

# 環境DNAデータのデータベースを用いた 魚類生息地域の可視化手法の検討

中島 悠貴<sup>1</sup> 伊藤 正彦<sup>1,a)</sup>

**概要：**近年地球温暖化による魚類の生息量や生息地域に変化が表れており、暖かい地域に生息している魚類が本来生息していないはずの地域で観測されることが増えてきている。しかしどのような魚類の生息域や生息地域がどのように変化しているかを理解するには報道等の情報で理解するのは難しい。そのため本研究では環境DNAという調査手法のデータベースを用い、検出された魚類によって地点の類似度を求め、採水された地点の相対的な関係を可視化し、専門的な知識を必要とせずに魚類の生息量と温暖化による生息地域を知ることのできるよう可視化する。

## 1. はじめに

近年地球温暖化による魚類の生息量や生息地域に変化が表れており、暖かい地域に生息している魚類が本来生息していないはずの地域で観測されることが増えてきている。この現状を調査する手法として環境DNAというものがある。従来の生物調査の手法は捕獲や視認が必要となり、時間と労力を多く必要とするものであったが、環境DNAによる調査はそれらを必要とせず、生物の種類や個体数の推測が可能になった。しかしこれらは研究者が使用することが多く、専門の知識が必要となる。またこれまでは環境DNAのオープンデータベースが一般には公開されていなかったため一般の人が生物の生息状況に詳しく知ることは難しいことであった。

しかし近年にオープンデータベースが公開されたため、広範囲かつ多様な地点での観測や定期的な観測等による視認することが難しい詳細な情報も確認することができるようになった。とはいえ魚類の学名やそこに生息している魚類なのかそうでない魚類なのかを理解するには専門的な知識を必要とする。

本研究では環境DNAのオープンデータベースを用い、検出された魚類によって地点の類似度を求める。これにより地球温暖化の影響で魚類の生息地域に変化があり、本来生息していない地域でも検出されていれば、可視化した際に遠い関係に表示されるはずの地点らが近い関係で表示されることが考えられる。そして専門的な知識を必要とせず、魚類の生息量と温暖化による生息地域の変化を知るこ

とのできるよう可視化する。

## 2. 関連研究

生物と環境の関係性は数多く調査されており、本研究のような生物の種類と生息量を可視化する研究も行われている。海老原らは市民科学による東京のチョウと植物の共生ネットワークモニタリングの可能性の研究を行っている [2]。この研究では市民により集められたチョウのデータと訪花吸汁に訪れられた植物のデータを使用し、植物に訪れたチョウの種類によるグループ化と、関係性のネットワーク図を作成している。これより東京のチョウに多く訪花吸汁されている植物とチョウと植物の共生的相互作用を研究している。

なお本研究では検出された魚類による地域の関連性を可視化し、生息している魚類の変化を探索する点で先行研究とは手法および目的が異なる。

## 3. 使用するデータ

本研究では環境DNAデータのデータベースである、ANEMONE DB [1] を用いる。データベースには和名のデータが存在しないため、日本産魚類全種リスト (JAF リスト) [3] から和名のデータを使用している。

環境DNAとは個々の生物個体からではなく、海・川・湖沼等の水、土壌、大気といった環境の中に存在する生物由来のDNAの総称である。環境DNAは生物の捕獲や視認が必要なく、だれでも調査に参加できることから新しい生態系の把握方法とされている。

