

河川水中の栄養塩類濃度は土地利用などの流域特性から機械学習で予測できる

機械学習手法を用いて、熱帯島嶼の河川水に含まれる栄養塩類（窒素、リン、ケイ素）の濃度は、土地利用などの流域特性を説明変数として高い精度で予測できる。また、同手法により算出される各流域特性の重要度から、栄養塩類に対して特に寄与の大きい流域特性を同定できる。本成果は、陸域から流入する栄養塩類の量を適正に管理し、健全な沿岸生態系を保全するための施策立案に活用できる。

キーワード：沿岸生態系、環境負荷、熱帯島嶼、ランダムフォレスト、流域管理

背景・ねらい

熱帯島嶼の周辺には、サンゴ礁など生物多様性が高く、観光・漁業資源としても重要な沿岸生態系が発達している。一方、近年の気候変動に伴う海洋環境の変化（海水温の上昇、海洋酸性化など）に加えて、陸域での農業生産や土地開発などに起因する過剰な栄養塩（窒素、リン）の流入による、沿岸域の水質悪化や生態系への悪影響が懸念されている。さらに、主要な植物プランクトンである珪藻の生育に必要なケイ素に対して窒素やリンが過剰に供給されると、珪藻以外の植物プランクトンが増え、漁業被害を引き起こす赤潮などの発生に繋がる可能性も指摘されている。

そこで本研究では、熱帯島嶼の一つである石垣島（沖縄県）内の複数の河川において定期的に水質調査を実施し（図1）、河川水中の栄養塩類（窒素、リン、ケイ素）濃度を測定する。そしてこれらの栄養塩類濃度と流域の土地利用や地質などの特性との関係について機械学習手法を用いて解析することにより、それらの流域特性が各栄養塩類濃度に対してどのように影響を及ぼしているかを評価する。

成果の内容・特徴

1. 水質調査結果をもとに、機械学習手法の一つであるランダムフォレスト(RF)を用いて、流域特性（土地利用および表層地質の各タイプの面積割合、人口密度）を説明変数として各栄養塩類濃度を予測するモデルが得られる。
2. 水質調査から得られた実測値に対するモデル予測値の回帰直線の決定係数(R^2)は、学習データで 0.87 ~ 0.96、検証データで 0.69 ~ 0.88 と高く、実測値とのずれを表す PBIAS（実測値の平均値に対する実測値と予測値との差の平均値の割合）も同じく -0.33 ~ 0.36%、-1.30 ~ -0.06% と小さい。したがって、RF モデルは栄養塩類濃度を高い精度で予測できることが分かる（図2）。
3. RF モデルにおける各流域特性の重要度、ならびに栄養塩類濃度と各流域特性との相関係数から、栄養塩類に対して特に寄与の大きい流域特性が同定でき

る。窒素およびリンに対しては、特にサトウキビ畑や畜舎など農業に関連する土地利用の寄与が大きく、それらの面積割合が大きい河川ほど濃度が高いことが分かる（図3）。石垣島ではサトウキビ栽培および肉牛の飼育が盛んに行われていることから、これらの農業生産活動が河川の栄養塩濃度に強い影響を与えていると考えられる。

4. 一方、ケイ素に対しては、特に森林（広葉樹林）の寄与が大きく、その面積割合が大きい河川ほど濃度が高いことが分かる（図3）。

成果の活用面・留意点

1. 本研究で作成した RF モデルは、対象とする流域の土地利用や地質などのデータがあれば比較的簡便な操作で予測値を算出できることから、栄養塩負荷に関して監視・対策が必要な流域の選定など、沿岸域の水質・生態系保全のための施策立案に活用できる。
2. 本 RF モデルは、今後の土地利用管理の変化に伴う栄養塩類濃度の変化の予測にも応用できる可能性があるが、その結果の妥当性については、流域における水や物質の移動、反応に係る各プロセスを数式化した分布型流出モデルによる計算結果等と比較、検証する必要がある。

その他

予算区分：交付金プロ [A5 熱帯島嶼環境保全]

研究期間：2020~2022 年度

研究担当者：菊地哲郎（生産環境・畜産領域）、安西俊彦（熱帯・島嶼研究拠点）

発表論文等：Kikuchi et al. (2023) *Environmental Pollution* 316: 120599. <https://doi.org/10.1016/j.envpol.2022.120599>.

