

**T.C.
ERCIYES ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

**BÜYÜK VERİLERİN ANALİZİ İÇİN COĞRAFI
AĞIRLIKLİ REGRESYON YAKLAŞIMLARININ
GELİŞTİRİLMESİ**

**Hazırlayan
Murat TAŞYÜREK**

**Danışman
Doç. Dr. Mete ÇELİK**

Doktora Tezi

**Eylül 2020
KAYSERİ**

**T.C.
ERCIYES ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

**BÜYÜK VERİLERİN ANALİZİ İÇİN COĞRAFİ
AĞIRLIKLİ REGRESYON YAKLAŞIMLARININ
GELİŞTİRİLMESİ**

**Hazırlayan
Murat TAŞYÜREK**

**Danışman
Doç. Dr. Mete ÇELİK**

Doktora Tezi

**Eylül 2020
KAYSERİ**

BİLİMSEL ETİĞE UYGUNLUK

Bu çalışmadaki tüm bilgilerin, akademik ve etik kurallara uygun bir şekilde elde edildiğini beyan ederim. Aynı zamanda bu kural ve davranışların gerektirdiği gibi, bu çalışmanın özünde olmayan tüm materyal ve sonuçları tam olarak aktardığımı ve referans gösterdiğimi belirtirim.

Murat TAŞYÜREK

İmza:



YÖNERGEYE UYGUNLUK

“Büyük Verilerin Analizi İçin Coğrafi Ağırlıklı Regresyon Yaklaşımlarının Geliştirilmesi” adlı Doktora tezi, Erciyes Üniversitesi Lisansüstü Tez Önerisi ve Tez Yazma Yönergesi’ne uygun olarak hazırlanmıştır.

Tezi Hazırlayan
Murat TAŞYÜREK

Danışman
Doç. Dr. Mete ÇELİK


Bilgisayar Mühendisliği ABD Başkanı
Prof. Dr. Veysel ASLANTAŞ

Doç. Dr. Mete ÇELİK danışmanlığında **Murat TAŞYÜREK** tarafından hazırlanan “**Büyük Verilerin Analizi İçin Coğrafi Ağırlıklı Regresyon Yaklaşımlarının Geliştirilmesi**” adlı bu çalışma, jürimiz tarafından Erciyes Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalında **Doktora** tezi olarak kabul edilmiştir.

04/09/2020

JÜRİ :

Danışman : Doç. Dr. Mete ÇELİK
Üye : Doç. Dr. Serkan ÖZTÜRK
Üye : Doç. Dr. Selçuk ASLAN
Üye : Dr. Öğr. Üyesi İlker DALKIRAN
Üye : Dr. Öğr. Üyesi Ahmet Şakir DOKUZ

ONAY :

Bu tezin kabulü Enstitü Yönetim Kurulunun tarih ve sayılı kararı ile onaylanmıştır.

.... / /2020
Prof. Dr. Mehmet AKKURT
Enstitü Müdürü

TEŞEKKÜR

Tez çalışmam boyunca ilgi ve katkılarıyla beni yönlendiren değerli danışmanım Doç. Dr. Mete ÇELİK'e teşekkürlerimi sunarım. Tez izleme komitemde bulunarak çalışmalarına katkıda bulunan değerli hocalarım Doç. Dr. Serkan ÖZTÜRK ve Dr. Öğr. Üyesi İlker DALKIRAN'a teşekkür ederim.

T.C. Tarım ve Orman Bakanlığı Meteoroloji Genel Müdürlüğü'ne meteoroloji istasyon bilgilerini ve ölçüm verilerini paylaştığı için için teşekkür ederim.

Tüm hayatım boyunca desteklerini arkamda gördüğüm ve haklarını ödeyemeyeceğim anneme, babama ve kardeşlerime şükranlarımı sunarım.

Tez çalışmam boyunca gösterdikleri sabırdan ve verdikleri destekten dolayı eşim Tuğba ile çocuklarım Ahmet Efe ve Mehmet Yiğit'e teşekkür etmeyi bir borç bilirim.

Murat TAŞYÜREK
Eylül 2020, KAYSERİ

BÜYÜK VERİLERİN ANALİZİ İÇİN COĞRAFI AĞIRLIKLIL REGRESYON YAKLAŞIMLARININ GELİŞTİRİLMESİ

Murat TAŞYÜREK

**Erciyes Üniversitesi, Fen Bilimleri Enstitüsü
Doktora Tezi, Eylül 2020**

Danışman : Doç. Dr. Mete ÇELİK

ÖZET

Coğrafi ağırlıklı regresyon (geographically weighted regression - GWR), ekoloji, çevre yönetimi, halk sağlığı, meteoroloji ve turizm gibi birçok uygulama alanında değişen ilişkileri modellemek için yaygın olarak kullanılan yerel mekânsal regresyon tekniğidir. GWR yaklaşımları yüksek bilgi işlem kaynakları ve geniş bellek alanlarına ihtiyaç duymaktadır. Veri boyutu büyüdükçe ve/veya veri güncellemeleri sıklaştıkça bu sorunlar daha kritik hale gelmektedir. Diğer taraftan literatürde yapılan çalışmalar genellikle mekânsal veya mekânsal ve rakımsal veya mekânsal ve zamansal ilişkilerini analiz etmektedir. Fakat gerçek dünyada mekânsal, rakımsal ve zamansal ilişkiler birçok veri kümesinde birlikte bulunmaktadır ve bu ilişkilerin birlikte analizi için önerilmiş bir GWR yaklaşımı bulunmamaktadır.

Bu tez çalışmasında, GWR yöntemi ile analizlerin etkili bir şekilde yapılması için çeşitli teknikler kullanan farklı yaklaşımlar geliştirilmiştir. Geliştirilen ilk yaklaşım, sık güncellenen veri kümelerini analiz etmek için ters en yakın komşu (reverse nearest neighbor - RNN) stratejisini kullanan hesaplama açısından etkili bir GWR algoritmasıdır (RNN-GWR). İkinci yaklaşım, coğrafi ve zamansal ağırlıklı regresyon yaklaşımının (geographically and temporally weighted regression - GTWR) veri boyutu kısıtlamasının üstesinden gelmek için hızlı GTWR yaklaşımıdır (FastGTWR). Üçüncü yaklaşım, veri kümelerindeki mekânsal, rakımsal ve zamansal olarak değişen ilişkileri ele almak için coğrafi, rakımsal ve zamansal ağırlıklı regresyon yaklaşımıdır (4D-GWR). Önerilen yaklaşımların başarımlarını değerlendirmek için sentetik ve gerçek veri kümeleri kullanılarak çeşitli deneyler yapılmıştır. Deneysel sonuçlar, önerilen yaklaşımların literatürdeki benzerlerine göre daha etkili bir şekilde çalıştıklarını göstermiştir.

Anahtar Kelimeler: Veri madenciliği, coğrafi ağırlıklı regresyon, ters en yakın komşu, sık güncellenen veriler, büyük veri, mekânsal-rakımsal-zamansal değişkenlikler

DEVELOPING GEOGRAPHICALLY WEIGHTED REGRESSION APPROACHES FOR ANALYSIS OF BIG DATA

Murat TAŞYÜREK

Erciyes University, Graduate School of Natural and Applied Sciences

Ph.D. Thesis, September 2020

Supervisor: Assoc. Prof. Dr. Mete ÇELİK

ABSTRACT

Geographically weighted regression (GWR) is a local spatial regression technique to model varying relationships in many application domains, such as ecology, environmental management, public health, meteorology, and tourism. GWR approaches require high computing resources and large memory spaces. These issues become more critical as data size grows and/or data updates become frequently. On the other hand, studies in the literature generally analyze spatial, spatial and altitudinal or spatial and temporal relationships. However in the real world, spatial, altitudinal, and temporal relationships exist together in many datasets, and there is not yet a proposed GWR approach to analyze datasets that contain these relationships together.

In this thesis, different approaches using various techniques have been developed in order to perform the analyzes with the GWR method effectively. The first approach is a computationally efficient GWR algorithm (RNN-GWR) that uses the reverse nearest neighbor (RNN) strategy to analyze frequently updated datasets. The second approach is a fast geographically and temporally weighted regression (FastGTWR) to overcome data size restriction of GTWR. The third approach is a geographically, altitudinal, and temporally weighted regression (4D-GWR) approach to handle spatial, altitudinal and temporal nonstationarities in datasets. In order to evaluate the performances of the proposed approaches, various experiments were conducted using synthetic and real datasets. Experimental results showed that the proposed approaches worked more effectively than their counterparts in the literature.

Keywords: Data mining, geographically weighted regression, reverse nearest neighbor, frequently updated data, big data, spatial-altitudinal-temporal nonstationarities

İÇİNDEKİLER

BÜYÜK VERİLERİN ANALİZİ İÇİN COĞRAFI AĞIRLIKLİ REGRESYON YAKLAŞIMLARININ GELİŞTİRİLMESİ

BİLİMSEL ETİĞE UYGUNLUK SAYFASI	i
YÖNERGEYE UYGUNLUK SAYFASI	ii
KABUL VE ONAY	iii
TEŞEKKÜR	iv
ÖZET	v
ABSTRACT	vi
İÇİNDEKİLER	vii
TABLolar LİSTESİ	x
ŞEKİLLER LİSTESİ	xi
KISALTMALAR	xiii

GİRİŞ	1
-----------------	---

1. BÖLÜM

GENEL BİLGİLER VE LİTERATÜR TARAMASI

1.1. Coğrafi Ağırlıklı Regresyon (Geographically Weighted Regression - GWR)	7
1.2. Literatür Taraması	13
1.2.1. Sık Güncellenen Veriler İçin GWR Yöntemleri	15
1.2.2. Zamansal GWR Yöntemleri	16
1.2.3. Çok Boyutlu GWR Yöntemleri	16
1.3. Veri Kümesi	18
1.3.1. Gerçek Veri Kümesi	18
1.3.2. Sentetik Veri Kümesi	19

2. BÖLÜM

RNN-GWR: SIK GÜNCELLENEN VERİLER İÇİN TERS EN YAKIN KOMŞU TABANLI GWR YAKLAŞIMI

2.1. Giriş	20
2.2. Toy-GWR Algoritması	22
2.3. RNN-GWR: Sık Güncellenen Veriler İçin Önerilen Ters En Yakın Komşu Tabanlı GWR Yaklaşımı	25
2.4. GWR Yaklaşımlarının Hesaplama Maliyetlerinin Analizi	30
2.5. Deneysel Değerlendirmeler	33
2.5.1. Veri Kümesi ve Model Ayarları	33
2.5.1.1. Veri Kümesi	34
2.5.1.2. Model Ayarları	35
2.5.2. Deneyler	37
2.5.2.1. Algoritmaların Her Bir İterasyon için Çalışma Zamanları	37
2.5.2.2. Veri Boyutunun Etkisi	39
2.5.2.3. Güncellenme Oranının Etkisi	40
2.5.2.4. Çok Çekirdekli Ortamda Algoritmaların Karşılaştırılması	42
2.5.2.5. Gerçek ve Tahmin Edilen Değerlerin Karşılaştırılması	43
2.6. Sonuçlar	45

3. BÖLÜM

FastGTWR: HIZLI COĞRAFI VE ZAMANSAL AĞIRLIKLIL REGRESYON YAKLAŞIMI

3.1. Giriş	46
3.2. Coğrafi ve Zamansal Ağırlıklı Regresyon	48
3.3. FastGTWR: Önerilen Hızlı Coğrafi ve Zamansal Ağırlıklı Regresyon Yaklaşımı	51
3.4. Deneysel Değerlendirmeler	52
3.4.1. Veri Kümesi ve Model Ayarları	53
3.4.1.1. Veri Kümesi	53
3.4.1.2. Model Ayarları	53
3.4.2. Deneyler	54
3.4.2.1. Yaklaşımların Doğruluğu	54
3.4.2.2. Yaklaşımların Çalışma Zamanı	56

3.4.2.3. Veri Boyutunun Yaklaşımlar Üzerindeki Etkisi	56
3.5. Sonuçlar	58

4. BÖLÜM

4D-GWR: COĞRAFI, RAKIMSAL VE ZAMANSAL AĞIRLIKLIL REGRESYON YAKLAŞIMI

4.1. Giriş	59
4.2. 4D-GWR: Önerilen Coğrafi, Rakımsal ve Zamansal Ağırlıklı Regresyon Yaklaşımı	61
4.2.1. GAWR Yaklaşımı	62
4.2.2. Önerilen 4D-GWR Yaklaşımı	63
4.2.3. Dengeleme Katsayılarının ve Optimum Bant Genişliğinin Bulunması	64
4.2.4. GWR Yaklaşımlarında Hesaplama Maliyeti	66
4.3. Deneysel Değerlendirmeler	67
4.3.1. Veri Kümesi ve Model Ayarları	68
4.3.1.1. Veri Kümesi	68
4.3.1.2. Model Ayarları	69
4.3.2. Deneyler	71
4.3.2.1. Zaman Aralıklarının Etkisi	71
4.3.2.2. Yaklaşımların Çalışma Zamanları	72
4.3.2.3. Veri Boyutunun Yaklaşımlar Üzerindeki Etkisi	74
4.3.2.4. Gerçek ve Tahmin Edilen Değerlerin Karşılaştırılması	75
4.4. Sonuçlar	78

5. BÖLÜM

TARTIŞMA, SONUÇ ve ÖNERİLER

5.1. Tartışma ve Sonuç	79
5.2. Öneriler	81
KAYNAKLAR	82
ÖZGEÇMİŞ	94

TABLolar LİSTESİ

Tablo 2.1.	1.000 ile 1.000.0000 arasında ölçüm noktası kullanan algoritmaların çalışma zamanları (saniye - sn)	39
Tablo 2.2.	Tek çekirdekli bilgisayar üzerinde 200.000 veri kümesinde farklı veri güncellenme oranları için algoritmaların çalışma zamanları (saniye) .	41
Tablo 3.1.	1.000 ile 1.000.0000 arasında ölçüm noktası kullanan algoritmaların çalışma zamanları (saniye - sn)	57
Tablo 4.1.	GWR yaklaşımları için bant genişliği değerleri ve dengeleme katsayıları	70
Tablo 4.2.	Sentetik veri kümeleri için yaklaşımların çalışma süreleri (saniye) . .	75

ŞEKİLLER LİSTESİ

Şekil 1.1.	GWR yaklaşımı için bi-square mekânsal mesafe fonksiyonu	9
Şekil 1.2.	GWR yaklaşımı için gaussian mekânsal mesafe fonksiyonu	10
Şekil 1.3.	Regresyon noktası i_1 ile referans noktası j_1 arasındaki mekânsal mesafe	11
Şekil 1.4.	GWR algoritmasının genel çalışma yapısı	13
Şekil 1.5.	MGM saatlik ölçüm istasyonları	19
Şekil 2.1.	$T1$ zaman diliminde Toy-GWR algoritmasının davranışı	23
Şekil 2.2.	$T2$ ve $T3$ zaman diliminde Toy-GWR algoritmasının davranışı	24
Şekil 2.3.	$T2$ zaman diliminde RNN-GWR algoritmasının davranışı	29
Şekil 2.4.	$T3$ zaman diliminde RNN-GWR algoritmasının davranışı	30
Şekil 2.5.	Hava sıcaklık değişim değerine göre meteoroloji istasyonları sayısı .	35
Şekil 2.6.	01 Aralık 2016'da üç zaman dilimi için en az 1°C sıcaklık değişimine sahip meteoroloji istasyonlarının konumları	36
Şekil 2.7.	Farklı sıcaklık değişimleri için algoritmaların çalışma zamanları . . .	38
Şekil 2.8.	Farklı veri boyutları için algoritmaların çalışma sürelerinin karşılaştırılması	40
Şekil 2.9.	Tek çekirdekli bilgisayar üzerinde farklı veri güncellenme oranları için RNN-GWR ve FastGWR'nin çalışma zamanı karşılaştırması . .	41
Şekil 2.10.	Çok çekirdekli bilgisayar üzerinde farklı veri güncellenme oranları için RNN-GWR ve FastGWR'nin çalışma zamanı karşılaştırması . .	42
Şekil 2.11.	Veri güncellenme oranı %100 olduğunda RNN-GWR ve FastGWR algoritmalarının artan çekirdek sayısına göre hızlanmaları	43
Şekil 2.12.	6:00 - 7:00 zaman diliminde sıcaklık değerleri en az 1°C değişen meteoroloji istasyonlarının tahmini ve gerçek sıcaklık değerleri . . .	44
Şekil 2.13.	6:00 - 7:00 zaman diliminde sıcaklık değerleri en az 1°C değişen meteoroloji istasyonlarının kalıntıları	44

Şekil 3.1.	Regresyon noktası i_1 ile referans noktası j_1 arasındaki mekân-zamansal mesafe	49
Şekil 3.2.	GTWR algoritması	50
Şekil 3.3.	GTWR modeli için parametre seçimi	54
Şekil 3.4.	Yaklaşımların doğruluğu	55
Şekil 3.5.	Yaklaşımların saat bazlı çalışma zamanı	56
Şekil 3.6.	Yaklaşımların çalışma zamanı	57
Şekil 4.1.	Rakım değerlerinin yoğunluğu	69
Şekil 4.2.	Zaman aralığı sayısının etkisi	72
Şekil 4.3.	Yaklaşımların çalışma zamanı karşılaştırmaları	73
Şekil 4.4.	Farklı sayıda ölçüm noktası için GWR yaklaşımlarının çalışma zamanları	76
Şekil 4.5.	Saat 00:00 (24) için 12 Aralık 2016 tarihinde GWR yöntemlerinin sonuçları ve kalıntıları	77

KISALTMALAR

a, b, c, d	: Dengeleme katsayıları (exaggeration coefficients)
AIC	: Akaike bilgi kriteri (Akaike information criterion)
AIC_c	: Düzeltilmiş Akaike bilgi kriteri (corrected Akaike information criterion)
BIC	: Bayesian bilgi kriteri (Bayesian information criterion)
bw	: Bant genişliği (bandwidth)
CV	: Çapraz doğrulama (cross-validation)
D	: Veri kümesi
d_{ij}	: i ve j noktaları arasındaki komşuluk mesafesi
F	: Uygunluk fonksiyonu (fitness function)
$FBGWR$	: Esnek bant genişlikli coğrafi ağırlıklı regresyon (geographically weighted regression with flexible bandwidths)
$FastGTWR$	: Hızlı coğrafi ve zamansal ağırlıklı regresyon (fast geographically and temporally weighted regression)
$FastGWR$	: Hızlı coğrafi ağırlıklı regresyon (fast geographically weighted regression)
$GAWR$	: Coğrafi ve rakım ağırlıklı regresyon (geographically and altitudinal weighted regression)
GCV	: Genelleştirilmiş çapraz doğrulama (generalized cross-validation)
GSS	: Altın kesit arama (golden section search)
$GTWR$	: Coğrafi ve zamansal ağırlıklı regresyon (geographically and temporally weighted regression)
GWR	: Coğrafi ağırlıklı regresyon (geographically weighted regression)
IoT	: Nesnelerin interneti (internet of things)
i	: Regresyon noktası (regression point)
IoT	: Nesnelerin interneti (internet of things)
j	: Referans noktası (reference point)
Kn	: Çekirdek fonksiyonu (kernel function)
Lb	: Alt sınırlar (lower bounds)
m	: Bağımsız değişken sayısı

<i>MGM</i>	: T.C. Tarım ve Orman Bakanlığı Meteoroloji Genel Müdürlüğü
<i>n</i>	: Ölçüm nokta sayısı (observation points)
<i>nD-GSS</i>	: N-boyutlu altın kesit arama (n-dimensional golden section search)
<i>OLS</i>	: Klasik en küçük kareler (ordinary least squares)
<i>PM</i>	: Partikül madde (particulate matter)
<i>PSDM-GWR</i>	: Parametreye özgü mesafe metrikli coğrafi ağırlıklı regresyon (geographically weighted regression model with parameter-specific distance metrics)
<i>Rf</i>	: Referans noktaları (reference points)
<i>RMSE</i>	: Kök ortalama karesel hata (root mean square error)
<i>Rn</i>	: Regresyon noktaları (regression points)
<i>RNN</i>	: Ters en yakın komşu (reverse nearest neighbor)
<i>RNN-GWR</i>	: Sık güncellenen veriler için ters en yakın komşu tabanlı coğrafi ağırlıklı regresyon yaklaşımı (geographically weighted regression approach for frequently updated data)
<i>S</i>	: Şapka matrisi (hat matrix)
<i>t_i</i>	: <i>i</i> noktasının zamanı
<i>Toy-GWR</i>	: Toy coğrafi ağırlıklı regresyon (naive geographically weighted regression)
<i>tr(S)</i>	: Şapka matrisinin izi
<i>u_i</i>	: <i>i</i> noktasının enlemi
<i>Ub</i>	: Üst sınırlar (upper bounds)
<i>uRf</i>	: Güncellenen referans nokta kümeleri
<i>v_i</i>	: <i>i</i> noktasının boyları
<i>W(i)</i>	: Komşuluk ağırlığının diyagonal matrisi
<i>w_{ij}</i>	: Komşuluk ağırlığı (neighborhood weight)
<i>X</i>	: Referans noktalarına ait bağımsız değişkenlerinin matrisi
<i>x_i</i>	: Regresyon noktası için bağımsız değişken
<i>Y</i>	: Referans noktalarına ait bağımlı değişken matrisi
<i>ŷ_i</i>	: Regresyon noktası için tahmin edilen değer
<i>y_i</i>	: Regresyon noktası için bağımlı değişken (ölçüm değeri)
<i>z_i</i>	: <i>i</i> noktasının rakımı

β_{ik}	: Regresyon katsayıları (regression coefficients)
ϵ_i	: i noktası için hata terimi
$\hat{\sigma}$	: Hata teriminin tahmini standart sapması
$2D-GWR$	: İki boyutlu coğrafi ağırlıklı regresyon (two dimensional geographically weighted regression)
$4D-GWR$	: Dört boyutlu coğrafi, rakımsal ve zamansal ağırlıklı regresyon (geographically, altitudinal, and temporally weighted regression)



GİRİŞ

Sensörlerin ve nesnelerin interneti (internet of things - IoT) teknolojilerinin artan kullanımı ile birlikte farklı uygulama alanlarında çok sayıda parametreye sahip sık güncellenen veya gerçek zamanlı mekânsal (veya konuma dayalı) veriler toplanmaktadır [1–5]. Bu veriler büyük veri (big data) karakteristiği göstermektedir. Büyük veri, geleneksel veri işleme uygulamaları tarafından ele alınamayacak kadar büyük veya karmaşık olan veri kümeleri olarak tanımlanır [6–8]. Büyük veriyi oluşturan 3 bileşen vardır. Bu bileşenler sırasıyla hacim (volume), çeşitlilik (variety) ve hız (velocity) olarak kabul edilir [9, 10] Büyük verilere örnek olarak meteorolojik parametreler (hava sıcaklığı, basınç, nem, güneş radyasyonu, yağış) ve çevresel parametreler (hava kalitesi, hava kirliliği, ozon, karbon monoksit, toz konsantrasyonu, gürültü vb.) listelenebilir [11–13]. Ancak ölçüm değeri elde edilmek istenen her noktaya, sensör yerleştirmek çok zor ve maliyetli bir işlemdir. Sensörün yerleştirilmediği bir yerde bir parametre değerine ihtiyaç duyulduğunda, referans alınan komşu noktaların değerlerine göre istenilen noktanın parametre değerinin tahmin edilmesi daha ekonomik ve uygulanabilir bir çözümdür. Belirli noktaların değerleri bilinirse, bilinmeyen noktaların değeri Tobler’e göre coğrafyanın birinci yasası olan “her şey her şey ile ilişkilidir ancak yakın şeyler uzak şeylerden daha ilişkilidir” prensibine dayanarak tahmin edilebilir [14].

Bir noktanın parametre değerlerini tahmin etmek için çeşitli yöntemler bulunmaktadır [15–19]. Coğrafya üzerinde konumu ve özellikleri bilinen referans noktaları temel olarak konumu bilinen diğer noktalar için regresyon yöntemi ile tahmini değerler üreten mekânsal veri manenciliği yöntemlerinden bir tanesi de coğrafi ağırlıklı regresyon (geographically weighted regression - GWR) yöntemidir [15, 19, 20]. GWR yöntemi mekânsal olarak değişen ilişkileri yerel (bölgesel) olarak modellemek için yaygın olarak kullanılan önemli bir tekniktir. GWR tekniği tüm veri kümesi için sadece bir regresyon denklemi üretmek yerine her bir regresyon noktası için ayrı bir denklem

üretim üzerine modellenmiştir ve bu özelliği ile genel çözüm üreten diğer regresyon tekniklerinden farklıdır. GWR yaklaşımında regresyon noktasına (değeri tahmin edilecek nokta) yakın olan referans noktaları (parametre değerleri bilinen ölçüm noktaları) regresyon noktasının parametrelerin tahmininde daha güçlü bir etkiye sahiptir. Bir regresyon noktasının parametrelerinin değerini tahmin etmek için regresyon noktasına komşu olan bütün referans noktalarının komşuluk ağırlığı belirlenir. Bu komşuluk ağırlığı genellikle Öklid mesafesi kullanılarak çekirdek fonksiyonu (gaussian veya bi-square en çok tercih edilenlerdendir) ile belirlenir. Bant genişliği, her yerel regresyon denklemi için kullanılan mesafe metriği veya komşu sayısıdır ve modeldeki yumuşatma derecesini kontrol ettiği için GWR yaklaşımında dikkate alınması gereken en önemli parametredir [21]. Bant genişliği değeri ne kadar büyük olursa ağırlıklar o kadar az olur, regresyon denklemi yerel bir denklemden ziyade genel bir denkleme dönüşür ve parametrelerin yerel olarak değişimi azalır. Bant genişliği değeri küçüldükçe ağırlıklar artar ve parametrelerin yerel olarak değişimi de artar fakat bu durumda az sayıda referans noktası dikkate alınacağından dolayı denklem doğru sonuç vermeyebilir. Bu nedenle GWR yaklaşımlarında bant genişliği değeri çok önemlidir ve veri kümesine uygun olan bant genişliği değeri seçilmelidir. Bant genişliği değeri genellikle kalibrasyon yöntemleri ile bulunmaktadır [22, 23].

GWR yaklaşımları her bir regresyon noktası için farklı katsayılarından oluşabildiğinden bu yaklaşımlar hesaplama maliyeti açısından zordur. Veri kümesinin boyutu büyüdükçe hesaplama maliyeti de artmaktadır. Diğer taraftan, GWR yaklaşımında bir regresyon noktasının parametrelerinin değerinin tahmin edilmesi diğer bir regresyon noktasının parametrelerinin değerinin tahmin edilmesinden bağımsız bir işlemdir ve bu nedenle GWR yaklaşımı dağınık ve/veya paralel olarak çalışabilmektedir [24–26].

GWR yaklaşımı, mekânsal durağanlığı ölçmek için model boyunca regresyon katsayılarının değişmesine izin verir. Mekansal durağanlık, küresel (global) modellerin bazı değişken kümeleri arasındaki ilişkileri açıklayamadığı bir durumdur [27]. Modelin doğası, veri içindeki yapıyı yansıtmak için katsayıların değişmesini gerektirmektedir. GWR yaklaşımı mekânsal durağanlığı modellemesine rağmen mekân-zamansal ilişkilerin olduğu veri kümelerinde zamansal durağanlığı modelleyemez. Bu sorunun üstesinden gelmek için coğrafi ve zamansal ağırlıklı regresyon (geographically and

temporally weighted regression - GTWR) [28–30] yaklaşımı literatüre sunulmuştur. GTWR yaklaşımı veri kümesindeki mekânsal ve zamansal ilişkileri belirlemek için mekân-zamansal bant genişliğini (spatio-temporal bandwidth) kullanarak komşuluk ağırlıklarını oluşturur. GTWR yaklaşımında, GWR yaklaşımında kullanılan mesafe hesaplama yöntemine zaman bilgisi de eklenmektedir. GTWR yaklaşımı mekân-zamansal ilişkileri birlikte içeren veri kümelerinde GWR yaklaşımına göre çok daha iyi modeller üretmesine rağmen [28–30] mekân-zamansal modellerin karmaşıklığı göz önüne alındığında zorluklar daha fazladır ve literatürde bulunan GTWR modelleri tarafından ele alınabilecek maksimum veri kümesi standart masaüstü bilgisayarlarda yaklaşık olarak 15.000 ölçüm noktası olarak tespit edilmiştir [26, 31]. Veri çeşitliliğinin ve boyutunun arttığı günümüz teknoloji çağında en fazla 15.000 ölçüm verisini analiz edebilme özelliği GTWR'nin kullanımı açısından ciddi bir sorun oluşturmaktadır.

GTWR yaklaşımı hem mekânsal (enlem ve boylam gibi) hem de zamansal ilişkileri dikkate alarak GWR yaklaşımından farklı olarak veri kümesini üç boyutta incelemektedir. Diğer bir üç boyutlu yaklaşım ise veri kümelerindeki hem mekânsal hem de rakımsal ilişkilerini dikkate alması için önerilen coğrafi ve rakımsal ağırlıklı regresyon (geographically and altitudinal weighted regression - GAWR) yöntemidir [32]. GAWR yaklaşımı klasik GWR yönteminin komşuluk ağırlığına üçüncü boyut olarak rakım mesafesini ekleyerek oluşturulmuştur. Bu yöntemler, mekânsal varyasyona ek olarak rakımsal veya zamansal varyasyonu dikkate aldığı için klasik GWR'den daha iyi performans gösterir. Bununla birlikte, bu yöntemler veri kümesindeki mekânsal, rakımsal ve zamansal değişiklikliğin birlikte olduğu durumları etkin bir şekilde yakalayamamaktadır.

Tezin Motivasyonu

GWR yaklaşımlarının hesaplama maliyetini azaltmak için literatürde paralel ve dağıtılmış platformlar çalışılmıştır. Harris ve arkadaşları [24] klasik GWR'nin kalibrasyon sürecini hızlandırmak için bir grid tabanlı GWR yaklaşımı geliştirmiştir. Tran ve arkadaşları [25], Spark altyapısından yararlanarak büyük veri kümesini işlemek için dağıtılmış GWR yaklaşımını geliştirmiştir. Pozdnoukhov ve Kaiser [33], akan verileri (streaming data) işlemek için mekânsal olarak dağıtılmış artımlı yerel regresyon modeli önermişlerdir. Önerilen yaklaşım, depolama yönetimi ve yoğun hesaplamaların yönetimi

için MapReduce'un dağıtılmış altyapısını kullanmıştır. Ancak, bu modeller GWR tekniği kullanarak akan verileri işlemek için dağıtılmış ve/veya paralel bilgi işlem kaynakları gerektirir. Ek olarak, Li ve arkadaşları [26], birden fazla CPU çekirdeği üzerinde çalışabilen klasik GWR yaklaşımının paralel bir versiyonu olan FastGWR algoritmasını önermiştir. FastGWR yaklaşımı, klasik GWR yaklaşımlarına göre normal özelliklere sahip bilgisayarlarda iyi performans gösterse de, veri kümesinin sık güncellenip güncellenmediğini dikkate almaz. Sık güncellenen büyük ölçekli veri kümeleri için yeni bir GWR yaklaşımı geliştirmeye ihtiyaç vardır.

Mekân-zamansal ilişkilerin olduğu veri kümelerinde GTWR yaklaşımı GWR yaklaşımına göre çok daha iyi modeller üretmesine rağmen [28–30] mekân-zamansal modellerin karmaşıklığı göz önüne alındığında zorluklar daha fazladır ve literatürde koşturulan GTWR modelleri tarafından ele alınabilecek maksimum veri kümesi standart masaüstü bilgisayarlarda yaklaşık olarak 15.000 ölçüm noktası olarak tespit edilmiştir [26, 31]. Veri çeşitliliğinin ve boyutunun arttığı günümüz teknoloji çağında en fazla 15.000 ölçüm verisini analiz edebilme özelliği GTWR kullanımı açısından ciddi bir sorun oluşturmaktadır ve klasik bilgisayarlarda bile veri boyutu kısıtlamasının üstesinden gelebilecek GTWR yaklaşımlarına ihtiyaç duyulmaktadır.

GWR yaklaşımlarının doğruluğunu artırmak için yapılan çalışmalar genellikle bant genişliklerinin kalibrasyonuna ve farklı mesafe metriklerinin kullanımına odaklanır [22, 23, 34–40]. Bu yaklaşımlar temel olarak veri kümeleri için uygun mesafe metrikleri ve bant genişliği değerlerini bulmaya odaklanmıştır ve veri kümelerinde mekânsal, rakımsal ve zamansal durağanlık olup olmadığı ile ilgilenmezler. Propastin [32] klasik GWR modelini (2D-GWR) genişletmiş ve veri kümesinde eşzamanlı olarak hem mekânsal hem de rakımsal değişiklikleri dikkate alarak GAWR yaklaşımını önermiştir. Ayrıca, GTWR modelleri, Fotheringham vd. [29], Ma vd. [30] ve Huang vd. [28] tarafından veri kümesindeki mekânsal değişiklik dışında zamansal değişikliği de dikkate alınarak önerilmiştir. GAWR kalibre edilmiş üç boyutlu Öklid mesafesini kullanır ve GTWR mekân-zamansal bant genişliğini kullanır. Ancak bu yaklaşımlar tahmin sürecinde mekânsal, rakımsal ve zamansal değişiklikleri aynı anda dikkate almazlar. Veri kümesindeki mekânsal, rakımsal ve zamansal değişen ilişkileri birlikte analiz edebilen bir GWR yaklaşımına ihtiyaç duyulmaktadır.

Tezin Amacı ve Katkısı

Bu tezin amacı farklı tipteki büyük ölçekli veri kümelerinde GWR yaklaşımı ile etkili bir şekilde analiz yapılmasını sağlayacak yeni ve özgün GWR yöntemlerinin geliştirilmesidir. Tez kapsamında ele alınan GWR yaklaşımları ve yapılan çalışmaların literatüre katkıları aşağıda listelenmiştir.

