

HABITUS

Habitat Analysis and Biodiversity Integrated Toolkit for Unified Species Distribution Modelling

Uygulamalı Kullanım Kılavuzu

Ömer K. Örücü · Seda Örücü

Süleyman Demirel Üniversitesi, Mimarlık Fakültesi,

Peyzaj Mimarlığı Bölümü, ISPARTA/Türkiye

omerorucu@sdu.edu.tr · ORCID 0000-0002-2162-7553

sedaorucu@sdu.edu.tr · ORCID 0000-0003-1592-5180

GİRİŞ

Bu kılavuz, tür dağılım modellemesi (TDM) için geliştirilmiş masaüstü yazılımı HABITUS'un uygulamalı bir tanıtımıdır. Programlama bilgisi ya da R, QGIS veya Python kurulumu gerektirmez. Aşağıdaki bütün adımlar grafik arayüz üzerinden yapılır; ekran görüntüleri de temsili değil, gerçek bir çözümlemeden alınmıştır.

HABITUS bütün TDM iş akışını tek pencerede toplar: veri hazırlama, çoklu bağlantı taraması, on üç algoritmayla model kurma, gelecek iklime projeksiyon, alan değişimi çözümlemesi, değerlendirme, bağımsız doğrulama ve otomatik raporlama. Ayırt edici özelliği bu aşamaların *sıralı* olmasıdır: her sekme, bir önceki başarıyla tamamlanana kadar kilitli kalır. Hangi değişkenleri kullanacağınıza karar vermeden model kuramaz, henüz üretilmemiş bir haritayı doğrulayamazsınız.

Kılavuz boyunca kullanılan örnek, Türkiye'de doğal yayılış gösteren *Pinus brutia* Ten. (kızılçam) türüdür: 108 varlık kaydı, 23 çevresel katman, CNRM-ESM2-1 iklim modelinin SSP5-8.5 senaryosu ve bağımsız referans olarak EUFORGEN yayılış haritası. Veri seti yazılımla birlikte dağıtıldığı için buradaki her şekli kendiniz üretebilirsiniz.

Yöntemlerin kuramsal arka planı için Guisan ve Thuiller (2005) ile Elith ve Leathwick (2009); raporlama standartları için Zurell ve ark. (2020) kaynaklarına bakınız. Kaynakların tamamı sonda listelenmiştir.

BAŞLARKEN

İndirme ve kurulum

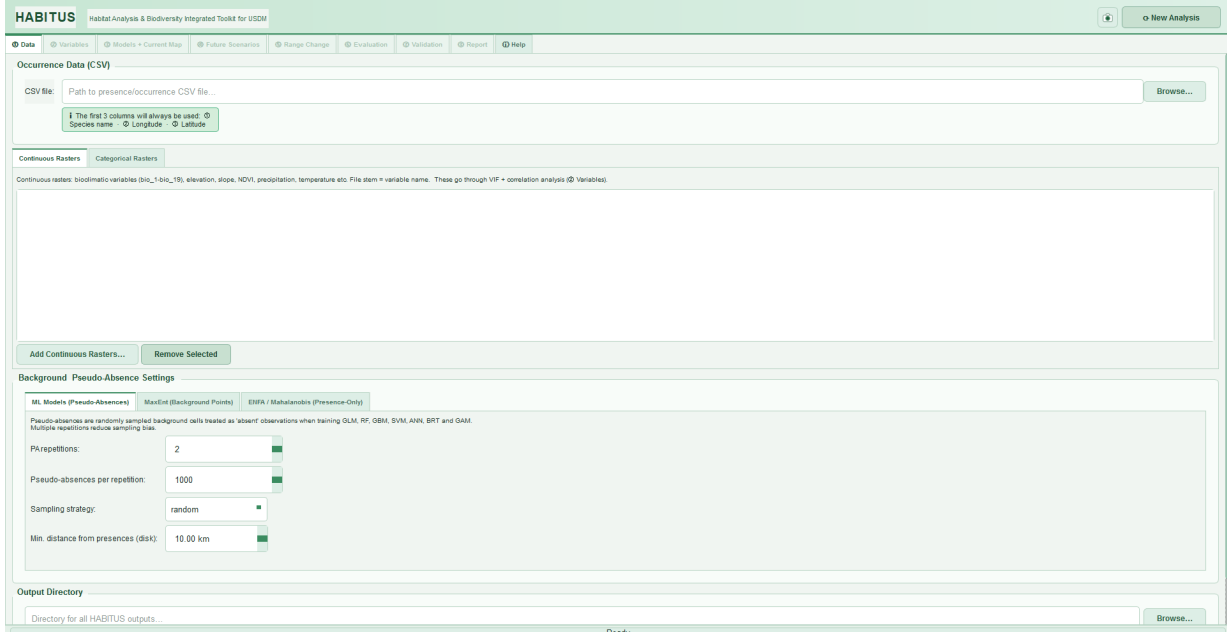
HABITUS kendi kendine yeten bir kurulum dosyası olarak dağıtılır. Başka hiçbir şey kurmanız gerekmez; Python çalışma zamanı ve bütün bilimsel kütüphaneler paketin içindedir. İşletim sisteminize uygun kurulum dosyasını şu adresten indirin:

<https://github.com/omerorucu/habitus/releases/latest>

Windows'ta HABITUS_Setup_v1.0.0.exe dosyasını çalıştırıp yönergeleri izleyin. Kurulum dosyası ve uygulama kod imzalıdır; bu nedenle Windows SmartScreen "bilinmeyen yayıncı" uyarısı yerine yayıncı adını gösterir. macOS'ta .dmg dosyasını açıp HABITUS'u Uygulamalar klasörüne sürükleyin; Linux'ta AppImage dosyasına çalıştırma izni verip başlatın.

İlk açılış

HABITUS'u Başlat menüsünden, Uygulamalar klasöründen ya da AppImage dosyasından başlatın. Şu pencere açılır:



Şekil 1. HABITUS'un açılış penceresi. Yalnızca Veri sekmesi ile Yardım sekmesinin etkin olduğuna dikkat edin; kalan yedi aşama, önkoşulları sağlanana kadar soluk görünür.

Üstteki sekiz numaralı sekme iş akışının kendisidir ve her aşamayı tamamladıkça sırayla açılır. Yardım sekmesi her zaman etkindir; tam kullanım kılavuzunu, önerilen en küçük örneklem büyüklüklerini ve atıf bilgisini içerir. Alttaki panel ise gündüktür: HABITUS'un yaptığı her şey zaman damgasıyla oraya yazılır ve aynı günlük çıktı klasörünüze kaydedilir.

Note. Önceki bir aşamayı değiştirirseniz — örneğin başka bir CSV yüklerseniz — sonraki bütün sonuçlar otomatik olarak temizlenir. Bu bilinçli bir davranıştır: bir veri setinden üretilmiş haritanın, başka bir veriden hesaplanmış istatistiklerle birlikte raporlanmasını engeller.

GEREKEN VERİ

İki şey gerekir. Birincisi, virgülle ayrılmış bir tür kayıt dosyası. HABITUS her zaman ilk üç sütunu şu sırayla kullanır: tür adı, boylam, enlem. Sütun başlıkları herhangi bir şey olabilir; önemli olan yalnızca sıradır. Bizim dosyamız şöyle başlıyor:

```
presence,long,lat
Pinus brutia,35.35278,37.59167
Pinus brutia,35.38333,37.54167
Pinus brutia,30.40833,40.38667
```

İkincisi, çevresel raster katmanlar. HABITUS ESRI ASCII (.asc) ve GeoTIFF biçimlerini okur. Bütün katmanların coğrafi sınırları ve hücre boyutu aynı olmalıdır. Dosya adı değişken adına dönüşür; yani bio_ (15).asc dosyası bio_15 olarak okunur. Güncel ve gelecek katman kümelerinde adları tutarlı tutun, çünkü HABITUS eşleştirmeyi ada göre yapar.

Katmanlar sürekli ya da kategorik olarak tanıtılır. Bitki örtüsü sınıfı, toprak tipi, arazi örtüsü gibi kategorik değişkenler metin değil tam sayı olarak kodlanmalı ve ikinci sekmeye eklenmelidir; böylece yazılım bunları sayısal gradyan gibi işlemez.

Note. Örneklem büyüklüğü, herhangi bir modelleme tercihinden daha belirleyicidir. Otuzun altındaki kayıt sayılarında modeller kararsızlaşır; küçük örneklerle nelerin yapılabileceğini Pearson ve ark. (2007) tartışır. Yardım sekmesinde algoritma başına önerilen en küçük değerleri gösteren bir tablo bulunur.

