

# 修 士 論 文

題 目 因果関係を考慮した最適化漁  
獲モデル構築手法の開発

指導教官 小山田 耕二 教授

京都大学大学院工学研究科修士課程 電気工学専攻

氏 名 高見 円仁

平成28年2月3日

# 目次

第1章 序論	1
第2章 関連研究	3
2.1 漁場形成過程に関する研究	3
2.2 漁場予測に関する研究	4
2.3 HSI モデル	4
2.4 HSI モデルの構築手法に関する研究	5
第3章 提案手法	7
3.1 使用データ	7
3.1.1 海洋環境変数データ	7
3.1.2 過去のデータの計算	7
3.1.3 ルンゲクッタ法	7
3.1.4 漁獲実績データ	9
3.2 HSI モデルから漁獲量予測への変換	10
3.3 SI モデル	10
3.4 従来手法による SI モデル構築	10
3.5 提案手法による SI モデル構築	11
3.6 オーバーフィッティング	13
3.7 提案手法のアルゴリズム	16
3.8 IRLS	16
3.9 最適な罰則項の重みの決定	16
3.10 正答率	17
3.11 HSI モデル構築	17
第4章 プログラムの実装	18
4.1 プログラム全体	18
4.2 流跡線計算プログラム	18

4.3	最適化計算プログラム . . . . .	18
4.4	地図描画プログラム . . . . .	18
<b>第 5 章</b>	<b>実験</b>	<b>19</b>
5.1	使用データ . . . . .	19
5.1.1	漁獲実績データ . . . . .	19
5.1.2	海洋環境変数データ . . . . .	19
5.2	過去のデータを用いることの有効性を確認する実験 . . . . .	20
5.2.1	実験方法 . . . . .	20
5.3	提案手法の有効性を確認する実験 . . . . .	20
5.3.1	実験方法 . . . . .	21
<b>第 6 章</b>	<b>結果及び考察</b>	<b>22</b>
6.1	実験結果及び考察 . . . . .	22
<b>第 7 章</b>	<b>結論</b>	<b>23</b>
7.1	結論 . . . . .	23
7.2	課題 . . . . .	23
	<b>参考文献</b>	<b>24</b>

# 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<sup>1)</sup>。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<sup>2)</sup>。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 モデルによって海水温や塩分濃度や流速などといった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<sup>3) 4)</sup>。

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ような因果関係が存在する。そのため、SI モデル構築に利用する海洋環境変数として、漁獲日当日の海洋環境変数データだけではなく、漁獲日から遡った過去の海洋環境変数データを利用したい。一方、従来行われてきた各 SI モデル構築ではパラメータをいくつか設定する操作を行う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、適切なモデル構築ができているかどうかを人が判断しモデル構築のための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、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 章ではプログラムをどのように実装したか説明し，5 章で実験を行い，6 章では考察を行う．7 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

本研究で漁場予測の対象とするアカイカは亜熱帯海域、温帯海域に広く分布するイカである。<sup>5)</sup> 亜熱帯海域、温帯海域は太平洋ではカムチャッカ半島から台湾付近の南北幅、日本からアメリカまでの東西幅がある広い領域である。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<sup>6)</sup>。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（黒潮）と寒流（親潮）の境目に集まる可能性について述べた。親潮は一般に栄養塩の濃度が高い。栄養塩とは窒素、リン、珪素の無機塩類の総称である<sup>7)</sup>。光と栄養塩は植物プランクトンの成長増殖にとって重要な因子である。その栄養塩を消費する植物プランクトンが親潮には少ないため栄養塩の濃度が高い。栄養塩は沈降するので特に海の深層に堆積する。

親潮と黒潮がぶつかる際に渦が発生し、その流れに乗って深層の栄養塩が深海から巻き上げられる。巻き上げられた栄養塩は深海に比べて太陽光が多い深度で植物プランクトンに消費され、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き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。

