



## FastGTWR: A fast geographically and temporally weighted regression approach

Murat Taşyürek<sup>1\*</sup>, Mete Çelik<sup>2</sup>

<sup>1</sup>Department of Computer Engineering, Kayseri University, Kayseri, 38280, Turkey

<sup>2</sup>Department of Computer Engineering, Erciyes University, Kayseri, 38039, Turkey

### Highlights:

- A new and fast GTWR approach, FastGTWR, was proposed.
- The performance of the proposed FastGTWR approach was compared with that of the classical GWR and GTWR approaches.
- Experimental evaluations conducted on real and synthetic datasets showed that FastGTWR approach outperformed other approaches.

### Keywords:

- Geographically weighted regression
- Spatial data mining
- Temporal data mining
- FastGWR

### Article Info:

Research Article  
Received: 25.06.2020  
Accepted: 11.10.2020

### DOI:

10.17341/gazimmfd.757131

### Correspondence:

Author: Murat Taşyürek  
e-mail:  
m\_tasyurek@hotmail.com  
phone: +90 352 207 1718

### Graphical/Tabular Abstract

Geographically Weighted Regression (GWR) method, which is one of the widely used spatial analysis methods, is the local spatial regression technique used to model the changing relationships on geography. Geographically and Temporal Weighted Regression (GTWR) is an approach developed by including temporal relations into the GWR approach. Although the GTWR approach produces much better models than the GWR approach in the dataset containing spatial-temporal heterogeneity, there are still challenges given the complexity of spatial-temporal approaches. Because of this reason, in the literature GTWR models can able to handle limited number of data. In this study, we propose the FastGTWR approach to reduce the algorithmic complexity of GTWR approach and to overcome data size restriction. The proposed FastGTWR approach was run on real data set. The performance of the proposed FastGTWR approach was compared with the performances of the classical GWR and GTWR approaches. Experimental results showed that the proposed FastGTWR approach works faster than the GWR and GTWR approaches.

	GWR	GTWR	FastGTWR
Handling large scale dataset	✗	✗	✓
Dataset size that can be handled on a standard computer	40,000	40,000	1,000,000+
Computation times for 40,000 observation points on a standard computer (seconds)	372	409	130
Analyzing spatial relationships	✓	✓	✓

Figure A. Comparisons of GWR approaches

**Purpose:** The purpose of this study is to develop a fast GTWR approach of FastGTWR to reduce the algorithmic complexity of GTWR approach and to overcome data size restriction

### Theory and Methods:

In this study, the FastGTWR algorithm is proposed to increase the speed of the GTWR algorithm and so to overcome data size restriction. It is observed that that the calculation and storage cost of GTWR approach is due to  $X^T W(u_i, v_i, t_i) X$  matrix multiplication and calculation of inverse of this matrix. The size of neighborhood matrix  $W(u_i, v_i, t_i)$  is  $n \times n$  ( $n$  is the number of regression points). In the proposed FastGTWR approach, the  $W$  matrix was reorganized and converted into a vector format and dot matrix multiplication was used to reduce the complexity of the matrix operations. In addition, considering the fact that  $X^T W(u_i, v_i, t_i) X$  matrix is a symmetrical matrix, only the values at the top or bottom of the diagonal of the matrix were calculated. Thus, the calculation and storage complexity of the GTWR approach was reduced.

### Results:

The performance of the proposed FastGTWR algorithm was compared with that of the GWR and GTWR approaches using synthetic and real meteorological data of Turkey. Experimental evaluations showed that on a standard computer, proposed FastGTWR algorithm can handle 1,000,000 observation points, while the GWR and GTWR algorithms can process a maximum of 40,000 observation points.

### Conclusion:

Experimental evaluations show that the proposed FastGTWR approach is computationally efficient than the GWR and GTWR approaches and can handle large scale datasets .



## FastGTWR: Hızlı coğrafi ve zamansal ağırlıklı regresyon yaklaşımı

Murat Taşyürek<sup>1\*</sup>, Mete Çelik<sup>2</sup>

<sup>1</sup>Kayseri Üniversitesi, Mühendislik Mimarlık ve Tasarım Fakültesi, Bilgisayar Mühendisliği Bölümü, 38280, Talas, Kayseri

<sup>2</sup>Erciyes Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, 38039, Melikgazi, Kayseri

### Ö N E Ç I K A N L A R

- Yeni ve hızlı GTWR yaklaşımı, FastGTWR, önerilmiştir
- Önerilen FastGTWR yaklaşımının performansı GWR ve GTWR yaklaşımlarının performansları ile karşılaştırılmıştır
- Gerçek ve sentetik veriler üzerinde yapılan deneysel çalışmalarda FastGTWR diğer yaklaşımlardan daha iyi performans sergilemiştir

### Makale Bilgileri

Araştırma Makalesi

Geliş: 25.06.2020

Kabul: 11.10.2020

DOI:

10.17341/gazimmfd.757131

### Anahtar Kelimeler:

Coğrafi ağırlıklı regresyon,  
mekansal veri madenciliği,  
zamansal veri madenciliği,  
FastGWR

### ÖZET

Mekânsal analizler günümüzde önemli hale gelmiştir ve çok farklı uygulama alanlarında kullanılmaktadır. Yaygın olarak kullanılan konum temelli analiz yöntemlerinden biri olan Coğrafi Ağırlıklı Regresyon (Geographically Weighted Regression - GWR) coğrafya üzerindeki değişen ilişkileri modellemek için kullanılan bir yerel mekânsal regresyon tekniğidir. Coğrafi ve Zamansal Ağırlıklı Regresyon (Geographically and Temporal Weighted Regression - GTWR) ise GWR yaklaşımının verideki zamansal ilişkileri göz önüne almasıyla geliştirilen bir yaklaşımdır. Veri kümesinde mekân-zamansal heterojenliğin olduğu durumlarda GTWR yaklaşımı GWR yaklaşımına göre daha iyi modeller üretmesine rağmen mekân-zamansal modellerin karmaşıklığı göz önüne alındığında algoritma zaman karmaşıklığı artmaktadır. Bu nedenle, literatürde bulunan GTWR algoritmaları sınırlı sayıda veri üzerinde çalışabilmektedir. Bu çalışmada, GTWR algoritmasının hızını arttırmak ve dolayısı ile veri boyutu kısıtlamasının üstesinden gelmek için hızlı bir GTWR algoritması olan FastGTWR yaklaşımı önerilmiştir. Önerilen FastGTWR yaklaşımının performansı gerçek ve sentetik veriler kullanılarak klasik GWR ve GTWR yaklaşımlarının performanslarıyla karşılaştırılmıştır. Deneysel sonuçlar önerilen FastGTWR yaklaşımının GWR ve GTWR yaklaşımlarına göre daha hızlı çalıştığını ortaya koymuştur.

## FastGTWR: A fast geographically and temporally weighted regression approach

### H I G H L I G H T S

- A new and fast GTWR approach, FastGTWR, was proposed
- The performance of the proposed FastGTWR approach was compared with that of the classical GWR and GTWR approaches
- Experimental evaluations conducted on real and synthetic datasets showed that FastGTWR approach outperformed other approaches

### Article Info

Research Article

Received: 25.06.2020

Accepted: 11.10.2020

DOI:

10.17341/gazimmfd.757131

### Keywords:

Geographically weighted  
regression,  
spatial data mining,  
temporal data mining,  
FastGWR

### ABSTRACT

Spatial analysis has become more important today and is used in several application domains. Geographically Weighted Regression (GWR) method, which is one of the widely used spatial analysis methods, is the local spatial regression technique used to model the changing relationships on geography. Geographically and Temporal Weighted Regression (GTWR) is an approach developed by including temporal relations into the GWR approach. Although the GTWR approach produces much better models than the GWR approach in the dataset containing spatial-temporal heterogeneity, there are still challenges given the complexity of spatial-temporal approaches. Because of this reason, in the literature, GTWR algorithms can able to handle limited number of data. In this study, we proposed the FastGTWR approach to reduce the algorithmic complexity of GTWR approach and overcome data size restriction. The proposed FastGTWR approach was run on real and synthetic dataset. The performance of the proposed FastGTWR approach was compared with the performances of the classical GWR and GTWR approaches. Experimental results showed that the proposed FastGTWR approach works faster than the GWR and GTWR approaches.