VERİYİ YÜKLEME

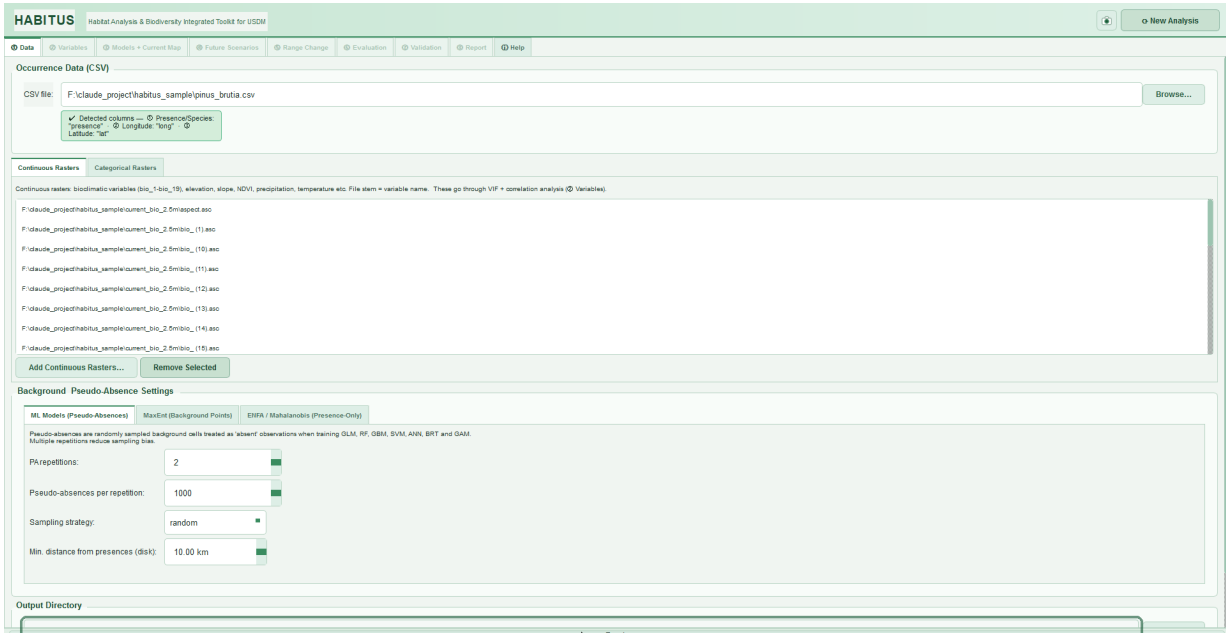
Veri sekmesini açın ve "CSV file" satırındaki **Browse...** düğmesine basarak tür kayıt dosyanızı seçin. HABITUS başlık satırını hemen okur ve ne bulunduğunu bildirir:



Şekil 2. CSV seçildikten sonra yeşil şerit, hangi sütunun tür adı, boylam ve enlem olarak yorumlandığını doğrular. Sütunlarınız farklı sıradaysa bunu, sonradan tuhaf bir haritayla fark etmek yerine burada görürsünüz.

Ardından çevresel katmanları ekleyin. **Add Continuous Rasters...** düğmesine basıp 23 dosyanın tamamını tek seferde seçebilirsiniz. Kategorik katmanlarınız varsa "Categorical Rasters" sekmesine geçip onları oraya ekleyin.

Son olarak çıktı klasörünü belirleyin. HABITUS bütün haritaları, tabloları, şekilleri ve günlüğü bu klasöre yazar. Dokunmazsanız Belgeler klasörünüzün altında tür adı ve zaman damgasıyla bir klasör otomatik oluşturulur; ardışık çözümlemeleri birbirinden ayrı tuttuğu için genellikle istenen de budur.



Şekil 3. Veri sekmesi tümüyle hazır: tür kayıt dosyası, 23 sürekli katman ve arka plan örnekleme ayarları. Altteki üç alt sekme, üç model ailesi için birbirinden bağımsız ayarlar tutar.

ARKA PLAN VE SÖZDE-YOKLUK

Tür kayıtları size türün nerede *bulunduğunu* söyler, nerede *bulunmadığını* asla. Buna karşılık çoğu algoritma varlıkları karşılaştıracak bir kümeye ihtiyaç duyar ve bu küme üretilmek zorundadır. Bütün iş akışının en sonuç belirleyici kararlarından biridir (Barbet-Massin ve ark., 2012); HABITUS'un buna gizli bir varsayılan yerine ayrı bir panel ayırmasının sebebi budur.

Üç alt sekmenin bulunma nedeni, üç model ailesinin farklı şeyler istemesidir:

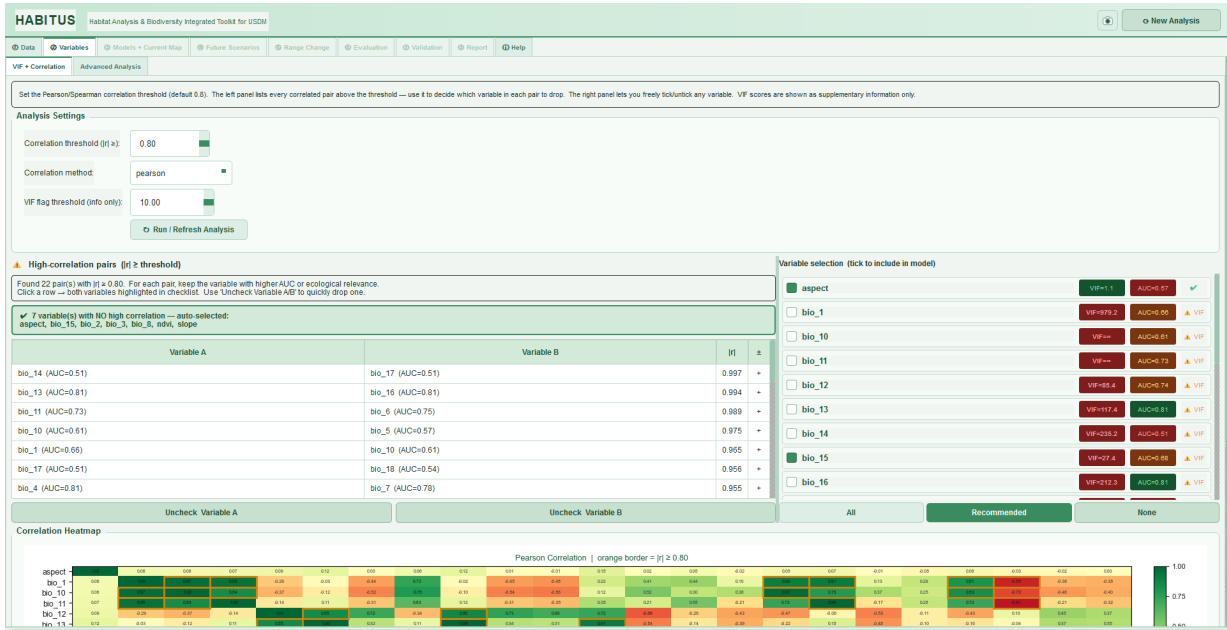
ML Models (Pseudo-Absences) — GLM, RF, GBM, SVM, ANN, BRT, GAM ve gradyan artırma algoritmaları, örneklenen arka plan hücrelerini yokluk gibi kullanır. Kaç yineleme çekileceğini (*PA repetitions*) ve yineleme başına kaç nokta alınacağını belirleyin. İki ya da daha fazla yineleme çekmek, sonucunuzun tek bir rastgele çekilişe bağlı olup olmadığını görmeyi sağlar.

MaxEnt (Background Points) — MaxEnt yokluk kullanmaz; varlıkları, bölge genelinde mevcut çevresel koşullarla karşılaştırır. Bu koşullar tipik olarak 10.000 arka plan noktasıyla temsil edilir (Phillips ve ark., 2006).

ENFA / Mahalanobis (Presence-Only) — bu yöntemler yalnızca varlıkların kendi oluşturduğu çevresel zarfı betimler.

Üç örnekleme stratejisi vardır. *Random*, çalışma alanının herhangi bir yerinden çeker. *Disk*, bilinen varlıkların çevresine bir tampon koyarak dışlar; kayıtlarınız mekânsal olarak kümelenmişse yerinde bir seçimdir. *SRE* (yüzey aralık zarfı) yalnızca gözlenen zarfın dışındaki çevresel koşullardan çeker ve daha temkinli bir karşılaştırma kümesi üretir.

Ayarlar tamamlandınca ► **Load Data & Generate Background/Pseudo-Absence Points** düğmesine basın. HABITUS her kayıt için çevresel değerleri çıkarır; raster sınırlarının dışında kalan, boş hücreye düşen ya da mevcut bir koordinatı yineleyen noktaları eler ve kaç kaydın hayatta kaldığını tam olarak bildirir:

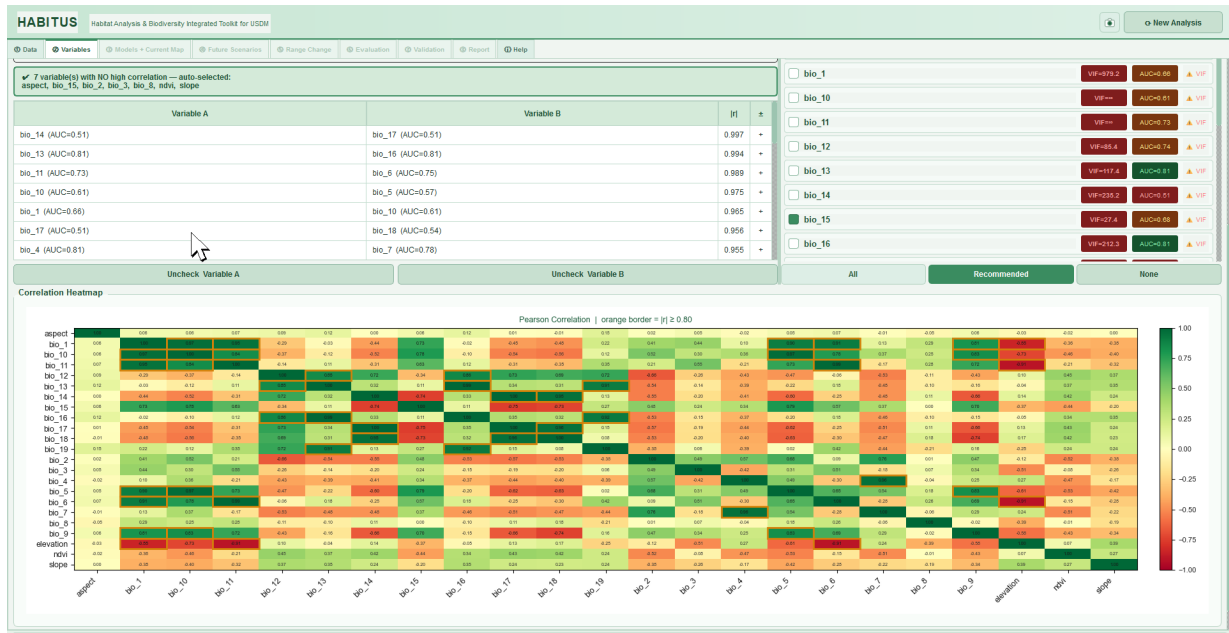


Şekil 4. Veri yükleme tamamlandı. 108 kaydın tamamı çözümlemeye girdi; 1.000'er noktalık iki sözde-yokluk yinelemesi ve 10.000 MaxEnt arka plan noktası üretildi. Değişkenler sekmesi açıldı ve çözümlemesi çoktan çalıştı.

