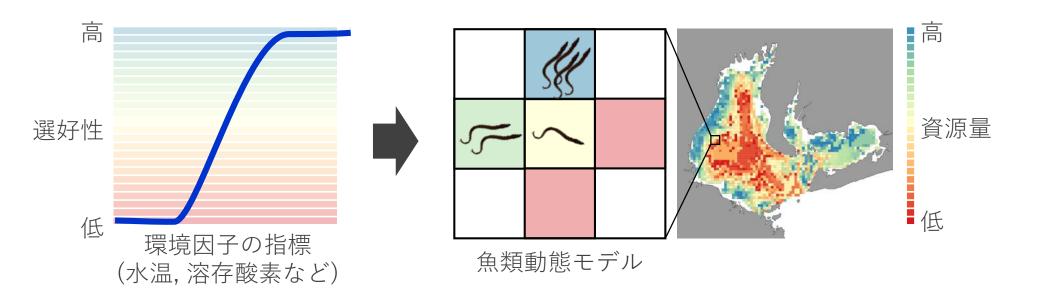
小型底びき網漁船における 操業日誌記録と水質ロガー記録との 突合手法について

鈴木 翔太・多部田 茂・水野 勝紀 (東京大学大学院)

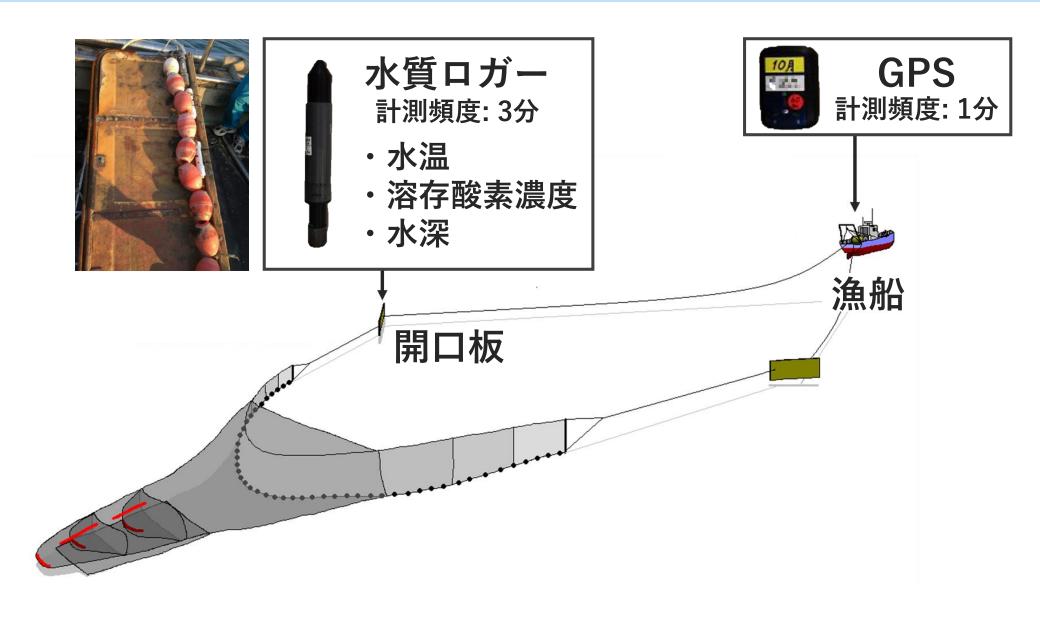
魚類資源の時空間的変動のシミュレーション



- ・環境因子への魚類の応答の知見が乏しく、精度向上の制約に
- → 伊勢湾の小型底びき網漁船に<u>水質ロガー+GPS</u>を着装 同時に<u>操業日誌に漁獲量を記録</u>

底層環境因子 と 漁獲量 の詳細な関係性を調べる!