## 1. GİRİŞ (INTRODUCTION)

Mekânsal analizler günümüzde önemli hale gelmiştir ve çok farklı uygulama alanlarında kullanılmaktadır. Sensör kullanımının ve nesnelerin internetinin yaygınlaşması [1], nesnelerinin akıllı şehir altyapısını oluşturmaya başlaması büyük miktarlarda [2] ve çok fazla sayıda parametresi olan verilerin toplanmasına imkân vermiştir. Hava kalitesini, hava kirliliğini ve iklim bilgilerini oluşturan sıcaklık, basınç, nem, rüzgâr yönü ve 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. Yaygın olarak kullanılan mekânsal analiz yöntemlerinden biri olan Coğrafi Ağırlıklı Regresyon (Geographically Weighted Regression - GWR), coğrafya üzerindeki değişen ilişkileri modellemek için kullanılan bir yerel mekânsal regresyon tekniğidir [3]. Veri kümesinde mekân-zamansal bilgilerin olduğu durumlarda Coğrafi ve Zamansal Ağırlıklı Regresyon (Geographically and Temporal Weighted Regression - GTWR) yaklaşımı GWR yaklaşımına göre daha iyi modeller üretmektedir [4]. GTWR yaklaşımı GWR yaklaşımına zamansal ilişkileri de ekleyerek geliştirilen bir yaklaşımdır [5, 6]. 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 [7, 8], bu yaklaşımların hesaplama maliyetleri yüksektir [9, 10]. GTWR yaklaşımı, GWR yaklaşımına göre mekân-zamansal ilişki içeren veri kümelerinde daha iyi modeller üretmesine rağmen [11, 12] geliştirilen mekân-zamansal modellerin karmaşıklığı göz önüne alındığında algoritma karmaşıklığı artmaktadır. Literatürde bulunan GTWR modelleri sınırlı sayıda veri üzerinde çalışabilmektedir. Yapılan bir çalışmada, GTWR modelleri tarafından ele alınabilecek maksimum veri kümesi standart masaüstü bilgisayarlarda yaklaşık olarak 15.000 ölçüm noktası olarak tespit edilmiştir [7].

Literatürde, GWR ve GTWR yaklaşımlarının hesaplama maliyetini azaltmak için dağıtık ve paralel platformlarda çalışmalar yapılmıştır. Ancak bu yöntemlerin klasik bilgisayarlarda çalışması zordur. Taşyürek ve Çelik [8] tarafından yapılan çalışmada, GWR yaklaşımının hesaplama maliyeti, komşuluk matrisi  $W$  vektör şeklinde düzenlenerek ve matrislerin simetrik özellikleri kullanılarak azaltılmıştır. Benzeri bir çalışma, GTWR yaklaşımı için yapılmamıştır ve bu çalışmada, bu özellikler kullanılarak GTWR algoritmasının hesaplama ve depolama maliyeti azaltılmıştır. Bu çalışmada, GTWR algoritmasının hızını arttırmak ve dolayısı ile 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  $n \times n$  boyutundaki komşuluk matrisidir ( $n$  regresyon nokta sayısıdır). Önerilen

FastGTWR yaklaşımında, belirtilen matris işlemlerinin karmaşıklığını azaltmak için komşuluk matrisi  $W$  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 sonucunun simetrik bir matris olduğunu göz önüne alınarak bu işlemdeki bütün değerleri hesaplamak yerine sadece diyagonalin alt veya üst tarafındaki değerler hesaplanmıştır. Sonuç olarak önerilen FastGTWR yaklaşımının, klasik GTWR yaklaşımından iki önemli farkı vardır: i) FastGTWR yaklaşımında komşuluk matrisi  $W$  vektör formatına çevrilmiştir ve dolayısı ile matris çarpımları noktasal yapılmıştır ve ii)  $X^T W(u_i, v_i, t_i) X$  matris çarpımı sonucu simetrik olduğundan dolayı, FastGTWR algoritmasında bu simetrik matrisin diyagonalinin alt veya üst tarafındaki değerler hesaplanmamıştır. Bu sayede GTWR yaklaşımının hesaplama ve depolama karmaşıklığı azaltılmıştır. Önerilen FastGTWR algoritmasının performansı, Meteoroloji Genel Müdürlüğü'nden (MGM) alınan gerçek ölçüm verileri ve sentetik veriler kullanılarak klasik GWR ve GTWR yaklaşımlarının performansları ile karşılaştırılmıştır. Deneysel sonuçlar önerilen FastGTWR yaklaşımının klasik yaklaşımlara göre daha hızlı çalıştığını ortaya koymuştur.

### 1.1. Katkılar (Contributions)

Bu çalışmanın temel katkıları aşağıdaki gibi özetlenebilir:

- Bu çalışmada hızlı bir GTWR algoritması olan FastGTWR yaklaşımı önerilmiştir. Önerilen algoritma, dağıtık veya eş zamanlı çalışmaya ihtiyaç duymamaktadır ve klasik GTWR algoritmasının doğruluğundan ödün vermemektedir.
- Önerilen FastGTWR yaklaşımının performansı GWR ve GTWR yaklaşımlarının performansları ile karşılaştırılmıştır.
- Meteoroloji Genel Müdürlüğü'nden elde edilen veriler üzerinde yapılan deneysel çalışmalar FastGTWR algoritmasının klasik yaklaşımlara göre daha hızlı çalıştığını ortaya koymuştur.

Bu makalenin bundan sonraki bölümlerinde sırası ile literatür araştırması sunulmuş, GWR ve GTWR modelleri ile ilgili temel tanımlar verilmiş, önerilen hızlı coğrafi ve zamansal ağırlıklı regresyon yaklaşımı FastGTWR tartışılmış, deneysel değerlendirmeler sunulmuş ve daha sonra sonuçlar ve gelecekteki çalışmalar tartışılmıştır.

## 2. LİTERATÜR ARAŞTIRMASI (LITERATURE RESEARCH)

GWR modelinin verimliliğini ve doğruluğunu arttırmak için farklı uygulama alanlarında çalışmalar yapılmıştır [13, 14]. GTWR yaklaşımı Huang vd. [4] tarafından emlak piyasası verilerinde eşzamanlı ve zamansal duraksama ile başa çıkmak için GWR modeline zamansal etkiler eklenerek geliştirilmiştir. GWR modelinden farklı olarak GTWR, mekânsal ve zamansal heterojenliği modellemek için ağırlık matrislerine hem mekânsal hem de zamansal bilgileri entegre

etmektedir. GTWR yaklaşımı, mekân-zamansal bilgilerin olduğu birçok veri kümesinde uygulanabilen bir yaklaşımdır. Fotheringham vd. [5] tarafından yapılan çalışmada, 1980-1998 yılları arasında Londra'da 19 yıllık ev fiyat verisi kullanarak ev fiyat modellemesi için GTWR modelinin geleneksel GWR daha iyi sonuç verdiği ortaya konulmuştur. Ma vd. [6] tarafından yapılan çalışmada, Pekin'e ait aylık transit akıllı kart ve trafik analiz verisi kullanılarak GTWR'nin performansı, en küçük kareler regresyonu ve GWR modellerinin performansları ile karşılaştırılmıştır ve GTWR modelinin iyi performans sergilediği ortaya konulmuştur. Chu vd. [15] tarafından yapılan çalışmada, GTWR, uydu görüntülerinden su kalitesini tahmin etmek için kullanılmıştır ve GWR ile karşılaştırıldığında, GTWR'nin su kalitesi haritalanmasında daha iyi olduğu ortaya konulmuştur.

GWR yaklaşımının hesaplama maliyetini azaltmak için eşzamanlı ve dağıtık platformlar üzerinde çalışmalar yapılmıştır. Harris vd. [16] GWR yaklaşımın kalibrasyon sürecini hızlandırmak için grid tabanlı GWR yaklaşımı geliştirmiştir. Tran vd. [17] Spark alt yapısından yararlanarak büyük veri kümesini işlemek için dağıtık GWR geliştirmiştir. Pozdnoukhov vd. [18] sürekli akan veriler için mekânsal olarak dağıtık artımlı yerel regresyon modeli önermişlerdir ve veri depolama ve yoğun hesaplamalar için MapReduce kütüphanesini kullanmışlardır. Bu çalışmalar GWR kullanarak akan verileri işlemek için dağıtık ve/veya paralel bilgi işlem kaynakları gerektirmektedir.