DEĞİŞKEN SEÇİMİ

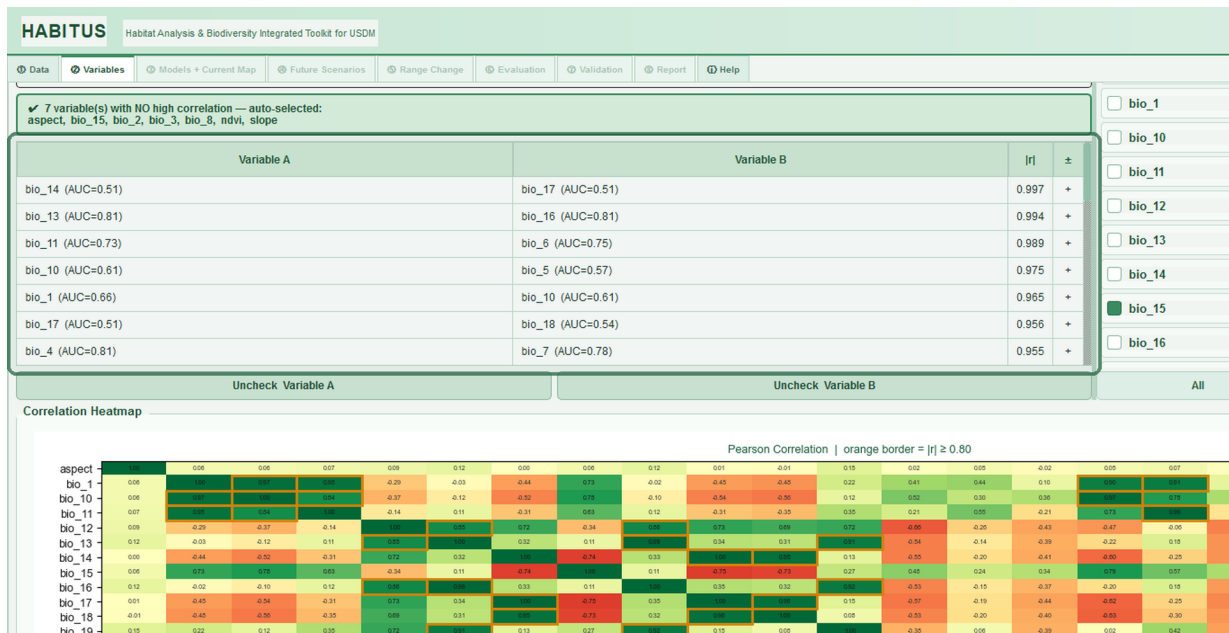
Biyoiklim değişkenleri aynı aylık sıcaklık ve yağış yüzeylerinden türetildiği için tanımı gereği birbiriyle güçlü ilişkilidir: en soğuk çeyreğin ortalama sıcaklığı ile yıllık ortalama sıcaklık neredeyse aynı bilgiyi taşır. Hepsini modele vermek katsayıları kararsızlaştırır ve değişken önemini yanıltıcı kılar (Dormann ve ark., 2013). Değişkenler sekmesi tam da bunu önlemek için vardır.

Taramayı iki eşik yönetir. **Korelasyon eşiği** (varsayılan $|r| \geq 0,80$) hangi çiftlerin işaretleneceğini belirler; **VIF eşiği** (varsayılan 10) ise varyansı çoklu bağlantı yüzünden şişmiş değişkenleri damgalar (O'Brien, 2007). Hesaplamak için **Run / Refresh Analysis** düğmesine basın.



Şekil 5. Korelasyon ısı haritası. Biyoiklim değişkenlerinde birbiriyle ilişkili blokların çıkması normaldir ve beklenir — tarama adımı zaten bu deseni ele almak için tasarlanmıştır.

Ayarların altında HABITUS, eşği aşan her çifti ve her iki üyenin tek değişkenli AUC değerini listeler; yani her değişkenin tek başına varlığı arka plandan ne kadar iyi ayırdığını. Bu, çiftin hangisini tutacağınıza karar verirken nesnel bir dayanak sağlar.



Şekil 6. Yüksek korelasyonlu çiftler. Bu veri setinde 22 çift $|r| \geq 0,80$ eşliğini aştı. Bir satıra tıklamak her iki değişkeni de sağdaki listede vurular; "Uncheck Variable A/B" düğmeleri seçilen birini anında listeden düşürür.

Sağdaki panel asıl seçimin yapıldığı yerdir. Üç düğme bu panel üzerinde çalışır: **All**, **None** ve **Recommended**. Sonuncusu, hiçbir yüksek korelasyon çiftine girmeyen değişkenleri seçer; savunulabilir bir başlangıç noktasıdır ama ekolojik muhakemenin yerini tutmaz. Belirli bir değişkenin türünüz için fizyolojik olarak anlamlı olduğunu biliyorsanız onu tutup ilişkili olduğu değişkeni düşürün.

Note. VIF=999 olarak işaretlenen bir değişken yalnızca ilişkili değil, neredeyse tam olarak diğerlerinin doğrusal bileşimidir. HABITUS $1,9 \times 10^{13}$ gibi bir sayı basmak yerine değeri sınırlar; ancak böyle bir değişken hemen her durumda çıkarılmalıdır.

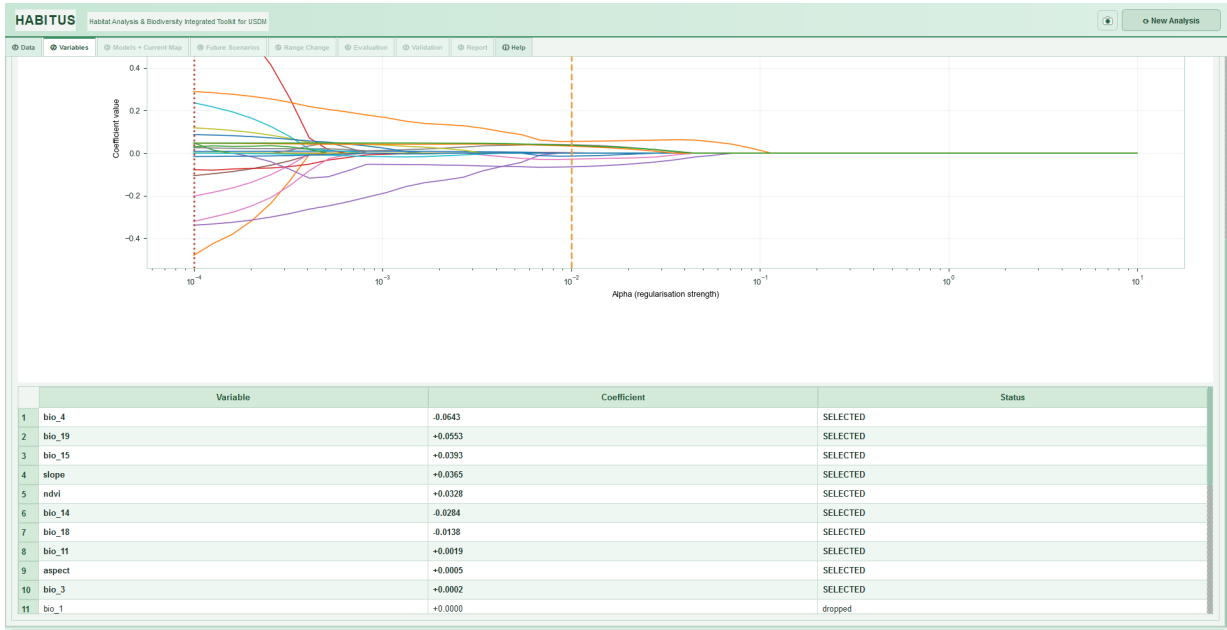
İLERİ TARAMA

VIF ve korelasyon ikili tanılardır; birden çok değişkeni birlikte ilgilendiren yapıyı kaçırabilirler. **Advanced Analysis** alt sekmesine geçip **Run All Analyses** düğmesine basın; dört ek denetim çalışır.

Koşul sayısı — bütün değişken matrisinin ne kadar kötü koşullandığını tek bir sayıyla verir (Belsley ve ark., 1980). Otuzun üzerindeki değerler, ikili korelasyonlar ne gösterirse göstereceği ciddi çoklu bağlantıya işaret eder.

PCA — değişkenlerinizin gerçekte kaç boyut kapladığını gösterir. İlk iki üç bileşen varyansın çoğunu açıklıyorsa, 23 katmanınız belki de yalnızca üç bağımsız şeyi ölçüyordur.

LASSO (L1) — katsayıları sıfıra doğru çeker ve gereksiz olanları tam olarak sıfırlar; yani seçimi kendisi yapar (Tibshirani, 1996). Alfa değeri çapraz doğrulamayla belirlenir.



