

İstatistik Ölçek Küçültme Çalışmaları

*Ozan Mert Göktürk, Avrasya Yer Bilimleri Enstitüsü

Giriş ve Yöntem

İstatistik ölçek küçültme, belirli bir bölgedeki yerel iklim değişkenleri ile o bölgeyi etkilediği varsayılan büyük ölçekli iklim değişkenleri arasındaki istatistik ilişkiler kullanılarak, büyük ölçekli iklimden bölgesel iklimin öngörülmesi olarak özetlenebilir. Kullanımındaki en büyük avantaj, dinamik ölçek küçültmeye göre çok daha az bilgisayar gücüne ve süreye ihtiyaç duyulmasıdır, performansı da dinamik ölçek küçültmeyle karşılaştırılabilir düzeydedir (Von Storch ve diğ., 1993). Wilby ve diğ. (2004), istatistik ölçek küçültme yöntemlerini üç ana grupta toplamıştır: 1. Sinoptik durumların, benzerliklerine ve bölgesel iklime etkilerine göre gruplandığı **hava sınıflama şemaları**. 2. Regresyon modelleri. 3. Bir yerel iklim değişkeninin istatistiki özelliklerini yeniden üreten ve parametreleri büyük ölçekli değişkenlerinin birer fonksiyonu olan, **hava üreticileri** (weather generators).

Büyük ölçekte ve çok düşük çözünürlükte veri üreten ve yerel iklimin özellikle istasyon bazında ne yönde değişeceği konusunda fikir vermeyen genel dolaşım modeli çıktılarının, dinamik yanında istatistik ölçek küçültmeye de tabi tutulması ve bu yolla yerel ölçekte yüksek çözünürlüklü veri elde edilmesi, Türkiye İklim Değişimi Senaryoları projesinin iş paketleri arasında yer almaktadır. Bu amaçla ilk aşamada, bir regresyon modeli örneği olan ve istatistik ölçek küçültme çalışmalarında yaygın olarak kullanılan **kanonik korelasyon analizi** (canonical correlation analysis) uygulanmış, oluşturulan istatistiki model test edilmiş, sıcaklık ve yağış gibi ana iklim değişkenlerini aylık ve mevsimlik zaman ölçeğinde simüle etmedeki performansı değerlendirilmiştir. İkinci aşamada, oluşturulan modelden, 21. yüzyıl için çalıştırılan genel dolaşım modellerinin çıktılarının istatistik ölçek küçültmesinde faydalanılacaktır.

Kanonik korelasyon analizi, klimatolojideki yaygın kullanımından yola çıkılarak tanımlanırsa, zaman ve mekan koordinatları bulunan iki veri setinin (örneğin büyük ölçekli ve yerel iklim veri setleri) her birinden, ilişkileri en yüksek düzeyde olan birer coğrafi paterni matematiksel olarak çıkarmaya denir. Analizin ana varsayımı, her bir veri setinin, **coğrafi paternler** ve bunların **zaman katsayılarının** lineer kombinasyonu olarak ifade edilebileceğidir. Elde edilen birbiriyle ilişkili paternler, bir **kanonik patern çifti** oluşturur. Bu patern çiftinin arasındaki ilişkinin derecesini, paternlerin zaman katsayıları arasındaki korelasyon belirler ki, buna **kanonik korelasyon** denir. Bir başka deyişle büyük ölçekli ve yerel coğrafi paternler, kanonik korelasyon ölçüsünde, birlikte gerçekleşirler. İlk patern çiftinden sonra, ilişki düzeyleri giderek azalan birkaç patern çifti daha bulunur. Daha sonra bu patern çiftlerinin en **baskınları** ve aralarındaki korelasyonlar kullanılarak, yerel iklim değişkenlerini büyük ölçekli değişkenlerden öngörmeye yarayan bir transfer fonksiyonu elde edilir. Bu nedenle, büyük ölçekli iklim değişkenlerinden her birine **öngören** (predictor), yerel iklim değişkenlerine de **öngörülen** (predictand) adı verilir. Kanonik patern çiftlerinin baskınlıkları, kanonik korelasyonlar ve her bir paternin kendi ana veri setinde temsil ettiği varyanslar tarafından belirlenir. Analizin ayrıntılı matematiksel açıklaması Von Storch ve Zwiers (1999), Barnett ve Preisendorfer (1987).

Analizden önce, kullanılacak veri setlerindeki her bir zaman serisinden kendi ortalamasını çıkararak değerleri sıfırdan fark (anomali) olarak ifade etmek, analiz sonucu elde edilen paternlerin ve zaman katsayılarının kolaylıkla yorumlanabilmesini sağlar; çünkü bu şekilde elde edilen kanonik paternler kullanılan değişkenin birimine sahip birer **anomali paternidir** ve büyük ölçekteki anomalilere yerel ölçekte hangi anomalilerin eşlik ettiğini açık hale getirir. Zaman katsayıları da bağlı olduğu paternin zaman içindeki gücünü ifade eden boyutsuz sayılardır. Öte yandan, yağış gibi, her bir noktadaki varyansı birbirinden çok farklı olabilen değişkenlerle çalışırken her bir zaman serisini **standart** hale getirmek, yani anomali olarak ifade ettikten sonra standart sapmasına da bölmek gereklidir. Aksi takdirde, bazı bölgelerdeki yüksek varyans değerlerinden ötürü, öteki bölgelerdeki anomali değerleri görece çok küçük görünmekte, bu durum yanlış yorumlara yol açabilmektedir.

Kanonik korelasyon analizi, çok fazla zaman serisi içeren büyük veri setleri için yapıyorsa, öngören ve öngörülen veri setlerine **ana bileşen analizi** uygulamak gerekir. Bu yolla, hem analizin ara aşamalarında oluşturulan matrislerin dejenere olma riski ortadan kalkar, hem de ana veri setlerinin değişkenliklerinin çoğu (yaklaşık %95) orijinal zaman serisi sayısının çok daha azıyla ifade edilmiş olur ki, bu, istatistik modelin sonuç üretme süresini önemli ölçüde kısaltır.