Bu çalışmalardan farklı olarak, Li vd. [7] birden fazla CPU çekirdeği üzerinde çalışabilen klasik GWR yaklaşımının paralel bir versiyonu olan FastGWR algoritmasını önermiştir. FastGWR yaklaşımı, klasik GWR yaklaşımlarına göre bağımsız bilgisayarlarda iyi performans göstermiştir ve 1.000.000 kayıta kadar veriyi işleyebilmektedir. Fakat FastGWR yaklaşımı mekân-zamansal bilgileri içeren veri kümesinde GTWR yaklaşımı kadar doğru sonuç gösterememektedir. Bununla birlikte Taşyürek vd. [8] sık güncellenen veriler için RNN-GWR (Reverse Nearest Neighbor-GWR) yaklaşımını önermiştir. RNN-GWR yaklaşımı veri kümesinde sıklıkla güncelleme olduğu durumda bile diğer GWR yaklaşımlarına göre doğruluktan taviz vermeden daha hızlı sonuç üretebilmektedir ve paralel işlem gerektirmemektedir. Fakat RNN-GWR'de klasik GWR yaklaşımının geliştirilmiş bir halidir ve mekân-zamansal bilgisi içeren veri kümesinde GTWR yaklaşımı kadar doğru sonuçlar üretememektedir.

GTWR yaklaşımı, mekân-zamansal veriler için GWR yaklaşımına göre çok daha iyi modeller üretmesine rağmen [11, 12] mekân-zamansal modellerin karmaşıklığı göz önüne alındığında algoritma maliyeti artmaktadır. Literatürde bulunan GTWR modelleri sınırlı sayıda 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 [7, 8].

Bu çalışmada, GTWR algoritmasının çalışma maliyetini azaltmak ve dolayısı ile veri boyutu kısıtlamasının üstesinden gelmek için dağıtık veya eşzamanlı çalışmaya ihtiyaç duymayan ve GTWR yaklaşımının doğruluğundan ödün vermeyen FastGTWR yaklaşımı önerilmektedir.

### 3. COĞRAFI VE ZAMANSAL AĞIRLIKLIL REGRESYON (GEOGRAPHICALLY AND TEMPORAL WEIGHTED REGRESSION)

GTWR yaklaşımı, GWR yaklaşımının mekân-zamansal veriler için genişletilmiş bir versiyonudur. Bu bölümde öncelikle GWR ve daha sonra ise GTWR yaklaşımı ile ilgili temel bilgiler sunulmuştur.

#### 3.1. GWR: Coğrafi ağırlıklı regresyon (Geographically weighted regression)

Sensör kullanımının ve nesnelerin internetinin yaygınlaşması ile büyük hacimli çok fazla sayıda parametresi olan veriler toplanmaktadır [19, 20]. Mekân ve zaman bilgilerini içeren bu veriler veri madenciliği yöntemleri ile analiz edildiğinde faydalı bilgiler elde edilmektedir [21, 22]. Veri madenciliği, büyük ölçekli verinin içindeki ilginç ve anlamlı ilişkilerin yararlı bilgiye dönüştürülmesi olarak tanımlanır [23, 24]. Bu veriler arasındaki ilişkiler genellikle regresyon yöntemi ile bulunur [25, 26]. Regresyon ise birbirine bağımlı ve birbirinden etkilenen iki veya daha fazla değişken arasındaki ilişkiyi bulmak için kullanılan istatistiksel bir yöntem olarak tanımlanır [27]. GWR yöntemi coğrafya üzerindeki değişen ilişkileri modellemek için kullanılan bir yerel mekânsal regresyon tekniğidir [28, 29].

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. GWR, klasik çoklu regresyon modelini temel almaktadır (Eş. 1). Klasik regresyon modelinden farklı olarak, GWR modelinde katsayılar sabit değildir (Eş. 2). Her mekânsal noktanın kendisine ait katsayıları oluşturulmaktadır [30, 31].

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

$$y_i(u_i, v_i) = \beta_{i0}(u_i, v_i) + \sum_{k=1}^m \beta_{ik}(u_i, v_i) x_{ik}(u_i, v_i) + \varepsilon_i(u_i, v_i) \quad (2)$$

Eş. 2'de  $(u_i, v_i)$   $i$  noktanın koordinatları göstermek üzere  $y_i$  bağımlı değişken,  $x_{ik}$  ( $k = 1, \dots, m$ ) bağımsız değişken,  $\beta_{ik}$  regresyon katsayısı ve  $\varepsilon_i$  hata terimi olarak adlandırılır.  $\beta_{ik}$  ( $k = 0, \dots, m$ ) regresyon katsayısı Eş. 3'te gösterildiği şekilde hesaplanır.

$$\beta(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) Y \quad (3)$$

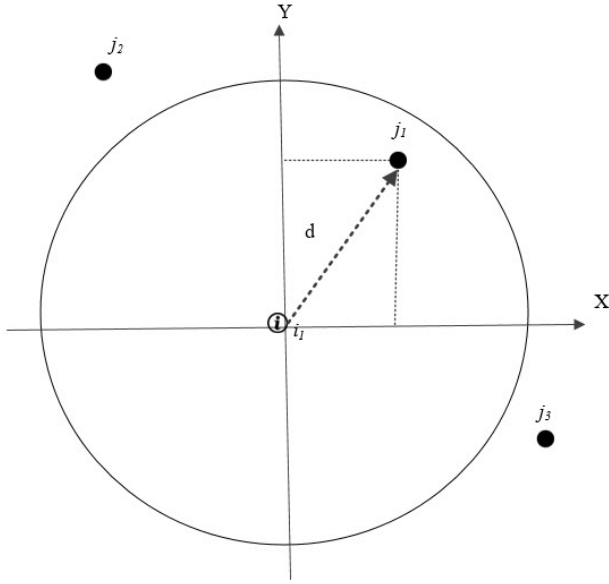
Eş. 3'te  $X$  bağımsız değişkenlerin matrisidir ve  $m + 1$  sütundan oluşur,  $Y = (Y_1, \dots, Y_n)^T$  bağımlı değişken matrisini göstermektedir ve  $W(i)$  ise  $w_{ij}$  değerlerinden oluşan diagonal matristir ve Eş. 4'teki gibi hesaplanır [32].

$$W(i) = \begin{bmatrix} w_{i1} & \dots & 0 \\ \vdots & w_{iz} & 0 \\ 0 & \dots & w_{in} \end{bmatrix} \quad (4)$$

Eş. 4'te  $w_{ij}$  regresyon noktası ile referans nokta arasındaki komşuluk oranıdır.  $w_{ij}$  komşuluk oranı Global Model, Box-Car, Exponential, Gaussian, Bi-Square ve Tri-Cube yöntemleri ile hesaplanır [33, 34]. Yaygın olarak Gaussian ve Bi-Square çekirdek fonksiyonları kullanılmaktadır. Bu çalışmada Gaussian çekirdek fonksiyonu kullanılmıştır. Gaussian çekirdek fonksiyonu ile  $i$  ve  $j$  noktaları arasındaki komşuluk oranı  $w_{ij}$  Eş. 5'deki gibi hesaplanır.

$$w_{ij} = \exp \left[ -\frac{1}{2} \left( \frac{d_{ij}}{bw} \right)^2 \right] \quad (5)$$

Eş. 5'te  $bw$  bant genişliği değerini ve  $d_{ij}$  ise  $i$  regresyon noktası ile  $j$  referans nokta arasındaki mesafeyi göstermektedir.  $i$  regresyon noktası ile  $j$  referans nokta arasındaki mekânsal mesafe Şekil 1'de gösterilmiştir.



Şekil 1 Mekânsal Mesafe (Spatial distance)

Şekil 1'de  $x$  ve  $y$  nokta koordinatlarını göstermek üzere  $d_{ij}$  mesafesi genellikle Öklid mesafesi ile hesaplanır (Eş. 6).

