

生き物の分布推定ツール オープン SDM

原題 : Wallace Ecological Modeling Application Vignette

Jamie M. Kass, Sarah I. Meenan, Gonzalo E. Pinilla-Buitrago, Cory Merow, Robert P. Anderson

06/08/2018; 07/12/18 更新

翻訳

石濱史子, 小出大

作図

趙学群

09/29/2023 ; 10/27/2025 更新

本文書は、R 言語のパッケージ wallace の解説文書 vignette v.1 を国立環境研究所が翻訳・解説を加筆するとともに、解析事例を国内向けに差し替えたものです。このファイルの元の文書は、以下の URL にあります。

https://wallaceecomod.github.io/vignettes/wallace_vignette.html

序文

本文書は *Wallace* v. 1.0.5 を日本語に翻訳し、バイアス補正機能を追加したオープン SDM v.1.1.3-1.3 についてのものであるため、異なるバージョンをお使いの場合は内容が一致しないことがあります。また、この文書およびその他の *Wallace* パッケージの内容は、今後の継続的な開発に従って定期的に更新されることが予想されます。

はじめに

オープン SDM は生態学的モデリングのための R をベースとした GUI アプリケーション *Wallace* を日本語訳したものであり、現在、生物種のニッチおよび分布に関するモデルの構築、評価および可視化を主眼としています。本文では、これらのモデルを種分布モデル (Species Distribution Model; SDM) と呼びます。SDM については、ここでは詳しい説明はしませんが、この後で紹介するように、アプリケーション中に詳細な参考資料のリストを載せています。(訳注：SDM に関する初心者向けの解説を気候変動適応プラットフォーム A-PLAT から提供しています)

Wallace とその翻訳版であるオープン SDM は、1) オープンシステム、2) 拡張可能、3) 柔軟性、4) インタラクティブ、5) 教育的、かつ 6) 再現可能など、次世代の科学計算ソフトウェアの好例と言える特色を多く持っています。このアプリケーションは拡大・縮小が可能な地図と動的なプロットと表を使って、途中処理や結果をわかりやすく可視化していることを特徴としています。モデルで使うデータは、オンラインデータベースからのダウンロード、またはユーザーによるアップロードが可能です。ほとんどの結果はダウンロード可能で、分析を再現できる R のコードを保存するオプションもあります。SDM についての解説を含め、詳しい解説は本文書の原文執筆者の *Methods in Ecology and Evolution* に発表の論文をご覧ください。引用は以下の通りです。

Kass JM, Vilela B, Aiello-Lammens ME, Muscarella R, Merow C, Anderson RP.
(2018). *Wallace: A flexible platform for reproducible modeling of species niches and distributions built for community expansion. Methods in Ecology and Evolution*. 9:1151-1156. <https://doi-org.ezproxy.gc.cuny.edu/10.1111/2041-210X.12945>

また、日本語の関連文献としては以下もご覧ください。本パッケージが対象としている、博物館標本などの在のみデータ（不在地点の情報がない分布情報）での種分布モデル構築に関する用語や技術、指標の解説が含まれています。

石濱 史子 (2017) 標本情報等の分布推定への活用と実際： バイアスの除去から精度評価まで, 保全生態学研究, 22 巻, 1 号, p. 21-40, 公開日 2018/04/01, Online ISSN 2424-1431, Print ISSN 1342-4327, https://doi.org/10.18960/hozen.22.1_21, https://www.jstage.jst.go.jp/article/hozen/22/1/22_21/_article/-char/ja

Wallace プロジェクトの[メインページ](#) には、[Google Group](#)、公式 [email アドレス](#)、安定バージョンをホスティングする [CRAN ページ](#)、および [Github 開発サイト](#)へのリンクがあります。

また、[A-PLAT のオープン SDM のページ](#)には、SDM に関する初心者向けの日本語解説文書と [github のオープン SDM ダウンロードサイト](#)へのリンクがあります。[github のオープン SDM サイトの wiki](#) にはインストールエラーへの対処法などを掲載しています。

セットアップについて

オープン *SDM* を実行するには、**最新バージョンの R**([Windows 版](#)または [Mac 版](#); または少なくともバージョン 3.5.0 以降のもの)**が必要で、加えて [RStudio](#) を使うことを推奨します。**

R と RStudio をインストールした後、RStudio を開きます。下記のコードを実行するだけで、オープン *SDM* (パッケージ名は *wallaceJP*)をインストールして読み込み、使用することができます。インストールがうまく行かない場合は [github のオープン SDM サイトの wiki](#) もご覧ください。

```
# github からのパッケージのインストールに必要な devtools をインストール
install.packages("devtools")

# オープン SDM (wallaceJP package) のインストール
devtools::install_github("opensdm/wallaceJP")

# パッケージの読み込み
library(wallaceJP)

# オープン SDM の起動
run_wallace()
```

使用しているデフォルトのブラウザでオープン *SDM* の GUI が開きます。R コンソールは占有された状態になります。オープン *SDM* を終了するには R コンソールの中で Escape キーを押すか、ブラウザのウィンドウを閉じます。

注意： オープン *SDM* を実行中にブラウザのウィンドウを閉じると、セッションが終了して作業中の内容がすべて失われます。

Wallace 導入 1 在データ 2 在データの処理 3 環境データ 4 環境データの処理 5 在データ分割 6 モデル 7 可視化 8 投影 セッションコード

ワークフロー

Wallace (v1.1.3-1.3) には現在、以下のコンポーネントとモジュールが含まれています。

1: 在データの取得

- データベースクエリ
- ユーザー指定

2: 在データの処理

- 地図上で選択
- IDで削除
- Spatial Thin

3: 環境データの取得

- WorldClim
- ユーザー指定の環境データ

4: 環境データの処理

- 対象地域を選択
- 背景加重法 (Background thickening)
- ユーザー指定の対象地域

5: 在データの分割

- 非空間分割
- 空間分割

6: ニッチモデルの構築と評価

- BIOCLIM
- Maxent

7: モデルの結果の可視化

- BIOCLIM Envelope Plot
- Maxent Evaluation Plots
- 応曲線のプロット
- 予測の地図化

8: モデルの投影

- 新規エリアへの投影
- 新規時間への投影
- 環境類似度を計算

セッションコードをダウンロード

導入 紹介

オープンSDMとは？

オープンSDMへようこそ。オープンSDMは、国際的なチームにより開発されたパッケージWallaceを日本語化したものであり、再現性のある生態学的モデリングのための柔軟なアプリケーションで、この分野のコミュニティを拡げるために構築されました。現在のバージョンのオープンSDM (v1.1.3-1.3)では、データ取得から結果の可視化まで、ニッチ/分布モデリング分析の全てのステップをユーザーに案内します。

このアプリケーションは R で書かれており、Rのウェブアプリ開発パッケージ [shiny](#) を使用しています。安定版の Wallace は [CRAN](#) に、開発版は [Github](#) にあります。また、私たちは [Wallace website](#) を維持しています。このサイトには基本的な情報やリンクがあり、近い将来チュートリアル資料も更新される予定です。

オープンSDMは生物多様性の空間的研究を促進するために設計されており、現在是在地点データセットと環境予測変数を用いた生物種のニッチと分布のモデル化に注力しています。これらのモデルは、環境条件に対する種の応答を推定するものであり、種の生育に適した地域を示す地図を作成するために使用することができます (すなわち、種の潜在的な地理的分布を示すマップを作成するために使用することができます; Guisan & Thuiller 2005; Elith & Leathwick 2009; Franklin 2010a; Peterson et al. 2011)。この研究分野は過去 20 年間で飛躍的に成長し、保全生物学 (Franklin 2010b)、侵略種 (Ficetola et al. 2007)、人獣共通感染症 (Gonzalez et al. 2010)、気候変動の影響 (Kearney et al. 2010) などの差し迫った環境問題に応用されている。

また、より詳しくは、Methods in Ecology and Evolutionの論文を参照してください。

Kass J. M., Vilela B., Aiello-Lammens M. E., Muscarella R., Merow C., Anderson R. P. (2018). Wallace: A flexible platform for reproducible modeling of species niches and distributions built for community expansion. *Methods Ecol. Evol.* 2018. 9: 1151-1156. DOI: 10.1111/2041-210X.12945

オープンSDMは誰のためのものか？

オープンSDMは、大学院生、生態学者、自然保護の実践者、自然資源管理者、教育者、プログラマーを含む幅広い人々に使ってもらえるように設計されています。プログラミング能力に関係なく、誰でもオープンSDMを使って解析を行い、手法を学び、結果を共有することができます。また、ある手法を普及させたい人は、Wallace用のモジュールを作成することができます。

オープンSDMの特徴

- オープンソース:** このコードは自由に使用・変更できます (GPL 3.0)。
- 拡張性:** ユーザーは、新しい方法についてのオプションを可能にするモジュールを作成したり、投稿したりすることができます。
- 柔軟性:** ユーザーによるアップロードと結果のダウンロードのためのオプション。
- 双方向性:** ズーム可能な [リーフレット](#) マップ、ソート可能な [DF](#) データ テーブル、および結果の可視化が含まれています。
- 学習支援:** ワークフローの各ステップの理論的および分析的な側面について、ユーザーの学習に役立つガイダンステキストを搭載しています。
- 再現可能:** ユーザーは [rmarkdown](#) .Rmd ファイルをダウンロードすることができ、このファイルを実行することで解析を再現できます。

オープン *SDM* の実行中に R コンソールを使用する場合、ターミナルウィンドウ(MacOS/Linus)またはコマンドプロンプト(Windows)を開き、Rを初期化した上で上記のラインを実行します。下記は MacOS のターミナルの例です。

```
R 4.2.2 . ~/
R version 4.2.2 (2022-10-31 ucrt) -- "Innocent and Trusting"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> library("wallaceJP")
要求されたパッケージ leaflet をロード中です
要求されたパッケージ shiny をロード中です
warning messages:
1: パッケージ 'leaflet' はバージョン 4.2.3 の R の下で造られました
2: パッケージ 'shiny' はバージョン 4.2.3 の R の下で造られました
> run_wallace()

Listening on http://127.0.0.1:4377
warning: パッケージ 'shiny.i18n' はバージョン 4.2.3 の R の下で造られました
```

また、オープン SDM で Maxent を使用する場合は次の点に留意して下さい。Maxent は Java プログラムなので、まず、自身の環境にあったバージョンの Java をインストールしておく必要があります。また、オープン SDM は dismo パッケージで `maxent()` 関数を使用しますが、そのためにはユーザー自身が `maxent.jar` ファイルを dismo パッケージのルートフォルダにある `/java` ディレクトリに置く必要があります。Maxent を [ここ](#) からダウンロードすると、フォルダの中に Maxent プログラムの本体である `maxent.jar` があります。R コンソールで `system.file('java', package="dismo")` を実行すると `dismo/java` ディレクトリへのパスが表示されるので、このディレクトリに `maxent.jar` をコピー・ペーストします。このファイルがない状態で Maxent をオープン SDM で実行しようとすると、ログウィンドウに操作を指示する警告メッセージが表示されますが、Maxent は実行されません。また、rJava をインストールして使用する上で問題がある場合、オープン SDM の Github レポジトリである [README](#) に有用なトラブルシューティング情報が若干あります。

オリエンテーション

まず「在データの取得（Obtain Occurrence Records）」の手順から始めます。その前に簡単にオープン SDM の各部を説明します。下図を見てください。

オープン *SDM* のインターフェース上にタブがあります。これが個別の分析ステップを表す「手順（コンポーネント）」で、それぞれの手順を順次実行して行きます。(1)まず「1 在データ」をクリックします。ボタンやテキスト入力などのツールバーが左側にあります。(2) **データベース検索** という「モジュール」が選択されているのがわかります。「モジュール」は各コンポーネント中の個別の分析オプションです。このコンポーネントにはもう1つ「**ユーザー指定**」というモジュールがあり、ユーザーはこのモジュールでは自分の在のみデータをアップロードすることができます。このモジュールを選択するとツールバーが変わります。

次に、クリックして**データベース検索**に戻ります。このツールバーでは使用するモジュール名と R パッケージ(2a)、選択したモジュール用のコントロールパネル(2b)が表示されます。

右側は可視化のスペース(3)で、ログウィンドウ(3a)といくつかのタブ(3b)があり、タブにはコンポーネントとモジュールの両方のレベルでインタラクティブな地図、在地点の表、結果を表示するウィンドウ、ガイダンスのウィンドウなどがあります。

分析のこの段階では、まだ結果が存在せず、表に表示するデータありませんが、ガイダンスを閲覧することができます。ガイダンスには、各コンポーネントとモジュールについて**方法論的**(ツールの機能)および**理論的**(使用する理由)な指針と、詳細を知るための参考文献リスト（生態学の学術論文）があります。基礎知識を強化するためにも、分析を始める前にこれらの文献を読むことを推奨します。

在データの取得

最初に“手順のガイダンス”をクリックして「手順 1：在データの取得」を開き、次に“モジュールガイダンス”を開きます。在データを取得するのに何が必要か、Wallace/オープン SDM がそれをどのように実装しているかをよりよく理解するために、これらを一読してください。

Wallace 導入 1 在データ 2 在データの処理 3 環境データ 4 環境データの処理 5 在データ分割 6 モデル 7 可視化 8 投影 セッションコード

在データの取得

選択可能なモジュール:

☒ データベース検索

☐ ユーザー指定

モジュール: データベース検索

spocc: 種の在データソースへのインターフェース

データベースを選択

☒ GBIF ☐ VertNet ☐ BISON

種の学名を入力してください

Fagus japonica

最大の在地点数の設定

2000

データベース検索

データベースから在地点情報 (csv形式) をダウンロードする

ダウンロード

Module Developers: Jamie M. Kass, Bruno Vilela, Robert P. Anderson

spocc references

Package Developers: Scott Chamberlain, Karthik Ram, Ted Hart

CRAN | documentation

> 合計 gbif から、Fagus japonica が返された数[2000]、それぞれのレコード数は、データベースの総件数 [2150]、検索の上限数 [2000]、座標がないため除去された数 [724]、重複により除去した数 [483]、残りのレコード数 [793]。

地図 在地点の表 結果 手順のガイダンス モジュールガイダンス

はじめに

ニッチ/分布のモデリング分析にはその生物種について位置情報付きの(例: 緯度/経度を伴う)在(occurrence)データを必要とします。現在、コンポーネント 在データの取得は生物種の「存在(presence)」を記録したデータに主眼を置いています(つまり、不在 (absence) または不検出 (non-detection)の情報は扱いません。Franklin 2010 chap. 4、Peterson et al. 2011 chapter 5、Anderson 2012)。現在Wallace でユーザーが利用できるのは、1)一部のオンライン生物多様性データベースから在データを取得して結果をダウンロードするか(モジュール データベース検索)、または2)ユーザー自身のデータセットをアップロードするか(モジュール ユーザー指定の在データ)のどちらかです。

引用文献

Anderson, R. P. (2012). Harnessing the world's biodiversity data: promise and peril in ecological niche modeling of species distributions. *Annals of the New York Academy of Sciences*. 1260: 66-80.