水質ロガー+GPSの記録



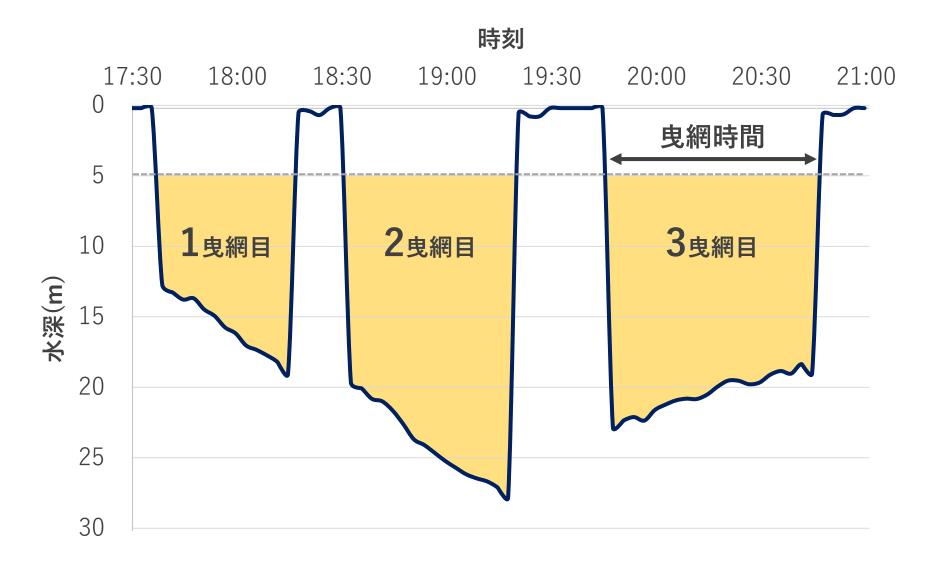
操業日誌による漁獲量の記録

	操	業状	入網情報					
回数	漁場 番号	曳網方向	曳網速度	時間	時間		シャコ	
1					漁獲	kg	kg	
Ι Τ		度	ノット	分間	放流	kg	kg	
2					漁獲	kg	kg	
		度	ノット	分間	放流	kg	kg	

- ・曳網ごとに魚種別漁獲量を漁業者が手書きで記録
- → 水質口ガーは 3分間隔の記録のため 記録を1曳網ごとに切り分ける

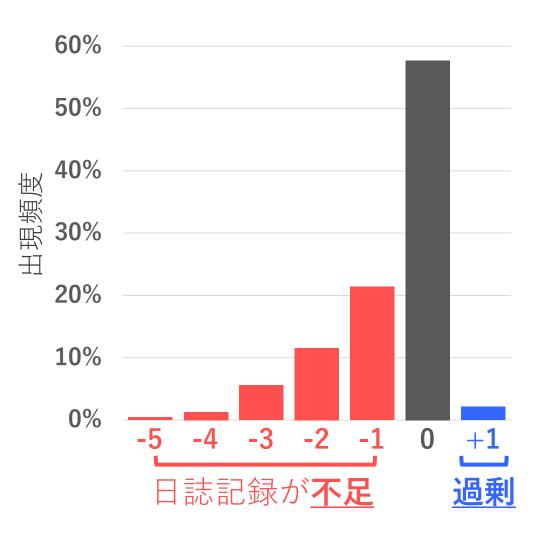
水質ロガー+GPSの記録

・<u>水深5m以深の連続記録</u>を 1曳網として集約



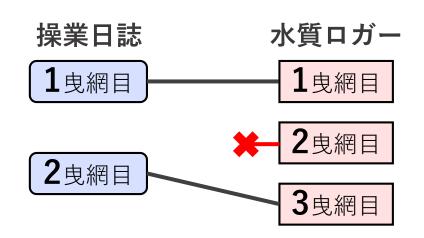
水質ロガーと操業日誌で曳網回数が一致しない!

操業日誌 水質ロガー 上の曳網回数を引いた値



曳網回数不合致の理由

- ・水質ロガーへの曳網以外の情報の記録 (網洗い、曳網の中断など)
- ・漁業者の記録忘れ



対応関係が分からなければ データが利用できない!

