

ランダムフォレストを用いた全国における哺乳類の空間分布モデリング Random forests as a tool for modelling mammal distributions in Japan

○福田信二¹・齋藤志保²・大里耕司¹・梶光一¹
○FUKUDA Shinji, SAITO Shiho, OSATO Koji, KAJI Koichi

1. はじめに

近年、野生動物の個体数増加と分布拡大に伴い、全国において農林業被害が拡大している。平成26年度の全国の野生動物による農作物被害は約191億円に上り、そのうちの約7割はシカ、イノシシ、サルによるものである（農水省、2016）。これに対し、分布域の変化や拡大に影響する諸要因に関する研究が行われており、ニホンジカの分布拡大域の解析（奥村ら、2009）や複数の哺乳類の分布拡大要因に関する検討（Saito *et al.*, 2016）等が例として挙げられる。一方、日本全国を対象にした研究は少ないことから、近年、容易に利用可能になった高精度な解析手法を用いた現状把握や原因の究明、将来予測等に向けた研究事例の集積が求められる。そこで本研究では、ランダムフォレスト（RF；Breiman, 2001）を用いて、全国レベルでの哺乳類（ニホンジカ、イノシシ、ニホンザル）の空間分布について解析した。その際、野生動物の分布の公開データから空間分布モデルを構築し、RFから得られる変数の重要度と応答曲線に基づいて、対象種の分布域の環境特性とその変化について検討した。

2. 方法

対象種の分布データには、環境省の自然環境保全基礎調査の第2回（1978年）と第6回（2003年）の結果を使用した。同データは、5倍地域メッシュ（約5 km四方）で、対象種の分布の確認状況により、出現メッシュ／非出現メッシュを定義した。環境データには、国土数値情報の土地利用、標高・傾斜度、30年気候平年値（いずれも3次メッシュ（約1 km四方））をそれぞれ5倍メッシュ化して使用した。以上の分布データと環境データを対象種と調査年ごとに統合した。

対象種の空間分布の解析には、統計ソフトR（R Core Team, 2015）のパッケージ「randomForest」（Liaw and Wiener, 2002）を使用した。RFは、多数の分類回帰木を用いた機械学習の一手法であり、再現精度が高く、計算速度が速いこと、説明変数が多い場合や説明変数間に相関がある場合でも効率的に動作すること、説明変数の重要度を評価できる等の特長がある（Cutler *et al.*, 2007）。ここでは、21項目の環境変数を説明変数とし、対象種の出現／非出現を応答変数とする空間分布モデルを対象種と調査年ごとに構築した。その際、RF内部で使用される乱数による結果のばらつきを評価するために、乱数の種を50回変化させた。モデルの再現性の評価には、AUC、正答率（CCI）、平均二乗誤差（MSE）、Nash-Sutcliffe Efficiency（NSE）を用いた。AUC、CCI、NSEは1に近いほど、MSEは0に近いほど精度が高いことを意味する。

本研究では、まず空間分布モデルを構築し、モデルから出力された出現確率を用いて分布ポテンシャルマップを作成した。さらに、ランダムフォレストから計算される変数の重要度と応答曲線により、対象種の空間分布特性を定量的に評価した。

3. 結果と考察

1978年と2003年のモデルは、全ての評価指標において非常に高い再現性を示したことから、野生動物の空間分布解析におけるRFの有用性が確認できた。RFで評価された変数の重要度をFig. 1に示す。両モデルとも、気温の重要度が最も高く、次いで積雪深や降水量、日射量が高か

¹ 東京農工大学大学院農学研究院 Institute of Agriculture, Tokyo University of Agriculture and Technology

² 東京農工大学農学部 Faculty of Agriculture, Tokyo University of Agriculture and Technology

った。これにより、対象種の分布は、気温や積雪深等の気象条件によって特徴づけられる可能性が示唆された。気温や降水量、日射量は植生タイプとの関係性が考えられ、ニホンジカの分布は積雪深やササのタイプによって制限されるという既往の研究 (Kaji *et al.*, 2000) と一致する。