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<sup>8)</sup>。アカイカは季節的に南北に回遊している。アカイカの産卵場は南の亜熱帯海域である。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大きくなる。そして秋に再び亜熱帯海域へ南下し産卵する。また、アカイカを追跡することで日中は300m付近の深さにいて夜間は水面近

くへ浮上するという、一日の中でも鉛直方向の周期的な移動を行うことがわかっている<sup>9)</sup>。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<sup>10)</sup>。このモデルは海洋物理環境モデル、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。

根本らはアカイカの卵稚仔が産卵から1ヶ月ほどは自走できず海流によって浮遊していることに着目し、その追跡シミュレーションシステムを構築した<sup>11)</sup>。卵稚仔を粒子とみなし、海洋の流速データを用いて粒子を追跡シミュレーションしている。また単に追跡するだけではなく、海洋環境のトレースを行い卵稚仔が生存しているかどうかをシミュレーションし、それをもとにした漁獲量予測を行うことも可能である。

## 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<sup>2)</sup>、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。算術平均法は各SIモデルが算出した0から1の数値を相加平均したものでSI値が全て算出できなくてもHSI値を計算可能である。幾何平均は各SI値を相乗平均したものであり、複数のSI値が対象生物の生息に適さない領域での値が小さくなる傾向にある。限定要因法は各SI値の中で最小値のSI値をHSI値とする統合手法である。植物の成長は必要とする栄養素の内

最も少ないものに影響されるというリービッツの最小律<sup>12)</sup>と同じ概念で構築される。加算要因法は各 SI 値を足したもので算術平均法と違い、HSI 値が 1 より大きくなると 1 に制限される。このように HSI モデルによる数値化の目的によって各統合手法を使い分ける必要がある。

Tian らは本研究と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<sup>13)</sup>。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<sup>4)</sup>。それらを踏襲し、本研究では幾何平均法を採用する。

HSI モデルはアサリ<sup>14)</sup>、メバル<sup>15)</sup> など他の海洋生物だけでなく、オオタカ<sup>16)</sup> などの陸上生物、海草<sup>17)</sup> <sup>18)</sup> などの植物にも適用事例があり、広く一般に使われている指標である。また、対象生物にとっての生息しやすさを数値化することから、開発事業によってどれだけ生息しやすさが変化するかの環境アセスメントの評価<sup>19)</sup> や生物多様性の評価<sup>20)</sup> にも使われる。

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 モデルの構築手法に関する研究

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<sup>13)</sup>。

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<sup>21)</sup>。

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<sup>3)</sup>。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、この負担をなくしさらに漁獲量と海洋環境変数の間の因果関係についてもモデル構築に組み込んだ手法を提案する。また、提案手法が先に説明した従来の 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、データの計算法や従来の HSI モデル構築手法及び提案する HSI モデル構築手法などの基礎技術について述べる。

### 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<sup>22)</sup>。

#### 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.  $k1$  から  $k4$  を計算する
2.  $k1$  から  $k4$  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  $\Delta t$  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

$k1$  は式 3.2 で計算する。

$$k1 = \mathbf{f}(\mathbf{X}, t) \quad (3.2)$$

ただし  $\mathbf{bfX} = (x, y, z)$  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。

$k2$  は式 3.3 で計算する。

$$k2 = \mathbf{f}\left(\mathbf{X} + k1 \frac{\Delta t}{2}, t - \frac{\Delta t}{2}\right) \quad (3.3)$$

$\mathbf{bfX} + \frac{\Delta t}{2} k1 = (x + \frac{\Delta t}{2} k1_u, y + \frac{\Delta t}{2} k1_v, z + \frac{\Delta t}{2} k1_w)$  であり、 $\mathbf{f}(\mathbf{X} + \frac{\Delta t}{2} k1, t - \frac{\Delta t}{2})$  は時刻  $t - \frac{\Delta t}{2}$ 、座標  $(x + \frac{\Delta t}{2} k1_u, y + \frac{\Delta t}{2} k1_v, z + \frac{\Delta t}{2} k1_w)$  における速度ベクトルである。

k3 は式 3.4 で計算する。

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

k4 は式 3.5 で計算する。

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

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。漁船ごとに釣り機の性能や数は異なり、操業を行う人間の釣り機を扱う技能によっても CPUE は変化する。しかしすべての漁船に搭載される釣り機が同じ性能であり、操業を行う人間の技能も同じ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。生息しやすい環境に対象生物が集まる、生息しやすい環境では対象生物の生存確率が上がるので生息しやすい環境では対象生物の数が多くなると考えられる。対象生物の数が多くなると、一般には対象生物の CPUE が高くなる。そのため HSI 値と CPUE 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 3.7 で行うことができる。