ANEMONE DB [1] は環境DNAのデータベースである。ANEMONE DBには界 (kingdom)、目 (order)、科

<sup>1</sup> 北海道情報大学

<sup>a)</sup> imash@do-johodai.ac.jp

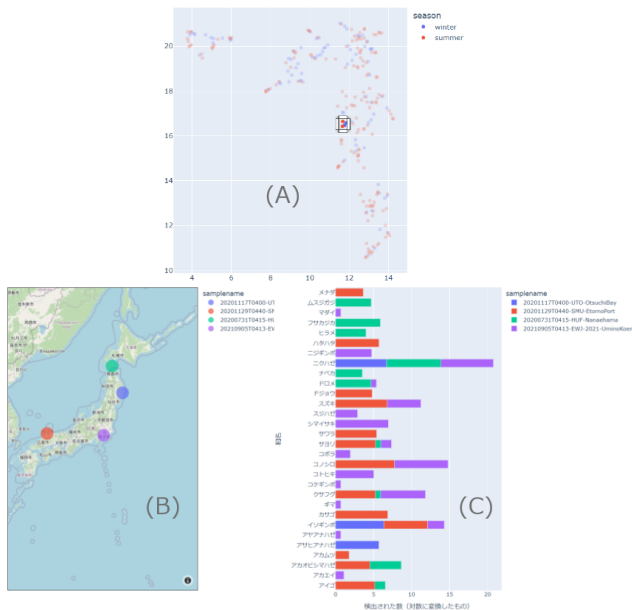


図 1 類似度を求められた地点に生息している魚類を見るシステム図

(family), 種 (species) や塩基配列, 検出された数 (nreads), などの検出された魚類の種類と数のデータがデータベースとして存在する. ここでの検出された数は採水された水の中にある DNA 情報をもとに生息していると考えられる魚類の推定数である. 著者らがデータベースからデータを取得した 2023 年 5 月には 2017 年から 2022 年までのデータが存在し, 魚の種類は約 4700 種ほどとなっている. データベースには日本全域分のデータがあり約 200 地点から採水されている. 海や川, 湖などのいろいろな場所のデータがあるが海で採水されたデータは主に港等のデータとなっている. また, 採集された回数は場所によって変わる.

日本産魚類全種リスト (JAF リスト) は一般の方々が最新の日本魚類の名称を調べることを目的としたデータとなっている [3].

本研究では ANEMONE DB からは年月日, 地名, 緯度経度, 種, 検出された数のデータ, 日本産魚類全種リスト (JAF リスト) からは種, 和名のデータを使用する. ANEMONE DB には何らかの理由により種等に unidentified (正体不明) とついたものが存在するがこれらは使用しない.

上記のデータをもとに必要なデータの抽出・追加を行った. 緯度経度のデータが一つのデータとして扱われていたため, 緯度と経度に分けたものを作成した. 年月日と季節のデータは 20210125 というような数字の列となっており, そのままでは年月日として使用することができない. しかし本研究では一地点の最も新しデータを使用するため, 識別するために作成した. また季節は 5 月から 9 月までを夏, 11 月から 3 月までを冬とし, 夏と冬の 2 つにしたものを使用している.

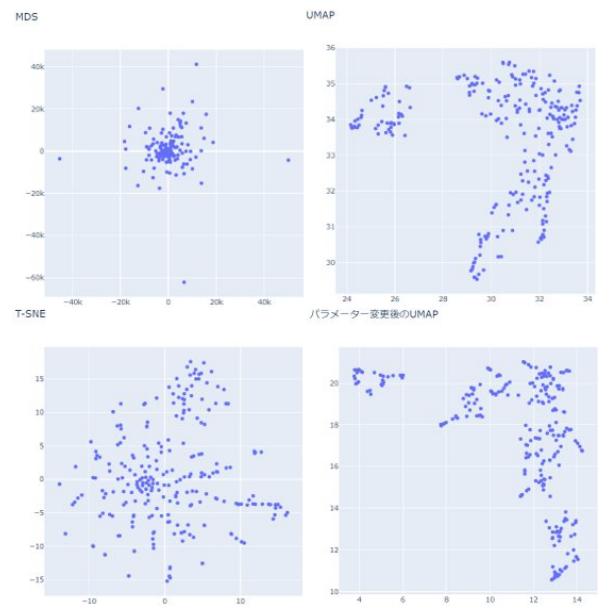


図 2 左上: MDS, 左下: T-SNE, 右上: UMAP, 右下: パラメータ変更後の UMAP

## 4. 実装システム

検出された魚類によって地点の類似度を求め, 魚類による採水された地点の相対的な関係を可視化する. また魚類の生息量と生息地域を知ることができるように可視化するための実装システムを作成した (図 1).