Kullanılan Veri Setleri ve Seçilen Parametreler

Kanonik korelasyon analizini temel alan istatistik ölçek küçültme modeli oluşturulurken kullanılan büyük ölçekli iklim değişkenleri NCEP-NCAR renaliz veri setlerinden (Kalnay ve diğ., 1996) alınmıştır. Bu veri setleri aylık ortalamalar halinde 2.5x2.5 derecelik gridler üzerinde bulunmakta ve 1949-2004 periyodunu kapsamaktadır. Yerel iklim değişkenleri ise, Devlet Meteoroloji İşleri'nden elde edilen istasyon veri setlerinden alınmıştır. Bu veri setlerinde 1949-2004 arasında aylık sıcaklık verileri tam olan 63, yağış verileri tam olan 126 istasyon bulunmaktadır. Veri setlerinin kalite kontrolleri projenin önceki altı aylık döneminde gerçekleştirilmiştir.

Büyük ölçekli (öngören) değişken olarak 500 hPa jeopotansiyel yükseklik ve deniz seviyesi basıncı kullanılmıştır. Bu iki değişkenin sırasıyla serbest atmosferin ve yer seviyesindeki atmosferin değişkenliğini iyi temsil ettikleri bilinmektedir (Von Storch, 1993). 500 hPa jeopotansiyel yüksekliğin kullanıldığı model test aşamasında bir miktar daha iyi performans gösterdiğinden, sonuç bölümünde yer verilen şekillerde öngören bu değişkendir. Öngörülen değişken olarak, aylık olmak üzere sıcaklık ortalaması, maksimum sıcaklık ortalaması, minimum sıcaklık ortalaması ve toplam yağış kullanılmıştır.

İstatistik ölçek küçültme modelini oluştururken kullanılan parametrelerden modelin performansını en fazla etkileyenlerin şunlar olduğu görülmüştür:

- 1) Büyük ölçekli değişkenlerden ve yerel verilerinden alınan ve kanonik korelasyon analizinde kullanılan **ana bileşen sayıları**. Model performansını test etmek için yapılan deneyler, hem büyük ölçekli değişkenler hem de istasyon verileri için, ana bileşen sayısının 10 civarında olması gerektiğini ortaya koymuştur.

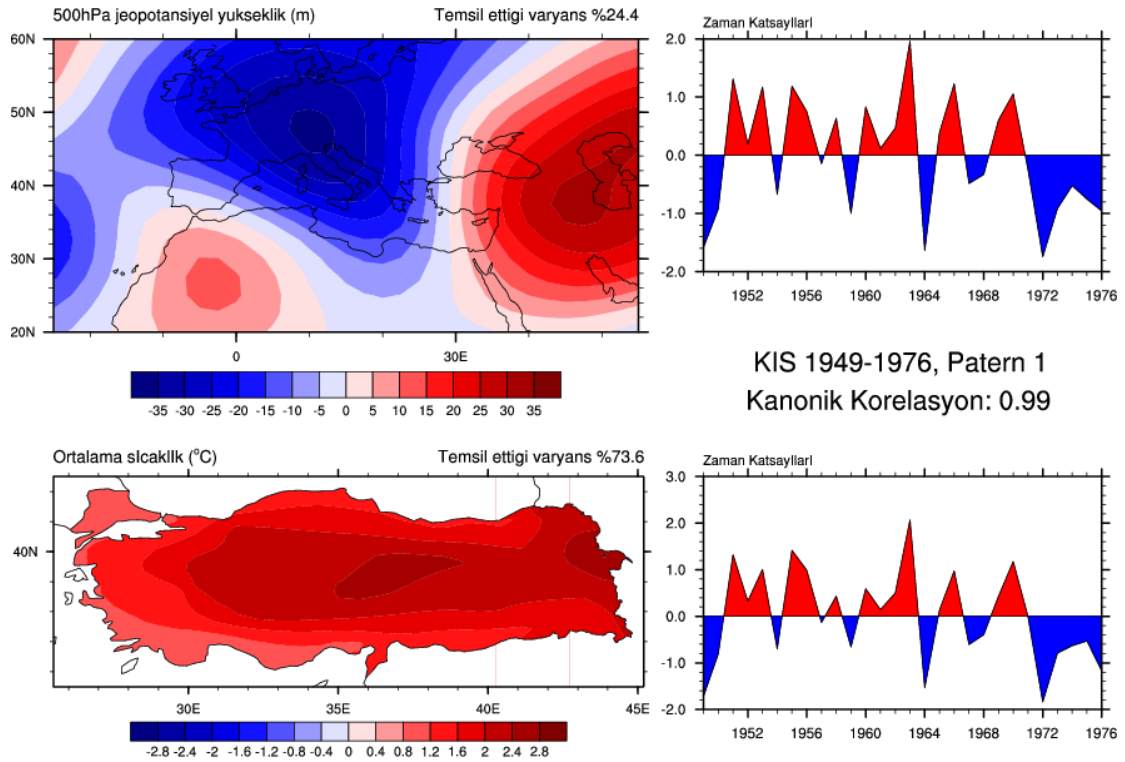
- 2) Modeli oluştururken kullanılan **kanonik patern sayısı**. Yapılan deneyler, bu sayının öngörülen değişkene göre belirlenmek üzere 2 ya da 3 olması gerektiğini ortaya koymuştur.
- 3) Türkiye istasyon verileriyle ilişkili olduğu varsayılan büyük ölçekli değişkenin kapsadığı coğrafi bölgenin sınırları. 25 derece batı, 55 derece doğu boylamları ve 20-60 derece kuzey enlemleri arasında kalan bölge kullanılarak oluşturulan modelin en yüksek performansı gösterdiği ortaya çıkmıştır.

Model performansı değerlendirilirken yukarıdaki parametreler sabit tutulmuş, veri setlerinin tam olduğu 1949-2004 dönemi ikiye bölünmüş, model önce 1949-1976 periyodunda oluşturulup 1977-2004 arasında test edilmiştir. Daha sonra, oluşturma ve test periyotları değiştirilerek sonuçların tutarlılığı da değerlendirilmiştir.

Sonuç, Çıkarım ve Öneriler

Aşağıda, sadece 500 hPa jeopotansiyel yüksekliğin öngören olarak kullanıldığı mevsimlik bazdaki model sonuçlarına yer verilmiştir.

