

Deniz yüzey sıcaklığının ARIMA yöntemiyle modellenmesi ve gelecek tahmini: Zonguldak ve Bartın uygulaması

Cemal ERDEM^{1*}

Zafer ASLAN²

Geliş tarihi / Received: 06.01.2023

Düzeltilerek geliş tarihi / Received in revised form: 11.03.2023

Kabul tarihi / Accepted: 13.03.2023

DOI: 10.17932/IAU.ABMYOD.2006.005/abmyod_v18i67004

Özet

Deniz yüzey sıcaklığı (DYS), okyanus ve atmosferik sistemlerin dinamiklerini anlamak ve gelecekteki iklim trendlerini tahmin etmek için kritik bir parametredir. Bu çalışmada, Avrupa Uzay Ajansı İklim Değişikliği Girişimi'nden, özellikle Deniz Yüzey Sıcaklığı İklim Değişikliği Girişimi (DYS IDG) projesinden elde edilen veriler kullanılarak, Zonguldak ve Bartın illerindeki DYS'nin ARIMA (otoregresif hareketli ortalamalar) yöntemini kullanarak modelliyoruz. Veri seti (boylam 31.25 ve enlem 40.95) 1981'den 2022'ye kadar olan 40 yıllık bir dönemi kapsamakta ve DYS trendleri ile mevsimsel varyasyonların bir değerlendirmesini içermektedir. Sonuçlar, çalışma dönemi boyunca DYS'de tutarlı bir artış göstermektedir ve mse (ortalama kare hatası) 0,07'dir. Zonguldak ve Bartın illerindeki değişen DYS trendleri, balıkçılık ve turizm de dahil olmak üzere birçok endüstri ve sektör için önemli sonuçları vardır. Bu çalışmanın sonuçları bu

^{1*}İstanbul Aydın Üniversitesi, Lisansüstü Eğitim Enstitüsü, Bilgisayar Mühendisliği, 34295 Küçükçekmece /İstanbul, ORCID ID 0009-0001-2632-0729 cemalerdem@stu.aydin.edu.tr

² İstanbul Aydın Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, 34295, Küçükçekmece, İstanbul, ORCID ID 0000-0001-7707-7370 zaferaslan@aydin.edu.tr

alanlardaki karar verme süreçlerine ve iklim değişikliği uyum ve azaltma stratejileriyle ilgili politika kararlarına yardımcı olabilir. Bulgularımız ayrıca, DYS verilerinin alındığı DYS IDG projesi ve ARIMA yönteminin DYS verilerini modelleme etkinliği ve potansiyel sınırlamaları hakkında değerli bir iç görüş sağlar.

Anahtar Kelimeler: *Deniz Yüzey Sıcaklığı, İklim Değişikliği, Modelleme, ARIMA, Tahmin*

Modeling of sea surface temperature with ARIMA method and future prediction in Zonguldak and Bartın

Abstract

Sea surface temperature (SST) is a critical parameter in understanding the dynamics of oceanic and atmospheric systems and predicting future climate trends. In this study, we use data obtained from the European Space Agency Climate Change Initiative, specifically from the Sea Surface Temperature Climate Change Initiative (SST CCI) project, to model SST in the Zonguldak and Bartın provinces using the autoregressive integrated moving average (ARIMA) method. The dataset covers 40 years from 1981 to 2022 (longitude 31.25 and latitude 40.95) and includes an assessment of FMS trends and seasonal variations. The results show a slight but consistent increase in SST over the study period, with a mean squared error of 0.07. The changing SST trends in the Zonguldak and Bartın provinces have important implications for several industries and sectors, including fisheries and tourism. The results of this study can help inform decision-making in these areas as well as policy decisions pertaining to climate change adaptation and mitigation strategies. Our findings also provide valuable insights into the effectiveness of the ARIMA method for modeling

SST data and the potential limitations of the data obtained from the SST CCI project.

