

ランダムフォレスト法を用いた釧路湿原の植生生育条件の分析

Research on Vegetation Growth Condition of Kushiro Mire Using the Random Forest Method

室蘭工業大学 ○学生員 佐久間 寛樹 (Kanju Sakuma)
 室蘭工業大学 正 員 中津川 誠 (Makoto Nakatsugawa)
 室蘭工業大学 非会員 小林 洋介 (Yosuke Kobayashi)
 室蘭工業大学 学生員 坂本 久宣 (Hisanobu Sakamoto)

1. はじめに

北海道の東部に位置する釧路湿原では、近年湿原面積が減少し、ハンノキ林の急激な拡大といった湿原本来の植生環境の変化が問題となっている。図-1は釧路湿原におけるハンノキ林の変遷を示した繁茂図であるが、1947年～2004年までの58年間で湿原面積は約30%減少し、2004年時点でハンノキ林の分布域は釧路湿原域の約46%に達している。特に1996年～2004年の8年間と1947年～1977年の30年間における湿原のハンノキ林が占める面積の増加率、すなわち湿原面積の減少率といった変化度合がほぼ等しく、このことから近年の植生変化が加速していることがわかる¹⁾。このような著しい植生の変化は、野生動物の生息環境の悪化、湿原景観の改変、湿原のもつ機能の低下など様々な問題を連鎖的に引き起こす要因となる。

ハンノキ林の拡大要因については、釧路湿原内の湿原植生と地下水変動パターンとの関係性について調べた辻井ら²⁾、また数量化Ⅱ類を用いてヨシ群落とハンノキ林の分類を行った羽石ら³⁾による報告がある。また、夏季における地下水位変動が植生の生育に影響を及ぼすことを指摘した坂本ら⁴⁾の研究があるが、湿原の植生生育条件の一般性を見出すところには至っていない。

そこで本研究では、機械学習法の一つである Random Forest 法(以下、RF 法)によって現況の湿原のヨシ群落、ハンノキ林の植生生育条件の分類モデルを構築し、未知地点の植生分類を行うことで、釧路湿原におけるヨシ群落、ハンノキ林といった植生生育条件を明確にすることを目指す。

2. 研究方法

2.1 RF 法における分類モデルの構築について⁵⁾

RF 法とは、多数の決定木を組み合わせることで汎化能力を高める方法である。図-2にRF法におけるクラス分類の概略図を示し、以下の方法で計算を行う。

1) 分類モデルの構築のための学習データとそのモデル

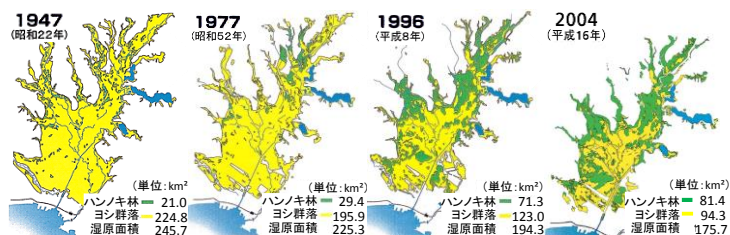


図-1 釧路湿原の植生変化
(釧路湿原自然再生プロジェクトHP¹⁾より)

の精度を確認するためのテストデータに分ける。

- 全説明変数 n 個から重複を許容しランダムに 1 つの説明変数を抽出する作業を m 回行う。 n 及び m の値が十分大きい場合 n の約 64%($=1-(1-1/n)^m$)の説明変数が抽出される。また、本研究において、抽出回数 m は n と同数($m=n$)とした。
 - 抽出された説明変数を用いて 1 本の決定木を作成する。決定木は(1)式に示す Gini 不純度 I が最小になる変数によってデータを 2 分し、同様の操作を決定木の構成要素の他の説明変数についても繰り返す。分割されたグループ内の全てのデータがヨシ群落かハンノキ林のどちらかの植生に分類されるまで行う。分割の操作を行う分岐点はノードと呼ばれる。
- $$I = 1 - \sum_{i=1}^c \left(\frac{N_i}{N_{\text{node}}} \right)^2 \quad (1)$$
- ここで、 c は分類するクラス数(本研究においてはヨシ群落とハンノキ林の2クラス)、 N_{node} は各ノード内に含まれる全地点数、 N_i は各ノード内に含まれるクラス i に属する地点数である。
- このようにして作成された決定木にテストデータに対応させ、どちらかの植生に分類された末端のノードまで辿らせる。辿り着いた末端のノードがヨシ群落かハンノキ林かの分類結果をもって決定木の分類結果とする。
 - この作業を繰り返し、本研究においては決定木を 1,500 本作成した。その後、ある箇所の条件を作用させた場合に各決定木から推定された結果の多数決を