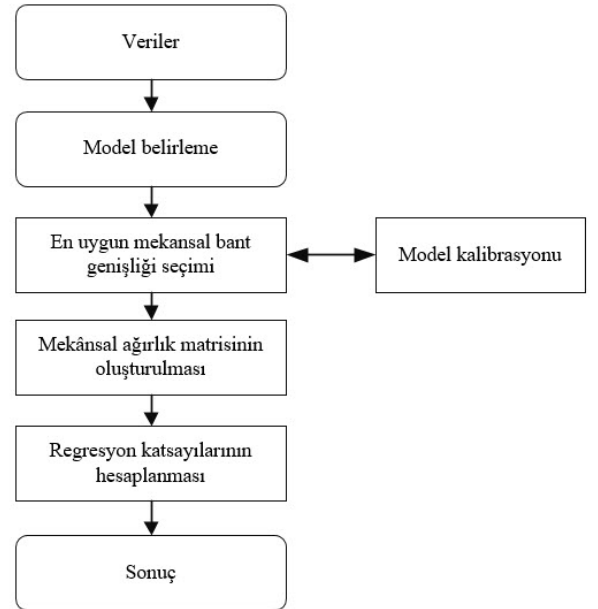
$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (6)$$

GWR modelinde  $bw$  bant genişliği parametresinin değeri bütün bir model için sabit olabilir ya da bölgedeki nokta yoğunluğuna göre değişken olabilir. Nokta dağılımın eşit olduğu veri kümesinde bant genişliğinin sabit ya da adaptif olması modelin doğruluğunu etkilemeyecektir. Nokta dağılımın bazı yerlerde yoğun bazı yerlerde seyrek olduğu

veri kümeleri için bant genişliğinin yoğunluk durumuna göre değişken olması modelin doğruluğunu artırmaktadır [34]. Eğer bant genişliği olarak komşu sayısı seçilirse ve bu komşu sayısı da bütün bir model için sabit olursa bu model de adaptif olarak adlandırılmaktadır [35, 36]. Veri kümesi için en uygun  $bw$  değerini bulmak için çapraz doğrulama (cross-validation - CV), genelleştirilmiş çapraz doğrulama (generalised cross-validation - GCV), Akaike bilgi kriteri (Akaike information Criterion - AIC), Bayes bilgi kriteri (Bayesian information Criterion - BIC) yöntemleri kullanılabilir [37,38]. AIC yönteminin daha doğru sonuç verdiği tespit edilmiştir [39] ve Eş. 7'de gösterildiği gibi hesaplanır. Eşitlikte  $n$  gözlem sayısını ve  $\hat{\sigma}$  hata teriminin tahmini standart sapmasını göstermektedir ve  $tr(S)$  şapka matrisi olarak bilinen  $S$  matrisinin izini göstermektedir.

$$AIC_c = 2nl n(\hat{\sigma}) + nln(2\pi) + n \left\{ \frac{n+tr(S)}{n-2-tr(S)} \right\} \quad (7)$$

GWR algoritmasının genel çalışma yapısı Şekil 2'de gösterilmiştir [5]. Şekil 2'de görüldüğü gibi giriş veri kümesine göre Eş. 1 kullanılarak klasik GWR modeli oluşturulur. Veri kümesine göre Eş. 1 kullanılarak hangi parametrelerin bağımsız ve hangi parametrenin bağımlı değişken olacağına karar verilir. Daha sonra, CV veya AIC metodu ile modele en uygun mekânsal bant genişliği belirlenir. Bu işlemden sonra mekânsal bant genişliği kullanılarak ağırlık matrisi oluşturulur (Eş. 4). Oluşturan bu ağırlık matrisi, bağımsız değişkenlerin matrisi ve bağımlı değişkenlerin matrisi kullanılarak Eş. 3 ile regresyon katsayıları belirlenir.



Şekil 2 GWR Yaklaşımı (GWR Approach)

### 3.2. GTWR: Coğrafi ve zamansal ağırlıklı regresyon (Geographically and temporally weighted regression)

GWR yaklaşımı mekânsal durağanlığını ölçmek için model boyunca  $\beta_{ik}$  regresyon katsayılarının değişmesine izin

verir. Bununla birlikte mekân-zamansal bilgisinin olduğu veri kümelerinde zamansal durağanlığı modelleyemez. Huang vd. [4] mekânsal ve zamansal durağanlıkların eşzamanlı olarak değerlendirilmesi için GWR yaklaşımını geliştirerek GTWR yaklaşımı önermiştir (Eş. 8). GTWR yaklaşımı mekân-zamansal ilişkiyi ölçmek için mekân-zamansal mesafeyi kullanarak bir mekân-zamansal ağırlık matrisi oluşturur.

$$y_i(u_i, v_i, t_i) = \beta_{i0}(u_i, v_i, t_i) + \sum_{k=1}^m \beta_{ik}(u_i, v_i, t_i)x_{ik}(u_i, v_i, t_i) + \varepsilon_i(u_i, v_i, t_i) \quad (8)$$

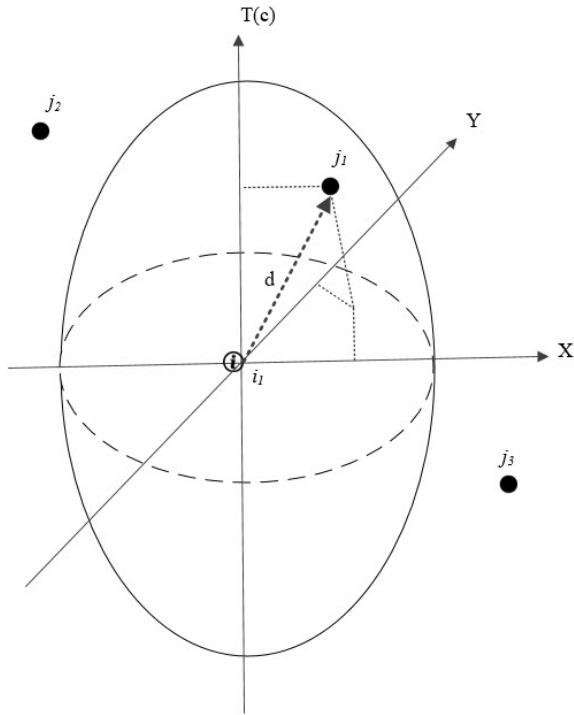
GWR formülünden farklı olarak formüle  $t$  zaman bilgisi eklenmiştir.  $\beta_{ik}(k = 0, \dots, m)$  regresyon katsayısı GWR yaklaşımında olduğu gibi hesaplanır (Eş. 9).

$$\beta(u_i, v_i, t_i) = (X^T W(u_i, v_i, t_i) X)^{-1} X^T W(u_i, v_i, t_i) Y \quad (9)$$

$W(i)$  ağırlık matrisi Eş. 3'te olduğu gibi  $w_{ij}$  değerlerinden oluşur. Fakat burada  $w_{ij}$  mekân-zamansal komşuluk oranı değeridir.  $x$  ve  $y$  nokta koordinatlarını,  $t$  zaman bilgisini ve  $c$  ise normalleştirme katsayısını göstermek üzere  $d_{ij}$  mekân-zamansal mesafesi Eş. 10'da gösterilen formül ile hesaplanır.

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + c(t_i - t_j)^2} \quad (10)$$

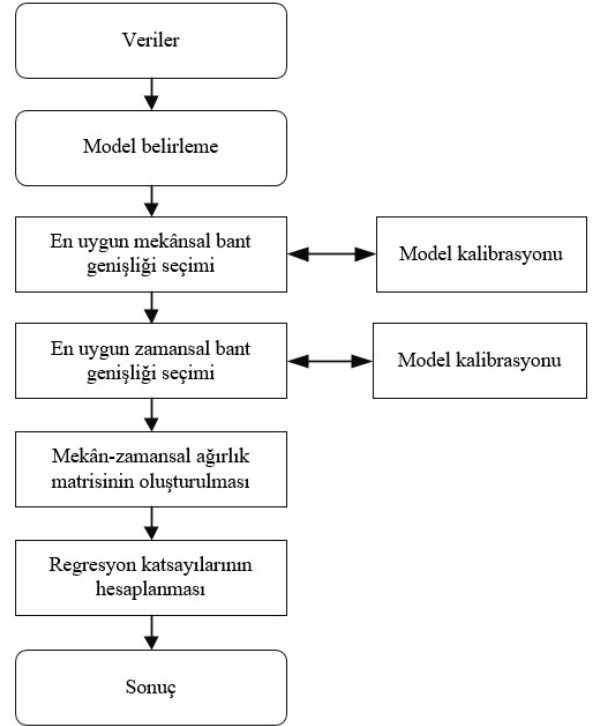
$i$  regresyon noktası ile  $j$  referans nokta arasındaki mekân-zamansal mesafe Şekil 3'te gösterilmiştir.