RFから得られた応答曲線のうち、Fig. 2に示す6変数について特徴的な傾向がみられた。例えば、水田、その他の農地、森林については、1978年のモデルでは、いずれも面積が大きい条件において分布ポテンシャルが高かったが、2003年のモデルでは、その他の農地および森林の面積の大きい条件で分布ポテンシャルの低下がみられた。気温と日射量に関する1978年のモデルでは、数値の小さい部分 (約5°C以下, 約11 MJm⁻²) と大きい部分 (約18°C以上, 約14.5 MJm⁻²) の両方で分布ポテンシャルが高かったが、2003年のモデルでは数値の小さい部分 (約3°C, 約12 MJm⁻²) にピークがみられた。積雪深については、1978年のモデルでは積雪深が小さいほど分布ポテンシャルが高い傾向がみられたが、2003年のモデルでは60~90 cmの間がピークとなっていた。これらの変化は、ニホンジカの分布拡大が小雪地帯から多雪地へ拡大途上にあることを反映したものと考えられる。ニホンジカは縄文時代には全国的に分布 (Tsujino *et al.*, 2010) していたことから、現在の分布拡大は、過去に空きニッチとなった空間への再移入と捉えることができる。つまり、現在は非出現地域であっても、潜在的には分布できる可能性が高いため、事前に対応を検討することが望ましい。

4. おわりに

本研究では、全国レベルでの哺乳類の空間分布についてRFを用いて解析し、対象種の分布域の環境特性を評価した。結果として、ニホンジカとイノシシについては、分布拡大に伴う空間分布特性に変化がみられた。哺乳類の分布の歴史的な経緯を考慮すると、分布拡大要因はマルチスケールで把握する必要があり、農業農村地域における持続的な農業生産と生態系管理のためにはトップダウンとボトムアップの両アプローチによる柔軟な対応策が求められる。

参考文献 Breiman (2001): Random Forests. *Mach. Learn.*, 45, 5–32. Cutler *et al.* (2007): Random forests for classification in ecology, *Ecology*, 88(11), 2783–2792. Kaji *et al.* (2000): Spatial distribution of an expanding sika deer population on Hokkaido Island, Japan. *Wildlife Soc. B.*, 28(3), 699–707. Liaw and Wiener (2002): Classification and regression by random forest, *R News*, 2(3), 18–22. 農林水産省 (2016): 全国の野生鳥獣による農作物被害状況について. URL: <http://www.maff.go.jp/j/press/nousin/saigai/160122.html>. 奥村ら (2009): ニホンジカ (*Cervus nippon*) の分布拡大に影響を与える要因. *環境科学会誌*, 22(6), 379–390. R Core Team (2015): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <http://www.R-project.org/>. Saito *et al.* (2016): Range-expanding wildlife: modelling the distribution of large mammals in Japan, with management implications. *Int. J. Geogr. Inf. Sci.*, 30(1), 20–35. Tsujino *et al.* (2010): Distribution patterns of five mammals in the Jomon period, middle Edo period, and the present, in the Japanese Archipelago. *Mamm. Study* 35, 179–189.

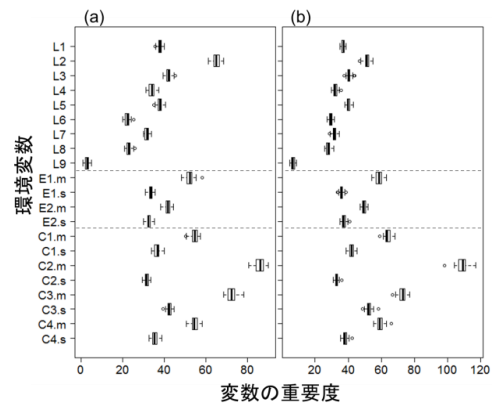


Fig. 1 Variable importance:
(a) 1978; (b) 2003

(L1: Paddy, L2: Other agricultural products, L3: Forest, L4: Abandoned land, L5: Buildings, L6: Major transport, L7: Other land use, L8: River & lakes, L9: Coastal area, E1: Mean elevation, E2: Mean slope, C1: Mean annual precipitation, C2: Mean annual temperature, C3: Maximum snow depth, C4: Mean solar radiation, m: mean, s: standard deviation)

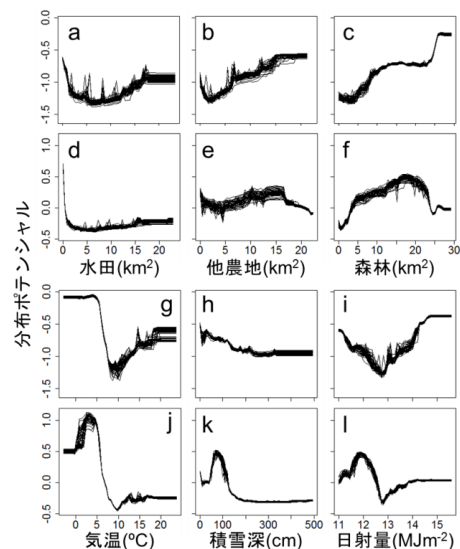


Fig. 2 Response curves:
(a-c, g-i) 1978; (d-f, j-l) 2003