Franklin J. (2010). Mapping Species Distributions: Spatial Inference and Prediction. Data for species distribution models: the biological data. In: Mapping species distributions: spatial inference and prediction. Cambridge: Cambridge University Press.

Peterson A.T., Soberón J., Pearson R.G., Anderson R.P., Martinez-Meyer E., Nakamura M., & Araujo M.B. (2011). Species Occurrence Data. In: *Ecological Niches and Geographic Distributions*. Princeton, New Jersey: Monographs in Population Biology, 49. Princeton University Press.

背景地図の切り替え
ESRI Topo

Wallace 導入 1 在データ 2 在データの処理 3 環境データ 4 環境データの処理 5 在データ分割 6 モデル 7 可視化 8 投影 セッションコード

在データの取得

選択可能なモジュール:

☒ データベース検索
☐ ユーザー指定

モジュール: データベース検索

spocc: 種の在データソースへのインターフェース

データベースを選択

☒ GBIF
☐ VertNet
☐ BISON

種の学名を入力してください

Fagus japonica

最大の在地点数の設定

2000

データベース検索

データベースから在地点情報 (csv形式) をダウンロードする

ダウンロード

Module Developers: Jamie M. Kass, Bruno Vilela, Robert P. Anderson

spocc references

Package Developers: Scott Chamberlain, Karthik Ram, Ted Hart

CRAN | documentation

背景地図の切り替え

ESRI Topo

合計 gbif から、Fagus japonica が返された数[2000]、それぞれのレコード数は、データベースの総件数 [2150]、検索の上限数 [2000]、座標がないため除去された数 [724]、重複により除去した数 [483]、残りのレコード数 [793]。

地図

在地点の表

結果

手順のガイダンス

モジュールガイダンス

モジュール: データベース検索

背景

過去20年の間に、世界の生物多様性情報学コミュニティは驚くべき進歩を遂げ、各種の重要なデータベースを通じて、大量の位置情報(例: 緯度/経度座標。Galji et al. 2013, Peterson et al. 2015, Walters and Scholes 2017)を含む、何百万という生物種の在データをオンラインで利用できるようになりました。これらのデータは特定の時点・地点における生物種の存在の記録で、他にも有用なメタデータが含まれているものもあります(例: 研究機関、標本/観察番号、標高など)。これらの多くが自然史博物館や植物標本室における研究資料中の標本をソースとしていますが、市民科学イニシアティブなどの新しいデータソースからの寄与も増加しつつあります(Sullivan et al. 2009)。

実装

このモジュールは、在データベースへのアクセスを効率化するRパッケージ spocc を利用して実装されています。ユーザーは最も一般的なデータベース3種(a href="http://www.gbif.org" target="_blank">GBIF、VertNet、BISON)のどれかを選択することができます。これらのデータベースは、無数の提供者からのデータを集約しています。ただし、現在のところ、3つのデータベースのうちどれか1つのみを選ばなければならず、後からダウンロードしたものは先にダウンロードしたものを上書きするので注意が必要です。

Wallace は、解析に使用するレコードのうち、位置情報(緯度/経度座標)のないものや他のレコードと座標が完全に一致するもの(少数位の桁数を含む)をフィルタして取り除きます。フィルタされたレコードは「Occs Tbl」タブに、name(種名)、longitude(経度)、latitude(緯度)、year(記録年)、institutionCode(標本所蔵機関名)、country(国名)、stateProvince(地域名)、locality(地名)、elevation(標高)、および basisOfRecord (観測方法)の主要フィールドと共に表示されます。csv形式のファイルでダウンロード可能なレコードには元のフィールドがすべて含まれ、位置情報のないレコードも入っています。

引用文献

Galji, S., Chavan, V., Ariño, A. H., Otegui, J., Hobern, D., Sood, R., & Robles, E. (2013). Content assessment of the primary biodiversity data published through GBIF network: status, challenges and potentials. *Biodiversity Informatics*. 8: 94-172.

Peterson, A. T., Soberón, J., & Krishtalka, L. (2015). A global perspective on decadal challenges and priorities in biodiversity informatics. *BMC Ecology*. 15: 15.

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*. 142: 2282-2292.

Walters, M., and Scholes, R. J., (Eds.). (2017). The GEO Handbook on Biodiversity Observation Networks. Springer International Publishing. Link: <http://link.springer.com/book/10.1007/978-3-319-27288-7>

理解が深まったところで、実際に在データを取得してみましょう。一例として、GBIF から *Fagus japonica* (イヌブナ)の在データのレコードをダウンロードします。これは日本だけに生育している固有種です。在データの最大数を 2000 に設定して、「データベース検索ボタンをクリックします。ダウンロードが完了したら、ログウィンドウにメッセージが表示されます。2000 件のレコードを検索しましたが、ここでは座標情報(緯度と経度)があるレコードで重複していないものは 792 件しか見つかりませんでした。SDM に不可欠な位置情報を持つレコードを 1000 件以上取得するには、検索する在のみデータの最大数を増やして、再度検索します。

Wallace 導入 1 在データ 2 在データの処理 3 環境データ 4 環境データの処理 5 在データ分割 6 モデル 7 可視化 8 投影 セッションコード

在データの取得

選択可能なモジュール:

☒ データベース検索
 ☐ ユーザー指定

モジュール: データベース検索

spocc: 種の在データソースへのインターフェース

データベースを選択

☒ GBIF
 ☐ VertNet
 ☐ BISON

種の学名を入力してください

Fagus japonica

最大の在地点数の設定

2000

データベース検索

データベースから在地点情報 (csv形式) をダウンロードする

ダウンロード

Module Developers: Jamie M. Kass, Bruno Vilela, Robert P. Anderson

spocc references

Package Developers: Scott Chamberlain, Karthik Ram, Ted Hart

CRAN | documentation

> 合計 gbif から、Fagus japonica が返された数[2000]、それぞれのレコード数は、データベースの総件数 [2149]、検索の上限数 [2000]、座標がないため除去された数 [725]、重複により除去した数 [483]、残りのレコード数 [792]。

背景地図の切り替え

ESRI Topo

地図 在地点の表 結果 手順のガイダンス モジュールガイダンス

次に、「在地点の表」タブをクリックして、レコードの詳細を閲覧します。たくさんあるフィールドのうち、開発者が選んだ生物種の分布の研究に関連性が高いと考えられるフィールドが表示されています。「ダウンロード」ボタンをクリックすると、これらのレコードを csv 形式でダウンロードすることができ、ダウンロードファイルは元のデータベースのフィールドすべてを網羅しています。

Wallace 導入 1 在データ 2 在データの処理 3 環境データ 4 環境データの処理 5 在データ分割 6 モデル 7 可視化 8 投影 セッションコード

在データの取得

選択可能なモジュール:

☒ データベース検索
 ☐ ユーザー指定

モジュール: データベース検索

spocc: 種の在データソースへのインターフェース

データベースを選択

☒ GBIF
 ☐ VertNet
 ☐ BISON

種の学名を入力してください

Fagus japonica

最大の在地点数の設定

2000

データベース検索

データベースから在地点情報 (csv形式) をダウンロードする

ダウンロード

Module Developers: Jamie M. Kass, Bruno Vilela, Robert P. Anderson

spocc references

Package Developers: Scott Chamberlain, Karthik Ram, Ted Hart

CRAN | documentation

> 合計 gbif から、Fagus japonica が返された数[2000]、それぞれのレコード数は、データベースの総件数 [2149]、検索の上限数 [2000]、座標がないため除去された数 [725]、重複により除去した数 [483]、残りのレコード数 [792]。

背景地図の切り替え

ESRI Topo

地図 在地点の表 結果 手順のガイダンス モジュールガイダンス

Show 10 entries Search:

name	occlID	longitude	latitude	year	institutionCode	catalogNumber	basisOfRecord	country
Fagus japonica Maxim.	1	139.11	35.85	2022	iNaturalist	131286873	HUMAN_OBSERVATION	Japan
Fagus japonica Maxim.	2	136.87	35.48	2021	GPM	91739	PRESERVED_SPECIMEN	Japan
Fagus japonica Maxim.	4	138.92	35.83	2020	iNaturalist	52118020	HUMAN_OBSERVATION	Japan
Fagus japonica Maxim.	5	137.46	35.14	2020	TCNS	4176	PRESERVED_SPECIMEN	Japan
Fagus japonica Maxim.	6	139.2	35.43	2019	iNaturalist	46374719	HUMAN_OBSERVATION	Japan
Fagus japonica Maxim.	7	138.08	36.35	2019	SUG	6886	PRESERVED_SPECIMEN	Japan
Fagus japonica Maxim.	8	138.95	35.89	2019	iNaturalist	41453866	HUMAN_OBSERVATION	Japan

9

在データの処理

次のコンポーネントである**在データの処理**では、いくつかのデータクリーニングのツールが用意されています。ダウンロードしたデータは未精査で、必ずと言って良いほどいくつかの欠陥を含んでいます。最も明らかな誤りを除去するには、研究対象種の生息範囲についての基礎知識が役立ちます。図の例では、日本の固有種であるイヌブナ *F. japonica* の在のみデータがヨーロッパにまで表示されています。多様なソースから大量のデータを蓄積している GBIF のようなデータベースでは、例えば標本に関する座標の代わりに博物館の位置など、何らかの理由による不正確な座標が含まれています。こうした明らかに誤りであるレコードを除去するには、モジュール「**地図上での在地点の選択**」をクリックして、分析対象として保持したい地点を選択します。ここでは地図上の多角形アイコンをクリックして、日本の各地点を多角形で囲みます。

Wallace 導入 1 在データ 2 在データの処理 3 環境データ 4 環境データの処理 5 在データ分割 6 モデル 7 可視化 8 投影 セッションコード

在データの処理

選択可能なモジュール:

☒ 地図上での在地点の選択

☐ IDによる在データの除去

☐ 空間的な在データの間引き

モジュール: 地図上での在データの選択

leaflet.extras: 'leaflet'パッケージの追加機能

描画したポリゴンと重なる在地点を選択
(**NOTE** 描画を開始するには、マップツールバーの六角形アイコンをクリックします。描画が終わったら、「Finish」ボタンを押してから「在地点の選択」ボタンを押してください。)

在地点を選択

在データを元のセットにリセットします。

リセットします

処理された在位置のダウンロード (.csv)

ダウンロード

Module Developers: Jamie M. Kass, Robert P. Anderson
leaflet.extras references
Package Developers: Bhaskar Karambelkar, Bangyou Zheng
CRAN | documentation

> 合計 gbif から、Fagus japonica が返された数[2000]、それぞれのレコード数は、データベースの総件数 [2149]、検索の上限数 [2000]、座標がないため除去された数 [725]、重複により除去した数 [483]、残りのレコード数 [792]。

地図 在地点の表 結果 手順のガイダンス モジュールガイダンス

背景地図の切り替え
ESRI Topo

地図

SWEDEN FINLAND NORWAY POLAND UKRAINE RUSSIA KAZAKHSTAN MONGOLIA GOBI DESERT CHINA INDIA MYANMAR BURMA INDONESIA

在地点の表

結果

手順のガイダンス

モジュールガイダンス

在地点を選択

在データを元のセットにリセットします。

リセットします

処理された在位置のダウンロード (.csv)

ダウンロード

Module Developers: Jamie M. Kass, Robert P. Anderson
leaflet.extras references
Package Developers: Bhaskar Karambelkar, Bangyou Zheng
CRAN | documentation

次に、多角形ツールバーの「Finish」をクリックして、さらに「**在地点を選択**」をクリックします。地図が選択した地点にズームインします。

10

Wallace 導入 1 在データ 2 在データの処理 3 環境データ 4 環境データの処理 5 在データ分割 6 モデル 7 可視化 8 投影 セッションコード

在データの処理

選択可能なモジュール:

☒ 地図上で在地点の選択
 ☐ IDによる在データの除去
 ☐ 空間的な在データの引き

モジュール: 地図上で在データの選択

leaflet.extras: 'leaflet'パッケージの追加機能

描画したポリゴンと重なる在地点を選択

(NOTE 描画を開始するには、マップツールバーの六角形アイコンをクリックします。描画が終わったら、「Finish」ボタンを押してから「在地点の選択」ボタンを押してください。)

在地点を選択

在データを元のセットにリセットします。

リセットします

処理された位置のダウンロード (.csv)

ダウンロード

Module Developers: Jamie M. Kass, Robert P. Anderson
 leaflet.extras references
 Package Developers: Bhaskar Karambelkar, Bangyou Zheng
 CRAN | documentation

のレコード数は、データベースの総件数 [2149]、検索の上限数 [2000]、座標がないため除去された数 [725]、重複により除去した数 [483]、残りのレコード数 [792]。

> 削除された在データの occID = c(495, 873, 879, 1339, 1789, 1804, 1831, 1865)。選択されたデータは n = 784 レコード。

背景地図の切り替え

ESRI Topo

地図 在地点の表 結果 手順のガイダンス モジュールガイダンス

他の方法として、モジュール **ID による在データの除去** を使用して在データを除去することもできます。*F. japonica* は、本州、四国、九州に生育しますが、少しズームインすると、在地点の一部が北海道にも出現します。これはジョリアファレンスの誤りで、対象種の一般に受け入れられている生育範囲を超えていると考えられます。**注意：** この文書では、これらの地点が北海道での *F. japonica* の真の目撃事例である可能性を無視していますが、研究においてはこのような憶測は安易にすべきではなく、よく検討する必要があります。これらの地点を除去するには、それぞれのポイントをクリックして固有の ID(「occID」と地理的座標を表示し、その ID を左のパネルに入力して、「**在データを削除**」をクリックします。

11

Wallace
導入
1 在データ
2 在データの処理
3 環境データ
4 環境データの処理
5 在データ分割
6 モデル
7 可視化
8 投影
セッションコード

在データの処理

選択可能なモジュール:

☐ 地図上での在地点の選択
☒ IDによる在データの除去
☐ 空間的な在データの間引き

IDによる在データの除去

削除するレコードIDを入力

1968

在データを削除

在データを元のセットにリセットします。

リセットします

処理された在位置のダウンロード (.csv)

ダウンロード

Module Developers: Jamie M. Kass, Robert P. Anderson

のレコード数は、データベースの総件数 [2149]、検索の上限数 [2000]、座標がないため除去された数 [725]、重複により除去した数 [483]、残りのレコード数 [792]。