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

$$= maxCPUE \times HSI \quad (3.8)$$

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。スプライン曲線は制御点を通る滑らかな曲線で、制御点を作る各区分ごとに多項式を立てている。図 3.1 は 2006 年 2 月の漁獲点データ 184

点を用いてそれぞれの漁獲点での CPUE を縦軸、その漁獲点での深さ 0.5m での塩分濃度を横軸とし、散布図を描いている。またさらに平滑化パラメータ  $=0.5$  に設定したスプライン曲線を重ねており、これを散布図の近似曲線として扱う。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。図 3.2 は同じように散布図を描いているが、平滑化パラメータ  $=0.999$  に設定したスプライン曲線を重ねている。このように海洋環境変数の分布が同じでも平滑化パラメータの設定によって近似曲線が大きく変化する。また、どのスプライン曲線が最もよく近似できているかの指標はなく、そのため利用者は最適な平滑化パラメータの設定を見て確認し決定する必要がある。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。

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。

$$SI = \sum_{j=0}^{depth} \sum_{i=0}^{depth} a_i j x_i j \quad (3.10)$$

$x_{ij}$  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 (3.9) 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 3.11)。

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

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

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 3.13)。

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

$\mathbf{Y}$  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 3.10 の  $x_{ij}$  を  $k$  行目に横に並べた行列である。 $\theta$  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 3.13 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、 $\theta$  は  $n \times m$  行列ではなく、 $n \times 1$  行列となる。式 3.13 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 3.13 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。

### 3.6 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 3.10 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  $\theta$  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<sup>23)</sup>、Lasso 回帰<sup>24)</sup> 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 3.13 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  $p \geq 1$  で定義されるノルムで、式 3.14 で表せる。

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

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。

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

これを罰則項とした上で罰則項そのものの重み  $\lambda$  も考慮し、式 3.13 に対し適用すると次式となる。

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

ただし、 $\theta_i$  は  $\theta$  の  $i$  番目の成分である。式 3.16 が Ridge 回帰である。式 3.16 は次のようにして解く。



式 3.16 は最小化を行いたいので  $\theta$  で偏微分すると 0 になる（式 3.17）。

$$2X^T X \theta - 2X^T Y + 2\lambda \theta = 0 \quad (3.17)$$

これを整理すると、解くべき方程式は式 3.18 となる。

$$(X^T X + \lambda I) \theta = X^T Y \quad (3.18)$$

式 3.18 を満たす  $\theta$  は  $X^T X + \lambda I$  の逆行列を求めることで求めることができる。これによって Ridge 回帰は解くことができる。

このようにして求めた解はオーバーフィッティングの問題が発生しないようなスパースなモデルになっているだろうか。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。しかし実際には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 3.19 となる。

$$\min \left( E(\mathbf{w}) + \lambda \sqrt{\sum_i |w_i|^2} \right) \quad (3.19)$$

この目的関数を  $w_i$  で微分すると次式になる。

$$\frac{\partial E(\mathbf{w})}{\partial w_i} + \lambda w_i \quad (3.20)$$

式 3.20 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。モデルに対する影響の大小によって罰則項の効果が変化し、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。

$$L1norm = \|x\|_1 = (|x_0|^1 + |x_1|^1 + |x_2|^1 + \dots)^{\frac{1}{1}} = \sum_i |x_i| \quad (3.21)$$