Keywords: *Sea Surface Temperature, Climate Change, Modeling, ARIMA, Future Prediction*

Giriş

Deniz Yüzey Sıcaklığı (DYS), okyanus ve atmosfer sistemlerinin dinamiklerini anlamak için temel bir parametredir ve gelecekteki iklim eğilimlerini tahmin etmede kritik bir rol oynamaktadır (Manabe, 1988). DYS, okyanusun üst tabakasının sıcaklığının bir ölçüsüdür ve güneş radyasyonu, rüzgâr ve akımlar dahil olmak üzere çeşitli faktörlerden etkilenir. DYS 'deki değişiklikler, okyanus dolaşımı, deniz ekosistemleri ve hava koşulları dahil olmak üzere çeşitli doğal süreçler üzerinde geniş kapsamlı etkilere sahip olabilir.

Ayrıca DYS, balıkçılık, turizm ve tarım gibi çeşitli insan faaliyetleri ve endüstrileriyle yakından bağlantılıdır. DYS'deki değişiklikler, balık popülasyonlarının dağılımını ve bolluğunu etkileyerek gelirleri ve gıda güvenliği için onlara bağımlı olanların geçim kaynaklarını etkileyebilir (Mohamed, 2022). Kıyı bölgelerindeki bir başka kilit sektörü olan turizm, plajların ve diğer kıyı cazibe merkezlerinin kalitesini etkileyebileceği için DYS'deki değişikliklere de duyarlıdır. Tarım, DYS'deki değişikliklerden de etkilenebilir, çünkü öngörülebilir yağış modellerine dayanan mahsullerin büyümesini ve verimini etkileyebilir.

DYS'nin hem doğal süreçler hem de insan faaliyetleri için önemi göz önüne alındığında, DYS eğilimlerinin doğru modellenmesi ve tahmini bilim adamları, politika yapımcılar ve paydaşlar için çok önemlidir. Bu çalışma, Türkiye'nin Karadeniz bölgesinde bulunan ve son birkaç on yılda DYS'de sıcaklık değişiklikleri yaşadığını tahmin ettiğimiz Zonguldak ve Bartın şehirlerine odaklanmaktadır. Avrupa Uzay Ajansı İklim Değişikliği

Girişimi'nden (European Space Agency Climate Change Initiative) elde edilen verileri analiz ederek, bu bölgedeki DYS eğilimlerinin değişmesinin anlaşılmasına katkıda bulunmayı ve iklim değişikliği uyarlaması ve hafifletme stratejileri ile ilgili politika kararları için değerli bilgiler sağlamayı amaçlıyoruz.

Materyal ve Metot

Bu çalışmada kullanılan veriler, Avrupa Uzay Ajansı İklim Değişikliği Girişimi'nden (ESA CCI) elde edilmiştir. (Merchant C. J., 2020) (Merchant C. E., 2019) Veri kümesi, Kelvin cinsinden günlük ortalama deniz yüzeyi sıcaklığı (DYS) değerlerinin yanı sıra santigrat derece cinsinden ortalama sıcaklığı, sıcaklık belirsizliğini ve deniz buzu ile kaplı okyanusun oranını içerir. Veri seti 1982'den 2022'ye kadar olan dönemi kapsamaktadır.

Bu çalışma için, veri setinden Kelvin cinsinden DYS değerlerini ve tarih bilgileri kullanıldı. Eksik veya hatalı değerleri kaldırmak için veri temizliği yapıldı. Gelecekteki DYS eğilimlerini modellemek ve tahmin etmek için temizlenmiş veri setine ARIMA (Oto-regresif Entegre Hareketli Ortalama) yöntemi uygulandı.

Mühendislik çalışmalarında performansın ve çalışma süresinin ulaşılan sonuçlar içerisinde önemi vardır. Uygun bulunan sonuçların performans ve çalışma zamanları ve bellek yönetimleri gibi kriterler dikkate alınarak seçilmelidir. (Cengiz, 2020) Broyden-Fletcher-Goldfarb-Shanno (BFGS) algoritması, kısıtlamasız optimizasyon problemlerini çözmek için yaygın olarak kullanılan bir optimizasyon algoritmasıdır.

