

第1章 序論

近年、地球温暖化による気候変動により、これまで見られなかった漁場の変化が起こっている。そのため、中期的な水産資源変動推定だけではなく、短期的な漁場予測が必要とされる¹⁾。短期的な漁場予測を行う際には、水産生物の資源分布予測モデルが作成される。そのようなモデルの一つとして、好的生息域 (HSI : Habitat Suitability Index) モデルが知られている²⁾。HSI モデルではまず、対象魚種の漁獲実績 (Catch Per Unit Effort) と各海域での海洋環境変数の関係を数式化した SI (Suitability Index) モデルを作成する。この SI モデルによって海洋環境変数ごとに、対象魚種にとっての生息しやすさが数値化される。海洋環境変数の数だけ作成された SI モデルを一つの式に統合することで、各海域がどの程度生息しやすいかを定式化した HSI モデルが構築できる。従来の研究では、深さの自由度を持つ各海洋環境変数と漁獲実績を散布図で表示し、各変数について一つ選択した深さの変数を平滑化スプラインで多項式近似することで各 SI モデルを構築する。そしてそれらを統合することで HSI モデルを構築している^{3) 4)}。

従来手法ではある一日の漁場予測を行う際、その日の海洋環境変数のみで SI モデルを構築している。しかし、実際の漁場は漁獲日当日の海洋環境だけではなく、漁獲日から遡った過去の海洋環境も影響する。そのため、SI モデル構築に利用する海洋環境変数として、過去のデータを利用したい。一方、従来行われてきた各 SI モデル構築では複数のパラメータを設定するが、構築したモデルについてオーバーフィッティングの問題があるので適切なパラメータを人が設定する必要がある。過去の海洋環境変数データを SI モデル構築に利用する場合、このパラメータ決定作業は利用者の負担となる問題がある。

そこで本研究では、計算機を用いた HSI モデルの自動生成手法を提案する。またいくつかの実験を行い、過去のデータを利用することが漁場予測精度の向上につながることを確認し、提案手法が従来手法より精度よく漁場予測ができることを確認し、提案手法が有効であると結論付けた。

各章は以下のように構成される。2 章では本研究での関連研究を挙げ、3 章では従来手法及び提案手法について説明する。4 章では実験を行い、5 章

では考察を行う。6 章で本研究の結論と今後の展望について述べる。

第2章 関連研究

本研究ではアカイカを対象として漁場予測を行う。本章では漁場形成に関する研究、漁場予測技術に関する研究及びその可視化技術に関する先行研究を述べる。

2.1 漁場形成過程に関する研究

本研究で対象とするアカイカについての漁場形成過程は鈴木が過去に研究している⁵⁾。鈴木は日本近海の太平洋側におけるアカイカの魚群が暖流と寒流の境目に集まる可能性について述べた。暖流と寒流がぶつかる際に渦が発生し、その流れに乗って植物プランクトンが深海から巻き上げられる。深海に比べて太陽光が多い深度に巻き上げられた植物プランクトンは光合成が活発となり、その数が増える。植物プランクトンが増えることで、植物プランクトンを餌とする動物プランクトンも増える。さらに、動物プランクトンを餌とする小魚が集まり、小魚を餌とするアカイカが集まる。このような食物連鎖が起こることでアカイカの魚群が暖流と寒流の境目に集まるとされている。また、村田らは北太平洋におけるアカイカの季節回遊性についてまとめた⁶⁾。アカイカの産卵場は亜熱帯海域である。この海域で孵化したアカイカは7月頃に索餌域である亜寒帯海域付近まで北上し、秋に亜熱帯海域へ南下する。漁場予測の研究では古くから、このような魚の生態をモデル化し漁場予測に取り入れる試みが行われてきた。

2.2 漁場予測に関する研究

P.Lehodeyらはカツオやマグロ類の分布を予測するSEAPODYM(Spatial ecosystem and population dynamics model)を構築した⁷⁾。このモデルは海洋物理環境、餌生物モデル、移動回遊モデルから構成されたモデルである。特徴として魚類の移動に影響する水温、流速、溶存酸素、基礎生産等の海洋環境要因を考慮した分布予測モデルである。