1. RNN-GWR: Sık güncellenen veriler için coğrafi ağırlıklı regresyon yaklaşımı:

Sık güncellenen verileri analiz etmek için ters en yakın komşu (reverse nearest neighbor - RNN) stratejini kullanan hesaplama açısından verimli olan RNN-GWR yaklaşımı geliştirilmiştir. Bu tez kapsamında, RNN-GWR yaklaşımının etkili bir şekilde çalışması için, ters en yakın komşu sorgusu, simetrik matris özelliği ve komşuluk ağırlığı sıfır olan noktaların hesaplama matrislerinden çıkartılması önerilmektedir. Önerilen algoritmanın detayları ve başarımını gösteren deneysel çalışmalar Bölüm 2’de sunulmuştur. RNN-GWR yaklaşımının etkili bir şekilde çözüm üretmesini gösteren bu çalışma derlenerek bir dergide yayımlanmıştır [31].

2. FastGTWR: Hızlı coğrafi ve zamansal ağırlıklı regresyon yaklaşımı:

Literatürde koşturulan coğrafi ve zamansal ağırlıklı regresyon (GTWR) modelleri tarafından ele alınabilecek maksimum veri kümesi standart masaüstü bilgisayarlarda yaklaşık olarak 15.000 ölçüm noktası olarak tespit edilmiştir. Veri boyutu kısıtlamasının üstesinden gelmek için daha hızlı bir yaklaşım olan FastGTWR modeli geliştirilmiştir. Önerilen algoritmanın detayları ve başarımını gösteren deneysel çalışmalar Bölüm 3’te sunulmuştur. FastGTWR yaklaşımının etkili bir şekilde çözümler ürettiğini gösteren bu çalışma derlenerek bir dergiye gönderilmiştir.

3. 4D-GWR: Coğrafi, Rakımsal ve Zamansal Ağırlıklı Regresyon:

Mekânsal, rakımsal ve zamansal ilişkileri aynı anda içeren veri kümelerini analiz etmek için dört boyutlu coğrafi, rakımsal ve zamansal ağırlıklı regresyon yaklaşımı (4D-GWR) geliştirilmiştir. Mekânsal, zamansal ve rakımsal ilişkileri aynı anda yakalayan ve hesaplama maliyeti açısından verimli olan 4D-GWR yaklaşımının detaylarını ve başarımını gösteren deneysel çalışmalar Bölüm 4’te sunulmuştur. 4D-GWR yaklaşımının etkili bir şekilde çalıştığını gösteren bu çalışma derlenerek bir dergiye gönderilmiştir.

Tezin Organizasyonu

Büyük verilerin analizi için coğrafi ağırlıklı regresyon yaklaşımlarının geliştirilmesini ele alan bu tez çalışmasının organizasyonu aşağıdaki gibidir.

Giriş bölümünde, tez konusunun tanıtımı yapılmış, hangi alt konuların çalışılacağı ifade edilmiş, tez kapsamında yapılan çalışmaların literatüre katkıları sunulmuş ve tezin takip eden bölümleri tanıtılmıştır.

Birinci bölümde, gerekli kavram ve tanımlarla beraber coğrafi ağırlıklı regresyon yaklaşımı tanımlanmıştır. Daha sonra literatürde yapılan çalışmalardan bahsedilmiştir. Ayrıca yapılan çalışmalarda kullanılan gerçek ölçüm verilerinden ve sentetik verilerden oluşan veri kümesi de tanıtılmıştır.

İkinci bölümde, sık güncellenen veri kümelerini analiz etmek için RNN stratejisini kullanan hesaplama açısından etkili bir GWR yaklaşımı, RNN-GWR, önerilmiştir. Önerilen algoritmanın kaba kodu verilerek çalışma prensibinden bahsedilmiş ve örnek bir veri kümesi üzerinde çalışma mantığı verilmiştir. Bölüm sonunda, önerilen algoritmanın başarımı test edilerek sunulmuştur.

Üçüncü bölümde, hızlı coğrafi ve zamansal ağırlıklı regresyon yaklaşımı, FastGTWR, önerilmiştir. Önerilen algoritmanın çalışma mantığından bahsedilmiş ve GTWR yaklaşımının nasıl hızlandırıldığı anlatılmıştır. Bölümü sonunda, önerilen algoritmanın başarımı test edilerek sunulmuştur.

Dördüncü bölümde, veri kümelerindeki mekânsal, rakımsal ve zamansal olarak değişen ilişkileri ele almak için coğrafi, rakımsal ve zamansal ağırlıklı regresyon yaklaşımı, 4D-GWR, önerilmiştir. Önerilen algoritmanın kaba kodu verilerek çalışma prensibinden bahsedilmiştir. Bölüm sonunda, önerilen algoritmanın başarımı test edilerek sunulmuştur.

Beşinci bölümde ise tez kapsamında yapılan çalışmalardan elde edilen sonuçlar değerlendirilmiş ve gelecekte yapılabilecek çalışmaların neler olabileceğine dair önerilere yer verilmiştir.

1. BÖLÜM

GENEL BİLGİLER VE LİTERATÜR TARAMASI

Bu bölümde, ilk önce GWR yöntemi ile ilgili temel kavramlar verilerek GWR algoritması sunulmuştur. Daha sonra literatürde yapılan çalışmalardan bahsedilmiştir. Son olarak da bu tez kapsamında kullanılan veri kümeleri hakkında bilgi verilmiştir.

1.1. Coğrafi Ağırlıklı Regresyon (Geographically Weighted Regression - GWR)

Coğrafi ağırlıklı regresyon (geographically weighted regression - GWR) yöntemi, mekânsal olarak değişen ilişkileri yerel olarak analiz etmek için geliştirilen bir regresyon tekniğidir [15]. Regresyon ise birbirine bağımlı ve birbirinden etkilenen iki veya daha fazla değişken arasındaki ilişkiyi bulmak için kullanılan istatistiksel bir yöntemdir [41–44]. Amacı, iki veya daha fazla bağımlı değişkenden yararlanmak suretiyle diğer bilinmeyen bağımlı değişken için yakın bir kestirim yapabilmektir. Regresyon analizinin temelinde gözlenen bir olay değerlendirilirken, hangi olayların etkisi içinde olduğunun araştırılması yatmaktadır. GWR yaklaşımı standart çoklu regresyon modelini temel alır. Doğrusal veya doğrusal olmayan yapı teşkil eden, iki veya daha fazla bağımsız değişkene sahip regresyon metotlarına çoklu regresyon modeli denir ve Eşitlik 1.1 ile ifade edilir.

$$y_i = \beta_{i0} + \sum_{k=1}^m (\beta_{ik} x_{ik} + \epsilon_i) \quad (1.1)$$

GWR yaklaşımı Eşitlik 1.1’de bulunan çoklu regresyon modelini temel alır. Fakat GWR yaklaşımında regresyon katsayıları model boyunca değişmektedir [15]. Klasik GWR yaklaşımı Eşitlik 1.2 ile ifade edilir.

$$y_i(u_i, v_i) = \beta_{i0}(u_i, v_i) + \sum_{k=1}^m (\beta_{ik}(u_i, v_i)x_{ik}(u_i, v_i) + \epsilon_i(u_i, v_i)) \quad (1.2)$$

Eşitlik 1.1 ve Eşitlik 1.2’de i , regresyon noktasını, (u_i, v_i) , regresyon noktasının koordinatını, y_i , regresyon noktası için bağımlı değişkeni, x_i ($k = 1, \dots, m$), regresyon noktasının bağımsız değişkenlerini, β_{ik} , regresyon katsayılarını ve son olarak da ϵ_i , hata terimini ifade eder. Regresyon katsayısı β_{ik} ($k = 0, \dots, m$) aşağıdaki eşitlik ile hesaplanır [15].

$$\beta(i) = (X^T W(i) X)^{-1} X^T W(i) Y \quad (1.3)$$

Eşitlik 1.3’te bulunan X matrisi Eşitlik 1.4’te gösterildiği gibi referans noktalarına ait bağımsız değişkenlerinin matrisini temsil eder, Y matrisi Eşitlik 1.5’te gösterildiği gibi $n \times 1$ boyuttan oluşmak üzere referans noktalarına ait bağımlı değişken matrisini temsil eder ve $W(i)$ Eşitlik 1.6’da gösterildiği gibi w_{ij} komşuluk ağırlıklarının diyagonal matrisini temsil eder [39].

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & \dots & \dots & \vdots \\ \vdots & \dots & \dots & \vdots \\ x_{n1} & \dots & \dots & x_{nm} \end{bmatrix} \quad (1.4)$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad (1.5)$$

$$W(i) = \begin{bmatrix} w_{i1} & \dots & \dots & 0 \\ 0 & w_{i2} & 0 & \vdots \\ \vdots & 0 & \dots & \vdots \\ 0 & \dots & \dots & w_{in} \end{bmatrix} \quad (1.6)$$

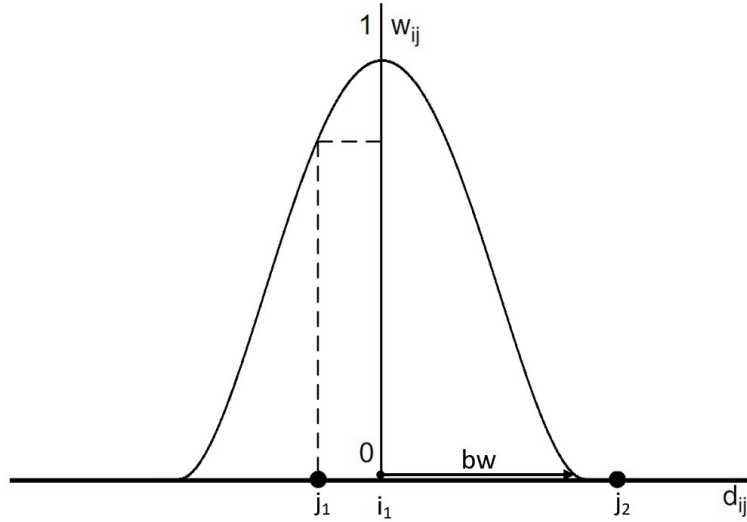
$W(i)$ matrisini oluşturan w_{ij} , regresyon noktası i ile referans noktası j arasındaki komşuluk ağırlığı (neighborhood weight) olarak adlandırılmaktadır. w_{ij} değeri küresel model (global model), box-cor, üstel (exponential), gaussian, bi-square ve tri-cube gibi

çekirdek fonksiyonları kullanılarak hesaplanabilir [45]. GWR yönteminde gaussian ve bi-square en çok kullanılan çekirdek fonksiyonlarıdır [34,35].

Bi-square çekirdek fonksiyonunda w_{ij} değerleri Eşitlik 1.1 ile hesaplanır.

$$w_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{bw}\right)^2\right]^2, & \text{eğer } d_{ij} < bw \\ 0, & \text{diğer durumda} \end{cases}$$

Şekil 1.1 bi-square çekirdek fonksiyonu kullanıldığında oluşan komşuluk ağırlığının grafiğini göstermektedir. Eşitlik 1.1 ve Şekil 1.1’de de görüldüğü gibi regresyon noktası ile referans noktası arasındaki mesafe d_{ij} , bant genişliği değeri bw ’den daha büyük ise bu durumda iki nokta arasındaki komşuluk ağırlığı 0 olur ve regresyon noktasının parametre değerleri tahmin edilirken bant genişliğinden daha uzakta olan referans noktaları dikkate alınmaz. Şekil 1.1’de j_2 referans noktası ile i_1 regresyon noktası arasındaki komşuluk mesafesi bant genişliği değerinden daha büyük olduğu için bi-square çekirdek fonksiyonuna göre komşuluk ağırlığı 0 olmaktadır.

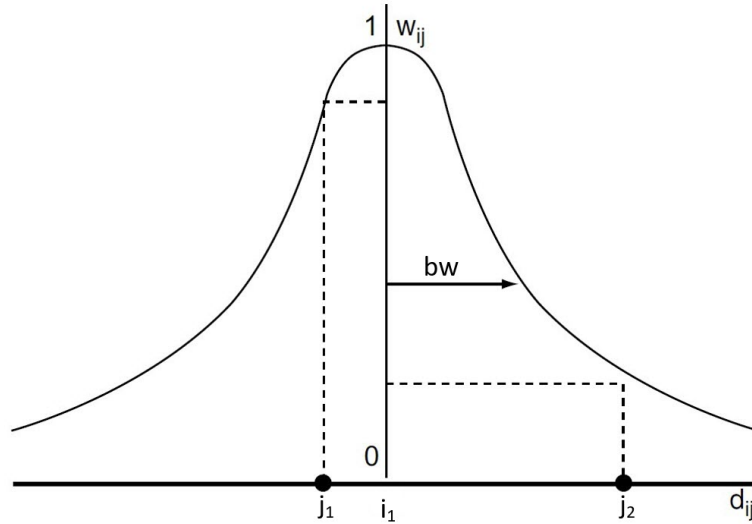


Şekil 1.1. GWR yaklaşımı için bi-square mekânsal mesafe fonksiyonu

Gaussian çekirdek fonksiyonunda w_{ij} değerleri ise Eşitlik 1.7 ile hesaplanır.

$$w_{ij} = \exp \left[- \left(\frac{d_{ij}}{bw} \right)^2 \right] \quad (1.7)$$

Şekil 1.2 gaussian çekirdek fonksiyonu kullanıldığında oluşan komşuluk ağırlığının grafiğini göstermektedir. Eşitlik 1.7 ve Şekil 1.2’de de görüldüğü gibi referans noktaları regresyon noktasından uzaklaştıkça referans noktasının regresyon noktasının parametre değerleri üzerindeki etkisi azalır fakat bu etki bi-square çekirdek fonksiyonun aksine hiç bir zaman 0 olmaz. Yani regresyon noktası ile referans noktası arasındaki mesafe d_{ij} bant genişliği değeri bw ’den daha küçük olsa dahi gaussian çekirdek fonksiyonu kullanıldığı zaman referans noktaları dikkate alınır. Şekil 1.2’de j_2 referans noktası ile i_1 regresyon noktası arasındaki komşuluk mesafesi bant genişliği değerinden daha büyük olmasına rağmen gaussian çekirdek fonksiyonuna göre komşuluk ağırlığının değeri Eşitlik 1.7 ile hesaplanır.

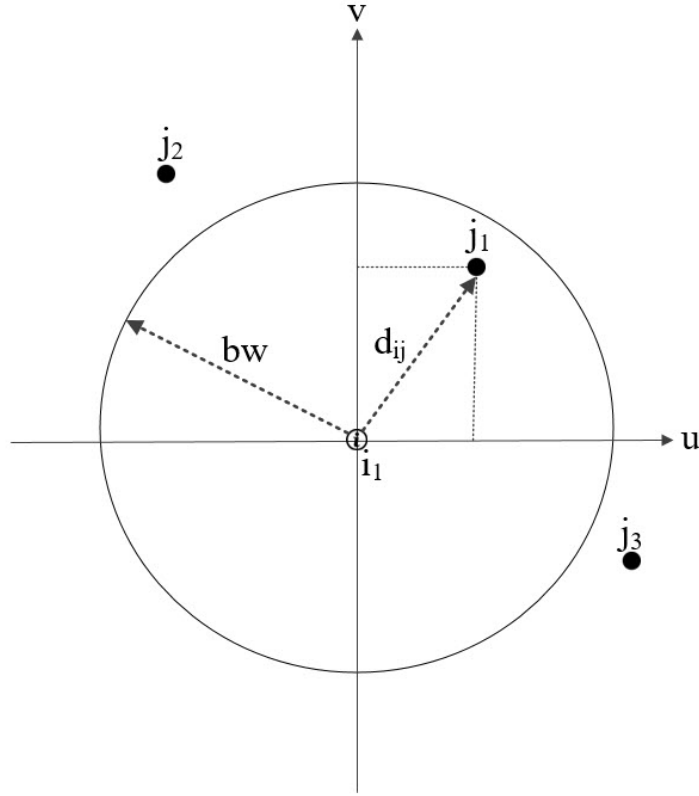


Şekil 1.2. GWR yaklaşımı için gaussian mekânsal mesafe fonksiyonu

Eşitlik 1.1 ve Eşitlik 1.7’de bulunan i regresyon noktası ile j referans noktası arasındaki d_{ij} mesafesinin hesabı için Öklid mesafesi en yaygın kullanılan yöntemdir [15, 26]. i regresyon noktası ile j referans noktası arasındaki Öklid mesafesi Eşitlik 1.8 ile hesaplanır.

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \quad (1.8)$$

Eşitlik 1.8’de u bir noktanın enlemini v ise bir noktanın boylamını göstermektedir. Regresyon noktası i_1 ile referans noktası j_1 arasındaki mekânsal mesafe değeri d_{ij} Öklid mesafesi ile hesaplandığında oluşan grafik Şekil 1.3’te gösterilmiştir. Şekil 1.3’te, bant genişliği değeri bw regresyon noktasının komşuluk yarıçapını göstermektedir.



Şekil 1.3. Regresyon noktası i_1 ile referans noktası j_1 arasındaki mekânsal mesafe

GWR çalışmalarında bant genişliği bw değeri veri kümesindeki konumların dağılım yoğunluklarına göre bütün veri kümesi için sabit bir değer olabilir veya adaptif bir değer olabilir [34, 35, 37]. Veri kümesindeki konum dağılımının eşit olduğu durumlarda bant genişliği parametresinin sabit veya adaptif olması modelin doğruluğunu etkilemez. Bant genişliği parametresinin nokta dağılımına bağlı olarak değişkenliği modelin doğruluğunu artırır [46]. Bant genişliği parametresi en yakın komşuların sayısı seçilir ise ve bu değer bütün bir veri kümesi boyunca sabit olur ise bu model adaptif bant genişliği olarak da adlandırılır [34, 35, 37].

Çapraz doğrulama (cross-validation - CV), genelleştirilmiş çağraz doğrulama (generalized cross-validation - GCV), Akaike bilgi kriteri (Akaike information criterion - AIC) ve Bayesian bilgi kriteri (Bayesian information criterion - BIC) yöntemi bir veri kümesi için en uygun bw değerini bulmak için kullanılabilir [22, 23]. Düzeltilmiş Akaike bilgi kriteri (corrected Akaike information criterion - AIC_c) yaygın olarak kullanılan kalibrasyon yöntemlerinden biridir [15, 39, 47] ve Eşitlik 1.9 ile ifade edilir.

$$AIC_c = 2n \ln(\hat{\sigma}) + n \ln(2\pi) + n \left\{ \frac{n + \text{tr}(S)}{n - 2 - \text{tr}(S)} \right\} \quad (1.9)$$

Eşitlik 1.9’da, n ölçüm sayısını, $\hat{\sigma}$ hata teriminin tahmini standart sapmasını ve $\text{tr}(S)$ şapka matrisi olarak bilinen S matrisinin izini gösterir [48]. Şapka matrisindeki bulunan her bir satır Eşitlik 1.10 ile ifade edilir.

$$S_i = X_i(X^T W(i) X)^{-1} X^T W(i) \quad (1.10)$$

Eşitlik 1.9’da bulunan σ ise Eşitlik 1.11 ile ifade edilir.

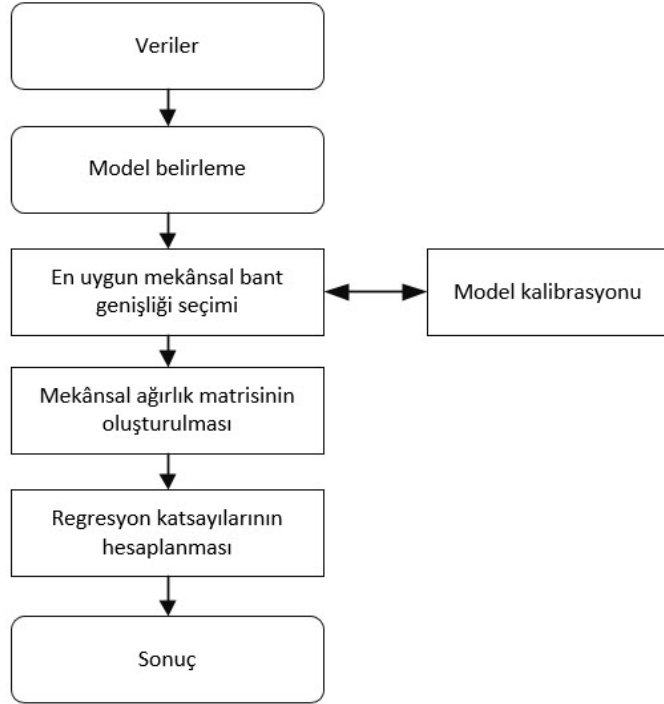
$$\sigma^2 = \sum_{i=1}^n (y_i + \hat{y}_i) / (n - 2\text{tr}(S) + \text{tr}(S^T S)). \quad (1.11)$$

Eşitlik 1.11’de $\hat{y}_i$ regresyon noktasının tahmini değerini ifade etmektedir.

Yaygın olarak kullanılan kalibrasyon yöntemlerinden bir diğeri olan çapraz doğrulama (cross-validation - CV) [15, 39, 47], $y_{\neq 1}(bw)$, bant genişliği bw olan GWR modelinden tahmin edilen değer olduğunu varsayarsak Eşitlik 1.12 ile ifade edilir.

$$CV(bw) = \sum_i [y_i - \hat{y}_{\neq 1}(bw)]^2 \quad (1.12)$$

GWR algoritmasının genel çalışma yapısı Şekil 1.4’te gösterilmiştir [29]. GWR çalışmalarında öncelikle veri kümesine uygun GWR modeli seçilmelidir. Şekil 1.4’de görüldüğü gibi giriş veri kümesine uygun GWR modeli seçilir ve Eşitlik 1.2 kullanılarak GWR modeli oluşturulur, hangi parametrelerin bağımsız ve hangi parametrenin bağımlı değişken olacağına karar verilir. Daha sonra, model için en uygun mekânsal bant genişliği belirlenir. En uygun bant genişliği değeri kalibrasyon yöntemleri ile bulunur. Genellikle CV veya AIC kalibrasyon yöntemi kullanılmaktadır. Bu işlemten sonra mekânsal bant genişliği kullanılarak ağırlık matrisi oluşturulur (Eşitlik 1.6). Oluşturan bu ağırlık matrisi, bağımsız değişkenlerin matrisi ve bağımlı değişkenlerin matrisi kullanılarak Eşitlik 1.3 ile regresyon katsayıları hesaplanır. Daha sonra bu regresyon katsayıları kullanılarak regresyon noktasının tahmini değeri bulunur.



Şekil 1.4. GWR algoritmasının genel çalışma yapısı

1.2. Literatür Taraması

GWR yaklaşımı mekânsal olarak değişen ilişkileri yerel olarak modellemek için yaygın kullanılan bir tekniktir [15, 39, 49]. Bu nedenle GWR modelinin verimliliğini ve doğruluğunu arttırmak için çeşitli çalışmalar yapılmıştır [20, 22, 24, 34, 37, 46, 50–55].

GWR yöntemleri çeşitli uygulama alanlarında kullanılmıştır. GWR yöntemi, Güney Avrupa’da yangın yoğunluğunun mekânsal modellerini keşfetmek için kullanılmıştır [56]. Wang ve arkadaşları [57] bölgesel olarak toplam toprak azotunu tahmin etmek için GWR’yi kullanmıştır. Gilbert ve Chakraborty [58] kümülatif kanser risklerini tahmin etmek için GWR kullanmıştır. Hsueh ve arkadaşları [59] Tayvan’ın Kaoshiung kentinde dang humması vakalarının mekân-zamansal modellerini bulmak için GWR’yi kullanmıştır. Zhang ve arkadaşları [60], GWR sayesinde otel oda fiyatları modeli oluşturmuştur. Liu ve arkadaşları [4] kamyon şoförü yaralanma şiddeti ve geçiş kontrol alanları arasındaki yerel ilişkileri ortaya çıkarmak için GWR’yi kullanmıştır. Lin ve Cromley [61] alansal enterpolasyon problemlerini çözmek için GWR ve kantil regresyon (quantile regression - QR) tabanlı enterpolatörleri entegre ederek çok kategorili bir

yöntem önermiştir.

GWR çalışmalarında öncelikle veri kümesine uygun GWR modeli seçilmelidir. Sürekli veri (continuous data) olarak adlandırılan ölçülebilen nicel veri tipi içeren veri kümeleri için klasik GWR yaklaşımı uygundur. Ayırık veri kümesi (discrete data) olarak adlandırılan sadece belirli değerler alabilen veri tipini içeren veri kümeleri için coğrafi ağırlıklı genelleştirilmiş doğrusal model (geographically weighted generalized linear) ve coğrafi ağırlıklı lojistik regresyon modeli (geographically weighted logistic regression) kullanılmaktadır [35, 46, 62–64]. Oran ve orantıları içeren veri kümeleri için coğrafi ağırlıklı beta regresyon modeli (geographically weighted beta regression) tercih edilmektedir [19]. Sıralı kategorik sonuçlar içeren veri kümeleri için ise coğrafi ağırlıklı sıralı regresyon modeli (geographically weighted ordinal regression) kullanılmaktadır [19]. Veri kümesinde bazı ilişkiler sabittir ve parametre katsayılarının da sabit olması gerekir. Bu tür veri kümelerinin analizi için bazı katsayıların model boyunca sabit kalmasını diğerlerinin değişmesini sağlayan karma coğrafi ağırlıklı regresyon yaklaşımı (mixed geographically weighted regression) Brundson vd. [65] ve Wei vd. [53] tarafından geliştirilmiştir. Veri kümesinde ilişkiler parametre bazında değişebilmektedir ve bu ilişkileri içeren veri kümelerini analiz etmek için, Leong ve Yue [52] koşullu coğrafi ağırlıklı regresyon (conditional geographically weighted regression) yaklaşımını, Yang [66] esnek bir bant genişliği tabanlı GWR (geographically weighted regression with flexible bandwidths - FBGWR) yaklaşımını ve Lu vd. [38–40] her bağımsız değişken için farklı bant genişlikleri seçilmesine imkan veren parametreye özgü mesafe metrikleri tabanlı modeli (geographically weighted regression model with parameter-specific distance metrics - PSDM-GWR) önermiştir.

Veri kümesine uygun GWR modeli seçildikten sonra çekirdek fonksiyonunu, çekirdek tipi, bant genişliği değeri ve bant genişliği kalibrasyon tekniği belirlenmelidir. Gaussian ve bi-square en yaygın kullanılan çekirdek fonksiyonlarıdır [34–36]. Çekirdek tipi ölçüm noktalarının mekânsal dağılımına göre sabit veya adaptif olabilir [34, 35, 37]. Bant genişliği değeri çok önemlidir ve çapraz doğrulama (cross-validation - CV), genelleştirilmiş çapraz doğrulama (generalized cross-validation - GCV), Akaike bilgi kriteri (Akaike information criterion - AIC) ve Bayesian bilgi kriteri (Bayesian information criterion - BIC) yöntemi bir veri kümesi için en uygun *bw* değerini bulmak

için kullanılan yöntemlerdir [22,23]. Bant genişliği değerini bulmak için genellikle Öklid mesafesi kullanılır. Lu ve arkadaşları [67] Öklid mesafesi yerine en kısa yol mesafesini kullanmıştır ve GWR yaklaşımının kalibrasyonunda önemli bir gelişme elde etmiştir. Lu ve arkadaşları [20] ev fiyatları analizinde Öklid mesafesi yerine en kısa yol mesafesini ve zaman metriğini birlikte kullanmıştır ve bu sayede GWR yaklaşımında önemli bir gelişme elde ederek ev fiyatları modeline dair yeni ilginç bulgular keşfetmiştir. Lu ve arkadaşları [50] özenle seçilmiş mesafe metriğinin GWR yaklaşımının doğruluğunu önemli ölçüde artırdığını keşfetmiştir ve bu nedenle GWR yaklaşımında veri kümesine en uygun mesafe metriğinin seçimi için Minkowski yaklaşımını önermiştir.

1.2.1. Sık Güncellenen Veriler İçin GWR Yöntemleri

Literatürde GWR yaklaşımlarının hesaplama maliyetini azaltmak için paralel ve dağıtılmış platformlar üzerinde çalışmalar yapılmıştır. Harris ve arkadaşları [24] klasik GWR yaklaşımının kalibrasyon sürecini hızlandırmak için grid tabanlı yeni bir GWR yaklaşımı geliştirmiştir. Tran ve arkadaşları [25] Spark kütüphanesinden yararlanarak büyük veri kümesini işlemek için dağıtılmış şekilde çalışan yeni bir GWR yaklaşımı geliştirmiştir. Pozdnoukhov ve Kaiser [33] sürekli olarak gelen yeni verileri (streaming data) yakalamak için mekânsal olarak dağıtılmış artırimsal lokal regresyon modeli önermiştir. Pozdnoukhov ve Kaiser [33] tarafından önerilen yaklaşım, depolama ve yoğun hesaplamaların yönetimi için MapReduce kütüphanesinin dağıtılmış bilgi işlem alt yapısını kullanmıştır. Fakat bu çalışmalar sürekli akan verilerini işlemek için GWR yaklaşımını dağıtılmış ve / veya paralel formasyonda çalıştırmaya ihtiyaç duyar ve bu nedenle yüksek özelliklere sahip bilgi işlem kaynakları gerektirir. Bu çalışmalardan farklı olarak Li ve arkadaşları [26] klasik GWR yaklaşımının paralel versiyonu olan birden fazla CPU çekirdeği üzerinde eş zamanlı olarak çalışabilen FastGWR algoritmasını önermiştir. FastGWR yaklaşımı, klasik GWR yaklaşımlarına göre normal özelliklere sahip standart bir masaüstü bilgisayarlarda bile iyi bir performans sergilemesine rağmen, veri kümesinin sık güncellenip güncellenmediğini de dikkate almaz.

Klasik bir bilgisayarda eş zamanlı ve paralel olarak çalışmaya ihtiyaç duymadan sık güncellenen büyük ölçekli verileri işleyebilmek için yeni bir GWR yaklaşımına ihtiyaç

vardır.

1.2.2. Zamansal GWR Yöntemleri

Coğrafi ve zamansal ağırlıklı regresyon (geographically and temporally weighted regression - GTWR) yaklaşımı ise Huang ve arkadaşları [28] tarafından emlak piyasası verilerinde eşzamanlı ve mekânsal duraksama ile başa çıkmak için GWR modeline zamansal etkiler ekleyerek genişletilmiş bir GWR yaklaşımıdır. Standart GWR modelinden farklı olarak GTWR, mekânsal ve zamansal heterojenliği yakalamak için ağırlık matrislerine hem mekânsal hem de zamansal bilgileri entegre eder. GTWR yaklaşımı mekân-zamansal bilgilerinin olduğu birçok veri kümesine uygulanabilen yaygın bir yaklaşımdır. Örneğin, Fotheringham ve arkadaşları [29] 1980-1998 yılları arasında Londra'da 19 yıllık bir dizi ev fiyat verisi kullanarak GTWR'nin ev fiyat modellemesinden elde edilen deneysel sonuçlar ile GTWR modelinin geleneksel GWR yaklaşımına olan üstünlüğünü göstermiştir. Ma ve arkadaşları [30] Pekin'de bir aylık transit akıllı kart ve trafik analiz bölgesi düzeyinde yapılan deneysel çalışmada klasik en küçük kareler yöntemi (ordinary least squares - OLS) ve GWR modellerine kıyasla, GTWR için çok daha iyi bir uyum sağladığını gözlemlemiştir. Chu ve arkadaşları [68] GTWR yaklaşımını uydu görüntülerinden su kalitesini tahmin etmek için kullanmıştır ve GWR ile karşılaştırıldığında, GTWR yaklaşımının su kalitesi haritalamasında son derece güvenilir bilgi sağladığını tespit etmiştir. Veri kümesinde mekân-zamansal bilgilerin olduğu durumlarda GTWR yaklaşımı GWR yaklaşımına göre çok daha iyi sonuçlar üretmektedir [28–30]. GTWR yaklaşımı GWR yaklaşımına göre çok daha iyi modeller üretmesine rağmen mekânsal-zaman modellerinin karmaşıklığı göz önüne alındığında zorluklar daha fazladır [30, 69, 70] ve literatürde koşturulan GTWR modelleri tarafından ele alınabilecek maksimum veri kümesi standart masaüstü bilgisayarlarda yaklaşık olarak 15.000 ölçüm noktası olarak tespit edilmiştir [26, 31].

Klasik bir bilgisayarda mekân-zamansal verileri içeren büyük ölçekli verileri işleyebilmek için yeni bir GTWR yaklaşımına ihtiyaç vardır.

1.2.3. Çok Boyutlu GWR Yöntemleri

GWR yaklaşımlarının doğruluğunu artırmak için yapılan çalışmalar genellikle bant genişliklerinin kalibrasyonuna ve farklı mesafe metriklerinin kullanımına odaklanmaktadır [22, 23, 34–37]. Lu vd. [67], GWR kalibrasyonunu iyileştirmek için sabit bir çekirdek ile Öklid mesafesi yerine en kısa yol mesafesini kullanmıştır. Lu vd. [20], konut fiyatı veri kümesindeki ilişkileri modellemek için GWR'deki Öklid mesafesi yerine yol ağı mesafesi ve seyahat süresi metriklerini kullanmıştır. Lu vd. [50], GWR modeli için optimum mesafe metriğinin seçilmesini sağlayan bir Minkowski yaklaşımı önermiştir. Yang [66] esnek bir bant genişliği tabanlı GWR (geographically weighted regression with flexible bandwidths - FBGWR) yaklaşımını önermiştir. Lu vd. [38–40] tarafından, her bağımsız değişken için farklı bant genişlikleri seçilmesine imkan veren parametreye özgü mesafe metrikleri tabanlı model (geographically weighted regression model with parameter-specific distance metrics - PSDM-GWR) önerilmiştir. FBGWR ve PSDM-GWR bant genişliklerini kalibre etmek için geri beslemeli algoritma kullanmaktadır ve bu nedenle bu algoritmaların hesaplama maliyetleri oldukça yüksektir. Yang vd. [71] tarafından önerilen modifiye edilmiş GWR (modified GWR) yaklaşımı komşuluk mesafesindeki enlem ve boylam ağırlıklarının uygun değerlerini belirlemek için bütün olası ihtimalleri denemektedir. Fakat bu yaklaşımlar temel olarak veri kümeleri için uygun mesafe metrikleri ve bant genişliği değerlerini bulmaya odaklanmıştır ve veri kümelerinde mekânsal, rakımsal ve zamansal özellik olup olmadığı ile ilgilenmezler.