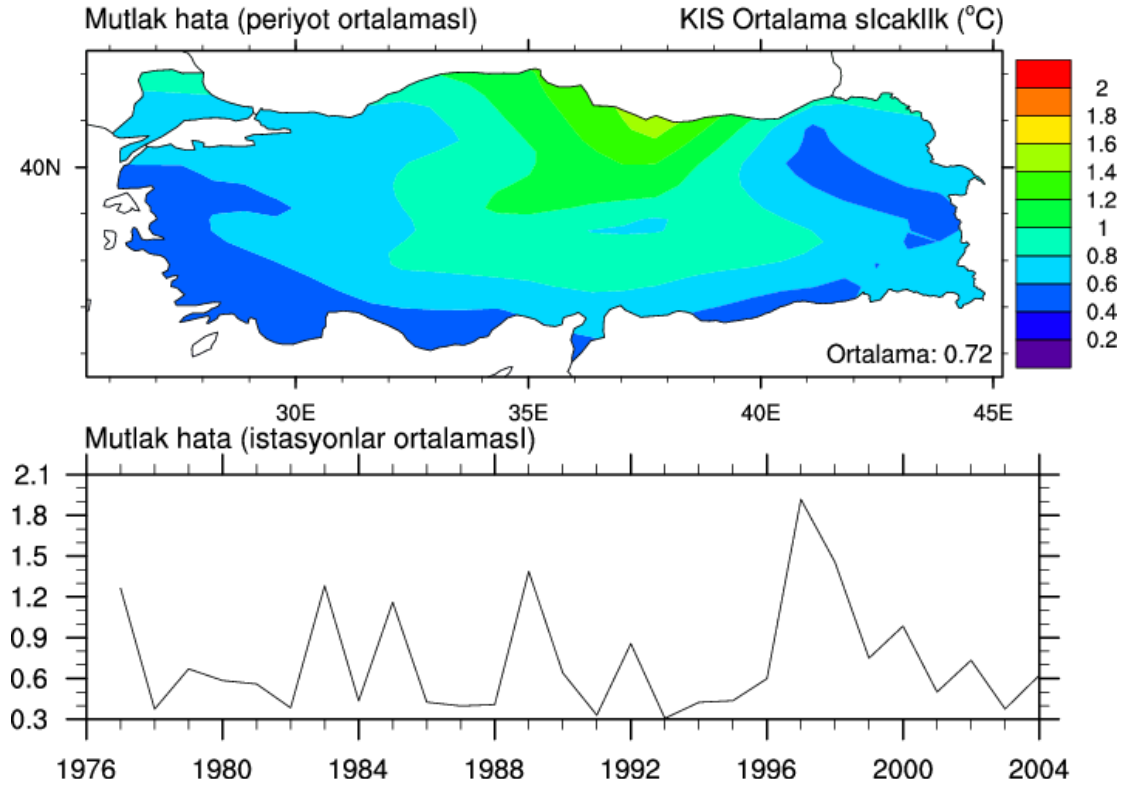
Ortalama Sıcaklık



Şekil 1: 500 hPa jeopotansiyel yükseklik ve ortalama sıcaklık için kanonik korelasyon paternleri. Kış mevsimi, 1949-1976 periyodu.

Şekil 1’de, kış mevsimi için, 500 hPa jeopotansiyel yüksekliğin öngören, istasyon ortalama sıcaklıklarının öngörülen olarak kullanıldığı istatistik ölçek küçültme

modeline temel oluşturan 3 kanonik patern çiftinden ilki görülmektedir. Öngörülen (predictand) paterni, kendi ana veri setinde temsil ettiği %74 varyansla değişkenliğin en baskın modudur. Türkiye genelinde kış ortalama sıcaklıklarının 1 ila 3 santigrad derece arttığı bu paterne eşlik eden büyük ölçekli patern, Orta Avrupa'da merkezlenen bir negatif, ve Hazar Denizi'nde merkezlenen bir pozitif jeopotansiyel yükseklik anomali alanı ile belirlenir. İki paternin arasındaki ilişki, gerçekte ikisinin zaman katsayıları arasındaki korelasyon olan kanonik korelasyonun 0.99 gibi çok yüksek bir değer olmasından anlaşılabilir. Bu patern çiftinin birlikte gerçekleşmesi gerçekten de akla yakındır, çünkü kışın ılık havayı tüm Avrupa'ya ve Türkiye'ye taşıyan ve gezici alçak basınç alanlarıyla karakterize olan zonal akış, Orta Avrupa merkezli bir negatif jeopotansiyel yükseklik alanının oluşmasına yol açar. Coğrafi paternlerin sağında ise her bir paterne ait zaman katsayıları görülmektedir. Zaman katsayısının 0'a yakın oluşu, o paternin o yıl baskınlığının azaldığına işaret etmekte, -1 civarında olması ise paternin tam tersinin görüldüğünü belirtmektedir.

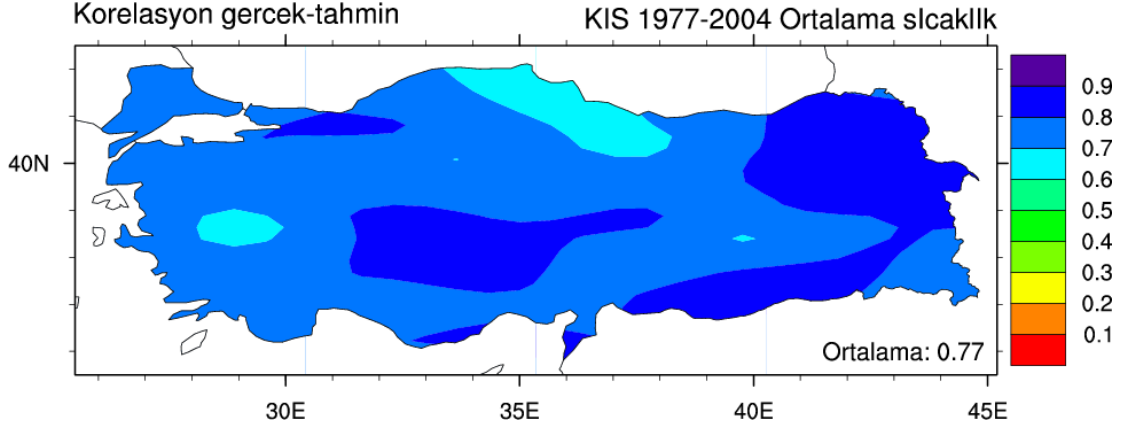


Şekil 2: Test periyodu için performans değerlendirmesi (mutlak hatalar).
Kış ortalama sıcaklığı, 1977-2004.