2.3 HSI モデル

HSI (Habitat Suitability Index) モデルとは、対象生物の生息地としての適性を 0 から 1 で数値化するモデルであり²⁾、各環境変数から算出される。HSI は一般に、単一の環境変数から算出される SI (Suitability Index) を統合した値として計算される。単一の環境変数で生息地としての適性を 0 から 1 で数値化したモデルである SI モデルを統合することによって、複数の環境変数で生息地としての適性を 0 から 1 で数値化したモデルである HSI モデルを構築する。その統合手法には主に、算術平均法、幾何平均法、限定要因法、加算要因法がある。Tian らは同じくアカイカを対象とした HSI モデルを構築したが、その際に幾何平均法を用いている³⁾。また、五十嵐らもアカイカを対象魚種として同様に幾何平均法で HSI モデルを構築している⁴⁾。それらを踏襲し、本研究では幾何平均法を採用する。対象生物の生息地として適しているということと、対象生物の数は比例関係にあると仮定すると、HSI モデルは対象生物の多さを数値化したモデルであるとみなすことができる。

2.4 HSI モデル構築に関する研究

Tian らはアカイカの漁獲量予測を目的とした HSI モデルを 2 種類構築し、その比較を行った⁸⁾。

Chen らは環境変数に重みを付けて和を取る HSI モデルによって CPUE が説明できることを示し、weighting method による影響を評価した⁹⁾。

櫛田らは HSI モデル構築のための支援システムを開発した³⁾。このシステムではシミュレーションから得られた海水温、塩分濃度、経度方向の流速、緯度方向の流速、深さ方向の流速、の 5 変数それぞれに SI モデルを構築し、その相乗平均を HSI モデルとしている。また、各 SI モデルを単一変数の平滑化スプラインによる多項式近似で構築している。海洋環境変数から漁獲実績を誤差なく説明する SI モデルを構築することもできるが、この場合オーバーフィッティングの問題から漁場予測の精度が悪くなる傾向にある。そのため、このシステムではモデルの滑らかさ (平滑化パラメータ) を設定できるようにしている。利用者は SI モデルを確認し、最適な滑らかさの係数を選択する必要がある。この HSI モデルでは漁獲日当日のデータのみを用いている。

本研究では、この負担をなく手法を提案する。また、提案手法がこの HSI モデル構築手法より精度よく漁場予測ができることを確認する。

第3章 提案手法

本章では、従来手法及び提案手法について述べる。

3.1 使用データ

HSI 値を求めるためには海洋環境変数データ、対象魚種の漁獲実績データが必要となる。本研究で利用したこれらのデータについて述べる。

3.1.1 海洋環境変数データ

本研究で使用する海洋環境変数データは、気象庁気象研究所で開発された海洋データ同化システムで作成された MOVE (MRI Multivariate Ocean Variational Estimation) データである。この海洋データ同化システムは数値海洋モデル及び同化システムから成るシステムである¹⁰⁾。

3.1.2 過去のデータの計算

本研究では過去の海洋環境変数を利用することを提案している。ある地点での漁獲を過去の海洋環境変数から説明する場合、その地点の海水がどこから来たものなのか計算する必要がある。本研究で使用する海洋環境データは3次元方向の速度も含んでいるので、海水を遡って計算することができる。そして深さごとに異なる速度を持っているので、それぞれの深さをスタート地点として遡ると全く別の地点にたどり着く。今回はより正確に過去の地点を遡るために4次のルンゲクッタ法を用いた。

3.1.3 ルンゲクッタ法

ルンゲクッタ法は常微分方程式の数値解を求める手法の一種である。同じように常微分方程式の数値解を求める手法として、オイラー法がある。海洋環境変数を使用することである地点から流れを遡る場合、その地点の速度に

ステップの時間幅をかけることで過去の地点を求める手法がオイラー法である（式 3.1）。

$$(x^{(t-1)}, y^{(t-1)}, z^{(t-1)}) = (x^t, y^t, z^t) - (u^t, v^t, w^t) \cdot \Delta t \quad (3.1)$$