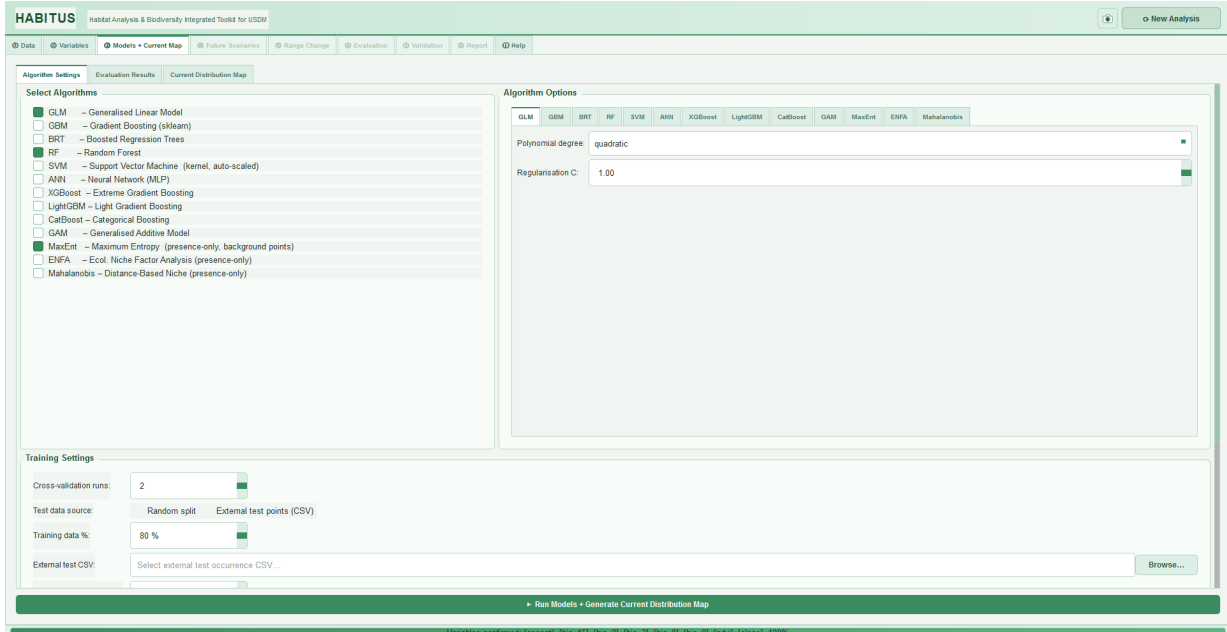
Şekil 7. LASSO düzenleme yolu. Her çizgi, ceza arttıkça bir değişkenin katsayısını gösterir; kesikli dikey çizgi çapraz doğrulamayla seçilen alfa değeridir. Altteki tablo hangi değişkenlerin kaldığını (SELECTED) ve hangilerinin sıfırlandığını (dropped) bildirir.

Ridge (L2) — katsayıları küçültür ama hiçbir zaman sıfırlamaz (Hoerl ve Kennard, 1970). İlişkili değişkenleri, bir çiftin rastgele birini seçmek yerine kararlı tuttuğu için Ridge'i LASSO ile birlikte okumak her ikisini ayrı ayrı okumaktan daha bilgilendiricidir: LASSO'nun düşürdüğü ama Ridge'in kayda değer katsayıyla tuttuğu değişkenler ikinci bir bakış hak eder.

Dört tanı aynı yönü işaret ettiğinde seçiminize makul ölçüde güvenebilirsiniz. İlk alt sekmeye dönüp **Confirm Variable Selection** → **Proceed to Models** düğmesine basın. Örneğimizde sekiz değişken kaldı: aspect, bio_2, bio_3, bio_8, bio_9, bio_15, NDVI ve slope.

MODELLERİ EĞİTME

Modeller sekmesi üç ailede on üç algoritma sunar. İstediklerinizi işaretleyin; her birinin, hiperparametrelerin (ağaç sayısı, öğrenme oranı, derinlik, çekirdek vb.) ayarlanabildiği kendi ayar alt sekmesi vardır. Varsayılanlar literatürde yaygın kullanılan değerleri izler ve makul bir başlangıç noktasıdır.



Şekil 8. Algoritma seçimi. Bu kılavuzda her aileden birer temsilci kullanıyoruz: GLM (klasik regresyon), RF (makine öğrenmesi) ve MaxEnt (yalnız varlık).

İki ayar daha önemlidir. **Eğitim/test bölmesi** tabakalıdır: varlıklar ve arka plan noktaları ayrı ayrı bölünür, böylece ayrılan test kümesi eğitim kümesiyle aynı oranı korur. **Ensemble** ise kurulan modelleri birleştirir; ya her modelin başarı puanıyla orantılı katkı verdiği ağırlıklı ortalama ya da ikili tahminlerin çoğunluk oyuyla toplandığı komite ortalaması olarak. Ensemble genellikle tek bir modelden daha kararlıdır (Araújo ve New, 2007).

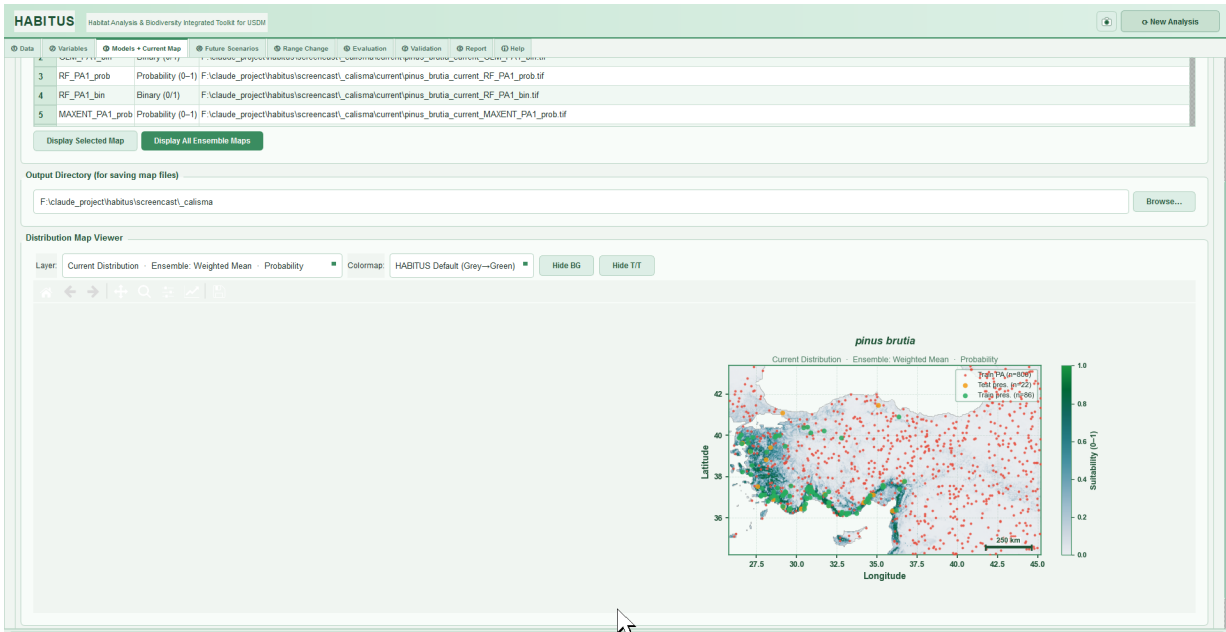
► **Run Models + Generate Current Distribution Map** düğmesine basın. Her algoritma her sözde-yokluk yinelemesi için eğitilir, çapraz doğrulanır, puanlanır ve son olarak güncel iklim katmanlarına yansıtılır. Algoritma sayısına ve raster boyutuna göre bu adım bir dakikanın altından birkaç on dakikaya kadar sürebilir; ilerleme durum çubuğunda ve günlükte bildirilir.

HABITUS						
Habitat Analysis & Biodiversity Integrated Toolkit for USDM						
Data	Variables	Models	Current Map	Future Scenarios	Range Change	Evaluation
1	USDM		PA1			PA1
3	RF		PA1			PA1
4	RF		PA1			PA1
5	MAXENT		PA1			PA1
6	MAXENT		PA1			PA1
7	GLM		PA2			PA2
8	GLM		PA2			PA2
9	RF		PA2			PA2
10	RF		PA2			PA2
11	MAXENT		PA2			PA2
12	MAXENT		PA2			PA2
13	GLM		PA1			PA1
14	RF		PA1			PA1
15	MAXENT		PA1			PA1
16	GLM		PA2			PA2
17	RF		PA2			PA2
18	MAXENT		PA2			PA2

Şekil 9. Değerlendirme sonuç tablosu. Her algoritma × sözde-yokluk yinelemesi × çapraz doğrulama çalıştırması, kendi AUC, TSS ve Boyce değeriyle ayrı bir satır olarak görünür; ensemble satırları en son da yer alır.

TAHMİNİ BAKMAK

Current Distribution Map alt sekmesine geçin. Ensemble tahmini, tür kayıtları üzerine bindirilmiş olarak otomatik görüntülenir.



