerek secilen_kurallar değişkenine atanır.

Buraya kadar elde edilen ikili kurallar üçlü kuralları üretmek için kullanılacaktır.

Adım 18-35 arasında oluşabilecek bütün kurallar üretilecektir daha sonra ikili kural üretiminde olduğu gibi üretilen kuralın destek ve hidro-meteorolojik yaygınlığı test edilerek dosyaya kaydedilecektir.

4.6. İlişki ve Örüntü Madenciliği

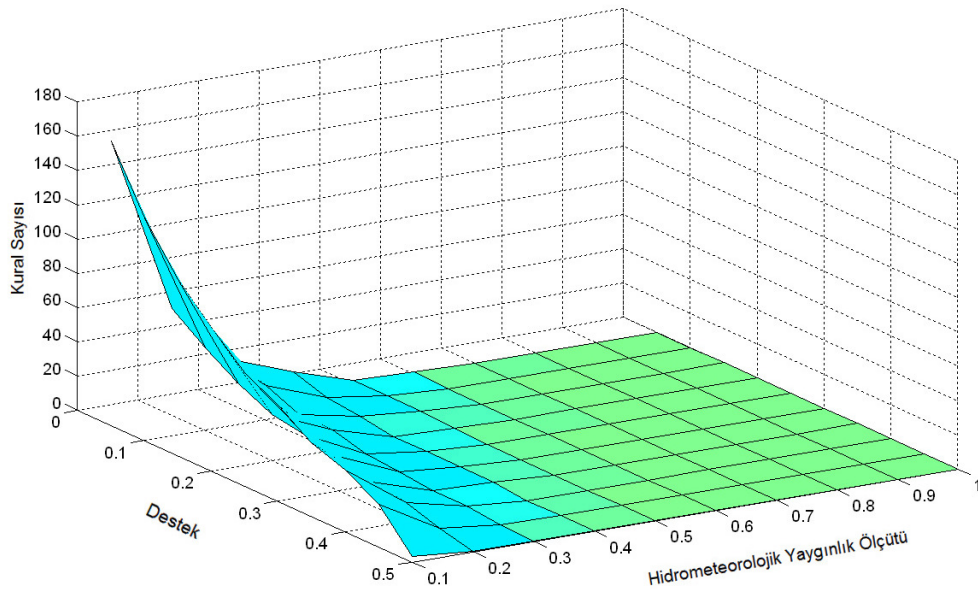
Proje kapsamında 3 farklı veri seti algoritma çalıştırılmıştır. Bunlar 1) Bütün veriler 2) Akımın yüksek ve düşük olduğu dönemler 3) Mevsimsel

4.6.1. Bütün veriler kullanılarak elde edilen sonuçlar

Aşağıda bütün veriler kullanılarak elde edilen sonuçlar sunulmaktadır. Analizin başlangıcında destek ve hidrometeorolojik yaygınlık ölçütü için hangi değerlerin en uygun olacağını belirlenmesi için farklı değerler kullanılarak elde edilen kural sayıları incelenmiştir. Bu amaçla destek değeri 0.00-0.50 arasında, hidrometeorolojik yaygınlık ölçütü değeri ise 0.00-1.00 arasında değiştirilerek algoritma çalıştırılmış ve her bir adımda sonuç olarak elde edilen kural sayısı hesaplanmıştır. Daha sonra bu kurallar grafiksel olarak ifade edilmiştir. Şekil 4.16'da gösterilen grafikte x eksenı destek değerlerini, z eksenı hidrometeorolojik yaygınlık ölçütü değerlerini, ve y eksenı ise kural sayısını ifade etmektedir. Aynı sonuçlar ayrıca Tablo 4.11'de de sunulmaktadır.

Tablo 4.11. Farklı destek ve hidrometeorolojik yaygınlık ölçütü değerleri için kural sayıları

Destek	Hidrometeorolojik Yaygınlık Ölçütü				
	0.50	0.60	0.70	0.80	0.90
0.05	66	56	43	36	25
0.10	24	20	14	11	5
0.15	7	5	3	3	1
0.20	2	2	2	0	0
0.25	0	0	0	0	0



Şekil 4.16. Farklı destek ve hidrometeorolojik yaygınlık ölçütü değerleri için kural sayıları

Şekil 4.16 ve Tablo 4.11’de görüleceği üzere en fazla kural düşük destek ve düşük hidrometeorolojik yaygınlık ölçütü değerleri için tespit edilmiştir. En az kural ise yüksek destek ve yüksek hidrometeorolojik yaygınlık ölçütü değerleri için tespit edilmiştir. Kural sayısının çok fazla olması ilginç olmayan kuralların da sonuçlara dahil edildiğinin göstergesi olabilir. Bu durum sonuçların değerlendirilmesini zorlaştırmaktadır. Kural sayısının çok az olması durumunda ise bazı ilginç kuralların ihmal edilmesi ile sonuçlanabilir. Elde edilen sonuçlar bu çerçevede değerlendirilerek, destek sınır değeri olarak 0.15’in ve hidrometeorolojik yaygınlık ölçütü sınır değeri olarak ise 0.50’nin seçilmesine karar verilmiştir. Diğer bir ifadeyle bütün veri seti içinde en az %15 oranında mevcut olan ve çalışmada kullanılan 64 istasyonun en az %50’sinde görülen kurallar ilginç olarak kabul edilmiş ve değerlendirilmiştir. Bu kriterleri sağlayan toplam 7 kural elde edilmiştir.

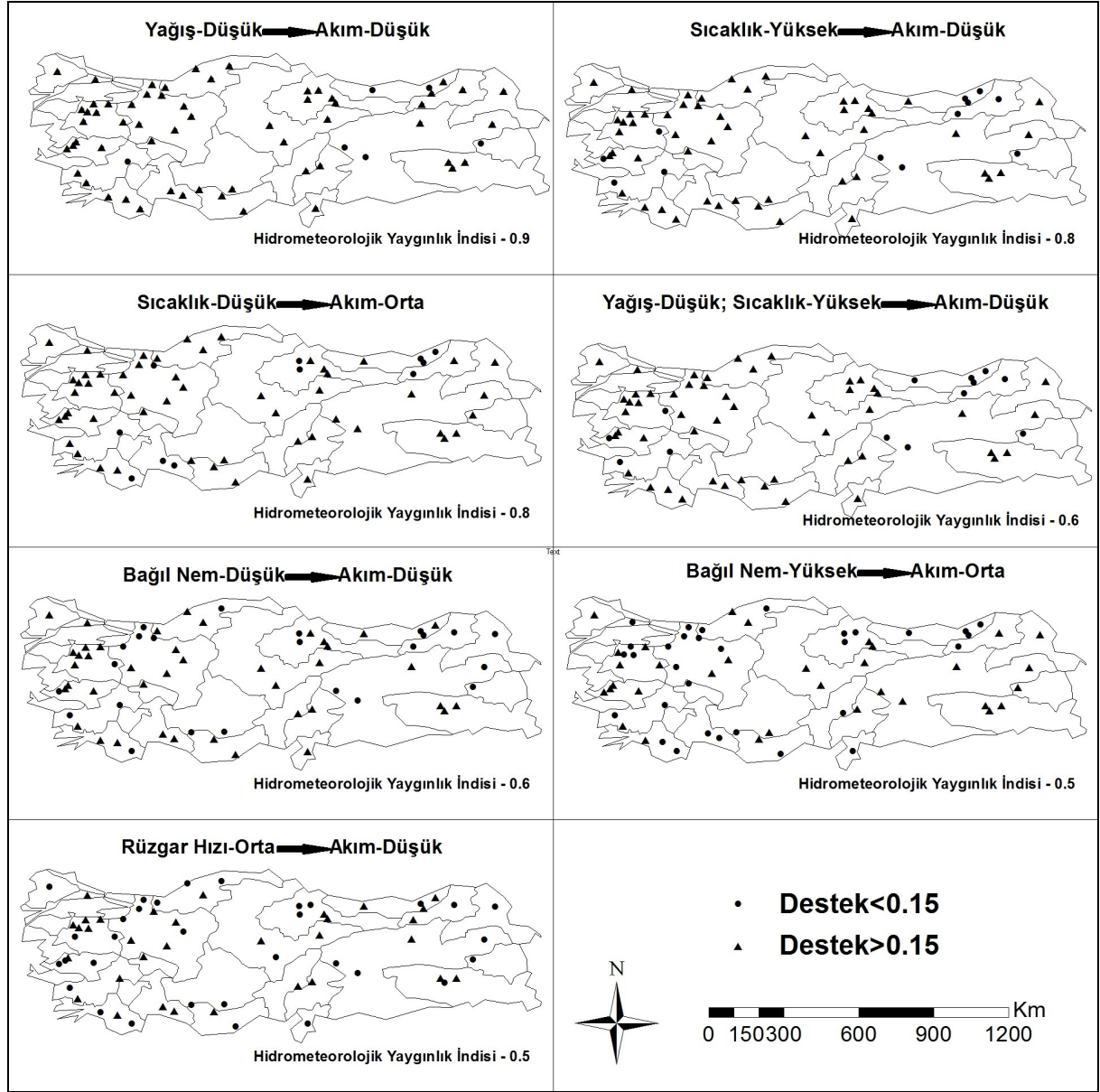
Tablo 4.12’de ve Şekil 4.17’de destek sınır değeri olarak 0.15 ve hidrometeorolojik yaygınlık ölçütü sınır değeri 0.50 alınmasıyla üretilen kurallar gösterilmektedir.