遡っているのに速度が負になっている。オイラー法では速度はそのままその地点の速度を用いて、ステップの時間幅で速度変化がないものとして扱っている。しかし、実際には微小時間ごとに速度変化は起きている。ルンゲクッタ法はオイラー法の式である式 3.1 における u^t と v^t と w^t をより正確に求める手法である。そのため、速度変化が一切ない領域ではどちらの手法も同じ結果となるが、本研究で使用する海洋環境変数データはそのような領域がないため、ルンゲクッタ法による精度向上が期待できる。

4 次のルンゲクッタ法は次のように計算する。

1. k_1 から k_4 を計算する
2. k_1 から k_4 を重みを付けて平均化する
3. それを速度としてステップの時間幅 Δt をかけることで、移動量とする
4. もとの位置に移動量を足すことで解とする

k_1 は式 3.2 で計算する。

$$k_1 = \mathbf{f}(\mathbf{X}, t) \quad (3.2)$$

ただし $\mathbf{f}(\mathbf{X}, t) = (u, v, w)$ であり、 $\mathbf{f}(\mathbf{X}, t)$ は時刻 t 、座標 (x, y, z) における速度ベクトルである。

k_2 は式 3.3 で計算する。

$$k_2 = \mathbf{f}\left(\mathbf{X} + k_1 \frac{\Delta t}{2}, t - \frac{\Delta t}{2}\right) \quad (3.3)$$

$\mathbf{f}(\mathbf{X} + k_1 \frac{\Delta t}{2}, t - \frac{\Delta t}{2}) = (u + \frac{\Delta t}{2} k_{1u}, v + \frac{\Delta t}{2} k_{1v}, w + \frac{\Delta t}{2} k_{1w})$ であり、 $\mathbf{f}(\mathbf{X} + k_1 \frac{\Delta t}{2}, t - \frac{\Delta t}{2})$ は時刻 $t - \frac{\Delta t}{2}$ 、座標 $(x + \frac{\Delta t}{2} k_{1u}, y + \frac{\Delta t}{2} k_{1v}, z + \frac{\Delta t}{2} k_{1w})$ における速度ベクトルである。

k3 は式 3.4 で計算する。

$$k3 = f(\mathbf{X} + k2\frac{\Delta t}{2}, t - \frac{\Delta t}{2}) \quad (3.4)$$

k4 は式 3.5 で計算する。

$$k4 = f(\mathbf{X} + k3\Delta t, t - \Delta t) \quad (3.5)$$

k1 から k4 を用いて、ステップ時間前の座標は次式のようにになる。

$$\mathbf{X}^{(t-1)} = \mathbf{X}^t - \frac{\Delta t}{6}(k1 + 2k2 + 2k3 + k4) \quad (3.6)$$

このようにして遡った地点を一日ごとにサンプリングする。サンプリングは線形補間で行う。

3.1.4 漁獲実績データ

本研究で使用する漁獲実績データは、日本近海の太平洋沖におけるアカイカの漁獲実績データある。漁獲実績データには日付、緯度、経度、漁獲量、操業開始時間、操業終了時間、CPUE（Catch Per Unit Effort）値が記録されている。CPUE 値は単位釣り機が単位時間に捕獲したアカイカの重量を表している。すべての漁船に搭載される釣り機が同じ性能であると仮定すると、CPUE 値はその海域での魚群の密度を反映する値であることができる。このことから、CPUE 値は漁場予測において重要な値である。また、この漁獲実績データは、HSI モデリング結果を可視化する際に漁を行った地点をマーキングするためにも使用する。例えば、太平洋の漁獲実績を基に日本海での漁獲量を予測しようとしても、それぞれの漁獲に関連がないので、有効な予測をすることはできない。漁を行った地点をマーキングすることで、HSI モデリングの有効な範囲を大まかに把握することができる。