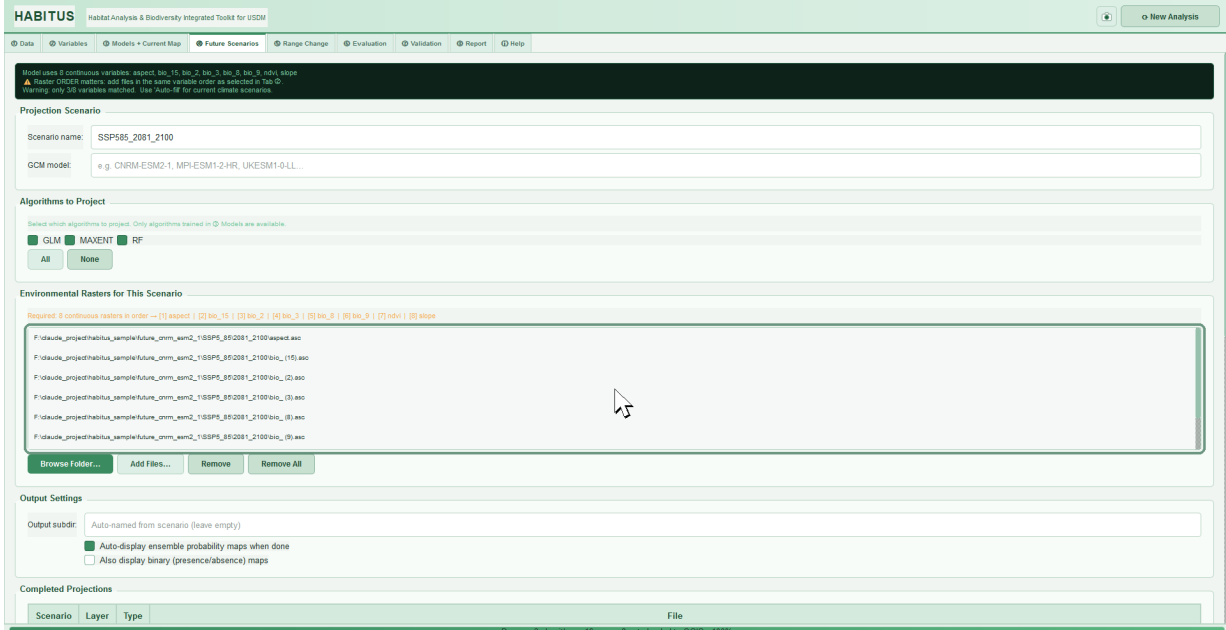
Şekil 10. Gömülü harita görüntüleyici. Katman seçici; tek tek algoritmalar, sözde-yokluk yinelemeleri, ensemble sonuçları ve bunların olasılık ile ikili sürümleri arasında geçiş yapar.

Birkaç denetim bilmeye değer. **Colormap** seçici, renk körlüğüne duyarlı (Viridis) ve ekoloji standardı (YlGnBu) seçenekleri de içeren sekiz palet sunar. **Hide BG** ve **Hide T/T** düğmeleri arka plan noktalarını ve eğitim/test varlık noktalarını gizler. Araç çubuğundaki kaydetme simgesi görünümü 300 DPI olarak, tür ve katman adıyla otomatik adlandırarak yazar.

Her harita ayrıca çıktı klasörünüze GeoTIFF olarak yazılır; aynı katmanı QGIS ya da ArcGIS'te açıp kartografik olarak bitirebilirsiniz. Her model için iki sürüm üretilir: **_prob** (sürekli uygunluk, 0-1) ve **_bin** (eşikleme sonrası varlık/yokluk).

GELECEĞE PROJEKSİYON

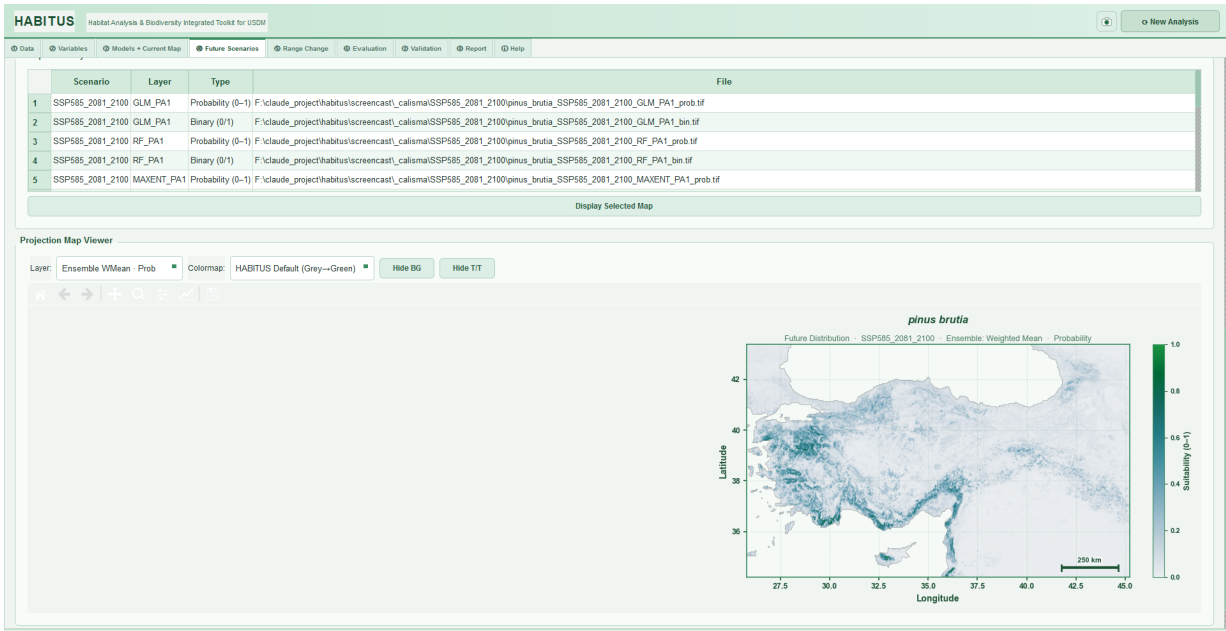
Gelecek Senaryoları sekmesi, kurduğunuz modelleri farklı bir iklim katmanı kümesine uygular. Senaryoya bir ad verin — bu ad hem çıktı klasörünü etiketler hem de alan değişimi sekmesinde görünür — sonra **Browse Folder...** ya da **Add Files...** ile katmanları ekleyin.



Şekil 11. SSP5-8.5, 2081–2100 senaryosunun katmanları ekleniyor. Yalnızca modelin gerçekten kullandığı değişkenler, aynı sırayla ekleniyor.

Note. Hataya en açık adım budur. Katman kümesi, modelin eğitildiği değişkenlerle *tam olarak* — hem içerik hem sıra bakımından — örtüşmelidir. Aksi hâlde HABITUS çalışmayı reddeder ve hangi değişkenleri beklediğini söyler. Kategorik değişkenlerin anlamlı bir gelecek karşılığı olmadığı için bunlar eğitim verisinden otomatik olarak yeniden kullanılır; buraya eklemeyin.

Hangi algoritmaların yansıtılacağını seçin (genellikle **All**) ve ► **Run Projection** düğmesine basın. Her algoritma için ayrı, ayrıca ensemble için birer harita üretilir; hepsi senaryo adını taşıyan bir alt klasöre GeoTIFF olarak yazılır.

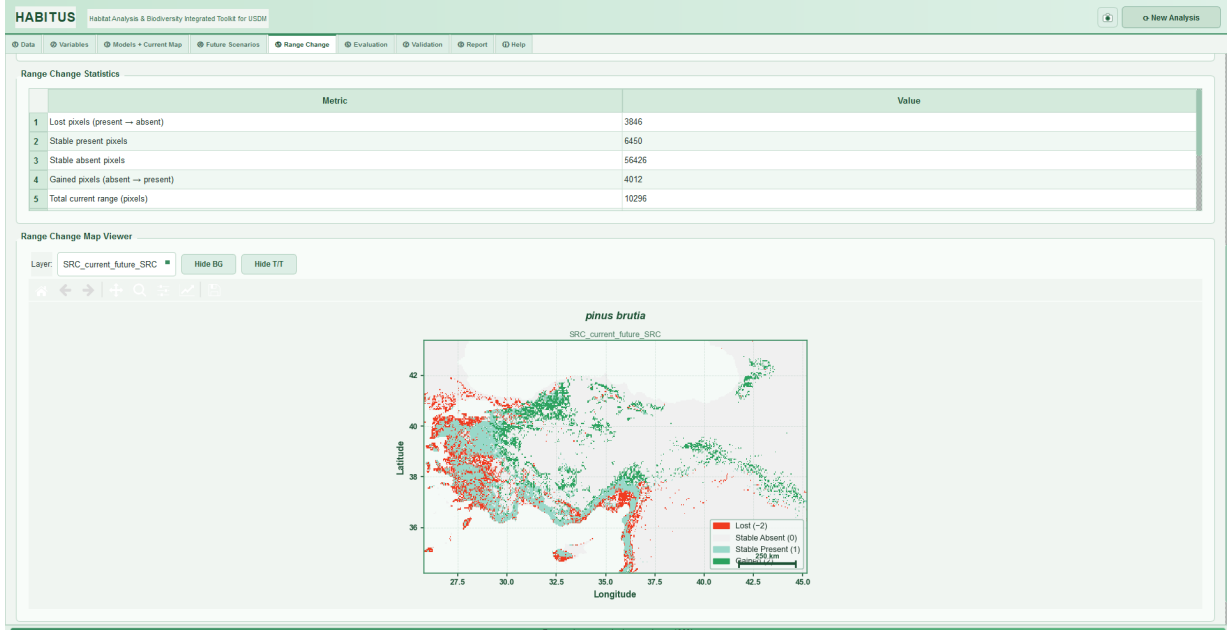


Şekil 12. SSP5-8.5 senaryosu altında 2081-2100 için öngörülen dağılım. Şekil 10 ile karşılaştırıldığında uygunluğun batı alçak kesimlerinde daraldığı, iç kesimlere ve kuzeye doğru genişlediği görülür.

İstediğiniz kadar senaryo-dönem birleşimi ekleyebilirsiniz. Her biri ayrı saklanır ve sonraki adımda bunlardan herhangi ikisi karşılaştırılabilir.