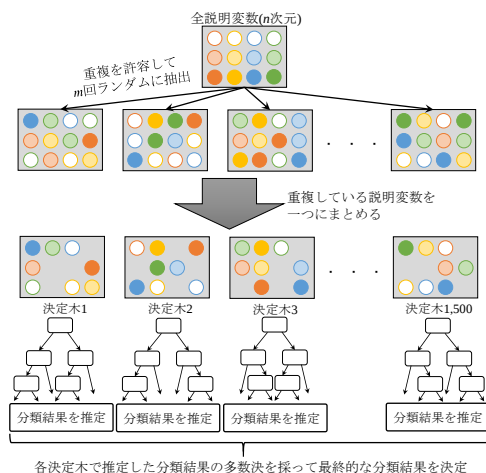


図-2 RF 法におけるクラス分類の概略図

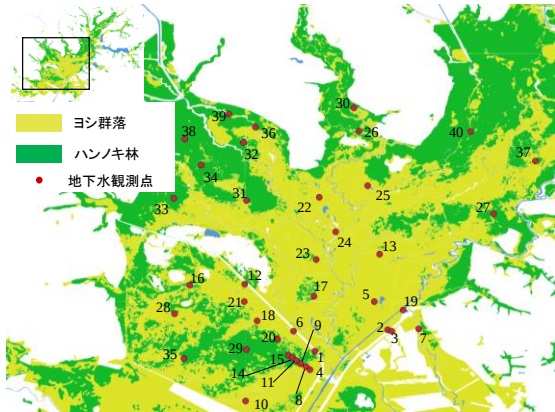
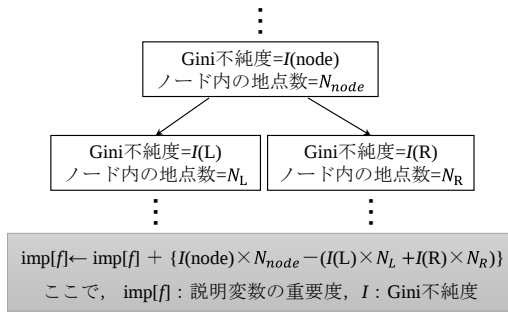


表-1 観測地点振り分け表

ヨシ				ハンノキ			
観測地名	番号	観測地名	番号	観測地名	番号	観測地名	番号
湿原No.5	37	湿原No.21	28	湿原No.40	16	T-4	1
湿原No.7	13	湿原No.22	35	昭和No.7	10	T-6	12
湿原No.8	5	湿原No.28	18	昭和No.9	21	鶴居No.2	39
湿原No.14	22	湿原No.29	20	昭和-A	4	鶴居No.8	40
湿原No.15	25	湿原No.37	26	昭和-D	8	遊水地1号	6
湿原No.19	23	湿原No.38	24	T-1	2	岩保木1号	19
湿原No.20	17	湿原No.39	36	T-2	3	岩保木2号	7
						昭和No.8	29
						昭和-O	14
						昭和-R	15
						昭和-G	9
						昭和-K	11
						鶴居No.5	30
						湿原No.26	34

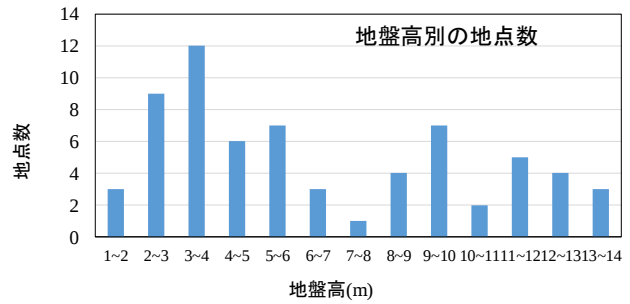
採って、最終的な分類結果を決定する。

また、図-3に示すよう変数重要度（以下、 $\text{imp}[f]$ ）を決定木の分割ごとに計算を行い、全ての $\text{imp}[f]$ の総和で各説明変数の $\text{imp}[f]$ を除して正規化したものが寄与度となる。よって寄与度の値は0～1の範囲となる。