3.2 HSI モデルから漁獲量予測への変換

先に説明したように HSI はその地点が対象生物の生息地として適しているかどうかを 0 から 1 で数値化したモデルである。一方、漁獲量を表す CPUE は

単位釣り機が単位時間に捕獲した対象生物の重量である。これらが比例関係にあると仮定することで漁獲予測に関する過去の研究が行われてきた。HSI=0の時は対象生物がおらず CPUE=0 となり、HSI=1 の時は対象生物が最大数おり CPUE=max となるとすると、その変換は式 3.6 で行うことができる。

$$\text{PredictedCPUE} = (\text{maxCPUE} - 0) \times \text{HSI} \quad (3.7)$$

$$= \text{maxCPUE} \times \text{HSI} \quad (3.8)$$

PredictedCPUE は漁獲量予測、maxCPUE は観測された最大 CPUE である。本研究では HSI と CPUE の変換をこの式で行う。

3.3 SI モデル

漁獲量を単一の海洋環境変数から説明するモデルが SI モデルである。以下では従来の SI モデル構築手法と提案手法を説明し、その違いを述べる。

3.4 従来手法による SI モデル構築

従来手法での SI モデル構築は次の手順で行う。まず、SI モデルを構築したい変数について、深さインデックスを一つ選ぶ。実測した漁獲点から垂直に見た時の深さインデックスでの変数値と、その漁獲点での CPUE を散布図に描点する。次に平滑化パラメータを一つ設定し、散布図を平滑化スプラインで近似する。深さインデックスの選択と、平滑化パラメータの設定を繰り返し、利用者が最適と判断したスプライン曲線を SI モデルとする。これを変数ごとに行うことで、各 SI モデルを構築する。この手法で SI モデルを構築する場合、変数値と CPUE の散布図を作成する時点で漁獲を行った時刻の情報が失われる。そのため、SI モデルの時間変化を考慮することができない。そこで、SI モデルが変化しないよう短い時間幅でモデルを構築する必要がある。この散布図をスプライン曲線で近似することで、SI モデルを構築する。次の変数でも同様に深さインデックスを選び直しスプライン曲線で近似する。こうすることで 5 種の SI モデルを構築できる。

3.5 提案手法による SI モデル構築

提案手法での SI モデルの構築を説明する。本研究で使用する海洋環境変数データは5種の変数、塩分濃度、温度、緯度方向の流速、経度方向の流速、深さ方向の流速からなる。また各変数は特定の深さごとに値を持っており、それらの線形和を使って SI モデル構築をする場合、次式で表すことができる。

$$SI = \sum_{i=0}^{depth} a_i x_i \quad (3.9)$$

x_i は漁獲点の鉛直方向深さ i での海洋環境変数の値、 a_i はその係数である。さらに過去のデータとして、先に説明した手法で計算した海洋環境変数データを用いると次式で SI モデルを表せる。

$$SI = \sum_{j=0} \sum_{i=0}^{depth} a_{ij} x_{ij} \quad (3.10)$$

x_{ij} は漁獲点の鉛直方向の深さ i を始点として j 日遡った地点での海洋環境変数の値、 a_{ij} はその係数である。

漁獲点における式(3.8)の SI 値を漁獲量予測に変換した値と漁獲実績 (CPUE) との差の絶対値を最小化する a_{ij} を求めることで SI モデルを構築できる (式 3.10)。

$$\arg \min_{a_{ij}} (|CPUE - PredictedCPUE|) \quad (3.11)$$

$$= \arg \min_{a_{ij}} \left(\left| CPUE - (maxCPUE * \sum_{j=0} \sum_{i=0}^{depth} a_{ij} x_{ij}) \right| \right) \quad (3.12)$$

また、実測した漁獲点は複数存在するのでこれを行列で表すことができる (式 3.12)。

$$\arg \min_{\theta} \|\mathbf{Y} - \mathbf{X}\theta\|_2^2 \quad (3.13)$$

\mathbf{Y} は $n \times 1$ 行列で、 i 行目は i 番目の漁獲点での CPUE となっている。 \mathbf{X} は k 番目の漁獲点で成立する式 3.9 の x_{ij} を k 行目に横に並べた行列である。 θ は $n \times 1$ 行列で \mathbf{X} における x_{ij} の並びと同じ順である。式 3.12 ではどの漁獲点でも利用可能なモデルを構築しようとしているので、 x_{ij} はどの漁獲点でも同じ係数となる。そのため、 θ は $n \times m$ 行列ではなく、 $n \times 1$ 行列となる。式 3.12 は機械学習の分野で教師あり学習と呼ばれたり、回帰問題と呼ばれる問題で最小二乗法などの解法によって解くことができる。式 3.12 を解くと全ての変数に 0 でない係数が現れる。これは全ての深さの全ての変数が漁獲量に影響するということを意味するが、例えばある深さで対象生物を漁獲する際に、同じ日の全く別の深さの海洋環境変数とその漁獲量に影響するとは考えにくく、海洋の専門家も否定している。また、すべての変数を漁獲モデルに使用するとオーバーフィッティングの問題が発生する。そのため、漁獲モデルに全ての変数を利用するのではなく、一部を削減する必要がある。