> 削除された在データの occID = c(495, 873, 879, 1339, 1789, 1804, 1831, 1865)。選択されたデータは n = 784 レコード。

地図

在地点の表

結果

手順のガイダンス

モジュールガイダンス

+

-

📍

🗑️

occID: 1968

Latitude: 43.05

Longitude: 141.29

Year: 1891

Inst. Code: MNHN

Basis of Record: PRESERVED_SPECIMEN

Occurrence ID: http://coldb.mnhn.fr/catalognumber/mnhn/ip06812018

Country: Japan

State/Prov.: Hokkaido(北海道)

Locality: NA

Elevation: NA

背景地図の切り替え

ESRI Topo

誤りと思われる地点を除去した後に残った地点が、調査範囲の偏りによってクラスター状になっていることがあります。これは分布の人為的に誇張された空間的な自己相関（訳注：空間的に集中した分布）につながる 경우가多く、このようなデータにモデルを当てはめると環境応答の推定にバイアスをかける可能性があります。例えば、データのソースが市民科学者によるものである場合、市民科学者はおおむね都市周辺に住んでいるため、在地点の都市近辺への集中が起きると考えられます。あるいは、データを収集した野外研究の生物学者が運転中に観察したか、または道路から現地へアクセスしたことで地点が道路周辺に集中する場合があります。こうした観測方法により生じた在データの集中は、対象種の本来の環境応答以上にこうした場所（この例では都市や道路）への好適度が高いとモデルで推定してしまう危険性があり、その補正がモデル構築前に必要とされます。

補正方法の1つが空間的に在地点を間引く方法です。モジュール「空間的な在データの間引き」をクリックします。この空間間引き関数では、在データ間の空間距離が近いものを除外することによって、各地点の間の距離が互いに定義した距離以上になるようにし、推定の空間バイアスを減らします。ここでは例として 10 キロを使用します。

12

Wallace 導入 1 在データ 2 在データの処理 3 環境データ 4 環境データの処理 5 在データ分割 6 モデル 7 可視化 8 投影 セッションコード

在データの処理

選択可能なモジュール:

☐ 地図上での在地点の選択
 ☒ IDによる在データの除去
 ☐ 空間的な在データの間引き

IDによる在データの除去

削除するレコードIDを入力

1968

在データを削除

在データを元のセットにリセットします。

リセットします

処理された在位置のダウンロード (.csv)

ダウンロード

Module Developers: Jamie M. Kass, Robert P. Anderson

た数 [483]、残りのレコード数 [792]。
 > 削除された在データの occID = c(495, 873, 879, 1339, 1789, 1804, 1831, 1865)。選択されたデータは n = 784 レコード。
 > 削除された在データのID = 1968。選択されたデータは n = 784 レコード。

地図 在地点の表 結果 手順のガイダンス モジュールガイダンス

背景地図の切り替え
 ESRI Topo

この分析では 278 地点が残りました(ユーザーによっては異なることがあります)。

Wallace 導入 1 在データ 2 在データの処理 3 環境データ 4 環境データの処理 5 在データ分割 6 モデル 7 可視化 8 投影 セッションコード

在データの処理

選択可能なモジュール:

☐ 地図上での在地点の選択
 ☐ IDによる在データの除去
 ☒ 空間的な在データの間引き

モジュール: 空間的な在データの間引き

spThin : 種の在記録の空間的間引き

間引き距離(km)

10

在地点のまびき

在データを元のセットにリセットします。

リセットします

処理された在位置のダウンロード (.csv)

ダウンロード

Module Developers: Jamie M. Kass, Matthew E. Aiello-Lammens, Robert P. Anderson

spThin references

Package Developers: Matthew E. Aiello-Lammens, Rob A. Boria, Alex Radosavljevic, Bruno Vilela, Robert P. Anderson

CRAN | documentation | software note

> 削除された在データの occID = c(495, 873, 879, 1339, 1789, 1804, 1831, 1865)。選択されたデータは n = 784 レコード。
 > 削除された在データのID = 1968。選択されたデータは n = 784 レコード。
 > 間引き後の合計レコード数は [278] 地点。

地図 在地点の表 結果 手順のガイダンス モジュールガイダンス

Occ Records

retained

removed

ズームインすると関数の実行結果が見られます。赤の地点が残り、青の地点が除去されました。こうして処理をした在のみデータのセットは「ダウンロード」ボタンをクリックすることでダウンロードできます。

Wallace 導入 1 在データ 2 在データの処理 3 環境データ 4 環境データの処理 5 在データ分割 6 モデル 7 可視化 8 投影 セッションコード

在データの処理

選択可能なモジュール:

☐ 地図上での在地点の選択
☐ IDによる在データの除去
☒ 空間的な在データの間引き

モジュール: 空間的な在データの間引き

spThin: 種の在記録の空間的な間引き

間引き距離 (km)

10

在地点のまびき

在データを元のセットにリセットします。

リセットします

処理された在位置のダウンロード (.csv)

ダウンロード

Module Developers: Jamie M. Kass, Matthew E. Aiello-Lammens, Robert P. Anderson
spThin references
Package Developers: Matthew E. Aiello-Lammens, Rob A. Boria, Alex Radosavljevic, Bruno Vilela, Robert P. Anderson
CRAN | documentation | software note

> 削除された在データの occID = c(495, 873, 879, 1339, 1789, 1804, 1831, 1865)。選択されたデータは n = 784 レコード。

> 削除された在データの ID = 1968。選択されたデータは n = 784 レコード。

> 間引き後の合計レコード数は [278] 地点。

地図

在地点の表

結果

手順のガイダンス

モジュールガイダンス

+

-

📍

🗑️

ESRI Topo

環境データの取得

次に、分析のために環境変数を取得する必要があります。在のみデータのある地点について環境変数の値を抽出し、モデルの構築に使用します。環境データはラスタ形式（訳注：位置情報付きデータの形式の1つで、空間が格子状に区切られている）で、グリッドのセルそれぞれが値を持っています。ラスタは地図上に面として表示することができます(後述)。

「環境データ」コンポーネントをクリックしましょう。最初のモジュールである **WorldClim Bioclims** で、グローバル気候データベース [WorldClim](#) から気候データをダウンロードします。複数の解像度のデータがあり、これらは気象観測所のデータを面的に補間したものです。(特に先進国の)気象観測所の多い地域はよくカバーされていますが、観測所の少ない地域ではデータの不確実性が高くなります。[bioclim](#) 変数は、一般的に生物学上重要と提唱された気温および降水量の要約変数です。

ここでは bioclim の空間解像度として「**2.5 arcmin**」を選択し、「**環境データ読み込**」ボタンをクリックします。初めてオープン SDM を使用する場合には、これらのデータはコンピュータのハードドライブにダウンロードされ、その後はこのローカルディレクトリから読み込まれます。解像度が細かければダウンロードに時間がかかります。最も解像度の細かいデータ(30 arcsec)はタイル形式で提供されるので、地図の中心に対応するタイルがダウンロードされ

14

ます。ラスタをダウンロードする他に、オープン *SDM* は環境値のない在地点(例：ラスタと重複しない地点)も除去します。右下隅に進捗バーが表示されます。

環境データの取得

選択可能なモジュール:

- ☒ WorldClim Bioclimes
- ☐ ユーザー指定

モジュール: WorldClim Bioclimes

raster: 地理データの分析とモデリング

WorldClimの生物気候変数の解像度を選択します。

2.5 arcmin

☐ 分析に使用する変数を指定しますか?

タイルのダウンロードに地図の中心座標を参照しています。
Using map center: 136.154, 35.852

環境データ読み込み

Module Developers: Jamie M. Kass, Gonzalo E. Pinilla-Buitrago, Robert P. Anderson

raster references

Package Developers: Robert J. Hijmans, Jacob van Etten, Joe Cheng, Matteo Mattiuzzi, Michael Sumner, Jonathan A. Greenberg, Oscar Perpinan Lamiguero, Andrew Bevan, Etienne B. Racine, Ashton Shortridge

CRAN | documentation | WorldClim

背景地図の切り替え

ESRI Topo

地図 在地点の表 結果 手順のガイダンス モジュールガイダンス

NA値がある地点をチェックしています...

ラスタが読み込まれると、「結果」タブにそのサマリー情報(解像度、範囲、セル番号などの)の一部が表示されます。

環境データの取得

選択可能なモジュール:

- ☒ WorldClim Bioclimes
- ☐ ユーザー指定

モジュール: WorldClim Bioclimes

raster: 地理データの分析とモデリング

WorldClimの生物気候変数の解像度を選択します。

2.5 arcmin

☐ 分析に使用する変数を指定しますか?

タイルのダウンロードに地図の中心座標を参照しています。
Using map center: 136.154, 35.852

環境データ読み込み

Module Developers: Jamie M. Kass, Gonzalo E. Pinilla-Buitrago, Robert P. Anderson

raster references

Package Developers: Robert J. Hijmans, Jacob van Etten, Joe Cheng, Matteo Mattiuzzi, Michael Sumner, Jonathan A. Greenberg, Oscar Perpinan Lamiguero, Andrew Bevan, Etienne B. Racine, Ashton Shortridge

CRAN | documentation | WorldClim

背景地図の切り替え

ESRI Topo

地図 在地点の表 結果 手順のガイダンス モジュールガイダンス

レコード。

> 間引き後の合計レコード数は [279] 地点。

> 環境予測変数: WorldClim 生物気候変数 bio1-19 解像度 2.5 分。

! WARNING: 環境値のないレコードをocclDとともに削除しました: 607, 632.

```
class      : RasterStack
dimensions : 3600, 8640, 31104000, 19 (nrow, ncol, ncell, nlayers)
resolution : 0.04166667, 0.04166667 (x, y)
extent      : -180, 180, -60, 90 (xmin, xmax, ymin, ymax)
crs         : +proj=longlat +datum=WGS84 +no_defs
names       : bio01, bio02, bio03, bio04, bio05, bio06, bio07, bio08, bio09, bio10, bio11, bio12, bio13, bio14, bio15, ...
min values  : -278, 9, 8, 64, -86, -559, 53, -278, -501, -127, -506, 0, 0, 0, 0, ...
max values  : 319, 213, 96, 22704, 489, 258, 725, 376, 365, 382, 289, 10577, 2437, 697, 265, ...
```

分析に使用する変数を指定するオプションがあることに注意します。(訳注：分析対象の種の分布と関連性が高いと考えられ、また、説明変数同士の相関が高くなりすぎないような変数セットを選ぶ。)

Wallace 導入 1 データ 2 データの処理 3 環境データ 4 環境データの処理 5 データ分割 6 モデル 7 可視化 8 投影 セッションコード

環境データの取得

選択可能なモジュール:

☒ WorldClim Bioclimes
 ☐ ユーザー指定

モジュール: WorldClim Bioclimes

raster: 地理データの分析とモデリング

WorldClimの生物気候変数の解像度を選択します。

2.5 arcmin

☒ 分析に使用する変数を指定しますか?

選択

☒ bio1
 ☐ bio2
 ☐ bio3
 ☐ bio4
 ☐ bio5
 ☒ bio6
 ☐ bio7
 ☐ bio8
 ☐ bio9
 ☐ bio10
 ☒ bio11
 ☐ bio12
 ☐ bio13
 ☐ bio14
 ☐ bio15
 ☐ bio16
 ☒ bio17
 ☒ bio18
 ☒ bio19

タイルのダウンロードに地図の中心座標を参照しています。
 Using map center: 136.154, 35.852

環境データ読み込み

Module Developers: Jamie M. Kass, Gonzalo E. Pinilla-Buitrago, Robert P. Anderson
 raster references
 Package Developers: Robert J. Hijmans, Jacob van Etten, Joe Cheng, Matteo Mattiuzzi, Michael Sumner, Jonathan A. Greenberg, Oscar Perpinan Lamiguero, Andrew Bevan, Etienne B. Racine, Ashton Shortridge
 CRAN | documentation | WorldClim

> 環境予測変数: WorldClim 生物気候変数 bio1-19 解像度 2.5 分。
 ! WARNING: 環境値のないレコードをocclDとともに削除しました: 607, 632.
 > 環境予測変数: WorldClim 生物気候変数 bio1, bio6, bio11, bio17, bio18, bio19 解像度 2.5 分。

地図 在地点の表 結果 手順のガイダンス モジュールガイダンス

```

Class      : RasterStack
dimensions : 3600, 8640, 31104000, 6 (nrow, ncol, ncell, nlayers)
resolution : 0.04166667, 0.04166667 (x, y)
extent     : -180, 180, -60, 90 (xmin, xmax, ymin, ymax)
crs       : +proj=longlat +datum=WGS84 +no_defs
names     : bio01, bio06, bio11, bio17, bio18, bio19
min values : -278, -559, -506, 0, 0, 0
max values : 319, 258, 289, 2319, 5040, 4580
          
```

背景地図の切り替え

ESRI Topo

環境データの処理

さて、モデルを構築するためには分析範囲を選択する必要があります。これはモデルの構築に使う「背景」地点を抽出する地域を定義するものです。Maxent のような方法は、背景地点における説明変数の値を、在地点での説明変数の値と比較することから、「在-背景（presence-background）法」として知られています（訳注：この方法では、在地点の環境が、周辺の環境に比べて、どのような条件に偏っているかを評価することで、生育に適した環境条件を検出する。このときの環境条件の比較対象とする地点を「背景」と呼び、オープンSDM では、背景地点を分析範囲の中からランダムにサンプリングする）。分析範囲を決めるにあたっては、対象種が過去に移入できなかった地域（例えば横断することができない山脈や大河などの障壁の外にある地域）、または調査が全く行われていない地域を避ける必要があります。こうした地域を含めると、その地域の環境条件が生息に不適であるという誤ったシグナルをモデルに送ってしまうことがあるからです（訳注：生息に適した環境であるにもかかわらず、障壁など他の理由で生息していないことがあるため）。詳しくはガイダンステキストを参照してください。

分析範囲の境界の設定するためのオプションを検討します。モジュール **分析範囲の選択**には 1)分析範囲の形状の選択、2)背景地点のサンプリングの 2 つのステップがあります。ステップ 1 で「最小凸多角形」を選択して、分析範囲のバッファ距離を 1 度に設定、「**選択**」ボタンをクリックすると、この形状が地図に描画されます。

Wallace 導入 1 在データ 2 在データの処理 3 環境データ 4 環境データの処理 5 在データ分割 6 モデル 7 可視化 8 投影 セッションコード

環境データの処理

選択可能なモジュール:

☒ 分析範囲の選択
☐ 背景加重法 (Background thickening)
☐ ユーザー指定

モジュール: 分析範囲の選択

sf: 空間データのタイトルクラスとメソッド

terra: 地理データの分析とモデリング

ステップ1: 背景の範囲を選択

背景の範囲:

☐ バウンディングボックス
☒ 最小凸多角形
☐ ポイントバッファ

分析範囲のバッファ距離 (度)

1

選択

ステップ2: 背景地点をサンプリングする

選択可能なモジュール:

☐ 対象群背景法 (Target group background)
☒ ランダムサンプリング
☐ ユーザー指定

背景地点数

> 環境予測変数: WorldClim 生物気候変数 bio1-19 解像度 2.5 分。

! WARNING: 環境値のないレコードをocclIDとともに削除しました: 900, 1106.