Şekil 2'de, test periyodu için (1977-2004) istatistik modelden hesaplanan tahmini kış mevsimi ortalama sıcaklık değerleriyle, aynı periyotta ölçülmüş gerçek değerler arasındaki farkların coğrafi dağılımı, ve bu farkların yıllar içerisinde nasıl değiştiği görülmektedir. Model, kış ortalama sıcaklıklarını tahmin etmede özellikle Türkiye'nin güney, batı ve doğu kesimleri için çok iyi performans göstermekte, kuzey ve orta kesimler içinse hatalar biraz artmaktadır. 1949-1976 için yapılan testlerde de benzer sonuçlar elde edilmiştir. Mutlak hatanın istasyonlar ortalamasının ise yıldan yıla çok fazla değişkenlik gösterebildiği görülmektedir. Hatanın fazla olduğu kışların genelde

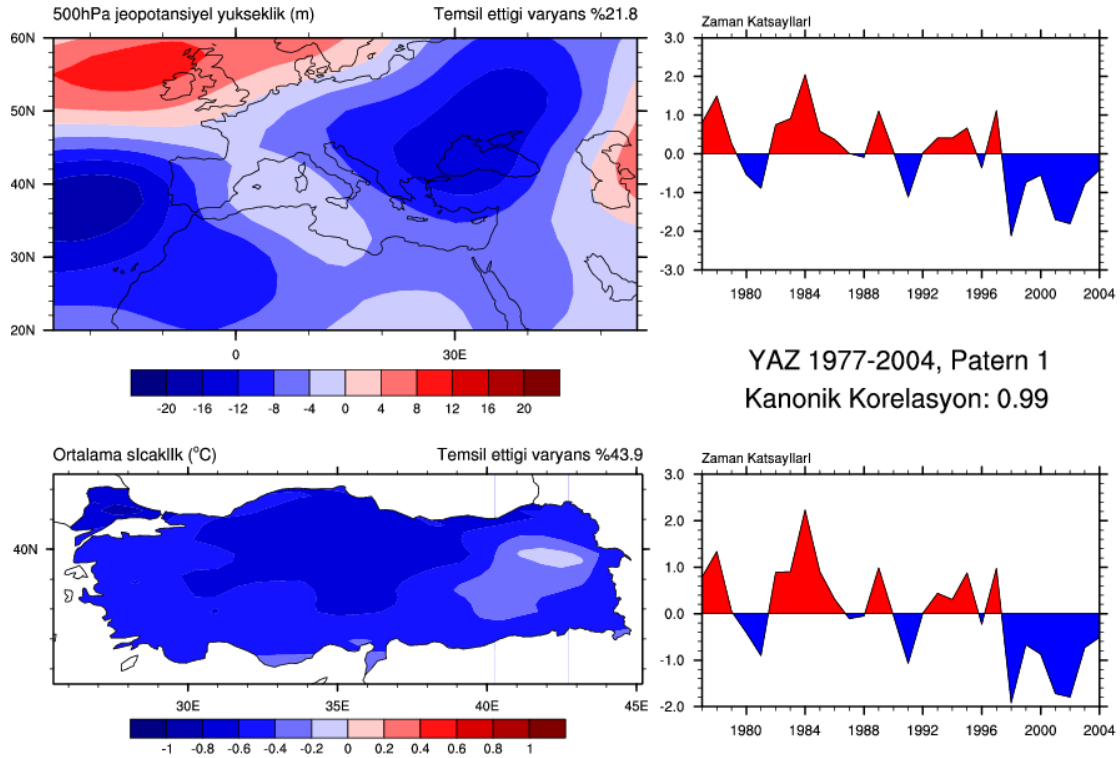
El Nino ya da La Nina görülen kışlar olması ilginç bir sonuçtur. Bu, Türkiye ikliminin sadece yakın çevresinin koşullarından etkilenmediğine kanıt olarak sunulabilir.

Şekil 3'te aynı değişkenler ve periyot için bu kez tahmini ve ölçülmüş değerler arasındaki korelasyonlar gösterilmiştir. Değerler tüm Türkiye için, %99 güvenirlilik düzeyinde anlamlı olan 0.48'in bile çok üzerindedir.

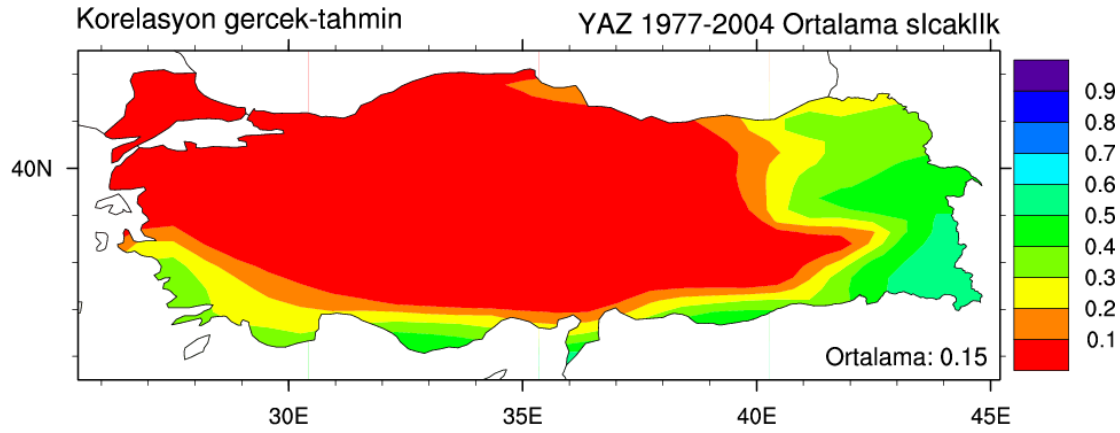


Şekil 3: Test periyodu için performans değerlendirmesi (korelasyonlar).
Kış ortalama sıcaklığı, 1977-2004.