3.5.1 オーバーフィッティング

構築したモデルが、構築に使用したデータに対してはよく適合しているが、未知のデータに対して適合していない状態がオーバーフィッティング（過学習）である。今回のように計算機による自動計算でモデルを構築する際によく問題となる状態である。今回の例であれば、実測した漁獲点で予測させた場合は実際の CPUE に近い値が出るが、それを実際の予測に使用すると実際の CPUE とかけ離れたモデルができてしまう。一般にオーバーフィッティングはモデルの自由度が高すぎる場合に発生する。そこで、式 3.9 のように使用可能な全ての変数を使用するのではなく、変数の数を減らす工夫をする。以下で説明する変数の数を減らすモデル構築手法はスパースモデリングとも呼ばれる。例えば係数行列 θ の成分に 0 の項があれば、実質的に変数の数が減っていることになる。その手法には Ridge 回帰⁷⁾、Lasso 回帰¹¹⁾ など様々な手法がある。Ridge 回帰は古くに提案された変数削減手法であり次のように行う。式 3.12 で示した最小化式に罰則項として L2 ノルムを加えたものである。Lp ノルムは一般に $p_i=1$ で定義されるノルムで、式 3.13 で表せる。

$$Lp\text{norm} = ||x||_p = (|x_0|^p + |x_1|^p + |x_2|^p + \dots)^{\frac{1}{p}} \quad (3.14)$$

p=2 の場合が L2 ノルムであり、次式で示すように各係数の二乗和の平方根となる。

$$L2norm = ||x||_2 = (|x_0|^2 + |x_1|^2 + |x_2|^2 + \dots)^{\frac{1}{2}} = \sum_i |x_i|^2 \quad (3.15)$$

これを罰則項とした上で罰則項そのものの重み λ も考慮し、式 3.12 に対し適用すると次式となる。

$$\arg \min_{\theta} \left(||\mathbf{Y} - \mathbf{X}\theta||_2^2 + \lambda \sum_i |\theta_i|^2 \right) \quad (3.16)$$

ただし、 θ_i は θ の i 番目の成分である。式 3.15 が Ridge 回帰である。Ridge 回帰では罰則項が係数の二乗和の平方根であり、変数の数（係数の非 0 の項の数）については一切制約がない。しかし、モデルに与える影響の小さい変数に対応する係数は 0 に近づき、閾値処理によって係数 0 にできる。これは次の理由で説明できる。まず、一般化した Ridge 回帰において、最小化式は式 3.16 となる。

$$\min E(\mathbf{w}) + \lambda \sum_i |w_i|^2 \quad (3.17)$$

この目的関数を w_i で微分すると次式になる。

$$\frac{\partial E(\mathbf{w})}{\partial w_i} + \frac{1}{2} \lambda w_i \quad (3.18)$$

式 3.17 の第一項は特定の w_i で 0 になり、それ未満では正、それより大きい時は負となる。第二項は $w_i = 0$ で 0 になり、それ未満では負、それより大きい時は正となる。また、線形回帰の場合第一項は階段関数となる。そして第一項はその変数がモデルに与える影響の大小に比例する。そのため、影響が小さい変数の場合、第二項が係数 w_i 決定に支配的となり w_i は 0 に近づく。一方、影響の大きい変数の場合、第二項が w_i 決定にはあまり寄与しなくなり、 w_i はあまり 0 に近づかない。そのため、モデル決定に対して影響力の小さい変数は w_i が 0 に近づく。しかし、この手法では w_i が 0 に近づくほど第二項の