> 分析範囲: 最小凸多角形。バッファ後の分析範囲 (バッファサイズ 1 度)。

背景地図の切り替え

ESRI Topo

地図 在地点の表 結果 手順のガイダンス モジュールガイダンス

次のステップ 2 では、ラスタを研究範囲に合わせて切り抜き、背景地点をサンプリングします。背景地点の数を 10,000(大きく広範なサンプルほど有用)に設定し、「**サンプリングを行う**」ボタンをクリックします。

17

Wallace 導入 1 在データ 2 在データの処理 3 環境データ 4 環境データの処理 5 在データ分割 6 モデル 7 可視化 8 投影 セッションコード

環境データの処理

選択可能なモジュール:

☒ 分析範囲の選択
☐ 背景加重法 (Background thickening)
☐ ユーザー指定

モジュール: 分析範囲の選択

sf: 空間データのタイトルクラスとメソッド

terra: 地理データの分析とモデリング

ステップ1: 背景の範囲を選択

背景の範囲:

☐ バウンディングボックス
☒ 最小凸多角形
☐ ポイントバッファ

分析範囲のバッファ距離 (度)

1

選択

ステップ2: 背景地点をサンプリングする

選択可能なモジュール:

☐ 対象群背景法 (Target group background)
☒ ランダムサンプリング
☐ ユーザー指定

背景地点数

10000

サンプリングを行う

環境予測変数: WorldClim 生物気候変数 bio1-19 解像度 2.5 分。

> 分析範囲: 最小凸多角形。バッファ後の分析範囲 (バッファサイズ 1 度)。

> マスクされた環境データ。

> ランダムにサンプリングされた背景地点 (n = 10000)。

背景地図の切り替え

ESRI Topo

地図 在地点の表 結果 手順のガイダンス モジュールガイダンス

+

-

Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ, TomTom, Intermap, iPC, USGS, FAO, NPS, NRCAN, GeoBase, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), and the GIS User Community

右下隅の進捗バーを確認します。処理の終了後に「ダウンロード」ボタンをクリックすると、切り抜いたラスタを圧縮ファイル形式でダウンロードすることができます。

18

分析範囲のバッファ距離 (度)

1

選択

ステップ2: 背景地点をサンプリングする

選択可能なモジュール:

☐ 対象群背景法 (Target group background)
☒ ランダムサンプリング
☐ ユーザー指定

背景地点数

10000

サンプリングを行う

ダウンロードファイルのタイプを選択

GeoTIFF

背景範囲でマスクされた説明変数のラスタをダウンロード

ダウンロード

Module Developers: Jamie M. Kass, Bruno Vilela, Robert P. Anderson

sf references

Package Developers: Edzer Pebesma, Roger Bivand, Etienne Racine, Michael Sumner, Ian Cook, Tim Keltz, Robin Lovelace, Hadley Wickham, Jeroen Ooms, Kirill Müller, Thomas Lin Pedersen, Dan Baston

CRAN | documentation

terra references

Package Developers: Robert J. Hijmans, Márcia Barbosa, Roger Bivand, Andrew Brown, Michael Chirico, Emanuele Cordano, Krzysztof Dyrba, Edzer Pebesma, Barry Rowlingson, Michael D. Sumner

CRAN | documentation

Map showing the distribution of target species (red dots) and background locations (blue dots) within a buffer zone around the Korean Peninsula. The map includes labels for Daegu, Ulsan, Changwon, Busan, and the Korea Strait. The background is a light blue map with a white buffer zone around the Korean Peninsula.

オリジナルの wallace にはない OpenSDM 固有の機能として、ステップ 2 には**対象群背景法** (Target Group Background) のオプションがあります。この方法は、モデルを構築しようとしている種の近縁種の分布情報を背景地点として用います。**対象群背景法**は、多数の種の分布が同一の調査スキームで記録されている場合のバイアス補正に効果的です。OpenSDM では、最小凸多角形で選択された範囲内にある、**同じ属**の生物の分布地点を GBIF データベースから取得します。以下の図では、モデル構築対象種の座位置点は赤色で、背景地点 (同属の種の在地点) は青色で示されています。

19

環境データの処理

選択可能なモジュール:

- 分析範囲の選択
- ☐ 背景加重法 (Background thickening)
- ☐ ユーザー指定

モジュール: 分析範囲の選択

sf: 空間データのタイトルクラスとメソッド

terra: 地理データの分析とモデリング

ステップ1: 背景の範囲を選択

背景の範囲:

- ☐ バウンディングボックス
- 最小凸多角形
- ☐ ポイントバッファ

分析範囲のバッファ距離 (度)

1

選択

ステップ2: 背景地点をサンプリングする

選択可能なモジュール:

- 対象群背景法 (Target group background)
- ☐ ランダムサンプリング
- ☐ ユーザー指定

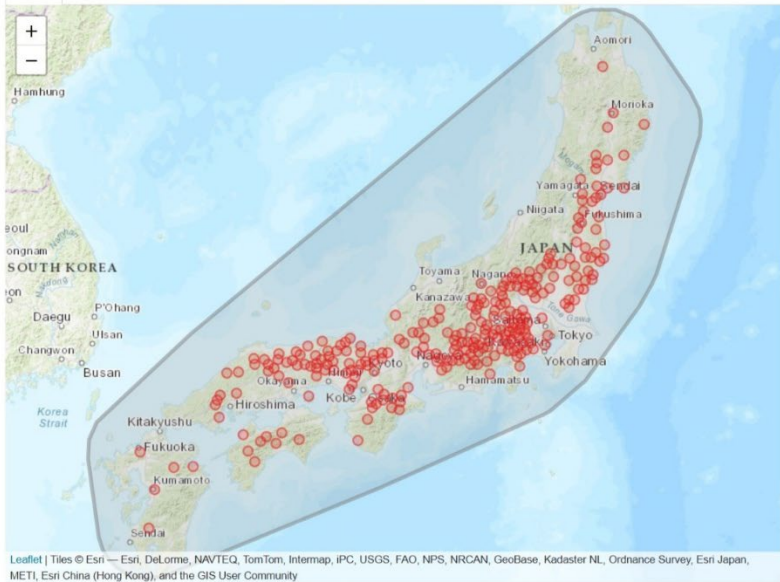
背景地点数

た: 874,900,1106,1126。
 > 間引き後の合計レコード数は [268] 地点。
 > 環境予測変数: WorldClim 生物気候変数 bio1-19 解像度 2.5 分。
 > 分析範囲: 最小凸多角形。バッファ後の分析範囲 (バッファサイズ 1 度)。

背景地図の切り替え

ESRI Topo

地図 在地点の表 結果 手順のガイダンス モジュールガイダンス



環境データの処理

選択可能なモジュール:

- 分析範囲の選択
- ☐ 背景加重法 (Background thickening)
- ☐ ユーザー指定

モジュール: 分析範囲の選択

sf: 空間データのタイトルクラスとメソッド

terra: 地理データの分析とモデリング

ステップ1: 背景の範囲を選択

背景の範囲:

- ☐ バウンディングボックス
- 最小凸多角形
- ☐ ポイントバッファ

分析範囲のバッファ距離 (度)

1

選択

ステップ2: 背景地点をサンプリングする

選択可能なモジュール:

- 対象群背景法 (Target group background)
- ☐ ランダムサンプリング
- ☐ ユーザー指定

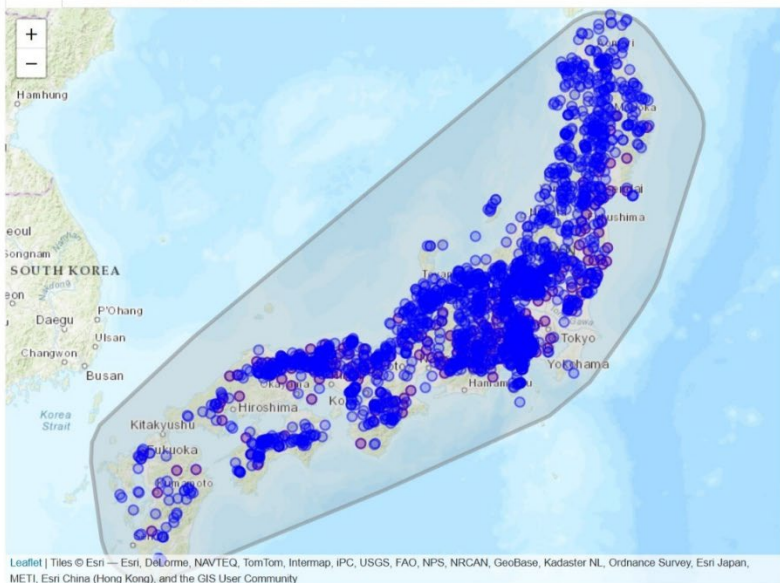
背景地点数

> ランダムにサンプリングされた背景地点 (n = 10000)。
 > マスクされた環境データ。
 > ターゲットグループの背景地点が要求されました。(n = 10000)、
 しかし 25 % のポイント (n = 2457) しかサンプリングできませんでした。

背景地図の切り替え

ESRI Topo

地図 在地点の表 結果 手順のガイダンス モジュールガイダンス



もう1つの OpenSDM 固有のバイアス補正機能が、**背景加重法（Background thickening）**です。この方法は、ポイントバッファに似ていますが、ポイントバッファは隣接した在地点に対するバッファの重複を考慮しないのに大して、背景加重法では重複するバッファが多い範囲内では、重複がないバッファ内の範囲より多くの背景地点がサンプリングされるという違いがあります。

以下の例では、まず在地点から 20km のバッファを発生させます。このバッファは測地バッファ（地球の曲率を考慮したバッファ）となっており、緯度の範囲が広い分析においても誤差が少ない方法です。

Wallace 導入 1 在データ 2 在データの処理 3 環境データ 4 環境データの処理 5 在データ分割 6 モデル 7 可視化 8 投影 セッションコード

環境データの処理

選択可能なモジュール:

☐ 分析範囲の選択

☒ 背景加重法 (Background thickening)

☐ ユーザー指定

モジュール: 背景加重法
(background thickening)

sf: 空間データのタイトルクラスとメソッド

ステップ1: 一定距離で測地線バッファを計算する

キロメートル単位での測地線バッファ距離の指定

実行

ステップ2: 背景地点をサンプリングする

背景地点数

サンプリングを行う

た: 900, 1106.
> 分析範囲: 最小凸多角形。バッファ後の分析範囲 (バッファサイズ 1 度)。
> 種に対して測地線バッファが計算された。バッファサイズ 20 キロメートル。

背景地図の切り替え
ESRI Topo

地図 在地点の表 結果 手順のガイダンス モジュールガイダンス

Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ, TomTom, Intermap, IPC, USGS, FAO, NPS, NRCAN, GeoBase, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), and the GIS User Community

ステップ2で指定した数の背景地点が、バッファ内からサンプリングされ、青色で表示されます。

環境データの処理

選択可能なモジュール:

- ☐ 分析範囲の選択
- ☒ 背景加重法 (Background thickening)
- ☐ ユーザー指定

モジュール: 背景加重法 (background thickening)

sf: 空間データのタイトルクラスとメソッド

ステップ1: 一定距離で測地線バッファを 計算する

キロメートル単位での測地線バッファ距離の指定

20

実行

ステップ2: 背景地点をサンプリングする

背景地点数

10000

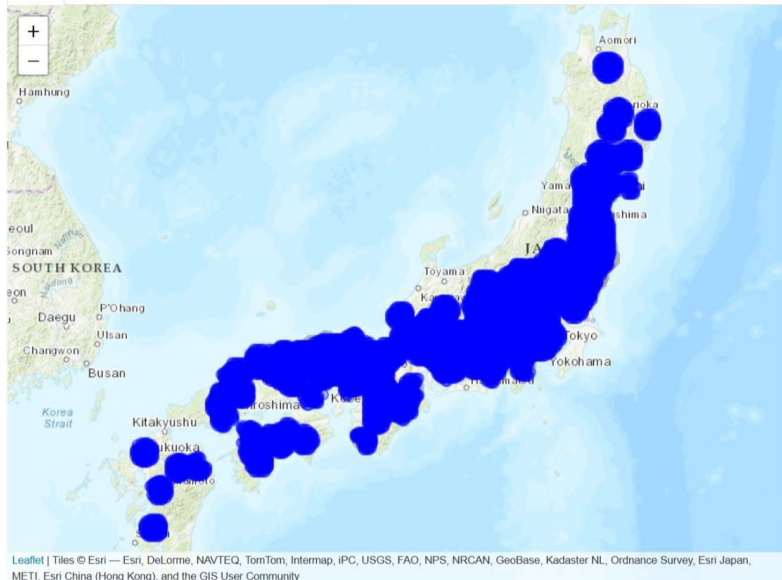
サンプリングを行う

ロメートル。
> マスクされた環境データ。
> 背景点のランダムサンプリングが要求されました (n = 10000)、
しかし 87 % のポイント (n = 8662) しかサンプリングできませんで
した。

背景地図の切り替え

ESRI Topo

地図 在地点の表 結果 手順のガイダンス モジュールガイダンス



環境データの処理

選択可能なモジュール:

- ☐ 分析範囲の選択
- ☒ 背景加重法 (Background thickening)
- ☐ ユーザー指定

モジュール: 背景加重法 (background thickening)

sf: 空間データのタイトルクラスとメソッド

ステップ1: 一定距離で測地線バッファを 計算する

キロメートル単位での測地線バッファ距離の指定

20

実行

ステップ2: 背景地点をサンプリングする

背景地点数

10000

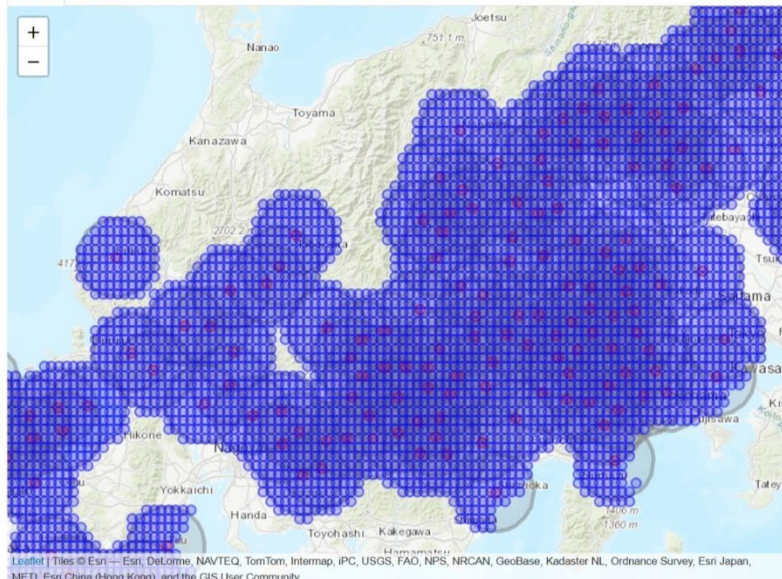
サンプリングを行う

ロメートル。
> マスクされた環境データ。
> 背景点のランダムサンプリングが要求されました (n = 10000)、
しかし 87 % のポイント (n = 8662) しかサンプリングできませんで
した。

背景地図の切り替え

ESRI Topo

地図 在地点の表 結果 手順のガイダンス モジュールガイダンス



在データの分割

この時点ではまだモデルの構築はしていませんが、後でモデルを評価するためにデータをどのように分割するかを決定します。モデルの予測能力を判定するには、理論的には検証するための独立したデータが必要です。独立したデータセットがない場合、1つの解決策として、持っているデータをそれぞれ独立していると考えられるサブセットに分割して、1組を除いたすべてのサブセットで順次モデルを構築し、モデル構築時に除外したサブセットによって評価する方法があります。これは k 倍交差検証（クロスバリデーション、 k は分割数）として知られる方法で、統計学の特に機械学習とデータ科学の分野で広く使われています。順次モデルを構築する作業が終わると、オープン *SDM* はすべての分割セットについて評価統計量を集計(平均)します。これはあくまでもモデルの評価のためのプロセスで、最終的なモデルは**すべてのデータ**を使って構築します。

モデリングのためにどのようにデータを分割するのが最適かについては多数の文献があります。ひとつの選択肢は、無作為に分割を行うことですが、空間データの場合は分割グループが空間的に互いに独立でないというリスクがあります。おそらく確実に良いと考えられる選択肢としては、例えばデータを分割する際に地図上に線を引くなど、空間ブロックの手法があります。空間分割による k 倍交差検証では、評価のための地点はモデルのトレーニングで使用した地域から離れた地域に必ずなります。*F. japonica* については、東北、関東、近畿、九州の各地方、また日本海側と太平洋側では、それぞれの環境条件はかなり異なると考えられます。空間的に分割されたデータのうち、評価用に残したデータをモデルが正確に予測したとすれば、モデルの移行性（訳注：あるデータで構築したモデルが他のデータセットの予測にどれくらい使えるか）は良好と考えられます。つまり、（遠く離れた区域は近い区域よりも環境的に異なるのが普通であるため）予測変数の新しい数値に良好に移行できるということです。オープン *SDM* に搭載しているすべての分割方法の詳細は、ガイダンスを参照してください。

以下は各地点を分割グループに無作為に割り当てる非空間的な無作為 k 倍分割の例です。ここでは $k=4$ とします。

Wallace 導入 1 在データ 2 在データの処理 3 環境データ 4 環境データの処理 5 在データ分割 6 モデル 7 可視化 8 投影 セッションコード

在データの分割

選択可能なモジュール:

☒ 非空間的分割
☐ 空間的分割

モジュール: 非空間的分割

ENMeval: 生態ニッチモデルの自動実行と評価

利用可能なオプション:

Random k-fold

分割数

4

分割

分割グループ番号付きの在地点と背景地点のデータをダウンロード(.csv)

ダウンロード

Module Developers: Jamie M. Kass, Bruno Vilela, Robert R. Anderson

ENMeval references

Package Developers: Robert Muscarella, Peter J. Galante, Mariano Soley-Guardia, Robert A. Boria, Jamie M. Kass, Maria Uriarte, Robert R. Anderson

CRAN | documentation | software note

> 分析範囲: 最小凸多角形。ハッファ後の分析範囲 (ハッファサイズ 1 度)。

> マスクされた環境データ。

> ランダムにサンプリングされた背景地点 (n = 10000)。

> ランダムなkフォールドで分割された在データ (k = 4)。

背景地図の切り替え

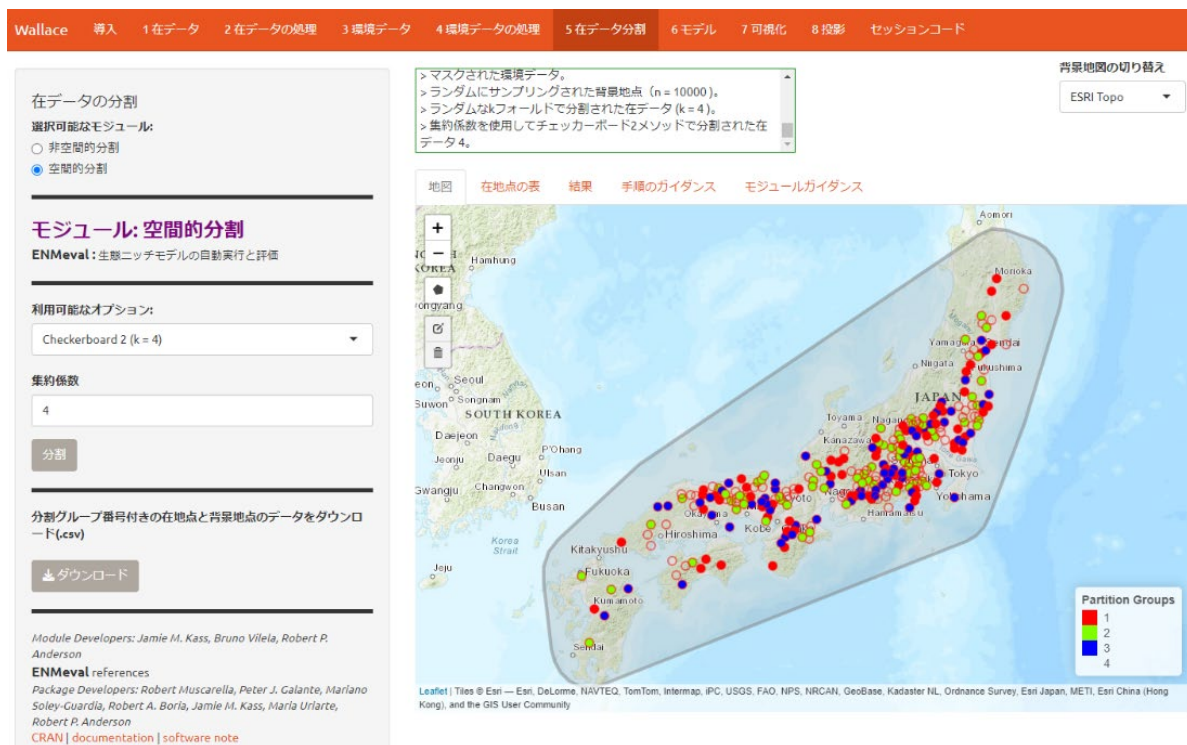
ESRI Topo

地図 在地点の表 結果 手順のガイダンス モジュールガイダンス

Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ, TomTom, Intermap, IPC, USGS, FAO, NPS, NRCAN, GeoBase, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), and the GIS User Community

24

以下は空間的分割の例です。各地点を 4 つの空間分割グループの 1 つに割り当てます。



モデル

さて、モデルを構築する準備ができました。この概要では、Maxent を使用します。Maxent は単純な直線から複雑な曲線まで、幅広い関数型を推定することができる機械学習法です。Maxent についての詳細はガイダンスを参照してください。

画像はより広範なモデル構築の例ですが、時間節約のため、ここでは以下のモデリングオプションを選択します。

- フィーチャクラスとして L、LQ、H を選択します。推定に用いる関数の形状です。
 - L = 線形(例：気温 + 降水量)
 - Q = 二次関数(例：気温² + 降水量²)
 - H = ヒンジ(区分直線関数)
- 正則化乗数を 1 から 3 の間、刻み値 1 で選択
 - 正則化はモデルの複雑性を低減する方法
 - 高い数値 = なめらかであり複雑でないモデル。基本的に、すべての予測変数の係数は漸次縮小され 0 に達するとその変数はモデルから除外され、予測貢献度が最大の変数のみがモデルに残ります。

3つのフィーチャクラスの組み合わせ(L、LQ、H) x 正則化乗数 3つ(1、2、3) = モデル 9種類となります。フィーチャクラス H はかなり複雑な応答曲線も可能であるため、実行にはより単純なモデルと比べて時間がかかります。

最初に実行する際に Maxent ソフトウェアが dismo パッケージのフォルダに入っていない場合、エラーメッセージが出ます。オープン SDM はこの dismo パッケージを使用して Maxent を実行します。この際、オープン SDM はログウィンドウでどこに maxent.jar ファイルを置くべきかを通知します。(これは dismo パッケージを使う以上どうしようもありませんが、wallace では他の方法も検討しています。) この[リンク](#)経由で maxent.jar ファイルをダウンロードして適切なディレクトリに挿入し、もう一度「実行」をクリックして下さい。

Wallace
導入
1 在データ
2 在データの処理
3 環境データ
4 環境データの処理
5 在データ分割
6 モデル
7 可視化
8 投影
セッションコード

ニッチモデルの構築と評価

選択可能なモジュール:

☐ BIOCLIM
☒ Maxent

モジュール: Maxent

ENMeval: Maxent GUI (Java) を使用した生態ニッチモデリングの自動実行と評価
dismo: 種分布モデリング
maxnet: glmnet パッケージを用いて「Maxent」種分布モデルを適用する

(NOTE: 問題がある場合は、モジュールガイダンスにあるトラブルシューティングを参照してください)。

アルゴリズムを選択

☒ maxnet
☐ maxent.jar

フィーチャクラスの選択 (モデルの応答の柔軟さ)

key: L linear, Q quadratic, H hinge, P product

☒ L
☒ LQ
☒ H
☐ LQH
☐ LQHP

正則化乗数の選択 (モデルの複雑さに対する罰則)

☒ 1
☐ 2
☐ 3
☐ 4
☐ 5
☐ 6
☐ 7
☐ 8
☐ 9
☐ 10

正則化乗数のステップ幅

ランダムなkフォールドで分割された在データ (k = 4)。
> 集約係数を使用してチェッカーボード2メソッドで分割された在データ 4。
> Maxentが正常に動作しました (使用パッケージ: maxnet)。9 の評価結果とクラмпされたモデルを出力します。

背景地図の切り替え

ESRI Topo

現在の環境変数

bio02

現在のモデル

rm_1_fc.L

地図

在地点の表

結果

手順のガイダンス

モジュールガイダンス

Evaluation

Lambdas

フルモデルとクロスバリデーション分割の平均評価統計量

	rm	fc	tune.args	auc.train	cbi.train	auc.diff.avg	auc.diff.sd	auc.val.avg	auc.val.sd	cbi.val.avg
1	1	L	rm_1_fc.L	0.744	0.985	0.034	0.016	0.732	0.028	0.91
2	2	L	rm_2_fc.L	0.735	0.986	0.027	0.016	0.724	0.021	0.92
3	3	L	rm_3_fc.L	0.722	0.987	0.022	0.019	0.71	0.017	0.92
4	1	LQ	rm_1_fc.LQ	0.77	0.992	0.039	0.026	0.759	0.038	0.94
5	2	LQ	rm_2_fc.LQ	0.767	0.988	0.036	0.028	0.757	0.036	0.9
6	3	LQ	rm_3_fc.LQ	0.763	0.975	0.034	0.03	0.753	0.035	0.91
7	1	H	rm_1_fc.H	0.795	0.972	0.043	0.043	0.757	0.034	0.91
8	2	H	rm_2_fc.H	0.781	0.985	0.038	0.033	0.754	0.034	0.94
9	3	H	rm_3_fc.H	0.772	0.984	0.036	0.029	0.749	0.031	0.93

Previous
1
Next

個々のクロスバリデーション分割の評価統計量

	tune.args	fold	auc.val	auc.diff	cbi.val	or.mtp	or.10p
1	rm_1_fc.L	1	0.732	0.028	0.91		
2	rm_2_fc.L	1	0.724	0.021	0.92		
3	rm_3_fc.L	1	0.71	0.017	0.92		
4	rm_1_fc.LQ	1	0.759	0.038	0.94		
5	rm_2_fc.LQ	1	0.757	0.036	0.9		
6	rm_3_fc.LQ	1	0.753	0.035	0.91		
7	rm_1_fc.H	1	0.757	0.034	0.91		
8	rm_2_fc.H	1	0.754	0.034	0.94		
9	rm_3_fc.H	1	0.749	0.031	0.93		

結果は2つの評価統計量の表として表示され、構築した複数のモデルを比較することができます。表は選択したフィーチャクラス／正則化乗数の組み合わせごとに、1行ずつ、合わせて9行あります。4個の分割グループから構築したモデルの統計量が第1の表では平均化されており、第2の表では2グループそれぞれからの統計量が別々に表示されます。評価用に留保しておいた分割データで算出した統計量には「test」のラベルが付いています。

いかにして「最良」のモデルを選ぶか。これについては山のような文献があり、実際のところ1つの答えがあるわけはありません。AUCとOR(オMISSION率、訳注：観測された在地点のうち、モデル予測では不在と予測されている地点の割合)は上記の分割データを使用して計算されますが、AICはそうではなくモデル構築した範囲全域(とすべての間引き後の在地点)での予測を使って計算されています。AICは交差検証の結果を組み入れませんが、モデルの複雑さをペナルティとして明示的に考慮しており、モデルのパラメータが多いほどAICスコアが悪くなります。いずれのモデルを最良として選ぶかの実際の判断はユーザーの判断に委ねられますが、ガイダンスにはさら

に判断のための知識を深めるのに有用な参考文献を収録しています。

この概要では、交差検証の結果について一連の基準を使用します。まず、予測において観測された在のみデータのオMISSIONが少なかったモデルを優先します。結果の表を or.10p.avg、つまり教師データの 10 パーセントル値をしきい値として適用したテストデータの値でソートします(詳しくはガイダンスを参照)。ここで使用しているモデルのセットでは、最低の or.10p.avg の値をとったのが L_1(一次関数と正則化乗数 1)の 1 つのモデルです。(訳注：在地点を漏れなく在と予測することを重視する場合、たとえば希少種の生育地を漏れなく発見したい場合などは、OR = omission rate の低さを基準としてモデルを選ぶのが妥当だが、そうした場合、在地点がない範囲を在と予測してしまう割合も高くなる。在を漏らさないことと、過剰に在と予測することのバランスがとれた、全体的な精度の高さを重視してモデルを選択する場合は、AIC c や AUC、TSS を優先的に考慮することも多い)