Şekil 3 Mekân-zamansal Mesafe (Spatio-temporal distance)

Mekân-zamansal ağırlık matrisleri oluşturulmadan önce mekân-zamansal mesafenin uyumlu hale getirilmesi

gerekmektedir. Farklı mekânsal ve zamansal birimleri dengelemek veya uyumlu hale getirmek için  $c$  parametresi kullanılır (Eş. 10) [4]. Bu nedenle, GTWR yaklaşımını uygulamadan önce  $c$  parametresinin optimize edilmesi gerekmektedir. Huang vd. [4] bu parametreyi bütün ihtimalleri deneyerek bulmayı önermiştir. Fotheringham vd. [5] ise bu işlemi algoritma adımları ile gerçekleştirmiştir. GTWR algoritmasının genel çalışma yapısı Şekil 4'te gösterilmiştir [5]. Şekil 4'te görüldüğü gibi, giriş veri kümesine göre Eş. 1 kullanılarak klasik GWR modeli oluşturulur. Burada hangi parametrelerin bağımsız değişken olacağına ve hangi parametrenin bağımlı değişken olacağına karar verilir. Daha sonra, CV veya 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 Eş. 9 ile regresyon katsayıları belirlenir. Daha sonra bu regresyon katsayıları kullanılarak regresyon noktası için değer üretilir.



Şekil 4 GTWR Yaklaşımı (GTWR Approach)

#### 4. FastGTWR: ÖNERİLEN HIZLI COĞRAFİ VE ZAMANSAL AĞIRLIKLILIK REGRESYON (FastGTWR: PROPOSED FAST GEOGRAPHICALLY AND TEMPORALLY WEIGHTED REGRESSION)

GTWR yaklaşımı, mekân-zamansal veriler için GWR yaklaşımına göre çok daha iyi modeller üretmesine rağmen [11, 12] mekânsal-zamansal modelin karmaşıklığı göz önüne

alındığında algoritma karmaşıklığı artmaktadır. Bu çalışmada, GTWR algoritma karmaşıklığını azaltmak ve dolayısı ile veri boyutu kısıtlamasının üstesinden gelmek için dağıtık veya eşzamanlı çalışmaya ihtiyaç duymayan ve GTWR yaklaşımının doğruluğundan ödün vermeyen FastGTWR modeli önerilmiştir.

GWR yaklaşımların hesaplama maliyetinin çoğunluğu matris işlemlerinden kaynaklanmaktadır (Eş. 9) [7, 8].  $n$  regresyon sayısını göstermek üzere, Eş. 9 bütün noktaların katsayılarını bulmak için  $n$  kez çalıştırılmaktadır. Eş. 9'da hesaplama maliyetini en çok oluşturan 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 [7, 8].  $m$  değeri genellikle  $n$  değerinden çok daha küçük olduğu için Eş. 9 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. Eş. 9'un hesaplama ve depolama maliyeti yüksektir ve bu yüzden klasik GWR ve GTWR yaklaşımları sınırlı sayıda veriyi işleyebilmektedir. Bu sorunun üstesinden gelmek için Li vd. [7] ve Taşyürek vd. [8] tarafından matris formlarının yeniden düzenlenmesi önerilmiştir.

$$W(u_i, v_i, t_i) = W(i) = \begin{bmatrix} w_{i1} & \dots & 0 \\ \vdots & w_{i2} & 0 \\ 0 & 0 & \dots \\ 0 & \dots & w_{in} \end{bmatrix}$$

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

$$X^T W(i) = (X \cdot w(i)^T)^T \quad (11)$$

Eş. 11'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 ( $w(i) = (w_{i1}, \dots, w_{in})$ ) çevrilip noktasal olarak çarpılması önerilmiştir. Bu işlem sayesinde Eş. 9'da karşılaşılan hafıza sorunu ortadan kalmaktadır ve hafızada daha az yer kapladığı için işlemler daha hızlı şekilde yapılabilmektedir. Önerilen FastGTWR yaklaşımında Eş. 11'de verilen formül kullanılmıştır. Önerilen FastGTWR algoritmasının kaba kodu Tablo 1'de verilmiştir.

FastGTWR algoritmasında,  $Rn$ 'de bulunan her bir regresyon noktası için Adım 1 ve 6 arasındaki adımlar çalıştırılır. Adım 2'de üzerinde işlem yapılan  $i$  regresyon noktası için bağımsız değişkenlerinin matrisi  $X$  ve referans noktalarının bağımlı değişkenlerinin matrisi  $Y$  oluşturulur. Adım 3'te regresyon noktası  $i$  ve referans noktaları  $Rf$  arasındaki mekân-zamansal mesafeler,  $bw$  ve  $Kn$  parametre değerleri kullanılarak komşuluk ağırlığı hesaplanır ve ağırlık matrisi  $W(i)$  oluşturulur. Adım 4'te  $i$  regresyon noktası için Eş. 9 kullanılarak regresyon katsayıları  $\beta(i)$  hesaplanır. Adım 5'te  $i$  noktası için  $\beta(i)$  ve  $X_i$  değerleri kullanılarak regresyon noktasının tahmini değeri  $\hat{y}_i$  hesaplanır. Daha sonra Adım

6'da  $R^2$  hesaplamasında kullanmak için tahmin edilen  $\hat{y}_i$  değeri ile ölçüm değeri  $y_i$  hafızada saklanır. Son olarak, Adım 7'de algoritma regresyon noktalarının tahmin edilen değerlerini çıktı olarak verir.

**Tablo 1.** FastGTWR Algoritması (FastGTWR Algorithm)

Girdiler:	$D$ : Veri kümesi
	$bw$ : bant genişliği
Değişkenler:	$Kn$ : Çekirdek fonksiyonu $Rn$ : Regresyon noktalar $Rf$ : referans noktaları
Çıktılar:	$\hat{Y}$ : Regresyon noktalarının tahmin edilen değerleri
1	for $i=1; i \leq$ Regresyon nokta sayısı; $i++$ do
2	Bağımsız değişken matrisini ( $X$ ) ve bağımlı değişken matrisini oluştur ( $Y$ );
3	Regresyon noktası $i$ ile referans noktaları $Rf$ arasındaki mekân-zamansal mesafeler $bw$ ve $Kn$ kullanarak ağırlık matrisini ( $W(i)$ ) oluştur;
4	Eşitlik 9 ile $i$ noktasının regresyon katsayılarını hesapla ( $\beta(i)$ );
5	$i$ noktası için $\beta(i)$ ve $X_i$ 'yi kullanarak tahmin değerini ( $\hat{y}_i$ ) hesapla;
6	$R^2$ hesabı için tahmin edilen $\hat{y}_i$ ve ölçüm $y_i$ değerlerini sakla;
7	Return $\hat{Y}$ ;

Taşyürek ve Çelik [8] GWR hesaplama maliyetinin büyük kısmının kaynaklandığı  $X^T W(u_i, v_i, t_i) X$  matrisinin simetrik bir matris olduğunu göz önüne alarak bütün değerleri hesaplamaya gerek olmadığını ve diyagonalin alt veya üst tarafını hesaplamamanın yeterli olduğunu ifade etmişlerdir. Önerilen FastGTWR yaklaşımında simetrik matris özelliğinden faydalanılmıştır ve bu sayede FastGTWR yaklaşımının hesaplama maliyeti  $O(((m^2 + m)/2)n^2)$  olmuştur.

## 5. DENEYSEL ÇALIŞMALAR (EXPERIMENTAL STUDIES)

Bu bölümde klasik GWR, GTWR ve önerilen FastGTWR yaklaşımlarının deneysel performansları değerlendirilmiştir. Deneysel çalışma aşağıdaki sorulara cevap vermek için yapılmıştır.

- Yaklaşımların doğruluğu nedir?
- Yaklaşımların çalışma süreleri ne kadardır?
- Veri boyutunun yaklaşımlar üzerindeki etkisi nedir?