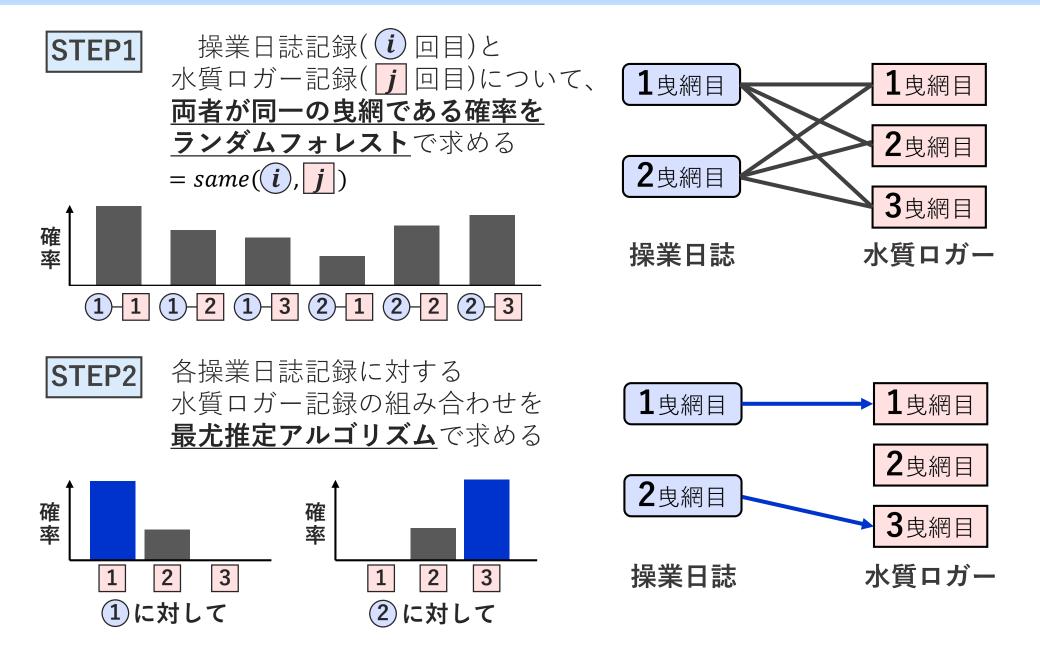
データの冗長性から同一曳網か判定

水質口ガー(+GPS)と 操業日誌 で 共通に記録している項目

- ・経度
- ・緯度
- ・曳網時間
- ・曳網方向
- ・曳網速度

上記項目の 水質ロガー と 操業日誌 の**差分を変数として、** 水質ロガー上のある曳網と、操業日誌上のある曳網とが **同一の曳網か否かを判定**するアルゴリズムを構築