BFGS algoritmasının ARIMA modeliyle kullanılmasının nedenlerinden biri, modelin parametrelerini tahmin etmek için kullanılabilmesidir. ARIMA modeli, geçmiş değerlerine dayanarak zaman serilerinin gelecekteki değerlerini tahmin etmek için kullanılan popüler bir zaman serisi modelidir. Modelin, oto-regresif (AR) katsayıları, hareketli ortalama

(MA) katsayıları ve fark alma sırası gibi verilerden tahmin edilmesi gereken birkaç parametresi vardır. ARIMA modelinin parametrelerini tahmin etmek için MLE (Maximum Likelihood Estimation) kullanılır. Bu verilerin olabilirlik fonksiyonunu maksimize eden parametreleri bulmak için sayısal optimizasyon yöntemi olan BFGS algoritması kullanılabilir.

BFGS algoritmasının diğer optimizasyon yöntemlerine göre çeşitli avantajları vardır. Bir avantajı, Hessian matrisini açıkça hesaplamadan yaklaşık olarak hesaplayabilen bir Quasi-Newton yöntemi olmasıdır. Hessian matrisinin hesaplanması çok büyük veya hesaplama maliyeti çok yüksek ise bu tür sorunlarda kullanışlı olabilir. Algoritma her zaman azalan değere doğru yön aldığından yerel minimuma yaklaşma olasılığı artmaktadır. (Bollapragada, 2018)

ARIMA ile LSTM'de zaman serileri için kullanılacak bir başka yöntemdir. Covid-19 için yapılan çalışmada günlük veriler ile yapılan çalışmada ARIMA ve LSTM yöntemleri ile çalışılmış olup ARIMA yönteminin daha performanslı çalıştığı sonucuna varılmıştır. Bu nedenle günlük verilerin benzerlik içermesi nedeniyle çalışmada ARIMA ile devam edilecektir. (Gülhan Toğa, 2021)

Son olarak, BFGS algoritmasının pratikte hızlı bir şekilde yakınsadığı gösterildi ve bu, onu bilgisayar mühendisliği ve finans da dahil olmak üzere birçok alanda optimizasyon problemlerini çözmek için popüler bir seçim haline getirdi.

ARIMA

Otoregresif Hareketli Ortalama anlamına gelen ARIMA, geçmiş gözlemlere dayalı olarak gelecekteki değerleri tahmin etmek için kullanılan popüler bir zaman serisi analiz tekniğidir. Ekonomik, finansal ve meteorolojik veriler dahil olmak üzere çok çeşitli zaman serisi verilerini analiz etmek ve modellemek için esnek ve güçlü bir araçtır. (Zhang, 2003)

ARIMA modeli üç ana bileşenden oluşur: otoregresif (AR) bileşeni, fark (I) bileşeni ve hareketli ortalama (MA) bileşeni. Otoregresif bileşen, gelecekteki değerleri tahmin etmek için geçmiş gözlemleri kullanırken, hareketli ortalama bileşeni, tahmini iyileştirmek için geçmiş tahminlerden gelen hataları kullanır. Fark bileşeni, verilerdeki herhangi bir trend veya mevsimsel etkiyi ortadan kaldırmak için kullanılır.

ARIMA, bir zaman serisi tahmin yöntemi olarak birçok avantaja sahiptir. İlk olarak, doğrusal olmayan ve durağan olmayan veriler de dahil olmak üzere geniş bir veri deseni yelpazesini ele alabilen esnek bir modele sahiptir. İkinci olarak, verilerden model parametrelerini tahmin edebilen bir parametrik modele sahiptir ve bu sayede veri kullanımını açısından daha verimlidir. Üçüncü olarak, karar verme süreci için önemli olan tahmin belirsizliğinin istatistiksel ölçüsünü sağlar. Son olarak, tahmin modeli yeni verilerle kolayca güncellenebilir ve böylece sürekli izleme ve uyarılama sağlanabilir.