Propastin [32] klasik GWR modelini (2D-GWR) genişletmiş ve veri kümesinde eşzamanlı olarak hem mekânsal hem de rakımsal değişiklikleri dikkate alarak coğrafi ve rakım ağırlıklı regresyon modelini, GAWR, önermiştir. Ayrıca, coğrafi ve zamansal ağırlıklı regresyon modelleri, GTWR, Fotheringham vd. [29], Ma vd. [30] ve Huang vd. [28] tarafından veri kümesindeki mekânsal özelliğe ek olarak zamansal özellik de dikkate alınarak önerilmiştir. GAWR kalibre edilmiş üç boyutlu Öklid mesafesini kullanmaktadır. GTWR mekân-zamansal bant genişliğini kullanmaktadır. Ancak bu yaklaşımlar, tahmin sürecinde mekânsal, rakımsal ve zamansal özellikleri aynı anda dikkate almazlar.

Literatürde klasik GWR (2D-GWR) yaklaşımları, optimum bant genişliği değerini belirlemek için genellikle tek boyutlu altın kesit arama yöntemini (golden section search - GSS) kullanır [26, 72–74]. GAWR, GTWR ve modifiye edilmiş GWR yaklaşımları çapraz doğrulama veya kaba kuvvet yaklaşımları kullanır ve bu nedenle optimum bant genişliği

değerlerini belirlemek için tüm olasılıkları denemektedirler. Bu nedenle, GAWR, GTWR ve modifiye edilmiş GWR yaklaşımları için optimum bant genişliği bulmanın hesaplama maliyeti çok yüksektir.

Bahsedilen eksiklikleri gidermek ve tahmin doğruluğunu artırmak için veri kümesindeki mekânsal, rakımsal ve zamansal özellikleri birlikte dikkate alan yeni bir GWR yaklaşımına ihtiyaç vardır.

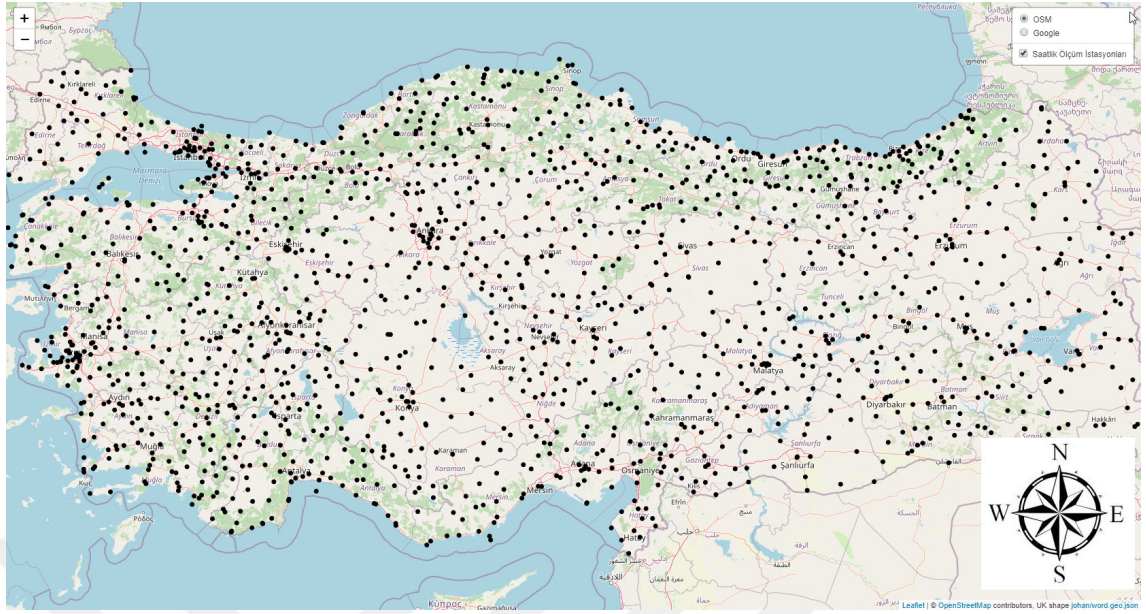
1.3. Veri Kümesi

Bu bölümde, ilk olarak yöntemlerin başarımlarını değerlendirmek için T.C. Tarım ve Orman Bakanlığı Meteoroloji Genel Müdürlüğü'nden (MGM) alınan gerçek veri kümesi tanıtılmış daha sonra da bu veri kümesinden oluşturulan sentetik veri kümesi sunulmuştur.

1.3.1. Gerçek Veri Kümesi

Bu tez kapsamında MGM'den alınan 1746 adet, saatlik ölçüm yapan istasyon bilgilerini ve gerçek ölçüm verilerini içeren meteoroloji veri kümesi kullanılmıştır. Veri kümesi 2016 yılı 01-30 Aralık tarih aralığına ait yaklaşık olarak 1,1 milyon saatlik ölçüm verilerini içermektedir. Veri kümesi 1746 adet meteoroloji istasyonlarının konumunu (enlem ve boylam), rakımını ve saatlik ölçüm olarak da hava sıcaklığı, bağıl nem, hava basınç değerlerini ve ölçüm işleminin hangi zamana ait olduğunu içerir. Bu tez çalışmasında kullanılan meteoroloji ölçüm istasyonlarının konumları Şekil 1.5'de gösterilmiştir.

Büyük veri (big data), geleneksel veri işleme uygulamaları tarafından ele alınamayacak kadar büyük veya karmaşık olan veri kümeleri olarak tanımlanır [6–8]. Büyük veriyi oluşturan 3 bileşen vardır. Bu bileşenler sırasıyla hacim (volume), çeşitlilik (variety) ve hız (velocity) olarak kabul edilir [9, 10] Bir verinin büyük veri olarak kabul edilmesi için veri boyutunun büyük olması (hacim), farklı parametrelerden oluşması (çeşitlilik) ve verinin sürekli güncellenmesi veya yeni verilerin sürekli gelmesi (hız) gerekmektedir [9, 10]. MGM'den alınan meteorolojik veriler büyük hacimlidir, farklı parametrelerden oluşmaktadır ve sık güncellenmektedir ve bu özelliklerinden dolayı MGM'den alınan meteorolojik veriler büyük veri karakteristiği göstermektedir.



Şekil 1.5. MGM saatlik ölçüm istasyonları

1.3.2. Sentetik Veri Kümesi

Çok fazla sayıda ölçüm noktası içeren veri kümesinde GWR algoritmalarının ölçeklenebilirlik performanslarını değerlendirmek için, MGM'den alınan gerçek veriler kullanılarak 1000 referans noktası ve 1000 regresyon noktasından başlayarak her bir zaman aralığında veya güncelleme işleminde 1000 referans noktası ve 1000 regresyon noktasından oluşan toplamda 1.000.000'a kadar ölçüm noktası içeren sentetik veri kümesi oluşturulmuştur. Ölçüm noktalarının konumu ve parametre değerleri rastgele belirlenmiştir.

2. BÖLÜM

RNN-GWR: SIK GÜNCELLENEN VERİLER İÇİN TERS EN YAKIN KOMŞU TABANLI GWR YAKLAŞIMI

2.1. Giriş

Sensörlerin ve IoT (nesnelerin interneti) teknolojilerinin artan kullanımı ile birlikte farklı uygulama alanlarında çok sayıda parametreye sahip sık güncellenen veya gerçek zamanlı mekânsal (veya konuma dayalı) veriler toplanmaktadır [1–5]. Örnek olarak meteorolojik parametreler (hava sıcaklığı, basınç, nem, güneş radyasyonu, yağış) ve çevresel parametreler (hava kalitesi, hava kirliliği, ozon, karbon monoksit, toz konsantrasyonu, gürültü vb.) olarak listelenebilir. Ancak ölçüm değeri elde edilmek istenen her noktaya sensör yerleştirmek çok zor ve maliyetli bir işlemdir. Sensörün yerleştirilmediği bir yerde bir parametre değerine ihtiyaç duyulduğunda, komşu noktalara göre parametre değerinin tahmin edilmesi daha ekonomik ve uygulanabilir bir çözümdür. Belirli noktaların değerleri bilinirse, bilinmeyen noktaların değeri mekânsal olarak değişen ilişkileri modellemek için yaygın olarak kullanılan GWR yöntemi ile bulunabilir [15, 39, 49] fakat referans olarak alınan bilinen noktaların parametre değeri sıklıkla güncellenmektedir.

GWR çalışmalarında sık güncellenen veriler analiz edilirken iki büyük zorluk vardır. Zorluklardan birincisi veri boyutunun zaman geçtikçe sürekli büyümesidir. Zorluklardan ikincisi ise Bölüm 1.2.1’de bahsedildiği üzere mevcut GWR yöntemlerinin sık güncellenen verileri işlemek için yeterli olmamasıdır ve bu nedenle hesaplama açısından verimli algoritmalar geliştirilmelidir.

Tezin bu bölümünde sık güncellenen veriler için hesaplama açısından verimli bir GWR yöntemi önerilmektedir. Önerilen yaklaşım dağıtılmış veya paralel şekilde çalışabilen bilgi işlem kaynaklarına ihtiyaç duymamaktadır ve yaklaşımın doğruluğundan ödün

vermemektedir. Önerilen yöntem ters en yakın komşu (reverse nearest neighbor - RNN) stratejini kullanmaktadır ve bu yüzden RNN-GWR olarak adlandırılmaktadır. RNN yaklaşımı [75] dinamik veriler ve hareketli nesneler [76–80] için sıklıkla kullanılan bir yöntemdir, sınıflandırma ve anamoli tesbitinde de kullanılabilir [54, 81, 82].

Önerilen RNN-GWR yaklaşımının performansı Bölüm 1.3.1 ve Bölüm 1.3.2’de bahsedilen veri kümeleri kullanarak Toy-GWR ve FastGWR [26] yaklaşımları ile karşılaştırılmıştır. Toy-GWR algoritması, sabit konumlardaki sık güncellenen verileri analiz etmek için klasik GWR algoritmasının bir uzantısıdır. Bu algoritma temel olarak veri kümesinde bulunan parametre değerlerinin her güncelleme veya toplu güncelleme işleminde tüm veri kümesi için klasik GWR algoritmasını çalıştırır. Her güncelleme için klasik GWR’nin tekrar tekrar çalışması nedeniyle, bu işlem hesaplama açısından karmaşıktır ve dolayısı ile çok fazla zaman alır. Bu eksikliklerin üstesinden gelmek için, bu çalışma sık güncellenen verilerin işlenmesi için RNN-GWR algoritması önermektedir. Önerilen RNN-GWR algoritması temel olarak referans noktalarındaki güncelleme nedeniyle etkilenen regresyon noktalarını belirler ve sadece etkilenen regresyon noktaları için değerleri tahmin eder. Etkilenen regresyon noktalarını belirlemek için, önerilen RNN-GWR yaklaşımı ters en yakın komşu (reverse nearest neighbor-RNN) sorgusunu kullanmaktadır.

Bu yaklaşımın ana katkıları aşağıdaki gibidir.

- Konumu sabit ve parametre değerleri sık güncellenen verileri analiz etmek için RNN-GWR olarak adlandırılan ters en yakın komşu stratejisine dayanan yeni bir GWR yaklaşımı önerilmiştir.
- Önerilen model Toy-GWR ve FastGWR [26] yaklaşımları ile karşılaştırılmıştır.
- Algoritmaların deneysel değerlendirmeleri Bölüm 1.3.1’de bahsedilen gerçek ölçüm verileri ve Bölüm 1.3.2’de bahsedilen büyük ölçekli sentetik veriler kullanarak yapılmıştır.
- Çalışma sonuçları derlenerek bir dergide yayımlanmıştır [31].

Bu bölümün geri kalan kısmı şu şekilde organize edilmiştir. Bölüm 2.2’de Toy-GWR

algoritmasından bahsedilmektedir. Bölüm 2.3'te önerilen RNN-GWR algoritması hakkında bilgi verilmiştir. Bölüm 2.4'te GWR yaklaşımlarının hesaplama maliyetlerinin analizi verilmiştir. Bölüm 2.5'te deneysel çalışmalar ve Bölüm 2.6'da sonuçlar sunulmuştur.

2.2. Toy-GWR Algoritması

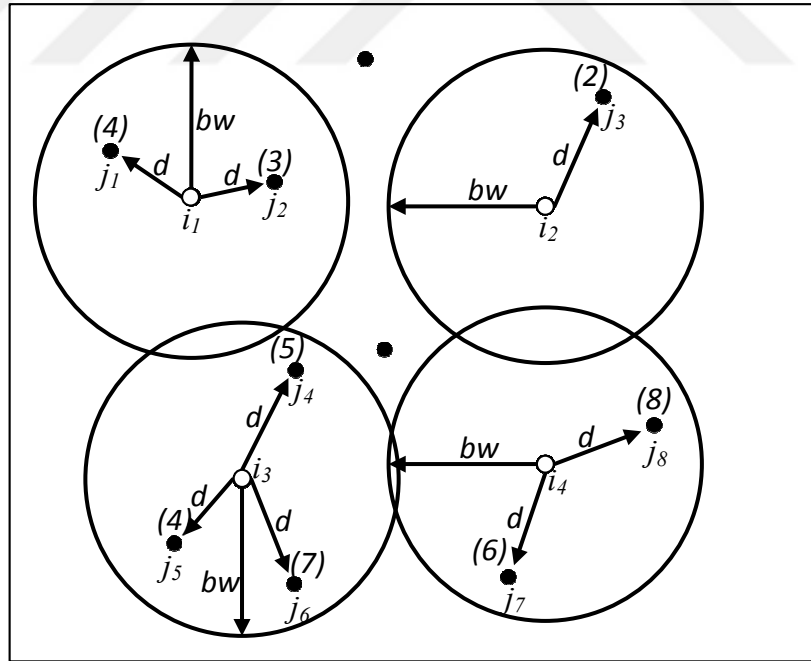
GWR yaklaşımı, değeri bilinen komşu referans noktalarının parametre değerlerini dikkate alarak bir regresyon noktasının parametre değerini tahmin etmek için kullanılmaktadır. Bu yaklaşım kullanıldığında bir referans noktasının parametre değerleri her hangi bir zamanda değiştiğinde etkilenen veya komşu regresyon noktalarının parametre değerleri de değiştirilmeli / güncellenmelidir. Bu durumda en basit yaklaşım (Toy-GWR algoritması) regresyon noktalarının değerlerini doğru bir şekilde tahmin etmek için her bir güncelleme anında bütün veri kümesini dikkate alarak GWR algoritmasını çalıştıracaktır. GWR algoritması sık güncellenen verilere uygulandığında komşu referans noktalarının değerlerinin değişip değişmediğini kontrol etmeden tüm regresyon noktalarının değerlerini yeniden hesaplar. Toy-GWR algoritmasının kaba kodu Algoritma 2.1'de verilmiştir.

Algoritma 2.1: Toy-GWR Algoritması	
Girdi	: D : Veri kümesi
Çıktı	: $\hat{Y}$: Regresyon noktalarının tahmin edilen değeri
Değişkenler:	bw : Bant genişliği, Kn : Çekirdek fonksiyonu, Rn : Regresyon noktaları, Rf : Referans noktaları, uRf : Güncellenen referans nokta kümeleri
1	for each Güncellenen referans nokta kümesi uRf do
2	for $i = 1, i \leq \text{Regresyon nokta sayısı}, i++$ do
3	Bağımsız değişken matrisini (X) ve bağımlı değişken matrisini (Y) oluştur;
4	Regresyon noktası i ile referans noktaları Rf arasındaki mesafeleri, bw ve Kn kullanarak ağırlık matrisini ($W(i)$) oluştur;
5	Eşitlik 1.3 ile i noktasının regresyon katsayılarını ($\beta(i)$) hesapla;
6	i noktası için $\beta(i)$ ve X_i 'yi kullanarak tahmin değerini ($\hat{y}_i$) hesapla;
7	RMSE hesabı için tahmin edilen $\hat{y}_i$ ve ölçüm y_i değerlerini sakla;

Toy-GWR algoritmasında uRf 'de bulunan her bir güncelleme kümesi için Adım 1 ve 7 arasındaki adımlar koşturulur, Rn 'de bulunan her bir regresyon noktası için ise adım

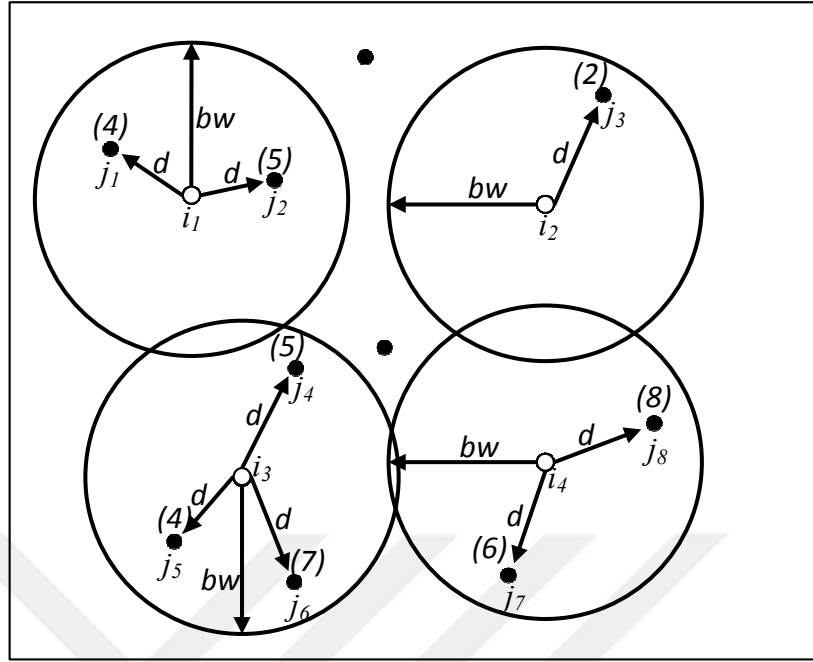
2 ve 7 arasındaki adımlar koşturulur. Adım 3'te işlem yapılan regresyon noktasına özel referans noktalarının bağımsız değişkenlerinin matrisi X ve referans noktalarının bağımlı değişkenlerinin matrisi Y oluşturulur. Adım 4'te regresyon noktası i ve referans noktaları Rf arasındaki mesafeler, bw ve Kn parametre değerleri kullanılarak komşuluk ağırlığı hesaplanır ve ağırlık matrisi $W(i)$ oluşturulur. Adım 5'te i regresyon noktası için Eşitlik 1.3 kullanarak regresyon katsayıları $\beta(i)$ hesaplanır. Adım 6'da i noktası için $\beta(i)$ and X_i değerleri kullanarak regresyon noktasının tahmini değeri $\hat{y}_i$ hesaplanır. Daha sonra Adım 7'de RMSE hesaplamasında kullanmak için tahmin edilen $\hat{y}_i$ değeri ile ölçüm değeri y_i hafızada saklanır.

Toy-GWR algoritmasının nasıl çalıştığı Şekil 2.1 ve Şekil 2.2'de gösterilmiştir. Şekillerde referans noktaları j_s ve regresyon noktaları i_s ile gösterilmiştir. Referans nokta değerleri parantez içerisinde verilmiştir. Regresyon noktaları etrafındaki daireler, regresyon noktalarının komşuluğunu oluşturmaktadır. Komşuluk yarıçapı bant genişliği bw 'dir. Regresyon noktası ile referans noktaları arasındaki komşuluk mesafesi d ile gösterilmiştir.

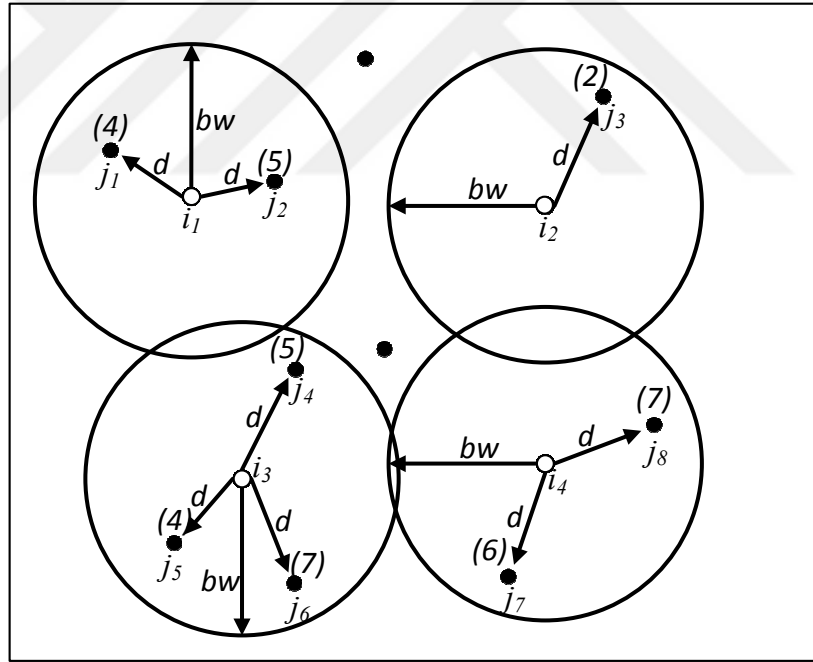


Şekil 2.1. $T1$ zaman diliminde Toy-GWR algoritmasının davranışı

$T1$ zaman diliminde Toy-GWR algoritması bütün i regresyon noktalarının değerlerini tahmin eder. Şekil 2.2a'da gösterildiği gibi $T2$ zaman diliminde j_2 referans noktasının değeri 5 olarak güncellenmiştir ve bu güncelleme işlemi komşu regresyon noktalarının tahmini değerini etkileyecektir. Bu durumda $T2$ zaman diliminde Toy-GWR yaklaşımı



(a) $T2$ zaman diliminde j_2 noktasının değeri güncellenir



(b) $T3$ zaman diliminde j_8 noktasının değeri güncellenir

Şekil 2.2. $T2$ ve $T3$ zaman diliminde Toy-GWR algoritmasının davranışı

regresyon noktalarının değerlerini yeniden tahmin etmek için tüm veri kümesini kullanarak yeniden tahmin işlemi yapar. Şekil 2.2b'de gösterildiği gibi $T3$ zaman diliminde j_8 referans noktasının değeri 7 olarak güncellenmiştir ve bu durumda Toy-GWR algoritması komşuluk ağırlıklarını dikkate alarak bütün regresyon noktaları için yeniden

tahmin işleminde bulunur.

Şekil 2.2a incelendiğinde $T2$ zaman diliminde j_2 referans noktasının değeri güncellendiğinde i_2 , i_3 ve i_4 regresyon noktaları güncelleme işleminden etkilenmez ve benzer şekilde $T3$ zaman diliminde j_8 referans noktasının değeri güncellendiğinde i_1 , i_2 ve i_3 regresyon noktaları güncelleme işleminden etkilenmez. Çünkü bu regresyon noktaları güncellenen referans noktalarının komşu sınırları içerisinde değildirler. Görüldüğü gibi, Toy-GWR yaklaşımı regresyon noktalarının güncellenen referans noktalarından etkilenip etkilenmediğini kontrol etmeksizin bütün veri kümesi için tahmin işleminde bulunmaktadır.

FastGWR algoritması [26] Toy-GWR algoritması ile aynı algoritmik adımları çalıştırır. FastGWR’de GWR yaklaşımının hafıza karmaşıklığını azaltmak için Eşitlik 1.3’de bulunan $\beta(i)$, Eşitlik 2.2’de görüldüğü gibi yeniden düzenlenir ve hesaplamalar eş zamanlı şekilde yapılabilir. Ancak FastGWR yaklaşımı da sık güncellenen bu veri kümesinde verinin güncellenip güncellenmediğini kontrol etmez. Veri kümesinde sık güncellenmenin olduğu durumlarda bu işlemler hesaplama maliyeti açısından çok karmaşıktır ve bu zorlukların üstesinden gelmek için bu tez çalışmasında sık güncellenen verileri etkili bir şekilde analiz etmek için RNN-GWR yaklaşımı önerilmiştir.

2.3. RNN-GWR: Sık Güncellenen Veriler İçin Önerilen Ters En Yakın Komşu Tabanlı GWR Yaklaşımı

Sabit konumlardaki sık güncellenen verileri GWR yaklaşımının doğruluğundan ödün vermeden etkili bir şekilde analiz etmek için RNN-GWR yaklaşımı önerilmiştir. Veri kümesinde bulunan bir referans noktasının parametre değerindeki bir güncelleştirme GWR algoritmasının bant genişliği nedeniyle veri kümesindeki tüm noktaları etkilemez. Bir referans noktasının değerinin güncellenmesi sadece referans noktasının komşuluğunda bulunan regresyon noktalarını etkiler. Bu nedenle önerilen RNN-GWR algoritması güncellenmiş referans noktalarının ters en yakın regresyon komşularını bulur. Bu işlem yapılarak sadece güncellemeden etkilenen regresyon noktaları belirlenir.

GWR modelinde çekirdek fonksiyonu olarak Eşitlik 1.1’de bulunan bi-square fonksiyonu seçildiği zaman sadece bant genişliği içerisinde kalan referans noktaların komşuluk

ağırlığı Eşitlik 1.6'de belirtilen $W(i)$ diyagonal matrisine değer olarak girilmektedir ve kalan diğer referans noktalarının komşuluk ağırlığı 0 olarak girilmektedir. Bu durumda $W(i)$ diyagonal matrisinin 0 değerine sahip satır ve sütunlarının çarpımının sonucu sıfır olmaktadır ve bu işlem gereksiz yere zaman almaktadır. Eşitlik 1.3'ün hesaplama sürecini hızlandırmak için $W(i)$ diyagonal matrisinin 0 değerine sahip satır ve sütunları ve bunlara karşılık gelen X matrisinin satırları (ve bu sayede X^T matrisinin de sütunları) ve Y matrisinin satırları hesaplama işlemlerinden çıkarılabilir. Sıfır değerine sahip satır ve sütunların çarpım işleminden çıkarmanın maliyeti $O(n)$ olmaktadır.

Ayrıca, Eşitlik 2.1'de bulunan $X^T W(i) X$ matris çarpım sonucu simetriktir. Çünkü herhangi bir matris (X) ile o matrisinin transpozunun (X^T) çarpımının sonucu simetrik matristir [83] ve $W(i)$ de kare diyagonal matris olduğu için hepsinin çarpımının sonucu da simetrik matristir. Bu nedenle $X^T W(i) X$ simetrik matrisinin bütün değerlerinin hesaplanmasına gerek yoktur. $X^T W(i) X$ simetrik matrisinde diyagonalın üstündeki değerler hesaplandığında diyagonalın altındaki değerler aynı olacağından dolayı hesaplanmalarına gerek yoktur. $X^T W(i) X$ matrisinin bütün değerleri hesaplanmadığında hesaplama maliyeti de azalmaktadır.

$$X^T W(i) X = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \dots & \dots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & a_{mm} \end{bmatrix} \quad (2.1)$$

GWR algoritmasında Eşitlik 1.3'ün en çok zaman alan kısmı $X^T W(i) X$ matrislerinin çarpımıdır ve hesaplama maliyeti $O(m^2 n)$ 'dir ve bu matrisin tersini $((X^T W(i) X)^{-1})$ almanın maliyeti $O(m^3)$ olarak hesaplanır [26]. n ölçüm sayısı ve m parametre sayısı olmak üzere X , $W(i)$ ve Y matrislerinin boyutları, sırasıyla, $n \times m$, $n \times n$ ve $n \times 1$ 'dir. Yukarıda bahsedilen bulgular sayesinde k değişkeni regresyon noktası için bant genişliği içerisinde kalan komşu referans nokta sayısını göstermek üzere X , $W(i)$ ve Y matrislerinin boyutları, sırasıyla, $k \times m$, $k \times k$ ve $k \times 1$ olmaktadır ve $X^T W(i) X$ matris çarpım hesaplama maliyeti de simetrik matris olduğundan düşecektir. Bu durumda $X^T W(i) X$ matris çarpımının hesaplama maliyeti $O\left(\left(\frac{m^2+m}{2}\right)k + n\right)$ olacaktır. k 'nın değeri n 'nin değerinden daha küçük olacağı ($k \ll n$) için Eşitlik 1.3'ün

hesaplama maliyeti klasik GWR yaklaşımının hesaplama maliyetinden daha düşük olacaktır. Algoritmaların hesaplama karmaşıklığının detayları Bölüm 2.4'te tartışılmıştır.

Önerilen RNN-GWR algoritmasının kaba kodu Algoritma 2.2'de verilmiştir.

Algoritma 2.2: RNN-GWR Algoritması

Girdi : D : Veri kümesi
Çıktı : $\hat{Y}$: Regresyon noktalarının tahmin edilen değeri
Değişkenler: bw : Bant genişliği, Kn : Çekirdek fonksiyonu, Rn : Regresyon noktaları, Rf : Referans noktaları, uRf : Güncellenen referans nokta kümeleri, $Etkilenen_Noktalar$: Etkilenen noktaların listesi

```

1 for each Güncellenen_Referans_Nokta_Kümesi  $uRf$  do
2    $Etkilenen\_Noktalar = \text{RNN\_Etkilenen\_Noktalar\_Getir}(Rn, uRf, bw)$ ;
3   for  $i = 1, i \leq Etkilenen\_Noktalar\_Sayısı, i++$  do
4     Bağımsız değişken matrisini ( $X$ ) ve bağımlı değişken matrisini ( $Y$ )
       oluştur;
5     Regresyon noktası  $Etkilenen\_Noktalar(i)$  ile referans noktaları  $Rf$ 
       arasındaki mesafeleri,  $bw$  ve  $Kn$  değerlerini kullanarak ağırlık matrisini
       ( $W(i)$ ) oluştur;
6     Denklem 1.3'yi kullanarak  $Etkilenen\_Noktalar(i)$  noktasının
       regersyon katsayılarını ( $\beta(i)$ ) hesapla;
7      $Etkilenen\_Noktalar(i)$  noktası için  $\beta(i)$  and  $X_i$ 'yi kullanarak tahmini
       değeri ( $\hat{y}_i$ ) hesapla;
8     RMSE hesabı için tahmin edilen  $\hat{y}_i$  ve ölçüm  $y_i$  değerlerini sakla;
```

Algoritma 2.3: RNN_Etkilenen_Noktalar_Getir Algoritması

Girdi: Rn : Regresyon noktaları, uRf : Güncellenen referans nokta kümeleri, bw : Bant genişliği
Çıktı : $Etkilenen_Noktalar$: Etkilenen noktaların listesi

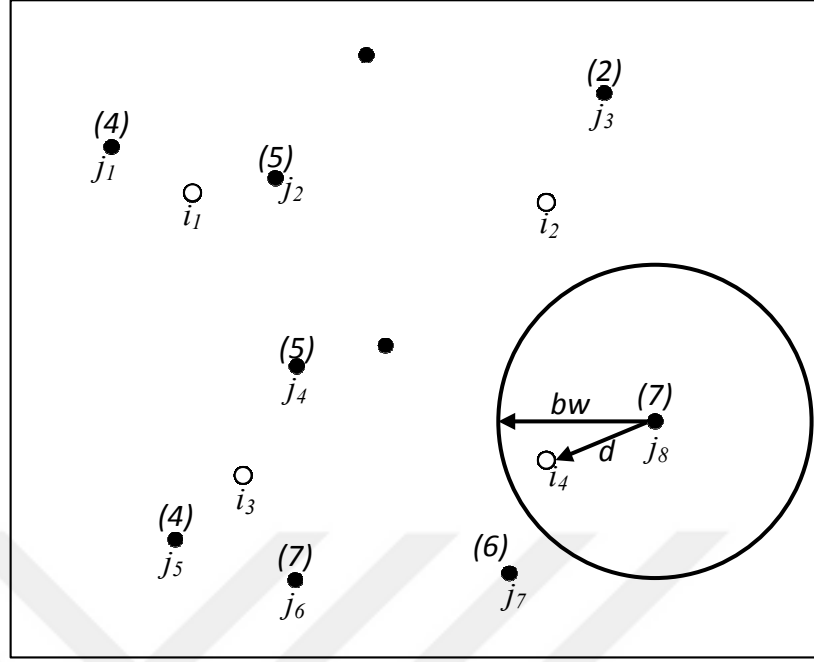
```

1  $Etkilenen\_Noktalar = []$ ;
2 for  $k = 1, k \leq Güncellenen\_Referans\_Nokta\_Sayısı, k++$  do
3   Regresyon noktaları  $Rn$  ve güncellenen referans noktası  $uRf(k)$  arasındaki
       mesafeleri hesapla ( $Mesafeler$ ) ;
4   Bant genişlik mesafesi içerisinde olan noktaları bul
       ( $Eklenecek\_Noktalar = Mesafeler < bw$ );
5   Bulunan noktaları etkilenen noktalar listesine ekle
       ( $Etkilenen\_Noktalar = [Etkilenen\_Noktalar \ Eklenecek\_Noktalar]$ );
6 Etkilenen noktaları tekilleştir
       ( $Etkilenen\_Noktalar = \text{unique}(Etkilenen\_Noktalar)$ );
7 return  $Etkilenen\_Noktalar$ ;
```

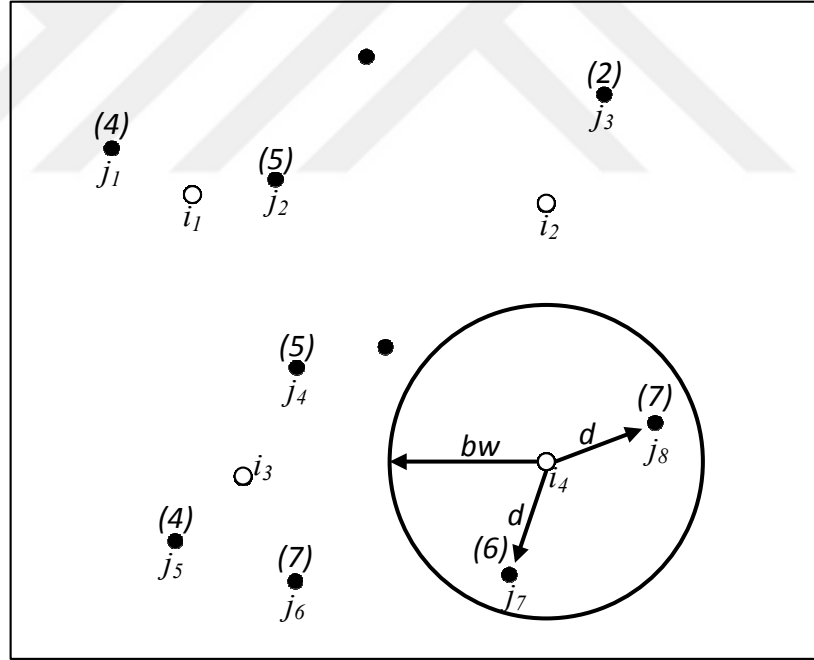
RNN-GWR algoritmasında uRf 'de bulunan her bir güncelleme kümesi için Adım 1 ve 8 arasındaki adımlar koşutur. Adım 2'de güncelleme işleminden etkilenen

regresyon noktaları (*Etkilenen_Noktalar*) Algoritma 2.3 kullanarak bulunur. Algoritma 2.3'te ilk olarak *Etkilenen_Noktalar* değişkeni boş bir liste olarak ayarlanır. Adım 2 ve Adım 5 arasında bulunan satırlar uRf 'de bulunan her bir referans noktası için çalıştırılır. Adım 3'te güncellenen her bir referans nokta $uRf(k)$ ile bütün regresyon noktaları arasındaki mesafe hesaplanır. Adım 4'te bant genişliği değerinden daha küçük mesafeye sahip olan regresyon noktaları tespit edilir ve Adım 5'te bu regresyon noktaları *Etkilenen_Noktalar* değişkenine eklenir. *Etkilenen_Noktalar* değişkeninde aynı regresyon noktası birden fazla eklenebileceği için Adım 6'da tekilleştirme işlemi yapılır ve Adım 7'de bu liste algoritmanın ürettiği sonuç değer olarak döner. RNN-GWR algoritmasında (Algoritma 2.2) değişimden etkilenen regresyon noktalarında (*Etkilenen_Noktalar*) bulunan her bir regresyon noktası için ise Adım 3 ve 8 arasındaki adımlar koşturulur. Adım 4'te işlem yapılan regresyon noktasına (*Etkilenen_Noktalar(i)*) özel referans noktalarının bağımsız değişkenlerinin matrisi X ve referans noktalarının bağımlı değişkenlerinin matrisi Y oluşturulur. Adım 5'te regresyon noktası *Etkilenen_Noktalar(i)* ve referans noktaları Rf arasındaki mesafeler, bw ve Kn parametre değerleri kullanılarak ağırlık matrisi $W(i)$ oluşturulur. Adım 6'da *Etkilenen_Noktalar(i)* regresyon noktası için Eşitlik 1.3 kullanarak regresyon katsayıları $\beta(i)$ hesaplanır. Adım 7'de *Etkilenen_Noktalar(i)* noktası için $\beta(i)$ and X_i formlarını kullanılarak regresyon noktasının tahmini değeri $\hat{y}_i$ hesaplanır. Daha sonra Adım 8'de RMSE hesaplamasında kullanmak için tahmin edilen $\hat{y}_i$ değeri ve ölçüm değeri y_i hafızada saklanır.

