



Hava Sıcaklık Değerlerinin Coğrafi ve Rakım Ağırlıklı Regresyon Yöntemi ile Tahmin Edilmesi

Murat Taşyürek^{1,2*}, Mete Çelik²

¹ Kayseri Büyükşehir Belediyesi, Bilgi İşlem Dairesi Başkanlığı, Yazılım Şube Müdürlüğü, Kayseri, Türkiye (ORCID: 0000-0001-5623-8577)

² Erciyes Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Kayseri, Türkiye (ORCID: 0000-0002-1488-1502)

(Bu yayın 26-27 Haziran 2020 tarihinde HORA-2020 kongresinde sözlü olarak sunulmuştur.)

(DOI: 10.31590/ejosat.779074)

ATIF/REFERENCE: Taşyürek, M. & Çelik, M. (2020). Hava Sıcaklık Değerlerinin Coğrafi ve Rakım Ağırlıklı Regresyon Yöntemi ile Tahmin Edilmesi. *Avrupa Bilim ve Teknoloji Dergisi*, (Special Issue), 81-86.

Öz

Nesnelerin interneti sayesinde farklı alanlarda çok fazla sayıda ve çeşitte mekânsal verinin toplanması konum temelli analizlere olan ilgiyi arttırmıştır. Yaygın olarak kullanılan konum temelli analiz yöntemlerinden birisi olan coğrafi ağırlıklı regresyon (Geographically Weighted Regression - GWR), coğrafya üzerindeki değişen ilişkileri modelleyen bir yerel mekânsal regresyon tekniğidir. Coğrafi ve rakım ağırlıklı regresyon (Geographically and Altitudinal Weighted Regression - GAWR) ise GWR yaklaşımına rakım (yükseklik) ilişkilerin eklenmesiyle geliştirilen bir yaklaşımdır. GAWR, GWR modelinden farklı olarak, verideki mekân (yatay) ve rakım (dikey) bilgilerini gözönüne alarak modelleme yapar ve bu nedenle, hem mekân ve hem de rakım ilişkilerinin olduğu veri kümelerinde (meteorolojik veri kümeleri gibi) başarılı sonuçlar verebilmektedir. Literatürde GWR ile sıcaklık tahmin çalışmaları yapılmıştır. Bu çalışmada, sıcaklık tahmini için GAWR algoritmasının kullanılması önerilmiştir. Çalışmada Meteoroloji Genel Müdürlüğü'nden (MGM) alınan veriler kullanılmıştır ve bir noktanın sıcaklığının tahmini için o noktanın rakım, basınç ve nem bilgisi kullanılmıştır. Sonuçlar GAWR algoritmasının GWR algoritmasına göre daha doğru sonuçlar ürettiğini göstermiştir.

Anahtar Kelimeler: Coğrafi ve rakım ağırlıklı regresyon, Dikey ağırlık, Hava sıcaklık tahmini, Veri madenciliği.

Estimation of Air Temperature Using Geographically and Altitudinal Weighted Regression

Abstract

The interest in location-based analysis has increased with the collection of data in different application domains using the internet of things technologies. Geographically Weighted Regression (GWR), one of the commonly used location-based analysis methods, is a local spatial regression technique that models changing relationships on geography. Geographically and Altitudinal Weighted Regression (GAWR) is an approach developed by adding altitude relationships to the GWR approach. The GAWR model does modeling by considering the spatial (horizontal) and altitude (vertical) information in the data, and therefore it can produce successful results in datasets (such as meteorological datasets) that have spatial and altitude relationships. In the literature, temperature estimation studies have been done by using GWR. This study proposes to use the GAWR algorithm to estimate the air temperature. In the study, the meteorological dataset, taken from the General Directorate of Meteorology (MGM), was used and altitude, pressure, and humidity information was used to estimate temperature value of a location. Results show that GAWR outperforms GWR algorithm.