図 1(A) では採水された地点で検出された魚類の種類や量などによる相対的な関係を表示している. ユーザーがこの散布図内で範囲選択をすることで (B) と (C) で詳細なデータを見ることができる. 採水された地点の相対的な関係は UMAP を用いて類似度を求め, 散布図に可視化している. (B) では (A) の散布図で選択されたデータの採水された地点をマップ上に表示している. (C) では (A) の散布図で選択されたデータの地点で検出された魚類とその魚類の検出された数を積み上げ棒グラフび表示している.

以上を Python 内で使用可能なウェブアプリケーションフレームである Dash を使用し, インタラクティブな可視化システムとして実装した.

### 4.1 採水された地点の相対的な関係の可視化

今回は採水された地点の相対的な関係を求める手法として UMAP を使用する. 本研究では採水された地点の検出された魚の種類と数の類似度を求め散布図に表示する. この際に表示されるのは地点である. 使用するデータは高次元データであるため直感的な理解が難しくなる. そのため次元削減を行い, 直感的に理解しやすい低次元データに変換する.

実装するにあたり, 手法の候補として, 多次元尺度構成

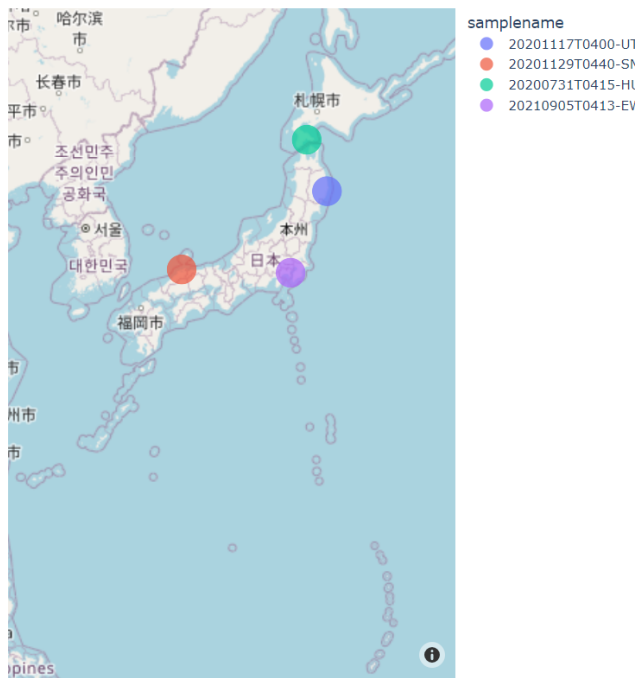


図 3 図 1(A) の散布図で選択されたデータの採水された地点を表示するマップ

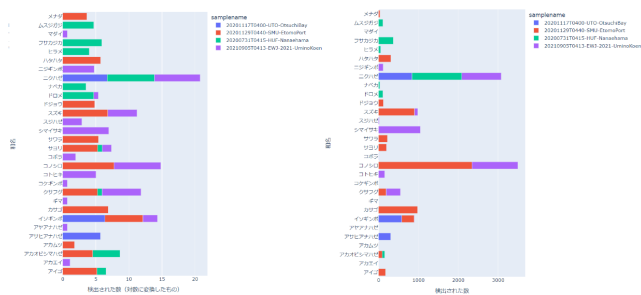


図 4 図 1(A) の散布図で選択されたデータの検出された魚類と数を表示する積み上げ棒グラフ (左: 対数表示, 右: 通常表示)

法である MDS と次元削減が可能である t-SNE と UMAP の 3 つを検討した。

MDS の可視化結果は図 2 の左上となっている。このように中心に多くまとまり、その周りに点々と表示されるというような大域的な位置関係を可視化することになる。本研究で点と点の近さやまとまりなどの局所的な位置関係を可視化するため本研究では MDS を使用しない。

t-SNE の可視化結果は図 2 の左下となっている。t-SNE では局所的な位置関係を可視化することができている。しかし局所的過ぎるため、多くの魚の種類で似ていないと近くて表示されない可能性がある。そのため本研究のような本来生息していない魚類が検出されている地域と本来生息している地域を近い関係で表示させることは難しいため本研究では使用しない。