Şekil 2.3 ve Şekil 2.4'te RNN-GWR algoritmasının nasıl çalıştığı gösterilmektedir. $T1$ zaman diliminde (Şekil 2.1) RNN-GWR algoritması Toy-GWR algoritmasında olduğu gibi bütün veri kümesi için tahmin değerleri üretecektir. $T2$ zaman diliminde sadece j_2 referans noktası güncellenir ve RNN-GWR algoritması Şekil 2.3a'da gösterildiği gibi ters en yakın komşu sorgusunu kullanarak değişimden etkilenen regresyon noktalarını belirler. Şekil 2.3a'da görüldüğü gibi değişimden sadece i_1 regresyon noktası etkilenmektedir. RNN-GWR algoritması Şekil 2.3b'de gösterildiği gibi sadece değişimden etkilenen i_1 regresyon noktası için yeniden değer üretir. Benzer şekilde $T3$ zaman diliminde sadece j_8 referans noktası güncellenir ve RNN-GWR algoritması Şekil 2.4a'da gösterildiği üzere ters en yakın komşu stratejisini kullanarak değişimden etkilenen regresyon noktalarını



(a) Değeri güncellenen j_8 referans noktasının ters en yakın komşuları



(b) Değişimden etkilenen i_4 regresyon noktasının değerinin hesaplanması

Şekil 2.4. T_3 zaman diliminde RNN-GWR algoritmasının davranışı

2.4. GWR Yaklaşımlarının Hesaplama Maliyetlerinin Analizi

GWR yaklaşımları yüksek bilgi işlem kaynakları ve geniş bellek alanları gerektirir. Veri boyutu büyüdükçe ve/veya veri güncellemeleri sıklaştıkça bu sorunlar kritik hale gelir.

GWR yaklaşımlarının bellek karmaşıklığını azaltmak için farklı matris depolama ve hesaplama stratejileri kullanılabilir. Li vd. [26] Eşitlik 1.3'te bulunan $\beta(i)$ değerlerinin hesaplarken hesaplama maliyetini azaltmak ve daha etkili hafıza yönetmek için Eşitlik 2.2'yi kullanmışlardır. Eşitlik 2.2 hesaplamalarında $n \times n$ boyutta $W(i)$ komşuluk matrisi kullanmak yerine $1 \times n$ boyutta olan $w(i) = (w_{i1}, \dots, w_{in})$ vektör formatında komşuluk matrisi kullanılmıştır. Buna ilaveten $w(i)$ vektör matrisinden yararlanmak için $R(i)$ noktasal (.) çarpımı kullanılmıştır.

$$\begin{aligned}
 \beta(i) &= (X^T W(i) X)^{-1} X^T W(i) Y \\
 R(i) &= X^T W(i) \\
 \beta(i) &= (R(i) X)^{-1} R(i) Y \\
 R(i) &= X^T W(i) = (X \cdot w(i)^T)^T
 \end{aligned} \tag{2.2}$$

Eşitlik 2.2'de kullanılan $R(i)$ matrisi Eşitlik 2.3 ile ifade edilir.

$$R(i) = \begin{bmatrix} x_{11}w_{i1} & x_{21}w_{i2} & \dots & x_{n1}w_{in} \\ x_{12}w_{i1} & x_{22}w_{i2} & \dots & \vdots \\ \vdots & \dots & \dots & \vdots \\ x_{1m}w_{i1} & \dots & \dots & x_{nm}w_{in} \end{bmatrix} \tag{2.3}$$

$X^T W(i)$ matris formunun $(X \cdot w(i)^T)^T$ olarak yeniden düzenlenmesi $n \times n$ boyutta olan diyagonal $W(i)$ matrisini $1 \times n$ boyutta $w(i)$ vektör satırına dönüştürür ve hafıza karmaşıklığını $O(n^2)$ boyuttan $O(mn)$ boyuta düşürür [26]. $W(i)$ matrisinin yeni hali sadece hafıza karmaşıklığını azaltmaz aynı zamanda GWR algoritmalarının hesaplama maliyetini de düşürür.

GWR algoritmalarının zaman karmaşıklığının büyük kısmı β_i hesaplamalarından kaynaklanmaktadır. Eşitlik 1.3'te bulunan β_i her bir regresyon işleminde tekrar hesaplanacaktır yani n ölçüm sayısı olmak üzere bu işlem n defa tekrar edilecektir. GWR algoritmasında Eşitlik 1.3'ün en çok zaman alan kısmı $X^T W(i) X$ matrislerinin çarpımıdır ve hesaplama maliyeti $O(m^2 n)$ 'dir ve bu matrisin tersini $((X^T W(i) X)^{-1})$ almanın maliyeti $O(m^3)$ 'dür [26]. n ölçüm sayısı ve m parametre sayısı olmak üzere X , $W(i)$ ve Y matrislerinin boyutları, sırasıyla, $n \times m$, $n \times n$ ve $n \times 1$ 'dir. X ,

$W(i)$ ve Y matrislerinin boyutu n ölçüm sayısı ve m tahmin edilecek parametre sayısı olmak üzere, sırasıyla, $n \times m$, $n \times n$ ve $n \times 1$ formatındadır. m genellikle n 'den daha küçük olacağından ve bu hesaplama işlemleri de n defa tekrar edeceğinden dolayı klasik GWR algoritmalarının hesaplama maliyeti $O(m^2n^2)$ olarak ifade ededilir. Yani en kötü durumda Toy-GWR ve FastGWR yaklaşımlarının hesaplama maliyeti $O(m^2n^2)$ olmaktadır.

Önerilen RNN-GWR algoritmasının hesaplama maliyeti RNN sorgusu ve Eşitlik 1.3'te bulunan β_i hesaplamasına bağlıdır. RNN sorgusunun hesaplama maliyeti $O(n^2)$ olarak ifade edilir [84]. Bölüm 2.3'te bahsedildiği üzere X , $W(i)$ ve Y matrislerinin boyutu k değeri regresyon noktası i için bant genişliği içerisinde kalan komşu sayısı olmak üzere sırasıyla $k \times m$, $k \times k$ ve $k \times 1$ haline gelmektedir ve $X^T W(i) X$ matrisi simetrik matris olduğundan dolayı hesaplama maliyeti düşmektedir. k değişkenin değeri n değişkeninin değerinden daha küçük olacağı ($k \ll n$) için Eşitlik 1.3'in hesaplama maliyeti $O\left(\left(\frac{m^2+m}{2}\right)k + n\right)$ olarak ifade edilir.

a parametresi değişimden etkilenen nokta sayısı *Etkilenen_Noktalar*'ı göstermek üzere n değerinden daha küçük olur ise RNN-GWR algoritmasında β_i denkleminin toplam hesaplama maliyeti $O\left(\left(\left(\frac{m^2+m}{2}\right)k + n\right)a\right)$ olacaktır. RNN sorgusu da dahil edilince RNN-GWR algoritmasının hesaplama maliyeti $O(n^2) + O\left(\left(\left(\frac{m^2+m}{2}\right)k + n\right)a\right)$ olarak ifade edilir. En kötü durumda, değişimden bütün regresyon noktaları etkilenebilir (yani $a = n$ olabilir), k parametresi ölçüm sayısına eşit olabilir (yani $k = n$ olabilir) ve bu durumda RNN-GWR algoritmasının hesaplama maliyeti $O\left(\left(\frac{m^2+m}{2}\right)n^2 + 2n^2\right)$ olacaktır. Fakat bu hesaplama maliyeti, değişim bütün regresyon noktalarını etkilerse ve bütün referans noktaları herhangi bir regresyon noktasının komşu olursa gerçekleşecektir. Çoğu veri kümesi kısmen güncellendiği için ve büyük veri kümelerinde komşu referans nokta sayısı toplam referans nokta sayısından az olacağı için RNN-GWR algoritmasının hesaplama maliyeti $O\left(\left(\frac{m^2+m}{2}\right)n^2 + 2n^2\right)$ değerinden daha küçük olacaktır.

Tüm algoritmaların maliyetleri optimum bant genişliği bulma maliyetini de içerebilir ve bu maliyet $O(\log(n))$ olarak ifade edilir [26, 85]. Bu durumda bant genişliği bir defa bulunur ve algoritmalara giriş parametresi olarak verilir. Algoritmaların içerisinde her bir iterasyonda bant genişliği bulma prosedürü yoktur. Algoritmalara en uygun bant

genişliği değerini bulma işleminin hesaplama maliyeti eklenirse Toy-GWR, FastGWR ve RNN-GWR yaklaşımlarının hesaplama maliyeti , sırasıyla, $O(m^2n^2\log n)$, $O(m^2n^2\log n)$ ve $O\left(\left(\frac{m^2+m}{2}\right)n^2 + 2n^2\right)\log n$ olacaktır [26].

2.5. Deneysel Değerlendirmeler

Bu bölümde Toy-GWR, FastGWR [26] ve önerilen RNN-GWR algoritmalarının başarısını göstermek için yapılan deneysel değerlendirmelerden bahsedilmektedir. Bölüm 1.3.1’de bahsedilen MGM’den alınan gerçek ölçüm verileri ve Bölüm 1.3.2’de bahsedilen sentetik veri kümesi kullanılmıştır. Veri kümesi ve model ayarları ve deneysel değerlendirme ile ilgili detaylı bilgiler sırasıyla Bölüm 2.5.1 ve 2.5.2’de verilmiştir.

Deneysel değerlendirmeler aşağıdaki sorulara cevap vermeye odaklanmıştır.

- Her bir iterasyon (güncelleme) için algoritmaların çalışma zamanı nedir?
- Veri boyutunun etkisi nedir?
- Güncellenme oranının etkisi nedir?
- Tahmin edilen değer ile gerçek değer arasındaki fark nedir?
- Çok çekirdekli ortamda algoritmaların performansları nedir?

İlk dört soruya cevap vermek için algoritmalar Intel Core i5-3210M 2.5 GHz 8GB RAM özelliklerine sahip bir bilgisayar üzerinde çalıştırılmıştır. Son soruya cevap vermek için algoritmalar, çok çekirdekli Intel Xenon E5-2660 V3 2.60 GHz (16 çekirdek) 16GB RAM ve SSD disk özelliklerine sahip olan bir bilgisayar üzerinde koşturulmuştur. Yaklaşımlar Matlab’da koşturulmuştur.

2.5.1. Veri Kümesi ve Model Ayarları

Bu bölümde, ilk önce deneysel değerlendirmelerde kullanılan veri kümesi hakkında bilgi verilmiştir. Daha sonra yaklaşımlar için kullanılan model ayarları sunulmuştur.

2.5.1.1. Veri Kümesi

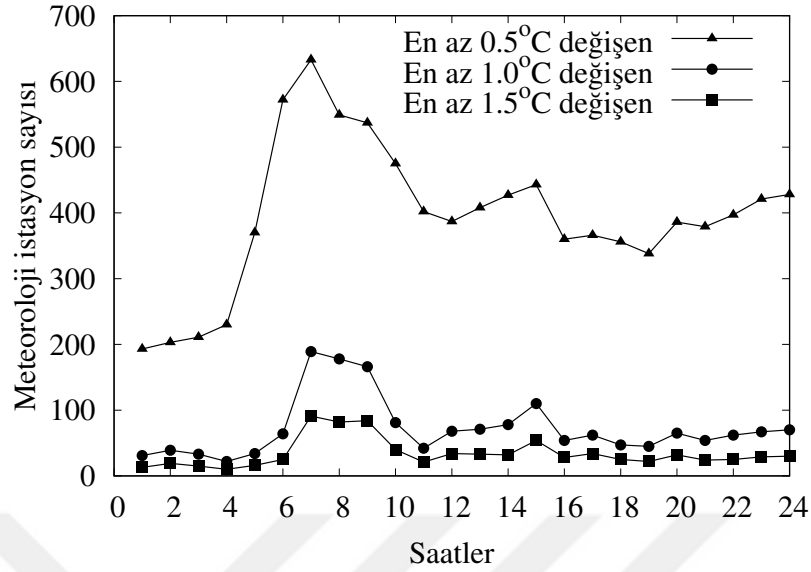
Bu tez kapsamında Bölüm 1.3.1’de bahsedilen MGM’den alınan 1746 adet saatlik ölçüm yapan istasyon bilgilerini ve gerçek ölçüm verilerini içeren meteoroloji veri kümesi kullanılmıştır. Bu çalışmanın amacı sık güncellenen veri kümeleri için etkili GWR yaklaşımları geliştirmektedir. Bu nedenle meteorolojik veri kümesi dışındaki sık güncellenen veri kümeleri önerilen RNN-GWR algoritması tarafından kullanılabilir.

Büyük veri kümesindeki algoritmaların ölçeklenebilirlik performanslarını değerlendirmek için Bölüm 1.3.2’de bahsedilen 1.000.000 ölçüm noktası içerecek şekilde sentetik veri kümesi oluşturulmuştur. Ölçüm noktalarının yeri ve parametre değerleri rastgele oluşturulmuştur.

Bu tez çalışmasında kullanılan meteoroloji ölçüm istasyonlarının konumları Şekil 1.5’de gösterilmiştir. Bu çalışmada 1600 adet ölçüm istasyonu deneysel değerlendirmelerde kullanılmak üzere rasgele seçilmiştir. Bu istasyonlarından 1200 referans noktası 400 tanesi de regresyon noktası olarak kullanılmıştır. Bu çalışmada, ardışık saatlerdeki minimum sıcaklık değişimine dayanarak üç farklı veri kümesi hazırlanmıştır. Deneysel çalışmalarda amaç regresyon noktasının hava sıcaklık değerini tahmin etmektir. Bir referans noktasının hava sıcaklığı değeri saat bazlı değişirse regresyon noktalarının değerleri önerilen algoritma kullanılarak güncellenir. İlk veri kümesi sıcaklık değerleri ardışık saatlerde en az 0.5°C değişirse güncellenen noktaları içermektedir, ikinci veri kümesi sıcaklık değerleri ardışık saatlerde en az 1°C değişirse güncellenen noktaları içermektedir ve üçüncü veri kümesi ise sıcaklık değerleri ardışık saatlerde en az 1.5°C değişirse güncellenen noktaları içermektedir.

Şekil 2.5 ardışık saatlerde sıcaklık değeri en az 0.5°C , 1.0°C ve 1.5°C değişen meteoroloji istasyon sayısını göstermektedir. Şekil 2.5’te görülebileceği gibi, hava sıcaklık değişim değeri azaldıkça değişen meteoroloji istasyonlarının sayısı artmaktadır.

Buna ek olarak sıcaklık değerleri sabah saat 06:00-10:00 arasında daha çok değişmektedir. Şekil 2.6’da üç farklı zaman dilimi için ardışık saatlerde sıcaklık değeri en az 1.0°C değişen istasyonların konumlarını göstermektedir.



Şekil 2.5. Hava sıcaklık değişim değerine göre meteoroloji istasyonları sayısı

2.5.1.2. Model Ayarları

Deneylerde amaç, bir noktanın bağıl nem, hava basıncı ve rakım değerini kullanarak o noktanın sıcaklık değerini tahmin etmektir. GWR yaklaşımında kullanılan sıcaklık tahmin etme modeli Eşitlik 2.4 ile ifade edilmiştir.

$$Sicaklik_i = \beta_{0i} + \beta_{1i}Rakim_i + \beta_{2i}Nem_i + \beta_{3i}Basinc_i \quad (2.4)$$

Eşitlik 2.4'te i regresyon noktasını ve β_1, β_2 ve β_3 regresyon katsayılarını göstermektedir.

GWR modellerinde öncelikle çekirdek fonksiyonu ve bant genişliği değeri belirlenmelidir. Bu çalışma kapsamında yaygın olarak kullanılan çekirdek fonksiyonlarından olan bi-square çekirdek fonksiyonu kullanılmıştır [34, 35]. Bant genişliği değeri sabit bir değer olabilir veya veri kümesi için en uygun bant genişliği değeri kalibrasyon yöntemleri ile bulunabilir [22, 23]. Bant genişliği değeri en uygun regresyon modeli için kullanılan düzeltilmiş Akaike bilgi kriteri, AIC_c , minimize edilerek bulunmuştur [15, 39, 47].

Bu çalışmada Toy-GWR, FastGWR ve önerilen RNN-GWR için en uygun bant genişliği değeri olarak AIC_c ile bulunan değer kullanılmıştır. Deneysel çalışmalarda ölçüm noktalarının konumu sabit olduğundan dolayı bant genişliği değeri algoritmanın başında



(a) 07:00 ve 08:00 arasında



(b) 15:00 ve 16:00 arasında



(c) 23:00 ve 24:00 arasında

Şekil 2.6. 01 Aralık 2016'da üç zaman dilimi için en az 1°C sıcaklık değişimine sahip meteoroloji istasyonlarının konumları

ayarlanmış ve algoritma çalışma süresi boyunca sabit kalmıştır. Yani herhangi bir iterasyonda bant genişliği değeri değişmemiştir. GWR algoritmaları için AIC_c yöntemi ile, Öklid mesafesi kullanılarak en uygun bant genişliği değeri 89 km bulunmuştur.

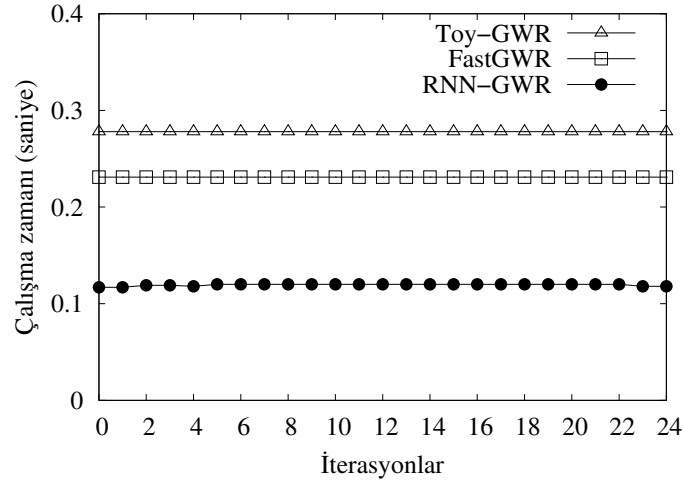
2.5.2. Deneyler

2.5.2.1. Algoritmaların Her Bir İterasyon için Çalışma Zamanları

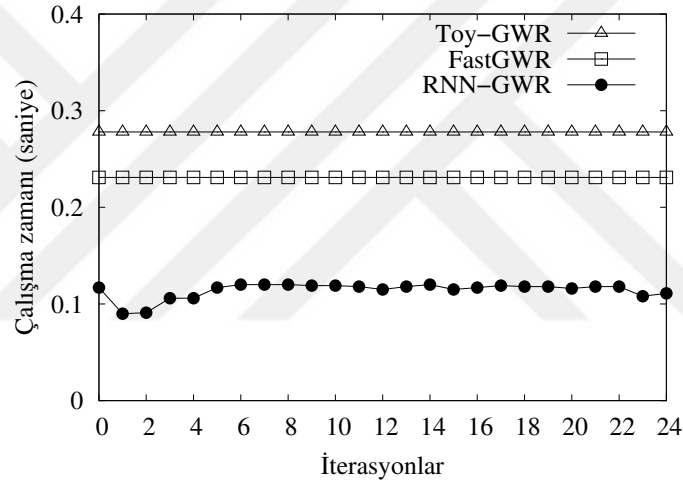
Bu deneyde Toy-GWR, FastGWR ve RNN-GWR algoritmalarının çalışma zamanı performansları her bir gerçek veri kümesi için değerlendirilmiştir. Referans nokta sayısı 1200 ve regresyon nokta sayısı 400 olarak ayarlanmıştır. Bu deneyde algoritmaların çalışma zamanları algoritmaların her bir iterasyonu (yani veri kümesinde saat bazlı) için incelenmiştir. Şekil 2.7, hazırlanan üç veri kümesi için tüm algoritmaların çalışma zamanlarını göstermektedir. Şekilde görülebileceği gibi önerilen RNN-GWR algoritması hem Toy-GWR hem de FastGWR yaklaşımlarından daha iyi performans göstermiştir. Toy-GWR ve FastGWR algoritmaları her bir iterasyonda (saat bazında) tüm veri kümesi için çalışmaktadır. Bu yüzden çalışma süreleri şekillerde görüldüğü gibi her iterasyonda sabittir. Çünkü her güncelleştirme işleminde tüm regresyon noktalarının değerlerini hesaplamak için bütün veri kümesini kullanırlar. Bunun aksine, önerilen RNN-GWR yaklaşımı her bir güncelleme işleminde tüm veri kümesini kullanmaz, yalnızca güncellenmiş referans noktaları ve RNN sorgusu kullanılarak belirlediği güncellemeden etkilenen regresyon noktaları için çalışır. RNN-GWR yaklaşımının çalışma zamanı güncellenen referans noktalarının sayısına ve etkilenen regresyon noktalarına bağlı olarak değişir. Şekil 2.7a'da görüleceği üzere veri kümesindeki dikkate alınan sıcaklık değişikliği küçükse, güncellenen referans noktalarının sayısı artar ve bu nedenle RNN-GWR algoritmasının çalışma zamanı artar. Şekil 2.7c'de gösterildiği gibi veri kümesindeki sıcaklık değişikliği büyükse, güncellenen referans noktalarının sayısı azalacak ve bu nedenle RNN-GWR algoritmasının çalışma zamanı azalacaktır.

Bu deneyde, kullanılan gerçek meteorolojik veri kümesi nedeniyle sıcaklık değerlerinin güncelleme aralığı bir saattir. Ancak, güncelleme aralığı küçükse ve / veya veri kümesi boyutu artarsa, algoritmaların zaman karmaşıklığı artar. Bu durumda, RNN-GWR algoritması diğer algoritmalarından daha iyi performans gösterir.

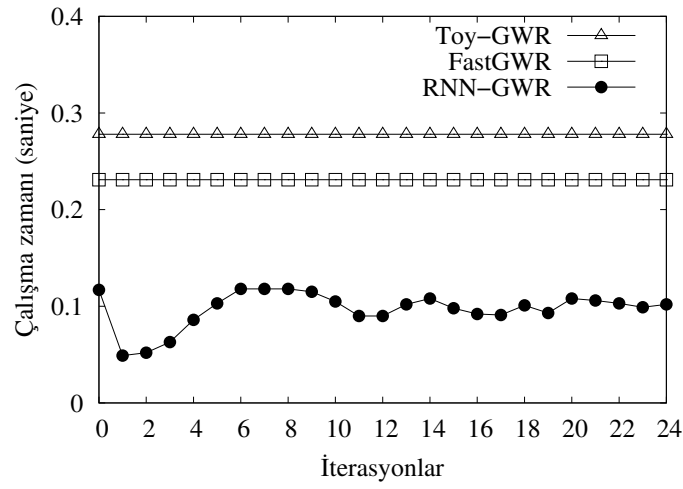
Algoritmaların veri boyutuna göre ölçeklenebilirliği gerçek meteoroloji veri kümesinden sentetik olarak üretilen büyük veri kümesi kullanarak bir sonraki deneyde incelenmiştir.



(a) En az 0.5°C sıcaklık değişimi



(b) En az 1°C sıcaklık değişimi



(c) En az 1.5°C sıcaklık değişimi

Şekil 2.7. Farklı sıcaklık değişimleri için algoritmaların çalışma zamanları

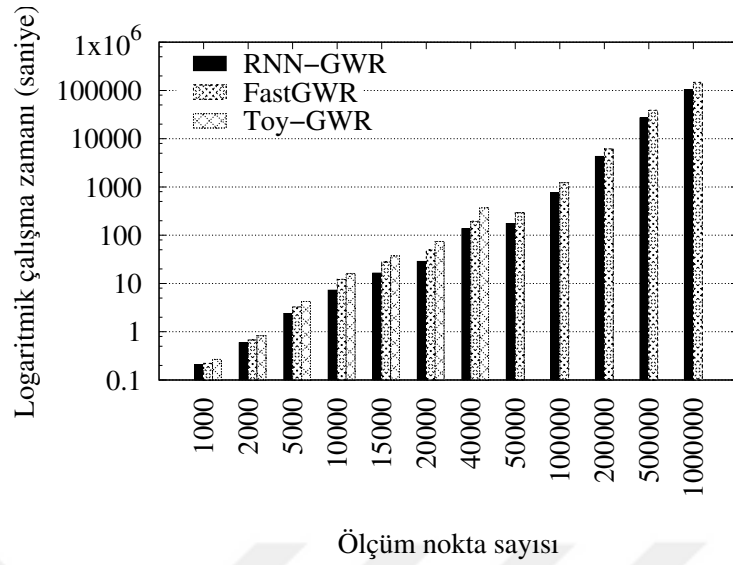
2.5.2.2. Veri Boyutunun Etkisi

Bu deneyde, veri kümesi boyutunun Toy-GWR, FastGWR ve RNN-GWR algoritmalarının performansları üzerindeki etkisi değerlendirilmiştir. Bölüm 1.3.2’de bahsedilen gerçek ölçüm verilerinden elde edilen 1.000 ve 1.000.000’a kadar ölçüm noktasına sahip sentetik veri kümesi kullanılmıştır. Deneylerde tüm algoritmalar önceden tanımlanmış bant genişliği değerlerini kullanmıştır. Deneysel sonuçlar Tablo 2.1 ve Şekil 2.8’de verilmiştir. Veri kümesi boyutları arttıkça tüm algoritmaların maliyeti de artmaktadır. Tablo 2.1 ve Şekil 2.8’de gösterildiği üzere Toy-GWR algoritması en fazla 40.000 ölçüm noktasına sahip veri kümesini işleyebilmektedir ve 40.000 ölçüm noktasından daha fazla ölçüm noktasına sahip veri kümelerini işleyememektedir. Bu nedenle 50.000 ölçüm noktası ve fazlası için Toy-GWR algoritmasının sonuçları Tablo 2.1 ve Şekil 2.8’de verilememiştir. FastGWR ve önerilen RNN-GWR algoritması 1.000.000 ölçüm noktasına sahip veri kümelerini işleyebilmektedir ve bu algoritmaların bütün sonuçları Tablo 2.1 ve Şekil 2.8’de gösterilmiştir.

Tablo 2.1. 1.000 ile 1.000.0000 arasında ölçüm noktası kullanan algoritmaların çalışma zamanları (saniye - sn)

Ölçüm nokta sayısı	RNN-GWR (sn)	FastGWR (sn)	Toy-GWR (sn)
1.000	0,11	0,22	0,26
2.000	0,31	0,67	0,82
5.000	1,54	3,25	4,23
10.000	5,72	12,19	16,09
15.000	12,83	27,78	37,21
20.000	22,29	49,37	74,04
40.000	87,17	195,29	368,85
50.000	129,83	292,62	-
100.000	524,47	1224,89	-
200.000	2541,60	6136,41	-
500.000	14794,93	38883,18	-
1.000.000	47819,89	146920,33	-

Önerilen RNN-GWR algoritması aşağıdaki nedenlerden dolayı Toy-GWR ve FastGWR algoritmalarından daha iyi performans göstermektedir. FastGWR algoritması ise Toy-GWR algoritmasından daha iyi performans göstermektedir. İlk sebep, RNN-GWR ve FastGWR algoritmalarının Eşitlik 2.2’de bulunan $X^T W(i)$ matrisinin yeni formasyonu



Şekil 2.8. Farklı veri boyutları için algoritmaların çalışma sürelerinin karşılaştırılması

olan $(X.w(i)^T)^T$ ifadesini kullanmasıdır. İkinci sebep, RNN-GWR algoritmasında komşuluk matrisi $W(i)$ 'nin sıfır değerine sahip satır ve sütunlar ve bunlara karşılık gelen X ve Y matrisinin ilgili satır ve/veya sütunları hesaplama işleminden çıkarılmasıdır. Üçüncü sebep, $X^T W(i) X$ matrisi simetrik bir matrisdir ve böylece sadece diyagonalın üstündeki değerlerin hesaplanması yeterli olmaktadır ve diyagonalın altındaki değerlerin hesaplanmasına ihtiyaç yoktur. Bu nedenle regresyon katsayısı $\beta(i)$ değerlerinin hesaplanması zaman karmaşıklığı açısından daha uygun hale gelmektedir. Ayrıca bu deneyde RNN-GWR algoritması Toy-GWR ve FastGWR algoritmalarında olduğu gibi bütün regresyon noktaları için değer üretmekdir.

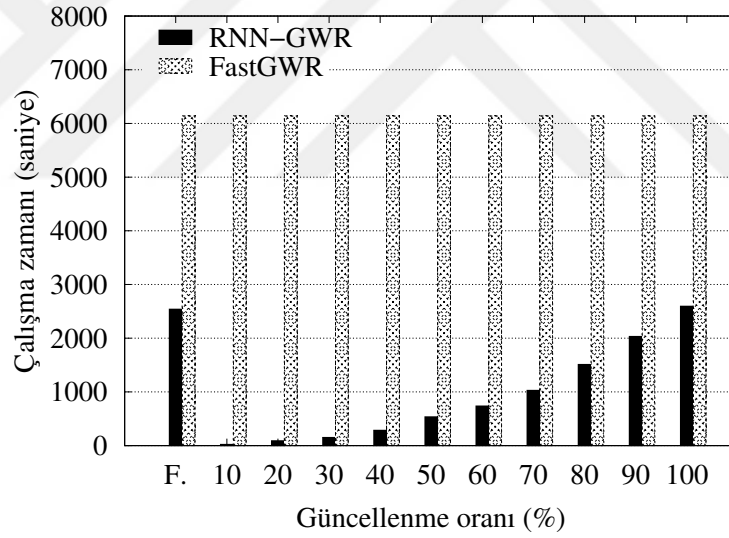
2.5.2.3. Güncellenme Oranının Etkisi

Bu deneyde veri kümesindeki güncellenme oranının FastGWR ve RNN-GWR algoritmalarının performansları üzerindeki etkisi değerlendirilmiştir. Ölçüm nokta sayısı 200.000 olarak ayarlanmıştır ve güncellenme oranı %10'dan %100'e kadar değiştirilmiştir. Deneysel çalışmalarda Toy-GWR algoritması en fazla 40.000 ölçüm noktasına kadar çalıştığı ve ölçüm nokta sayısı 200.000 olarak arayarlandığı için Toy-GWR algoritması bu deneysel çalışmaya dahil edilememiştir.