Tablo 4.12. Tüm verilerin kullanılması durumunda elde edilen kurallar (destek: 0.15, hidrometeorolojik yaygınlık ölçütü :0.50). D: Düşük, O: Orta, Y: Yüksek

Destek	0.15		0.20	
Hidrometeorolojik Yaygınlık Ölçütü	Ön Şart	Sonuç	Ön Şart	Sonuç
0.50	R. Hızı-O	Akım-D	Yağış-D	Akım-D
	B. Nem-D	Akım-D	Sıcaklık-Y	Akım-D
	B. Nem-Y	Akım-O		
	Yağış-D	Akım-D		
	Sıcaklık-D	Akım-O		
	Sıcaklık-Y	Akım-D		
	Yağış-L; Sıcaklık-Y	Akım-D		
0.60	B. Nem-D	Akım-D	Yağış-D	Akım-D
	Yağış-D	Akım-D	Sıcaklık-Y	Akım-D
	Sıcaklık-D	Akım-O		
	Sıcaklık-Y	Akım-D		
	Yağış-L; Sıcaklık-Y	Akım-D		
0.70	Yağış-D	Akım-D	Yağış-D	Akım-D
	Sıcaklık-D	Akım-O	Sıcaklık-Y	Akım-D
	Sıcaklık-Y	Akım-D		
0.80	Yağış-D	Akım-D		
	Sıcaklık-D	Akım-O		
	Sıcaklık-Y	Akım-D		
0.90	Yağış-D	Akım-D		

Tüm veriler kullanıldığı durumda ortaya çıkan en güçlü kural “Yağış-Düşük → Akım-Düşük” kuralıdır. Bu kural “yağış düşükse, akım düşüktür” şeklinde de ifade edilebilir. Bu kural bütün istasyonların en az %90’ında mevcuttur. Türkiye’nin doğusunda yer alan kuzey-güney doğrultusunda uzanan bir hat dışında bu kural ağırlıklı olarak yer almıştır. Ortaya çıkan diğer önemli iki kural “Sıcaklık-Yüksek → Akım-Düşük” ve “Sıcaklık-Düşük → Akım-Orta” kurallarıdır. Her iki kural için de hidrometeorolojik yaygınlık indisi değeri en az 0.8’dir. “Sıcaklık-Yüksek → Akım-Düşük” kuralı kuzeydoğuda ve doğuda yer alan bazı bölgeler dışında her bölgede mevcuttur. “Sıcaklık-Düşük → Akım-Orta” kuralı ise kuzeydoğu ve güneybatıda yer aşan bazı istasyonlarda görülmemiştir. Yağış ve sıcaklık parametrelerinin her ikisinin birlikte bulunduğu kural olan “Yağış-Düşük; Sıcaklık-Yüksek → Akım-Düşük” kuralı ise bütün istasyonların %60’ında görülmüştür. Bu kural özellikle Türkiye’nin batı kısmında yer alan istasyonlarda yaygın olarak mevcuttur. Türkiye’nin kuzeydoğu bölgesinde yer alan istasyonlarda ve doğuda yeralan kuzey-güney yönünde uzanan hat boyunca mevcut değildir. Bağıl nem ve rüzgar hızı ile akım arasındaki ilişkiler nispeten zayıftır. Bu nedenle elde edilen kuralların yaygınlık düzeyleri (yani kuralların görüldüğü istasyon sayıları) de düşüktür. Bunun yanında her iki parametre ile elde edilen kurallar coğrafi olarak bir örüntü oluşturmamaktadır.

Elde edilen sonuçlar genel olarak akım ve yağış ve sıcaklık arasındaki bağlantıların daha güçlü olduğunu göstermektedir. Elde edilen kurallar “yağış düşükse, sıcaklık yüksekse, akım düşüktür” şeklinde özetlenebilir. Söz konusu kurallar hemen hemen bütün Türkiye için geçerlidir. Sadece kuzeybatıda yer alan istasyonlar için akım ile yağış ve sıcaklık ilişkisi zayıftır.

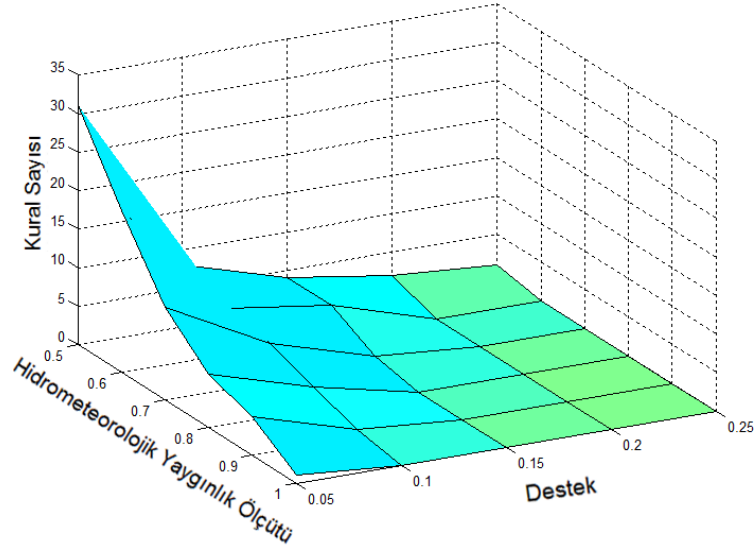


Şekil 4.17. Tüm verilerin kullanılması durumunda elde edilen kuralların coğrafi dağılımı (destek: 0.15, hidrometeorolojik yaygınlık ölçütü :0.50)

4.6.2. Düşük ve Yüksek Akımlar için Sonuçlar

Çalışmanın ikinci kısmında geliştirilen algoritma sadece düşük ve sadece yüksek akımlar için çalıştırılmıştır. Böylece yaygın olarak görülen ortalama değerlerden arındırılan veritabanlarında yapılan çalışmada, ekstrem akımlara ve bu akımların meteorolojik faktörlerle ilişkilerine odaklanılmıştır. Bu kısımda da kurallar belirlenirken ilginç kuralların seçilmesine dikkat edilmiştir. Bu amaçla destek değeri 0.00-0.50 arasında, hidrometeorolojik yaygınlık

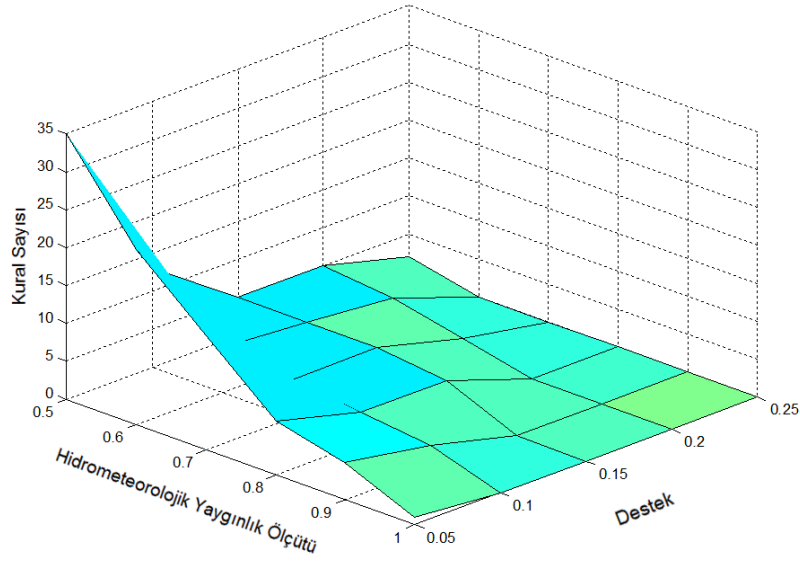
ölçütü değeri ise 0.00-1.00 arasında değiştirilerek algoritma çalıştırılmış ve her bir adımda sonuç olarak elde edilen kural sayısı hesaplanmıştır (Tablo 4.13 ve Tablo 4.14, Şekil 4.18 ve Şekil 4.19). Tüm verilerde elde edilen sınır değerlerin (destek-0.15; hidrometeorolojik yaygınlık indisi-0.50) bu kısım için de uygun olduğuna karar verilmiştir. Düşük akımlar için bu kriterleri sağlayan dört kural, yüksek akımlar için ise beş kural tespit edilmiştir.