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。

$$\min E(\mathbf{w}) + \lambda \sum_i |w_i| \quad (3.22)$$

この目的関数を  $w_i$  で微分すると次式になる。

$$\frac{\partial E(\mathbf{w})}{\partial w_i} + \lambda \text{sgn}(w_i) \quad (3.23)$$

$\text{sgn}(x)$  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  $\theta$  の L0 ノルムと重み  $\lambda$  の積を加えた次式を  
考える。

$$\arg \min_{\theta} \left( \|\mathbf{Y} - \mathbf{X}\theta\|_2^2 + \lambda \sum_i \|\theta_i\|_0 \right) \quad (3.24)$$

式 3.14 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。

$$L0norm = \|x\|_0 = (|x_0|^0 + |x_1|^0 + |x_2|^0 + \dots)^{\frac{1}{0}} = \sum_i (|x_i|^0)^{\infty} \quad (3.25)$$

$0^0 = 0$  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 (3.25) 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 3.24 をどのようにして解くか説明する。

### 3.7 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。

### 3.8 IRLS

IRLS とは

### 3.9 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。

### 3.10 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。

### 3.11 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。

$$HSI = \sqrt{\sum SI} = \sqrt{\sum_{k=0}^{depth} a_k x_k} \quad (3.26)$$

## 第4章 プログラムの実装

### 4.1 プログラム全体

本研究で提案する HSI モデルの新しい構築手法について先の章で説明した。この章ではその手法を実際にどのように実装したかについて説明する。

本研究では提案手法を3つのプログラムに分けて実装した。

- 流跡線計算プログラム
- 最適化計算プログラム
- 地図描画プログラム

各プログラムについて次節から説明する。

### 4.2 流跡線計算プログラム

流跡線計算プログラムではルンゲクッタ法を用いた流跡線計算を行い、計算結果に沿ったサンプリングを行う。

### 4.3 最適化計算プログラム

最適化計算プログラムでは流跡線計算プログラムの結果を入力とし、HSI モデル構築に必要な各 SI モデル式を計算する。

### 4.4 地図描画プログラム

地図描画プログラムでは最適化計算プログラムの結果として得られた SI モデル式を基に HSI モデルを構築し、地図に描画する。

## 第5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。

### 5.1 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。

#### 5.1.1 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。

#### 5.1.2 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。

## 5.2 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どうかを確認する実験を行った。

### 5.2.1 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。

## 5.3 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。

### 5.3.1 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6章 結果及び考察

### 6.1 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7章 結論

### 7.1 結論

従来の HSI モデル構築手法では CPUE と海洋環境変数データの間の因果関係を考慮することができない、HSI モデルを構築する際に利用者の負担が大きい、という問題があった。本研究では、それらの問題を解決するために最適化手法を用いた 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正答率の高い 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。

### 7.2 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。

海洋環境変数データの中で、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, (朝倉書店, 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) Roper C.F.E., Sweeney M.J., Nauen C.E, Fao species catalogue: an annotated and illustrated catalogue of species of interest to fisheries, *FAO Fish. Synop.*, Vol. 3, (1984), pp. 175–176.
- 6) 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.
- 7) 宮尾孝, 藤原弘行, 川村知裕, 北川隆洋, 永井直樹, 栄養塩分析の高精度化, 測候時報, Vol. 80, (2013), pp. 139–147.
- 8) 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.
- 9) 中村好和, イカ類の野外行動観察の手法特にバイオテレメトリーを中心に, 平成4年度イカ類資源研究会議報告, (1994), pp. 118–128.