Keywords: Geographically and altitudinal weighted regression, Altitudinal weight, Estimation of Air Temperature, Data mining

1. Giriş

Nesnelerin interneti sayesinde farklı alanlarda birçok veri toplanması konum temelli analizlere olan ilgiyi artırmaktadır. Hava kalitesini, hava kirliliğini ve iklim bilgilerini oluşturan sıcaklık, basınç, nem, rüzgar yönü/hızı, ozon, karbondioksit, karbonmonoksit,

* Sorumlu Yazar: Erciyes Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Kayseri, Türkiye, ORCID: 0000-0001-5623-8577, m_tasyurek@hotmail.com

toz konsantrasyonu, gürültü, güneş radyasyon miktarı ve yağış miktarı gibi veriler nesnelerin interneti ile toplanan verilere örnek gösterilebilir [Perera ve ark., 2014; Prasad ve ark., 2017; Novo, 2018; Stergiou ve ark., 2018; Yıldırım ve ark., 2019]. Elde edilen bu veriler veri madenciliği yöntemleri [Celik ve ark., 2006; Dadaser-Celik ve ark., 2012; Kazar ve ark., 2012; Shekhar ve ark., 2009] ile analiz edildiğinde faydalı bilgiler ortaya çıkmaktadır. Yaygın olarak tercih edilen veri madenciliği yöntemlerinden biri olan ve her şey her şey ile ilişkilidir ama yakın şeyler uzak şeylerden daha ilişkilidir felsefesi üzerine kurulu olan coğrafi ağırlıklı regresyon (GWR) yöntemi coğrafya üzerindeki değişen ilişkileri modellemek için kullanılan yerel mekânsal regresyon tekniğidir [Fotheringham ve ark., 2003; Tasyurek ve Celik., 2020; Tobler, 1970]. Coğrafi ve rakım ağırlıklı regresyon (GAWR) yöntemi ise klasik GWR yaklaşımına dikey (yükseklik veya rakım) ilişkilerin eklenmesiyle geliştirilen bir yaklaşımdır [Propastin, 2012]. GAWR modeli, GWR modelinden farklı olarak, ilişkileri mekân (yatay) ve rakım (dikey) bilgilerini gözönüne alarak modeller ve ağırlık matrislerine hem mekânsal komşuluk bilgilerini hem de rakım komşuluk bilgilerini entegre eder. GAWR yaklaşımı veriler arasında hem mekânsal ilişkinin hem de yükseklik ilişkisinin olduğu veri kümesinde GWR yaklaşımına göre çok daha iyi sonuçlar üretebilmektedir.

Meteoroloji veri kümesi veriler arasında yükseklik ve mekânsal ilişkileri içeren bir veri kümesidir. Tasyurek ve Celik GWR yaklaşımı ile meteoroloji veri kümesini kullanarak sıcaklık tahmini yapmış ve başarılı sonuçlar elde etmiştir [Tasyurek ve Celik, 2020]. Sıcaklık parametresi yaşam koşullarını etkileyen ve sıcaklık değerine göre özel olarak önlem alınması gereken bir değer olduğundan doğru olarak tespit edilmesi/ölçülmesi önemlidir. Bununla birlikte bütün istasyonlara ya da konumlara sıcaklık sensörü takmak maliyetli bir işlemdir. Propastin [Propastin, 2012] tarafından önerilen GAWR yaklaşımının hem mekânsal ve hem de yükseklik ilişkilerin olduğu veri kümesinde GWR yaklaşımına göre çok daha iyi sonuçlar ürettiği ispatlanmıştır ve GAWR'yi biyokütle tahmininde kullanmışlardır.

Bu çalışmada, hava sıcaklığının tahmini için GAWR algoritmasının kullanılması önerilmiştir. Çalışmada Meteoroloji Genel Müdürlüğü'nden (MGM) alınan 1746 adet saatlik ölçüm yapan istasyon bilgilerini ve gerçek ölçüm verilerini içeren meteoroloji veri kümesi kullanılmıştır. Önerilen modelde çoklu regresyon denklemi kullanılarak sıcaklık değeri tahmin edilecek noktanın rakım (yükseklik), basınç ve nem bilgisi kullanılarak o noktanın sıcaklık bilgisi GAWR yaklaşımı ile tahmin edilmektedir. GAWR algoritmasının sonuçları GWR algoritmasının sonuçları ile karşılaştırılmıştır. Deneysel sonuçlar GAWR algoritmasının meteorolojik veriler kullanarak sıcaklık değeri tahmin etmede klasik GWR algoritmasına göre daha doğru sonuçlar ürettiğini göstermiştir. Bu çalışmanın bundan sonraki bölümlerinde GWR ve GAWR yöntemleri tanıtılmış, hava sıcaklık değerlerinin GAWR yaklaşımı ile tahmin edilmesi tartışılmış ve sonuçlar bölümünde ise yapılan deney sonuçları paylaşılmıştır.