ARIMA'daki her bileşen, standart notasyonlu bir parametre olarak işlev görür. ARIMA modelleri için standart notasyon, p , d ve q ile ARIMA olacaktır; burada kullanılan ARIMA modelinin türünü belirtmek için parametrelerin yerine tamsayı değerler geçmektedir. Parametreler şu şekilde tanımlanabilir:

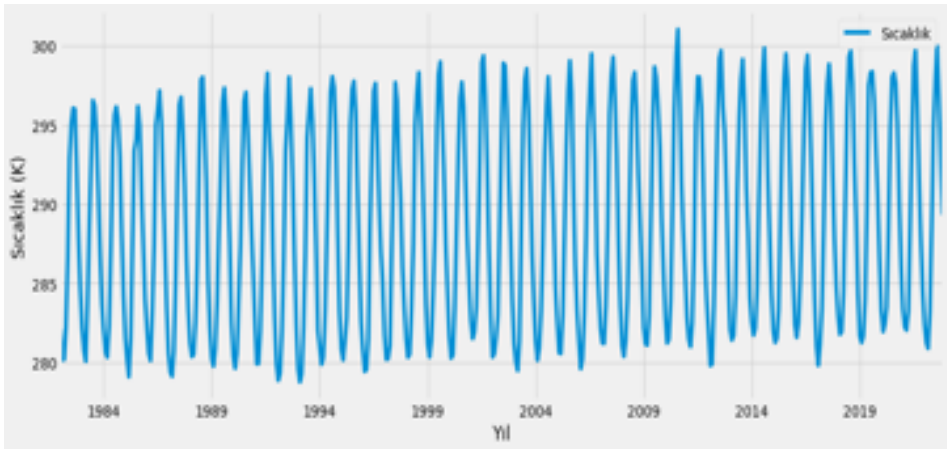
- p , modelin otomatik gerileyen kısmıdır. Geçmiş değerlerin etkisini modelimize dahil etmemizi sağlar. Sezgisel olarak bu, son 3 gündür hava sıcaksa yarın da sıcak olacağını söylemeye benzer.
- d , modelin bütünleşmiş parçasıdır. Bu, zaman serisine uygulanacak fark miktarını (yani mevcut değerden çıkarılacak geçmiş zaman noktalarının sayısı) içeren modeldeki terimleri içerir. Sezgisel olarak bu, son üç gündeki sıcaklık farkı çok küçükse, yarın da aynı sıcaklık olacağını söylemeye benzer.

• q , modelin hareketli ortalama kısmıdır. Bu, modelimizin hatasını geçmişte önceki zaman noktalarında gözlemlenen hata değerlerinin doğrusal bir kombinasyonu olarak belirlememize olanak tanır. (K. Kalpakis, 2001)

ARIMA Uygulama

Elde edilen verilerin analizi sonucunda, deniz yüzey sıcaklıklarının artış trendinde olduğu gözlemlenmiştir. Grafik üzerinde yapılan incelemelerde, en düşük sıcaklığın 26.02.1992 tarihinde 278,21 Kelvin derece, en yüksek sıcaklığın ise 14.08.2010 tarihinde 302,47 Kelvin derece olarak kaydedildiği görülmüştür. Ayrıca, son yıllara bakıldığında sıcaklıkların artış trendinin daha belirgin hale geldiği tespit edilmiştir.

Bu sonuçlar, iklim değişikliği ve küresel ısınma gibi çevresel faktörlerin deniz yüzey sıcaklıklarına olan etkisini ortaya koyabileceği gibi çevresel faktörler ve insan aktiviteleri nedeniyle oluşabilir. Örneğin, fosil yakıtların kullanımı, sera gazı emisyonlarına neden olarak atmosferdeki sıcaklığı artırır ve sonuçta deniz yüzey sıcaklıklarını da etkiler. Bu nedenle nüfusun ve sanayinin artmasının deniz yüzeyi sıcaklıklarına olan etkisini araştırmak için etkili bir neden olabilir. Bu veri setindeki sıcaklık trendleri, deniz yüzey sıcaklıklarının arttığını göstermektedir.