Şekil 4'te yaz ortalama sıcaklıkları ve 500 hPa jeopotansiyel yükseklikleri arasındaki ilişkiyi temsil eden kanonik patern çiftlerinin ilki görülmektedir. Öngörülen paterni, kış mevsiminde olduğu gibi yine yüksek sayılabilecek bir varyans temsil etmesine rağmen, modelin performansı Şekil 5'ten anlaşılacağı üzere ülkenin güney kıyıları ve en doğusu haricinde gayet düşüktür. Deniz seviyesi basıncının öngören olarak kullanıldığı modelden de benzer bir performansın elde edilmiş olması, yaz ortalama sıcaklıklarının ölçek küçültme çalışmalarında farklı öngörenlerin denenmesi gerektiği sonucunu ortaya koymuştur.

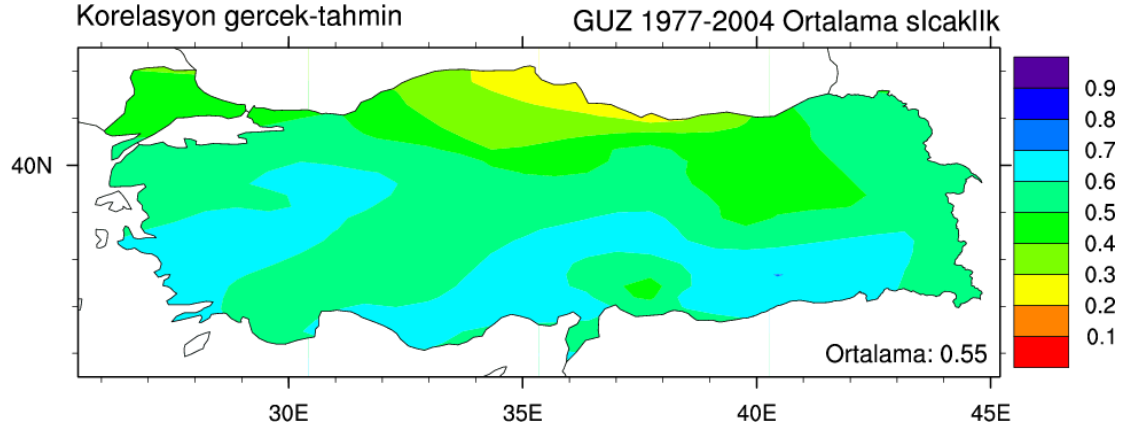


Şekil 4: 500 hPa jeopotansiyel yükseklik ve ortalama sıcaklık için kanonik korelasyon paternleri. Yaz mevsimi, 1949-1976 periyodu.

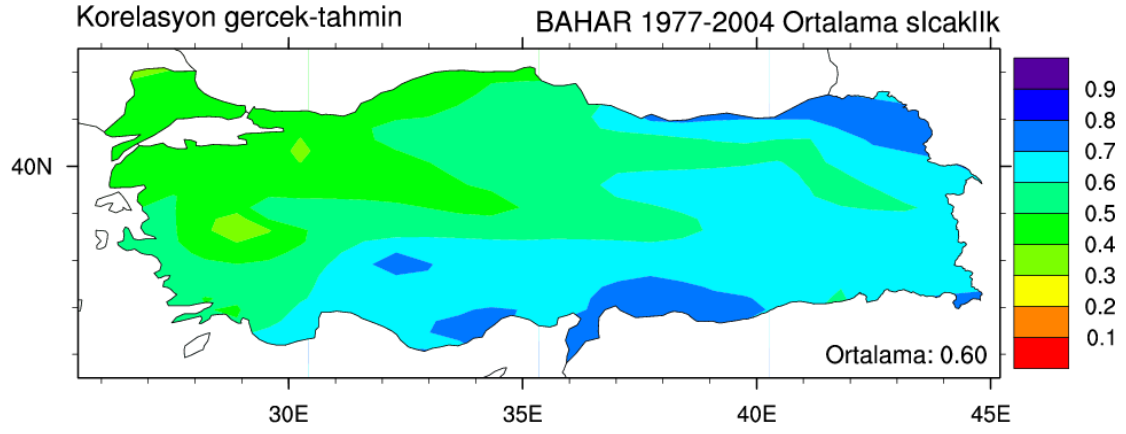


Şekil 5: Test periyodu için performans değerlendirmesi (korelasyonlar). Yaz ortalama sıcaklığı, 1977-2004.

Şekil 6 ve 7'de modelin ortalama sıcaklıkları tahmin etmedeki performansının geçiş mevsimlerinde de kış için olduğu kadar olmasa da yüksek olduğu görülmektedir. Sonbaharda özellikle güney, ilkbaharda ise özellikle güney ve doğu bölgelerdeki korelasyonlar üst seviyededir.

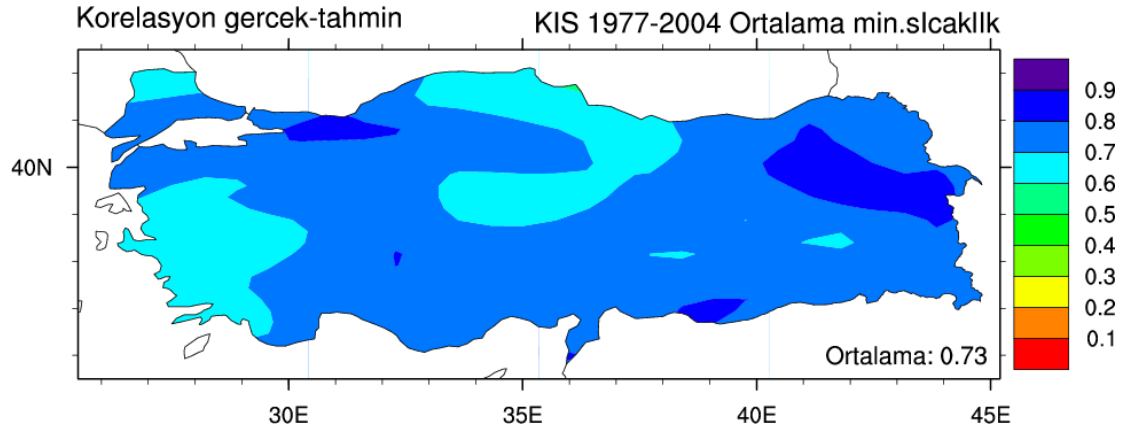


Şekil 6: Test periyodu için performans değerdendirme (korelasyonlar).
Güz ortalama sıcaklığı, 1977-2004.