UMAP の可視化結果は図 2 の右上となっている。UMAP では局所的な位置関係を可視化することが可能でありなが

ら大域的な構造を残すことが可能である。またパラメータでの調整が容易であることから本研究では UMAP を使用する。

本研究では `n_neighbors` のパラメータを変更している。このパラメータは `umap-learn` というライブラリ内のパラメータを使用している。`n_neighbors` は近傍に関するパラメータとなっている。値を小さくすると点のまとまりが小さくなり、大きくすることでばらけたような表示となる。初期値は使用するライブラリによって異なる。本研究で使用しているライブラリでの初期値は 15 となっているが図 2 の右下のパラメータ変更後の UMAP では 12 に変更している。

本研究の実装システムでは図 1(A) の散布図でデータを選択することにより、(B) のマップと (C) の積み上げ棒グラフに選択されたデータの詳細なデータが表示される。また (B) のマップはマウスホイールにより拡大縮小が可能となっており、(C) の積み上げ棒グラフは範囲選択をすることで拡大が可能である。

(A) の散布図の色は冬が青色、夏が赤色で表示している。(A) の散布図で選択されたデータの詳細なデータを表示している (B) のマップと (C) の積み上げ棒グラフでは採水された地点を元に色を割り当てている (図 1)。

## 4.2 採水された地点の可視化

採水された地域の検出された魚類の相対的な関係を可視化するにあたり、採水された地点を理解する必要がある。そのため散布図で選択されたデータ内にある、採水された地点の緯度経度のデータを使用しマップ上に表示させている。マップには散布図で選択されたデータの緯度経度によって採水された地点を表示し、色を付けることで右上にある名前がどこを指しているのかを表示している (図 3)。

## 4.3 検出された魚類の名前と数の可視化

どの地点にどのような魚類が生息しているかを理解する為に、散布図で選択された類似度のデータの地点で検出された魚類を表示させる必要がある。そのため積み上げ棒グラフを使用し表示させている。積み上げ棒グラフの X 軸で検出された数を表示させ、Y 軸が和名を表示させており、検出された数が選択した採水された地域で積み上げられるようになっている。図 1 の (B) と (C) の採水された地点の色は同じである。検出された魚類の数は採水された地点によって異なり大きな差が生まれ、大きな値を表示するために小さな値が確認できなくなり、本来あるはずのデータが確認することができなくなる (図 4 右)。本来あるはずのデータを確認するために、検出された魚類の数を対数でも表示可能にした (図 4 左)。

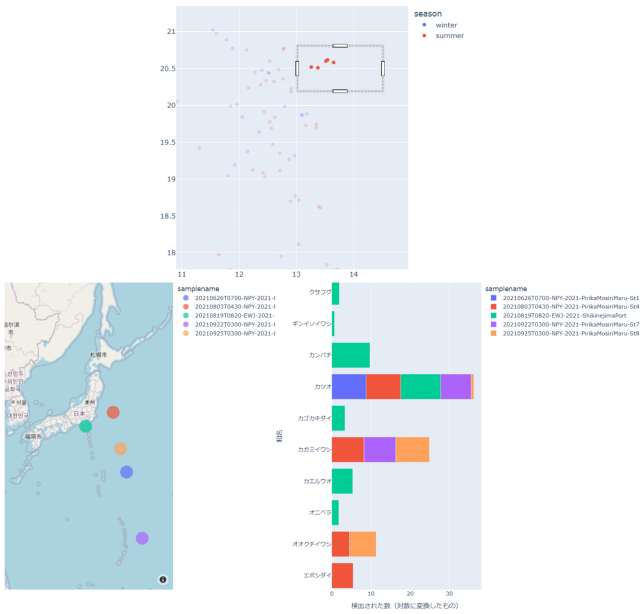


図 5 探索事例 1(右上は拡大した散布図)