2. 2 構築したモデルにおける交差検証⁶⁾

RF法を含む機械学習手法は、前述のように統計モデルの作成に利用する学習データとモデルの評価に利用するテストデータを用意する必要がある。十分なデータ数がある場合、学習データで分類モデルを構築し、テストデータで分類モデルのテストを行えばよい。しかし、本研究で収集したデータ数は後述する40地点であり、その方法では十分な分類が行えない。そのため、機械学習分野ではこのような少数のデータのときに交差検証によってモデルを評価する。本研究では、2.1節で述べた分類モデルを用いて、以下の手順で交差検証を行った。

- 1) 全40地点のデータからテストデータとして1地点のデータを取り除き、残りの39地点を学習データとする。
- 2) 残りの39地点からなる学習データから2.1節で述べたRF法により分類モデルを作成する
- 3) 作成した分類モデルで1)で取り除いた1地点のデータを分類する。
- 4) 1), 2)の作業を全40地点について繰り返す。



- 5) 全40モデルによる植生分類及び寄与度の平均を本研究による分類結果とする。

2. 3 RF法に用いるデータと解析範囲

北海道開発局釧路開発建設部の湿原内118地点において測定されている地下水位観測点のうち、以下の条件を満たす40地点について分析を行った。観測点の位置関係においては図-4に示す。

- 1) 植生がヨシ群落、またはハンノキ林であるもの。
- 2) 2005～2007年において一ヶ月にわたる連続した欠測がないもの。
- 3) 地盤高が7m未満であるもの。

地盤高については図-5に条件1), 2)を満たす67地点の頻度分布を示すが、地盤高7m未満の群と地盤高7m以上の群がみられる。2つの群を含めたモデルでの評価では地盤高へ大きく依存した結果が算出され、植生生育条件を正しく評価できないことが懸念されることから、3つの条件を設定した。また、地盤高7m未満の群は今後植生変化が懸念される湿原中心部付近に分布している。観測地点は植生別にヨシ群落28地点、ハンノキ林が12地点となった。図-4の観測点の位置は表-1の通し番号に対応しており、通し番号は地盤高の低い順番に振られている。図-4は環境省生物多様センター⁷⁾の第7回植生調査によって作成された2007年次の1/25,000植生図GISデータを使用し、本研究のために作成・加工したものであり、これを用いて植生の分類を行った。また、RF法を用いた説明変数の項目は羽石ら³⁾と坂本ら⁴⁾の先行研究を活かし、表-2に示す5つを採用した。また、各項目においてその季節別の寄与度から植生に対する季節別の影響度を比較するために1～3月を第1期、4～6月を第2期、7～9月を第3期、10～12月を第4期と分けてそれぞれを説明変数とし、全17次元で分析を行なった。表-2に本研究で用いた説明変数の概要について示す。

- 1) 平均相対水位

観測水位の期間平均値から地盤高を差し引いたものであり、相対水位が正であれば水面が地表面より高く、負であれば水面が地下に位置することを意味する。

- 2) 水位分散

任意の1点における地下水位観測点の期間内日平均観測水位を標本としたときの標本分散とした。

- 3) 地表面冠水頻度

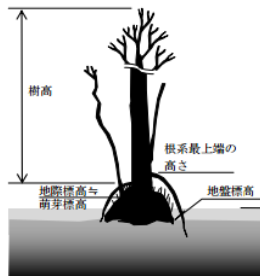
期間内における観測日数に対する日平均相対水位が0m以上となる日数の比である。

- 4) 30cm冠水頻度

期間内における観測日数に対する日平均相対水位が30cmを超えている日数の比である。矢野ら⁸⁾によると、

表-2 RF 法に用いた説明変数

項目	説明変数(2005~2007年)			
平均相対水位(m)				
水位分散	第1期	第2期	第3期	第4期
地表面冠水頻度	1~3月	4~6月	7~9月	10~12月
30cm冠水頻度				
地盤高(m)	通年一定			