Algoritmaların çalışma zamanları Tablo 2.2 ve Şekil 2.9'da verilmiştir. FastGWR algoritması, veri kümesindeki herhangi bir güncelleme işleminde tüm regresyon

Tablo 2.2. Tek çekirdekli bilgisayar üzerinde 200.000 veri kümesinde farklı veri güncellenme oranları için algoritmaların çalışma zamanları (saniye)

Güncellenme oranı	RNN-GWR	FastGWR
İlk çalışma	2541,60	6136,41
%10	23,02	6136,41
%20	89,72	6136,41
%30	152,88	6136,41
%40	285,04	6136,41
%50	535,06	6136,41
%60	737,97	6136,41
%70	1028,26	6136,41
%80	1512,62	6136,41
%90	2034,03	6136,41
%100	2595,92	6136,41



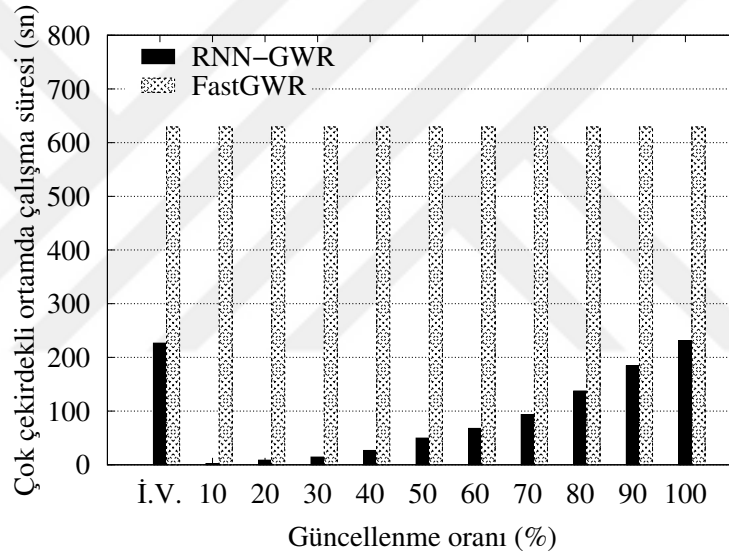
Şekil 2.9. Tek çekirdekli bilgisayar üzerinde farklı veri güncellenme oranları için RNN-GWR ve FastGWR'nin çalışma zamanı karşılaştırması

noktası için çalıştığından hesaplama maliyeti Şekil 2.9'da gösterildiği gibi tüm veri güncellenme oranları için her zaman aynıdır. RNN-GWR modelinde güncellemeden etkilenen regresyon noktaları RNN sorgusu sayesinde bulunur ve sadece bu noktaların değerleri yeniden hesaplanır. 200.000 ölçüm noktası için deneysel çalışmalar incelendiğinde RNN-GWR algoritması, güncellenme oranı %10 olduğu durumda FastGWR algoritmasından 266 kat daha hızlı, %20 olduğu durumda yaklaşık olarak 68 kat daha hızlı, %50 olduğu durumda yaklaşık 11 kat daha hızlı ve %90 olduğu durumda

ise yaklaşık 3 kat daha hızlı çalışmaktadır.

2.5.2.4. Çok Çekirdekli Ortamda Algoritmaların Karşılaştırılması

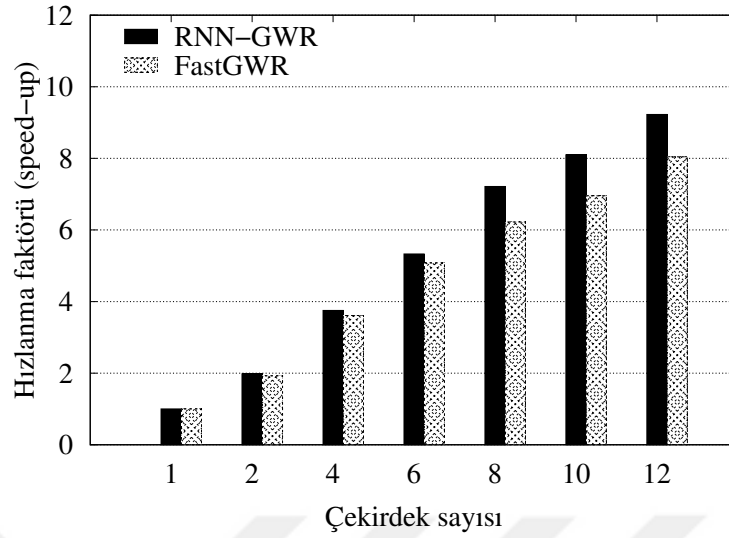
Bu çalışmada sık güncellenen verileri analiz etmek için RNN-GWR algoritması önerilmiştir ve FastGWR algoritması, klasik GWR algoritmasının çok çekirdekli veya paralel sistemlerde paralel olarak çalışabilen bir uzantısıdır. Bu bölümde, FastGWR ve önerilen RNN-GWR algoritmalarının performansları çok çekirdekli ortamda değerlendirilmiştir. Şekil 2.10 gösterildiği üzere ölçüm noktalarının sayısı 200.000 olarak ayarlanmıştır ve veri güncelleme oranları %10'dan %100'e kadar değiştirilmiştir.



Şekil 2.10. Çok çekirdekli bilgisayar üzerinde farklı veri güncellenme oranları için RNN-GWR ve FastGWR'nin çalışma zamanı karşılaştırması

Herhangi bir regresyon noktasının değerini tahmin etmek, başka bir regresyon noktasının değerini tahmin etme işleminden bağımsızdır ve bu nedenle regresyon noktaları aynı anda tahmin edilebilir. RNN-GWR yaklaşımında Algoritma 2.2'de Adım 3 ve Algoritma 2.3'te Adım 2 çok çekirdekli ortamda paralel olarak çalışmaktadır. Şekil 2.11 güncellenme oranı %100 olduğu durumda her iki algoritmanın artan çekirdek sayısına göre hızlanmalarını (speed-up) gösterir.

Şekil 2.10'de 12 çekirdekli ortamda her iki yaklaşımın değişen güncellenme oranlarına göre çalışma zamanları sunulmuştur. Şekillerden anlaşılacağı üzere RNN-GWR algoritması FastGWR algoritmasından daha iyi performans göstermektedir.

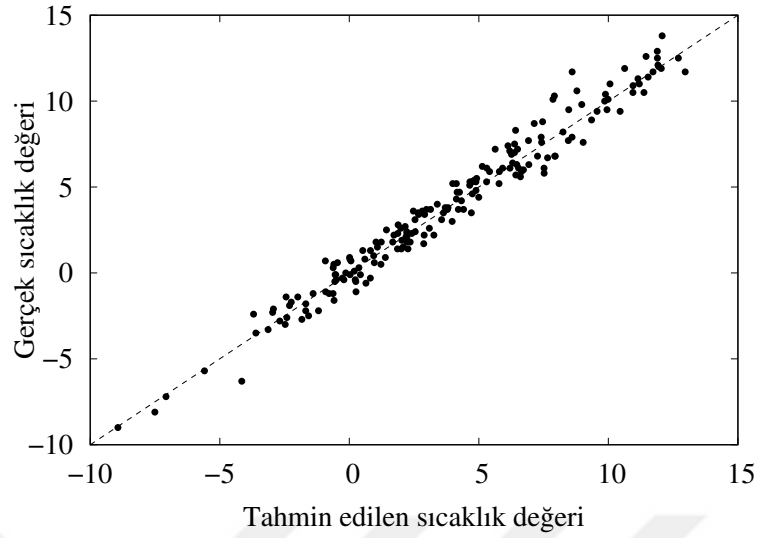


Şekil 2.11. Veri güncellenme oranı %100 olduğunda RNN-GWR ve FastGWR algoritmalarının artan çekirdek sayısına göre hızlanmaları

2.5.2.5. Gerçek ve Tahmin Edilen Değerlerin Karşılaştırılması

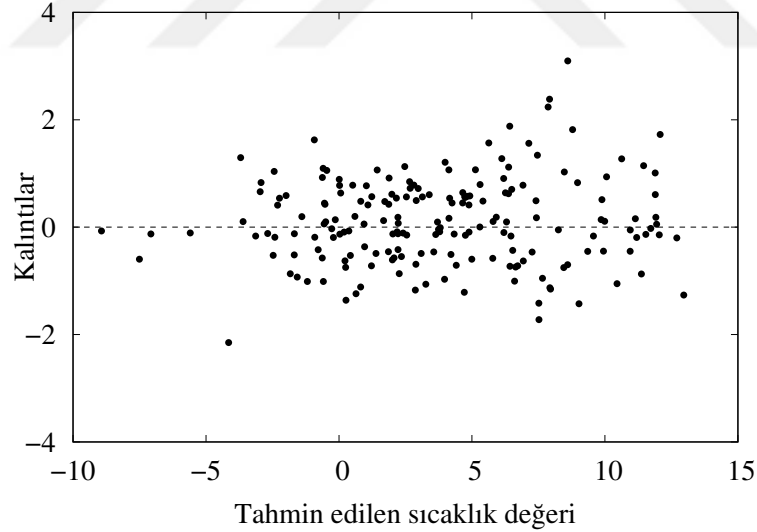
Bu bölüm, tahmin edilen ve gerçek sıcaklık değerlerinin karşılaştırmasını sunmaktadır. Toy-GWR, FastGWR ve RNN-GWR algoritmaları Eşitlik 2.4’de aynı değeri üretirler. Çünkü bu yaklaşımlar, Eşitlik 1.3’de aynı bant genişliği değerini, referans noktalarını ve regresyon noktalarını kullanmaktadır. Bu yaklaşımlar arasındaki tek fark, hesaplama maliyetlerinin farklı olmasıdır. Aslında, RNN-GWR algoritması ilk olarak RNN sorgusunu kullanarak güncellemeden etkilenen regresyon noktalarını belirler. Daha sonra RNN-GWR yaklaşımı da diğer yaklaşımlar gibi tüm regresyon noktaları için değerler üretir. Karşılaştırmada GWR yaklaşımının meteorolojik veri kümesine uygunluğunu görmek için meteoroloji parametrelerinin çoğunun değiştiği zaman dilimi seçilmiştir. Bu analizde sıcaklık değerleri saat 06:00 ile 07:00 arasında en az 1°C değişen 189 adet istasyon kullanılmıştır ve Şekil 2.6’da gösterilmiştir. Şekil 2.6’da görüldüğü gibi sıcaklığı değişen istasyon sayısı bu zaman diliminde maksimum seviyeye ulaşmıştır.

Gerçek sıcaklık değerlerine karşılık tahmin edilen sıcaklık değerleri Şekil 2.12’de gösterilmiştir. Şekil 2.12’de görüldüğü üzere değerler diyagonal etrafında toplanmaktadır. Şekil 2.12’de görüldüğü üzere gerçek ve tahmin edilen değerler genellikle iyi eşleşmiştir. Gerçek ve tahmin edilen sıcaklık değerleri arasında korelasyon katsayısı 0.98 olarak hesaplanmıştır.



Şekil 2.12. 6:00 - 7:00 zaman diliminde sıcaklık değerleri en az 1°C değişen meteoroloji istasyonlarının tahmini ve gerçek sıcaklık değerleri

Tahmin edilen ve gerçek sıcaklık değerleri arasındaki fark olarak tanımlanan kalıntılar hesaplanarak Şekil 2.13'te gösterilmiştir.



Şekil 2.13. 6:00 - 7:00 zaman diliminde sıcaklık değerleri en az 1°C değişen meteoroloji istasyonlarının kalıntıları

Şekil 2.13'te kalıntılar x eksenine etrafında toplanmıştır. RMSE $\pm 0.81^{\circ}\text{C}$ olarak hesaplanmıştır. Bu, tahmin edilen sıcaklık değeri gerçek sıcaklık değerinden en fazla $\pm 0.81^{\circ}\text{C}$ farklı olacağını anlamına gelmektedir.

2.6. Sonular

Bu alıřmada, sık gncellenen veri kmelerini analiz etmek iin ters en yakın komřu (RNN) stratejisini kullanan hesaplama aısından etkili bir GWR algoritması (RNN-GWR) nerilmiřtir. nerilen RNN-GWR algoritması, MGM'nin gerek veri kmesini ve bu kmeden oluřturulan sentetik veri kmesini kullanarak Toy-GWR ve FastGWR algoritmaları ile karřılařtırılmıřtır. Deneysel deęerlendirmeler, nerilen yaklařımın sık gncellenen verilerin analiz edilmesi aısından Toy-GWR ve FastGWR yaklařımlarından daha etkili olduęunu gstermektedir. RNN-GWR algoritması daha byk ve sık gncellenen veri kmelerine uygulanabilirken, Toy-GWR yaklařımı byk veri kmelerini analiz edememektedir. Ayrıca, nerilen RNN-GWR algoritması, byk veri kmelerinde FastGWR algoritmasından daha iyi performans gstermektedir.

3. BÖLÜM

FastGTWR: HIZLI COĞRAFI VE ZAMANSAL AĞIRLIKLIL REGRESYON YAKLAŞIMI

3.1. Giriş

Mekânsal analizler günümüzde önemli hale gelmiştir ve çok farklı uygulama alanlarında kullanılmaktadır. Sensör kullanımının ve nesnelerin internetinin yaygınlaşması, nesnelerinin akıllı şehir altyapısını oluşturmaya başlaması büyük miktarlarda ve çok fazla sayıda parametresi olan verilerin toplanmasına imkân vermiştir [1, 31, 86–89]. Hava kalitesini, hava kirliliğini ve iklim bilgilerini oluşturan sıcaklık, basınç, nem, rüzgâr yönü-hızı, azot miktarı, ozon, karbondioksit, karbon monoksit, toz konsantrasyonu, gürültü, güneş radyasyon miktarı ve yağış miktarı gibi veriler toplanan verilerden bazılarıdır. Mekân ve zaman bilgilerini içeren ve çok fazla sayıda parametresi olan bu veriler analiz edildiğinde faydalı bilgiler elde edilmektedir [90–94]. Yaygın olarak kullanılan mekânsal analiz yöntemlerinden biri olan GWR, coğrafya üzerindeki değişen ilişkileri modellemek için kullanılan bir yerel mekânsal regresyon tekniğidir [15]. Veri kümesinde mekân-zamansal bilgilerin olduğu durumlarda coğrafi ve zamansal ağırlıklı regresyon yaklaşımı, GTWR, GWR yaklaşımına göre daha iyi modeller üretmektedir [28–30]. GTWR yaklaşımı GWR yaklaşımına zamansal ilişkileri de ekleyerek geliştirilen bir yaklaşımdır [28, 29]. Klasik GWR modelinden farklı olarak GTWR modeli, mekân-zamansal heterojenliği yakalamak için ağırlık matrislerine hem mekânsal hem de zamansal bilgileri entegre etmektedir.

GWR yaklaşımları her bir regresyon noktası için farklı katsayılarından oluşabildiğinden, bu yaklaşımların hesaplama maliyetleri yüksektir [26, 31, 74, 95]. GTWR yaklaşımı, GWR yaklaşımına göre mekân-zamansal veriler için daha iyi modeller üretmesine

rağmen [30, 39, 69, 70] geliştirilen mekân-zamansal modellerin karmaşıklığı göz önüne alındığında algoritma karmaşıklığı artmaktadır. Literatürde koşturulan GTWR modelleri sınırlı sayıdaki veri üzerinde çalışabilmiştir. Yapılan bir çalışmada GTWR modelleri tarafından ele alınabilecek maksimum veri kümesi standart masaüstü bilgisayarlarda yaklaşık olarak 15.000 ölçüm noktası olarak tespit edilmiştir [26, 31]. Veri çeşitliliğinin ve boyutunun arttığı günümüzde sınırlı sayıda ölçüm verisini analiz edebilmesi GTWR kullanımı açısından ciddi bir sorun oluşturmaktadır.

Literatürde, GWR ve GTWR yaklaşımlarının hesaplama maliyetini azaltmak için dağıtılmış ve paralel platformlarda çalışmalar yapılmıştır. Ancak bu yöntemlerin klasik bilgisayarlarda çalışması zordur. Bu tez çalışmasında GWR yaklaşımının hesaplama maliyeti, W komşuluk matrisi vektör formatında düzenlenerek ve matrislerin simetrik özellikleri kullanılarak azaltılmıştır [31]. Bu bölümde, bahsedilen özellikler kullanarak GTWR algoritmasının hesaplama ve depolama maliyeti azaltılmıştır.

Bu çalışmada, GTWR algoritmasının hızını arttırmak ve veri boyutu kısıtlamasının üstesinden gelmek için FastGTWR algoritması önerilmiştir. Yapılan incelemede GTWR yaklaşımının hesaplama ve depolama maliyetinin $X^T W(u_i, v_i, t_i) X$ matris çarpımından ve bu matrisin tersinin alınmasından kaynaklandığı tespit edilmiştir. Bu matris çarpımındaki $W(u_i, v_i, t_i)$ matrisi $n \times n$ boyutunda komşuluk matrisidir (n regresyon nokta sayısıdır). Önerilen FastGTWR yaklaşımında, belirtilen matris işlemlerinin karmaşıklığını azaltmak için W matrisi yeniden düzenlenerek vektör formatına çevrilmiştir ve matris çarpımları noktasal olarak yapılmıştır. Buna ek olarak $X^T W(u_i, v_i, t_i) X$ matris işleminin simetrik bir matris olduğunu göz önüne alarak bütün değerleri hesaplamak yerine sadece diyagonalın alt veya üst tarafındaki değerler hesaplanmıştır. Bu sayede GTWR yaklaşımının hesaplama ve depolama karmaşıklığı azaltılmıştır.

Bu çalışmanın ana katkıları aşağıdaki gibidir.

- GTWR algoritmasının veri boyutu kısıtlamasının üstesinden gelmek için FastGTWR algoritması önerilmiştir.
- Önerilen FastGTWR algoritmasının performansı, Bölüm 1.3.1’de bahsedilen

MGM'den alınan alınan gerçek ölçüm verileri kullanarak klasik GWR ve GTWR yaklaşımlarının performansları ile karşılaştırılmıştır.

- Deneysel sonuçlar önerilen FastGTWR yaklaşımının klasik yaklaşımlara göre daha hızlı çalıştığını ortaya koymuştur.
- Çalışma sonuçları derlenerek dergiye gönderilmiştir.

Bu çalışmanın bundan sonraki bölümlerinde, Bölüm 3.2'de GTWR yöntemleri tanıtılmış, Bölüm 3.3'te önerilen FastGTWR algoritması tartışılmış ve Bölüm 3.4'te deneysel çalışmalar paylaşılmış ve Bölüm 3.5'te sonuçlar sunulmuştur.

3.2. Coğrafi ve Zamansal Ağırlıklı Regresyon

GWR yöntemi coğrafya üzerindeki değişen ilişkileri modellemek için kullanılan bir yerel mekânsal regresyon tekniğidir [15]. Diğer bir ifade ile GWR coğrafya üzerinde konumu ve özellikleri bilinen referans noktaları temel alarak konumu bilinen diğer noktalar için regresyon yöntemi ile tahmini değerler üreten bir mekânsal veri madenciliği yöntemidir [16, 18, 90–93]. GWR, Eşitlik 1.1'de bulunan klasik çoklu regresyon modelini temel alır. Eşitlik 1.2'de ifade edildiği üzere klasik regresyon modelinden farklı olarak, GWR modelinde katsayılar sabit değildir. Her mekânsal noktanın kendisine ait katsayıları oluşturulmaktadır [15, 39, 40, 96]. GWR yaklaşımı mekânsal değişiklikleri ölçmek için model boyunca β_{ik} regresyon katsayılarının değişmesine izin verir. Bununla birlikte mekân ve zaman bilgisinin olduğu veri kümelerinde zamansal değişiklikleri modelleyemez. Huang vd. [28] mekânsal ve zamansal değişikliklerin eşzamanlı olarak değerlendirilmesi için GWR modelinden genişletilen ve Eşitlik 3.1 ile ifade edilen GTWR modelini önermiştir. GTWR yaklaşımı mekân-zamansal ilişkiyi ölçmek için mekân-zamansal mesafeyi kullanarak bir mekân-zamansal ağırlık matrisi oluşturur.

$$y_i(u_i, v_i, t_i) = \beta_{i0}(u_i, v_i, t_i) + \sum_{k=1}^m (\beta_{ik}(u_i, v_i, t_i)x_{ik}(u_i, v_i, t_i) + \epsilon_i(u_i, v_i, t_i)) \quad (3.1)$$

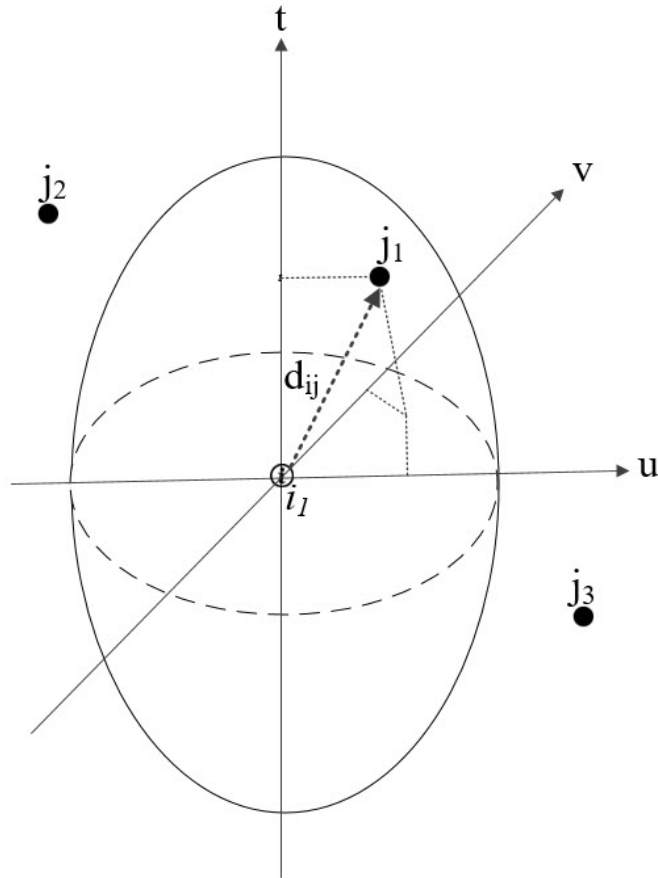
GWR formülünden farklı olarak formüle t zaman bilgisi eklenmiştir. $\beta_{ik}(k = 0, \dots, m)$ regresyon katsayısı GWR yaklaşımında olduğu gibi Eşitlik 3.2 ile hesaplanır.

$$\beta(u_i, v_i, t_i) = (X^T W(u_i, v_i, t_i) X)^{-1} X^T W(u_i, v_i, t_i) Y \quad (3.2)$$

$W(i) = W(u_i, v_i, t_i)$ ağırlık matrisi, klasik GWR olduğu gibi w_{ij} değerlerinden oluşur. Fakat burada w_{ij} mekân-zamansal komşuluk oranı değeridir. x ve y nokta koordinatlarını, t zaman bilgisini ve d ise dengeleme katsayısını göstermek üzere d_{ij} mekân-zamansal mesafesi Eşitlik 3.3'te gösterilen formül ile hesaplanır [28].

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2 + d(t_i - t_j)^2} \quad (3.3)$$

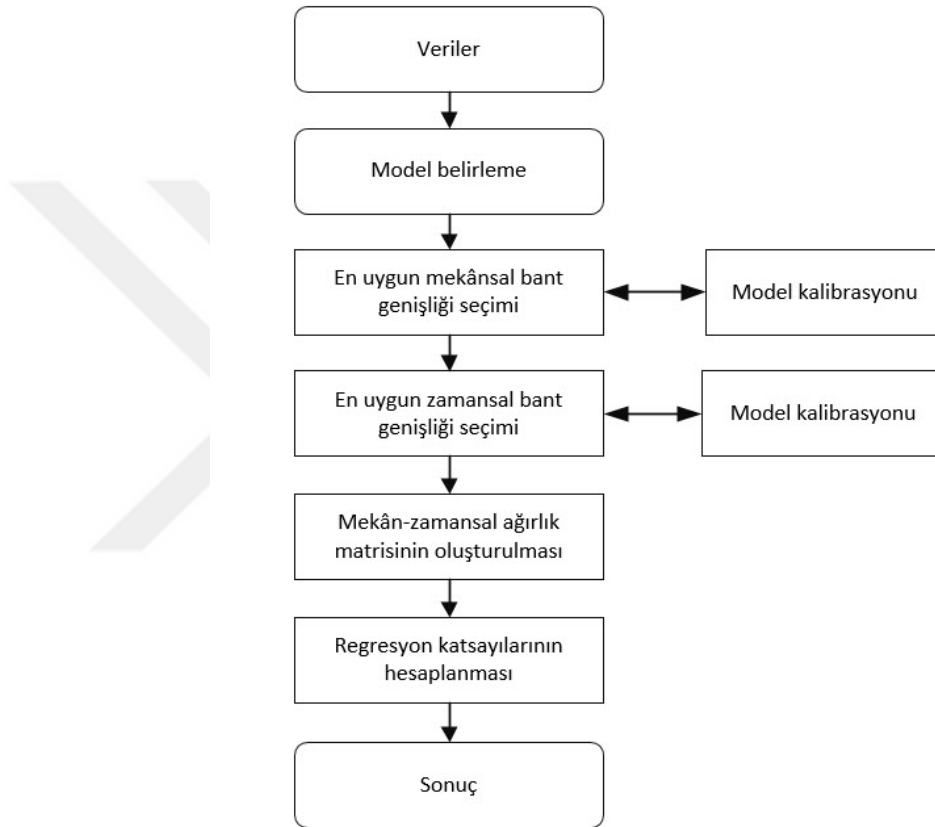
i_1 regresyon noktası ile j_1 referans nokta arasındaki mekân-zamansal mesafe Şekil 3.1'de gösterilmiştir.



Şekil 3.1. Regresyon noktası i_1 ile referans noktası j_1 arasındaki mekân-zamansal mesafe

Mekân-zamansal ağırlık matrisleri oluşturulmadan önce mekân-zamansal mesafenin uyumlu hale getirilmesi gerekmektedir. Eşitlik 3.3'te farklı mekânsal ve zamansal

birimleri dengelemek veya uyumlu hale getirmek için d parametresi eklenmiştir [28]. Bu nedenle, GTWR yaklaşımını uygulamadan önce d parametresinin optimize edilmesi gerekmektedir. Huang vd. [28] bu parametreyi bütün ihtimalleri deneyerek bulmayı önermiştir. Fotheringham vd. [29] ise bu işlemi algoritma adımları ile gerçekleştirmiştir. GTWR algoritmasının genel çalışma yapısı Şekil 3.2’de gösterilmiştir [29].



Şekil 3.2. GTWR algoritması

Şekil 3.2’de görüldüğü üzere giriş veri kümesine göre Eşitlik 3.1 kullanılarak GTWR modeli oluşturulur. Burada hangi parametrelerin bağımsız değişken olacağına ve hangi parametrenin bağımlı değişken olacağına karar verilir. Daha sonra, genellikle çapraz doğrulama (cross-validation - CV) veya Akaike bilgi kriteri (Akaike information criterion - AIC) metodu ile modele en uygun mekânsal bant genişliği belirlenir. Bulunan en uygun mekânsal bant genişliği de kullanılarak CV veya AIC yöntemi ile modele en uygun zamansal bant genişliği bulunur. Bu işlemden sonra mekân-zamansal bant genişliği kullanılarak mekân-zamansal komşuluğu içeren ağırlık matrisi oluşturulur. Oluşturan

bu ağırlık matrisi, bağımsız değişkenlerin matrisi ve bağımlı değişkenlerin matrisi de kullanılarak Eşitlik 3.2 ile regresyon katsayıları belirlenir. Daha sonra bu regresyon katsayıları kullanılarak regresyon noktası için değer üretilir.

3.3. FastGTWR: Önerilen Hızlı Coğrafi ve Zamansal Ağırlıklı Regresyon Yaklaşımı

GTWR yaklaşımı, mekân-zamansal veriler için GWR yaklaşımına göre çok daha iyi modeller üretmesine rağmen mekânsal-zamansal modelin karmaşıklığı göz önüne alındığında algoritma karmaşıklığı artmaktadır [30, 69, 70]. Bu çalışmada, GTWR algoritma karmaşıklığını azaltmak ve dolayısı ile veri boyutu kısıtlamasının üstesinden gelmek için dağıtılmış veya eş zamanlı çalışmaya ihtiyaç duymadan ve GTWR yaklaşımının doğruluğundan ödün vermeyen FastGTWR modeli önerilmiştir.

GWR yaklaşımların hesaplama maliyetinin çoğunluğu Eşitlik 3.2’de bulunan matris işlemlerinden kaynaklanmaktadır [26, 31]. n regresyon sayısını göstermek üzere Eşitlik 3.2 bütün noktaların katsayılarını bulmak için n kez çalıştırılmaktadır. Eşitlik 3.2’de hesaplama maliyetini en çok etkileyen işlem $X^T W(u_i, v_i, t_i) X$ matris çarpımı ve bu matrisin tersinin $(X^T W(u_i, v_i, t_i) X)^{-1}$ hesaplanması işlemidir. m bağımsız değişken sayısını göstermek üzere $X^T W(u_i, v_i, t_i) X$ işleminin hesaplama maliyeti $O(m^2 n)$ ve $(X^T W(u_i, v_i, t_i) X)^{-1}$ hesaplama maliyeti de $O(m^3)$ olarak ifade edilmektedir [26, 31]. m değeri genellikle n değerinden çok daha küçük olduğu için Eşitlik 3.2 işleminin hesaplama maliyeti $O(m^2 n)$ olarak kabul edilir ve bu işlemde her bir regresyon nokta sayısı olan n kadar tekrar edeceği için GTWR yaklaşımın hesaplama maliyeti $O(m^2 n^2)$ olarak ifade edilir. Eşitlik 3.2 hesaplama maliyetinden farklı olarak hafızda çok yer kaplayan bir işlemdir ve dolayısı ile depolama maliyeti yüksektir ve bu yüzden klasik GWR ve GTWR yaklaşımları sınırlı sayıda veriyi işleyebilmektedir. Bu sorunun üstesinden gelmek için Li vd. [26] matris formlarının yeniden düzenlenmesini önermiştir.

$$\begin{aligned}
W(u_i, v_i, t_i) &= W(i) \\
\beta(i) &= (X^T W(i) X)^{-1} X^T W(i) Y \\
X^T W(i) &= (X \cdot w(i)^T)^T \\
w(i) &= (w_{i1}, \dots, w_{in})
\end{aligned} \tag{3.4}$$

Eşitlik 3.4’de gösterildiği gibi $W(u_i, v_i, t_i)$ matrisinin $n \times n$ boyutta hafızada yer kaplayıp X matrisi ($n \times m$ boyutta) ile çarpmak yerine $W(i)$ matrisini vektör formatına ($w(i) = (w_{i1}, \dots, w_{in})$) çevirip noktasal olarak çarpılması önerilmiştir. $X^T W(i)$ matris formunun $(X \cdot w(i)^T)^T$ olarak yeniden düzenlenmesi $n \times n$ boyutta olan diyagonal $W(i)$ matrisini $1 \times n$ boyutta $w(i)$ vektör satırına dönüştürür ve hafıza karmaşıklığını $O(n^2)$ boyuttan $O(mn)$ boyuta düşürür [26]. Bu işlem sayesinde Eşitlik 3.2’de bulunan hafıza sorunu çözülmektedir ve ihtiyaç olmayan işlemler ortadan kalkacağından dolayı ($W(i)$ diyagonal matrisinde bulunan 0 değer ile çarpılan işlemlerin kaldırılması) çarpım işlemi daha hızlı olarak yapılabilir. Önerilen FastGTWR yaklaşımında Eşitlik 3.4’te verilen formül kullanılmıştır.

Bölüm 2’de GWR hesaplama maliyetinin büyük kısmını oluşturan $X^T W(u_i, v_i, t_i) X$ matrisinin simetrik bir matris olduğu, bütün değerleri hesaplamaya gerek olmadığı ve diyagonalın (köşegenin) alt veya üst tarafını hesaplamamanın yeterli olduğu belirtilmiştir. Önerilen FastGTWR yaklaşımında simetrik matris özelliğinden faydalanılmıştır ve bu sayede FastGTWR’nin maliyeti $O\left(\left(\frac{m^2+m}{2}\right) n^2\right)$ olmaktadır [26].

3.4. Deneysel Değerlendirmeler

Bu bölümde GWR, GTWR ve önerilen FastGTWR algoritmalarının başarısını göstermek için yapılan deneysel değerlendirilmelerden bahsedilmektedir. Bölüm 1.3.1’de bahsedilen MGM’den alınan gerçek ölçüm verileri ve Bölüm 1.3.2’de bahsedilen sentetik veri kümesi kullanılmıştır. Veri kümesi ve model ayarları ve deneysel değerlendirme ile ilgili detaylı bilgiler sırasıyla Bölüm 3.4.1 ve 3.4.2’de verilmiştir.

Deneysel değerlendirmeler aşağıdaki sorulara cevap vermeye odaklanmıştır.

- Yaklaşımların doğruluğu nedir?
- Yaklaşımların çalışma zamanı nedir?
- Veri boyutunun yaklaşımlar üzerindeki etkisi nedir?

Yaklaşımlar Intel Core i5-3210M 2.5 GHz 8GB RAM özelliklerine sahip bir bilgisayar üzerinde çalıştırılmış ve Matlab’da koşturulmuştur.

3.4.1. Veri Kümesi ve Model Ayarları

Bu bölümde, ilk önce deneysel değerlendirmelerde kullanılan veri kümesi hakkında bilgi verilmiştir. Daha sonra yaklaşımlar için kullanılan model ayarları sunulmuştur.

3.4.1.1. Veri Kümesi

Bölüm 1.3.1’de bahsedilen veri kümesi meteoroloji istasyonlarının rakımını ve saatlik ölçüm olarak hava sıcaklığı, bağıl nem ve hava basınç değerlerini içermektedir. 1746 adet MGM istasyonu arasından seçilen 1600 adet ölçüm noktasından oluşan veri kümesi 01-30 Aralık 2016 tarih aralığına ait yaklaşık olarak 1.1 milyon saatlik ölçüm verilerinden oluşmaktadır. Bu çalışmada, deneylerde kullanılmak üzere rastgele 1600 istasyon seçilmiştir. İstasyonların 1200’ü referans noktası, 400’ü regresyon noktası olarak kullanılmıştır. Bu çalışmada kullanılan istasyon noktaları Şekil 1.5’te gösterilmiştir.

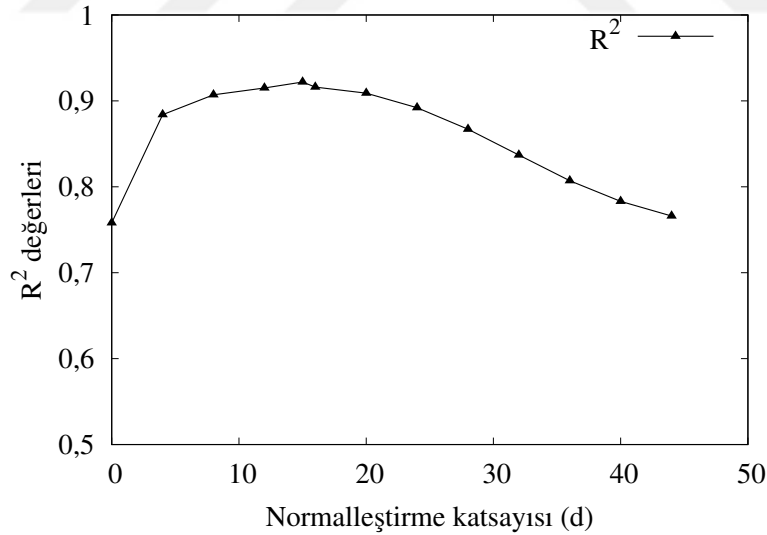
Ölçüm nokta sayısının (veri boyutu) algoritmaların performansına etkisini görmek için 1.000.000 ölçüm noktası içerecek şekilde sentetik veri kümesi oluşturulmuştur. Ölçüm noktalarının yeri ve parametre değerleri rastgele oluşturulmuştur.