Şekil 4.18. Farklı destek ve hidrometeorolojik yaygınlık ölçütü değerleri için kural sayıları (Düşük Akım)

Tablo 4.13. Farklı destek ve hidrometeorolojik yaygınlık ölçütü değerleri için kural sayıları (Düşük Akım)

Destek	Hidrometeorolojik Yaygınlık Ölçütü				
	0.50	0.60	0.70	0.80	0.90
0.05	31	21	12	7	5
0.10	8	6	5	3	1
0.15	4	4	1	0	0
0.20	2	0	0	0	0
0.25	1	0	0	0	0



Şekil 4.19. Farklı destek ve hidrometeorolojik yaygınlık ölçütü değerleri için kural sayıları (Yüksek Akım)

Tablo 4.14. Farklı destek ve hidrometeorolojik yaygınlık ölçütü değerleri için kural sayıları (Yüksek Akım)

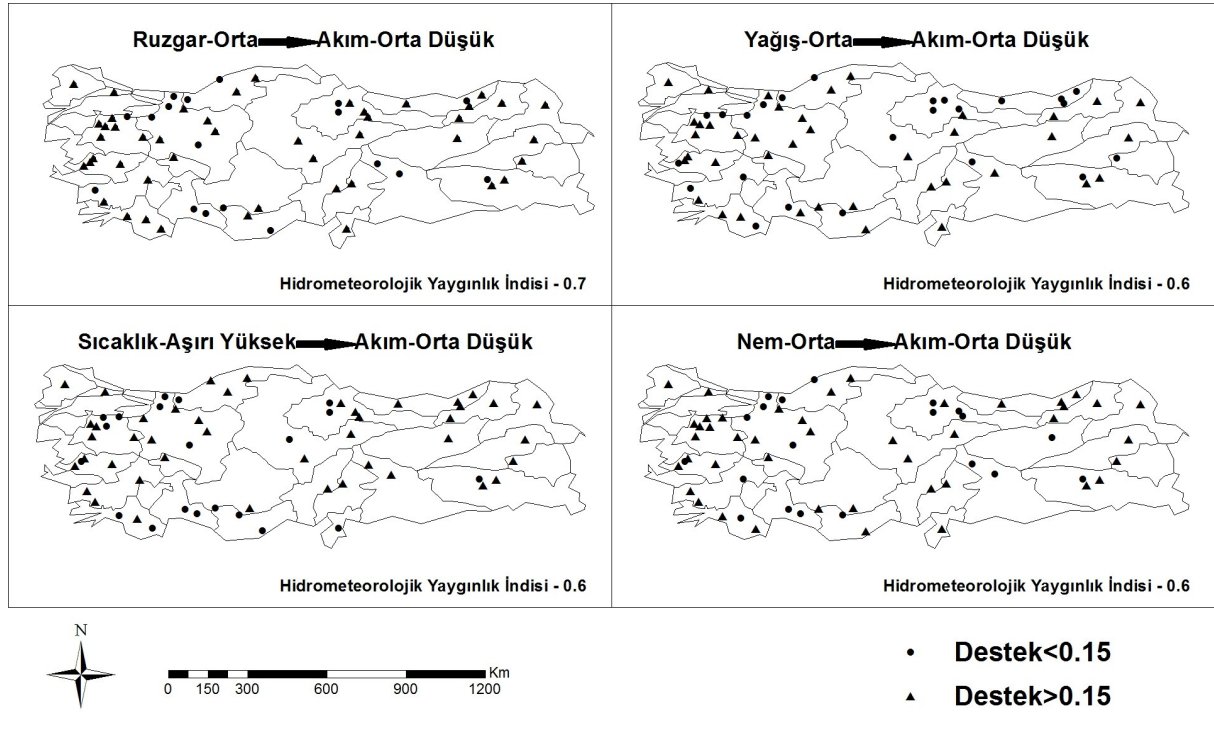
Destek	Hidrometeorolojik Yaygınlık Ölçütü				
	0.50	0.60	0.70	0.80	0.90
0.05	35	23	15	7	5
0.10	13	6	5	4	3
0.15	5	5	5	4	0
0.20	5	4	2	0	0
0.25	2	0	0	0	0

Düşük akımlar için destek sınır değerinin 0.15 ve hidrometeorolojik yaygınlık ölçütü sınır değeri 0.50 alınmasıyla üretilen kurallar Tablo 4.15’de ve Şekil 4.20’de gösterilmektedir.

Tablo 4.15. Tüm verilerin kullanılması durumunda elde edilen kurallar (Düşük Akım) (destek: 0.15, hidrometeorolojik yaygınlık ölçütü :0.50). D: Düşük, O: Orta, Y: Yüksek, AY: Aşırı Yüksek, OD: Orta Düşük

Destek	0.15		0.20		0.25	
Hidrometeorolojik Yaygınlık Ölçütü	Ön Şart	Sonuç	Ön Şart	Sonuç	Ön Şart	Sonuç
0.50	R. Hızı-O	Akım-OD	R. Hızı-O	Akım-OD	R. Hızı-O	Akım-OD
	B. Nem-O	Akım-OD	Yağış-O	Akım-OD		
	Yağış-O	Akım-OD				
	Sıcaklık-AY	Akım-OD				
0.60	R. Hızı-O	Akım-OD				
	B. Nem-O	Akım-OD				
	Yağış-O	Akım- OD				
	Sıcaklık-AY	Akım- OD				
0.70	R. Hızı-O	Akım- OD				

Düşük akımlar için elde edilen en güçlü kuralların tamamı akım orta düzeyde düşük olduğu duurm için üretilmiştir. Akım, rüzgar orta, yağış-orta, sıcaklık-aşırı yüksek ve bağıl nem-orta olduğu durumlarda orta düzeyde düşük çıkmıştır. Elde edilen kurallar arasında en güçlü kurual “Rüzgar-Orta → Akım-Orta Düşük” kuralıdır. Bu kuralın destek değeri 0.25 ve hidrometeorolojik yaygınlık indisi değeri ise 0.7’dir.

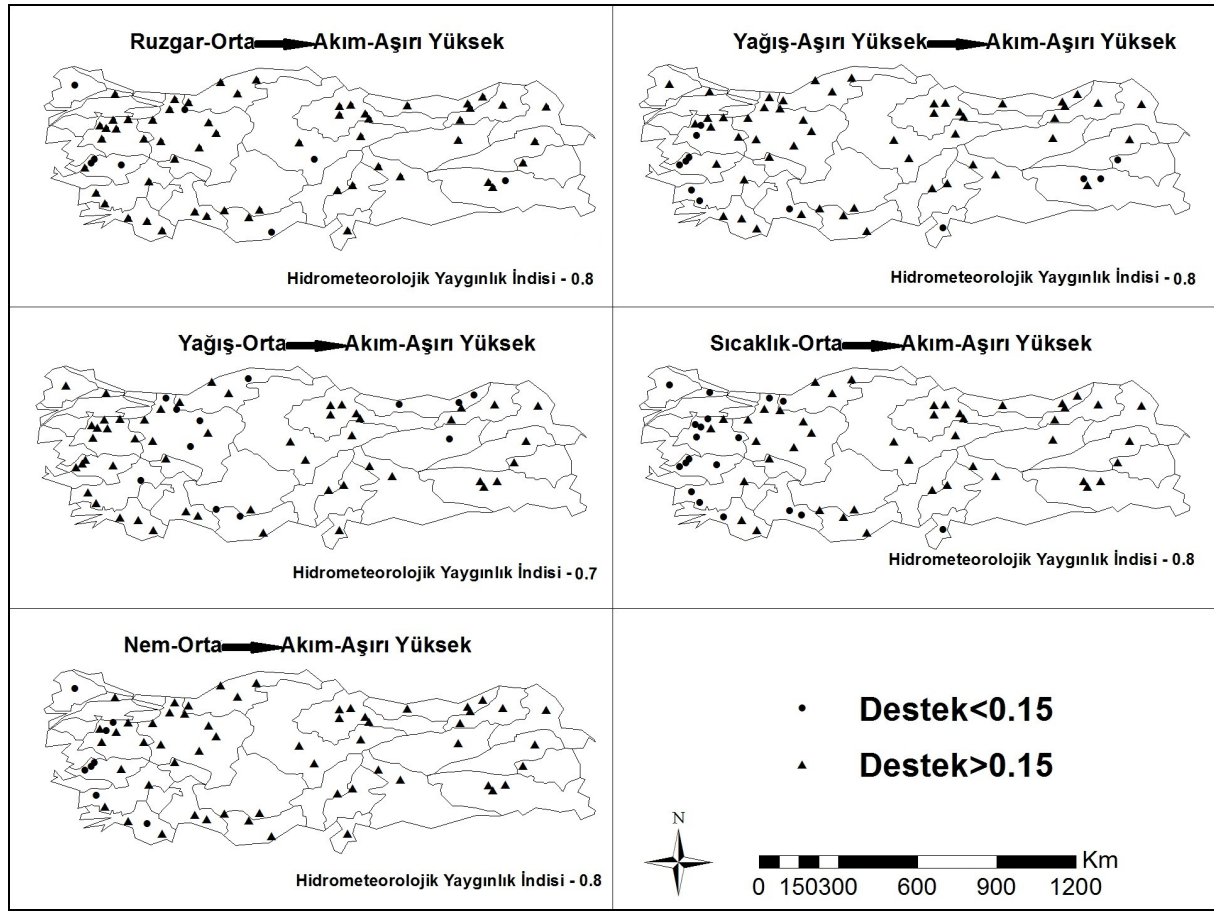


Şekil 4.20. Düşük akımları için elde edilen kuralların coğrafi dağılımı (destek: 0.15, hidrometeorolojik yaygınlık ölçütü :0.50)

Yüksek akımlar için destek sınır değerinin 0.15 ve hidrometeorolojik yaygınlık ölçütü sınır değeri 0.50 alınmasıyla üretilen kurallar Tablo 4.16’de ve Şekil 4.21’de gösterilmektedir. Yüksek akımlar için elde edilen kurallar sayıca daha fazladır. Destek değeri en yüksek (0.25) kurallar “B. Nem-Orta → Akım-Aşırı Yüksek” ve “Yağış-Orta → Akım-Aşırı Yüksek” kurallarıdır. Buna karşılık Türkiye için en yaygın olarak görülen dört adet kural bulunmaktadır. Bunlar B. Nem-Orta → Akım-Aşırı Yüksek”, “Yağış-Orta → Akım-Aşırı Yüksek”, “ Rüzgar Hızı-Orta → Akım-Aşırı Yüksek” ve “Yağış-Aşırı Yüksek → Akım-Aşırı Yüksek” kurallarıdır.