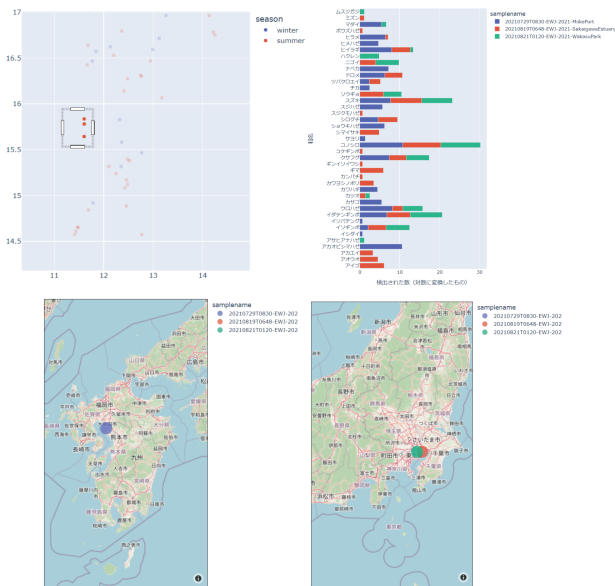


図 6 探索事例 2

## 5. 探索事例

探索事例 1 では散布図の中心的な塊の右上に表示されている小さな塊を選択した。ここではカツオが多く地点で検出されている(図 5)。しかしカガミイワシは緑色と青色の地点では検出されていない。これを踏まえ散布図の位置関係を確認すると、カツオ以外で同じ魚類が検出されていない黄色と緑色の地点は散布図で選択した中でもっとも離れている。そのため類似度による生息している魚類の相違を見ることは可能だと考えられる。

探索事例 2 では特徴的な魚類や類似性は見つけられないが、散布図では近く表示され、地図上では同じ内海である

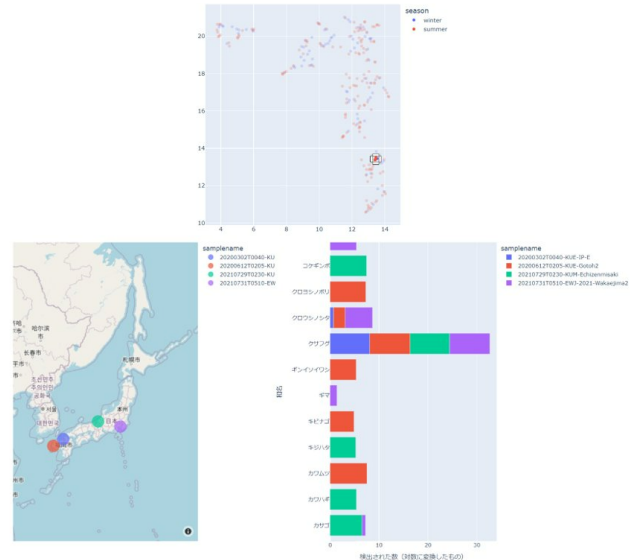


図 7 探索事例 3

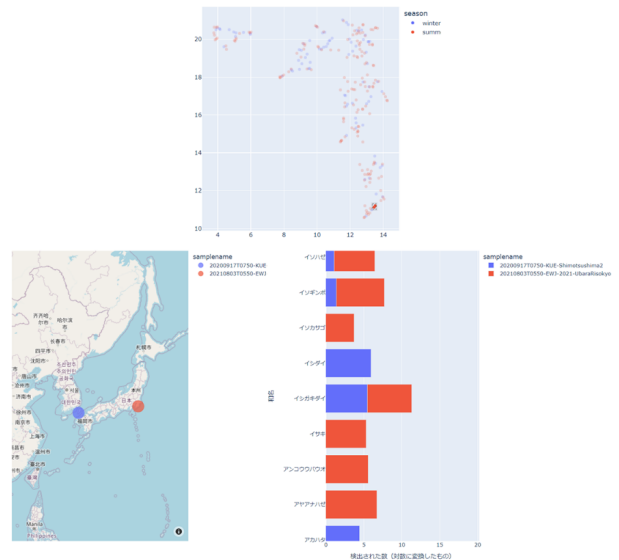


図 8 探索事例 4

ことがわかる(図 6)。そのためなんらかの特徴があると考えられるがクサフグやギンポ系などの多くの場所で確認できる魚類が多いため、確認するのが困難である。

探索事例 3 ではオレンジ色の地点では「ギマ」という魚が見受けられるが静岡県以南や熱帯海域に生息している魚とされている(図 7)。しかしこの地点は静岡県と隣接した場所にあるため、時期や海流によっては観測される可能性があるため一概に温暖化の影響とすることは難しい。