Wallace 導入 1 在データ 2 在データの処理 3 環境データ 4 環境データの処理 5 在データ分割 6 モデル 7 可視化 8 投影 セッションコード

二ツチモデルの構築と評価

選択可能なモジュール:

☐ BIOCLIM

☒ Maxent

モジュール: Maxent

ENMeval: Maxent GUI (Java) を使用した生態ニッチモデリングの自動実行と評価

dismo: 種分布モデリング

maxnet: glmnet パッケージを用いて「Maxent」種分布モデルを適用する

(NOTE: 問題がある場合は、モジュールガイダンスにあるトラブルシューティングを参照してください。)

アルゴリズムを選択

☒ maxnet ☐ maxent.jar

フィーチャークラスの選択 (モデルの応答の柔軟性)

key: L linear, Q quadratic, H inge, P product

☒ L ☒ Q ☒ H ☐ LQH ☐ LQHP

正則化乗数の選択 (モデルの複雑さに対する罰則)

0.5 1 3 10

0.5 1.5 2.5 3.5 4.5 5.5 6.5 7.5 8.5 9.5 10

正則化乗数のステップ値

1

> ランダムなkフォールドで分割された在データ (k=4)。
> 集約係数を使用してチェックボード2メソッドで分割された在データ 4。
> Maxentが正常に動作しました (使用パッケージ: maxnet)。9 の評価結果とクランプされたモデルを出力します。

背景地図の切り替え
ESRI Topo

現在の環境変数
bio02

現在のモデル
rm.1_fc.L

地図 在地点の表 結果 手順のガイダンス モジュールガイダンス

Evaluation Lambdas

フルモデルとクロスバリデーション分割の平均評価統計量

c.diff.avg	auc.diff.sd	auc.val.avg	auc.val.sd	cbi.val.avg	cbi.val.sd	or.10p.avg	or.10p.sd	or.mtp.avg	o
0.034	0.016	0.732	0.028	0.911	0.086	0.099	0.042	0.004	
0.027	0.016	0.724	0.021	0.928	0.06	0.11	0.036	0.004	
0.036	0.028	0.757	0.036	0.93	0.011	0.115	0.08	0.007	
0.038	0.033	0.754	0.034	0.944	0.016	0.117	0.049	0.011	
0.034	0.03	0.753	0.035	0.913	0.014	0.127	0.109	0.003	
0.022	0.019	0.71	0.017	0.922	0.016	0.128	0.028	0.004	
0.039	0.026	0.759	0.038	0.947	0.019	0.128	0.054	0.007	
0.036	0.029	0.749	0.031	0.934	0.009	0.133	0.06	0.008	
0.043	0.043	0.757	0.034	0.916	0.043	0.139	0.058	0.022	

Previous 1 Next

個々のクロスバリデーション分割の評価統計量

tune.args	fold	auc.val	auc.diff	cbi.val	or.mtp	or.10p
-----------	------	---------	----------	---------	--------	--------

もし同じスコアがある場合、同じスコアのモデルの中で、auc.val.avg(平均評価 AUC、訳注: val とは、交差検証で評価用に残したデータを使って算出された統計量であること、平均とは交差検証の全ての分割での平均値であることを示す)が最大のモデルを選ぶことができます。

Wallace 導入 1 在データ 2 在データの処理 3 環境データ 4 環境データの処理 5 在データ分割 6 モデル 7 可視化 8 投影 セッションコード

ニッチモデルの構築と評価

選択可能なモジュール:

☐ BIOCLIM
☒ Maxent

モジュール: Maxent

ENMeval: Maxent GUI (Java) を使用した生態ニッチモデリングの自動実行と評価
dismo: 確分布モデリング
maxnet: glmnet パッケージを用いて「Maxent」確分布モデルを適用する

(NOTE: 問題がある場合は、モジュールガイダンスにあるトラブルシューティングを参照してください)。

アルゴリズムを選択

☒ maxnet
☐ maxent.jar

フィーチャークラスの選択 (モデルの応答の柔軟さ)

key: L linear, Q quadratic, H Inge, P roduct

☒ L
☒ Q
☒ H
☐ LQH
☐ LQHP

正規化乗数の選択 (モデルの複雑さに対する罰則)

1

正規化乗数のステップ値

1

クラumpingしますか?

☒

実行

CSVをダウンロード

Module Developers: Jamie M. Kass, Gonzalo E. Pinilla-Buitrago, Robert Muscarella, Bruno Vilela, Robert P. Anderson

ENMeval references

Package Developers: Robert Muscarella, Peter J. Galante, Mariano Soley-Guardia, Robert A. Borja, Jamie M. Kass, Maria Uriarte, Robert P. Anderson

CRAN | documentation | software note

dismo references

Package Developers: Robert J. Hijmans, Steven Phillips, John Leathwick, Jane Elith

CRAN | documentation

maxnet references

Package Developers: Steven Phillips

CRAN | documentation

ランダムなkフォールドで分割された在データ (k = 4)。

> 集約係数を使用してチェックボード2メソッドで分割された在データ。

> Maxentが正常に動作しました (使用パッケージ: maxnet)。9の評価結果とクラumpされたモデルを出力します。

現在の環境変数

bio02

現在のモデル

rm.1_fc.L

地図

在地点の表

結果

手順のガイダンス

モジュールガイダンス

Evaluation

Lambdas

フルモデルとクロスバリデーション分割の平均評価統計量

	fc	tune.args	auc.train	cbi.train	auc.diffavg	auc.diffsd	auc.valavg	auc.valsd	cbi.valavg	cbi.v
L	rm.1_fc.L	0.744	0.985	0.034	0.016	0.732	0.028	0.911		
L	rm.2_fc.L	0.735	0.986	0.027	0.016	0.724	0.021	0.928		
LQ	rm.2_fc.LQ	0.767	0.988	0.036	0.028	0.757	0.036	0.93		
H	rm.2_fc.H	0.781	0.985	0.038	0.033	0.754	0.034	0.944		
LQ	rm.3_fc.LQ	0.763	0.975	0.034	0.03	0.753	0.035	0.913		
L	rm.3_fc.L	0.722	0.987	0.022	0.019	0.71	0.017	0.922		
LQ	rm.1_fc.LQ	0.77	0.992	0.039	0.026	0.759	0.038	0.947		
H	rm.3_fc.H	0.772	0.984	0.036	0.029	0.749	0.031	0.934		
H	rm.1_fc.H	0.795	0.972	0.043	0.043	0.757	0.034	0.916		

Previous 1 Next

個々のクロスバリデーション分割の評価統計量

	tune.args	fold	auc.val	auc.diff	cbi.val	or.mtp	or.10p
1	rm.1_fc.L	1	0.711	0.041	0.785	0	0.075
2	rm.1_fc.L	2	0.757	0.017	0.933	0	0.099
3	rm.1_fc.L	3	0.704	0.052	0.953	0.016	0.159
4	rm.1_fc.L	4	0.757	0.024	0.975	0	0.066
5	rm.2_fc.L	1	0.716	0.026	0.839	0	0.075
6	rm.2_fc.L	2	0.744	0.015	0.944	0	0.113
7	rm.2_fc.L	3	0.698	0.05	0.962	0.016	0.159
8	rm.2_fc.L	4	0.738	0.016	0.967	0	0.092
9	rm.3_fc.L	1	0.71	0.02	0.901	0	0.134
10	rm.3_fc.L	2	0.725	0.01	0.936	0	0.127

Previous 1 2 3 4 Next

また、上の表が平均値のみを列記しているのに対し、下の表では交差検証の分割ごとの評価統計が表示されています。

アルゴリズムを選択

☒ maxnet
☐ maxent.jar

フィーチャークラスの選択 (モデルの応答の柔軟さ)

key: L linear, Q quadratic, H Inge, P roduct

☒ L
☒ Q
☒ H
☐ LQH
☐ LQHP

正規化乗数の選択 (モデルの複雑さに対する罰則)

1

正規化乗数のステップ値

1

クラumpingしますか?

☒

実行

CSVをダウンロード

Module Developers: Jamie M. Kass, Gonzalo E. Pinilla-Buitrago, Robert Muscarella, Bruno Vilela, Robert P. Anderson

ENMeval references

Package Developers: Robert Muscarella, Peter J. Galante, Mariano Soley-Guardia, Robert A. Borja, Jamie M. Kass, Maria Uriarte, Robert P. Anderson

CRAN | documentation | software note

dismo references

Package Developers: Robert J. Hijmans, Steven Phillips, John Leathwick, Jane Elith

CRAN | documentation

maxnet references

Package Developers: Steven Phillips

CRAN | documentation

LQ

rm.3_fc.LQ

0.763

0.975

0.034

0.03

0.753

0.035

0.913

L

rm.3_fc.L

0.722

0.987

0.022

0.019

0.71

0.017

0.922

LQ

rm.1_fc.LQ

0.77

0.992

0.039

0.026

0.759

0.038

0.947

H

rm.3_fc.H

0.772

0.984

0.036

0.029

0.749

0.031

0.934

H

rm.1_fc.H

0.795

0.972

0.043

0.043

0.757

0.034

0.916

Previous 1 Next

個々のクロスバリデーション分割の評価統計量

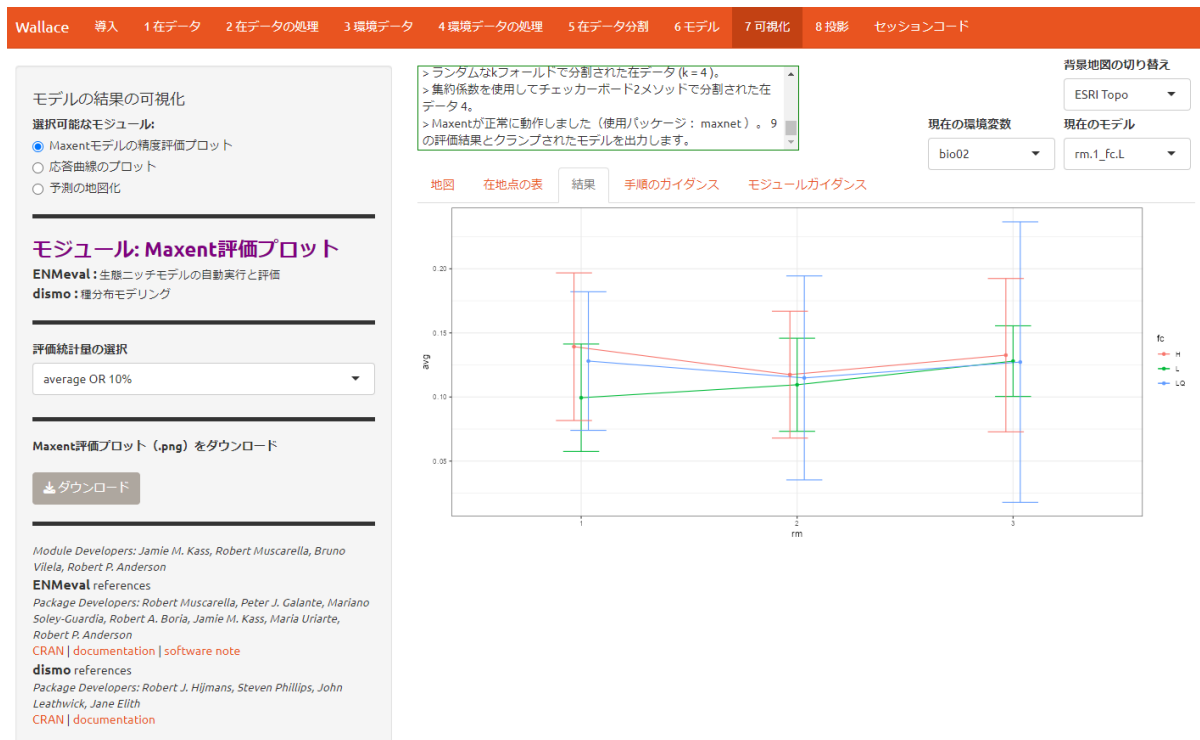
	tune.args	fold	auc.val	auc.diff	cbi.val	or.mtp	or.10p
1	rm.1_fc.L	1	0.711	0.041	0.785	0	0.075
2	rm.1_fc.L	2	0.757	0.017	0.933	0	0.099
3	rm.1_fc.L	3	0.704	0.052	0.953	0.016	0.159
4	rm.1_fc.L	4	0.757	0.024	0.975	0	0.066
5	rm.2_fc.L	1	0.716	0.026	0.839	0	0.075
6	rm.2_fc.L	2	0.744	0.015	0.944	0	0.113
7	rm.2_fc.L	3	0.698	0.05	0.962	0.016	0.159
8	rm.2_fc.L	4	0.738	0.016	0.967	0	0.092
9	rm.3_fc.L	1	0.71	0.02	0.901	0	0.134
10	rm.3_fc.L	2	0.725	0.01	0.936	0	0.127

Previous 1 2 3 4 Next

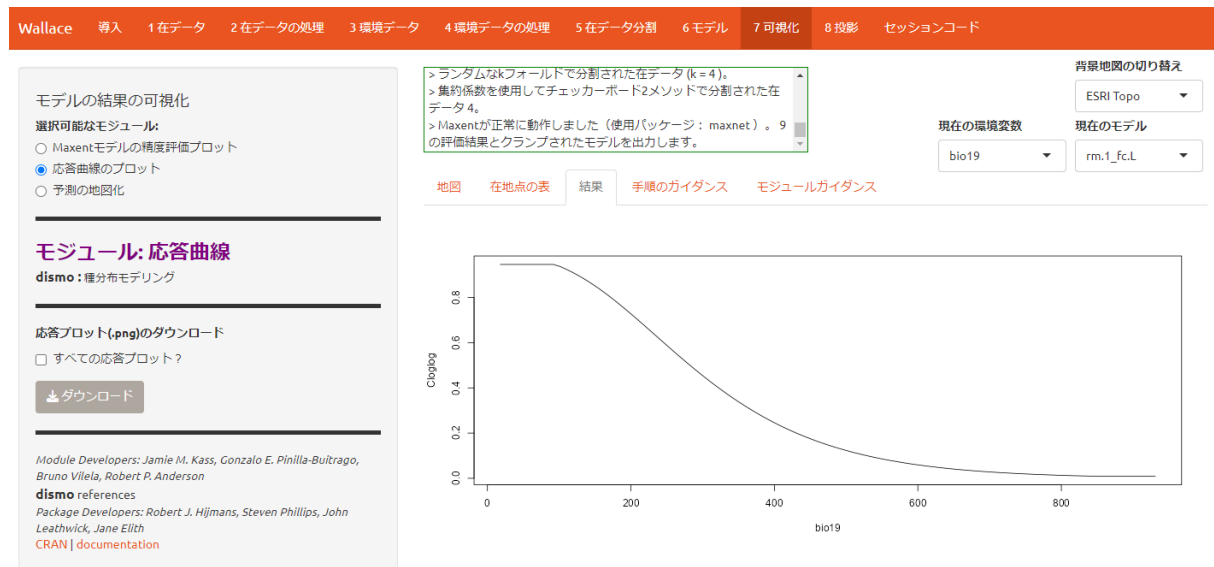
もし AICc が最小のモデルを選んだ場合はどれが選ばれるかなど、自身で分析を行って比べてみて下さい。