2. Materyal ve Metod

Bu bölümde öncelikle GWR ve GAWR modelleri tanıtılmış ve daha sonra sıcaklık değerlerinin tahmin edilmesi tartışılmıştır.

2.1. Coğrafi Ağırlıklı Regresyon (GWR)

GWR yöntemi coğrafya üzerindeki değişen ilişkileri modellemek için kullanılan yerel mekânsal regresyon tekniğidir [Fotheringham ve ark., 2003]. Diğer bir ifade ile GWR coğrafya üzerinde konumu ve özellikleri bilinen referans noktaları temel olarak konumu bilinen diğer noktalar için regresyon yöntemi ile tahmini değerler üreten bir mekânsal veri madenciliği yöntemidir. GWR, klasik çoklu regresyon modelini temel alır. Klasik regresyon modelinden farklı olarak, coğrafi GWR modelinde katsayılar sabit değildir. Her mekânsal noktanın kendisine ait katsayıları oluşturulmaktadır [Fotheringham ve ark., 2003; Lu ve ark., 2017; Lu ve ark., 2018; Fotheringham ve ark., 2017]. Coğrafi ağırlıklı regresyon modeli Eşitlik 1'de gösterildiği şekilde ifade edilir.

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

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

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

$$\beta(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) Y \quad (2)$$

Eşitlikte, X bağımsız değişkenlerin matrisidir ve $m + 1$ sütundan oluşur, $Y = (Y_1, \dots, Y_n)^T$ bağımlı değişken matrisini göstermekte ve $W(i)$ ise w_{ij} değerlerinden oluşan diagonal matristir ve Eşitlik 3'de gösterildiği gibidir [Fotheringham ve ark., 2003].

$$W(i) = \begin{bmatrix} w_{i1} & \dots & 0 \\ \vdots & w_{i2} & 0 \\ 0 & 0 & \dots \\ 0 & \dots & w_{in} \end{bmatrix} \quad (3)$$

Eşitlikte, 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 [Gollini ve ark., 2013; Guo ve ark., 2008]. Yaygın olarak Gaussian ve Bi-Square çekirdek fonksiyonları kullanılır. Gaussian çekirdek fonksiyonu ile i . noktası ve j . noktası arasındaki komşuluk oranı w_{ij} Eşitlik 4'te gösterildiği şekilde hesaplanır:

$$w_{ij} = \exp \left[-\frac{1}{2} \left(\frac{d_{ij}}{bw} \right)^2 \right] \quad (4)$$

Eşitlikte, bw bant genişliği değerini ve d_{ij} ise i regresyon noktası ile j referans nokta arasındaki mesafeyi göstermektedir. x ve y nokta koordinatlarını göstermek üzere d_{ij} mesafesi genellikle Öklid mesafesi Eşitlik 5'de gösterildiği şekilde hesaplanır:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (5)$$

Eşitlikte, bw bant genişliği parametresinin değeri GWR modelinde bütün bir model için sabit de olabilir ya da bölgedeki nokta yoğunluğuna göre değişken de olabilir. Veri kümesi için en uygun bw değerini bulmak için Cross-validation (CV), Generalised cross-validation (GCV), Akaike Information Critation (AIC), Bayesian Information Critation (BIC) yöntemleri kullanılabilir [Cho ve ark., 2010; Kauermann ve ark., 2004; Zougab ve ark., 2004]. AIC yönteminin daha doğru sonuç verdiği tespit edilmiştir [Fotheringham ve ark., 2003; Lu ve ark., 2017] ve Eşitlik 6'da gösterildiği şekilde hesaplanır.

$$AIC_c = 2nl n(\hat{\sigma}) + n \ln(2\pi) + n \left\{ \frac{n + tr(S)}{n - 2 - tr(S)} \right\} \quad (6)$$

Eşitlikte, n gözlem sayısını, $\hat{\sigma}$ hata teriminin tahmini standart sapmasını göstermekte, $tr(S)$ şapka matrisi olarak bilinen S matrisinin [Brook ve ark., 2018] izini göstermektedir.