Şekil 1. Yıllara Göre Sıcaklık (Kelvin)

ARIMA modellerinde AIC (Akaike Information Criterion) Akaike ölçütü bir modelin kalitesinin değerlendirilmesinde kullanılan bir ölçüttür. AIC, modelin performansını hem uyum hem de aşırı uyum açısından değerlendirir. AIC, düşük hatayla uyumlu bir modeli tercih ederken aşırı uyumdan kaçınır. Bu nedenle, AIC değeri düşük olan bir model, daha iyi bir model olarak kabul edilir.

AIC değeri, modelin karmaşıklığı ile log-likelihood değerinin bir fonksiyonudur. AIC değerinin hesaplanması için, öncelikle modelin log-likelihood değeri hesaplanır. Log-likelihood, modelin uyum kalitesinin bir ölçüsüdür ve logaritmik bir olasılık değeridir. Modelin karmaşıklığı arttıkça, log-likelihood değeri azalır. AIC, log-likelihood değerine, modelin karmaşıklığının (örneğin, p, d ve q değerleri gibi) bir cezası olarak eklenen bir terimle hesaplanır. Bu nedenle, daha karmaşık bir modelin AIC değeri daha yüksek olacaktır.

AIC, özellikle regresyon analizi gibi büyük veri setleri üzerinde çalışan araştırmacılar tarafından sıklıkla kullanılır. Ancak, yalnızca iki model arasındaki karşılaştırmalar için kullanılması önerilir.

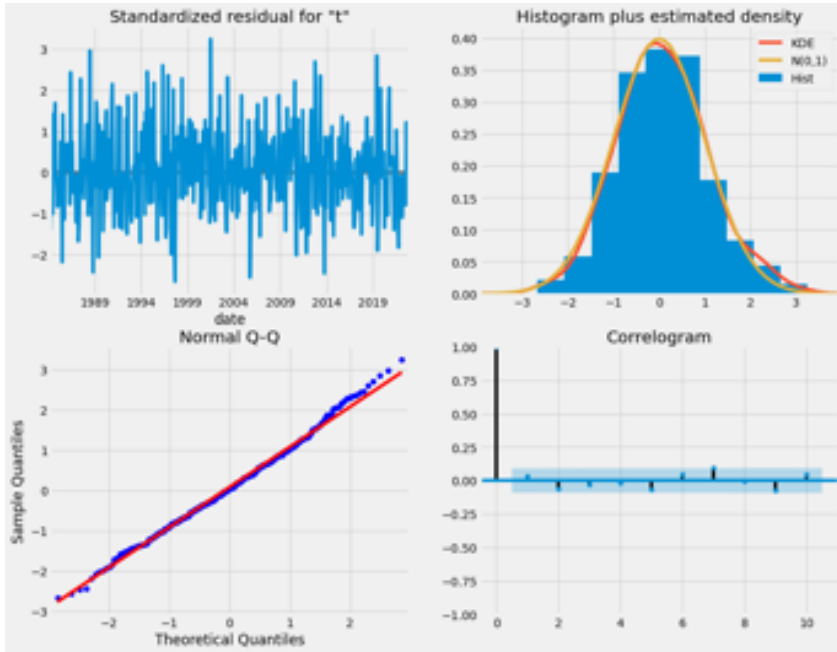
ARIMA modelin de order (p, q, d) ve seasonal order için hesaplanan AIC skorları hesaplanmıştır. En düşük AIC skoru, ARIMA (1, 0, 1) x (0, 1, 1, 12) uygulama içerisinde kullanılmıştır.

ARIMA PARAMETRELERİ	AIC SKOR
ARIMA (0, 0, 0) x (0, 0, 0, 12)	6944.87
ARIMA (0, 0, 0) x (0, 1, 0, 12)	1511.11
ARIMA (0, 0, 0) x (1, 0, 1, 12)	1271.88
ARIMA (0, 0, 1) x (0, 0, 0, 12)	6262.79
ARIMA (0, 0, 1) x (0, 1, 1, 12)	1133.97
ARIMA (0, 0, 1) x (1, 0, 1, 12)	1134.24
ARIMA (0, 1, 0) x (0, 0, 0, 12)	2613.1
....