Bu çalışmada mekân-zamansal ilişki içeren meteorolojik veri kümeleri kullanılmıştır. Ancak, mekân-zamansal ilişki içeren diğer veri kümelerine de önerilen FastGTWR algoritması uygulanabilir. Veri kümesi ve model ayarları ile ilgili bilgiler, Bölüm 5.1 ve 5.2'de verilmiştir.

Bu çalışma, Intel Core i5-3210M 2.5 GHz 8GB RAM özelliklerine sahip bir bilgisayar üzerinde Matlab kullanılarak gerçekleştirilmiştir.

### 5.1. Veri (Data)

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 verilerini içermektedir. Bu çalışmada kullanılan istasyon noktaları Şekil 5'te gösterilmiştir. Veri kümesi meteoroloji istasyonlarının rakımını ve saatlik ölçüm olarak hava sıcaklığı, bağıl nem ve hava basınç değerlerini içerir [8]. Bu çalışmada, deneysel değerlendirmelerde kullanılmak üzere rastgele 1600 istasyon seçilmiştir. İstasyonların 1200'ü referans noktası, 400'ü regresyon noktası olarak kullanılmıştır.

MGM' den alınan gerçek veriler 1 aylıktır ve 1746 istasyonun verisini içermektedir ve dolayısı ile veri kümesi boyutunun algoritmalarının performansları üzerindeki etkisini incelemede yetersiz kalabilecektir. Çok fazla sayıda ölçüm noktası içeren veri kümesinde algoritmaların ölçeklenebilirlik performanslarını değerlendirmek için, MGM'den alınan gerçek veriler kullanılarak 1000 referans noktası ve 1000 regresyon noktasından başlayarak 1.000.000 referans noktası ve 1.000.000 regresyon noktası içeren sentetik veri kümesi oluşturulmuştur. Ölçüm noktalarının konumu ve parametre değerleri rastgele belirlenmiştir. Tablo 2'de 1.000.000 referans noktası ve 1.000.000 regresyon noktası içeren sentetik veri kümesinin özet bilgileri sunulmuştur. Tablo 2'de, rakım, bağıl nem, basınç ve sıcaklık parametrelerinin en düşük, ortalama ve en yüksek değerleri gösterilmektedir.

### 5.2. Model ayarları (Model settings)

Deneysel hedeflenen amaç bir noktanın ölçülen bağıl nem, basınç ve rakım değerlerini kullanarak o noktanın sıcaklık değerini tahmin etmektir. Sıcaklık değerini tahmin etmek

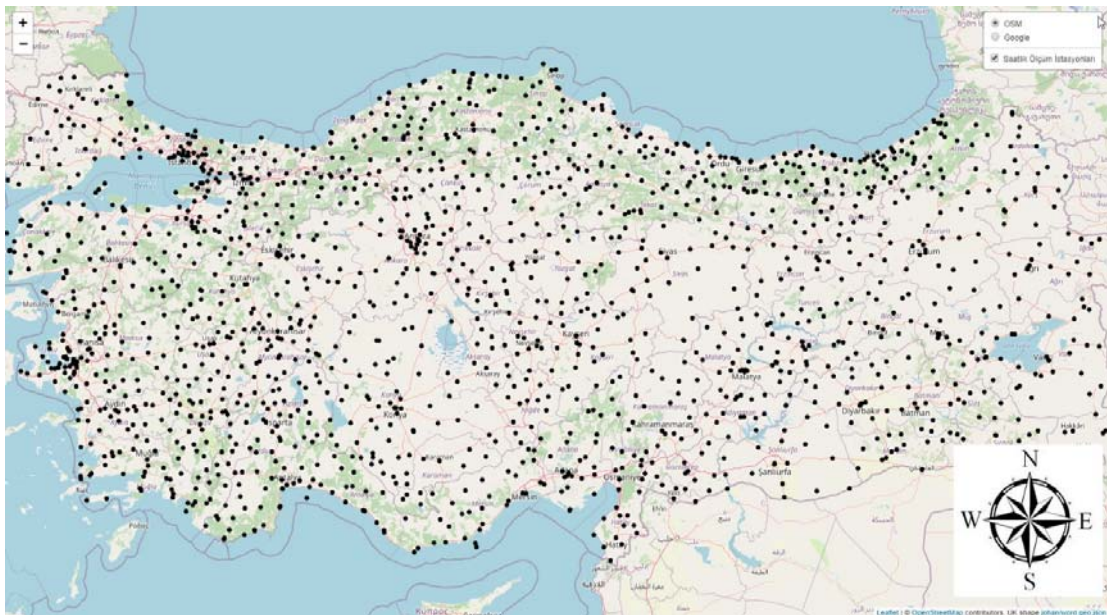
için GWR, GTWR ve FastGTWR yaklaşımlarında kullanılan model Eş. 12'de gösterilmiştir. Eş. 12'de  $i$  regresyon noktasını ve  $\beta$  regresyon katsayılarını temsil etmektedir [8].

**Tablo 2.** Sentetik Veri Kümesi Özet Bilgileri (Synthetic Dataset Summary)

Öznitelik Adı	En Düşük	Ortalama	En Yüksek
Rakım(m)	0	725	2.286
BağılNem(%)	20	75	100
Basınç(hPa)	762,1	889,3	1.017,5
Sıcaklık(°C)	-9,2	13	28

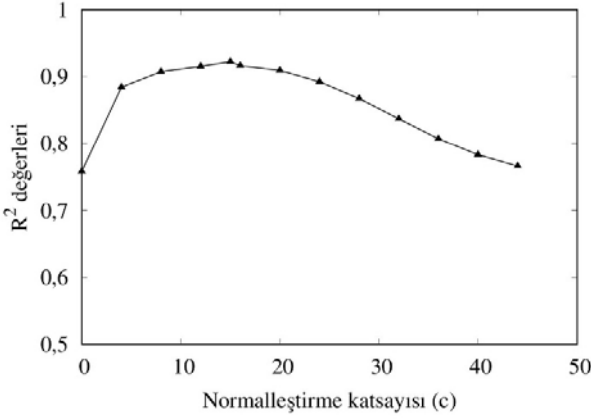
$$Sıcaklık_i = \beta_{0i} + \beta_{1i}Rakım_i + \beta_{2i}Nem_i + \beta_{3i}Basınç_i \quad (12)$$

GWR modelinde öncelikle çekirdek fonksiyonu ve bant genişliği değeri belirlenmelidir. Bu çalışmada GWR yaklaşımlarında yaygın olarak kullanılan ve GTWR yaklaşımlarında da Huang vd. [4] ve Fotheringham vd. [5] tarafından kullanılan Gaussian çekirdek fonksiyonu kullanılmıştır. En uygun mekânsal bant genişliği değerini bulmak için sık kullanılan AIC yöntemi tercih edilmiştir [3] ve bant genişliği 82 km bulunmuştur. GTWR ve FastGTWR yaklaşımları için en uygun mekân-zamansal bant genişliği değeri  $\beta$ , normalleştirme katsayısı  $c$  değeri (Eş. 10) 15,01 olduğunda bulunmuştur. Huang vd. [4]  $c$  değerinin uyum grafiğini  $R^2$  değerine göre göstermiştir. Bu çalışmada kullanılan veri kümesine göre  $c$  değerlerinin oluşturduğu  $R^2$  grafiği Şekil 6'da gösterilmiştir.  $R^2$  bağımsız bir değişken veya bir regresyon modelindeki değişkenler tarafından açıklanan bağımlı bir değişkenin varyans oranını temsil eden istatistiksel bir ölçüdür [24].  $R^2$  değeri 1'e ne kadar yakınsa model o kadar iyi temsil edilmiştir. Eş. 10'da  $t_i - t_j$  arasındaki zaman farkı saniye cinsinden alınmıştır.