2.2. Coğrafi ve Dikey Ağırlıklı Regresyon

Klasik GWR yaklaşımı mekânsal ilişkileri bulmada çok iyidir ve iki boyutlu uzayda çalışır. Fakat karmaşık arazi koşullarının olduğu veya yükseklik verisinin önemli bir etken olduğu durumlarda iki boyutlu uzayda incelemek yetersiz olabilmektedir ve üçüncü boyuta ihtiyaç duyulmaktadır. Propastin [Propastin, 2012] bu sorunun üstesinden gelmek için 3. boyutu (rakım veya yükseklik) da dikkate alarak coğrafi ve dikey ağırlıklı regresyon (GAWR) yaklaşımını önermiştir. GAWR yaklaşımı klasik GWR yönteminin ağırlık matrisine üçüncü boyut olarak rakımı ekleyerek oluşturulmuştur. GAWR yaklaşımı Eşitlik 7'de gösterildiği şekilde ifade edilebilir.

$$y_i(u_i, v_i, z_i) = \beta_{i0}(u_i, v_i, z_i) + \sum_{k=1}^m \beta_{ik}(u_i, v_i, z_i)x_{ik}(u_i, v_i, z_i) + \varepsilon_i(u_i, v_i, z_i) \quad (7)$$

GWR formülünden farklı olarak formüle rakım (yükseklik) bilgisi eklenmiştir ve z yükseklik bilgisini göstermektedir. β_{ik} ($k = 0, \dots, m$) regresyon katsayısı GWR yaklaşımında olduğu gibi Eşitlik 8'de gösterildiği şekilde hesaplanır.

$$\beta(u_i, v_i, z_i) = (X^T W(u_i, v_i, z_i) X)^{-1} X^T W(u_i, v_i, z_i) Y \quad (8)$$

$W(i) = W(u_i, v_i, z_i)$ ağırlık matrisi klasik GWR olduğu gibi w_{ij} değerlerinden oluşur fakat burada w_{ij} komşuluk oranı 2 boyutlu uzayda değil 3 boyutlu uzayda ölçülüdür. x , y nokta koordinatlarını, z yükseklik (rakım) bilgisini ve a ise normalleştirme katsayısını göstermek üzere d_{ij} mesafesi Eşitlik 9'de gösterildiği şekilde hesaplanır:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + a(z_i - z_j)^2} \quad (9)$$

GAWR yaklaşımında komşuluk oranlarının 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 a parametresi eklenmiştir [Propastin, 2012]. Bu nedenle, GAWR yaklaşımını uygulamadan önce a parametresinin optimize edilmesi gerekmektedir. Propastin [Propastin, 2012] a parametresini bütün ihtimalleri deneyerek bulmayı önermiştir.

2.3. Hava Sıcaklık Değerlerinin Tahmin Edilmesi

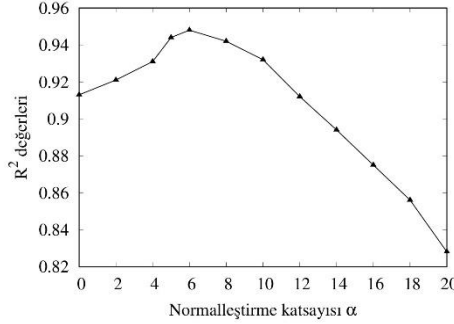
Bu çalışmanın amacı hava sıcaklık değerini GAWR algoritması kullanarak tahmin etmektir. Hedeflenen amaç bir noktanın ölçülen nem, basınç değeri ve o noktanın rakım değerini kullanarak o noktanın sıcaklık değerini tahmin etmektir. Sıcaklık değerini tahmin etmek için klasik GWR yaklaşımı aşağıdaki çoklu regresyon denklemini kullanır (Eşitlik 10).

$$Sicaklik_i = \beta_{0i} + \beta_{1i}Rakim_i + \beta_{2i}Nem_i + \beta_{3i}Basinc_i \quad (10)$$

Eşitlikte, i regresyon noktasını ve β_s regresyon katsayılarını temsil eder. GAWR yaklaşımında rakım değeri komşuluk oranında kullanıldığından dolayı çoklu regresyon denkleminde kullanılmamıştır ve GAWR için çoklu regresyon denklemi Eşitlik 11'de gösterildiği şekilde ifade edilir.