3.4.1.2. Model Ayarları

Deneylerdeki amaç bir noktanın ölçülen nem, basınç ve rakım değerini kullanarak sıcaklık değerini tahmin etmektir. Sıcaklık değerini tahmin etmek için GWR, GTWR ve FastGTWR yaklaşımlarında kullanılan model Eşitlik 2.4’de gösterilmiştir [31].

GWR modelinde öncelikle çekirdek fonksiyonu ve bant genişliği değeri belirlenmelidir. Bu çalışmada, GWR yaklaşımlarında yaygın olarak kullanılan ve GTWR yaklaşımlarında

da Huang vd. [28] ve Fotheringham vd. [29] tarafından kullanılan Gaussian çekirdek fonksiyonu kullanılmıştır. En uygun mekânsal bant genişliği değerini bulmak için sık kullanılan düzeltilmiş Akaike bilgi kriteri, AIC_c , yöntemi tercih edilmiştir [15,39]. GWR yaklaşımı için AIC_c yöntemi ile en uygun mekânsal bant genişliği 82 km (Öklid mesafesi) bulunmuştur. GTWR ve FastGTWR yaklaşımları için en uygun mekân-zamansal bant genişliği değeri 80 km, Eşitlik 3.3'te dengeleme katsayısı olan $d = 15.01$ olduğunda bulunmuştur. Eşitlik 3.3'te bulunan d parametresi farklı mekânsal ve zamansal birimleri dengelemek veya uyumlu hale getirmek için kullanılmaktadır [28]. Huang vd. [28] d değerinin uyum grafiğini R^2 değerine göre göstermiştir. Bu çalışmada kullanılan veri kümesine göre d değerlerinin oluşturduğu R^2 grafiği Şekil 3.3'te gösterilmiştir. R^2 değerleri hesaplanırken Eşitlik 3.3'te $t_i - t_j$ arasındaki zaman farkı saniye cinsinden alınmıştır. R^2 bağımsız bir değişken veya bir regresyon modelindeki değişkenler tarafından açıklanan bağımlı bir değişkenin varyans oranını temsil eden istatistiksel bir ölçüdür [97–100]. R^2 değeri 1'e ne kadar yakınsa model o kadar iyi temsil edilmiştir. Şekil 3.3'te $d = 15.01$ olduğunda R^2 değeri en yüksek seviyeye ulaşmıştır.



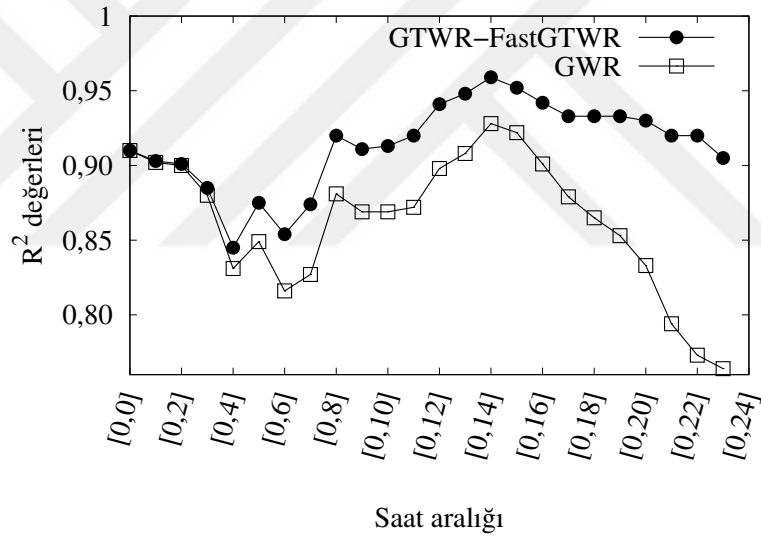
Şekil 3.3. GTWR modeli için parametre seçimi

3.4.2. Deneyler

3.4.2.1. Yaklaşımların Doğruluğu

Bu bölümde, yaklaşımlar tarafından tahmin edilen sıcaklık değeri ile gerçek sıcaklık

değerlerinin karşılaştırması yapılmıştır. Bu karşılaştırma genellikle R^2 değerine göre yapılmaktadır [97–100]. GWR, GTWR ve FastGTWR yaklaşımları Eşitlik 2.4'ü kullanmaktadır. GTWR ve FastGTWR yaklaşımları aynı doğruluk değerlerini oluşturmaktadır. Bu nedenle Şekil 3.4'te GTWR ve FastGTWR yaklaşımlarının davranışı aynıdır. GWR ile FastGTWR yaklaşımları farklı sonuçlar üretmektedir. Bunun sebebi GWR yaklaşımı sadece mekânsal ilişkileri incelerken GTWR ve FastGTWR yaklaşımlarının mekânsal ve zamansal ilişkileri incelemesidir. Bu deneyde, algoritmaların doğruluğu her bir saat için incelenmiştir. Referans ve regresyon noktalarının sayısı, sırasıyla, 1200 ve 400 olarak ayarlanmıştır. Şekil 3.4'te gösterildiği üzere, her bir saat için regresyon nokta sayısı sabit kalırken dikkate alınan ölçüm sayısı 1200 artmaktadır.

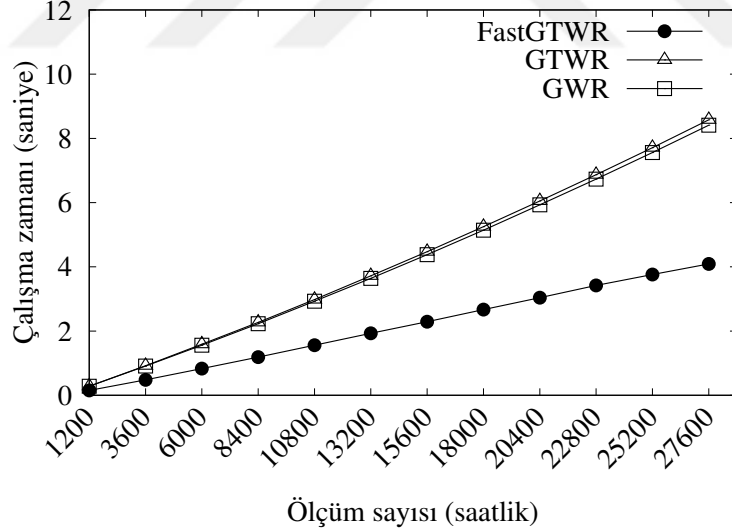


Şekil 3.4. Yaklaşımların doğruluğu

Şekil 3.4'te ilk çalışma saatinde geçmişe dair bir veri olmadığı için GWR, GTWR ve FastGTWR yaklaşımları aynı sonuçları üretmektedir. İlerleyen saatlerde geçmiş veriler dikkate alınırken GTWR ve FastGTWR yaklaşımları geçmiş verileri mekân-zamansal ağırlıkları ile dikkate almaktadır. GWR yaklaşımı ise verilerde zamansal fark olmasına rağmen sadece mekânsal ağırlıkta dikkate almaktadır. Şekil 3.4'te görüldüğü üzere zaman süresi arttıkça GTWR ve FastGTWR yaklaşımları GWR yaklaşımına göre daha doğru sonuçlar üretmektedir.

3.4.2.2. Yaklaşımların Çalışma Zamanı

Bu deneyde, her saat için GWR, GTWR ve FastGTWR algoritmalarının çalışma zamanı performansları gerçek veri kullanılarak değerlendirilmiştir. Referans ve regresyon noktalarının sayısı, sırasıyla, 1200 ve 400 olarak ayarlanmıştır. Ölçüm noktalarının konumu sabit olduğundan dolayı Bölüm 3.4.1.2’de bahsedilen bant genişliği değerleri kullanılmıştır ve dolayısı ile en uygun bant genişliği hesaplama işlemi maliyeti algoritma maliyetlerine dâhil edilmemiştir. Önerilen FastGTWR yaklaşımında simetrik matris özelliğinden faydalanılmıştır ve bu sebeple Şekil 3.5’te gösterildiği üzere hem klasik GWR hem de GTWR yaklaşımından daha iyi performans göstermiştir. GWR yaklaşımı GTWR yaklaşımına göre daha hızlıdır çünkü GTWR yaklaşımın komşuluk ağırlıklarının hesaplama maliyeti GWR yaklaşımın komşuluk ağırlığı hesaplama maliyetine göre az da olsa fazladır. Her bir saat için regresyon nokta sayısı sabit kalırken dikkate alınan ölçüm sayısı 1200 artmaktadır. Bu deneyde regresyon nokta sayısı sabit kaldığından deney veri kümesi boyutunun algoritmaların performansı üzerindeki etkisini göstermemektedir.



Şekil 3.5. Yaklaşımların saat bazlı çalışma zamanı

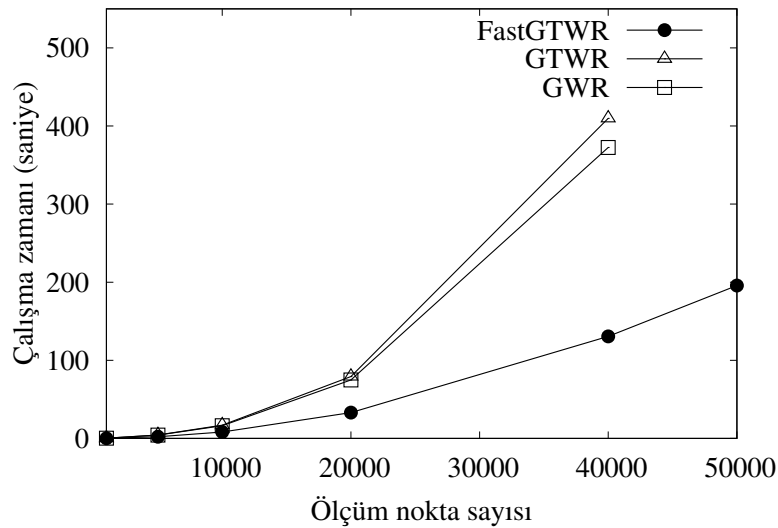
3.4.2.3. Veri Boyutunun Yaklaşımlar Üzerindeki Etkisi

Bu deneyde, veri boyutunun GWR, GTWR ve FastGTWR yaklaşımlarının üzerindeki etkisi değerlendirilmiştir. MGM’den alınan veri kümesindeki istasyon sayısı, algoritmaların veri büyüklüğüne göre performansını ölçmek için azdır ve bu sebeple konumları ve ölçüm değerleri rastgele oluşturularak bu verilerden 1.000.000 adet ölçüm

noktası içeren sentetik veri kümesi oluşturulmuştur. Deneylerde önceden tanımlanmış bant genişliği değerleri kullanılmıştır. Deney sonuçları Tablo 3.1 ve Şekil 3.6'da verilmiştir. Veri kümesi boyutları arttıkça tüm algoritmaların hesaplama maliyeti de artmaktadır. Li vd. [26] klasik GWR yaklaşımlarının normal bir bilgisayarda en fazla 15.000 ölçüm noktası işleyebildiğini tespit etmişlerdir. Fakat bu özellik bilgisayara göre değişebilmektedir ve bizim kullandığımız bilgisayarda 40.000 ölçüm noktası olarak tespit edilmiştir. Klasik GWR ve GTWR algoritması, Tablo 3.1 ve Şekil 3.6'de görüleceği gibi 40.000 ölçüm noktasına kadar veri kümelerini işleyebilmektedir.

Tablo 3.1. 1.000 ile 1.000.0000 arasında ölçüm noktası kullanan algoritmaların çalışma zamanları (saniye - sn)

Ölçüm nokta sayısı	FastGTWR (sn)	GWR (sn)	GTWR (sn)
1.000	0,14	0,27	0,27
5.000	2,18	4,27	4,36
10.000	8,15	16,25	19,90
20.000	33,01	74,78	79,27
40.000	130,58	372,54	409,80
50.000	195,66	-	-
100.000	819,03	-	-
200.000	4.103,17	-	-
500.000	25.999,63	-	-
1.000.000	98.239,75	-	-



Şekil 3.6. Yaklaşımların çalışma zamanı

Önerilen FastGTWR algoritması, GWR ve GTWR algoritmalarından daha iyi performans

göstermiştir. Bunun sebebi, ilk olarak, Eşitlik 3.4'te bulunan yeni dönüşümü kullanmasıdır. Eşitlik 3.4 sayesinde FastGTWR algoritması 1.000.000 ölçüm noktasına kadar veri kümesini işleyebilmektedir. Bu nedenle önerilen FastGTWR yaklaşımı büyük veri üzerinde çalışabilmekte ve daha az hafıza kullandığından dolayı hafıza yönetimini daha etkin yaparak hızlı cevap verebilmektedir. İkinci olarak ise önerilen FastGTWR yaklaşımında, simetrik matris özelliğinden faydalanılmıştır ve bu sayede simetrik matris olan $X^T W(u_i, v_i, t_i) X$ matrisinde diyagonalın alt veya üst kısmı hesaplanmadığından önerilen yaklaşımının hesaplama maliyeti ciddi oranda azalmaktadır.

3.5. Sonuçlar

Bu çalışmada, mekân-zamansal ilişkileri içeren veri kümeleri için hesaplama açısından etkili bir GTWR (FastGTWR) algoritması önerilmiştir. Önerilen FastGTWR yaklaşımının performansı, GWR ve GTWR yaklaşımlarının performansları ile karşılaştırılmıştır. Deneysel değerlendirmeler, önerilen FastGTWR yaklaşımının klasik GWR ve GTWR yaklaşımlarından hesaplama olarak etkili olduğunu göstermiştir.

4. BÖLÜM

4D-GWR: COĞRAFİ, RAKIMSAL VE ZAMANSAL AĞIRLIKLIL REGRESYON YAKLAŞIMI

4.1. Giriş

Teknolojinin ilerlemesi ile meteoroloji, çevre yönetimi vb. birçok uygulama alanında çok sayıda parametre değerine sahip veri kümeleri toplanmıştır. Bu veri kümeleri mekânsal (enlem, boylam), rakımsal ve zamansal ilişkileri içerir ve bu tür veri kümelerinde bulunan parametrenin değerleri mekân ve zamana göre değişir. Meteorolojik veri kümelerinde, meteorolojik parametrelerin (hava sıcaklığı, nem, basınç, vb.) değerleri konum, yükseklik ve zamana göre değişir. Farklı rakımlar için, hava kirliliği verileri zaman içinde bir şehrin PM konsantrasyonlarını ve ekolojik veriler ormanların veya bitki örtüsünün net birincil üretim değerlerini içerir. Bu tür veri kümelerinin mekânsal, rakımsal ve zamansal ilişkileri dikkate alınarak analiz edilmesi değerli bilgileri ortaya çıkarabilir ve gerekli önlemlerin alınmasına yardımcı olabilir. Örneğin, meteorolojik veri kümelerinin analizi konum, rakım ve zamanı bilinen bir noktanın hava sıcaklığı değerinin tahminine veya bir şehrin partikül madde (particulate matter - PM) konsantrasyon değerlerinin tahmini bir şehirdeki trafik akışının düzenlenmesine yardımcı olabilir.

Literatürde bu tür veri kümelerini analiz etmek için çeşitli tahmin teknikleri önerilmiştir [16]. Değişen ilişkileri modellemek ve belirli bir yerin parametre değerlerini tahmin etmek için yaygın olarak kullanılan yerel mekânsal regresyon tekniği olan coğrafi ağırlıklı regresyon, GWR, bu tekniklerden biridir. Klasik GWR esas olarak mekânsal (enlem ve boylam) veya iki boyutlu ilişkinin modellenmesi ile ilgilenir [15, 19, 20]. Veri kümelerinin rakım ve zaman gibi diğer boyutlarını dikkate almak için, literatürde yapılan birkaç çalışma bulunmaktadır. Üç boyutlu bir GWR yaklaşımı olan coğrafi ve zamansal

ağırlıklı regresyon, GTWR, veri kümelerinin hem mekânsal (enlem ve boylam) hem de zamansal ilişkilerini dikkate almak için önerilmektedir [28–30]. Diğer bir üç boyutlu yaklaşım ise veri kümelerindeki hem mekânsal hem de rakımsal ilişkilerini dikkate alması için önerilen coğrafi ve rakımsal ağırlıklı regresyon (geographically and altitudinal weighted regression - GAWR) yöntemidir [32]. Bu yöntemler, mekânsal varyasyona ek olarak rakımsal veya zamansal varyasyonun dikkate alındığı klasik GWR'den (iki boyutlu coğrafi ağırlıklı regresyon - 2D-GWR) daha iyi performans gösterir. Bu yöntemlerde, mekânsal, rakımsal ve zamansal birimleri dengelemek veya uyumlu hale getirmek için dengeleme katsayıları (exaggeration coefficients) kullanılmaktadır [28, 32]. Ayrıca, GAWR ve 2D-GWR yaklaşımlarının meteorolojik veriler üzerindeki performansının sonuçları derlenerek bir dergide yayınlanmıştır [101]. Ancak, bu yöntemler veri kümesindeki mekânsal, rakımsal ve zamansal değişiklikleri etkin bir şekilde yakalayamaz. Bu çalışma, dört boyutlu veri kümelerini analiz etmek için etkili bir GWR teknikleri geliştirmeyi ele almaktadır.

GWR yöntemi kullanılarak dört boyutlu veri kümelerini analiz etmek ile ilgili üç büyük zorluk vardır. Birincisi, mevcut yöntemler dört boyutlu (mekânsal, rakımsal ve zamansal) veri kümelerini işlemek için yeterli değildir ve bu nedenle bu boyutları dikkate almak için yeni bir yaklaşım geliştirilmelidir. İkincisi, veri kümelerinin boyutu zamanla artmaktadır ve bu nedenle hesaplama açısından verimli algoritmalar geliştirilmelidir. Üçüncüsü, GWR'nin dengeleme katsayılarının uygun değerlerini bulmak kolay değildir ve bu nedenle bu katsayıların değerlerini bulmak için yeni stratejiler geliştirilmelidir.

Tezin bu bölümünde, veri kümelerinin mekânsal, rakımsal ve zamansal değişiklikleri ele almak için dört boyutlu GWR (4D-GWR) yaklaşımını önerilmiştir. Önerilen yaklaşım, mekânsal, zamansal ve rakımsal olan yeni bir bant genişliği stratejisi kullanmaktadır. Önerilen 4D-GWR'nin performansı, ölçülen meteorolojik parametrelerin mekân (boylam ve enlem), rakım ve zaman bilgileri içeren gerçek zamanlı meteorolojik veri kümesinde klasik GWR (2D-GWR), GAWR ve GTWR yaklaşımlarının performansı ile karşılaştırılır. Deneysel sonuçlar, 4D-GWR'nin hava sıcaklığı değerlerini tahmin etmek için doğruluk ve hesaplama maliyeti açısından diğer yaklaşımlardan daha iyi performans elde ettiğini göstermektedir.

Önerilen yöntemin ana katkıları aşağıdaki gibidir.

- Veri kümelerinin mekânsal, rakımsal ve zamansal değişiklikleri dikkate alan yeni bir dört boyutlu GWR yaklaşımı (4D-GWR) önerilmektedir.
- 4D-GWR parametrelerini tahmin etmek için, n-boyutlu altın kesit arama (n-dimension golden section search - nD-GSS) algoritmasını kullanan yeni bir parametre tahmin stratejisi önerilmektedir.
- Önerilen 4D-GWR'nin performansı, mekân (boylam ve enlem), rakım ve zaman bilgileri içeren gerçek zamanlı meteorolojik veri kümesinde 2D-GWR, GAWR [32] ve GTWR [30] yaklaşımlarının performansı ile karşılaştırılmıştır.
- Deneysel sonuçlar, 4D-GWR'nin hava sıcaklığı değerlerini tahmin etmek için doğruluk ve hız açısından diğer yaklaşımlardan daha iyi performans elde ettiğini göstermektedir.

Bu bölümün geri kalanı şu şekilde organize edilmiştir. Bölüm 4.2'de önerilen 4D-GWR algoritması hakkında bilgi verilmiştir. Bölüm 4.3'te deneysel çalışmalar ve Bölüm 4.4'te sonuçlar sunulmuştur.

4.2. 4D-GWR: Önerilen Coğrafi, Rakımsal ve Zamansal Ağırlıklı Regresyon Yaklaşımı

Bu çalışmada, mekânsal, rakımsal ve zamansal değişiklikleri yakalamak için yeni bir dört boyutlu coğrafi ağırlıklı regresyon (4D-GWR) yaklaşımı önerilmektedir. Önerilen 4D-GWR yaklaşımı hesaplama maliyetini azaltırken aynı zamanda tahmin doğruluğunu da artırmaktadır. Hesaplama maliyetini azaltmak için FastGWR'de [26] olduğu gibi matris işlemlerinde noktasal çarpım işlemi ve Bölüm 2'de bahsedilen RNN-GWR'de olduğu gibi simetrik matris işlemleri kullanılmıştır.

Önerilen 4D-GWR yaklaşımı klasik GWR (2D-GWR)'nin genişletilmiş bir versiyonudur ve mekânsal, rakımsal ve zamansal özellikleri aynı anda yakalamak için GAWR ve GTWR yaklaşımlarının birleşimi olarak tanımlanabilir ve Eşitlik 4.4 ile ifade edilir. Önerilen 4D-GWR yaklaşımının detayları verilmeden önce GTWR ve GAWR yaklaşımları

incelenmiştir. GTWR yaklaşımı Bölüm 3.2’de anlatılmıştır ve GAWR yaklaşımı ise aşağıda bahsedilmiştir. Daha sonra önerilen 4D-GWR yaklaşımının detayları verilmiştir.

4.2.1. GAWR Yaklaşımı

Klasik GWR yaklaşımı mekânsal ilişkileri bulmada çok iyidir ve iki boyutlu uzayda çalışır. Fakat yükselteli arazi koşullarının olduğu veya yükseklik verisinin önemli bir etken olduğu durumlarda iki boyutlu uzayda inceleme yetersiz olabilmektedir ve üçüncü boyuta ihtiyaç duyulmaktadır. Propastin [32], bu sorunun üstesinden gelmek için üçüncü boyutu (rakım veya yükseklik) da dikkate alarak coğrafi ve rakım ağırlıklı regresyon (geographically and altitudinal weighted regression - GAWR) yaklaşımını önermiştir. GAWR yaklaşımı, klasik GWR yönteminin ağırlık matrisine üçüncü boyut olarak rakımı ekleyerek oluşturulmuştur. GAWR yaklaşımı Eşitlik 4.1’de gösterildiği şekilde ifade edilebilir.

$$y_i(u_i, v_i, z_i) = \beta_{i0}(u_i, v_i, z_i) + \sum_{k=1}^m (\beta_{ik}(u_i, v_i, z_i)x_{ik}(u_i, v_i, z_i) + \epsilon_i(u_i, v_i, z_i)) \quad (4.1)$$

GWR formülünden farklı olarak formüle rakım (yükseklik) bilgisi eklenmiştir ve z yükseklik bilgisini göstermektedir. $\beta_{ik}(k = 0, \dots, m)$ regresyon katsayısı GWR yaklaşımında olduğu gibi Eşitlik 4.2’de gösterildiği şekilde hesaplanır.

$$\beta(u_i, v_i, z_i) = (X^T W(u_i, v_i, z_i) X)^{-1} X^T W(u_i, v_i, z_i) Y \quad (4.2)$$

$W(i) = W(u_i, v_i, z_i)$ ağırlık matrisi, klasik GWR olduğu gibi w_{ij} değerlerinden oluşur. Fakat burada w_{ij} komşuluk oranı 2 boyutlu uzayda değil 3 boyutlu uzayda ölçülüdür. u, v nokta koordinatlarını, z yükseklik (rakım) bilgisini ve c ise dengeleme katsayısını göstermek üzere d_{ij} mesafesi Eşitlik 4.3’te gösterildiği şekilde hesaplanır:

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2 + c(z_i - z_j)^2} \quad (4.3)$$

GAWR yaklaşımında komşuluk ağırlıklarının hesaplanarak ağırlık matrisleri

oluşturulmadan önce mekânsal ve yükseklik (rakım) mesafesinin hesaplamada uyumlu hale getirilmesi gerekmektedir. Farklı mekânsal ve rakım birimleri dengelemek veya uyumlu hale getirmek için dengeleme parametresi c eklenmiştir [32]. Bu nedenle, GAWR yaklaşımını uygulamadan önce c dengeleme parametresinin optimize edilmesi gerekmektedir. Propastin [32], c parametresini bütün ihtimalleri deneyerek bulmayı önermiştir.

4.2.2. Önerilen 4D-GWR Yaklaşımı

Önerilen 4D-GWR yaklaşımı klasik GWR (2D-GWR), GAWR ve GTWR yaklaşımlarından geliştirilmiş bir GWR yaklaşımıdır. Önerilen 4D-GWR yaklaşımının regresyon formülasyonu Eşitlik 4.4 ile ifade edilir.

$$y_i(u_i, v_i, t_i, z_i) = \beta_{i0}(u_i, v_i, t_i, z_i) + \sum_{k=1}^m (\beta_{ik}(u_i, v_i, t_i, z_i) x_{ik}(u_i, v_i, t_i, z_i) + \epsilon_i(u_i, v_i, t_i, z_i)) \quad (4.4)$$

burada u_i ve v_i , i noktasının enlemi ve boylamıdır, z_i , i noktasının rakımıdır ve t_i , i noktasında alınan ölçümün zaman bilgisidir.

4D-GWR yaklaşımında komşuluk mesafesi, sırasıyla, Eşitlik 1.8, 4.3 ve 3.3'te verilen klasik GWR (2D-GWR), GAWR ve GTWR mesafelerinin birleşmiş halini kullanmaktadır. 4D-GWR yaklaşımında mekânsal, rakımsal ve zamansal terimleri içeren, Eşitlik 4.5'te gösterildiği gibi ifade edilen ve yeniden yazılmış d_{ij} mesafe metriğini kullanarak bir ağırlık matrisi oluşturur. Veri kümelerindeki enlem, boylam, rakım ve zaman bilgileri vurgulamak için a , b , c ve d dengeleme katsayıları (exaggeration coefficients) kullanılır.

$$d_{ij} = \sqrt{a(u_i - u_j)^2 + b(v_i - v_j)^2 + c(z_i - z_j)^2 + d(t_i - t_j)^2} \quad (4.5)$$

GWR yaklaşımlarında optimum bant genişliği değerlerini bulmak zor bir süreçtir. 2D-GWR yaklaşımında altın kesit arama (golden section search - GSS) genellikle optimum bant genişliğini verimli bir şekilde aramak için kullanılır [15, 26, 39, 47].

GAWR ve GTWR’de optimum bant genişliğinin değerini belirlemek için tüm olası çapraz doğrulama (cross-validation - CV) değerleri hesaplanır. Bu süreç hesaplama açısından karmaşıktır.

Ayrıca, 4D-GWR yaklaşımı için optimum bant genişliği bw değerini ve komşuluk mesafesi d_{ij} değerini oluşturan dengeleme katsayıları a , b , c ve d değerlerini belirlemek için yeni bir strateji önerilmektedir. Önerilen strateji, dengeleme katsayılarının değerlerini belirlemek için n-boyutlu altın kesit arama (n-dimension golden section search - nD-GSS) algoritmasını kullanmaktadır ve daha sonra bu dengeleme katsayıları kullanılarak optimum bant genişliğini değeri bulunmaktadır. Önerilen strateji, bütün CV değerlerini hesaplamadığından çalışma süresi açısından daha etkilidir. nD-GSS algoritmasının ayrıntıları aşağıdaki bölümde verilmiştir.

4.2.3. Dengeleme Katsayılarının ve Optimum Bant Genişliğinin Bulunması

Bu çalışmada, dengeleme katsayıları a , b , c ve d ve bant genişliği bw değerlerini bulmak için hem nD-GSS hem de 1D-GSS algoritmaları kullanılmaktadır. GSS, bir fonksiyonun minimum veya maksimumunu bulmak için yaygın olarak kullanılan bir yöntemdir [102–107]. Bu stratejinin kaba kodu Algoritma 4.1’de verilmiştir.

Algoritma parametrelerin alt ve üst limit dizilerini giriş parametresi olarak alır, dengeleme katsayıları ve optimum bant genişliği değerini de çıktı olarak verir. Alt ve üst sınırlar rastgele seçilir ve altın kesit arama adımıyla kullanılır. Algoritmanın temel olarak üç aşaması vardır. Bunlar, klasik GSS kullanarak ilk bant genişliğinin bulunması (Adım 1), Algoritma 4.2’de gösterilen 4D-GSS fonksiyonunu kullanan dengeleme katsayılarının (Adım 2) değerlerinin tahmini ve optimum bant genişliğinin bulunmasıdır (Adım 3). Her aşamanın ayrıntıları aşağıda verilmiştir.

İlk bant genişliğini bulma: a , b , c ve d dengeleme katsayıları bant genişliği (bw) değerini etkileyen parametrelerdir ve bu nedenle bu beş parametrenin değerlerini klasik GSS algoritmaları ile aynı anda bulmak mümkün değildir [102]. Klasik GSS yöntemi, $[Lb, Ub]$ aralığında bir tek modlu fonksiyonun minimum veya maksimum değerini bulmak için kullanılır. Lb parametre için bir alt sınırı ve Ub parametre için bir üst sınırı göstermektedir [108]. Optimizasyon işlemi, alt sınır ve üst sınır, optimal değere

Algoritma 4.1: 4D-GWR Algoritması**Girdi** : Lb : alt sınır dizisi, Ub : üst sınır dizisi**Çıktı** : dengeleme katsayıları (a, b, c, d) ve bant genişliği (bw) **Değişkenler:** r : altın oran (0.618), t : durdurma kriteri (0.0001)

- 1 Klasik GSS algoritması vasıtasıyla Eşitlik 1.12 ve Eşitlik 1.8 kullanarak 2D-GWR yaklaşımında optimum bant genişliğinin (bw) bulunması;
- 2 $(a, b, c, d) = 4D-GSS(Lb, Ub, bw, r, t)$;
- 3 Klasik GSS algoritması vasıtasıyla Eşitlik 1.12, ve Eşitlik 4.5 ve dengeleme katsayıları a, b, c, d kullanarak 4D-GWR yaklaşımı için optimum bant genişliğinin (bw) bulunması;

Algoritma 4.2: 4D-GSS**Çıktı** : a, b, c, d **Girdi** : Lb, Ub, bw, r, t **Değişkenler:** *Indeks*: Lb ve Ub dizilerinin indeksi

```

1 while  $abs(Ub_s - Lb_s) > t$  do
2    $a_1 = Lb[0] + (1 - r)(Ub[0] - Lb[0]), a_2 = Lb[0] + r(Ub[0] - Lb[0]);$ 
3    $b_1 = Lb[1] + (1 - r)(Ub[1] - Lb[1]), b_2 = Lb[1] + r(Ub[1] - Lb[1]);$ 
4    $c_1 = Lb[2] + (1 - r)(Ub[2] - Lb[2]), c_2 = Lb[2] + r(Ub[2] - Lb[2]);$ 
5    $d_1 = Lb[3] + (1 - r)(Ub[3] - Lb[3]), d_2 = Lb[3] + r(Ub[3] - Lb[3]);$ 
6    $f_1 = F(a_1, b_1, c_1, d_1, bw), f_2 = F(a_1, b_1, c_2, d_1, bw);$ 
7    $f_3 = F(a_1, b_2, c_1, d_1, bw), f_4 = F(a_1, b_2, c_2, d_1, bw);$ 
8    $f_5 = F(a_2, b_1, c_1, d_1, bw), f_6 = F(a_2, b_1, c_2, d_1, bw);$ 
9    $f_7 = F(a_2, b_2, c_1, d_1, bw), f_8 = F(a_2, b_2, c_2, d_1, bw);$ 
10   $f_9 = F(a_1, b_1, c_1, d_2, bw), f_{10} = F(a_1, b_1, c_2, d_2, bw);$ 
11   $f_{11} = F(a_1, b_2, c_1, d_2, bw), f_{12} = F(a_1, b_2, c_2, d_2, bw);$ 
12   $f_{13} = F(a_2, b_1, c_1, d_2, bw), f_{14} = F(a_2, b_1, c_2, d_2, bw);$ 
13   $f_{15} = F(a_2, b_2, c_1, d_2, bw), f_{16} = F(a_2, b_2, c_2, d_2, bw);$ 
14  On altı  $f$  arasından minumum  $f$  değerinin ve yeniden düzenlenecek Indeks
    indeksinin bulunması;
15   $f$  değerine göre  $Lb[Indeks]$  veya  $Ub[Indeks]$  değerinin değiştirilmesi;
16  $a = \frac{Lb[0] + Ub[0]}{2}, b = \frac{Lb[1] + Ub[1]}{2}, c = \frac{Lb[2] + Ub[2]}{2}, d = \frac{Lb[3] + Ub[3]}{2};$ 
17 return  $a, b, c, d$ ;

```

yeterince yaklaşıp kadar sınırların tekrar tekrar birbirine doğru hareket ettirilmesiyle gerçekleştirilir. Bu nedenle, veri kümesinin mekânsal özelliklerini dikkate alarak başlangıç bant genişliği değerinin tahmin edilmesi amaçlanmıştır ve Eşitlik 1.8’de verilen mesafe tanımı kullanılmıştır. Çapraz doğrulama (cross-validation - CV), bant genişliği değerlerini tahmin etmek için kullanılır ve klasik GSS, bant genişliği arama uzayını daraltmak için kullanılır. Bu ilk bant genişliği değeri, algoritmanın (Algoritma 4.1) bir sonraki aşamasında, 4D-GWR’nin dengeleme katsayılarının değerlerini tahmin etmek

Algoritma 4.3: F Uygunluk Fonksiyonu**Girdi:** a, b, c, d, bw

- 1 Komşuluk mesafesi olarak Eşitlik 4.5 kullanılarak Eşitlik 1.12 değerinin hesaplanması;
- 2 **return** Hesaplanan değer;

için kullanılır.