影響が小さくなるので w_i は完全には 0 にならず、閾値処理をしないと w_i を 0 にできない。

そこで、Lasso 回帰が考案された。Lasso 回帰は罰則項に L1 ノルムを用いた手法である。

$$L1norm = ||x||_1 = (|x_0|^1 + |x_1|^1 + |x_2|^1 + \dots)^{\frac{1}{1}} = \sum_i |x_i| \quad (3.19)$$

L1 ノルムとはつまり、係数行列の絶対値の和である。こちらに変数の数（係数の非 0 の項の数）についての制約はない。しかし、実際には係数の一部が 0 になる。それを以下で説明する。

一般化した Lasso 回帰の最小化式は次式となる。

$$\min E(\mathbf{w}) + \lambda \sum_i |w_i| \quad (3.20)$$

この目的関数を w_i で微分すると次式になる。

$$\frac{\partial E(\mathbf{w})}{\partial w_i} + \lambda \text{sgn}(w_i) \quad (3.21)$$

$\text{sgn}(x)$ は符号関数である。これも先と同様に考えると w_i は第二項を入れな
い場合と比較し、0 に近づく。しかし、符号関数は w_i の正負のみで w_i の大小
に関わらない値を持つ。そのため第一項が小さく、モデルに対する影響の小
さい変数は $w_i = 0$ に収束する。これにより、Ridge 回帰のような閾値処理を用
いることなく、変数の削減が可能である。

Lasso 回帰は影響の小さい変数の係数が 0 に収束するが、あくまで罰則項は
係数の絶対値の和である。そのため、ある変数が別の変数と従属の関係にあ
り、モデルに与える影響が大きい場合、どちらの変数もモデルに残ってしま
う。つまり、Lasso 回帰では削減できないが、モデル構築に不要な変数が存在
してしまう場合がある。

そこで罰則項として、係数行列 θ の L0 ノルムと重み λ の積を加えた次式を
考える。

$$\arg \min_{\theta} \left(\|\mathbf{Y} - \mathbf{X}\theta\|_2^2 + \lambda \sum_i \|\theta_i\|_0 \right) \quad (3.22)$$

式 3.13 は $p=1$ の場合で定義されているが、それを $p=0$ の場合にも拡張することで L0 ノルムは次式で表せる。

$$L0norm = \|x\|_0 = (|x_0|^0 + |x_1|^0 + |x_2|^0 + \dots)^{\frac{1}{0}} = \sum_i (|x_i|^0)^{\infty} \quad (3.23)$$

$0^0 = 0$ と定義した場合、 $x_i = 0$ のとき $0^{\infty} = 0$ 、 $x_i \neq 0$ のとき $1^{\infty} = 1$ となる。つまり、式 (3.22) は x_i の非ゼロ成分の数となる。この場合、罰則項は変数の数に比例するので Lasso 回帰の場合の問題も発生しない。また、変数の数を削減することでオーバーフィッティングの問題も発生しにくい。

本研究は以上の理由より、L0 ノルムを罰則項としてモデル構築に使用する。次に、最適化式 3.21 をどのようにして解くか説明する。

3.6 提案手法のアルゴリズム

先に述べた計算手法を計算機上で実現するアルゴリズムについて説明する。

3.7 IRLS

IRLS とは

3.8 最適な罰則項の重みの決定

先に説明した IRLS アルゴリズムによって、L0 ノルムを罰則項としたモデル構築が可能である。一方、モデルにはまだ罰則項の重みが自由度として残っている。そのため、これも最適化する必要がある。この最適化は次に説明する正答率が最大となる重みを選択することによって行う。選ばれた重みを用いて最適化された漁獲モデルが、その海洋環境変数の SI モデルとなる。

3.9 正答率

本研究では正答率によって重みの選択を行う。ここでは正答率の説明を行う。まずテストデータの漁獲点情報を用いて CPUE を予測する。また、学習データ内の実測値としての CPUE の平均値を求める。予測した CPUE と学習データの平均 CPUE を比較した時の大小を求める。次に、先ほど予測した漁獲点で観測された実際の CPUE と学習データの平均 CPUE を比較した時の大小を求める。これらが一致するかどうかを全てのテストデータに対し求めることで正答率とした。