図-6 ハンノキ計測部位の模式図(矢野ら⁸⁾より)

予測結果			
	ヨシ群落	ハンノキ林	各植生の再現率
実際の植生			
ヨシ群落	23	5	82.1%
ハンノキ林	7	5	41.7%

図-7 RF 法における植生分類の結果

図-6 に示すようにハンノキの根茎最上端の高さが地盤高から平均 24cm 高い位置にあることが指摘されており、また根茎最上端の冠水がハンノキの樹高成長を妨げるとの見解を示しているため、このような指標を考えた。

5) 地盤高

国土交通省北海道開発局釧路開発建設部から提供を受けた LP で計測された地盤標高(TP)である。

3. 結果と考察

3.1 RF 法による植生生育条件の結果

図-7 に各観測地点における 2005 年~2007 年の説明変数を学習させ、植生の分類を行った結果を示す。ここで図-7 における行列は、行側に実際の植生分類、列側に分類された植生分類を示しており、正答数は同じ植生が行と列に交差する部分を表し、一方、誤答数は植生が誤分類された部分を表している。評価の指標には正答率と再現率を採用した。(3)式に正答率を、式(4)にあるクラス*i*における再現率を示す。

$$\text{正答率} = \frac{\sum R_i}{N} \quad (3)$$

$$\text{クラス}i\text{における再現率} = \frac{R_i}{R_i + R_j} \quad (4)$$

ここで、 R_i はクラス*i*(*i*はヨシ群落、ハンノキ林)における実際の植生と分類結果が一致した箇所数であり、 N は全箇所数、 R_j はクラス*i*における実際の植生と分類結果が不一致であった箇所数である。今回のモデルでは正答率が 70.0%であるため、今回採用した説明変数では説明しきれない事例が 30%存在していることとなる。また、ヨシ群落の再現率は 82.1%、ハンノキ林の再現率は 41.7%であるため、実際の植生のうちヨシ群落のほうが良く再現できている結果となった。これはデータの学習の段階で植生がヨシ

全17次元の説明変数の寄与度

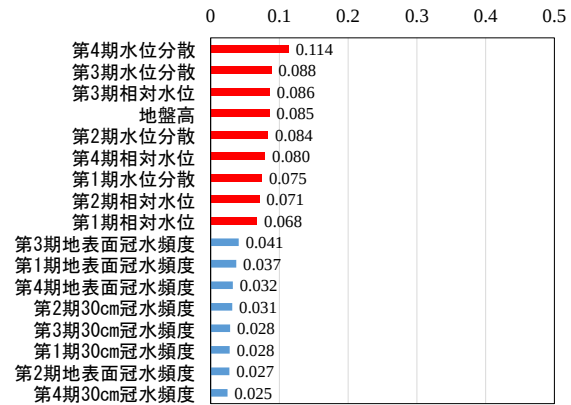


図-8 RF 法で算出した寄与度

群落である事例の方が多く学習されることが関係していると考えられる。

3.2 RF 法による植生生育条件の寄与度の結果と考察

図-8 に RF 法における植生分類の結果の寄与度を示す。第4期と第3期の水位分散の寄与度が高く、全ての地点が7m 未満と限定しているにも拘らず地盤高の寄与度が高い結果となった。また、坂本ら⁹⁾によると夏季における地下水位の変動がハンノキの生長に影響を及ぼすと指摘しており、RF 法における寄与度においても夏季である第3期の相対水位が上位に位置している。一方、30cm 冠水頻度の寄与度が下位となる結果を示した。これは、30cm 以上の冠水が起こる地点が少なく、あるいは解析対象期間中に冠水が少なかったことも考えられ、大きな寄与度とならなかった可能性が考えられる。