**Şekil 5** MGM Saatlik Ölçüm İstasyonları (MGM Hourly Observation Stations)





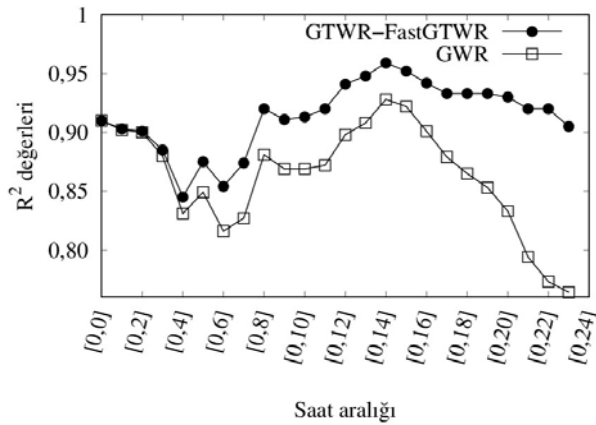
**Şekil 6.** GTWR Modeli için Parametre Seçimi  
(Parameter Selection For GTWR Model)

### 5.3. Deneyler (Experiments)

Bu bölümde GWR, GTWR ve önerilen FastGTWR yaklaşımlarının deneysel değerlendirilmesi sunulmuştur.

#### 5.3.1. Yaklaşımların doğruluğu (Accuracy of approaches)

Bu bölümde, yaklaşımlar tarafından tahmin edilen sıcaklık değeri ile gerçek sıcaklık değerlerinin karşılaştırması yapılmıştır. Bu karşılaştırma  $R^2$  değerine göre yapılmıştır (Şekil 7). GWR, GTWR ve FastGTWR yaklaşımları Eş. 12'yi kullanmaktadır. GTWR ve FastGTWR yaklaşımları aynı doğruluk değerlerine sahiptir. Bu nedenle Şekil 7'de GTWR ve FastGTWR yaklaşımlarının davranışı aynıdır ve her ikisinin değerleri beraber gösterilmiştir.



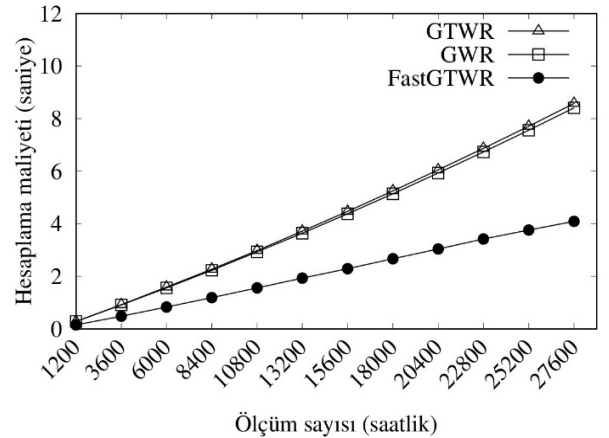
**Şekil 7.** Yaklaşımların Doğruluğu (Accuracy of Approaches)

GWR, GTWR ve FastGTWR yaklaşımlarının sonucunu etkileyecek fark daha önce bahsedildiği üzere GWR yaklaşımı sadece mekânsal ilişkileri incelerken GTWR ve FastGTWR yaklaşımları mekânsal ve zamansal ilişkileri inceler. Bu deneyde, algoritmaların doğruluğu her bir saat için incelenmiştir. Referans ve regresyon noktalarının sayısı sırasıyla 1200 ve 400 olarak ayarlanmıştır. Her bir saat için regresyon nokta sayısı sabit kalırken dikkate alınan ölçüm

sayısı 1200 artmaktadır (Şekil 7). İlk çalışma saatinde geçmişe dair bir veri olmadığı için GWR, GTWR ve FastGTWR yaklaşımları aynı sonuçları üretmektedir. Saat aralığı arttırıldığında, GTWR ve FastGTWR yaklaşımları geçmiş verileri mekân-zamansal ağırlıkları ile dikkate almaktadır. Buna karşılık, GWR yaklaşımı geçmiş verileri sadece mekânsal ağırlıkları ile dikkate almakta ve zamansal ağırlıkları dikkate almamaktadır. Şekil 7'de görüldüğü gibi saat aralığı arttıkça GTWR ve önerilen FastGTWR yaklaşımı GWR yaklaşımına göre daha doğru sonuçlar üretmektedir.

#### 5.3.2. Yaklaşımların çalışma süresi (Run-time of approaches)

Bu deneyde, her saat için gerçek 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 5.2'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 [8] ve bu sebeple hem klasik GWR hem de GTWR yaklaşımından daha iyi performans göstermiştir (Şekil 8). GWR yaklaşımı çok az da olsa GTWR yaklaşımına göre daha hızlıdır. Çünkü GTWR yaklaşımın komşuluk oranının hesaplama maliyeti GWR yaklaşımın komşuluk oranı 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 kaldığı için, bu deneyde algoritmaların veri kümesinin büyüklüğüne göre performanslarını göstermemektedir. Bu nedenle veri boyutunun yaklaşımlar üzerine etkisi bir sonraki deneyde incelenmiştir.



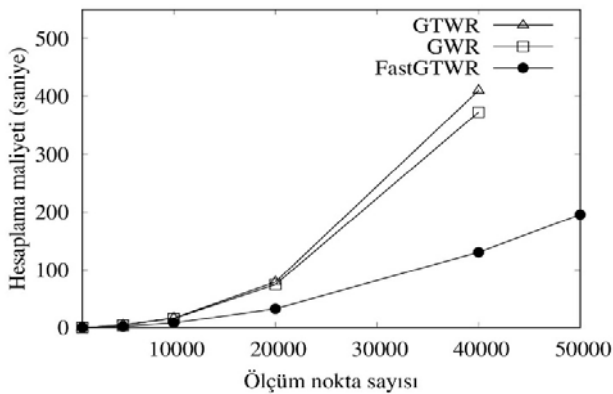
**Şekil 8.** Yaklaşımların Saat Bazlı Hesaplama Maliyeti  
(Hourly Calculation Cost of Approaches)

#### 5.3.3. Veri boyutunun yaklaşımlar üzerindeki etkisi (The effect of data size on approaches)

Bu deneyde, veri boyutunun GWR, GTWR ve FastGTWR yaklaşımlarının üzerindeki etkisi değerlendirilmiştir.

MGM'den alınan gerçek veriler bir aylıktır ve 1746 adet istasyona ait meteorolojik verileri içermektedir ve bu veri kümesi veri boyutunun algoritmalarının performansları üzerindeki etkisini incelemekte yetersiz kalmaktadır. Bu nedenle, bu deney çalışması için Bölüm 5.1'de anlatıldığı gibi üretilen sentetik veri kümesi kullanılmıştır. Her çalışmada bant genişliği değerini bulma işlem süresini uzatacağından deneylerde önceden tanımlanmış bant genişliği değerlerini kullanmıştır. Deney sonuçları Tablo 3 ve Şekil 9'da verilmiştir. Veri kümesi boyutları arttıkça tüm algoritmaların hesaplama maliyeti de artmaktadır. Li vd. [7] klasik GWR yaklaşımlarının normal bir bilgisayarda en fazla 15.000 ölçüm noktası işleyebildiğini tespit etmişlerdir. Fakat bu özellik kullanılan bilgisayara göre değişebilmektedir. Bu çalışmada, bu değer 40.000 ölçüm noktası olarak tespit edilmiştir [7, 8]. Klasik GWR ve GTWR algoritması, Tablo 3 ve Şekil 9'da görüleceği gibi 40.000 ölçüm noktasına kadar veri kümelerini işleyebilmektedir. Önerilen FastGTWR algoritması, GWR ve GTWR algoritmalarından daha iyi performans göstermiştir. Bunun sebebi ilk olarak Eş. 11'de bulunan yeni dönüşümü kullanmasıdır. Eş. 11 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 [8] ve bu sayede  $X^T W(u_i, v_i, t_i) X$  işlemi sonucunda üretilen simetrik matrisin diyagonalinin alt veya üst kısmı hesaplanmadığından önerilen yaklaşımın hesaplama maliyeti ciddi oranda azalmıştır.

FastGTWR algoritması GTWR algoritmasının hesaplama karmaşıklığının düşürülmesi için önerilmiştir. Deneylere GWR algoritması da dahil edilmiştir. GWR algoritması mekânsal uzaklığı kullanırken GTWR algoritması mekân-zamansal uzaklığı kullanmaktadır. Bu durum GWR algoritmasının daha hızlı olmasına neden olmaktadır. Ancak yapılan deneylerde, mekân-zamansal uzaklığı kullanmasına rağmen, FastGTWR algoritmasının hem GTWR'den hem de GWR'den daha hızlı çalıştığı ve GWR'den daha doğru sonuç ürettiği ortaya konulmuştur.