Tablo 4.16. Tüm verilerin kullanılması durumunda elde edilen kurallar (Yüksek Akım) (destek: 0.15, hidrometeorolojik yaygınlık ölçütü :0.50). D: Düşük, O: Orta, Y: Yüksek, AY: Aşırı Yüksek, AY: Aşırı Yüksek

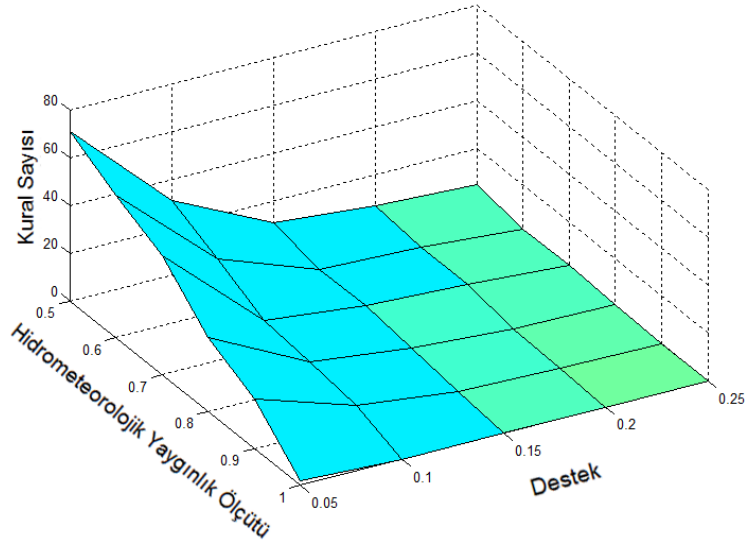
Destek	0.15		0.20		0.25	
Hidrometeorolojik Yaygınlık Ölçütü	Ön Şart	Sonuç	Ön Şart	Sonuç	Ön Şart	Sonuç
0.50	R. Hızı-O	Akım-AY	R. Hızı-O	Akım-AY	B. Nem-M	Akım-EH
	B. Nem-O	Akım-AY	B. Nem-O	Akım-AY	Yağış-M	Akım-EH
	Yağış-O	Akım-AY	Yağış-O	Akım-AY		
	Yağış-AY	Akım-AY	Yağış-AY	Akım-AY		
	Sıcaklık-O	Akım-AY	Sıcaklık-O	Akım-AY		
0.60	R. Hızı-O	Akım-AY	R. Hızı-O	Akım-AY		
	B. Nem-O	Akım-AY	B. Nem-O	Akım-AY		
	Yağış-O	Akım-AY	Yağış-AY	Akım-AY		
	Yağış-AY	Akım-AY	Sıcaklık-O	Akım-AY		
	Sıcaklık-O	Akım-AY				
0.70	R. Hızı-O	Akım-AY	B. Nem-O	Akım-AY		
	B. Nem-O	Akım-AY	Yağış-AY	Akım-AY		
	Yağış-O	Akım-AY				
	Yağış-AY	Akım-AY				
	Sıcaklık-O	Akım-AY				
0.80	R. Hızı-O	Akım-AY				
	B. Nem-O	Akım-AY				
	Yağış-O	Akım-AY				
	Yağış-AY	Akım-AY				



Şekil 4.21. Yüksek akımları için elde edilen kuralların coğrafi dağılımı (destek: 0.15, hidrometeorolojik yaygınlık ölçütü :0.50)

4.6.3. Yaz ve Kış Dönemleri için Elde Edilen Sonuçlar

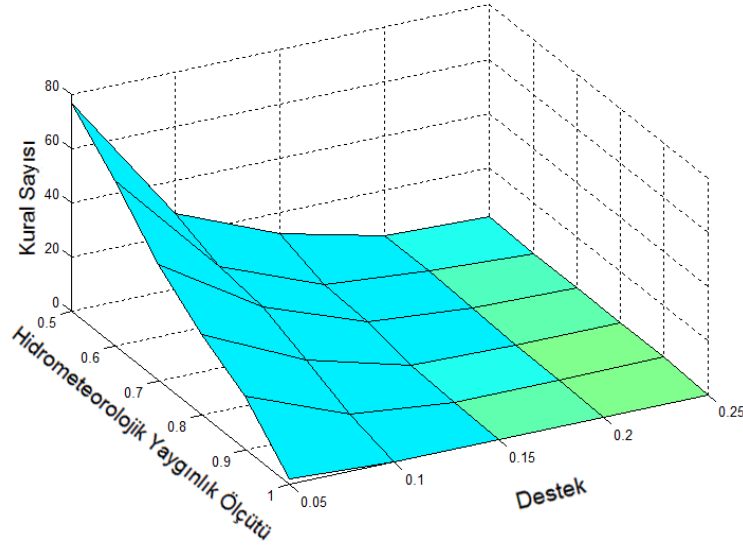
Çalışmanın üçüncü kısmında geliştirilen algoritma sadece yaz ve sadece kış mevsimleri için çalıştırılmıştır. Böylece mevsimsel olarak akımın meteorolojik faktörlerle ilişkilerine odaklanılmıştır. Bu kısımda da kurallar belirlenirken ilginç kuralların seçilmesine dikkat edilmiştir. Bu amaçla destek değeri 0.00-0.50 arasında, hidrometeorolojik yaygınlık ölçütü değeri ise 0.00-1.00 arasında değiştirilerek algoritma çalıştırılmış ve her bir adımda sonuç olarak elde edilen kural sayısı hesaplanmıştır (Tablo 4.17 ve Tablo 4.18, Şekil 4.22 ve Şekil 4.23). Tüm verilerde elde edilen sınır değerlerin (destek-0.15; hidrometeorolojik yaygınlık indisi-0.50) bu kısım için de uygun olduğuna karar verilmiştir. Yaz mevsimi için bu kriterleri sağlayan on bir kural, kış mevsimi için ise on iki kural tespit edilmiştir.



Şekil 4.22. Farklı destek ve hidrometeorolojik yaygınlık ölçütü değerleri için kural sayıları (Yaz Mevsimi)

Tablo 4.17. Farklı destek ve hidrometeorolojik yaygınlık ölçütü değerleri için kural sayıları (Yaz Mevsimi)

Destek	Hidrometeorolojik Yaygınlık Ölçütü				
	0.50	0.60	0.70	0.80	0.90
0.05	71	60	50	31	21
0.10	31	22	12	10	7
0.15	11	7	7	5	2
0.20	7	5	5	2	1
0.25	5	2	2	1	0



Şekil 4.23. Farklı destek ve hidrometeorolojik yaygınlık ölçütü değerleri için kural sayıları (Kış Mevsimi)

Tablo 4.18. Farklı destek ve hidrometeorolojik yaygınlık ölçütü değerleri için kural sayıları (Kış Mevsimi)