$$Sicaklik_i = \beta_{0i} + \beta_{1i}Nem_i + \beta_{2i}Basinc_i \quad (11)$$

GWR yaklaşımlarında öncelikle çekirdek fonksiyonu ve bant genişliği değeri belirlenmelidir. Bu çalışmada GWR yaklaşımlarında yaygın olarak kullanılan Gaussian çekirdek fonksiyonu kullanılmıştır. En uygun mekânsal bant genişliği değerini bulmak için sık kullanılan AIC yöntemi tercih edilmiştir [Fotheringham ve ark., 2003; Lu ve ark., 2017]. GWR yaklaşımı için AIC yöntemi ile en uygun mekânsal bant genişliği 82 km (Öklid mesafesi) olarak bulunmuştur. GAWR için en uygun mekân-zamansal bant genişliği 70 normalleştirme katsayısı olan $a = 6.05$ olduğunda bulunmuştur. a değerinin uyum grafiğini R^2 değerine göre göstermiştir. R^2 değeri 1'e ne kadar yakınsa model o kadar iyi temsil edilmiştir. Bu çalışmada kullanılan veri kümesine göre a değerlerinin oluşturduğu R^2 grafiği Şekil 1'te gösterilmiştir.

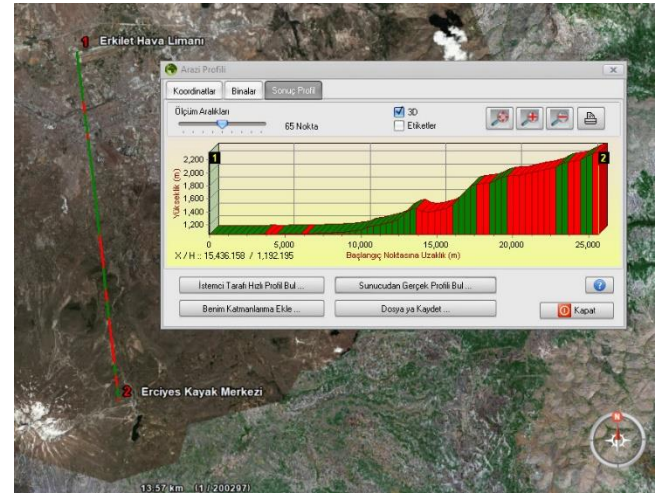


Şekil 1 GAWR Modeli İçin Parametre Seçimi (a)

3. Deneysel Değerlendirme

Bu çalışmada Meteoroloji Genel Müdürlüğü'nden (MGM) alınan 1746 adet saatlik ölçüm yapan istasyon bilgilerini ve gerçek ölçüm verilerini içeren meteoroloji veri kümesi kullanılmıştır. MGM istasyonlarından 1200 adet ölçüm noktası referans noktası 400 adet nokta ise regresyon noktası olarak seçilmiştir. MGM veri kümesi saatlik sıcaklık bilgisi, nem bilgisini, basınç bilgisini ve istasyona ait rakım, konum ve istasyon id bilgilerini içermektedir. Yapılan araştırmalara göre rakım değerinde 200 metre değişim olunca sıcaklıkta da yaklaşık 1 derece değişim olabilmektedir [Pabuççu, 2016]. Bu kural her zaman geçerli olmasa da sonuç olarak rakım sıcaklığı etkileyen bir faktördür. GAWR yaklaşımının çalışması için konum içerisinde rakım (yükseklik) değeri (z) de bulunmalıdır. Şekil 2'de Kayseri iline ait MGM istasyon bilgileri gösterilmiştir. Şekil 2'de görüldüğü üzere 17195 numaralı istasyonun (Kayseri Erkilet Havalimanı) rakım değeri 1053, 18149 numaralı istasyonun (Erciyes Kayak Merkezi) rakım değeri ise 2210'dur. Kayseri Büyükşehir Belediyesi [Kayseri Büyükşehir Belediyesi, 2020] CitySurf programı kullanarak yapılan ölçümde bu iki nokta arasındaki mekânsal mesafe (öklid mesafesi) 25 kilometredir (Şekil 3).