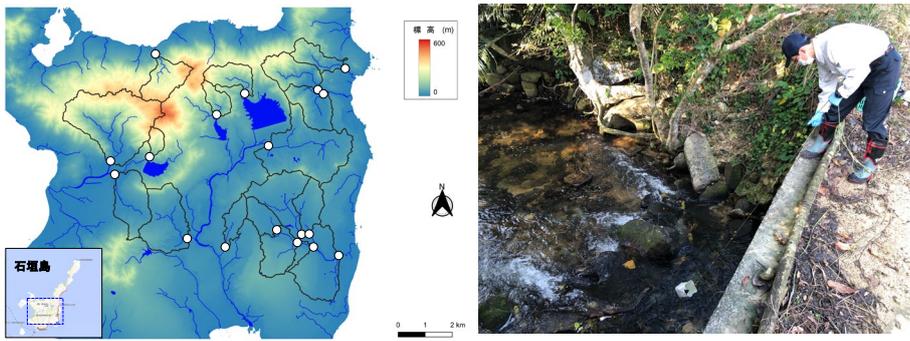


図1 調査対象地域の地図
（左）おおよび水質調査の様子（右）

地図上の○は調査地点を、黒の実線で囲まれた領域は各調査地点における河川流域を示す。調査は、約2ヶ月に1回の頻度で1年間実施。

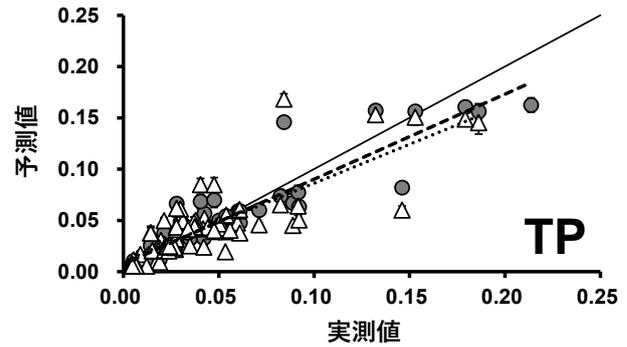
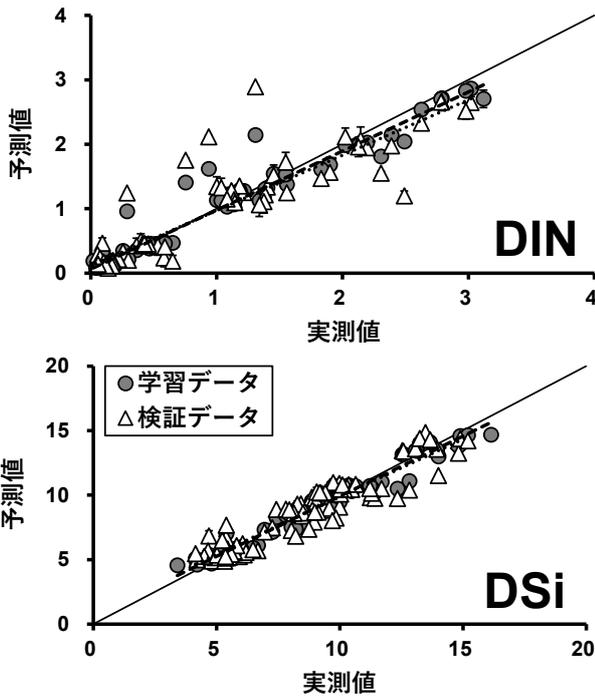


図2 各栄養塩類濃度（単位はいずれも mg/L）の実測値と RF モデルによる予測値との関係

DIN：溶存無機態窒素、TP：全リン、DSi：溶存態ケイ素。全データセットをランダムに5等分し、そのうちの4つを統合したものを学習データ、残りの1つを検証データとして、学習データ（計5通りの組み合わせのそれぞれ）についてのモデルを作成する。この操作を3回繰り返し、得られた計15個のモデルによる計算値の平均値を予測値とする。実線は実測値 = 予測値、破線は学習データについての回帰直線、点線は検証データについての回帰直線。

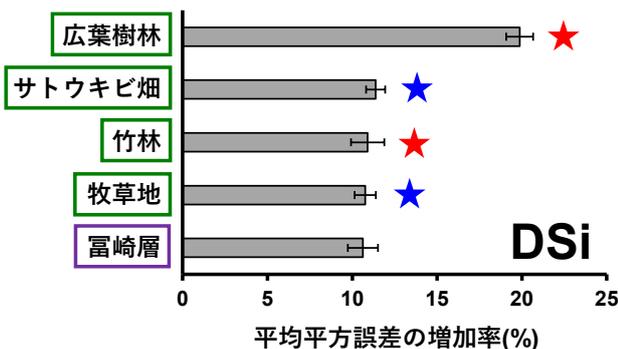
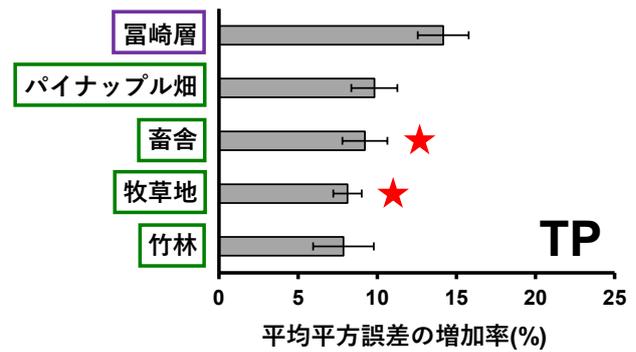
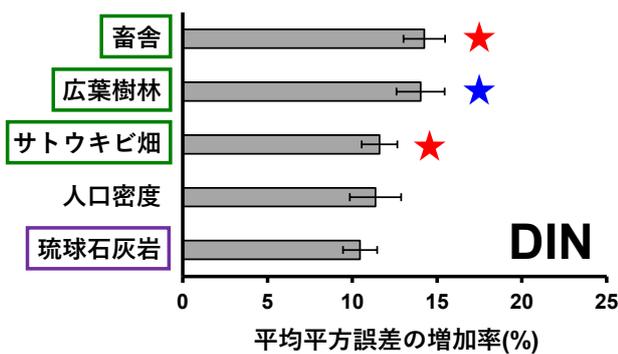


図3 各栄養塩類濃度に対する各流域特性の重要度（上位5特性のみ表示）

当該特性の値をランダムに入れ替えた場合の予測値の平均平方誤差の増加率が大きい特性ほど、栄養塩濃度に対する重要度が大きいと判断される。□は土地利用、□は表層地質。★は当該水質と強い正の相関（スピアマンの順位相関係数 r_s が0.7以上）がある特性、★は当該水質と強い負の相関（ r_s が-0.7以下）がある特性をそれぞれ表す。

図は Kikuchi et al. (2023)より引用（転載・改変許諾済）