3.10 HSI モデル構築

先程の SI モデル構築を各海洋環境変数ごとに適用することで、対象魚種にとっての好適な各海洋環境変数ができる。それらの相乗平均を HSI モデルとすることで、対象魚種の最適な海洋環境の分布ができる。

$$HSI = \sqrt{\sum SI} = \sqrt{\sum_{k=0}^{depth} a_k x_k} \quad (3.24)$$

第4章 実験

本章では、過去のデータを用いることで漁獲モデルの精度が向上することを確認するための実験、及び提案手法の有効性を確認するための実験について説明する。

4.1 使用データ

この節では本実験で使用するデータについて説明する。

4.1.1 漁獲実績データ

本研究ではアカイカを対象とした漁獲実績データ（CPUE データ）を用いる。このデータは青森県八戸港に所属するイカ釣り漁船が、2006 年 1 月及び 2 月に行った操業を記録したものである。データには操業日、操業を行った緯度経度、漁獲量が記録されている。漁獲量に影響を与えるのはあくまで海洋環境変数なので、本実験では緯度経度情報は使用しない。ただし、このデータ全体を一回の実験で使用するのではない。先の章で述べたように、提案手法と従来手法ともに SI モデルを構築する時に時刻の情報が失われ、SI モデルの時間変化は考慮していない。そのため、このデータを月ごとに別個のデータとして扱うことで SI モデルの時間変化は無いものと仮定する。

4.1.2 海洋環境変数データ

本実験で使用する海洋環境変数として、気象庁気象研究所で開発された海洋データ同化システムで作成された MOVE(MRI Multivariate Ocean Variational Estimation) データを用いる。このデータでは海水温、塩分濃度、経度方向の流速、緯度方向の流速、深さ方向の流速が利用できる。また、各変数は 54 段階の深さを持っているが最大深さは 6000m にもなり、操業する地点ではそれほど深さはない。そのため、本実験では全ての漁獲点で最も浅い 300m の深さまでを使用する。これは MOVE データでは 20 段階目の深さである。先の

章で説明したように、過去のデータとして流跡線をたどりサンプリングしたデータを用いる。流跡線の開始点は漁獲点の鉛直方向で MOVE データの各段階ごとの深さの点である。そのため一つの漁獲点では 20 点の流跡線開始点が存在する。これをルンゲクッタ法によって 60 日遡り、一日ごとにサンプリングする。さらに各点では 5 種類の海洋環境変数を得ることができる。そのため、過去のデータまで利用する場合 6000 種類の変数が利用可能である。一方当日のデータのみを利用する場合は 100 種類の変数までである。

4.2 過去のデータを用いることの有効性を確認する実験

過去のデータを用いることでそもそも漁獲モデルの精度が向上するのかわかりを確認する実験を行った。

4.2.1 実験方法

一ヶ月分の漁獲実績データを一回の実験で使用する。まず、2006 年 1 月の漁獲実績データのみを用いる。このデータを日時の順に並べ、奇数番目を学習データとして扱う。学習データのみでモデルを構築する。構築するモデルは、過去のデータを用いた場合と当日のみのデータを用いた場合で、それぞれ提案手法を用いて SI モデル構築を行う。これを各海洋環境変数に対し行うことで 5 種類の SI モデルが構築できる。SI モデルを統合した、最終的な漁獲モデルである HSI モデルを評価する。そのために、偶数番目の漁獲実績データをテストデータとし、これをもとにそれぞれの HSI モデルの正答率を求める。先に説明した正答率は SI モデルを評価するためのものだが、これはそのまま HSI モデルの評価にも利用できる。次に、学習データとテストデータを入れ替えて同じ実験を行った。さらに、2006 年 2 月の漁獲実績データを用いて同じ実験を行った。

4.3 提案手法の有効性を確認する実験

提案手法と従来手法を比較する実験を行った。

4.3.1 実験方法

提案手法、従来手法それぞれについて先の実験と同じように漁獲実績データを月ごとに分割し、奇数番目の漁獲実績データのみを使って HSI モデルを構築する。ただし、使用する海洋環境変数データは従来手法に合わせるために当日のデータのみを使用する。それぞれの手法で構築した HSI モデルを先の実験と同じように正答率で評価する。こちらも学習データとテストデータを入れ替えて同じ実験を行い、月を変えて同じ実験を行った。