ARIMA (1, 1, 1) x (0, 1, 1, 12)	1088.62
ARIMA (1, 1, 1) x (1, 0, 0, 12)	1360.91
ARIMA (1, 1, 1) x (1, 1, 1, 12)	1098.14

Tablo 1. ARIMA Değerlerinin AIC Skoru

Ana hedefimiz, modelimizin tahmin hatalarının ilintisiz olmasını ve ortalamasının sıfır ortalama ile normal dağılmasını sağlamaktır. Eğer mevsimsel ARIMA modeli bu özellikleri karşılamıyorsa, bu modelin daha da iyileştirilmesi gerektiği anlamına gelir.

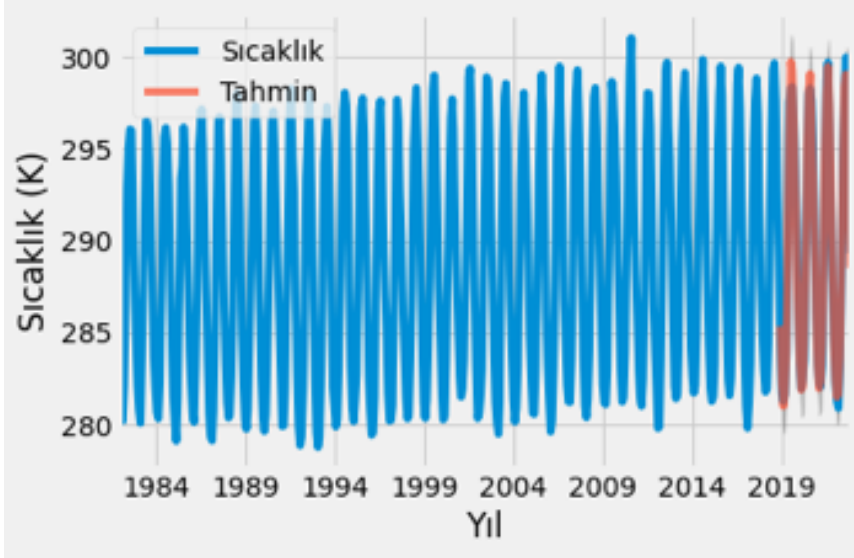


Şekil 2. Eğitilmiş ARIMA Modeli Sonuçları

- Sağ üstteki çizimde, kırmızı KDE çizgisi normal dağılımın ortalama değeri 0 ve standart sapması 1 olan $N(0,1)$ çizgisine yakın bir şekilde seyretmektedir. Burada $N(0,1)$ terimi normal dağılımın standart gösterimidir ve bu, tahmin hatalarının normal bir dağılım gösterdiğinin güçlü bir göstergesidir.
- Sol alttaki Q-Q grafiği, tahmin hatalarının sıralı dağılımının (mavi

noktalar), $N(0, 1)$ ile standart bir normal dağılımdan alınan örneklerin doğrusal eğilimini takip ettiğini gösterir. Yine, bu tahmin hatalarının normal dağıldığının güçlü bir göstergesidir.

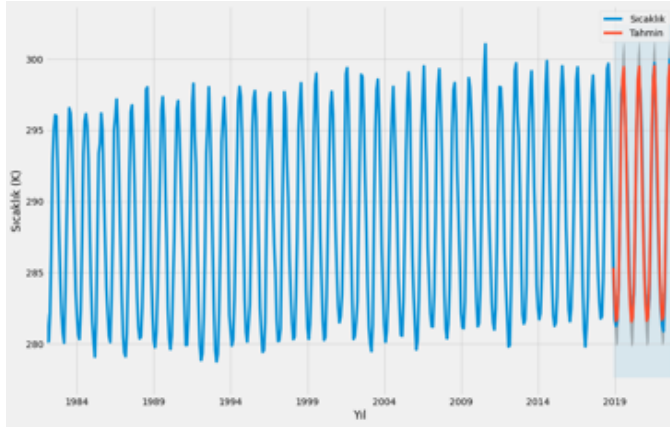
• Zaman serisi artıkları zaman içinde rastgele dalgalanmalar gösterir ve mevsimsel bir kalıp izlemezler. Bu özellik, oto-korelasyon (korelasyon grafiği) grafiğinde de görülebilir. Grafiğin sağ alt kısmı, artıkların kendilerinin gecikmeli versiyonları ile düşük korelasyona sahip olduğunu gösterir, bu da artıkların beyaz gürültü gibi davrandığını ortaya koyar.