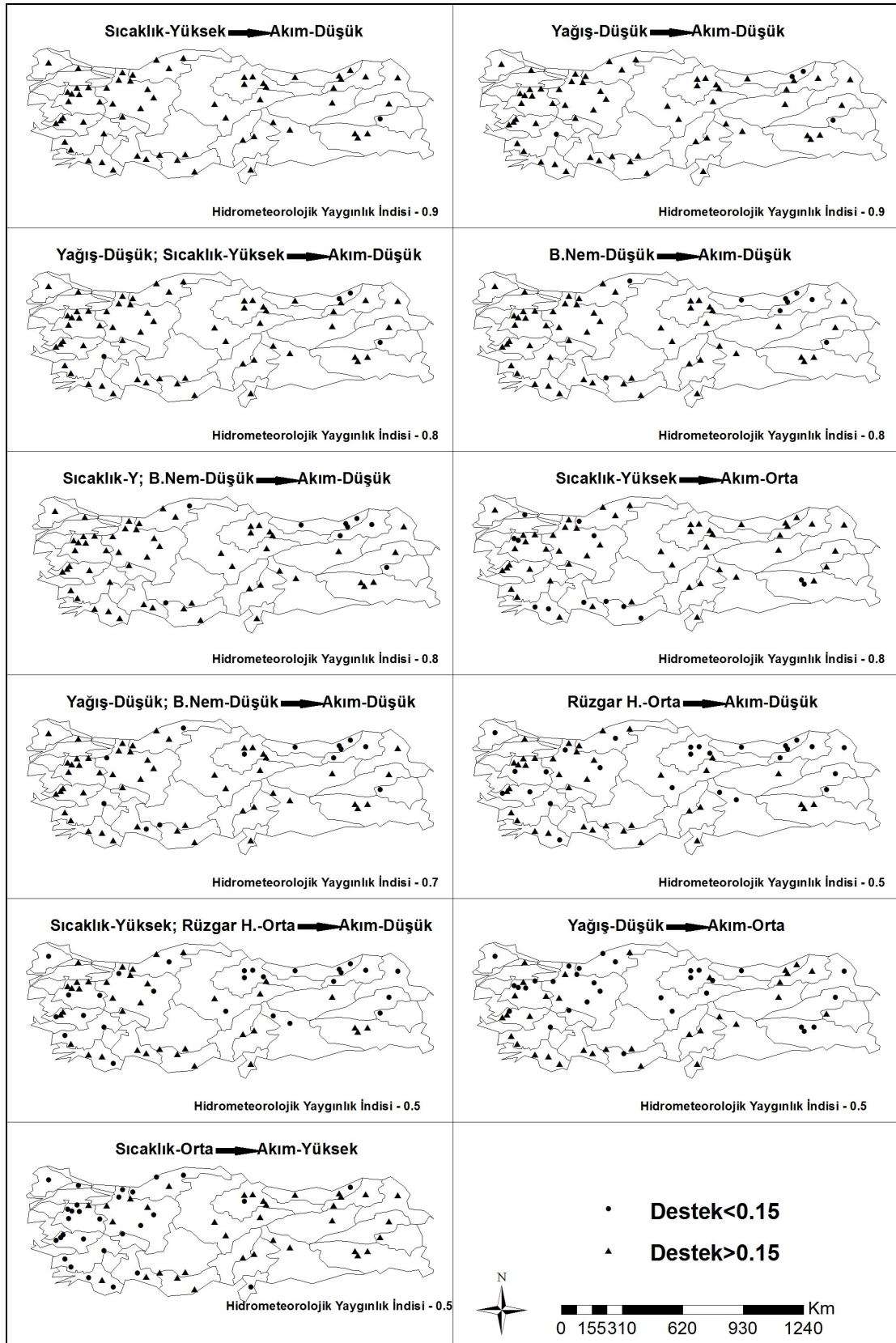
Destek	Hidrometeorolojik Yaygınlık Ölçütü				
	0.50	0.60	0.70	0.80	0.90
0.05	77	61	43	30	20
0.10	28	21	19	12	5
0.15	12	6	5	2	1
0.20	3	3	2	1	1
0.25	2	1	1	1	1

Yaz mevsimi için akımlar için destek sınır değerinin 0.15 ve hidrometeorolojik yaygınlık ölçütü sınır değeri 0.50 alınmasıyla üretilen kurallar Tablo 4.19’de ve Şekil 4.24’de gösterilmektedir. Yaz ayları akımları için elde edilen kurallar arasında destek değeri en yüksek kural “Sıcaklık-Yüksek → Akım-Düşük” kuralıdır. İkinci sırada ise “Yağış-Düşük → Akım-Düşük” kuralı yer almaktadır. Bu iki kural için hidrometeorolojik yaygınlık indisi değerleri de hayli yüksektir (sırasıyla 0.9 ve 0.8). Bu kuralların birlikte oluşturduğu “Sıcaklık-Yüksek, Yağış-Düşük → Akım-Düşük” kuralı, “ B. Nem-D, Sıcaklık-Yüksek → Akım-Aşırı Düşük” kuralı da hem destek değeri hem de yaygınlık değeri bakımından üst sıralarda yer almaktadırlar.

Tablo 4.19. Yaz mevsimi için elde edilen kurallar (destek: 0.15, hidrometeorolojik yaygınlık ölçütü :0.50). D: Düşük, O: Orta, Y: Yüksek

Destek	0.15		0.20		0.25	
Hidrometeorolojik Yaygınlık Ölçütü	Ön Şart	Sonuç	Ön Şart	Sonuç	Ön Şart	Sonuç
0.50	R. Hızı-O	Akım-D	B. Nem-D	Akım-D	B. Nem-D	Akım-D
	B. Nem-D	Akım-D	Yağış-D	Akım-D	Yağış-D	Akım-D
	Yağış-D	Akım-D	Sıcaklık-Y	Akım-D	Sıcaklık-Y	Akım-D
	Yağış-D	Akım-O	Sıcaklık-Y	Akım-O	B. Nem-D; Sıcaklık-Y	Akım-D
	Sıcaklık-O	Akım-Y	B. Nem-D; Yağış-D	Akım-D	Yağış-D; Sıcaklık-Y	Akım-D
	Sıcaklık-Y	Akım-D	B. Nem-D; Sıcaklık-Y	Akım-D		
	Sıcaklık-Y	Akım-O	Yağış-D; Sıcaklık-Y	Akım-D		
	R.Hızı-O; Sıcaklık-Y	Akım-D				
	B. Nem-D; Yağış-D	Akım-D				
	B. Nem-D; Sıcaklık-Y	Akım-D				
	Yağış-D; Sıcaklık-Y	Akım-D				
0.60	B. Nem-D	Akım-D	B. Nem-D	Akım-D	Yağış-D	Akım-D
	Yağış-D	Akım-D	Yağış-D	Akım-D	Sıcaklık-Y	Akım-D
	Sıcaklık-Y	Akım-D	Sıcaklık-Y	Akım-D		
	Sıcaklık-Y	Akım-O	B. Nem-D; Sıcaklık-Y	Akım-D		
	B. Nem-D; Yağış-D	Akım-D	Yağış-D; Sıcaklık-Y	Akım-D		
	B. Nem-D; Sıcaklık-Y	Akım-D				
	Yağış-D; Sıcaklık-Y	Akım-D				
0.70	B. Nem-D	Akım-D	B. Nem-D	Akım-D	Yağış-D	Akım-D
	Yağış-D	Akım-D	Yağış-D	Akım-D	Sıcaklık-Y	Akım-D
	Sıcaklık-Y	Akım-D	Sıcaklık-Y	Akım-D		
	Sıcaklık-Y	Akım-O	B. Nem-D; Sıcaklık-Y	Akım-D		
	B. Nem-D; Yağış-D	Akım-D	Yağış-D; Sıcaklık-Y	Akım-D		
	B. Nem-D; Sıcaklık-Y	Akım-D				
	Yağış-D; Sıcaklık-Y	Akım-D				

Destek	0.15		0.20		0.25	
Hidrometeorolojik Yaygınlık Ölçütü	Ön Şart	Sonuç	Ön Şart	Sonuç	Ön Şart	Sonuç
0.80	B. Nem-D	Akım-D	Yağış-D	Akım-D	Sıcaklık-Y	Akım-D
	Yağış-D	Akım-D	Sıcaklık-Y	Akım-D		
	Sıcaklık-Y	Akım-D				
	Sıcaklık-Y	Akım-O				
	B. Nem-D; Yağış-D	Akım-D				
	B. Nem-D; Sıcaklık-Y	Akım-D				
0.90	Yağış-D	Akım-D	Sıcaklık-Y	Akım-D		
	Sıcaklık-Y	Akım-D				