第5章 結果及び考察

5.1 実験結果及び考察

先の章で行った実験の結果を述べ、考察する。

第6章 結論

6.1 結論

本研究では、最適化を用いた HSI モデル構築の新しい手法を提案した。また提案手法の有効性を確認するために2つの実験を行った。その結果、過去のデータを用いたほうがより精度の高い HSI モデルを構築できることを確認した。また、提案手法のほうが従来手法と比べ、より正答率の高い HSI モデルを構築できることを確認した。

6.2 課題

一方で課題もある。従来手法ではスプライン曲線で SI モデルを構築しているが、提案手法ではより単純な線形回帰で SI モデルを構築している。スプライン曲線を最適化することでより精度の高いモデル構築が可能になることが期待できる。

深さ方向の流速は南北方向の流速に比べ、HSI モデルに対する寄与は小さいはずという専門家の意見がある。しかし本研究では各 SI モデルを統合し HSI モデルを構築する際、各 SI モデルを全く等価に扱っている。ここで各 SI モデルに重みを導入することによって、各海洋環境変数の HSI に対する寄与を変えることが可能である。これにより、例えば精度の悪い SI モデルの重みを減らし HSI モデルの精度に対する悪影響を減らすことが可能となる。

参 考 文 献

- 1) 気候変動に伴う水産資源・海況変動予測技術の革新と実利用化, 気候変動適応研究推進プログラム平成 24 年度報告書, (2012), pp. 4–16.
- 2) 田中章, HEP 入門―“ハビタット評価手続き”マニュアル, (2006).
- 3) 櫛田将史, 上中隆史, 坂本尚久, 小山田耕二, 海洋データ同化結果を使ったアカイカ好適生息域モデル作成支援可視化環境の開発, 第 27 回数値流体力学シンポジウム講演論文集, (2013).
- 4) 五十嵐弘道, 淡路敏之, 石川洋一, 蒲地政文, 碓氷典久, 酒井光夫, 加藤慶樹, 齊藤誠一, 清藤真樹, 海洋再解析データを用いたアカイカ好適生息域モデルの構築とその利用 recca アカイカプロジェクトの事例, *JAMSTEC Rep. Res. Dev.*, Vol. 18, (2014), pp. 89–101.
- 5) SUZUKI TSUNEYOSHI, Studies on the relationship between current boundary zones in waters to the southeast of hokkaido and migration of the squid, *ommastrephes sloani pacificus* (steenstrup), *MEMOIRS OF THE FACULTY OF FISHERIES HOKKAIDO UNIVERSITY*, Vol. 11, (1963), pp. 75–153.
- 6) 村田守, 北太平洋におけるいか流し網漁場の海洋環境及びアカイカの分布・回遊, 日本海ブロック試験研究集録, Vol. 17, (1990), pp. 144–148.
- 7) P.Lehodey, A spatial ecosystem and populations dynamics model (seapodym) - modeling of tuna and tuna-like populations, *Progress in Oceanography*, Vol. 78, (2008), pp. 304–318.
- 8) Siquan Tian, Xinjun Chena, Yong Chenc, Liuxiong Xu, Xiaojie Dai, Evaluating habitat suitability indices derived from cpue and fishing effort data for *ommastrephes bratramii* in the northwestern pacific ocean, *Fisheries Research*, Vol. 95, (2009), pp. 181–188.
- 9) C. Xinjun, T. Siquan, L. Bilin, C. Yong, Modeling a habitat suitability index for the eastern fall cohort of *ommastrephes bratramii* in the central north pacific

ocean, *Chinese Journal of Oceanology & Limnology*, Vol. 29, (2011), pp. 493–504.

- 10) Norihisa Usui, Shiro Ishizaki, Yosuke Fujii, Hiroyuki Tsujino, Tamaki Yasuda, Masafumi Kamachi, Meteorological research institute multivariate ocean variational estimation (move) system: Some early results, *Advances in Space Research*, Vol. 37, (2006), pp. 806–822.
- 11) Robert Tibshirani, Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 58, (1996), pp. 267–288.