Şekil 3. ARIMA Modeli Eğitim / Test

Eldeki veri setini kullanarak ARIMA eğitim modelini uyguladığımızda, 1982-2018 yılları arasındaki verileri eğitim amaçlı kullanırken, 2018 ve 2022 yılları arasındaki verileri test etmek için ayırdık. Bu süreçte, modelimizin performansını ölçmek için ortalama kare hatası (MSE) oranını hesapladık ve sonuç olarak 0.0713 değerini elde ettik. Şekilden de anlaşılacağı üzere, modelin oldukça başarılı olduğu gözlemlenmektedir. Bu sonuçlar, modelin kullanılabilirliğini ve gelecekteki deniz yüzey sıcaklığı tahminlerinde doğru sonuçlar sağlayabileceğini göstermektedir.

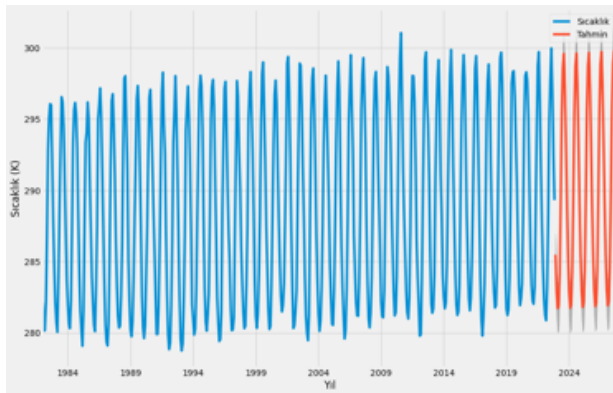
Ancak, modelimizden tamamen emin olmak için dinamik tahminler veri seti olarak kullanılabilir. Bu durumda, zaman serisinden belirli bir noktaya kadar olan bilgileri kullanıp, bundan sonrası için önceki tahmin edilen zaman noktalarından alınan değerler kullanılarak tahminler oluşturulur.



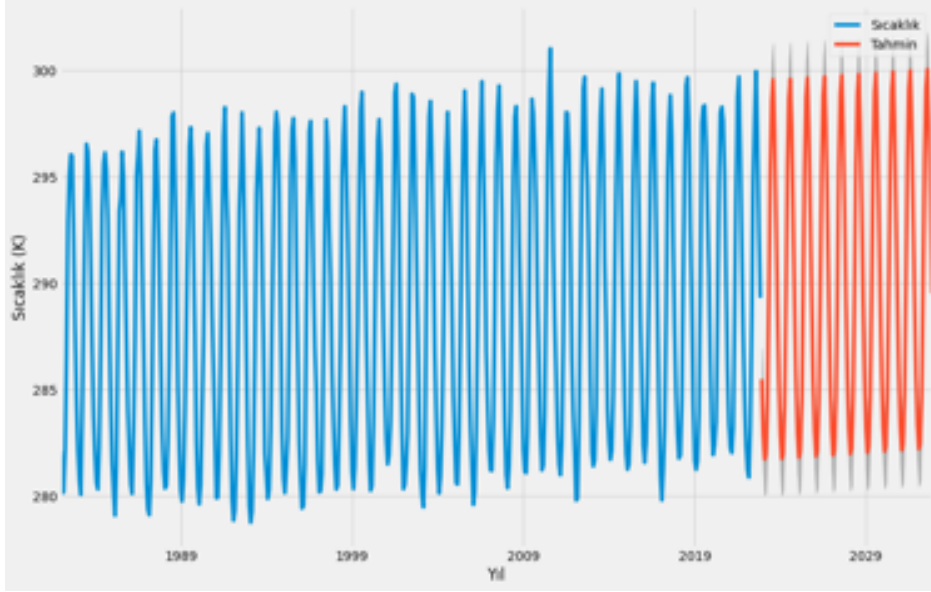
Şekil 4. Model Oluşumu Sonrası Doğrulama Tahmini