水質ロガーと操業日誌の各記録の突合方法



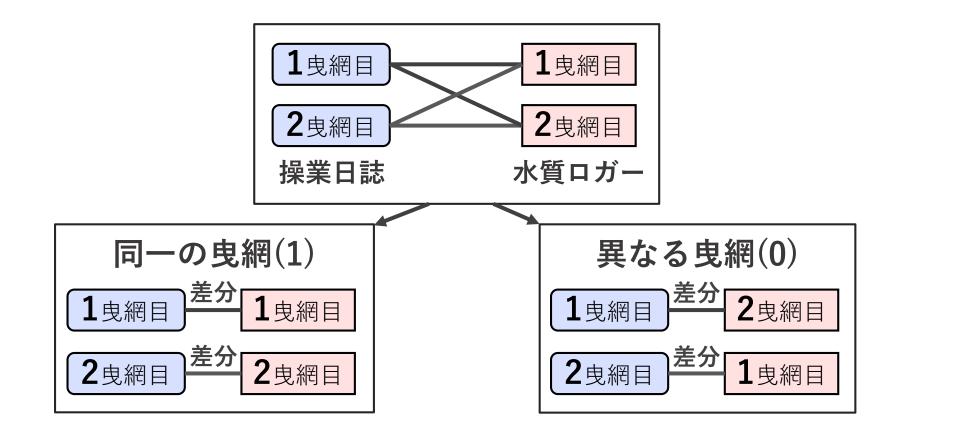
STEP1-1 教師データの作成

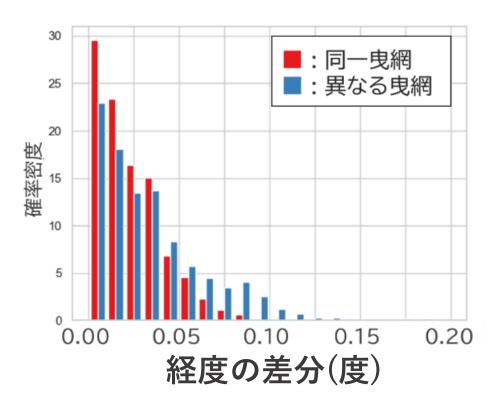
説明変数 経度, 緯度, 曳網時間, 曳網方向, 曳網速度の差分

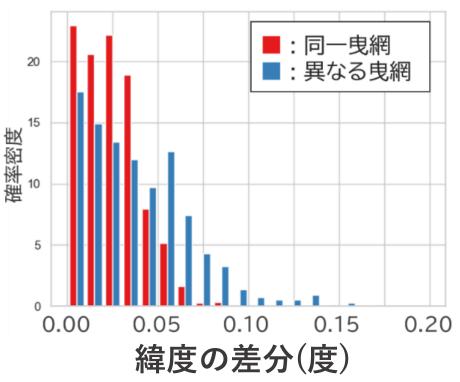
目的変数 同一の曳網(1) か、**異なる曳網**(0)か

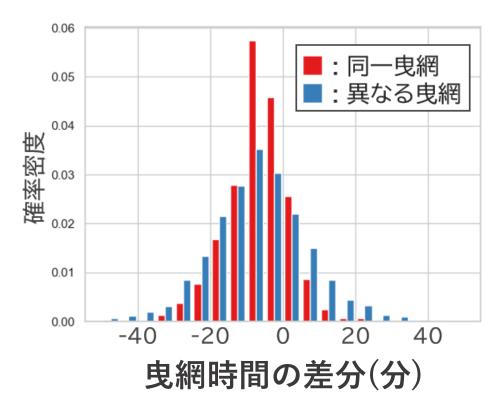
操業日誌上と 水質口ガー上 の操業回数が等しいデータを総当たり

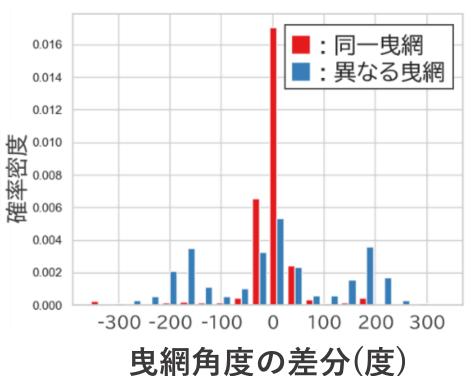
→「同一の曳網」と「異なる曳網」に分割

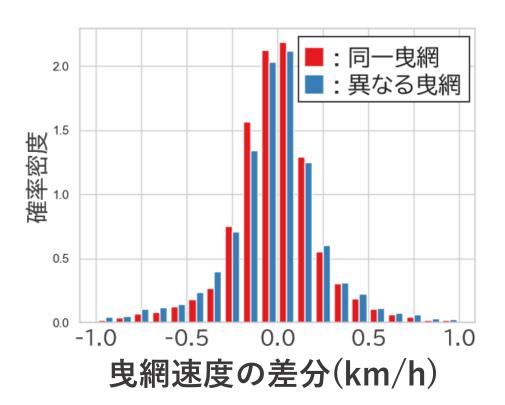








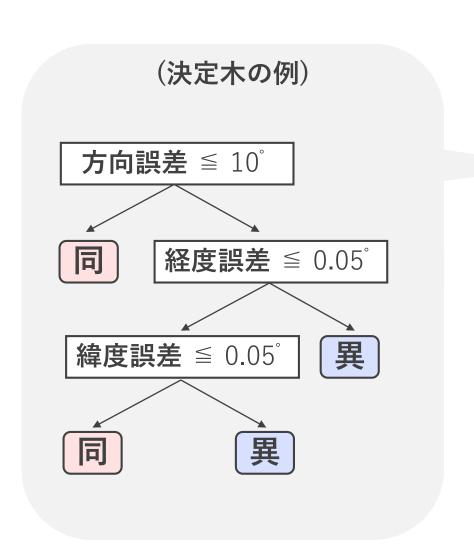