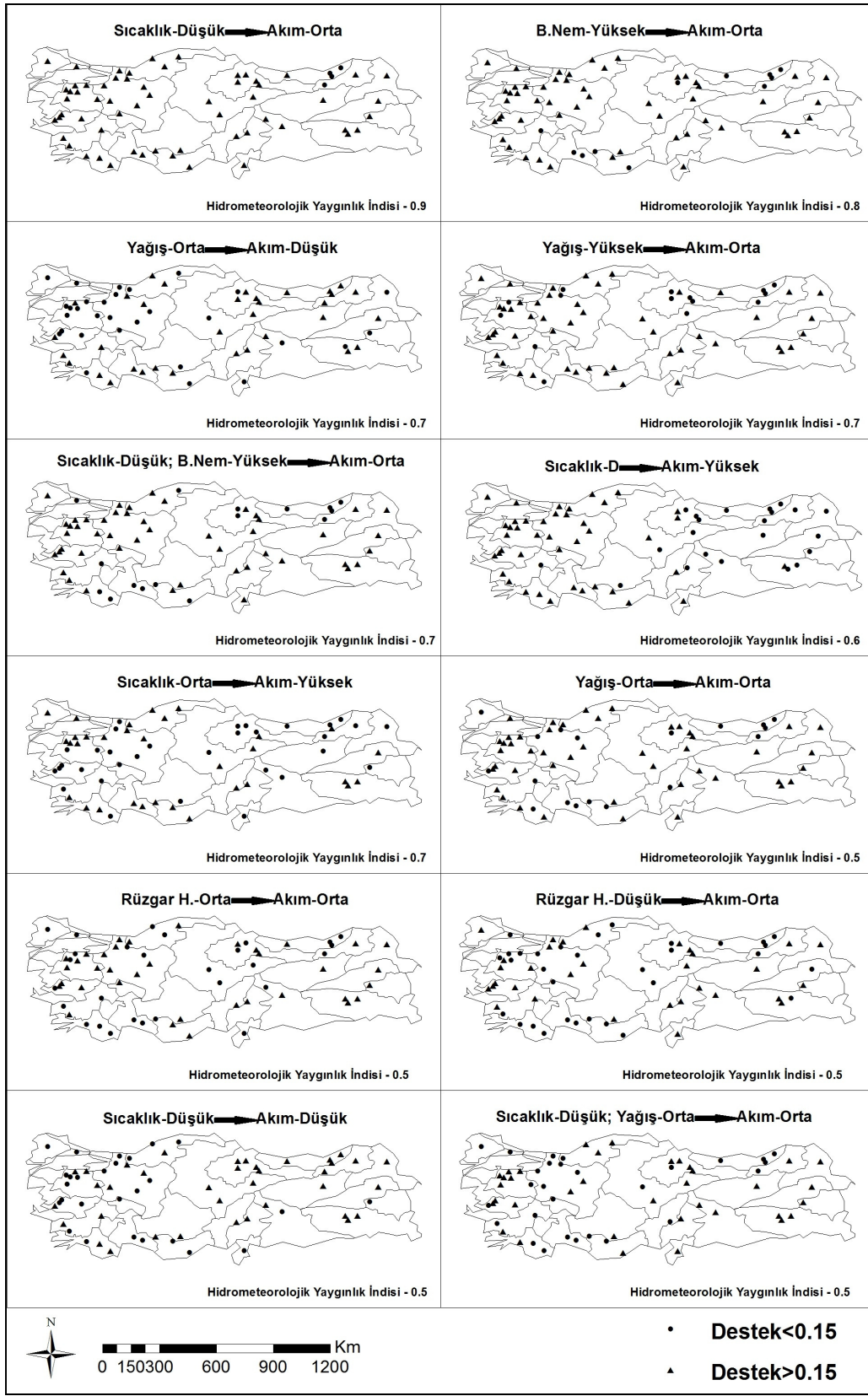


Şekil 4.24. Yaz mevsimi akımları için elde edilen kuralların coğrafi dağılımı (destek: 0.15, hidrometeorolojik yaygınlık ölçütü :0.50)

Kış mevsimi için akımlar için destek sınır değerin 0.15 ve hidrometeorolojik yaygınlık ölçütü sınır değeri 0.50 alınmasıyla üretilen kurallar Tablo 4.20’de ve Şekil 4.25’de gösterilmektedir. Kış mevsimi akımlar için elde edilen kurallar arasında “Sıcaklık-Orta → Akım-Orta” kuralı ve “B.Nem-Yüksek → Akım-Orta” kuralı, hem destek hem de hidrometeorolojik yaygınlık indisi bakımından üst sırada yer almaktadırlar.

Tablo 4.20. Kış mevsimi için elde edilen kurallar (destek: 0.15, hidrometeorolojik yaygınlık ölçütü :0.50). D: Düşük, O: Orta, Y: Yüksek

Destek	0.15		0.20		0.25	
Hidrometeorolojik Yaygınlık Ölçütü	Ön Şart	Sonuç	Ön Şart	Sonuç	Ön Şart	Sonuç
0.50	R. Hızı-D	Akım-O	B. Nem-Y	Akım-O	B. Nem-Y	Akım-O
	R. Hızı-O	Akım-O	Sıcaklık-D	Akım-O	Sıcaklık-D	Akım-O
	B. Nem-Y	Akım-O	Sıcaklık-D	Akım-Y		
	Yağış-O	Akım-D				
	Yağış-O	Akım-O				
	Yağış-Y	Akım-O				
	Sıcaklık-D	Akım-D				
	Sıcaklık-D	Akım-O				
	Sıcaklık-D	Akım-Y				
	Sıcaklık-O	Akım-Y				
	B. Nem-Y; Sıcaklık-D	Akım-O				
	Yağış-O; Sıcaklık-D	Akım-O				
0.60	B. Nem-Y	Akım-O	B. Nem-Y	Akım-O	Sıcaklık-D	Akım-O
	Yağış-O	Akım-D	Sıcaklık-D	Akım-O		
	Yağış-Y	Akım-O	Sıcaklık-D	Akım-Y		
	Sıcaklık-D	Akım-O				
	Sıcaklık-D	Akım-Y				
	B. Nem-Y; Sıcaklık-D	Akım-O				
0.70	B. Nem-Y	Akım-O	B. Nem-Y	Akım-O	Sıcaklık-D	Akım-O
	Yağış-O	Akım-D	Sıcaklık-D	Akım-O		
	Yağış-Y	Akım-O				
	Sıcaklık-D	Akım-O				
	Yağış-O; Sıcaklık-D	Akım-O				
0.80	B. Nem-Y	Akım-O	Sıcaklık-D	Akım-O	Sıcaklık-D	Akım-O
	Sıcaklık-D	Akım-O				
0.90	Sıcaklık-D	Akım-O	Sıcaklık-D	Akım-O	Sıcaklık-D	Akım-O



Şekil 4.25. Kış mevsimi akımları için elde edilen kuralların coğrafi dağılımı (destek: 0.15, hidrometeorolojik yaygınlık ölçütü :0.50)

4.7. Yöntemlerinin ve Örüntülerin Değerlendirilmesi

Yapılan analizler sonucunda birliktelik analizinin ortaya çıkardığı sonuçlar Türkiye’de akım ve meteorolojik faktörler arasındaki ilişkiyi belirlemekte katkı sağlamıştır. Akım parametresinin genel olarak en fazla yağış ve sıcaklıkla ilgili olduğu tespit edilmiştir.

5. TARTIŞMA VE SONUÇ

Bu projenin Türkiye için hidrolojik ve iklimsel parametreler arasındaki ilişkilerin ve bu parametrelerin birlikte oluşturdukları mekan-zamansal örüntülerin tespit edilmesini ve bu ilişki ve örüntülerin tespit edilmesi için verilerin mekansal ve mekan-zamansal özelliklerini göz önüne alan veri madenciliği yöntemlerinin geliştirilmesini amaçlamıştır.

Hidrolojik ve iklimsel verilerin analizleri için bugüne kadar Türkiye’de yapılan çalışmalar beş konu üzerinde yoğunlaşmışlardır. Bunlar 1) verilerin genel karakterlerinin ve mekansal değişimlerinin belirlenmesi, 2) homojenlik analizleri, 3) eğilim analizleri, 4) veri serilerinde büyük ölçekli atmosferik salınım olaylarının (El Nino gibi) etkilerinin araştırılması, ve 5) kümeleme analizleri olarak sınıflandırılabilir. Bunlara karşılık, hidrolojik ve iklimsel parametreler arasındaki ilişkilerin ve birlikte oluşturdukları örüntülerin (zamanla sıklaşan ilişkiler, birbirlerini tetikleyen ilişkiler, periyodik ilişkiler gibi) tespiti için henüz bir çalışma yapılmamıştır. İkinci olarak bugüne kadar yapılan çalışmalar da istatistiksel yöntemler ağırlıklı olarak kullanılmıştır. Fakat istatistiksel yöntemler büyük-ölçekli verilerin analizinde yetersiz kalmaktadır. Üçüncü olarak bugüne kadar yapılan çalışmalar genel olarak verilerin zamansal boyutuna odaklanmış ve mekansal ve mekan-zamansal boyutlarını yeterli düzeyde incelememişlerdir.

Bu projede hidrolojik ve iklimsel veriler beraber analiz edilerek, bu verilerin birlikte oluşturdukları ilişkilerinin ve mekan- zamansal hidrometeorolojik örüntülerin keşfi sağlanacaktır. Bu amaçla veri madenciliği yöntemleri (birliktelik analizi) kullanılacaktır. Veri madenciliği, analizi manuel olarak mümkün olmayan veya analizi çok zaman alan büyük ve çok değişkenli veri kümelerinden daha önce keşfedilmemiş, ilginç ve anlamlı bilgileri çıkarmak için otomatik veya yarı-otomatik yöntemleri geliştirilmiştir. Birliktelik analizi bu kapsamda değişik parametreler arasındaki ilişki ve örüntülerin belirlenmesinde neden-sonuç ilişkilerini de ifade eden kuralların keşfinde kullanılmıştır. Bu çalışmada kullanılan veriler çeşitli devlet kurumlarından (Devlet Meteoroloji İşleri Genel Müdürlüğü-DMİ, Elektrik İşleri Etüd İdaresi Genel Müdürlüğü-EİE) elde edilmiştir. Veri kalitesinin değerlendirilmesi ve veri madenciliği için hazırlanmasından sonra, verilerin mekansal ve mekan-zamansal özellikleri göz önüne alınarak analizleri gerçekleştirilecek ve hidrometeorolojik örüntülerin sınıflandırılması ve tanımlanması yapılacaktır. Tanımlanan örüntülerin keşfi için değerlendirme ölçütleri geliştirmiş ve bu ölçütler kullanılarak ilişkilerin ve örüntülerin

madenciliği için etkin, hızlı ve özgün algoritmalar hazırlanmıştır. Projenin son kısmında geliştirilen yöntemler Türkiye hidrolojik ve iklimsel verileri üzerinde uygulanmıştır. Uygulama için 64 istasyona ait akım verileri ve bu istasyonlarda ölçüden akım verileri ile en yüksek korelasyona sahip meteorolojik parametrelere sahip meteoroloji istasyonundan elde edilmiş yağış, sıcaklık, bağıl nem ve rüzgar hızı verileri kullanılmıştır.