ID	İSTASYON	RAKIM	İL	İLÇE
17195	KAYSERİ ERKİLET HAVALİMANI	1053	Kayseri	Kocasinan
18148	KOCASINAN/YAMULA BARAJI	1075	Kayseri	Kocasinan
17196	KAYSERİ BÖLGE	1094	Kayseri	Melikgazi
18461	YAHYALI	1125	Kayseri	Yahyalı
18457	İNCESU	1130	Kayseri	İncesu
18207	YEŞİLHISAR	1141	Kayseri	Yeşilhisar
18873	SARIOĞLAN/TUZLA GÖLÜ	1166	Kayseri	Sarıoğlan
17836	DEVELİ	1204	Kayseri	Develi
18459	SARIOĞLAN	1220	Kayseri	Sarıoğlan
18923	KOCASINAN/HİMMETDEDE KÖYÜ	1234	Kayseri	Kocasinan
18455	BÜNYAN	1335	Kayseri	Bünyan
18456	FELAHİYE	1370	Kayseri	Felahiye
18458	ÖZVATAN	1375	Kayseri	Özvatani(çukur)
18454	AKKIŞLA	1390	Kayseri	Akkışla
17837	TOMARZA	1402	Kayseri	Tomarza
18661	YAHYALI/AVLAĞA KÖYÜ	1469	Kayseri	Yahyalı
18662	AKKIŞLA/ORTAKÖY KÖYÜ	1530	Kayseri	Akkışla
17802	KAYSERİ/PINARBAŞI	1542	Kayseri	Pınarbaşı
18874	TALAS/KAMBER KÖYÜ	1592	Kayseri	Talas
17840	SARIZ	1599	Kayseri	Sarız
18663	DEVELİ/KÜÇÜKKÜNYE KÖYÜ	1619	Kayseri	Develi
18665	TOMARZA/GÜZELSU KÖYÜ	1733	Kayseri	Tomarza
18460	TALAS/ALI DAĞI	1745	Kayseri	Talas
18875	SARIZ/YEDİOLUK KÖYÜ	1833	Kayseri	Sarız
18664	PINARBAŞI/YUKARIKIZILÇEVRIK KÖYÜ	1845	Kayseri	Pınarbaşı
18660	YAHYALI/DEREBAĞ KÖYÜ	1970	Kayseri	Yahyalı
18149	MELİKGAZI/ERCIYES KAYAK MERKEZİ	2210	Kayseri	Kocasinan



Şekil 3 İki İstasyon Noktası Arasındaki Profil Görüntüsü

Şekil 2 Kayseri Meteoroloji İstasyonları Rakım Değerleri

Şekil 3'de gösterildiği gibi Kayseri ili incelendiğinde 1 numaralı konum (Erkilet Havalimanı) regresyon noktası olduğu durumda GWR yaklaşımında 2 numaralı referans noktasının (Erciyes Kayak Merkezi) etkisi sadece mekânsal mesafe olmaktadır fakat 1157 metre rakım farkı yaklaşık olarak 6°C fark oluşturmaktadır ve bunu klasik GWR ile belirlemek zor olabilmektedir. Meteorolojik veri kümesini karmaşık arazi koşullarının olduğu veya yükseklik verisinin önemli bir etken olduğu durumlarda iki boyutlu uzayda incelemek yetersiz olacaktır ve üçüncü boyuta ihtiyaç duyulmaktadır ve bu nedenle, bu çalışmada, GWR yaklaşımı yerine GAWR yaklaşımı kullanılarak sıcaklık tahmini önerilmiştir.

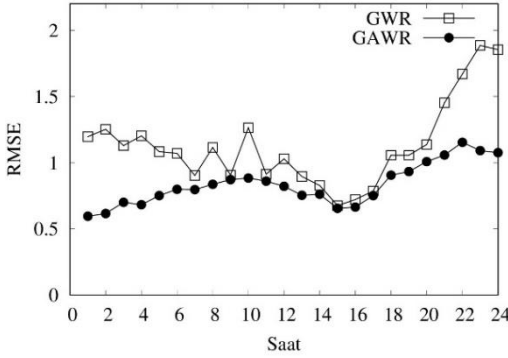
Bu çalışmada yapılan deneyler Intel Core i7-9750H CPU 2.60 GHz 16 GB Ram 500 GB SSD disk özellikleri ve Windows 10 işletim sistemine sahip dizüstü bilgisayar üzerinde yapılmıştır.