寄与度が上位に位置している水位分散と地盤高との関係を図-9、図-10 に示す。ヨシ群落が繁茂している観測点には水位分散が 0.05m²を超える箇所が地盤高 2~4m の箇所に散見されているが、ハンノキ林が繁茂している観測点には水位分散が 0.025m²を超える箇所はみられない。このことから水位分散の値が大きく冠水が起こりやすい地点ではハンノキ林が生育できないことが示唆される。また、地盤高が低い地点において冠水が起こりやすいことが示され、地盤高の寄与度の高さもこのことに関係があると考えられる。

図-11 に、RF 法による植生分類の分類結果を正誤別に地点表示する。図-11 は表-1 の通し番号に対応している。また、図-11 は図-4 の基図を用いている。それぞれの植生の正答地点において2地点ずつを抽出し、期間内における相対水位の出現頻度分布を比較したものを図-12 に示す。図-12 よりヨシ群落の繁茂する観測点においては相対水位 0.4m 付近の値が最も多く出現しており、ハンノキ林の繁茂する観測点においては相対水位 0m 付近の値が最も多く出現していた。これらのことからハンノキ林が生育できる水位の限界点が 0m~0.4m に存在することが示唆され、これは根茎最上端の冠水がハンノキの生育を阻害するという矢野ら⁹⁾の見解と一致する。ただし、その一般性を確立するには、より多くのデータを分析した上で検証することが必要である。

また、図-11 をみると実際の植生がヨシ群落に対しハンノキ林と判定された地点は通し番号で示すと、5, 20, 22, 35, 37 である。このうち 20, 35, 37 の観測点はハン

ノキ林との境界部に位置している。同様に、実際の植生がハンノキ林に対しヨシ群落と判定された地点は通し番号 9, 11, 27, 29, 30, 31, 32 である。このうち 29 以外の地点はヨシ群落との境界部に存在している。これらのことから、ヨシ群落、ハンノキ林の説明変数における固有な特徴が表れにくく、実際の植生とは異なる分類結果が算出されたことが考えられる。さらに、実際の植生がヨシ群落であるにも関わらず、ハンノキ林と判定された箇所においては、説明変数の特徴がハンノキ林と明瞭な区別ができないことから、潜在的にはハンノキ林の生育条件に近いことも考えられる。よって、今後ハンノキ林への変遷も懸念される。いずれにしても数値だけでなく、踏査もしくはドローン等による現地の確認が不可欠といえる。

4. まとめ

本研究で得られた結果を下記に記す。

- 1) 統計的手法である RF 法による植生生育条件の分類法を提案した。
- 2) RF 法における未知データの植生分類において 70% の正答率が得られた。また、RF 法で算出された寄与度より、地盤高が低く、水位分散が卓越した地点において、ヨシ群落が繁茂するという傾向がみられた。
- 3) 地盤高より 40cm 高い地点における水位の出現頻度が高い地点ではヨシ群落の生育が分類された。
- 4) ヨシ群落とハンノキ林の境界部に位置する観測点において、各植生の固有な特徴をみるのが難しく、十分な分類を行えなかった。