可視化

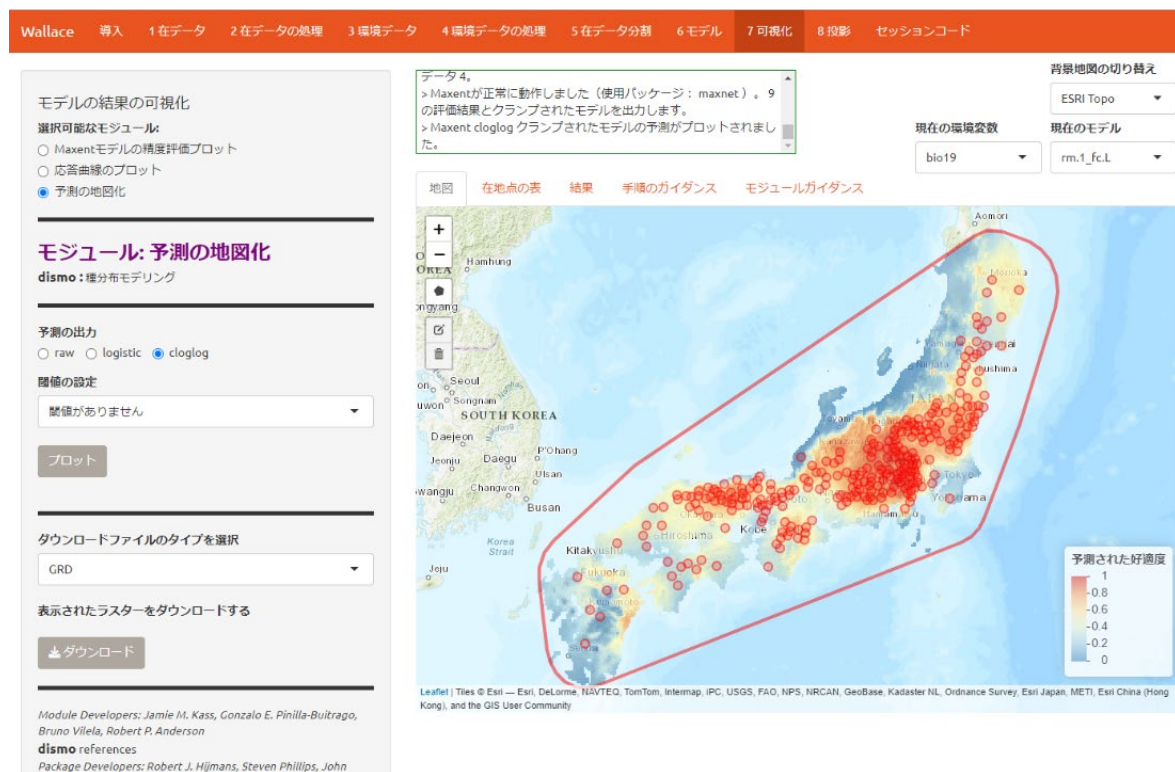
「Maxent モデルの制度評価プロット」モジュールはユーザーが複数のモデル全体でパフォーマンス統計量を評価することを可能とします。以下で、(10 パーセンタイルにトレーニング地点閾値 average OR 10%を使って)正則化乗数がどのように平均テストオMISSION率に作用するかを見てみます。



ここでは**応答曲線**についても調べてみます。これは、各変数(x 軸)の値の変化に伴って予測される生育地としての好適度(y 軸)がどのように変化するかを示すものです。特定のモデルの結果を見るには、「現在のモデル」のドロップダウンメニューからそのモデルを選択します。下図はモデル rm.1_fc.L(L_1 一次関数と正則化乗数 1)の応答曲線の 1 つ、bio19 (最寒四半期の降水量) についての応答曲線です。降水量の低い範囲では応答曲線が平坦で好適度は安定しているものの、200 以上では *F. japonica* の生育の好適度が急激に低下します。

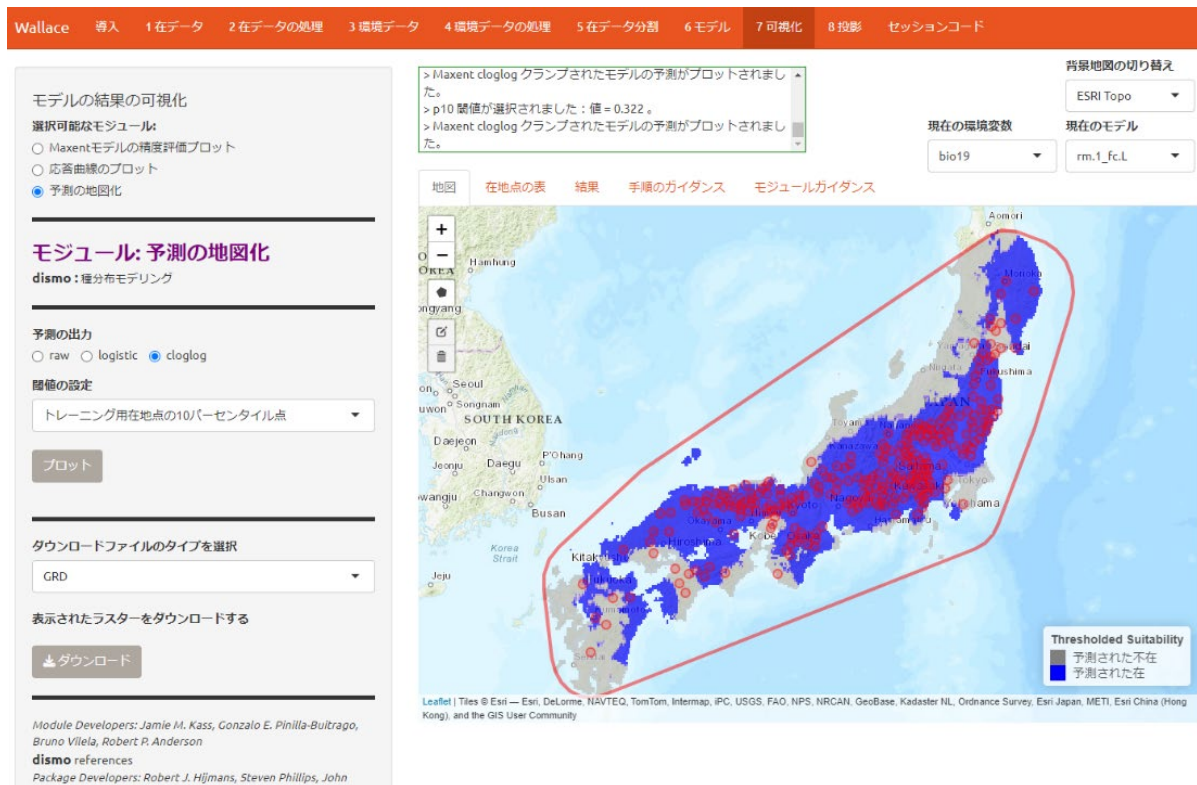


地図上でモデル予測を可視化することもできます。好適度の予測は(一定範囲の)連続値またはバイナリ(連続値を、閾値によって 0 と 1 に二値化したもの)のどちらでも可視化することができます。Maxent モデルの出力スケールと閾値化のルールについては、モジュールのガイダンスを参照してください。下図はモデル LQ_1 の閾値なし（訳注：0-1 の間の値を取るように cloglog 変換された連続値）の予測地図です。



下の画面は同じモデルの予測地図ですが、今度は閾値をトレーニング用地点の 10 パーセンタイル点に設定します。これは現在使用可能な 2 つの二値化ルールのうち、厳格な方です。在地点の中には *F. japonica* の生育に適した区域を表す青の区域から外れているものがあります。閾値をより厳格でない最小トレーニング地点閾値に設定して、予測地図を作成してみるとまた違う結果になります。（訳注：10 パーセンタイル閾値は、教師デ

ータとして用いた在地点での予測値のうち、小さいほうから 10%に相当する値を閾値とするものであるため、当然、1 割程度の在地点が好適範囲から外れると予期される。最小トレーニング地点閾値は、教師在地点の予測値で最小のものを閾値とするので、ほぼすべての在地点が好適範囲に含まれる。他にも閾値は多数あるため、ガイダンスの参考文献を参照することが望ましい。)



投影

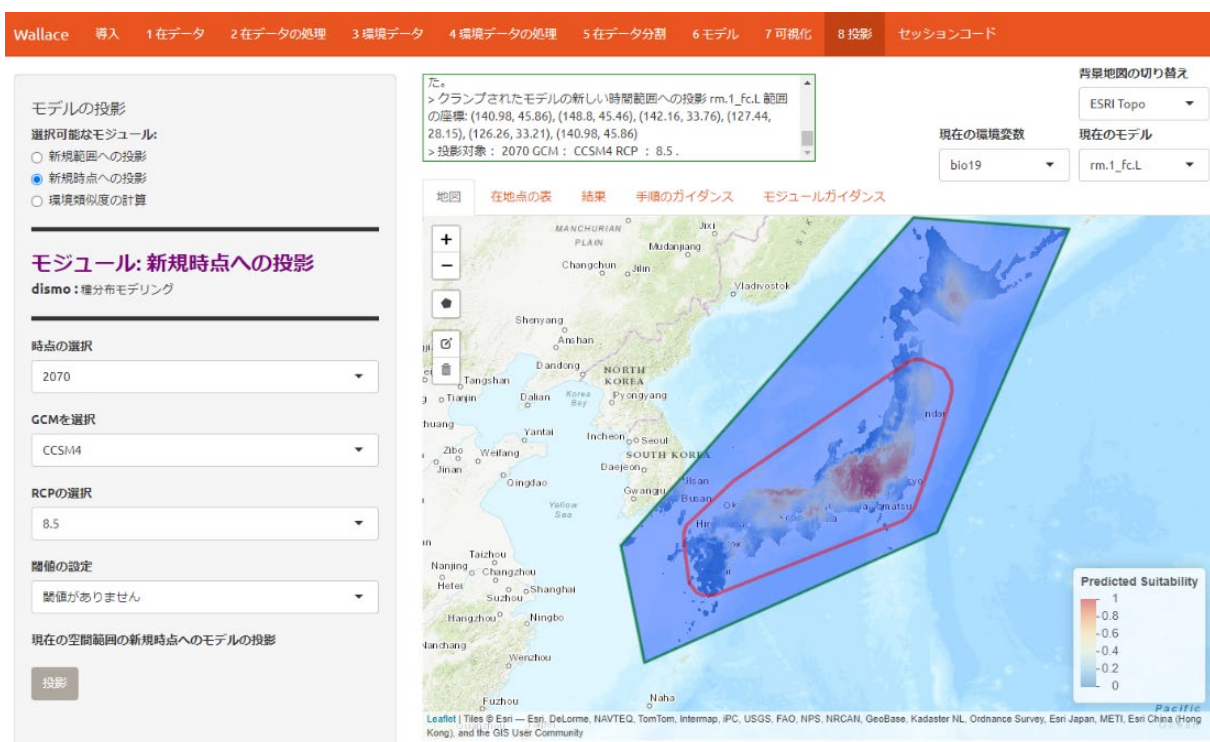
構築したモデルは別の場所や未来の気候シナリオを使って投影することができます。「投影」とは、単に(モデル構築に使用されたものではない)新しい環境値をモデルに挿入して、新しい応答(新しい変数範囲における好適度予測)を得ることです。

「交差検証のステップでもこれをしたのではなかったか？」などと、混乱するかも知れません。空間分割による交差検証ではモデルに新たな区域を順次予測させ、モデルの移行性を評価統計量により要約しました。しかし、今ここで見ている予測に使用した最終的なモデルは、**すべてのデータを使って**(つまり分割グループを一切除外せずに)構築されたものです。したがって、データセットの中の背景地点すべてにおける環境変数の範囲が、モデル構築のプロセスで使われています。

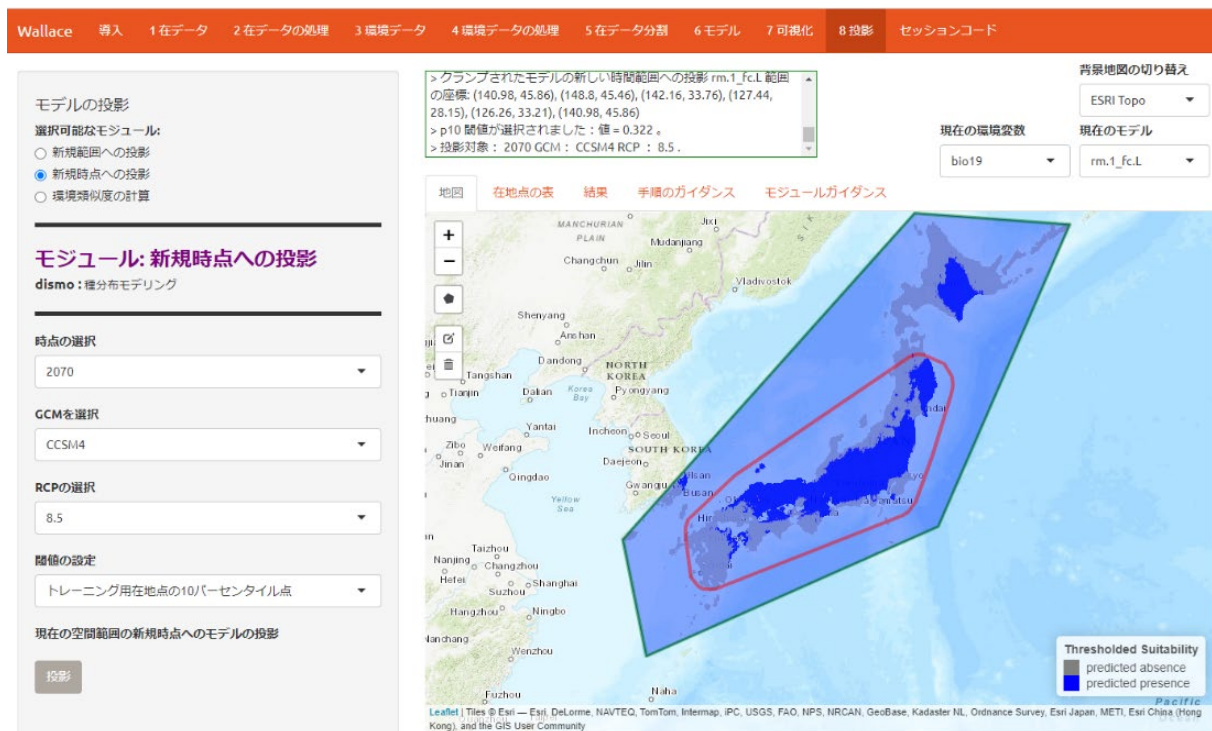
投影では、場所や時間の異なる環境値を使うため、環境変数の値の範囲が、モデル構築したときとはまったく異なるものになり得ます。あまりに違う環境値での投影は、その結果の正確性に疑問が生じることもあります。こうした「非アナログ条件」の詳細については、ガイダンスを参照してください。