探索事例 4 では青色の地点でも赤色の地点でもイシガキダイが検出されている(図 8)。イシガキダイは本州中部以南や熱帯に生息しているとされているが、無効分散<sup>\*1</sup>として観測されることがまれにあるため生息地域の変化の裏付けにすることは難しい。

\*1 海流などの影響により本来の生息地域から離れてしまい、季節や環境の変化によって死滅し定着できないこと。

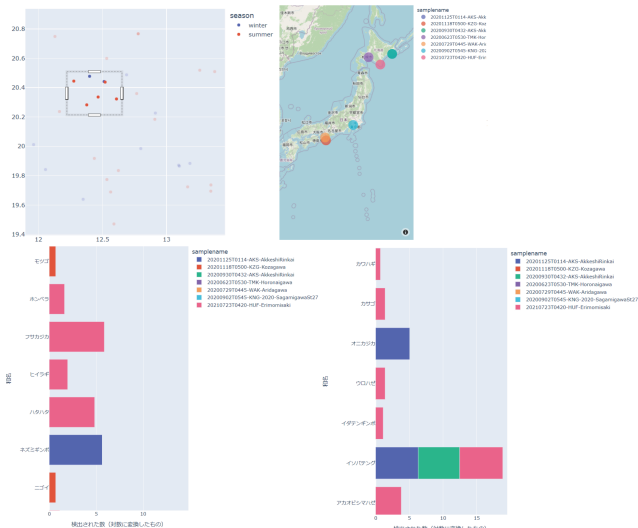


図 9 探索事例 5

探索事例 5 では襟裳岬にてヒイラギとウロハゼが検出されている (図 9)。ヒイラギは南日本に生息しているとされる魚であり、ウロハゼは茨城県や新潟県より南に生息している魚となっている。今回の結果では定着しているのか、何らかの間違いによりデータとして残っているのかわからないため生息地域の変化と言い切れない。またこの種類は別の場所で検出されており、そのため近い位置に表示されているわけではないので単なる偶然で見つけられたものだと考えられる。

探索事例 3, 5 からわかるように本来生息していないはずの魚類が検出されたことが確認できたのはその魚類が別の場所で検出されていないことから偶然見つけられたものであることがわかる。また探索事例 4 のイシガキダイのように無効分散か実際に定着しているか定かではないものも存在するため、信憑性に関し生息地域の変化と断定することはできない。

## 6. おわりに

本研究では専門的な知識を必要とせずに魚類の生息地域が変化していることを理解できるものの作成を目的としていた。しかし今回の結果では生息地域の変化を理解することができるかは結論付けられない。しかし今回の手法である検出された魚類で類似度を計算し相対的な位置関係としてプロットする方法により、本来生息していない魚類を見つけることが可能という結果は探索事例 3 の結果から得ることができた。

今後の課題としてはデータを年代別に用意し、そこに長期にわたり検出されているかを調べることができるように変更する必要がある。これにより探索事例 3 のような無効分散の成否を判別することができない場合にも長期にわたり検出されていれば生息地域の変化として表すことができる。

また今回の使用したデータには川や湖、池のデータが存在しているが今回は主に海に生息している魚類の変化を見ることが目的のため必要なデータではない。そのため、海以外のデータの削除等が必要だと考えられる。

さらに、クサフグなどの多くの地点で検出されているような魚類は、TF-IDF と呼ばれる単語の重要性を計算する手法を参考に、多くの地点で検出されている場合は重要性を下げ、局所的な地域でのみ検出される魚の重要性を上げるという対策が必要だと考えられる。

## 参考文献

- [1] ANEMONE DB, <https://db.anemone.bio/>, 参照 Dec. 15, 2023.
- [2] 海老原健吾, 安川雅紀, 永井美穂子, 喜連川優, 鷺谷いづみ: 市民科学による東京のチョウと植物の共生ネットワークモニタリングの可能性, 保全生態学研究 25 : 177-190 (2020), pp 177-190(2020).
- [3] 本村浩之: 日本産魚類全種目録, これまでに記録された日本産魚類全種の現在の標準和名と学名, ver. 22, 2023.