ALAN DEĞİŞİMİ

Alan Değişimi sekmesi iki ikili haritayı hücre hücre karşılaştırır. İki açılır listeden güncel ve gelecek katmanları seçin — listede yalnızca ikili (_bin) katmanlar görünür, çünkü alan değişimi sürekli uygunluğa değil varlık-yokluk kararına dayanır — sonra **Fill Paths** ve ► **Compute Range Change** düğmelerine basın.



Şekil 13. Güncel iklim ile SSP5-8.5 (2081-2100) arasındaki alan değişimi. Kırmızı kaybedilen, koyu yeşil kazanılan, açık turkuaz ise her iki dönemde de uygun kalan alanı gösterir.

İstatistik tablosu her sınıfın piksel sayısını bildirir. Örneğimizde: hâlihazırda 10.296 hücre uygun, öngörülen 10.462 hücre, 6.450 hücre sabit; %37,4 kayıp, %39,0 kazanç.

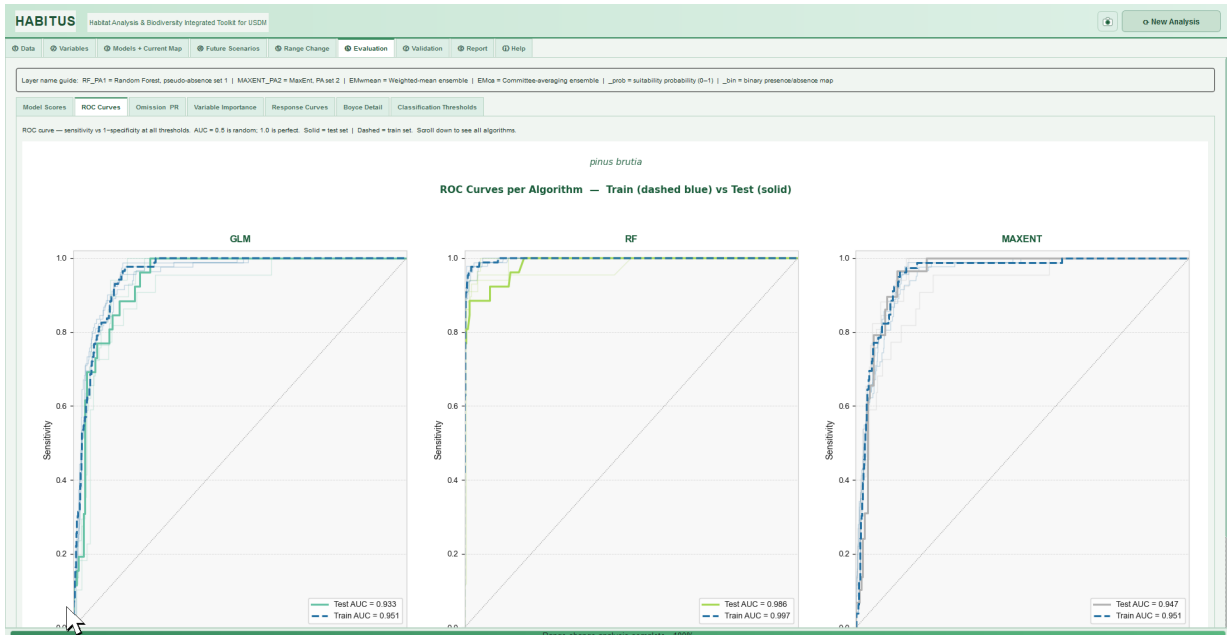
Note. Bu iki sayıyı birlikte okuyun. *Net* değişim +%1,6'dır ve tek başına türün neredeyse etkilenmediğini düşündürür. *Devir* ise %40'a yakındır; uygun alanın üçte birinden fazlası yer değiştirir. Yalnızca net değeri raporlamak yanıltıcı olur; koruma açısından asıl mesajı Şekil 13'teki mekânsal desen taşır: alçak kesimlerde kayıp, iç kesimlerde ve yüksek rakımda kazanç.

Piksel sayılarını alana çevirmek için katmanlarınızın hücre alanıyla çarpın. 2,5 yay dakikalık çözünürlükte, 38° kuzey enlemi civarında bir hücre kabaca 21 km²'dir.

İSTATİSTİKSEL ÇÖZÜMLEME

Değerlendirme sekmesi yedi alt sekmelik tanı barındırır. Bir kez **Refresh All Charts** düğmesine basın; her şey ayrılan test verisinden hesaplanır.

Model Scores bütün algoritmaların AUC, TSS ve Boyce değerlerini tek grafikte karşılaştırır. **ROC Curves** ayrımı ayrıntılı gösterir:



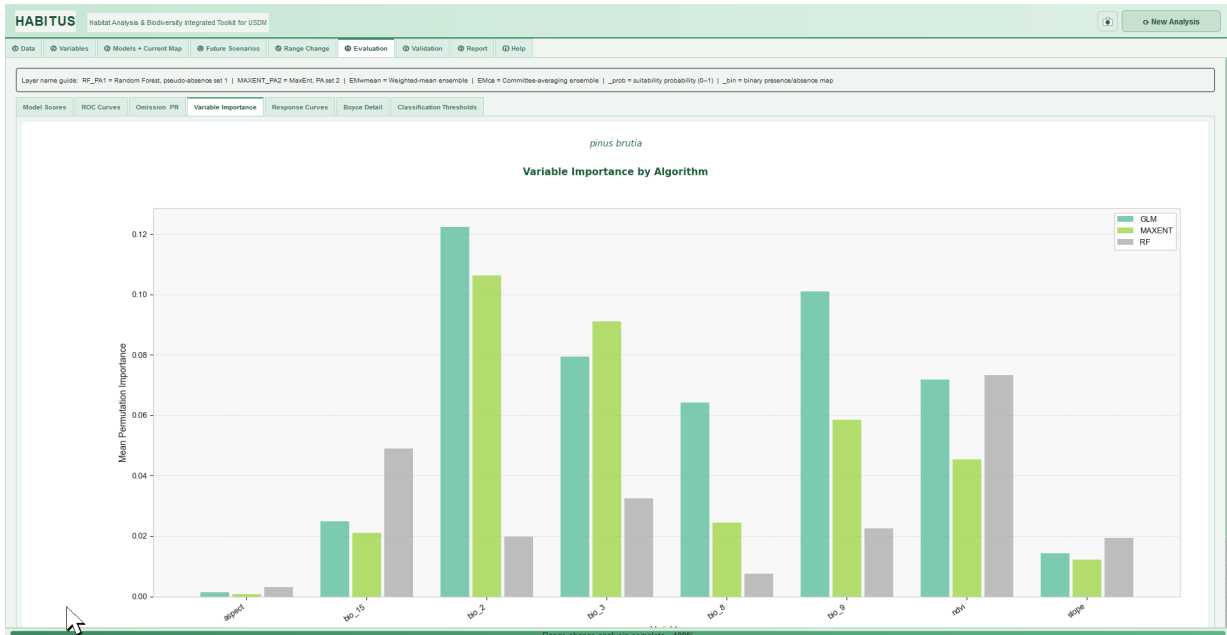
Şekil 14. Algoritma başına ROC eğrileri; eğitim (kesikli) ve test (sürekli) eğrileri üst üste çizilmiştir. Sürekli eğri sol üst köşeye ne kadar yakınsa o kadar iyidir; köşegen rastgele tahmindir. Kesikli ile sürekli arasındaki büyük açıklık aşırı uyuma işaret eder.

Üç ölçüt raporlanır ve farklı sorulara yanıt verirler. **AUC**, rastgele seçilen bir varlığın rastgele seçilen bir arka plan noktasının üstünde sıralanma olasılığıdır; eşikten bağımsız bir sıralama ölçüsüdür. **TSS** (duyarlılık + seçicilik – 1) eşiğe bağlıdır ama kappadan farklı olarak yaygınlıktan bağımsızdır (Allouche ve ark., 2006). **Sürekli Boyce indeksi** yalnızca varlık verisi ister; bu yüzden AUC'nin yorumlanmasının güç olduğu yalnız-varlık modelleri için uygun ölçüt olur (Boyce ve ark., 2002; Hirzel ve ark., 2006).

Note. Tek bir çapraz doğrulama katmanından gelen tek bir Boyce değerini, modelin bir özelliğiymiş gibi raporlamayın. Örneğimizde AUC on sekiz model gerçeklemesi boyunca yalnızca 0,054 aralıkta değişirken Boyce indeksi 0,261 ile 0,916 arasında savruldu — kötü bir modelle mükemmel bir model arasındaki farktan daha geniş bir aralık. HABITUS her gerçeklemeyi tam da bu değişkenlik görünsün diye gösterir; tek bir sayıyı değil, aralığı bildirir.

HANGİ DEĞİŞKENLER ÖNEMLİ

Variable Importance alt sekmesi önemi permütasyonla kestirir: bir değişkenin değerleri karıştırılır ve model başarısındaki düşüş ölçülür. Modelin yoğun olarak dayandığı bir değişken büyük bir düşüş üretir.