LQ_1 はオMISSION率が低く(既知の在のみデータほとんどすべてを予測可能)、平均テスト AUC が高く(移行性に優れている)なっています。下の画像では、L_1 モデルが厳しい気候シナリオ(代表濃度経路シナリオ(RCP)8.5)の下で 2070 年に投影されています。将来気候のモデル化は様々な試みがあり、数種類の全球気候モデル(GCM)からの選択が可能です。このチュートリアルでは CCSM4 を使用しています。RCP および GCM についての詳細は、モジュールのガイダンスを参照してください。

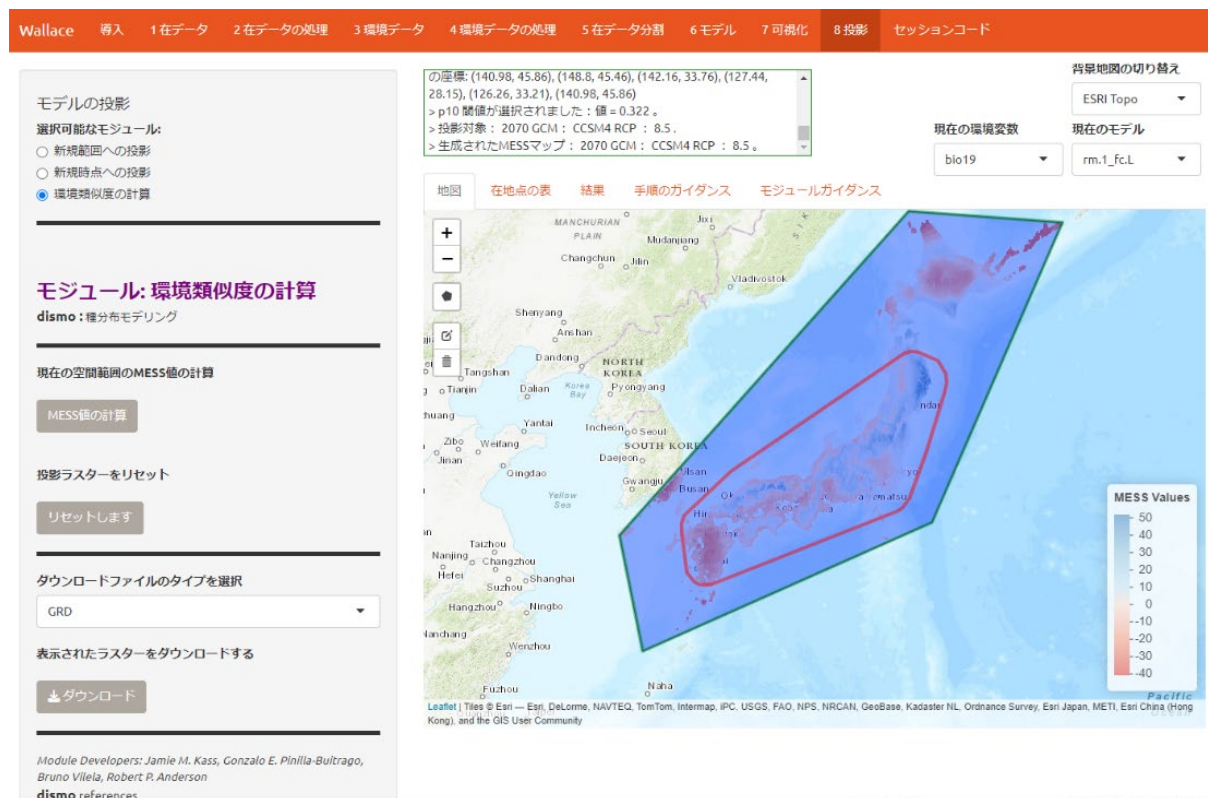
構築したモデルを投影するには、描画ツールバーを使用して多角形を描き、「Finish」をクリックします。次に年、GCM および RCP を選択し、「**投影**」ボタンをクリックして新しい地図を構築します。ラスタは WorldClim のデータベースから取得しますが、すべての GCM にそれぞれの RCP についてのラスタデータが存在するわけではありません。以下は連続予測の例です。



下の画像は、トレーニング 10 パーセンタイル閾値によって二値変換した投影です。現在と未来の投影図の間の差異の一部がやや明確になったのがわかります。



先に述べたように、ここで設定した新しい環境値の範囲には、モデル構築に使用した値とはかけ離れた値が含まれている可能性があり、その場合、予測に大きな不確実性が生じます。このような環境値をとる区域を可視化するには、MESS 地図を描く方法があります。MESS は多変量(Multivariate)、環境(Environmental)、類似性(Similarity)、面(Surface)の頭文字を取ったものです(詳細はモジュールのガイダンスを参照)。下图がここで用いた例での MESS です。色の濃い値はすべてモデル構築で使用したものと類似した環境値を持つ区域をしめしており、色が薄いほど差異が大きくなります。未来の気候値は本州以南において、高標高域では類似しているものの、特に九州のように、西部から沿岸部の低標高域では場所によって大きな差異があります。また、現時点で分布がない北海道は広い範囲で大きな差異があります。したがって、投影した好適度はこれら区域では不確実性が高いと解釈することができます。



コードの抽出

他の GUI ベースのソフトウェアと比較したときの *Wallace* の大きな利点は、分析で実行したコードをすべて抽出できることにあります。*Wallace* を使用している間、背景で多くの R コードが実行されます。このオプションでは、実行したコードすべての簡略バージョンを圧縮して、注釈付きのスクリプト形式でダウンロードし、このスクリプトを使って分析の再実行、共有、または修正が可能になります。スクリプトはいくつかの方法でダウンロードできます。R コードとテキストを組み合わせるのに便利な **R Markdown** 形式は、R で直接実行することができますが、.pdf 形式でダウンロードする場合は、使用するシステムに TeX のいずれかのバージョンがインストールされている必要があります。詳細はこのページのテキストを参照してください。

スクリプトをダウンロードするには、「Rmd」を選択して、「セッションコードのダウンロード」をクリックします。

セッションコードのダウンロード

rmarkdown: Rの動的ドキュメント
knitr: Rで動的レポートを生成するための汎用パッケージ

ダウンロードファイルのタイプを選択

Rmd

セッションコードのダウンロード

Module Developers: [Jamie M. Kass](#), [Gonzalo E. Pinilla-Buitrago](#), [Bruno Vilela](#), [Robert P. Anderson](#)
rmarkdown references
 Package Developers: [JJ Allaire](#), [Joe Cheng](#), [Yihui Xie](#), [Jonathan McPherson](#), [Winston Chang](#), [Jeff Allen](#), [Hadley Wickham](#), [Aron Atkins](#), [Rob Hyndman](#), [Ruben Arslan](#)
[CRAN | documentation](#)
knitr references
 Package Developers: [Yihui Xie](#)
[CRAN | documentation](#)

R パッケージ **rmarkdown** を使用すると、*Wallace* ユーザーは 現在のセッションで実行されるアクション (および基になるコード) の履歴をダウンロードできます。これには、複数のフォーマット (Rmd (R Markdown)、.pdf、.html、または.doc) が含まれます。.Rmd 形式は、R セッションで実行した場合に分析を再現する実行可能な R スクリプト ファイルです。このファイルは、プレーンテキストとRコード「チャンク」で構成されています。

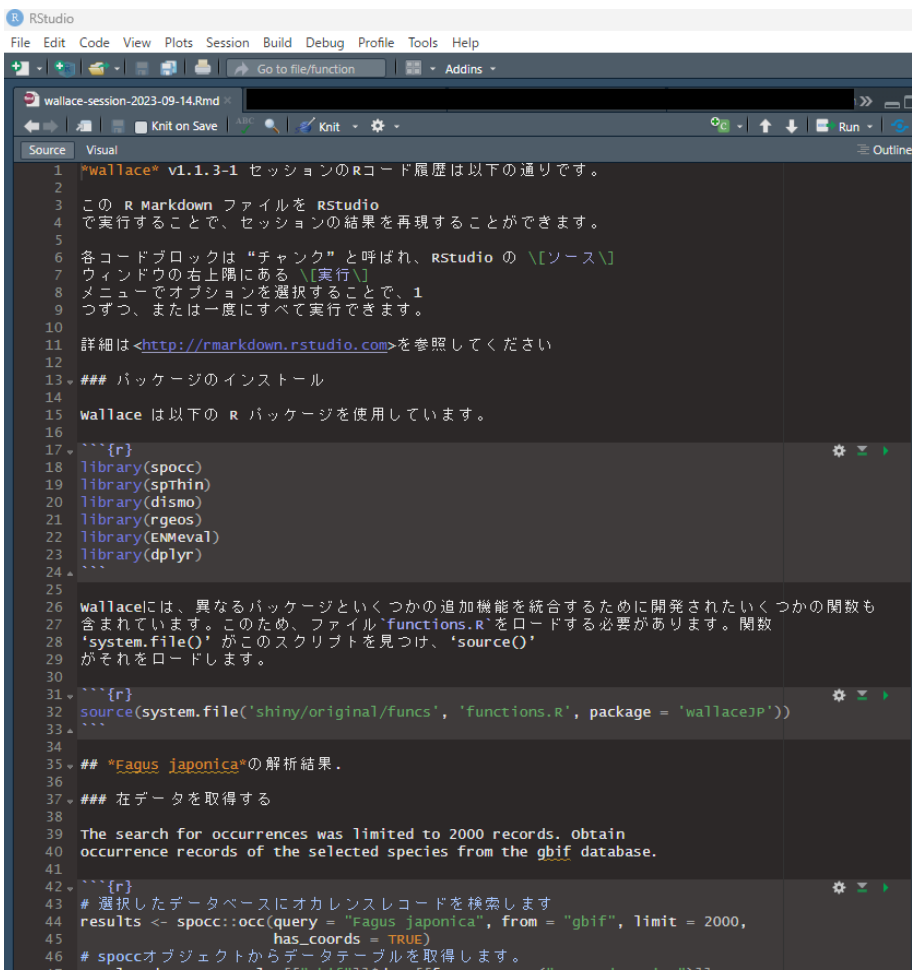
RStudio には R Markdown ファイルのための拡張機能があります。RStudio で .Rmd を開き、右上の「実行」をクリックするだけで、チャンクごと、または一度にすべてのチャンクを実行することができます。詳細については、RStudio [チュートリアル](#)をご覧ください。

注意事項

セッションコードのPDFを作成するには、動作するバージョンのTeXがインストールされている必要があります。以下の手順を実行してください。

1. Rstudio で `install.packages("tinytex")` を実行。このコマンドは、tinytex という R パッケージをインストールします。
2. Rstudio で `tinytex::install_tinytex()` を実行。このコマンドは、TinyTeX 本体をインストールします。

これで分析を完全収録した.Rmd ファイルを取得できました。**R Markdown** ファイルは通常のテキストをひと塊のコードと組み合わせたものです。*Wallace* のモジュールは###を添えたヘッダとして表示されます。[ここ](#)に Rmd の構文についての簡単な参考例があります。



新しい R ウィンドウを開いてこれらのコードの一部を実行しようとする場合、コードの後半部分はそれより前の部分の結果に依存しているため、省略して実行すると全体が実行されない場合があることに留意してください。また、オープン SDM のセッションを終了すると、ウェブブラウザではそこまでの解析結果が失われることにも留意してください(ただし.Rmd には影響ありません)。RStudio を使用している場合は、この Rmd を開いて「**knit**」をクリックすると、ワークフローを共有可能な html 文書にコンパイルすることができます。

作業工程を進める上で、このコードでは自由に変更できます。*Wallace* の後続バージョンでは、修正した.Rmd をアップロードして、選択したすべてのオプションを取り込み、GUI でそれまでの分析を中断したところから再開できるようにすることが検討されています。

オープン SDM の現行バージョンではモデルのあと処理についてまだ何も組み込まれていませんが、セッション終了後に.Rmd を修正して分析に組み込む作業を R ですることができます。後処理の例としては、複数の種のモデルを積み重ねて種多様性を推定したり、複数モデルの比較による重複ニッチの推定などがあります。現在、wallace チームはこの分野を専門とするパートナーとの開発作業を進めており、*Wallace* の後続バージョンにはモデルの後処理の機能が含まれる予定です。

結論

以上がオープン SDM の概要です。その用途、特色、生物種分布とニッチモデリングの概要について知識を深めていただけたかと思います。繰り返しになりますが、ガイダンスを読み、推奨されている出版物をフォローアップし、そこから他の関連する出版物に目を通してさらに知識を広げていかれることを強く奨励します。

また、E メールあるいは「はじめに」の部分で挙げたウェブサイトから、お気軽にお問合せ下さい（訳注：オープン SDM についての問い合わせ先：opensdm@nies.go.jp）。*Wallace* とオープン SDM をすべてのユーザーにとってより良いものにするため、ぜひ皆様のアイデアや意見、提案をお寄せください。

謝辞

オープン SDM は国立環境研究所の気候変動適応研究プログラムにおいて作成されました。

Wallace は 2015 年度 Ebbe Nielsen Challenge of the Global Biodiversity Information Facility (GBIF) のファイナリストに認定され、開発資金を授与されました。

本文書の原文は、National Science Foundation の交付金番号 DBI-1661510 (RPA), DBI-1650241 (RPA), DEB-1119915 (RPA), DEB-1046328 (MEA)および DBI-1401312 (RM)による支援により実施された研究に基づいています。本文書中の意見、知見、および結論または推奨事項は著者のものであり、必ずしも National Science Foundation の見解を反映したものではありません。

上記助成金の他、JMK は CUNY Science Scholarship および CUNY Graduate Center Provost Digital Innovation Grant Additional を授与されています。