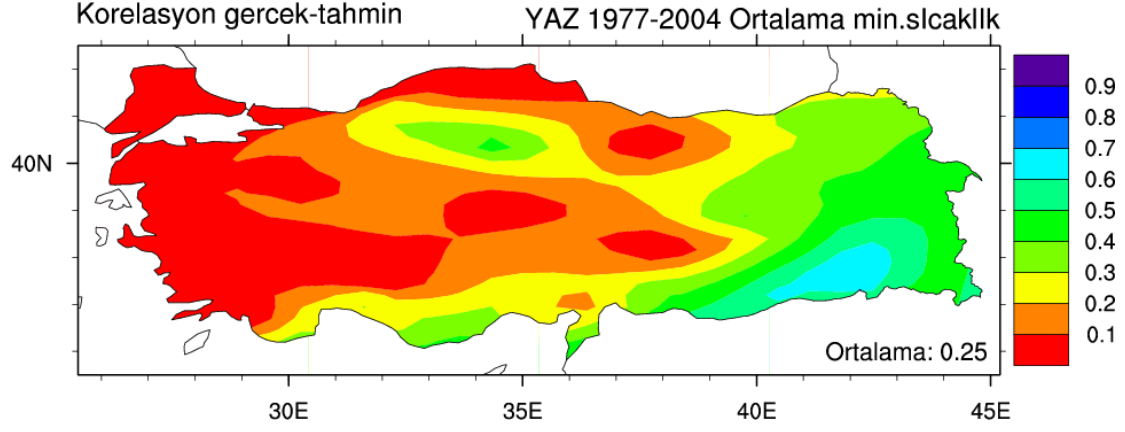


Şekil 7: Test periyodu için performans değerdendirme (korelasyonlar).
Bahar ortalama sıcaklığı, 1977-2004.

Ortalama Minimum Sıcaklık



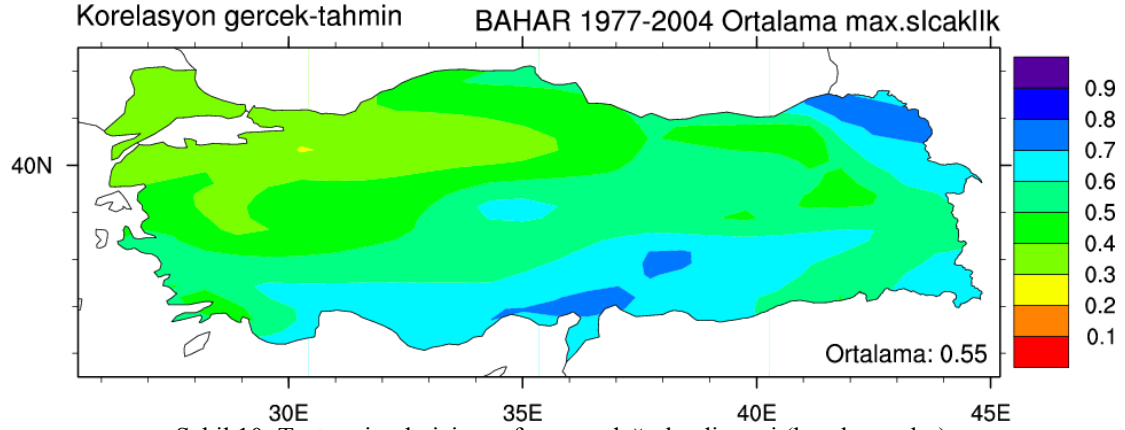
Şekil 8: Test periyodu için performans değerdendirme (korelasyonlar).
Kış ortalama minimum sıcaklığı, 1977-2004.



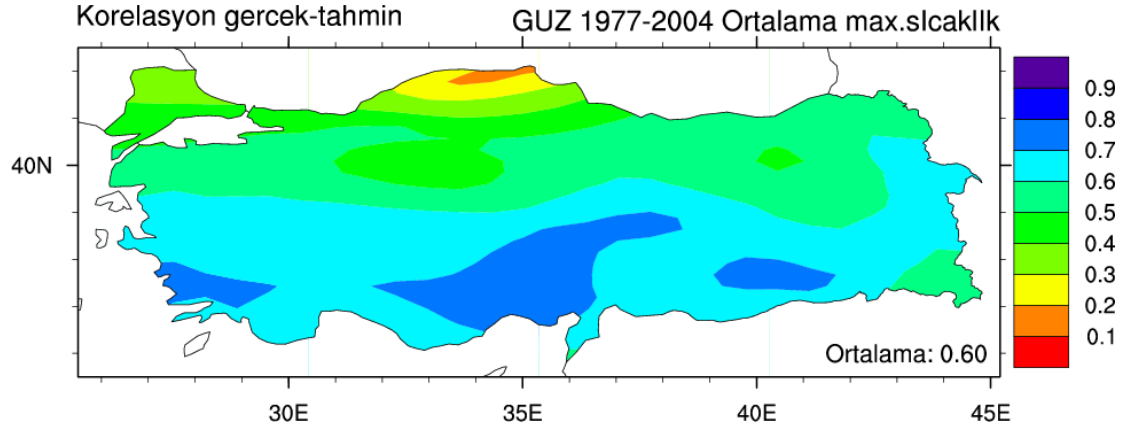
Şekil 9: Test periyodu için performans değerlendirmesi (korelasyonlar).
Yaz ortalama minimum sıcaklığı, 1977-2004.

Modelin mevsimlik ortalama minimum ve maksimum sıcaklıkları tahminindeki performansı, mevsimlik ortalama sıcaklıkları tahmin performansı ile paralellik göstermektedir. Korelasyonların değerleri ve coğrafi dağılımları benzerdir. (Şekil 8, 9, 10 ve 11.)