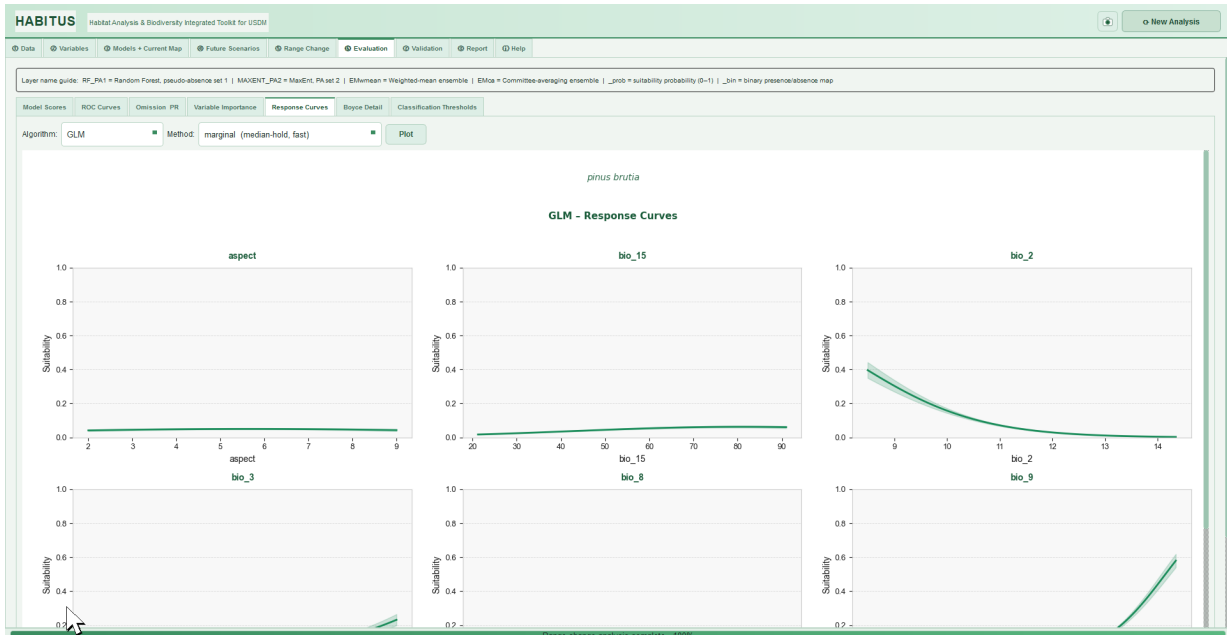
Şekil 15. Permütasyon temelli değişken önemi, her algoritma için ayrı gösterilmiştir. Algoritmalar arasındaki uyuşma — burada bio_2 ve bio_9'da görüldüğü gibi — tek bir modeldeki yüksek puandan daha ikna edicidir.

Değişkenler birbiriyle ilişkiliyken önem sıralamalarını temkinli değerlendirin. İki değişken aynı bilgiyi taşıyorsa, birini karıştırmak başarıyı neredeyse hiç etkilemeyebilir; çünkü diğeri açığı kapatır ve ikisi birden önemsiz görünür. Değişken seçimi adımının neden önce geldiğinin sebebi tam olarak budur.

TAHMİN DEĞİŞKENLERE NASIL BAĞLI

Önem, bir değişkenin *ne kadar* etkili olduğunu söyler; tepki eğrileri ise *hangi yönde* etkilediğini. **Response Curves** alt sekmesi açılır listeden seçilebilen üç yöntem sunar:

Marjinal, diğerlerini ortancada tutarak tek bir değişkeni değiştirir; hızlıdır ve okunması en kolaydır. **Kısmi bağımlılık**, diğer değişkenlerin gözlenen dağılımı üzerinden ortalama alır (Friedman, 2001) ve değişkenler etkileşimliyse daha titizdir. **ICE** ortalama yerine her gözlem için ayrı bir çizgi çizer (Goldstein ve ark., 2015); ortalama tepkinin, verinin farklı kesimlerindeki zıt davranışı gizlediği durumları açığa çıkarır.



Şekil 16. Tepki eğrileri. Ekolojik olarak makul olmayan biçimler — örneğin uç değerlerde sınırsız yükselen uygunluk — genellikle biyolojik bir bulgudan çok, örneklenen aralığın dışına genişletmenin işaretidir.

EŞİKLER

Sürekli bir uygunluk yüzeyini varlık/yokluk haritasına çevirmek bir eşik gerektirir ve bu seçim haritalanan alanı belirgin biçimde değiştirir. **Classification Thresholds** alt sekmesi her model için beş kural hesaplar:

HABITUS Habitat Analysis & Biodiversity Integrated Toolkit for USDM							
Layer name guide: RF_PA1 = Random Forest, pseudo-absence set 1 MAXENT_PA2 = MaxEnt, PA set 2 ERMmean = Weighted-mean ensemble ERMoe = Committee-averaging ensemble _prob = suitability probability (0-1) _bin = binary presence/absence map							
Model Scores ROC Curves Omission PR Variable Importance Response Curves Biotype Detail Classification Thresholds							
Algorithm: GLM Method: marginal (median-hold, fast) Plot							
Optimal binary classification thresholds for converting probability maps (0-1) to presence/absence maps. Each method uses a different statistical criterion. Choose based on your objective.							
Threshold Methods							
max_tss: Maximise TSS (Sensitivity + Specificity - 1). Prevalence-independent. Recommended for SDM (Allouche 2006)							
max_kappa: Maximise Cohen's Kappa. Prevalence-sensitive. (Fleiss & Bell 1997)							
sens_spec_eq: Equal Sensitivity & Specificity. Use when omission = commission error (Liu 2005)							
p10: 10th percentile of presence predictions. Minimal predicted area. Recommended for MaxEnt (Pearson 2007)							
min_roi: Minimum distance to perfect ROC point (0,1). Threshold-independent of prevalence.							
Algorithm	PA Set	max_tss	max_kappa	sens_spec_eq	p10	min_roi	Used (current)
1 GLM	PA1	0.1134 (metric=0.796)	0.3006 (metric=0.621)	0.1577 (metric=0.771)	0.1551 (metric=0.788)	0.1548 (metric=0.796)	0.1134
2 GLM	PA2	0.0937 (metric=0.802)	0.3547 (metric=0.612)	0.1676 (metric=0.778)	0.1548 (metric=0.782)	0.1542 (metric=0.791)	0.1282
3 MAXENT	PA1	0.1725 (metric=0.818)	0.5320 (metric=0.611)	0.2759 (metric=0.793)	0.2628 (metric=0.787)	0.1977 (metric=0.812)	0.2168
4 MAXENT	PA2	0.1725 (metric=0.818)	0.5271 (metric=0.611)	0.2759 (metric=0.793)	0.2635 (metric=0.787)	0.2471 (metric=0.800)	0.2464
5 RF	PA1	0.4187 (metric=0.982)	0.6305 (metric=0.952)	0.6010 (metric=0.981)	0.6638 (metric=0.892)	0.6100 (metric=0.981)	0.1036
6 RF	PA2	0.5025 (metric=0.986)	0.6010 (metric=0.965)	0.5616 (metric=0.981)	0.6640 (metric=0.903)	0.6280 (metric=0.985)	0.1134

Şekil 17. Algoritma başına hesaplanan beş eşik kuralı ve bunların ürettiği duyarlılık ile seçicilik değerleri.

max_TSS duyarlılık + seçicilik - 1 değerini en büyükler ve TDM'de standart varsayılandır. **max_Kappa** Cohen kappasını en büyükler. **sens_spec_eq** duyarlılığı seçiciliğe eşitler; atlama ve yanlış dahil etme hatalarının eşit maliyetli olduğu durumlar için uygundur (Liu ve ark., 2005). **P10**, varlık noktalarındaki tahmin değerlerinin onuncu yüzdeliğini kullanır; en düşük tahminli %10'luk kaydın hatalı olabileceğini kabul eder ve MaxEnt ile küçük örneklemelerde yaygındır (Pearson ve ark., 2007). **min_roi** ise kusursuz ROC köşesine en yakın noktayı alır.

Note. Hangi kuralı kullanırsanız kullanın, yöntem bölümünüzde belirtin ve eşik değerini raporlayın. Aynı türe ilişkin iki çalışma, yalnızca biri P10 değeri max_TSS kullandığı için alan büyüklüğünde anlaşmazlığa düşebilir.

BAĞIMSIZ DOĞRULAMA

Çapraz doğrulama iç tutarlılığı ölçer: modelin, kendisinin kurulduğu örneklemden gelen kayıtları ne kadar iyi kestirdiğini. Haritanın, türün yayılışına ilişkin bağımsız bir betimlemeyle uyuşup uyuşmadığını söyleyemez. Doğrulama sekmesi bunun içindir.

Açılır listeleri doldurmak için **Load HABITUS Output Layers...** düğmesine basın; sonra **Reference Map** alanını dış kaynağınıza, **Validation Map** alanını ise kendi çıktılarınızdan birine ayarlayın. Burada referans, EUFORGEN'in *Pinus brutia* yayılış haritasıdır — uzman görüşüyle çizilmiş ve model tarafından hiç görülmemiştir.

Note. Her iki açılır liste de aynı listeyle doldurulur; dolayısıyla aynı katmanı iki kez seçmek tümüyle mümkündür. Bunu yaparsanız genel doğruluk 1,000 ve kappa 1,000 çıkar ki bu hiçbir şey ifade etmez. Çalıştırmadan önce iki alanın farklı olduğunu doğrulayın.

EUFORGEN haritası zaten sıfır-bir biçiminde olduğu için **Use binary comparison** kutusunu işaretleyin. Sürekli bir referans için bu kutuyu boş bırakıp yeniden sınıflandırma eşiklerini belirleyin. Bir örnekleme deseni — rastgele, tabakalı, sistematik ya da kendi nokta CSV'niz — ve bir nokta sayısı seçip **Run Validation** düğmesine basın.

HABITUS Habitat Analysis & Biodiversity Integrated Toolkit for USDM

1 Map Selection

Reference Map:

Validation Map:

2 Reclassification (continuous 0-1 → classes)

Continuous (reclassify 0-1 → N classes)

Class 1 = 0, lower threshold. The remaining range is divided equally. Set the threshold from Classification Thresholds (max_TSS, max_Kappa, etc.). Example: 5 classes, threshold 0.20 → 0-0.20, 0.20-0.40, 0.40-0.60, 0.60-0.80, 0.80-1.0

Number of classes:

Reference — lower threshold:

Validation — lower threshold:

☐ Binary mode active - reclassification skipped.