Bu projenin literatüre katkısı iki şekilde olmuştur. Birinci olarak bu çalışmayla Türkiye için hidrolojik ve iklimsel parametreler arasındaki ilişkilerin ve beraber oluşturdıkları örüntülerin keşfi sağlanmıştır. İkinci olarak bu çalışma hidrolojik ve iklimsel parametreler arasındaki ilişkilerin tespiti için yeni ve özgün yöntemler ortaya çıkarmıştır. Bu çalışmada elde edilen sonuçlar Türkiye’de iklim değişikliği konusunda karar-destek mekanizmalarını güçlendirecektir. Hidrolojik ve iklimsel verilerin analizleri için yeni bir yöntem önererek bu verilerin analizlerini daha farklı boyutlara taşıyacaktır. Çalışma sonuçları EİE, DMİ gibi devlet kurumlarına hidrolojik afetlerin önceden tespit edilmesi, iklim değişikliği ve su kaynakları planlaması gibi konularda yol gösterici olacaktır. Bu çalışmada geliştirilecek yöntemlerin diğer uygulama alanlarına (stok verileri, hareketli nesne verileri gibi) uygulanabilmesi mümkündür.

Sonuç olarak bu proje veri madenciliği tekniklerinin hidrolojik ve iklimsel parametreler arasındaki ilişkilerin keşfinde kullanılabilme potansiyel olduğunu göstermiştir.

Proje kapsamında şu ana kadar yapılan çalışmalar çeşitli konferanslarda bildiri olarak sunulmuş ve bilimsel dergilere yayınlanmak üzere gönderilmiştir. Aşağıda bu çalışmalar listelenmektedir.

1. Dadaser-Celik, F., Celik, M., ve Dokuz, A.S. (incelemede). Associations between stream flow and climatic parameters at Kızılırmak River Basin in Turkey. GNEST.
2. Dadaser-Celik, F., Cengiz E. (2012). Correlations of stream flow and climatic variables in Turkey. BALWOIS, 28 Mayıs-2 Haziran 2012, Ohrid, Makedonya.
3. Dadaser-Celik, F., Celik, M., ve Dokuz, A.S. (2011). Associations between stream flow and climatic parameters at Kızılırmak River Basin in Turkey. CEST 2011 – 12th International Conference on Science and Technology, 8-10 Eylül 2011, Rodos, Yunanistan.
4. Celik, M., Dadaser-Celik, F. ve Dokuz, A.S. (2011). Anomaly detection in temperature data using DBSCAN algoritma. INISTA 2011 - International Symposium on Innovations in Intelligent Systems and Applications, 15-18 Haziran 2011, s.91-95.

5. Dadaser-Celik, F. ve Cengiz, E. (2011). Türkiye meteorolojik verilerinin homojenlik analizi. ATMOS 2011 - 5. Atmosferik Bilimler Sempozyumu Bildiriler Kitabı, 27-29 Nisan 2011, s. 601-610.

KAYNAKLAR

1. IPCC, *Climate Change 2007: Synthesis Report. Summary for Policymakers*, 2007, Cambridge University Press: Cambridge, UK.
2. Göktürk, O.M., et al., *Quality control and homogeneity of Turkish precipitation data*. Hydrological Processes, 2008. **22**(16): p. 3210-3218.
3. IPCC, *Climate Change and Water*, 2008, Cambridge University Press,: Cambridge, UK.
4. Tan, P.-N., M. Steinbach, and V. Kumar, *Introduction to Data Mining*2006: Addison-Wesley Companion.
5. Han, J. and M. Kamber, *Data Mining: Concepts and Techniques*2001: Morgan Kaufmann.
6. Cressie, N.A.C., *Statistics for Spatial Data*1991, New York, NY: Wiley and Sons.
7. Turkes, M., U.M. Sumer, and G. Kilic, *Variations and trends in annual mean air temperatures in Turkey with respect to climatic variability*. International Journal of Climatology, 1995. **15**(5): p. 557-569.
8. Turkes, M., *Spatial and temporal analysis of annual rainfall variations in Turkey*. International Journal of Climatology, 1996. **16**(9): p. 1057-1076.
9. Turkes, M., *Influence of geopotential heights, cyclone frequency and Southern Oscillation on rainfall variations in Turkey*. International Journal of Climatology, 1998. **18**(6): p. 649-680.
10. Kutiel, H., T.R. Hirsch-Eshkol, and M. Türkeş, *Sea level pressure patterns associated with dry and wet monthly rainfall conditions in Turkey*. Theoretical and Applied Climatology, 2001. **69**: p. 39-67.
11. Türkeş, M., U.M. Sümer, and G. Kilic, *Persistence and periodicity in the precipitation series of Turkey and*. Climate Research, 2002. **21**(1): p. 59-81.
12. Turkes, M., U.M. Sumer, and G. Kilic, *Observed changes in maximum and minimum temperatures in Turkey*. International Journal of Climatology, 1996. **16**(4): p. 463-477.
13. Tayanc, M., et al., *A comparative assessment of different methods for detecting inhomogeneities in turkish temperature data set*. International Journal of Climatology, 1998. **18**(5): p. 561-578.
14. Kahya, E. and M.C. Karabörk, *The analysis of El Nino and La Nina signals in streamflows of Turkey*. International Journal of Climatology, 2001. **21**(10): p. 1231-1250.
15. Cıgızoğlu, K., *Türkiye genelinde yağış akış ilişkisinin yapay sinir ağları ile modellenmesi*. TÜBİTAK Proje Raporu, Proje No: İÇTAG 1841, 2005.

16. Karabörk, M.C., *Trends in drought patterns of Turkey*. Journal of Environmental Engineering and Science, 2007. **6**(1): p. 45-52.
17. Türkes, M., T. Koc, and F. Sarış, *Spatiotemporal variability of precipitation total series over Turkey*. International Journal of Climatology, 2009. **29**: p. 1056-1074.
18. Tayanç, M., et al., *Climate change in Turkey for the last half century*. Climatic Change, 2009. **94**(3-4): p. 483-502.
19. Kadioglu, M., *Trends in Surface Air Temperature Data Over Turkey*. International Journal of Climatology, 1997. **17**: p. 511-520.
20. Tayanç, M. and H. Toros, *Urbanization effects on regional climate change in the case of four large cities of Turkey*. Climatic Change, 1997. **35**(4): p. 501-524.
21. Türkes, M., U.M. Sümer, and I. Demir, *Re-evaluation of trends and changes in mean, maximum and minimum temperatures of Turkey for the period 1929-1999*. International Journal of Climatology, 2002. **22**(8): p. 947-977.
22. Kahya, E. and S. Kalayci, *Trend analysis of streamflow in Turkey*. Journal of Hydrology, 2004. **289**(1-4): p. 128-144.
23. Türkes, M. and U.M. Sümer, *Spatial and temporal patterns of trends and variability in diurnal temperature ranges of Turkey*. Theoretical and Applied Climatology, 2004. **77**(3-4): p. 195-227.
24. Yıldırım, Y.E., M. Türkes, and M. Tekiner, *Time series anaysis of long-term variations in stream-flow data of some stream-flow stations over the Gediz Basin and in Precipitation of the Akhisar station*. Pakista Journal of Biological Sciences, 2004. **7**(1): p. 17-24.
25. Partal, T. and E. Kahya, *Trend analysis in Turkish precipitation data*. Hydrological Processes, 2006. **20**(9): p. 2011-2026.
26. Kahya, E. and T. Partal, *Is seasonal precipitation decreasing or increasing in Turkey?* Online Journal of Earth Sciences, 2007. **1**(1): p. 43-46.
27. Kutiel, H., et al., *North Sea-Caspian Pattern (NCP) - an upper leveş atmospheric teleconnection affecting the eastern Mediterranean - implications on the regional climate*. Theoretical and Applied Climatology, 2002. **72**: p. 173-192.
28. Karabork, M.C. and E. Kahya, *The teleconnections between the extreme phases of the Southern Oscillation and precipitation patterns over Turkey*. International Journal of Climatology, 2003. **23**(13): p. 1607-1625.
29. Türkes, M. and E. Erilat, *Precipitation changes and variability in Turkey linked to the North Atlantic oscillation during the period 1930-2000*. International Journal of Climatology, 2003. **23**(14): p. 1771-1796.