Ortalama Maksimum Sıcaklık



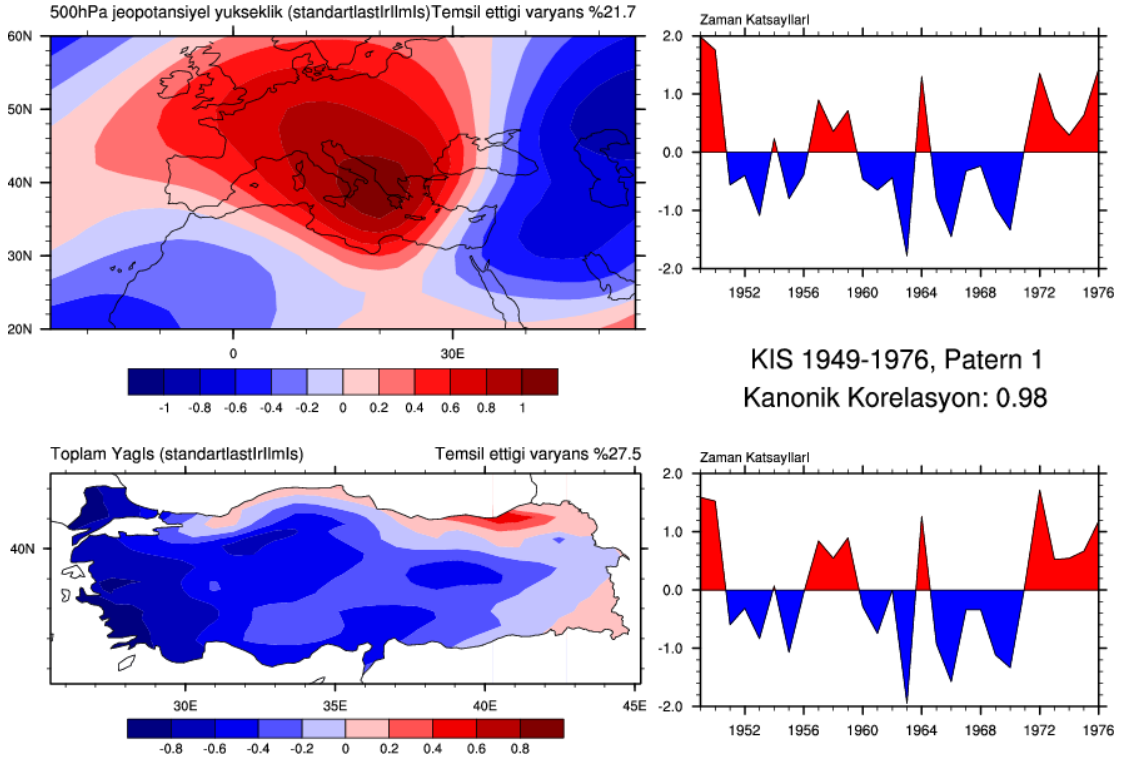
Şekil 10: Test periyodu için performans değerlendirmesi (korelasyonlar).
Bahar ortalama maksimum sıcaklığı, 1977-2004.



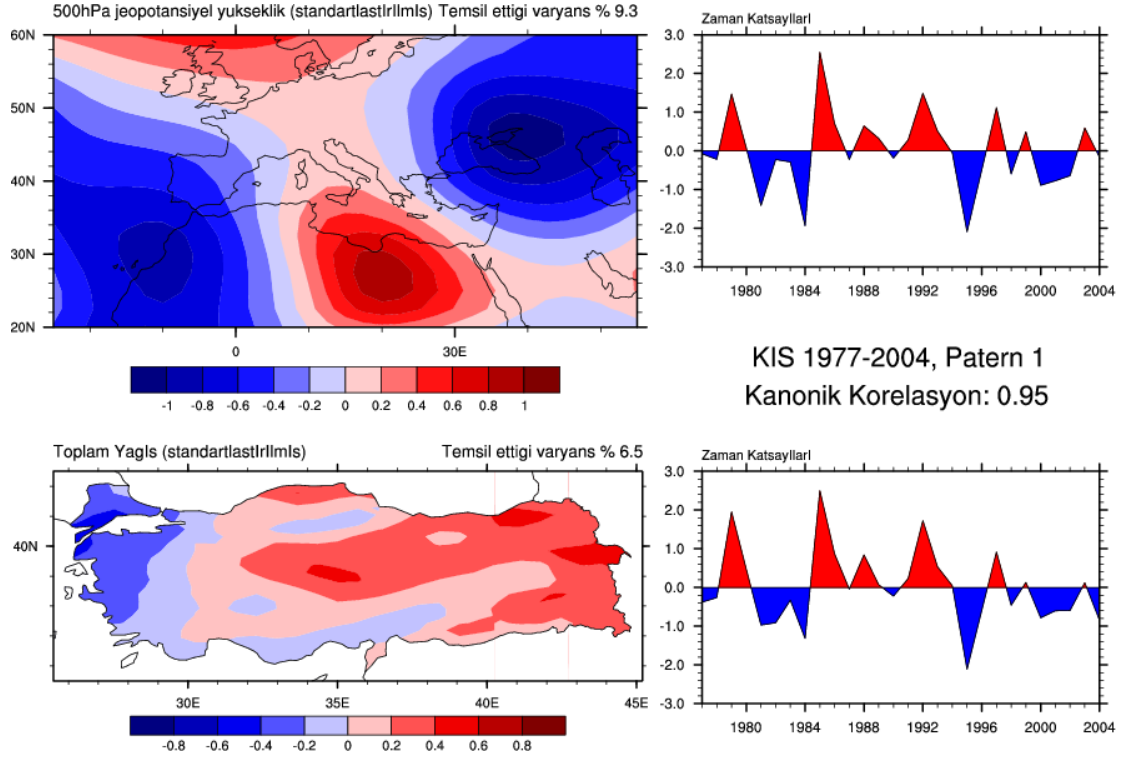
Şekil 11: Test periyodu için performans değerlendirmesi (korelasyonlar).
Güz ortalama maksimum sıcaklığı, 1977-2004.