Binary comparison

Binary mode compares two raster maps whose cells are already categorical — typically 0 = absent and 1 = present (binary presence/absence maps).

No reclassification or threshold is applied: the raw integer classes are compared cell-by-cell. Use this mode when you have an expert-delineated polygon converted to a raster, a published SDM binary output, or a field-validated presence/absence layer.

☐ Use binary comparison (0/1 maps)

3 Sampling Settings

Random ☒ Stratified ☐ Systematic ☐ CSV File ☐

Number of points:

4 Results

HABITUS Validation — BINARY Classification Accuracy Assessment

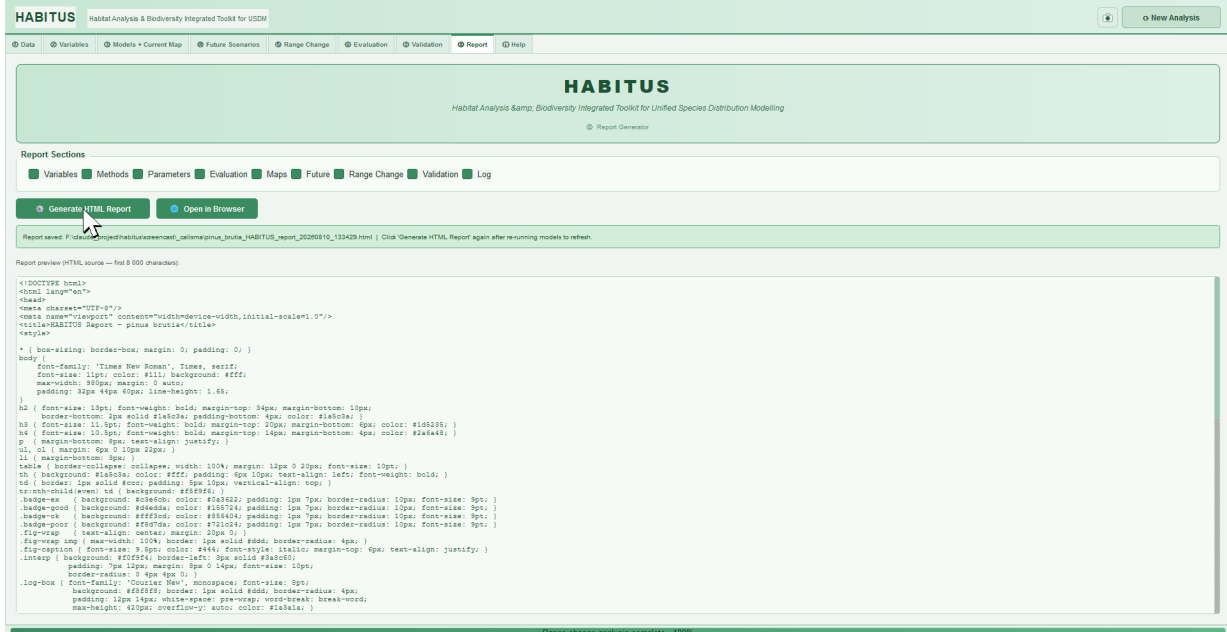
Mode	: Binary
Date	: 2026-08-10 13:34:19
Reference	: pinus_brutia_euforgen_distribution.tif
Validation	: pinus_brutia_current_EMmean_bin.tif
Classes	: 2 (binary / categorical - raw cell values)
Points	: 500 Method: Random
Time	: 0.1s

Şekil 18. Doğrulama çıktısı: genel doğruluk, F1 puanı ve Cohen kappası ile birlikte hata matrisi. Bizim ensemble modelimiz 500 rastgele noktada EUFORGEN'e karşı OA = 0,898 ve κ = 0,648 değerine ulaştı.

Vurgulanması gereken sayı kappadır. Genel doğruluk, sınıflardan biri baskın olduğunda şişer — çoğu dağılım haritasında yokluk ezici biçimde baskındır — oysa kappa tesadüfen beklenen uyuşmayı düşer (Cohen, 1960; Fielding ve Bell, 1997). 0,9'un üzerinde bir iç AUC ile 0,65'lik bir dış kappa çelişki değildir: iki istatistik farklı sorulara yanıt verir.

RAPOR

Rapor sekmesi her şeyi on bölümlük tek bir HTML belgesinde toplar: özet, değişken seçimi, yöntem, model parametreleri, değerlendirme sonuçları, güncel dağılım, gelecek projeksiyonları, alan değişimi, doğrulama ve oturum kaydı. İstemediğiniz bölümün işaretini kaldırıp **Generate HTML Report** düğmesine basın.



Şekil 19. Üretilen rapor. Her ölçüt literatüre dayalı bir yorumla birlikte verilir ve bütün şekiller base64 olarak gömüldüğü için rapor kendi kendine yeten tek bir dosyadır.

Asıl değeri tekrarlanabilirliktir. ODMAP gibi bir protokol kapsamında (Zurell ve ark., 2020) bir hakemin soracağı parametre ayarları, eşikler, örneklem büyüklükleri ve tanılar; aylar sonra hatırlanmaya çalışılmak yerine çözümlemenin çalıştığı anda otomatik olarak kaydedilir. Dosyayı tarayıcıda açıp okuyun ve kalıcı kayıt için PDF olarak yazdırın; gömülü şekiller 300 DPI olduğundan doğrudan bir makaleye taşınabilir.

ÇIKTI DOSYALARI

Her şey, Veri adımıyla seçtiğiniz çıktı klasörüne yazılır:

```
current/      güncel iklim GeoTIFF'leri (model başına _prob ve _bin)
<senaryo>/    her gelecek projeksiyonu için bir klasör
figures/      300 DPI değerlendirme grafikleri
csv/          puanlar, eşikler, VIF sıralaması, ilişkili çiftler
validation/   yeniden sınıflandırılmış referans ve doğrulama rasterları
occurrence_train.csv, occurrence_test.csv   kullanılan bölme
*.log         algoritma başına süreleri içeren zaman damgalı günlük
*.report.html kendi kendine yeten rapor
```

Eğitim/test CSV dosyaları ayrıca anılmayı hak eder: hangi kayıtların model kurmakta kullanıldığını, hangilerinin ayrıldığını tam olarak belgelerler; böylece bir hakem değerlendirmenizi birebir yineleyebilir.

SIK YAPILAN HATALAR

Projeksiyonda katman uyumsuzluğu. Uygulamalı TDM'de en sık yapılan hata, modeli eğitim kümesinden içerik ya da sıra bakımından farklı bir katman yığınına yansıtmaktır. HABITUS bu durumda çalışmayı reddeder; "Raster count mismatch" iletisini görürseniz yazılımın listelediği değişkenlerle Değişkenler adımıyla onayladıklarınızı karşılaştırın.

Yalnızca net alan değişimini raporlamak. Yukarıda görüldüğü gibi, sıfıra yakın bir net değişim %40'a yaklaşan bir devri gizleyebilir. Kayıp, kazanç ve sabit alanı ayrı ayrı bildirin.

Kendi çıktınıza karşı doğrulama. Bir haritayı kendisiyle karşılaştırmak kusursuz uyuma verir ve hiçbir bilgi taşımaz.

Tek bir katmana güvenmek. Özellikle küçük örneklemelerde Boyce indeksi için tek bir değer yerine yinelemeler arasındaki yayılımı bildirin.

Modelin bilemeyeceğini göz ardı etmek. Bunlar gerçekleşen nişi betimleyen ilişkisel modellerdir. Niş korunumunu varsayarlar; yayılma yeteneğini, biyotik etkileşimleri, toprağı ya da yangın gibi bozunma rejimlerini hesaba katmazlar. Seksen yıl sonrası için yeni uygun alan gösteren bir projeksiyon, türün oraya ulaşacağı anlamına gelmez.

KULLANIM KOŞULLARI

HABITUS akademik kullanım için MIT lisansı altında ücretsiz ve açık kaynaktır. Windows, macOS ve Linux kurulum dosyaları, kaynak kod ve bu kılavuz boyunca kullanılan *Pinus brutia* örnek veri seti şu adreste bulunabilir:

<https://github.com/omerorucu/habitus>

HABITUS bir yayına katkı sağlarsa lütfen şu şekilde atıf yapın:

Note. Örucü, Ö. K., & Örucü, S. (2026). HABITUS: Habitat Analysis and Biodiversity Integrated Toolkit for Unified Species Distribution Modelling. *Ecological Perspective*, [Technical Report]. <https://doi.org/10.53463/ecopers.20260435>

Sorular, hata bildirimleri ve özellik istekleri için depodaki konu takip sistemini kullanabilirsiniz.

KAYNAKLAR

- Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43(6), 1223-1232.
- Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, 22(1), 42-47.
- Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, 3(2), 327-338.
- Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.
- Boyce, M. S., Vernier, P. R., Nielsen, S. E., & Schmiegelow, F. K. A. (2002). Evaluating resource selection functions. *Ecological Modelling*, 157(2-3), 281-300.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46.
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., ... Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27-46.
- Elith, J., & Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40, 677-697.
- Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24(1), 38-49.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*, 29(5), 1189-1232.
- Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics*, 24(1), 44-65.

- Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8(9), 993–1009.
- Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, 199(2), 142–152.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67.
- Liu, C., Berry, P. M., Dawson, T. P., & Pearson, R. G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, 28(3), 385–393.
- O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41(5), 673–690.
- Pearson, R. G., Raxworthy, C. J., Nakamura, M., & Peterson, A. T. (2007). Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, 34(1), 102–117.
- Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3–4), 231–259.
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B*, 58(1), 267–288.
- Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., ... Merow, C. (2020). A standard protocol for reporting species distribution models. *Ecography*, 43(9), 1261–1277.