4D-GWR yaklaşımı için dengeleme katsayı değerlerinin tahmini: Bu aşamada a , b , c ve d parametrelerinin değerlerini tahmin etmek için nD-GSS kullanılmaktadır. Başlangıç bant genişliği değeri ile dengeleme katsayılarının alt ve üst sınırları nD-GSS algoritmasına girdi olarak verilir.

Algoritma 4.2’de, 1 ile 15 arasındaki adımlar, alt ve üst sınırların mutlak farkı t ’den büyük olduğu sürece yinelenir. Her yinelemede, bir parametrenin alt sınırı veya üst sınırı değişir. While döngüsü tamamlandıktan sonra Adım 16’da üst ve alt sınırların ortalamaları hesaplanır ve dengeleme katsayılarının değerleri olarak atanır. Algoritmada F uygunluk fonksiyonudur ve detayları Algoritma 4.3’de verilmiştir.

Model için optimum bant genişliği bulma: Bu aşamada (Adım 3), dengeleme katsayıları değerleri bulunduğundan sonra, bu değerler kullanılarak optimum bant genişliği bw değeri bulunur. Bu adım, bant genişliğini tahmin etmek için tek boyutlu klasik GSS kullanır. Optimum bant genişliği bw Eşitlik 1.12’de verilen çapraz doğrulama (cross-validation - CV), Eşitlik 4.5’de verilen mesafe ve bulunan dengeleme katsayıları (a, b, c, d) kullanılarak hesaplanır.

4.2.4. GWR Yaklaşımlarında Hesaplama Maliyeti

GWR yaklaşımları yüksek bilgi işlem kaynakları ve geniş bellek alanları gerektirir. Veri kümesi boyutu büyüdükçe bu sorunlar kritik hale gelir.

GWR modellerinin zaman karmaşıklıkları çoğunlukla Eşitlik 1.3’de bulunan β_i ’nin hesaplanmasına bağlıdır. GWR algoritmasında Eşitlik 1.3’in en çok zaman alan kısmı $X^T W(i) X$ matrislerinin çarpımıdır, n ölçüm sayısı olmak üzere ve m parametre sayısı olmak üzere hesaplama maliyeti $O(m^2 n)$ olarak ifade edilir [26]. β_i hesaplaması ve

değişkenleri karşılaştırılan tüm yaklaşımlar için aynı olduğundan, n ölçüm noktasına göre 2D-GWR, GTWR ve GAWR yaklaşımlarını için β_i 'lerin toplam hesaplama maliyeti $O(m^2n^2)$ olarak ifade edilir.

GWR'nin bellek karmaşıklığını azaltmak için Eşitlik 1.3'de bulunan β_i Li vd. [26] ve bu tez kapsamında geliştirilen RNN-GWR [31] ile yeniden düzenlenmiştir. Eşitlik 1.3'in en çok zaman alan $X^TW(i)X$ matrisi simetrik matristir [31]. Bu matrisin hesaplama maliyeti Li vd. [26] ve RNN-GWR'de (Bölüm 2) olduğu gibi, matris işlemlerinde noktasal çarpım kullanılarak ve simetrik olduğu için matrisin köşegen (veya diagonal) öğelerinin alt veya üst kısmının hesaplanmamasıyla azaltılır. Bu tekniklerin kullanılmasıyla 4D-GWR'nin hesaplama maliyeti $O\left(\left(\frac{m^2+m}{2}\right)n^2\right)$ haline gelir.

Bu çalışmada, optimum bant genişliğini bir kez bulunmuştur ve algoritmalara girdi olarak verilmiştir. En uygun bant genişliğini bulmak için GSS yaklaşımı kullanıldığında, zaman karmaşıklığı $O(\log n)$ [26, 85] olarak kabul edilir. Bu durumda, 2D-GWR'nin hesaplama karmaşıklığı $O(m^2n^2 \log n)$ olmaktadır. GAWR ve GTWR yaklaşımları GSS algoritmasını kullanmaz ve olası tüm çapraz doğrulama (CV) kombinasyonları değerlendirir. Bütün çapraz doğrulama (cross-validation - CV) işleminin maliyeti n^2 'dir ve böylece GAWR ve GTWR yaklaşımlarının hesaplama maliyet $O(m^2n^4)$ olur. 4D-GWR'de, 1D-GSS kullanarak ilk bant genişliğini bulma maliyeti $O(\log n)$, 4D-GSS kullanarak normalleştirme katsayılarının değerlerini bulma maliyeti $O(4 \log n)$ ve 1D-GSS kullanarak optimum bant genişliği bw değerini bulma maliyeti $O(\log n)$ olur. Bu nedenle 4D-GWR'de normalleştirme katsayıları ve optimum bant genişliği sürecinin değerlerini bulma maliyeti $O(6 \log n)$ haline gelir, böylece 4D-GWR'nin zaman karmaşıklığı $O\left(6\left(\frac{m^2+m}{2}\right)n^2 \log n\right)$ haline gelir.

4.3. Deneysel Değerlendirmeler

Bu bölümde klasik GWR (2D-GWR), GAWR [32], GTWR [28–30] ve önerilen 4D-GWR algoritmalarının başarısını göstermek için yapılan deneysel değerlendirmelerden bahsedilmektedir. Bölüm 1.3.1'de bahsedilen MGM'den alınan gerçek ölçüm verileri ve gerçek ölçüm verilerinden elde edilen sentetik veri kümesi kullanılmıştır. Veri kümesi ve model ayarları ve deneysel değerlendirme ile ilgili detaylı bilgiler sırasıyla Bölüm 4.3.1

ve 2.5.2’de verilmiştir.

Deneysel değerlendirmeler, gerçek veri kümeleri kullanılarak GWR modellerinin doğruluğunu ve hızını ve sentetik veri kümeleri kullanılarak veri kümesi boyutunun etkisini değerlendirmek için yapılmıştır. Deneysel değerlendirmeler aşağıdaki sorulara cevap vermeye odaklanmıştır.

- Zaman aralığı sayısının algoritma performanslarına etkisini nedir?
- Yaklaşımların çalışma zamanları ne kadardır?
- Veri boyutunun algoritma performanslarına etkisi nedir?
- Tahmin edilen değer ile gerçek değer arasındaki fark nedir?

Algoritmalar Intel Core i5-3210M 2.5 GHz 8GB RAM özelliklerine sahip bir bilgisayar üzerinde çalıştırılmıştır. Yaklaşımlar Matlab’da çalıştırılmıştır.

4.3.1. Veri Kümesi ve Model Ayarları

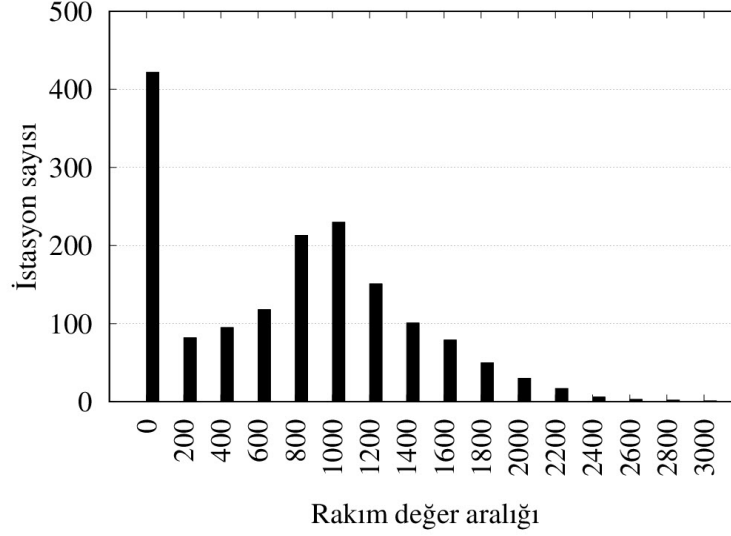
Bu bölümde, ilk önce deneysel değerlendirmelerde kullanılan veri kümesi hakkında bilgi verilmiştir. Daha sonra yaklaşımlar için kullanılan model ayarları sunulmuştur.

4.3.1.1. Veri Kümesi

Bu çalışmada Bölüm 1.3.1’de bahsedilen MGM’den alınan gerçek ölçüm verileri kullanılmıştır [31]. Veri kümesi 1600 adet istasyona ait mekân (enlem ve boylam), rakım ve saatlik olarak ölçülmüş sıcaklık, basınç ve nem bilgilerini içermektedir. Bu çalışmada kullanılan meteoroloji ölçüm istasyonlarının konumları Şekil 1.5’de gösterilmiştir. Veri kümesinde, 30 gün boyunca toplanan 1.152.000 saatlik ölçüm verisi bulunmaktadır. Şekil 4.1’de gösterildiği üzere konumların rakım değeri çoğunlukla 200 metreden küçüktür veya 600 metre ile 1200 metre arasındadır. Deneylerde 1200 meteoroloji istasyonu referans noktası, 400 meteoroloji istasyonu regresyon noktası olarak kullanılmıştır.

Bu çalışmada, önerilen 4D-GWR algoritmasının performansını değerlendirmek için

meteorolojik veri kümesi kullanılmaktadır. Ancak, mekânsal, rakımsal ve zamansal bilgi içeren diğer veri kümeleri de önerilen 4D-GWR algoritması tarafından kullanılabilir.



Şekil 4.1. Rakım değerlerinin yoğunluğu

4.3.1.2. Model Ayarları

Deneysel çalışmalarda amaç regresyon noktasının basınç, nem ve rakım verilerini kullanarak o noktanın hava sıcaklık değerini tahmin etmektir. 2D-GWR ve GTWR yaklaşımları sıcaklık tahminini Eşitlik 2.4 ile yapmaktadır. GAWR ve önerilen 4D-GWR yaklaşımları sıcaklık tahminini Eşitlik 4.6 ile yapmaktadır.

$$Sicaklik_i = \beta_{0i} + \beta_{1i}Nem_i + \beta_{2i}Basinci_i \quad (4.6)$$

Yaklaşımlar tarafından tahmin edilen sıcaklık değeri ile gerçek sıcaklık değerlerinin karşılaştırması kök ortalama karesel hata (root mean square error-RMSE) değerlerine göre yapılmıştır. RMSE (Eşitlik 4.7) model tarafından tahmin edilen değer ($\hat{y}_i$) ile gerçekte ölçülen değer (y_i) arasındaki farkı ifade eden ve önerilen yaklaşımın kalitesini gösteren bir metriktir [31, 109, 110]. RMSE tahmin hatalarının (kalıntıların) standart sapmasıdır. Yani, kalıntılar, regresyon hattının veri noktalarından ne kadar uzakta olduğunun bir ölçüsüdür. RMSE değeri sıfıra ne kadar yakınsa önerilen yaklaşım o kadar iyidir.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4.7)$$

Bir GWR modelinde, bant genişliği parametresinin değeri ve çekirdek fonksiyonu belirlenmelidir. Bu çalışmada en yaygın kullanılan çekirdek fonksiyonlarından biri olan gaussian fonksiyonu kullanılmıştır [34, 35]. GAWR [32] ve GTWR [30] yaklaşımlarında gaussian çekirdek fonksiyonunu kullanmaktadır.

2D-GWR, GAWR, GWTR ve önerilen 4D-GWR yaklaşımlarının bant genişliği ve dengeleme katsayılarının değerini belirlemek için çapraz doğrulama (cross-validation - CV) yöntemi kullanılmıştır. 4D-GWR yaklaşımında nD-GSS algoritması kullanılmaktadır. Böylece çalıştırılan çapraz doğrulama sayısı azalmaktadır ve dolayısı ile 4D-GWR'nin hesaplama maliyeti azalmaktadır. Deneysel çalışmanın başlangıcında, algoritmalar için, dengeleme katsayıları ve bant genişliği değerleri bulunmuştur ve ölçüm noktalarının konumları sabit olduğu için bu değerler değiştirilmemiş ve tekrar hesaplamaya ihtiyaç duyulmamıştır. Algoritmaların bant genişliği değerleri ve dengeleme katsayıları Tablo 4.1'de verilmiştir. Tabloda *bw* değeri kilometre olarak ifade edilmiştir.

Tablo 4.1. GWR yaklaşımları için bant genişliği değerleri ve dengeleme katsayıları

Yaklaşım	<i>bw</i> (km)	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
2D-GWR	82	1	1	n/a	n/a
GAWR	70	1	1	6.5	n/a
GTWR	80	1	1	n/a	15.01
4D-GWR	74	1.36	0.70	6.13	14.07

2D-GWR'de bant genişliği bulma maliyeti $O(\log n)$ olarak ifade edilmektedir [26, 85]. Önerilen 4D-GWR, Algoritma 4.1'de açıklandığı gibi n-boyutlu altın kesit arama algoritmasını (n-dimension golden section search-nD-GSS) kullanır. Bu nedenle, 4D-GWR için bant genişliği ve dengeleme katsayılarını bulma maliyeti $O(6 \log n)$ olmaktadır. GAWR ve GTWR yaklaşımları için bant genişliği ve dengeleme katsayılarını bulma maliyetleri, bu parametrelerin değerlerini bulmak için bütün olası ihtimalleri denemelerinden dolayı $O(n^2)$ olmaktadır.

4.3.2. Deneyler

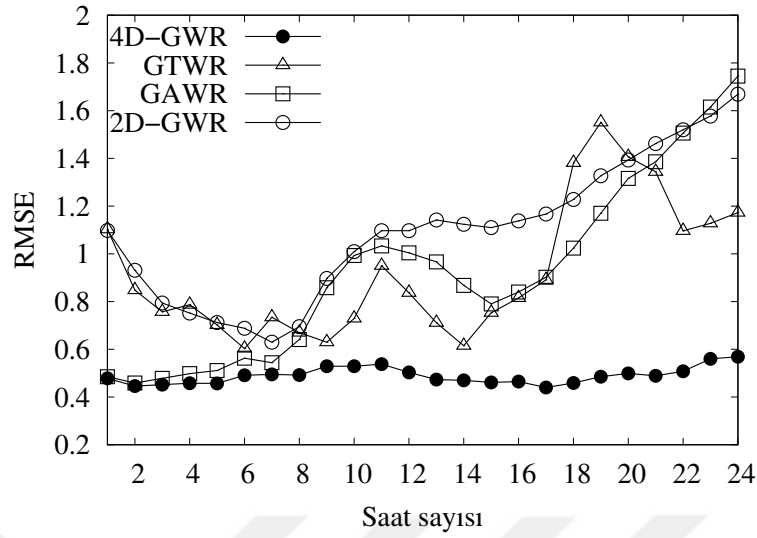
4.3.2.1. Zaman Aralıklarının Etkisi

Bu deney çalışmasında zaman aralıkları sayısının (saat ve günlerin) 2D-GWR, GAWR, GTWR ve önerilen 4D-GWR yaklaşımları üzerindeki etkisini doğruluk açısından gerçek saatlik meteorolojik veri kümesi kullanarak değerlendirilmiştir. Bu kapsamda iki deney yapılmıştır. İlk deneyde, saat sayısının etkisini değerlendirilmiş ve bu yüzden saat bazlı artan veri kümesi kullanılmıştır (Şekil 4.2a). İkinci deneyde, gün sayısının etkisi değerlendirilmiş ve bu nedenle gün bazlı artan veri kümesi kullanılmıştır (Şekil 4.2b). Deneylerde, zaman aralığı sayısı arttıkça algoritmaların kök ortalama karesel hata (root mean square error - RMSE) değerleri incelenmiştir.

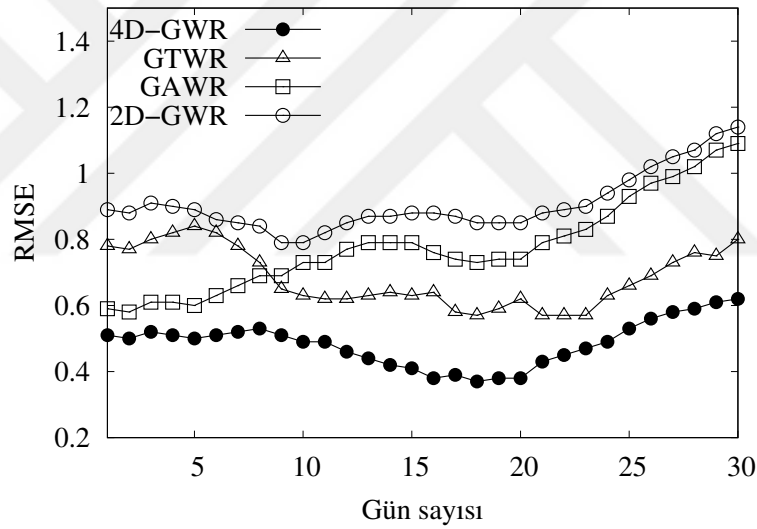
İlk deneyde kullanılan veri kümesi 12 Aralık 2016 gününe ait olup saatlik meteorolojik verileri içermektedir (Şekil 4.2a). Minimum saat sayısı 1 ve maksimum saat sayısı 24'tür. İlk zaman aralığı saat 0 verilerini kapsar ve bu nedenle ölçülen parametreler arasında zaman farkı yoktur. Yani, ilk zaman diliminde zaman bilgisinin tahmin edilen değerler üzerinde bir etkisi yoktur. Bu durumda, 2D-GWR ve GTWR yaklaşımlarının RMSE değerleri, her ikisi de veri kümesinin mekânsal bilgilerini (Eşitlik 2.4) dikkate aldığı için sonuçları birbirine yakındır. Diğer bir taraftan, Şekil 4.2a'da gösterildiği üzere zaman bilgisinin ilk zaman diliminde bir etkisi olmadığından, mekânsal ve rakımsal bilgiler (Eşitlik 4.6) kullanıldığından, GAWR ve 4D-GWR yaklaşımlarının RMSE değerleri birbirine yakın çıkmaktadır. Saat sayısı arttıkça, zaman bilgisi GTWR ve önerilen 4D-GWR algoritmasının tahmin sonuçlarını etkilemektedir.

İkinci deneyde, kullanılan veri kümesi Aralık 2016'ya aittir ve saatlik meteorolojik verileri içermektedir (Şekil 4.2b). Minimum gün sayısı 1, maksimum gün sayısı 30'dur. Her gün 24 saatlik meteorolojik verileri içerir. Bu deneyde gün sayısı birer birer artmaktadır.

Sonuçlar, önerilen 4D-GWR'nin, zaman aralığı sayısı arttıkça diğer yaklaşımlardan daha iyi performans gösterdiğini göstermektedir (Şekil 4.2a ve Şekil 4.2b). Zaman aralıklarının sayısı arttıkça, diğer yöntemlerde olan dalgalanmanın aksine önerilen 4D-GWR'nin RMSE değeri neredeyse sabittir. Bu sonuç önerilen 4D-GWR yaklaşımının diğer



(a) Saat sayısının etkisi



(b) Gün sayısının etkisi

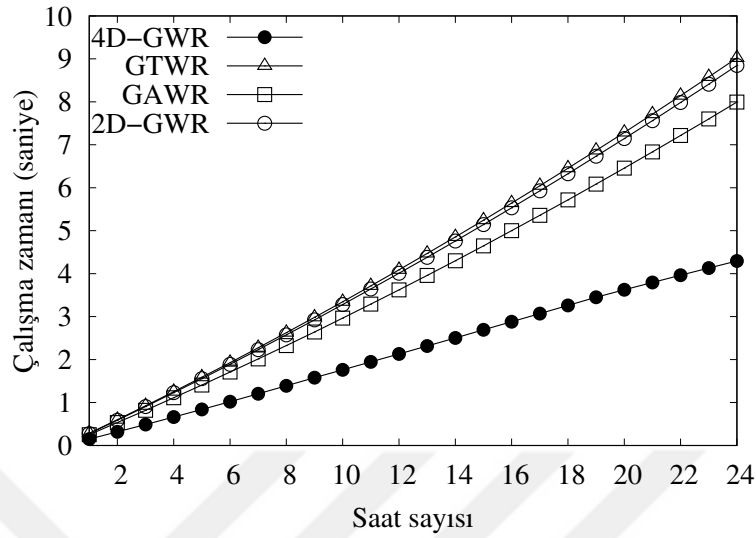
Şekil 4.2. Zaman aralığı sayısının etkisi

yaklaşımlardan daha gürbüz olduğunu göstermektedir. Deneysel sonuçlar, veri kümesinde bulunan mekânsal, rakımsal ve zamansal bilgilerinin birlikte dikkate alınmasının GWR modelinin tahmin doğruluğunu arttırdığını göstermektedir.

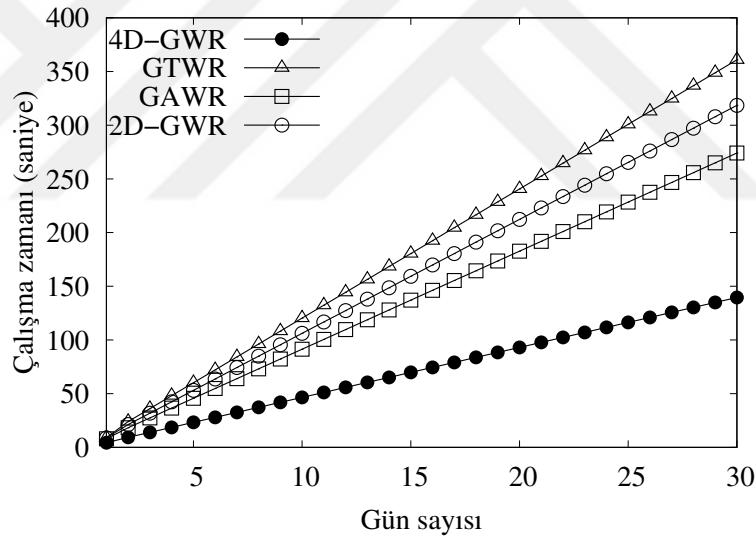
4.3.2.2. Yaklaşımların Çalışma Zamanları

Bu bölümde, önceki bölümde yapılan deneyler kullanılarak 2D-GWR, GAWR, GTWR ve önerilen 4D-GWR yaklaşımlarının çalışma zamanları değerlendirilmiştir. Şekil 4.3a, saat sayısı arttıkça her algoritmanın çalışma zamanlarını göstermektedir ve Şekil 4.3b ise

gün sayısı arttıkça her algoritmanın çalışma zamanlarını göstermektedir.



(a) Saat sayısı arttıkça çalışma zamanı karşılaştırmaları



(b) Gün sayısı arttıkça çalışma zamanı karşılaştırmaları

Şekil 4.3. Yaklaşımların çalışma zamanı karşılaştırmaları

Şekil 4.3a'da, GTWR ve 2D-GWR yaklaşımlarının hesaplama maliyeti Bölüm 4.2.4'de anlatıldığı üzere $\beta(i)$ hesaplaması için aynı formülasyonu ve aynı m değerini kullandıklarından dolayı birbirine çok yakın çıkmaktadır. GAWR yaklaşımı, $\beta(i)$ 'ın hesaplanması için aynı formülasyonu kullanmasına rağmen Eşitlik 4.6 bulunan m değeri daha küçük olduğundan hesaplama maliyeti GTWR ve 2D-GWR yaklaşımlarından daha düşüktür. Önerilen 4D-GWR algoritması, aşağıdaki nedenlerden dolayı 2D-GWR, GAWR ve GTWR algoritmalarından daha iyi performans göstermektedir.

İlk olarak 4D-GWR FastGWR [26] ve RNN-GWR (Bölüm 2) çalışmalarında olduğu gibi matris işlemlerinde noktasal çarpım işlemlerini kullanarak $\beta(i)$ 'ın yeniden ifade edilmiş sürümünü kullanmaktadır. İkisi ise bu tez çalışması kapsamında Eşitlik 1.3'in en çok zaman alan kısmının $(X^T W(i) X)$ simetrik bir matris olduğunu anlaşılmıştır [31]. 4D-GWR $X^T W(i) X$ matrisinin tüm değerlerini hesaplamamaktadır ve bu simetrik matris özelliğinden dolayı hesaplama maliyeti az olmaktadır. Üçüncüsü, GAWR yaklaşımında olduğu gibi 4D-GWR yaklaşımında kullanılan m değeri (Eşitlik 4.6) 2D-GWR ve GTWR yaklaşımlarında kullanılan m değerinden (Eşitlik 2.4) daha küçüktür. Bu nedenle $\beta(i)$ regresyon katsayılarının hesaplanması 4D-GWR yaklaşımında hesaplama açısından verimli hale gelmektedir.

Şekil 4.3a'da gösterilen deneyde, regresyon noktası 400'e sabitlenirken, ölçüm verileri her yinelemede 1200 artmaktadır. Şekil 4.3b'de gösterilen deneyde regresyon noktası 400 iken ölçüm verileri her bir yenileme işleminde 28800 (1200×24 (bir gün)) artmaktadır. Bu deney algoritmaların hızını gösterse de veri boyutu büyüdüğünde algoritmaların performansını göstermez ve bu nedenle algoritmaların performansını göstermek için veri boyutu sentetik olarak artırılarak yeni bir deney gerçekleştirilmiştir.

4.3.2.3. Veri Boyutunun Yaklaşımlar Üzerindeki Etkisi

Bu deneyde, veri kümesinin boyutunu artırarak 2D-GWR, GAWR, GTWR ve RNN-GWR algoritmalarının performansları üzerindeki etkisini değerlendirilmiştir. Her zaman dilimi için sentetik veri kümesine 1000 referans ve 1000 regresyon noktası eklenmiştir. Bu deneyde, 1000 referans noktası ve 1000 regresyon noktasından başlayarak her bir zaman aralığında 1000 referans noktası ve 1000 regresyon noktasından oluşan yeni bir veri kümesi oluşturulmuştur. Örneğin, ilk deney bir zaman aralığına sahiptir ve 1000 referans noktası ve 1000 regresyon noktası içerir, üçüncü deneyde 5 zaman aralığı vardır ve her bir zaman aralığı 1000 referans noktası ve 1000 regresyon noktası içerir ve bu şekilde devam etmektedir. Deneylerde, tüm algoritmalar önceden tanımlanmış bant genişliği ve normalleştirme katsayı değerlerini kullanmıştır. Deney sonuçları Tablo 4.2, Şekil 4.4a ve Şekil 4.4b'de verilmiştir.

Veri kümesi boyutları arttıkça, tüm algoritmaların maliyeti artmaktadır. 2D-GWR,

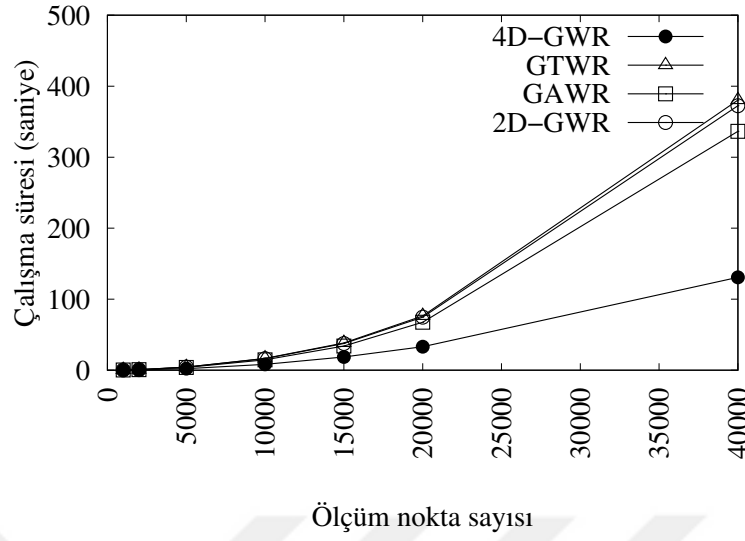
Tablo 4.2. Sentetik veri kümeleri için yaklaşımların çalışma süreleri (saniye)

Ölçüm noktalarının sayısı	Zaman dilimi sayısı	4D-GWR	2D-GWR	GAWR	GTWR
1.000	1	0,148	0,271	0,244	0,276
2.000	2	0,451	0,836	0,755	0,854
5.000	5	2,182	4,273	3,859	4,363
10.000	10	8,167	16,253	14,676	16,594
15.000	15	18,613	37,588	33,942	38,378
20.000	20	33,078	74,781	67,528	76,352
40.000	40	130,844	372,547	336,410	380,370
50.000	50	196,055	n/a	n/a	n/a
100.000	100	820,676	n/a	n/a	n/a
200.000	200	4111,397	n/a	n/a	n/a
500.000	500	26051,733	n/a	n/a	n/a
1.000.000	1000	98436,624	n/a	n/a	n/a

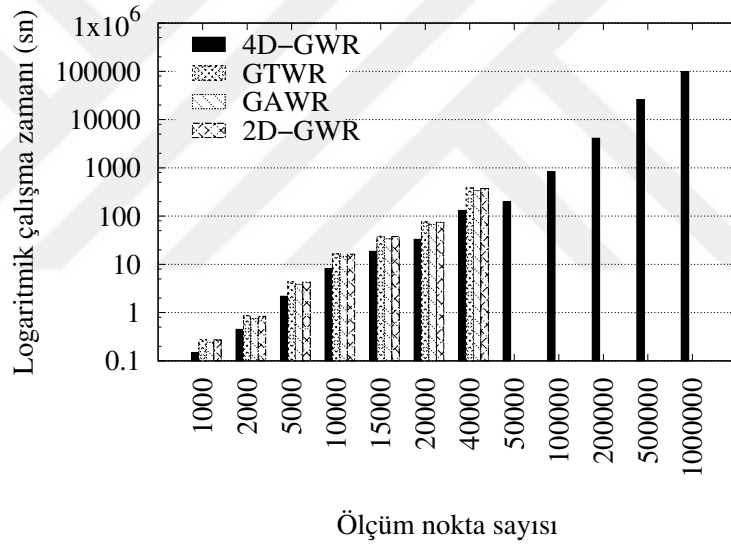
GAWR ve GTWR algoritması, Tablo 4.2, Şekil 4.4a ve Şekil 4.4b’de görülebileceği gibi veri boyutu 40.000 ölçüm noktası olana kadar çalışabilmektedir. Bu nedenle, 40.000 ölçüm noktasından sonra 2D-GWR, GAWR ve GTWR algoritmalarının sonuçları Tablo 4.2 ve Şekil 4.4b’de verilmemiştir. Önerilen 4D-GWR algoritması yeni $\beta(i)$ sürümü [26, 31] kullanılarak bellek karmaşıklığının azaltılmasıyla 1.000.000 ölçüm noktasına sahip veri kümesini işleyebilmektedir ve bunların çalışma zamanları Tablo 4.2 ve Şekil 4.4b’de verilmiştir.

4.3.2.4. Gerçek ve Tahmin Edilen Değerlerin Karşılaştırılması

Bu bölümde, GWR yöntemlerinin tahmini ve gerçek sıcaklık değerleri ve kalıntıları 12 Aralık 2016 tarihinde saat 24 için karşılaştırılmıştır. Bölüm 4.3.2.1’de görülebileceği gibi önerilen 4D-GWR yaklaşımı saat 24’te ve diğer saatlerde RMSE açısından diğer GWR yaklaşımlarından daha iyi performans göstermektedir. Şekil 4.5a, 4.5c, 4.5e ve 4.5g, sırasıyla, 2D-GWR, GAWR, GTWR ve 4D-GWR için 12 Aralık 2016 tarihinde saat 24’te tahmin edilen sıcaklık değerini ve gerçek sıcaklık değerlerini sunar. Gerçek ve tahmin edilen değerler arasında hesaplanan korelasyon katsayısı 2D-GWR, GAWR, GTWR ve 4D-GWR için, sırasıyla, 0.84, 0.82, 0.91 ve 0.98’dir. Sonuçlar 4D-GWR’nin diğer yöntemlerden daha iyi performans elde ettiğini göstermektedir. Önerilen 4D-GWR’nin gerçek ve tahmin edilen değerleri genellikle iyi eşleşir ve köşegen etrafında toplanmaktadır.



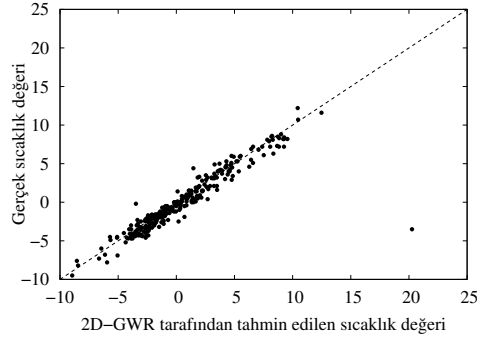
(a) 1.000-40.0000 arası ölçüm noktası



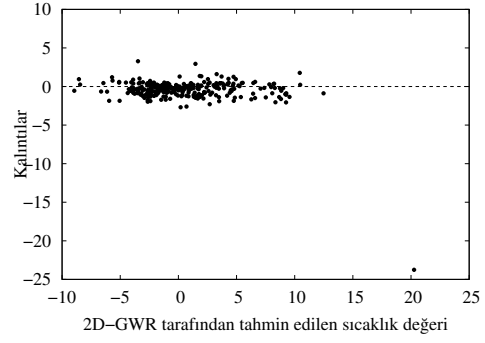
(b) 1.000-1.000.0000 arası ölçüm noktası

Şekil 4.4. Farklı sayıda ölçüm noktası için GWR yaklaşımlarının çalışma zamanları

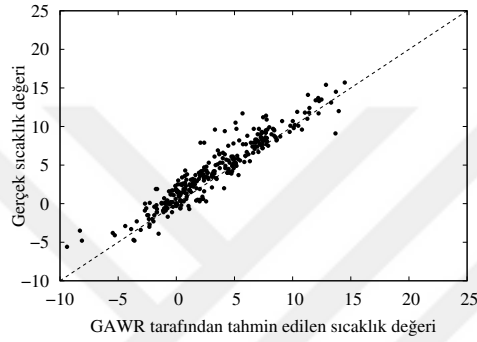
Şekil 4.5b, 4.5d, 4.5f, ve 4.5h, sırasıyla, 2D-GWR, GAWR, GTWR ve 4D-GWR için 12 Aralık 2016 tarihinde saat 24'te tahmin edilen sıcaklık değeri ve gerçek sıcaklık değerleri arasındaki kalıntıları sunar. Kalıntılar, tahmin edilen ve gerçek sıcaklık değerleri arasındaki farkı temsil eder. Önerilen 4D-GWR yaklaşımının kalıntıları diğer yöntemlere göre x eksenine daha yakın olduğu için diğer yaklaşımlardan daha iyi performans göstermektedir.



