

データ駆動解析手法を用いた湧水小河川におけるカワムツの生息環境評価

Habitat evaluation of the dark chub in a small spring-fed river using four different data-driven models

○地引 汰一¹・福田 信二²

○JIBIKI Taichi, FUKUDA Shinji

1. はじめに

2001 年には土地改良法で「環境との調和に配慮すること」が圃場整備事業実施の原則として位置づけられており、生態系に配慮した水辺空間の整備が求められている。自然との共生を可能にする環境整備の促進には、水辺の生態系保全機能の把握及び人為的環境変化が生態系に及ぼす影響の事前予測が重要である。水辺の有する生態系保全機能、すなわち生物の生息場としての質の評価には、対象種の生息環境条件を把握することが重要であり、それには生息場モデルが有効である。生息場モデルはそれぞれ計算方法や前提とする仮説が異なり対象とする種や環境により適切なモデルを選択する必要があるため、それぞれのモデルの特性について比較・検討しそれらの性質を整理する意義は大きい。そこで本研究では、数多く存在するデータ駆動解析手法の中から一般化線形モデル (GLM)、一般化加法モデル (GAM)、ランダムフォレスト (RF)、サポートベクターマシン (SVM) を用いて、湧水小河川におけるカワムツの生息環境を解析し、モデルごとの特性を比較する。

2. 方法

調査対象地とした矢川は、東京都国立市を流れる湧水起源の都市小河川である。この対象水域内に区間長 10 m の調査区を 15 地点設定し、魚類相調査と物理環境調査を 2015 年 6 月から 2020 年 3 月にかけて月 1 回の頻度で実施した。魚類相調査では、調査区の上流に定置網を設置し、タモ網および投網を用いて調査区内の魚類採捕を実施した。物理環境調査では、調査区内の物理環境として水深、流速、河床材料を計測した。

3. 解析方法

本研究の各種の解析に統計ソフト R (R Core Team, 2020) を使用した。カワムツの生息環境の把握に加えて、異なる手法 (GLM, GAM, RF, SVM) で構築した生息場モデルの特性を比較するために、各調査区のカワムツの在・不在データを応答変数として解析した。説明変数には、調査年 (Yr)、調査月 (MON)、平均水深 (Dm)、平均流速 (Vm)、流量 (Q)、植生被覆度 (VEGm)、中礫被覆度 (MGm)、小礫被覆度 (SGm)、砂泥被覆度 (SCm) を使用した。

10 分割交差確認法を行い、CCI (Correctly Classified Instances) による各モデルの精度評価及びハイパーパラメータの最適化を行った。CCI は次式により算出され、全データに対するモデルが在・不在を適切に予測できたデータの割合であり、値が大きいくほどモデルの精度が高いことを示す。

$$CCI = \frac{TP + TN}{TP + FN + TN + FP}$$

ここで、TP (True Positive) は在データをモデルが在と判別した数、FN (False Negative) は在データをモデルが不在と誤判別した数、TN (True Negative) は不在データをモデルが不在と判別した数、FP (False Positive) は不在データをモデルが在と誤判別した数である。

¹ 東京農工大学大学院農学府 Faculty of Agriculture, Tokyo University of Agriculture and Technology

² 東京農工大学大学院農学研究院 Institute of Agriculture, Tokyo University of Agriculture and Technology