Toplam Yağış

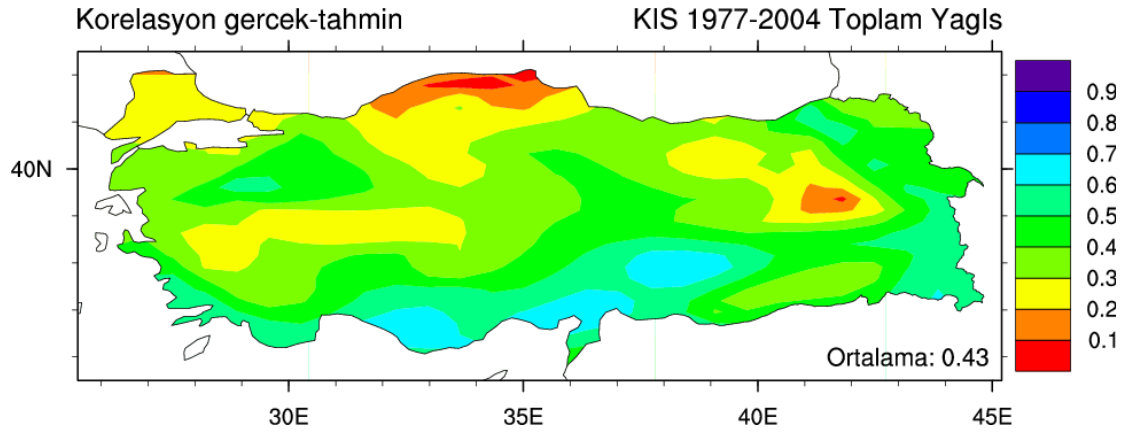
Şekil 12 ve 13'te sırasıyla 1949-1976 ve 1977-2004 periyotlarının kış mevsimleri için, 500 hPa jeopotansiyel yükseklikle yağış arasındaki ilişkiyi ortaya koyan ilk kanonik patern çiftleri gösterilmektedir. Dikkat edileceği üzere, hem coğrafi paternlerde hem de temsil edilen varyanslarda gözle görülür bir fark vardır. Bu fark model performansına da yansımaktadır. Şekil 14 ve 15'te görülebileceği gibi, iki test periyodunda modelin performansının yüksek olduğu bölgeler birbirinden farklıdır. Bu, yağış için olan modelin oluşturulma periyodundan büyük ölçüde etkilendiğini ve sonuçların bundan dolayı tutarsızlaştığını göstermektedir. Benzer durumlar, diğer mevsimler için yapılan analizler için de geçerlidir. Yağış değişkeninin ölçek küçültmesinde model oluşturulurken, mümkün olduğunca uzun bir periyot kullanmak gerektiği açıktır. Bu amaçla, projenin bundan sonraki aşamasında yağış için, modeli eldeki bütün verilerle oluşturan ve daha sonra her bir deneyde tek bir yılı test yılı olarak kullanan 'jack-knifing' yöntemi kullanılacaktır. (Barnett and Preisendorfer, 1987)



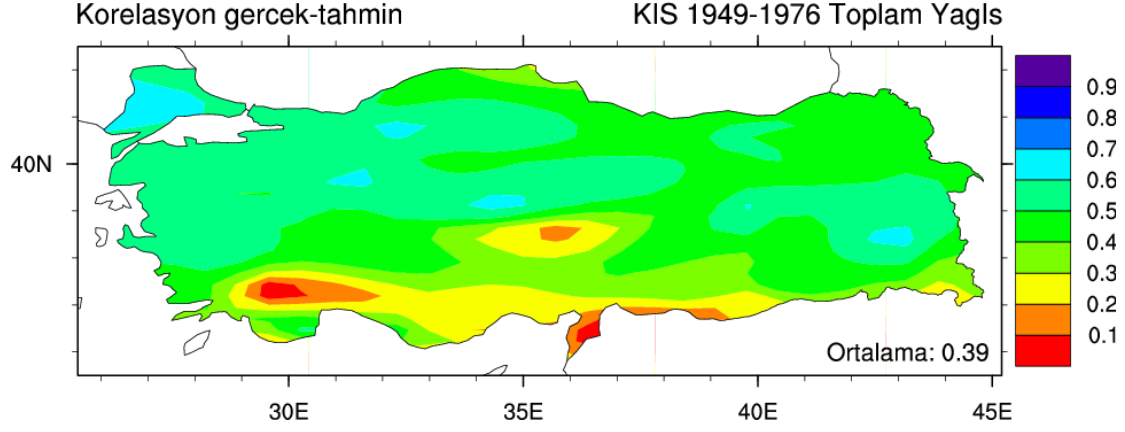
Şekil 12: 500 hPa jeopotansiyel yükseklik ve yağış için kanonik korelasyon paternleri.
Kış mevsimi, 1949-1976 periyodu.



Şekil 13: 500 hPa jeopotansiyel yükseklik ve yağış için kanonik korelasyon paternleri. Kış mevsimi, 1977-2004 periyodu.



Şekil 14: Test periyodu için performans değerlendirmesi (korelasyonlar). Kış toplam yağışı, 1977-2004.



Şekil 15: Test periyodu için performans değerlendirmesi (korelasyonlar).
Kış toplam yağışı, 1949-1976.

Projenin bundan sonraki aşamalarında istatistik ölçek küçültme yöntemlerinden **çoklu lineer regresyona** da yer verilecek, ayrıca her iki yöntem çeşitli başka öngören ve öngörülen değişkenlerle, farklı model parametreleri kullanılarak denenecektir.

REFERANSLAR

Barnett, T. P., Preisendorfer, R.W., 1987. Origins and levels of monthly and seasonal forecast skill for United States air temperatures determined by canonical correlation analysis. *Monthly Weather Review*, **115**: 1825-1850.

Barret, B. S., 2006. Relationship between sea surface temperature anomalies and precipitation across Turkey. In Ünal, Y., Kahya, C., Barl, D. D. (Eds.), *Proceedings of the International Conference on Climate Change and the Middle East: Past, Present and Future.*, İstanbul Teknik Üniversitesi, İstanbul.

Kalnay, E., ve diğerleri. 1996. The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of American Meteorological Society*, **77**: 437-471.

Karaca, M., A. Deniz, and M. Tayanç, 2000. Cyclone track variability over Turkey in association with regional climate, *International Journal of Climatology*, **20**:, 122-136.

Von Storch, H., Zorita, E., Cubasch, U, 1993. Downscaling of global climate change estimates to regional scales: An application to Iberian rainfall in wintertime. *Journal of Climate*, **6**: 1161-1171.

Von Storch, H., Zwiers, F.W., 1999. *Statistical Analysis in Climate Research.* Cambridge University Press, UK.

Wilby, R. L., Charles, S.P., Zorita, E., Timbal, B., Whetton, P., Mearns, L. O., 2004. Guidelines for use of climate scenarios developed from statistical downscaling methods. *Supporting material for IPCC reports.*