Şekil 9. Yaklaşımların Hesaplama Maliyeti  
(Calculation Cost of Approaches)

Tablo 3. 1.000'den 1.000.000'a kadar ölçüm noktası için yaklaşımların çalışma zamanları (saniye)  
(Run-times of approaches for observation points from 1,000 to 1,000,000 (second))

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

## 6. SONUÇLAR VE TARTIŞMALAR (RESULTS AND DISCUSSIONS)

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 GWR ve GTWR yaklaşımlarından hesaplamalı olarak etkili olduğunu göstermiştir.

Gelecekteki çalışmalarda önerilen algoritmanın paralel ve dağıtık versiyonunun geliştirilmesi ve etkili bant genişliği optimizasyonu yapan yaklaşımlarının geliştirilmesi planlanmaktadır.

## KAYNAKLAR (REFERENCES)

- Perera, C., Zaslavsky, A., Christen, P., Georgakopoulos, D., Sensing as a service model for smart cities supported by internet of things, Transactions on Emerging Telecommunications Technologies, 25 (1), 81-93, 2014.
- Prasad, A. V., Exploring the convergence of big data and the Internet of Things, IGI Global, 2017.
- Fotheringham, A. S., Brunson, C. ve Charlton, M., Geographically weighted regression: the analysis of spatially varying relationships, John Wiley & Sons, 2003.
- Huang, B., Wu, B., Barry, M., Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices, International Journal of Geographical Information Science, 24 (3), 383-401, 2010.
- Fotheringham, A. S., Crespo, R., Yao, J., Geographical and temporal weighted regression (GTWR), Geographical Analysis, 47 (4), 431-452, 2015.
- Ma, X., Zhang, J., Ding, C., Wang, Y., 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, 2018.
- Li, Z., Fotheringham, A. S., Li, W., Oshan, T., 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, 2019.
8. Taşyürek, M., Çelik, M., RNN-GWR: A geographically weighted regression approach for frequently updated data, *Neurocomputing*, 399, 258-270, 2020.
  9. Li, Z., Fotheringham, A. S., Computational improvements to multi-scale geographically weighted regression, *International Journal of Geographical Information Science*, 1-20, 2020.
  10. Oshan, T. M., Li, Z., Kang, W., Wolf, L. J., Fotheringham, A. S., 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, 2019.
  11. Guo, Y., Tang, Q., Gong, D. Y., Zhang, Z., Estimating ground-level PM<sub>2.5</sub> concentrations in Beijing using a satellite-based geographically and temporally weighted regression model, *Remote Sensing of Environment*, 198, 140-149, 2017.
  12. He, Q., Huang, B., Satellite-based high-resolution PM<sub>2.5</sub> estimation over the Beijing-Tianjin-Hebei region of China using an improved geographically and temporally weighted regression model, *Environmental Pollution*, 236, 1027-1037, 2018.
  13. Leong, Y. Y., Yue, J. C., A modification to geographically weighted regression, *International Journal of Health Geographics*, 16 (1), 11, 2017.
  14. Zou, B., Pu, Q., Bilal, M., Weng, Q., Zhai, L., Nichol, J. E., High-resolution satellite mapping of fine particulates based on geographically weighted regression, *IEEE Geoscience and Remote Sensing Letters*, 13 (4), 495-499, 2016.
  15. Chu, H. J., Kong, S. J., Chang, C. H., 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, 2018.
  16. Harris, R., Singleton, A., Grose, D., Brunson, C., Longley, P., Grid enabling geographically weighted regression: a case study of participation in higher education in England, *Transactions in GIS*, 14 (1), 43-61, 2010.
  17. Tran, H. T., Nguyen, H. T., Tran, V. T., Large-scale geographically weighted regression on Spark, In 2016 Eighth International Conference on Knowledge and Systems Engineering (KSE), IEEE, 127-132, 2016.
  18. Pozdnoukhov, A., Kaiser, C., Scalable local regression for spatial analytics, In Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 361-364, 2011.
  19. Stergiou, C., Psannis, K. E., Kim, B. G., Gupta, B., Secure integration of IoT and cloud computing. *Future Generation Computer Systems*, 78, 964-975, 2018.
  20. Yıldırım G., Tatar Y., Remote user supported IoT-WSN Laboratory and testbed platform: FiratWSN, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 34 (4), 1831-1846, 2019.
  21. Shekhar, S., Vatsavai, R.R., Çelik, M., Spatial and spatiotemporal data mining: Recent advances, *Data Mining: Next Generation Challenges and Future Directions*, 1-34, 2009.
  22. Çelik M., Dadaser-Çelik, F., Dokuz A.S., Discovery of hydrometeorological patterns, *Turkish Journal Of Electrical Engineering and Computer Sciences*, 22, 840-857, 2014.
  23. Türker A., Göleç A., Aktepe A., Ersöz S., İpek M., Çağıl G., A real-time system design using data mining for estimation of delayed orders and application, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 35 (2), 709-724, 2019.
  24. Tan, P. N., Steinbach, M. Karpatne, A. ve Kumar, V., *Introduction to data mining*. Pearson, 2018.
  25. Çelik, M., Kazar, B. M., Shekhar, S. ve Boley, D., Parameter Estimation for the Spatial Autoregression Model: A Rigorous Approach, *Proceedings of the Second NASA Data Mining Workshop: Issues and Applications in Earth Science*, Pasadena, A.B.D., 2006.
  26. Kazar, B. M. ve Çelik, M., Spatial autoregression (SAR) model: Parameter estimation techniques, *Springer Briefs in Computer Science*, ISBN:978-1461418412, Springer, March 2012.
  27. Başar G., Mıstıkoğlu S., Determination of the optimum welding parameters for ultimate tensile strength and hardness in friction stir welding of Cu/Al plates using Taguchi method, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 34 (3), 1595-1608, 2019.
  28. Lu, B., Charlton, M., Brunson, C., Harris, P., The Minkowski approach for choosing the distance metric in geographically weighted regression, *International Journal of Geographical Information Science*, 30 (2), 351-368, 2016.
  29. Fotheringham, A. S., Oshan, T. M., Geographically weighted regression and multicollinearity: dispelling the myth, *Journal of Geographical Systems*, 18 (4), 303-329, 2016.
  30. Lu, B., Yang, W., Ge, Y., Harris, P., Improvements to the calibration of a geographically weighted regression with parameter-specific distance metrics and bandwidths, *Computers, Environment and Urban Systems*, 71, 41-57, 2018.
  31. Fotheringham, A. S., Yang, W., Kang, W., Multiscale geographically weighted regression (MGWR), *Annals of the American Association of Geographers*, 107 (6), 1247-1265, 2017.
  32. Da Silva, A. R. , De Oliveira Lima, A., Geographically weighted beta regression, *Spatial Statistics*, 21, 279-303, 2017.
  33. Gollini, I., Lu, B., Charlton, M., Brunson, C., Harris, P., GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models, *arXiv preprint arXiv:1306.0413*, 2013.
  34. Guo, L., Ma, Z., Zhang, L., Comparison of bandwidth selection in application of geographically weighted

- regression: a case study, *Canadian Journal of Forest Research*, 38 (9), 2526-2534, 2008.
35. Hadayeghi, A., Shalaby, A. S., Persaud, B. N., Development of planning level transportation safety tools using Geographically Weighted Poisson Regression, *Accident Analysis & Prevention*, 42 (2), 676-688, 2010.
  36. Bidanset, P. E., Lombard, J. R., 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*, 11 (3), 5-14, 2014.
  37. Kauermann, G., Opsomer, J. D., Generalized cross-validation for bandwidth selection of backfitting estimates in generalized additive models, *Journal of Computational and Graphical Statistics*, 13 (1), 66-89, 2004.
  38. Zougab, N., Adjabi, S., Kokonendji, C. C., Bayesian estimation of adaptive bandwidth matrices in multivariate kernel density estimation, *Computational Statistics & Data Analysis*, 75, 28-38, 2014.
  39. Taşyürek, M., Çelik, M., Hava Sıcaklık Değerlerinin Coğrafi ve Rakım Ağırlıklı Regresyon Yöntemi ile Tahmin Edilmesi, *Avrupa Bilim ve Teknoloji Dergisi*, 81-86, 2020.