3.1. Deneysel Sonuçlar

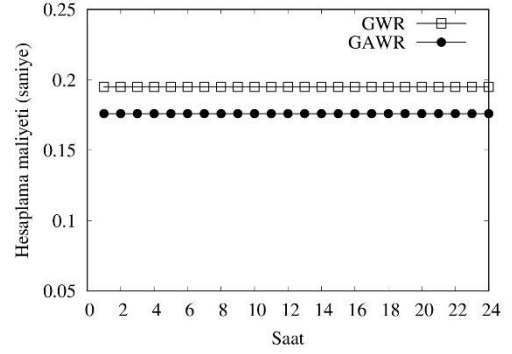
Bu bölümde, klasik GWR ve GAWR yaklaşımları tarafından tahmin edilen sıcaklık değeri ile gerçek sıcaklık değerlerinin karşılaştırması yapılmıştır. Önerilen modellerin uygunluğu hata karelerinin ortalamalarının karekökü (RMSE) değerlerine göre karşılaştırılmıştır. RMSE (eşitlik 12) model tarafından tahmin edilen değer (\hat{y}_i) ile gerçekte ölçülen değer (y_i) arasındaki farkı ifade eden ve önerilen yaklaşımın kalitesini gösteren bir metriktir [Taşyürek ve Çelik, 2020]. 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.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (12)$$

Bu deneyde, algoritmaların doğruluğu her bir saat için incelenmiştir. Referans noktalarının ve regresyon noktalarının sayısı sırasıyla 1200 ve 400 olarak ayarlanmıştır. Ölçüm değeri bilinen gerçek istasyon noktaları regresyon noktası olarak alınmıştır. Şekil 4’de saat bazlı GWR ve GAWR yaklaşımlarının RMSE sonuçları gösterilmiştir. GAWR algoritması Şekil 4’de görüldüğü üzere GWR yaklaşımına göre çok daha doğru tahminler üretmektedir.



Şekil 4 GWR ve GAWR Yaklaşımlarının Saat Bazlı RMSE Sonuçları



Şekil 5 GWR ve GAWR Yaklaşımlarının Hesaplama Maliyeti

Klasik GWR ve GAWR yaklaşımlarını hesaplama maliyeti açısından incelendiğinde m parameter sayısını (bağımsız değişken sayısı) ve n ölçüm nokta sayısını göstermek üzere GWR yaklaşımlarının hesaplama maliyeti $O(m^2n^2)$ olarak ifade edilir [Li ve ark., 2019]. Klasik GWR yaklaşımı Eşitlik 10’u kullanır ve burada $m = 3$ değerini almaktadır, GAWR yaklaşımı ise Eşitlik 11’i kullanır ve GAWR yaklaşımı için $m = 2$ değerini almaktadır. Klasik GWR ve GAWR yaklaşımları hesaplama maliyeti saniye cinsinden Şekil 5’de gösterilmiştir. GAWR yaklaşımı m değerinin küçük olmasından dolayı hesaplama maliyeti daha düşük çıkmıştır.

4. Sonuçlar

Bu çalışmada, karmaşık arazi koşullarının olduğu veya rakım (yükseklik) verisinin önemli bir etken olduğu durumlarda GWR yaklaşımı yerine GAWR yaklaşımı kullanılarak hava sıcaklık değerlerinin tahmin edilmesi önerilmiştir. Çalışmada Meteoroloji Genel Müdürlüğü’nden (MGM) alınan 1746 adet saatlik ölçüm yapan istasyon bilgilerini ve gerçek ölçüm verilerini içeren meteoroloji veri kümesi kullanılmıştır. Önerilen modelde çoklu regresyon denklemi kullanılarak sıcaklık değeri tahmin edilecek noktanın rakım (yükseklik), basınç ve nem bilgisi kullanılarak o noktanın sıcaklık bilgisi GAWR yaklaşımı ile tahmin edilmektedir. GAWR algoritmasının sonuçları GWR algoritmasının sonuçları ile karşılaştırılmıştır. Deneysel sonuçlar GAWR algoritmasının meteorolojik veriler kullanarak sıcaklık değeri tahmin etmede klasik GWR algoritmasına göre daha doğru sonuçlar ürettiğini göstermiştir. Gelecek çalışma planlarında GAWR yaklaşımı ile büyük veri kümeleri üzerinde çalışma planlanmaktadır.