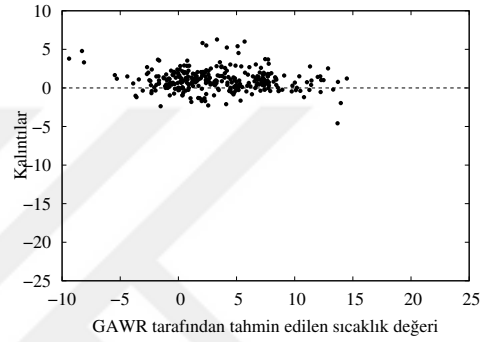
(a) Sıcaklık değerleri



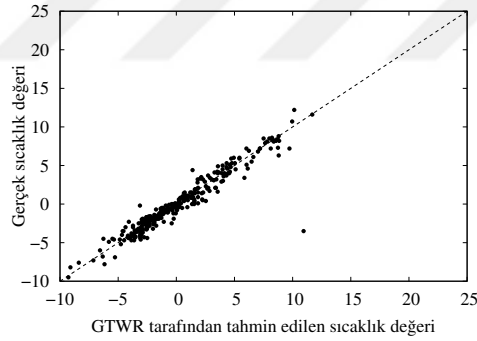
(b) Kalıntılar



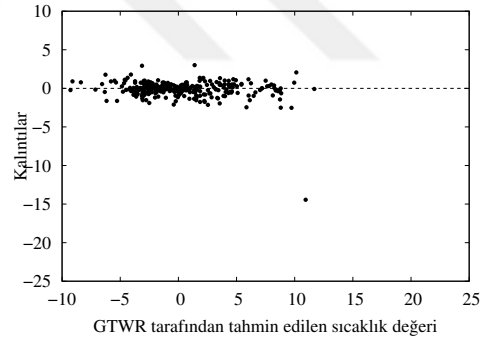
(c) Sıcaklık değerleri



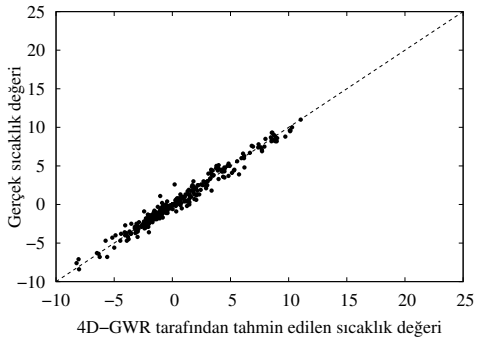
(d) Kalıntılar



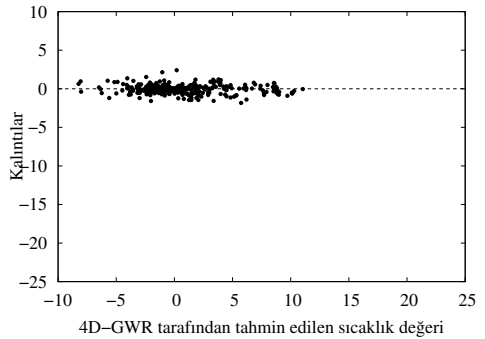
(e) Sıcaklık değerleri



(f) Kalıntılar



(g) Sıcaklık değerleri



(h) Kalıntılar

Şekil 4.5. Saat 00:00 (24) için 12 Aralık 2016 tarihinde GWR yöntemlerinin sonuçları ve kalıntıları

4.4. Sonular

Bu alıřmada veri kmelerindeki meknsal, rakımsal ve zamansal olarak deėiřen iliřkileri ele almak iin coėrafi, rakımsal ve zamansal aėırlıklı regresyon (4D-GWR) yaklařımı nerilmiřtir. Bilgimiz dahilinde, literatrde veri kmelerindeki meknsal, rakımsal ve zamansal deėiřiklikleri ele alan bir yaklařım yoktur. 4D-GWR'nin normalleřtirme katsayısı ve bant geniřliėi deėerlerini tahmin etmek iin n-boyutlu altın kesit arama (n-dimension golden section search- nD-GSS) algoritmasını kullanan yeni bir parametre tahmin stratejisi nerilmiřtir. Bu alıřmada, nerilen 4D-GWR'nin bellek karmařıklıėının stesinden gelmek ve matris hesaplamalarını hızlandırmak iin FastGWR [26] ve RNN-GWR'de olduėu gibi noktasal arpım iřlemi kullanılmaktadır. Ayrıca, nerilen 4D-GWR yaklařımı dengeleme katsayıları ve bant geniřliėi deėerlerini verimli bir řekilde bulmak iin n-boyutlu altın kesit arama yntemini [102, 105, 106] kullanmaktadır. nerilen 4D-GWR algoritmasının performansı sentetik veri kmeleri ve gerek meteorolojik veri kmesi kullanılarak klasik GWR (2D-GWR), GAWR [32] ve GTWR [30] yaklařımlarının performansları ile karřılařtırılmıřtır. Deneyisel deėerlendirmeler, nerilen 4D-GWR yaklařımının doėruluk ve hız aısından diėer yaklařımlardan daha iyi performans elde ettiėini gstermektedir.

5. BÖLÜM

TARTIŞMA, SONUÇ ve ÖNERİLER

5.1. Tartışma ve Sonuç

Bu tez çalışmasında, büyük verileri coğrafi ağırlıklı regresyon yöntemi (geographically weighted regression - GWR) ile etkin bir şekilde analiz etmek için farklı yaklaşımlar geliştirilmiştir. Geliştirilen yöntemlerin her biri farklı tipteki veri kümesini daha etkin bir şekilde analiz etmek için tasarlanmıştır. Geliştirilen yöntemler ve literatüre katkıları aşağıdaki şekilde sıralanabilir.

- Bölüm 2’de, sık güncellenen veri kümelerini analiz etmek için ters en yakın komşu (RNN) stratejisini kullanan hesaplama açısından etkili bir GWR algoritması (RNN-GWR) önerilerek literatüre kazandırılmıştır. Önerilen RNN-GWR algoritmasının başarımı literatürdeki Toy-GWR ve güncel FastGWR [26] algoritmaları ile kıyaslanarak değerlendirilmiştir. Yapılan deneysel çalışmalar, önerilen RNN-GWR algoritmasının literatürdeki Toy-GWR ve FastGWR algoritmalarına göre oldukça etkili bir şekilde çalıştığını göstermiştir. Yapılan çalışmanın sonuçları derlenerek bir dergide yayımlanmıştır [31].
- Bölüm 3’te, hızlı coğrafi ve zamansal ağırlıklı regresyon yaklaşımı (FastGTWR) önerilmiştir. Önerilen FastGTWR algoritmasının başarımı literatürdeki GTWR algoritmaları ile kıyaslanarak değerlendirilmiştir. Yapılan deneysel çalışmalar, önerilen FastGTWR algoritmasının literatürdeki GTWR algoritmalarına göre oldukça etkili bir şekilde çalıştığını göstermiştir. Yapılan çalışmanın sonuçları derlenerek bir dergiye gönderilmiştir.
- Bölüm 4’de veri kümelerindeki mekânsal, rakımsal ve zamansal olarak değişen

ilişkileri ele almak için coğrafi, rakımsal ve zamansal ağırlıklı regresyon (4D-GWR) önerilerek literatüre kazandırılmıştır. Önerilen GAWR algoritmasının başarısı gerçek ve sentetik veri kümesi üzerinde literatürdeki GWR, GAWR ve GTWR algoritmaları ile kıyaslanarak değerlendirilmiştir. Yapılan deneysel çalışmalar, önerilen 4D-GWR algoritmasının literatürdeki GWR, GAWR ve GTWR algoritmalarına göre oldukça etkili bir şekilde çalıştığını göstermiştir. Yapılan çalışmanın sonuçları derlenerek bir dergiye gönderilmiştir.

Geliştirilen yöntemler kendi aralarında kıyaslandığında öne çıkan özellikleri aşağıdaki şekilde sıralanabilir.

- RNN-GWR yaklaşımı
 - Mekânsal veri kümelerini analiz eder.
 - Bi-square çekirdek fonksiyonu kullanır.
 - Büyük veri kümelerini analiz edebilmektedir.
 - Ters en yakın komşu stratejisinden faydalanır.
 - Sık güncellenen mekânsal veri kümelerinde FastGTWR ve 4D-GWR yaklaşımlara göre daha hızlı çalışır.
- FastGTWR yaklaşımı
 - Mekânsal ve zamansal ilişki içeren veri kümelerini analiz eder.
 - Gaussian çekirdek fonksiyonu kullanır.
 - Büyük veri kümelerini analiz edebilmektedir.
 - Mekânsal ve zamansal veri kümelerinde RNN-GWR yaklaşımına göre daha hızlı çalışır ve daha doğru sonuçlar üretir.
 - Mekânsal ve zamansal veri kümelerinde 4D-GWR yaklaşımına göre daha yavaş çalışır ve aynı sonuçları üretir.
- 4D-GWR yaklaşımı
 - Mekânsal, rakımsal ve zamansal ilişki içeren veri kümelerini analiz eder.

- Gaussian çekirdek fonksiyonu kullanır.
- Büyük veri kümelerini analiz edebilmektedir.
- Mekânsal, rakımsal ve zamansal ilişkileri bulmak için yeni bir arama stratejisi kullanır.
- Mekânsal, rakımsal ve zamansal veri kümelerinde RNN-GWR ve FastGTWR yaklaşımına göre daha hızlı çalışır ve daha doğru sonuçları üretir.

5.2. Öneriler

- RNN-GWR yaklaşımında ölçüm noktalarının konumları sabit kabul edilmektedir. Gelecekteki çalışmalarda konumları değişen ölçüm noktalarına sahip büyük veri kümelerinde önerilen RNN-GWR algoritması ile çalışma yapmak planlanmaktadır.
- Önerilen FastGWR yaklaşımı tek çekirdekli ortamda çalışmaktadır. Gelecekteki çalışmalarda önerilen algoritmanın paralel ve dağıtık versiyonunun geliştirilmesi ve etkili bant genişliği optimizasyonu yapan yaklaşımlarının geliştirilmesi planlanmaktadır.
- Önerilen 4D-GWR algoritması meteoroloji veri kümesi üzerinde koşturulmuştur. Gelecekteki çalışmalarda önerilen 4D-GWR yaklaşımını genişletmeyi, akan (streaming) veri kümesinde ve hidrolojik veriler [93, 111] gibi diğer veri kümelerinde önerilen yaklaşımı koşturma planlanmaktadır. 4D-GWR yaklaşımının dengeleme katsayıları ve bant genişliği değerlerini aynı anda bulmak için çok amaçlı optimizasyon yöntemlerinden faydalanma planlanmaktadır.

KAYNAKLAR

1. Perera, C., Zaslavsky, A., Christen, P., and Georgakopoulos, D., 2014. Sensing as a service model for smart cities supported by internet of things, **Transactions on Emerging Telecommunications Technologies**, **25**(1):81–93.
2. Khare, P. and Khare, A., 2018. Internet of Things for Smart Cities, Exploring the Convergence of Big Data and the Internet of Things, 96–112, IGI Global.
3. Zanella, A., Bui, N., Castellani, A., Vangelista, L., and Zorzi, M., 2014. Internet of things for smart cities, **IEEE Internet of Things**, **1**(1):22–32.
4. Liu, J., Wang, X., Khattak, A.J., Hu, J., Cui, J., and Ma, J., 2016. How big data serves for freight safety management at highway-rail grade crossings? A spatial approach fused with path analysis, **Neurocomputing**, **181**:38–52.
5. Jiang, X., Abdel-Aty, M., Hu, J., and Lee, J., 2016. Investigating macro-level hotzone identification and variable importance using big data: A random forest models approach, **Neurocomputing**, **181**:53–63.
6. Allam, Z. and Dhunny, Z.A., 2019. On big data, artificial intelligence and smart cities, **Cities**, **89**:80–91.
7. Members, B.D.C., 2019. Database resources of the BIG Data Center in 2019, **Nucleic Acids Research**, **47**(Database issue):D8.
8. Price, W.N. and Cohen, I.G., 2019. Privacy in the age of medical big data, **Nature medicine**, **25**(1):37–43.
9. Tole, A.A. et al., 2013. Big data challenges, **Database systems journal**, **4**(3):31–40.
10. Lu, R., Zhu, H., Liu, X., Liu, J.K., and Shao, J., 2014. Toward efficient and privacy-preserving computing in big data era, **IEEE Network**, **28**(4):46–50.
11. Chen, X., Sun, L., Zhu, H., Zhen, Y., and Chen, H., 2012. Application of internet of things in power-line monitoring, *2012 International conference on cyber-enabled distributed computing and knowledge discovery*, 423–426, IEEE.

12. Xiaojun, C., Xianpeng, L., and Peng, X., 2015. IOT-based air pollution monitoring and forecasting system, *2015 international conference on computer and computational sciences (ICCCS)*, 257–260, IEEE.
13. Reda, H.T., Daely, P.T., Kharel, J., and Shin, S.Y., 2018. On the application of IoT: Meteorological information display system based on LoRa wireless communication, **IETE Technical Review**, **35**(3):256–265.
14. Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region, **Economic Geography**, **46**(sup1):234–240.
15. Fotheringham, A.S., Brunsdon, C., and Charlton, M., 2002. Geographically weighted regression: the analysis of spatially varying relationships, Wiley.
16. Kazar, B.M. and Celik, M., 2012. Spatial Autoregression (SAR) Model: Parameter Estimation Techniques, Springer.
17. Celik, M., Kazar, B.M., Shekhar, S., Boley, D., and Lilja, D.J., 2006. Spatial Dependency Modeling Using Spatial Auto-regression, *Workshop on Geospatial Analysis and Modeling with Geoinformation Connecting Societies (GICON)*, *International Cartography Association (ICA)*, 186–197.
18. Celik, M., Kazar, B.M., Shekhar, S., and Boley, D., 2006. Parameter Estimation for the Spatial Auto-Regression Model: A Rigorous Approach, *In Proc. of the The Second NASA Data Mining Workshop: Issues and Applications in Earth Science*, 1–8, Pasadena , California.
19. Dong, G., Nakaya, T., and Brunsdon, C., 2018. Geographically weighted regression models for ordinal categorical response variables: An application to geo-referenced life satisfaction data, **Computers, Environment and Urban Systems**, **70**:35–42.
20. Lu, B., Charlton, M., Harris, P., and Fotheringham, A.S., 2014. Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data, **International Journal of Geographical Information Science**, **28**(4):660–681.

21. Wheeler, D.C. and Páez, A., 2010. Geographically weighted regression, *Handbook of applied spatial analysis*, 461–486, Springer.
22. Kauermann, G. and Opsomer, J., 2004. Generalized cross-validation for bandwidth selection of backfitting estimates in generalized additive models, **Journal of Computational and Graphical Statistics**, **13**(1):66–89.
23. Zhang, X., King, M.L., and Hyndman, R.J., 2006. A Bayesian approach to bandwidth selection for multivariate kernel density estimation, **Computational Statistics & Data Analysis**, **50**(11):3009–3031.
24. Harris, R., Singleton, A., Grose, D., Brunsdon, C., and Longley, P., 2010. Grid-enabling Geographically Weighted Regression: A Case Study of Participation in Higher Education in England, **Transactions in GIS**, **14**(1):43–61.
25. Tran, H.T., Nguyen, H.T., and Tran, V.T., 2016. Large-scale geographically weighted regression on Spark, *Eighth International Conference on Knowledge and Systems Engineering (KSE)*, 127–132.
26. Li, Z., Fotheringham, A.S., Li, W., and Oshan, T., 2019. Fast Geographically Weighted Regression (FastGWR): a scalable algorithm to investigate spatial process heterogeneity in millions of observations, **International Journal of Geographical Information Science**, **33**(1):155–175.
27. Brunsdon, C., Fotheringham, A.S., and Charlton, M.E., 1996. Geographically weighted regression: a method for exploring spatial nonstationarity, **Geographical analysis**, **28**(4):281–298.
28. Huang, B., Wu, B., and Barry, M., 2010. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices, **International Journal of Geographical Information Science**, **24**(3):383–401.
29. Fotheringham, A.S., Crespo, R., and Yao, J., 2015. Geographical and Temporal Weighted Regression(GTWR), **Geographical Analysis**, **47**:431–452.

30. Ma, X., Zhang, J., Ding, C., and Wang, Y., 2018. A geographically and temporally weighted regression model to explore the spatiotemporal influence of built environment on transit ridership, **Computers, Environment and Urban Systems**, **70**:113–124.
31. Tasyurek, M. and Celik, M., 2020. RNN-GWR: A Geographically Weighted Regression Approach For Frequently Updated Data, **Neurocomputing**, **33**(1):155–175.
32. Propastin, P., 2012. Modifying geographically weighted regression for estimating aboveground biomass in tropical rainforests by multispectral remote sensing data, **International Journal of Applied Earth Observation and Geoinformation**, **18**:82–90.
33. Pozdnoukhov, A. and Kaiser, C., 2011. Scalable local regression for spatial analytics, *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 361–364.
34. Guo, L., Ma, Z., and Zhang, L., 2008. Comparison of bandwidth selection in application of geographically weighted regression: a case study, **Canadian Journal of Forest Research**, **38**(9):2526–2534.
35. Hadayeghi, A., Shalaby, A.S., and Persaud, B.N., 2010. Development of planning level transportation safety tools using Geographically Weighted Poisson Regression, **Accident Analysis & Prevention**, **42**(2):676–688.
36. Gollini, I., Lu, B., Charlton, M., Brunsdon, C., and Harris, P., 2013. GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models, **arXiv preprint arXiv:1306.0413**.
37. Bidanset, P.E. and Lombard, J.R., 2014. The effect of kernel and bandwidth specification in geographically weighted regression models on the accuracy and uniformity of mass real estate appraisal, **Journal of Property Tax Assessment & Administration**, **10**(3):5–14.

38. Lu, B., Harris, P., Charlton, M., and Brunsdon, C., 2015. Calibrating a geographically weighted regression model with parameter-specific distance metrics, **Procedia Environmental Sciences**, **26**:109–114.
39. Lu, B., Brunsdon, C., Charlton, M., and Harris, P., 2017. Geographically weighted regression with parameter-specific distance metrics, **International Journal of Geographical Information Science**, **31**(5):982–998.
40. Lu, B., Yang, W., Ge, Y., and Harris, P., 2018. Improvements to the calibration of a geographically weighted regression with parameter-specific distance metrics and bandwidths, **Computers, Environment and Urban Systems**, **71**:41–57.
41. Fox, J., 1997. Applied regression analysis, linear models, and related methods., Sage Publications, Inc.
42. Draper, N.R. and Smith, H., 1998. Applied regression analysis, volume 326, John Wiley & Sons.
43. Chatterjee, S. and Hadi, A.S., 2015. Regression analysis by example, John Wiley & Sons.
44. Montgomery, D.C., Peck, E.A., and Vining, G.G., 2012. Introduction to linear regression analysis, volume 821, John Wiley & Sons.
45. Gollini, I., Lu, B., Charlton, M., Brunsdon, C., and Harris, P., 2015. GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models, **Journal of Statistical Software**, **63**(17):1–50.
46. Nakaya, T., Fotheringham, A.S., Brunsdon, C., and Charlton, M., 2005. Geographically weighted Poisson regression for disease association mapping, **Statistics in Medicine**, **24**(17):2695–2717.
47. Windle, M.J., Rose, G.A., Devillers, R., and Fortin, M.J., 2009. Exploring spatial non-stationarity of fisheries survey data using geographically weighted regression (GWR): an example from the Northwest Atlantic, **ICES Journal of Marine Science**, **67**(1):145–154.

48. Hoaglin, D.C. and Welsch, R.E., 1978. The hat matrix in regression and ANOVA, **The American Statistician**, **32**(1):17–22.
49. Nunes, A. and Lourenço, L., 2018. Spatial Association Between Forest Fires Incidence and Socioeconomic Vulnerability in Portugal, at Municipal Level, Integrating Disaster Science and Management, 83–97, Elsevier.
50. Lu, B., Charlton, M., Brunsdon, C., and Harris, P., 2016. The Minkowski approach for choosing the distance metric in geographically weighted regression, **International Journal of Geographical Information Science**, **30**(2):351–368.
51. da Silva, A.R. and de Oliveira Lima, A., 2017. Geographically Weighted Beta Regression, **Spatial Statistics**, **21**:279–303.
52. Leong, Y.Y. and Yue, J.C., 2017. A modification to geographically weighted regression, **International journal of health geographics**, **16**(1):11.
53. Wei, C.H. and Qi, F., 2012. On the estimation and testing of mixed geographically weighted regression models, **Economic Modelling**, **29**(6):2615–2620.
54. Li, J., Wang, B., Wang, G., and Zhang, Y., 2016. Probabilistic threshold query optimization based on threshold classification using ELM for uncertain data, **Neurocomputing**, **174**:211–219.
55. Garcia-Gutierrez, J., Martinez-Alvarez, F., Troncoso, A., and Riquelme, J., 2015. A comparison of machine learning regression techniques for LiDAR-derived estimation of forest variables, **Neurocomputing**, **167**:24–31.
56. Oliveira, S., Pereira, J.M., San-Miguel-Ayanz, J., and Lourenço, L., 2014. Exploring the spatial patterns of fire density in Southern Europe using Geographically Weighted Regression, **Applied Geography**, **51**:143–157.
57. Wang, K., Zhang, C., and Li, W., 2013. Predictive mapping of soil total nitrogen at a regional scale: A comparison between geographically weighted regression and cokriging, **Applied Geography**, **42**:73–85.

58. Gilbert, A. and Chakraborty, J., 2011. Using geographically weighted regression for environmental justice analysis: Cumulative cancer risks from air toxics in Florida, **Soc. Sci. Res.**, **40**(1):273–286.
59. Hsueh, Y.H., Lee, J., and Beltz, L. Spatio-temporal patterns of dengue fever cases in Kaoshiung City, Taiwan, 2003–2008, **Applied Geography**, volume=**34**, pages=**587-594**, doi=**10.1016/j.apgeog.2012.03.003**, year=**2012**.
60. Zhang, H., Zhang, J., Lu, S., Cheng, S., and Zhang, J., 2011. Modeling hotel room price with geographically weighted regression, **International Journal of Hospitality Management**, **30**(4):1036–1043.
61. Lin, J. and Cromley, R.G., 2015. A local polycategorical approach to areal interpolation, **Computers, Environment and Urban Systems**, **54**:23–31.
62. Nakaya, T., Fotheringham, S., Charlton, M., and Brunsdon, C., 2009. Semiparametric geographically weighted generalised linear modelling in GWR 4.0.
63. Li, Y., Jiao, Y., and Browder, J.A., 2016. Modeling spatially-varying ecological relationships using geographically weighted generalized linear model: A simulation study based on longline seabird bycatch, **Fisheries Research**, **181**:14–24.
64. Zhang, C. and Yang, Y., 2020. Modeling the spatial variations in anthropogenic factors of soil heavy metal accumulation by geographically weighted logistic regression, **Science of The Total Environment**, 137096.
65. Brunsdon, C., Fotheringham, A.S., and Charlton, M., 1999. Some notes on parametric significance tests for geographically weighted regression, **Journal of Regional science**, **39**(3):497–524.
66. Yang, W., 2014. An extension of geographically weighted regression with flexible bandwidths, Doktora Tezi, University of St Andrews.

67. Lu, B., Charlton, M., and Fotheringham, A.S., 2011. Geographically weighted regression using a non-Euclidean distance metric with a study on London house price data, **Procedia Environmental Sciences**, 7:92–97.
68. Chu, H.J., Kong, S.J., and Chang, C.H., 2018. Spatio-temporal water quality mapping from satellite images using geographically and temporally weighted regression, **International journal of applied earth observation and geoinformation**, 65:1–11.
69. Guo, Y., Tang, Q., Gong, D.Y., and Zhang, Z., 2017. Estimating ground-level PM_{2.5} concentrations in Beijing using a satellite-based geographically and temporally weighted regression model, **Remote Sensing of Environment**, 198:140–149.
70. He, Q. and Huang, B., 2018. Satellite-based high-resolution PM_{2.5} estimation over the Beijing-Tianjin-Hebei region of China using an improved geographically and temporally weighted regression model, **Environmental pollution**, 236:1027–1037.
71. Yang, Q., Yuan, Q., Yue, L., and Li, T., 2020. Investigation of the spatially varying relationships of PM_{2.5} with meteorology, topography, and emissions in China in 2015 by using modified geographically weighted regression, **Environmental Pollution**, 114257.
72. Wheeler, D.C., 2007. Diagnostic tools and a remedial method for collinearity in geographically weighted regression, **Environment and Planning A**, 39(10):2464–2481.
73. da Silva, A.R. and Mendes, F.F., 2018. On comparing some algorithms for finding the optimal bandwidth in geographically weighted regression, **Applied Soft Computing**, 73:943–957.
74. Oshan, T.M., Li, Z., Kang, W., Wolf, L.J., and Fotheringham, A.S., 2019. mgwr: A Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale, **ISPRS International Journal of Geo-Information**, 8(6):269.

75. Korn, F. and Muthukrishnan, S., 2000. Influence sets based on reverse nearest neighbor queries, *ACM Sigmod Record*, volume 29, 201–212.
76. Benetis, R., Jensen, C.S., Karciauskas, G., and Saltenis, S., 2002. Nearest neighbor and reverse nearest neighbor queries for moving objects, *In Proceedings of International Database Engineering and Applications Symposium*, 44–53.
77. Stanoi, I., Agrawal, D., and El Abbadi, A., 2000. Reverse nearest neighbor queries for dynamic databases, *ACM SIGMOD workshop on research issues in data mining and knowledge discovery*, 44–53.
78. Chung, Y.C., Su, I.F., and Lee, C., 2018. K-most suitable locations selection, **GeoInformatica**, 1–32.
79. Bryant, A. and Cios, K. RNN-DBSCAN: A Density-Based Clustering Algorithm Using Reverse Nearest Neighbor Density Estimates, **IEEE Transactions on Knowledge and Data Engineering**.
80. Mostafiz, M.I., Mahmud, S., Hussain, M.M.u., Ali, M.E., and Trajcevski, G., 2017. Class-based conditional MaxRS query in spatial data streams, *Proceedings of the 29th International Conference on Scientific and Statistical Database Management*, 13.
81. Li, J., Xia, X., Liu, X., Wang, B., Zhou, D., and An, Y., 2018. Probabilistic group nearest neighbor query optimization based on classification using ELM, **Neurocomputing**, **277**:21–28.
82. Tang, B. and He, H., 2017. A local density-based approach for outlier detection, **Neurocomputing**, **241**:171–180.
83. Friedland, S. and Aliabadi, M., 2018. Linear algebra and matrices, volume 156, SIAM.
84. Cabello, S., Díaz-Báñez, J.M., Langerman, S., Seara, C., and Ventura, I., 2010. Facility location problems in the plane based on reverse nearest neighbor queries, **European Journal of Operational Research**, **202**(1):99–106.

85. Iyer, K., 2015. Computational complexity of data mining algorithms used in fraud detection, Yüksek Lisans Tezi, The Pennsylvania State University.
86. Prasad, A., 2017. Exploring the Convergence of Big Data and the Internet of Things, IGI Global.
87. Novo, O., 2018. Blockchain meets IoT: An architecture for scalable access management in IoT, **IEEE Internet of Things Journal**, 5(2):1184–1195.
88. Stergiou, C., Psannis, K.E., Kim, B.G., and Gupta, B., 2018. Secure integration of IoT and cloud computing, **Future Generation Computer Systems**, 78:964–975.
89. Yıldırım, G. and Tatar, Y., 2019. Uzak kullanıcı destekli bir IoT-WSN sanal laboratuvarı ve test platformu: FıratWSN., **Journal of the Faculty of Engineering & Architecture of Gazi University**, 34(4).
90. Shekhar, S., Vatsavai, R., and Celik, M., 2009, Spatial and Spatiotemporal Data Mining: Recent Advances, Data Mining: Next Generation Challenges and Future Directions.
91. Taşyürek, M. and Çelik, M. Akıllı Durak Sistemindeki Araç Seyahat Sürelerinin Birleşik Yapay Sinir Ağları Kullanarak Tahmini, **Avrupa Bilim ve Teknoloji Dergisi**, 72–79.
92. Çelik, M., Dokuz, A.Ş., and Dadaşer-Çelik, F., 2018. Emerging and Vanishing Association Pattern Mining in Hydroclimatic Datasets, **Karaelmas Science and Engineering Journal**, 8(1):30–37.
93. Celik, M., Dadaser-Celik, F., and Dokuz, A.S., 2014. Discovery of hydrometeorological patterns, **Turkish Journal of Electrical Engineering & Computer Sciences**, 22(4):840–857.
94. Celik, M. and Dokuz, A.S., 2017. Discovering socio-spatio-temporal important locations of social media users, **Journal of Computational Science**, 22:85–98.

95. Li, Z. and Fotheringham, A.S., 2020. Computational improvements to multi-scale geographically weighted regression, **International Journal of Geographical Information Science**, 1–20.
96. Fotheringham, A.S., Yang, W., and Kang, W., 2017. Multiscale geographically weighted regression (MGWR), **Annals of the American Association of Geographers**, **107**(6):1247–1265.
97. Miles, J. and Shevlin, M., 2001. Applying regression and correlation: A guide for students and researchers, Sage.
98. Israeli, O., 2007. A Shapley-based decomposition of the R-square of a linear regression, **The Journal of Economic Inequality**, **5**(2):199–212.
99. Tan, P.N., Steinbach, M., and Kumar, V., 2018. Introduction to Data Mining, Pearson.
100. Chan, Y., 2004. Biostatistics 201: linear regression analysis, **Singapore Medical Journal**, **45**(2):55–61.
101. Taşyürek, M. and Çelik, M. Hava Sıcaklık Değerlerinin Coğrafi ve Rakım Ağırlıklı Regresyon Yöntemi ile Tahmin Edilmesi, **Avrupa Bilim ve Teknoloji Dergisi**, 81–86.
102. Chang, Y.C., 2009. N-dimension golden section search: Its variants and limitations, *2009 2nd International Conference on Biomedical Engineering and Informatics*, 1–6, IEEE.
103. Egrioglu, E., Aladag, C.H., Yolcu, U., Uslu, V.R., and Basaran, M.A., 2010. Finding an optimal interval length in high order fuzzy time series, **Expert Systems with Applications**, **37**(7):5052–5055.
104. Sharma, T.K., Pant, M., and Singh, V.P., 2012. Improved local search in artificial bee colony using golden section search, **arXiv preprint arXiv:1210.6128**.
105. Holländer, J., 2015. Investigating the performance of matrix factorization techniques applied on purchase data for recommendation purposes.

106. Kristianto, R.P. and Setyanto, A., 2018. Golden Section Search-Multi Variable Algorithm for Optimization Parameter of Triple Exponential Smoothing Algorithm to Predict Sufferers of Lungs Disease, *2018 3rd International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE)*, 194–198, IEEE.
107. Cuevas, E., Enríquez, L., Zaldívar, D., and Pérez-Cisneros, M., 2018. A selection method for evolutionary algorithms based on the Golden Section, **Expert Systems with Applications**, **106**:183–196.
108. Tsai, C., Kolibal, J., and Li, M., 2010. The golden section search algorithm for finding a good shape parameter for meshless collocation methods, **Engineering Analysis with Boundary Elements**, **34**(8):738–746.
109. Willmott, C.J. and Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, **Climate research**, **30**(1):79–82.
110. Chai, T. and Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature, **Geoscientific model development**, **7**(3):1247–1250.
111. Dadaser-Celik, F., Celik, M., and Dokuz, A., 2012. Associations between stream flow and climatic variables at Kızılırmak River Basin in Turkey, **Global Nest Journal**, **14**:354–361.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı, Soyadı: Murat TAŞYÜREK
Uyruğu: Türkiye (T.C.)
Doğum Tarihi ve Yeri: 1986 - Kayseri
Email: m_tasyurek@hotmail.com
Telefon: +90 539 287 54 22
Yazışma Adresi: Erciyes Üni. Müh. Fakültesi, Bilgisayar Müh. Bölümü,
 Melikgazi, Kayseri

EĞİTİM

Derece	Kurum	Mez.Yılı
Lisans	Erciyes Üni., Bilgisayar Mühendisliği, Kayseri, Türkiye	2009
Yüksek Lisans	Erciyes Üni., Bilgisayar Mühendisliği ABD, Kayseri, Türkiye	2011

İŞ DENEYİMİ

Yıl	Kurum	Görev
2008-2010	İltelis İleri Teknoloji Elektronik ve Elektronik Sistemleri, Erciyes Teknopark, Kayseri	Yazılım Geliştirme Uzmanı - Bilgisayar Mühendisi
2010-2015	Kayseri Büyükşehir Belediyesi Coğrafi Bilgi Sistemleri Şube Müdürlüğü	Bilgisayar Mühendisi
2015-2019	Kayseri Büyükşehir Belediyesi Bilgi İşlem Dairesi Başkanlığı	Bilgi Teknolojileri Şube Müdürü
2015-Hâlen	Kayseri Büyükşehir Belediyesi Bilgi İşlem Dairesi Başkanlığı	Yazılım Şube Müdürü

YAYINLAR

SCI, SSCI, AHCI İndekslerine Giren Dergilerde Yayınlanan Makaleler

- Tasyurek, M., Celik M., "RNN-GWR: A geographically weighted regression approach for frequently updated data" Neurocomputing (2020): 399(258-270)

Diğer İndekslerine Giren Dergilerde Yayınlanan Makaleler

- Tasyurek, M., Celik M., "Akıllı Durak Sistemindeki Araç Seyahat Sürelerinin Birleşik Yapay Sinir Ağları Kullanarak Tahmini." Avrupa Bilim ve Teknoloji Dergisi (Nisan 2020):72-79

- Tasyurek, M., Celik M., "Hava Sıcaklık Değerlerinin Coğrafi ve Rakım
- Ağırlıklı Regresyon Yöntemi ile Tahmin Edilmesi" Avrupa Bilim ve Teknoloji Dergisi (Ağustos 2020):81-86