30. Tatli, H., H.N. Dalfes, and Ş.S. Menten, *Surface air temperature variability over Turkey and its connection to large-scale upper air circulation via multivariate techniques*. International Journal of Climatology, 2004. **25**: p. 331-350.
31. Kutiel, H. and M. Türkeş, *New evidence for the role of the North Sea - Caspian Pattern on the temperature and precipitation regimes in continental Central Turkey*. Geografiska Annaler, Series A: Physical Geography, 2005. **87**(4): p. 501-513.
32. Türkeş, M. and E. Erat, *Climatological responses of winter precipitation in Turkey to variability of the North Atlantic Oscillation during the period 1930-2001*. Theoretical and Applied Climatology, 2005. **81**(1-2): p. 45-69.
33. Karabörk, M.C. and E. Kahya. *The links between the categorized Southern Oscillation indicators and precipitation patterns over Turkey*. in *Hydrology Days 2007*. 2007. Fort Collins, CO, ABD.
34. Martı, A.İ. and E. Kahya. *ENSO effects on mean temperature in Turkey*. in *Hydrology Days 2007*. 2007. Fort Collins, CO, ABD.
35. Kahya, E. and T. Cengiz. *North Atlantic Oscillation signals in the series of Beyşehir lake levels (Turkey)*. in *Hydrology Days 2007*. 2007. Fort Collins, CO, ABD.
36. Türkeş, M. and E. Erat, *Influence of Arctic Oscillation on the variability of winter mean temperatures in Turkey*. Theoretical and Applied Climatology, 2008. **92**: p. 75-85.
37. Karabörk, M.C. and E. Kahya, *The links between the categorised Southern Oscillation indicators and climate and hydrologic variables in Turkey*. Hydrological Processes, 2009. **23**(13): p. 1927-1936.
38. Küçük, M., et al., *North Atlantic Oscillation influences on Turkish lake levels*. Hydrological Processes, 2009. **23**(6): p. 893-906.
39. Unal, Y., T. Kindap, and M. Karaca, *Redefining the climate zones of Turkey using cluster analysis*. International Journal of Climatology, 2003. **23**(9): p. 1045-1055.
40. Kalaycı, S. and E. Kahya, *Assessment of streamflow variability modes in Turkey: 1964-1994*. Journal of Hydrology, 2006. **324**(1-4): p. 163-177.
41. Kahya, E., M.C. Demirel, and T.C. Piechota. *Spatial grouping of annual streamflow patterns in Turkey*. in *Hydrology Days 2007*. 2007. Fort Collins, CO, ABD.
42. Demirel, M.C. and E. Kahya. *Hydrological determination of hierarchical clustering scheme by using small experimental matrix*. in *Hydrology Days 2007*. 2007. Fort Collins, CO, ABD.
43. Kahya, E. and M.C. Demirel, *A comparison of low-flow clustering methods: streamflow grouping*. Journal of engineering and applied sciences, 2007. **2**(3): p. 524-530.

44. Kalayci, S. and E. Kahya. *Comparison of rotated and unrotated principal components of Turkish streamflow*. in *Hydrology Days 2007*. 2007. Fort Collins, CO, ABD.
45. Kahya, E., S. Kalayci, and T.C. Piechota, *Streamflow regionalization: Case study of Turkey*. Journal of Hydrologic Engineering, 2008. **13**(4): p. 205-214.
46. Ganguly, A.R. and K. Steinhäuser. *Data mining for climate change and impacts*. in *International Workshop on Spatial and Spatiotemporal Data Mining, IEEE International Conference on Data Mining*. 2008. Pisa, Italy.
47. Kundzewicz, Z.W. and J.J. Napiórkowski, *Nonlinear models of dynamic hydrology*. Hydrological Sciences, 1986. **31**(2): p. 163-185.
48. Kletti, L.L. and H.G. Stefan, *Correlations of Climate and Streamflow in Three Minnesota Streams*. Climatic Change, 1997. **37**(4): p. 575-600.
49. Groisman, P.Y., R.W. Knight, and T.R. Karl, *Heavy precipitation and high streamflow in the contiguous United States: Trends in the twentieth century*. Bulletin of the American Meteorological Society, 2001. **82**(2): p. 219-246.
50. Burn, D.H. and M.A.H. Elnur, *Detection of hydrologic trends and variability*. Journal of Hydrology, 2002: p. 107-122.
51. Novotny, E.V. and H.G. Stefan, *Stream flow in Minnesota: Indicator of climate change?* Journal of Hydrology, 2007. **334**: p. 319– 333.
52. Dadaser-Celik, F. and H.G. Stefan, *Lake level response to climate in Minnesota.*, 2007, University of Minnesota, St. Anthony Falls Laboratory, Report No: 502.: Minneapolis, MN.
53. Dadaser-Celik, F. and H.G. Stefan, *Lake evaporation response to climate in Minnesota*, 2008, University of Minnesota, St. Anthony Falls Laboratory, Report No: 506.: Minneapolis, MN.
54. Shekhar, S. and S. Chawla, *Spatial Databases: A Tour*2003: Prentice Hall.
55. Shekhar, S., R.R. Vatsavai, and M. Celik, *Spatial and Spatiotemporal Data Mining: Recent Advances*, in *Next Generation of Data Mining*, H. Kargupta, et al., Editors. 2009, CRC Press: Boca Raton, FL, ABD.
56. Kargupta, H., et al., *Next Generation of Data Mining*2009, Boca Raton, FL, ABD: CRC Press.
57. Tan, P.-N., et al., *Finding Spatio-Temporal Patterns in Earth Science Data*, in *KDD 2001 Workshop on Temporal Data Mining*2001: San Francisco, CA, USA.
58. Agrawal, R. and R. Srikant. *Fast Algorithms for Mining Association Rule*. in *20th Int'l Conf. on Very Large Data Bases*. 1994 Santiago, Chile.

59. Han, J., J. Pei, and Y. Yin. *Mining Frequent Patterns Without Candidate Generation in Proceedings of 2000 ACM-SIGMOD International Conference on Management of Data (SIGMOD'00)*. 2000.
60. Shu, H., X. Zhu, and S. Dai, *Mining Association Rules in Geographical Spatio-Temporal Data*. 2008.
61. Tan, P.-N., et al. *Finding Spatio-Temporal Patterns in Earth Science Data*. in *KDD Workshop on Temporal Data Mining*. 2001.
62. Srikant, R. and R. Agrawal, *Mining Sequential Patterns: Generalizations and Performance Improvements*. Proceedings of the Fifth International Conference on Extending Database Technology, 1996.
63. Tadesse, T., et al., *Discovering associations between climatic and oceanic parameters to monitor drought in Nebraska using data-mining techniques*. Journal of Climate, 2005. **18**: p. 1541-1550.
64. Tadesse, T., et al., *Drought monitoring using data mining techniques: A case study for Nebraska, USA*. Natural Hazards, 2004. **33**: p. 137-159.
65. Nagesh Kumar, D., M. Ish, and C.T. Dhanya, *Data mining and it's applications for modelling rainfall extremes*. ISH Journal of Hydraulic Engineering, 2009. **15**: p. 25-51.
66. Dhanya, C.T. and D. Nagesh Kumar, *Data minig for evolution of association rules for droughts and floods in India using climate inputs*. Journal of Geophysical Research, 2009. **114**: p. D02102.
67. Lin, F., *Discovery of teleconnections using data mining technologies in global climate datasets*. Data Science Journal, 2007. **6**: p. 749-755.
68. Mannila, H., H. Toivonen, and A.I. Verkamo, *Discovery of Frequent Episodes in Event Sequences* Data Mining and Knowledge Discovery 1997. **1**(3): p. 259-289.
69. Koperski, K. and J. Han, *Discovery of Spatial Association Rules in Geographic Information Databases*. 4th International Symposium on Large Spatial Databases (SSD95), 1995.
70. Shekhar, S. and Y. Huang. *Discovering Spatial Co-location Patterns: A Summary of Results*. in *Proc. of 7th International Symposium on Spatial and Temporal Databases (SSTD01)*. 2001. L.A., CA,.
71. Conrad, V. and C. Pollak, *Methods in climatology*1950, Cambridge: Harvard University Press. 459
72. Auer, I., et al., *A new instrumental precipitation dataset for the greater alpine region for the period 1800-2002*. International Journal of Climatology, 2005. **25**(2): p. 139-166.

73. Peterson, T., et al., *Homogeneity adjustments of in situ atmospheric climate data: A review*. International Journal of Climatology, 1998. **18**(13): p. 1493-1517.
74. Costa, A.C. and Soares, *Homogenization of climate data: review and new perspectives using geostatistics*. Mathematical Geosciences, 2009. **41**: p. 291-305.
75. Costa, A.C. and A. Soares, *Homogenization of climate data: Review and new perspectives using geostatistics*. Mathematical Geosciences, 2009. **41**: p. 291-305.
76. Wijngaard, J., A. Klein Tank, and G. Können, *Homogeneity of 20th century European daily temperature and precipitation series*. International Journal of Climatology, 2003. **23**(6): p. 679-692.
77. Buishand, T.A., *Some methods for testing the homogeneity of rainfall records*. Journal of Hydrology, 1982. **58**: p. 11-27.
78. Ghysels, E., N.R. Swanson, and M.W. Watson, eds. *Essays in Econometrics, Volume I: Spectral Analysis, Seasonality, Nonlinearity, Methodology, and Forecasting*. 2001, Cambridge University Press.
79. Gomes, M.C., J.J. Gomes, and A.C. Paulo, *Diphtheria, pertussis, and measles in Portugal before and after mass vaccination: A time series analysis*. European Journal of Epidemiology, 1999. **15**: p. 791-798.
80. Kumar, V., et al., *Mining Scientific Data: Discovery of Patterns in the Global Climate System*, in *Joint Statistical Meeting 2001*: Atlanta, Georgia, USA.
81. Pyrcie, R.S., *Hydrological Low Flow Indices and their Uses*, 2004, Watershed Science Centre: Peterborough, Ontario. p. WSC Report No.04-2004.
82. Ester, M., et al. *A density based algorithm for discovering clusters in large spatial databases with noise*. in *KDD-96 Proceedings*. 1996.
83. Kumar, V., et al. *Mining Scientific Data: Discovery of Patterns in the Global Climate System*. in *Joint Statistical Meeting*. 2001.