Kaynakça

- Brook, R. J., & Arnold, G. C. (2018). *Applied regression analysis and experimental design*. Routledge.
- Celik, M., Kazar, B. M., Shekhar, S., & Boley, D. (2006). Parameter Estimation for the Spatial Autoregression Model: A Rigorous Approach. In *Proceedings of the 2nd NASA Data Mining Workshop: Issues and Applications in Earth Science with the 38th Symposium on the Interface of Computing Science, Statistics and Applications*.
- Cho, S. H., Lambert, D. M., & Chen, Z. (2010). Geographically weighted regression bandwidth selection and spatial autocorrelation: an empirical example using Chinese agriculture data. *Applied Economics Letters*, 17(8), 767-772.

- Dadaser-Celik F., Celik M., Dokuz A. (2012). Associations between stream flow and climatic parameters at Kızılırmak River Basin in Turkey. *Global Nest Journal*, 14, 354-361.
- Fotheringham, A. S., Brunson, C., & Charlton, M. (2003). Geographically weighted regression: the analysis of spatially varying relationships, John Wiley & Sons.
- Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale geographically weighted regression (MGWR). *Annals of the American Association of Geographers*, 107(6), 1247-1265.
- Gollini, I., Lu, B., Charlton, M., Brunson, C., & Harris, P. (2013). GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models. *arXiv preprint arXiv:1306.0413*.
- Guo, L., Ma, Z., & 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.
- Kauermann, G., & Opsomer, J. D. (2004). Generalized cross-validation for bandwidth selection of backfitting estimates in generalized additive models. *Journal of Computational and Graphical Statistics*, 13(1), 66-89.
- Kayseri Büyükşehir Belediyesi. (2020, March 10). Retrieved from <https://www.kayseri.bel.tr>.
- Kazar, B. M., & Celik, M. (2012). Spatial Autoregression (SAR) Model: Parameter Estimation Techniques, Springer Briefs in computer Science, ISBN:978-1461418412, Springer.
- Li, Z., Fotheringham, A. S., Li, W., & 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.
- Lu, B., Brunson, C., Charlton, M., & Harris, P. (2017). Geographically weighted regression with parameter-specific distance metrics. *International Journal of Geographical Information Science*, 31(5), 982-998.
- Lu, B., Yang, W., Ge, Y., & 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.
- Novo, O. (2018). Blockchain meets IoT: An architecture for scalable access management in IoT. *IEEE Internet of Things Journal*, 5(2), 1184-1195.
- Pabuçcu, A. (2016). Fen bilgisi öğretmen adaylarının gaz basıncıyla ilgili bilgilerini günlük hayatla ilişkilendirebilme seviyeleri. *Türkiye Kimya Derneği Dergisi Kısım C: Kimya Eğitimi*, 1(2), 1-24.
- Perera, C., Zaslavsky, A., Christen, P., & 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.
- Prasad, A. V. (2017). Exploring the Convergence of Big Data and the Internet of Things, IGI Global.
- 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.
- Shekhar, S., Vatsavai, R.R., & Celik, M. (2009). Spatial and Spatiotemporal Data Mining: Recent Advances, as a chapter of Next Generation of Data Mining, H. Kargupta, J. Han, P.S. Yu, R. Motwani, and Vipin Kumar (Eds.), ISBN:978-1-4200-8586-0, CRC Press
- Stergiou, C., Psannis, K. E., Kim, B. G., & Gupta, B. (2018). Secure integration of IoT and cloud computing. *Future Generation Computer Systems*, 78, 964-975.
- Tasyurek, M., & Celik, M. (2020). RNN-GWR: A geographically weighted regression approach for frequently updated data, *Neurocomputing*, 399, 258-270.
- Taşyürek, M., & Çelik, M. (2020). Akıllı Durak Sistemindeki Araç Seyahat Sürelerinin Birleşik Yapay Sinir Ağları Kullanarak Tahmini. *Avrupa Bilim ve Teknoloji Dergisi*, 72-79. DOI: 10.31590/ejosat.araconf10.
- Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography*, 46(sup1), 234-240.
- Yıldırım, G., & Tatar, Y. (2019). Uzak kullanıcı destekli bir IoT-WSN sanal laboratuvarı ve test platformu: FıratWSN. *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, 34(4), 1831-1846.
- Zougab, N., Adjabi, S., & Kokonendji, C. C. (2014). Bayesian estimation of adaptive bandwidth matrices in multivariate kernel density estimation. *Computational Statistics & Data Analysis*, 75, 28-38.