- 10) 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.
- 11) 根本利弘, 五十嵐弘道, 淡路敏之, 喜連川優, 水産資源研究のための粒子追跡シミュレーション: インタフェースの作成, 情報処理学会全国大会講演論文集, Vol. 72, (2010), pp. 41–42.
- 12) Justus von Liebig, *Die Grundsätze der Agricultur-Chemie mit Rücksicht auf die in England angestellten Untersuchungen*, (Friedrich Vieweg und Sohn Publisher, 1855).
- 13) Siquan Tian, Xinjun Chena, Yong Chenc, Liuxiong Xu, Xiaojie Dai, Evaluating habitat suitability indices derived from cpue and fishing effort data for *ommatrephes bratramii* in the northwestern pacific ocean, *Fisheries Research*, Vol. 95, (2009), pp. 181–188.
- 14) 新保裕美, 田中昌宏, 池谷毅, 越川義功, アサリを対象とした生物生息地適性評価モデル, 海岸工学論文集, Vol. 47, (2000), pp. 1111–1115.
- 15) 日本環境アセスメント協会・研究部会自然環境影響評価技法研究会, メバル稚魚・幼魚 hsi モデル, 自然環境影響評価技法研究会報告書, (2004), pp. 143–162.
- 16) 大野剛, 藤原靖, Gis による自然度評価およびオオタカ hsi モデルからみた建設事業地選定の可能性について, 地球環境シンポジウム講演集, Vol. 17, (2009), pp. 103–108.
- 17) 田中昌宏, 上野成三, 林文慶, 新保祐美, 高山百合子, 沿岸自然再生の計画・設計を支援する環境評価手法に関する一考察, 土木学会論文集, Vol. 741, (2003), pp. 89–94.
- 18) 高山百合子, 上野成三, 勝井秀博, 林文慶, 山木克貝, 田中昌宏, 江奈湾の藻場分布データに基づいたアマモの hsi モデル, 海岸工学論文集, Vol. 50, (2003), pp. 1136–1140.

- 19) 田中章, 大澤啓志, 吉沢麻衣子, 環境アセスメントにおける日本初のH E P適用事例, ランドスケープ研究, Vol. 71, (2009), pp. 543–548.
- 20) 浜本洋, 風間聡, 澤本正樹, Hsl モデルを用いた流域生物多様性の評価, 環境工学研究論文集, Vol. 43, (2006), pp. 559–567.
- 21) C. Xinjun, T. Siquan, L. Bilin, C. Yong, Modeling a habitat suitability index for the eastern fall cohort of *ommastrephes bartramii* in the central north pacific ocean, *Chinese Journal of Oceanology & Limnology*, Vol. 29, (2011), pp. 493–504.
- 22) 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.
- 23) Arthur E. Hoerl, Robert W. Kennard, Ridge regression: Biased estimation for nonorthogonal problems, *Technometrics*, Vol. 12, (1970), pp. 55–67.
- 24) Robert Tibshirani, Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 58, (1996), pp. 267–288.

## Scatter Plot Graph

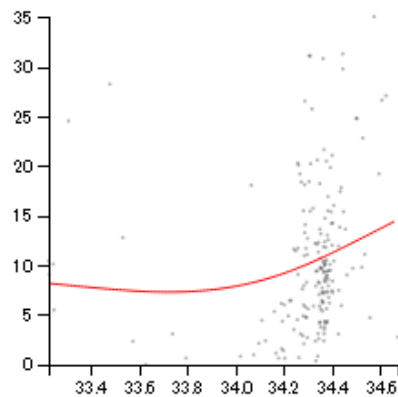


図 3.1: 2006 年 2 月の漁獲点における塩分濃度と CPUE の散布図と、スプライン係数=0.5 に設定したスプライン曲線による近似曲線

## Scatter Plot Graph

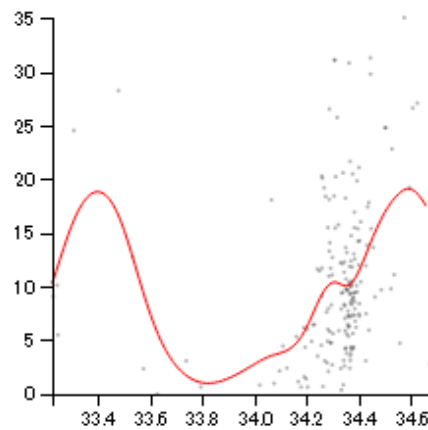


図 3.2: 2006 年 2 月の漁獲点における塩分濃度と CPUE の散布図と、スプライン係数=0.999 に設定したスプライン曲線による近似曲線