謝辞: 本研究を遂行するにあたり、国土交通省北海道開発局釧路開発建設部には釧路湿原関連のデータを提供して頂いた。また、室蘭工業大学情報電子工学系学科データサ

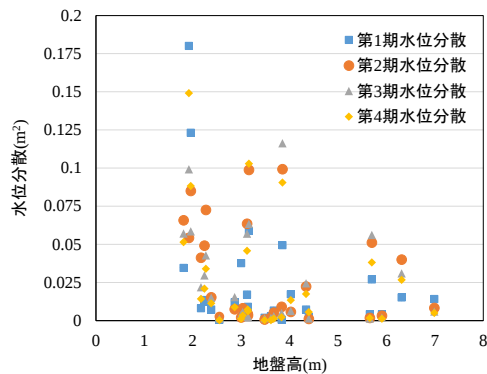


図-9 ヨシ群落における地盤高と水位分散の関係

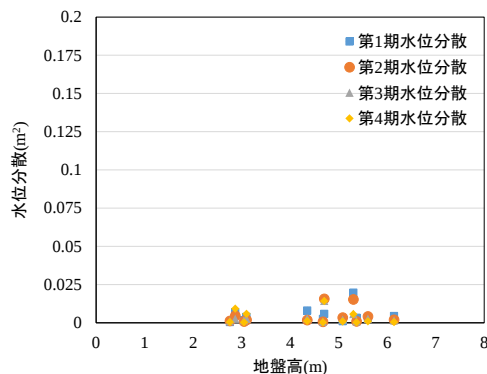


図-10 ハンノキ林における地盤高と水位分散の関係

イエンス研究室の山田恒輝氏には RF 法における助言を受けた。ここに記して謝意を表す。

参考文献

- 1) 環境省湿原自然再生プロジェクト湿原データセンターHP: <http://kushiro.env.gr.jp/> (2017/10/5 閲覧)
- 2) 辻井達一, 梅田安治, 櫻田純司: 泥炭地の植生と地下水位, 釧路湿原自然生態基礎調査, 北海道大学農学部, pp.36-105, 1974.
- 3) 羽石嵩 (2010), 釧路湿原のハンノキ林拡大に及ぼす地形と地下水位の影響に関する研究, 室蘭工業大学建築社会基盤系専攻水環境研究室, 修士学位論文.
- 4) 坂本久宣, 谷口陽子, 中津川誠: 釧路湿原における近年の植生状況と地下水位の関係についての考察, 土木学会北海道支部, 第 72 号, B-24, 2015
- 5) 速水悟: 事例+演習で学ぶ機械学習 ビジネスを支えるデータ活用のしくみ, 第 5 章, 森北出版株式会社, 2016.
- 6) C.M.ビショップ: パターン認識と機械学習 上 ベイズ理論による統計的予測, 丸善出版, 2014
- 7) 環境省生物多様センターHP: <http://www.biodic.go.jp/> (2017/11/7 閲覧)
- 8) 矢野雅昭, 水垣滋, 林田寿文, 村上泰啓: 釧路湿原におけるハンノキの形態と冠水環境への適応について, 湿地研究, Vol.1, pp.43-53, 2010.

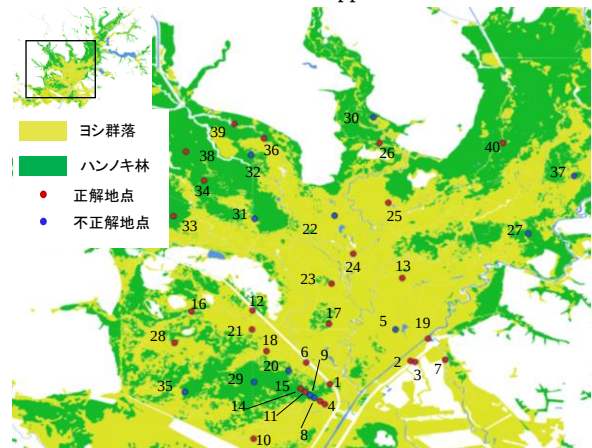


図-11 正誤別観測点図(環境省生物多様センター⁷⁾より)

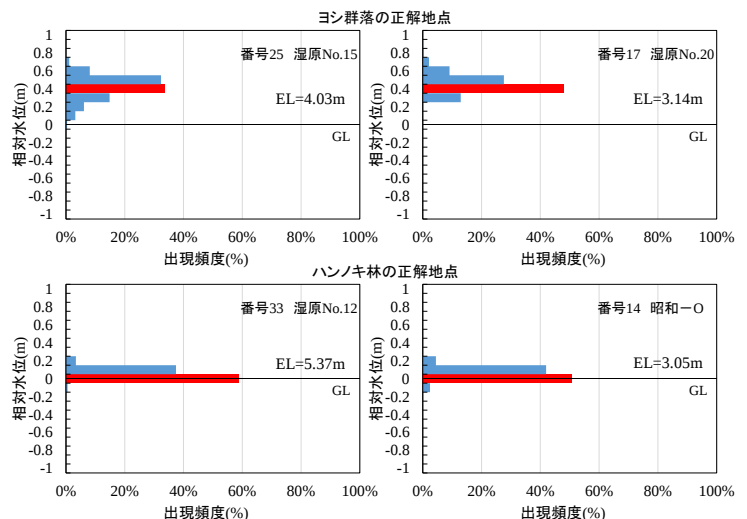


図-12 正解地点における相対水位の出現頻度分布