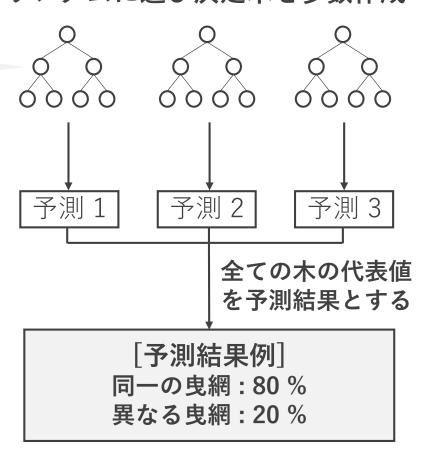
12

STEP1-2 ランダムフォレストモデルの構築

ランダムフォレスト … 決定木を用いた機械学習の手法の1つ

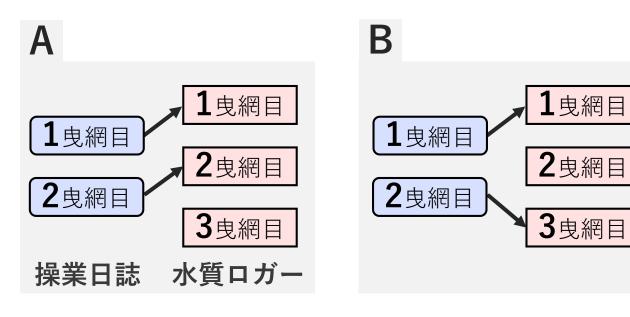


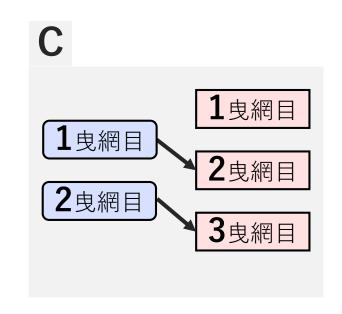
特徴量とサンプルデータを ランダムに選び決定木を多数作成



STEP2 組み合わせの最尤推定

・パターンを全列挙



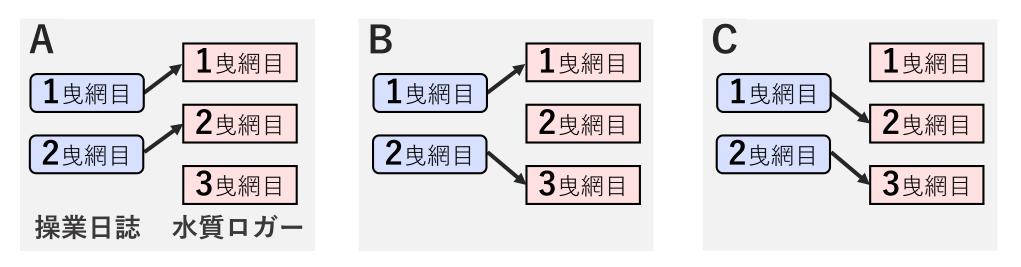


・STEP1で求めた確率が互いに独立として、 各パターンの確率を計算

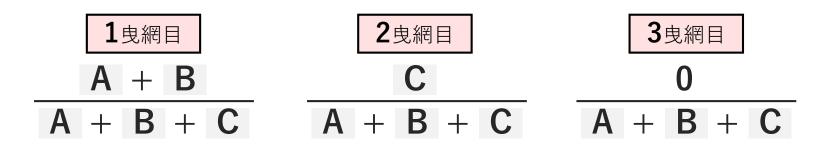
$$same(1, 1)$$
 $\times same(2, 2)$

$$\times same(2, 3)$$

 \times same((2), (3))