Dinamik tahminlerden elde edilen tahmin edilen değerler, 0.16'lık bir MSE verir. Bu bir önceki MSE oranından 2 kat daha yüksek olmasına rağmen hala çok düşük olduğu için bu modelimizin geçerli olduğunu doğrulamaktadır.

Elde ettiğimiz geçerli ARIMA modelimiz ile 2022 yılından başlamak üzere önümüzdeki 5 ve 10 yıllık deniz yüzeyi sıcaklıkları aşağıdaki gibidir.



Şekil 5. ARIMA 5 Yıllık Deniz Yüzey Sıcaklığı Tahmini



Şekil 6. ARIMA 10 Yıllık Deniz Yüzey Sıcaklığı Tahmini

Şekil 6 ve 7 de anlaşılacağı üzere önümüzdeki 5 ve 10 yıllık süreçte deniz yüzeyi sıcaklıklarının yükseliş trendinde olduğu açıkça gözlemlenmektedir.

Tartışma ve sonuç

Bu çalışmada, Deniz Yüzey Sıcaklığının (DYS) ARIMA yöntemi kullanılarak Zonguldak ve Bartın illerindeki sıcaklık trendleri modellenmiştir. Çalışma 1982-2022'ye kadar olan dönemi kapsamaktadır. Model 0.07 gibi düşük bir MSE oranına sahiptir. Modelin sonuçlarına göre, çalışma dönemi boyunca DYS'da yavaş bir artış gözlemlenmiştir. Karşılaştırmalı analizler yapıldığında, başka bir çalışmada da Karadeniz bölgelerindeki deniz yüzey sıcaklığının istikrarlı bir ısınma eğilimi yaşadığını göstermiştir. (Tokat, 2023) Önümüzdeki 5 ve 10 yıllık süre boyunca yapılan deniz yüzey sıcaklığı tahminleri bize gösteriyor ki sıcaklıklar çalışma bölgesi için artmaya devam edecektir.

Kaynaklar

- [1] Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 159-175.
- [2] Merchant, C. J. (2020). Adjusting for desert-dust-related biases in a climate data record of sea surface temperature. *Remote Sensing*, 2554.
- [3] Merchant, C. E. (2019). Satellite-based time-series of sea-surface temperature since 1981 for climate applications. *Scientific Data*.
- [4] Manabe, S. a. (1988). Two stable equilibria of a coupled ocean-atmosphere model. *Journal of Climate*, 841-866.
- [5] Mohamed, B. &. (2022). Sea Surface Temperature Variability and Marine Heatwaves in the Black Sea. *Remote Sensing*, 2383.
- [6] Tokat, E. a. (2023). Climatology and Variability of Sea Surface Temperature in the Region of Turkish Straits System, 1982-2021. *EGU General Assembly 2023*, EGU23-15096.
- [7] Cengiz, M. &. (2020). A solution of some commonly used optimization functions by a hybrid BFGS-PSO algorithm. *ournal of the Faculty of Engineering and Architecture of Gazi University*, 925-938.
- [8] Gülhan Toğa, B. A. (2021). COVID-19 prevalence forecasting using Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Networks (ANN): Case of Turkey. *ournal of Infection and Public Health*, 811-816.
- [9] Bollapragada, R. &.-J. (2018). A Progressive Batching L-BFGS Method for Machine Learning.
- [10] K. Kalpakis, D. G. (2001). Distance measures for effective clustering of ARIMA time-series,” Proceedings 2001 IEEE International Conference on Data Mining. *Proceedings 2001 IEEE International Conference on Data Mining, San Jose*, (s. 273-280). San Jose, CA, USA.