キーワード: 生態系, 生物多様性, 水環境, 環境保全, 環境影響評価

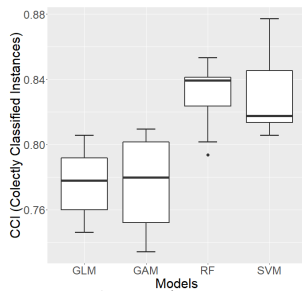


図1 各モデルのCCI

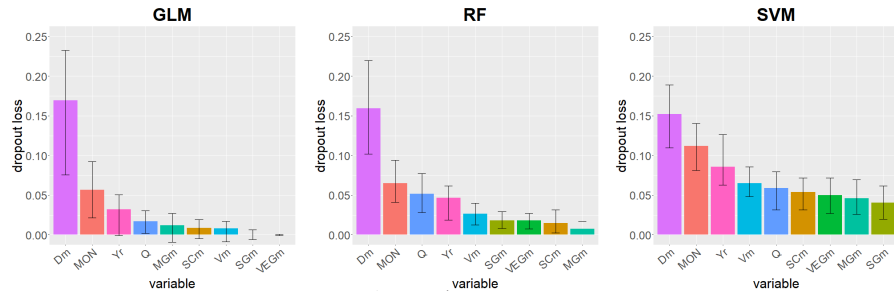


図2 各モデルの変数の重要

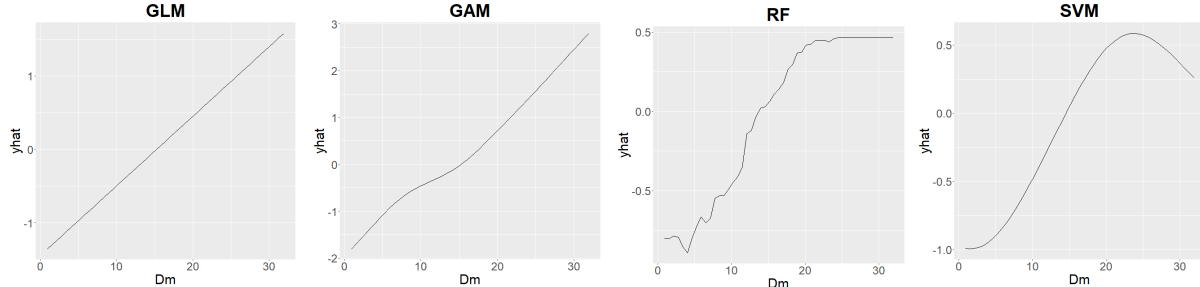


図3 各モデルの平均水深についての応答曲線

各モデルのCCIの差を二元配置の分散分析及びHolm法を用いて、有意水準 $p=0.05$ で検定した。次に、GLMとRF、SVMについては変数の重要度を算出した。また、全モデルについて対象種の生息環境特性の指標となる応答曲線を算出した。

4. 結果と考察

交差確認法を用いて算出したCCIの結果を図1に示す。モデルの精度を示すCCIの平均値は大きい順に $RF > SVM > GAM > GLM$ となった。また、統計検定では、GLMとGAM、RFとSVM以外の全ての組み合わせで有意な差があり、統計的手法と機械学習の再現性が異なることが確認された。変数の重要度の評価結果を図2に示す。いずれのモデルにおいても平均水深が最も重要な変数となった。上位に入った変数には離散変数の調査月と調査年、連続変数の平均流量や平均流速がある。最も重要度が高かった平均流速の応答曲線を図3に示す。いずれのモデルにおいても右上がりの傾向が見られるが、RFとSVMの応答曲線にピークが見られるのに対して、GLMとGAMは大きな水深においても単調増加の応答を示す。機械学習におけるピークの水深は20cm程度になっているが、これは中規模河川における先行研究の選好水深よりも数十cmほど小さい(鬼東ら, 2017)。その理由として、矢川の流路長や環境傾度が小さいことが挙げられる。

5. おわりに

本研究では、湧水起源の都市小河川におけるカワムツの生息環境及び解析手法間の特性の差異を明らかにした。水辺空間の整備における意思決定支援ツールとして生息場モデルを考える際に、モデルの精度だけでなくモデルの解釈可能性が重要となる。一般に、機械学習を用いた生息場モデルは統計的手法よりも精度が高いが解釈が難しいと認識されている。しかし、機械学習でも適切な手法を用いれば統計的手法と同様に解釈が可能であることが示された。今後は各個体の身体条件、種間関係等を考慮に入れてより精緻な生息場評価を行い、豊かな水辺空間の創設に貢献したい。

引用文献

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>

鬼東幸樹ら (2012) 板櫃川における魚類生息場の季節変化の調査. 土木学会論文集 B1 (水工学), 68(4), I_703-I_708.