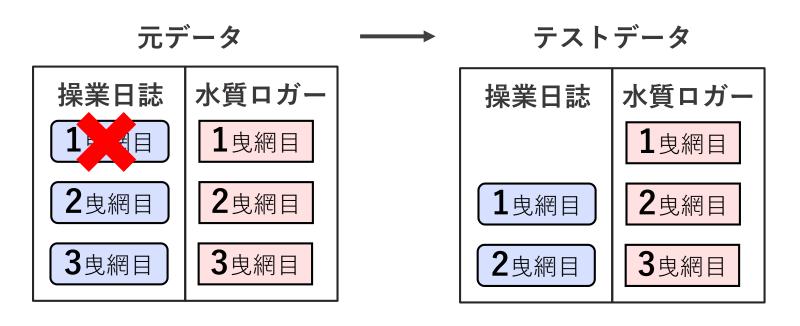
操業日誌記録 に対して 各ロガー記録 が対応する確率を求める例: **1**曳網目 に対して…



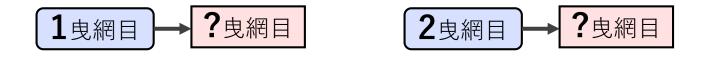
確率の最大値が 閾値 λ 以上 … 最大値をとる組み合わせで突合処理 閾値 λ 未満 … 判別不能 (突合処理を行わない)

精度検証用テストデータの作成

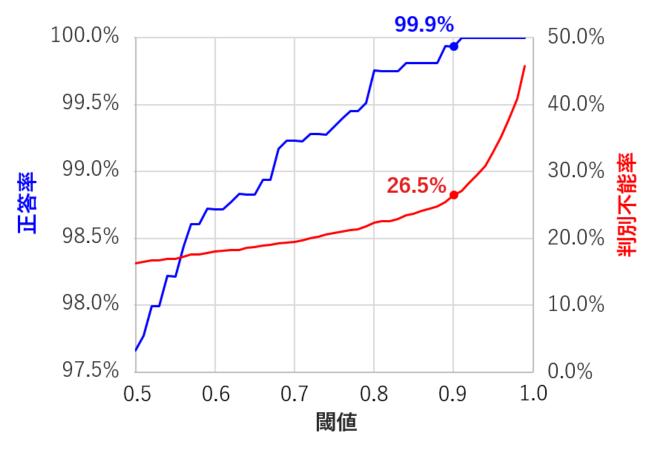
操業日誌上 と 水質口ガー上 の操業回数が等しいデータを用いる **操業日誌記録のうち任意の1個を欠落させ、テストデータを生成**



各操業日誌記録 に対応する 水質ロガー記録 を予想させ、 **正答率** および **判別不能率** を算出



精度検証結果



- ・閾値λを上げると正答率が上昇する一方、判別不能率も上昇
- ・ $\lambda = 0.9$ で正答率がほぼ 100% $\lambda > 0.9$ で判別不能率が急激に上昇
- →実際のデータに適用する際には $\lambda = 0.9$ とするのが妥当

実際の操業日誌欠損データへの適用

- ・水質ロガー上の曳網回数に対し 操業日誌上の曳網回数が 1回足りないデータを使用
- ・183 / 241曳網で記録の突合に成功 (判別不能率: 24.1%)
- ・28操業日中 15操業日で 最後の曳網が欠損したと推定
- →帰港前に「網洗い」をしている?

曳網回数/日 欠損回数	4	5	6	7	8	9	10
101	0	0	0	0	0	0	0
2回目	0	0	0	0	0	0	1
3回目	1	0	0	0	0	3	2
4回目	0	0	0	1	1	0	1
5回目		1	0	0	0	0	0
6回目			1	1	1	1	0
7回目				0	0	0	0
8回目					4	0	0
9回目						2	1
10回目							6

欠損したと推定される曳網は何回目か? (数字は出現回数)

- ・<u>ランダムフォレスト</u> と <u>最尤推定アルゴリズム</u>を用いて **欠損を含む操業日誌**を**水質ロガーと突合**する手法を開発
- ・閾値の適切な設定により、非常に高い正答率での突合が可能

課題等

- ・操業日誌記録は、数値の丸め方等の個人の癖を反映
- →本研究での訓練済モデルが他漁業者の操業日誌でも同様の突合性能を持つとは限らない
- ・1操業日に複数回の欠損が起きた場合の精度の検証
- ・漁獲量を説明変数として利用する (漁獲量が多いと選別に時間を要し記録漏れが増加する可能性?)