

生成AI時代を見据えた海象学習モデルの構築

堆積泥のC, N組成による生物生息場の推定 -ポアソン分布の学習による環境場の評価-

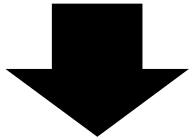
参考文献：土居田祐希, 井川広之, 岡田知也, 日比野忠史：堆積泥のC, N組成による生物生息環境の推定-ポアソン分布の学習による環境場の評価-, 土木学会論文集（2024）80巻17号. 論文ID：24-17240

研究開発の目的・意義

生成AIは港湾事業において今後向かえる「省人化の必要性」, 「DXの遅れ問題」を解決していくための新しい技術である. 生成AIの活用環境を整備することは, 日本の港湾行政を革新的に成長させるためのチャンスになる. 生成AIであるChatGPTは生まれたばかりの技術であり, 港湾分野においても先導的に活用することが望まれる. 本研究では構造物の設計外力への生成AIの適用法を検討している. 潮位, 波等の観測値（海象情報）を学習させた海象学習RAG（検索拡張生成）を構築して, 予測値による造成干潟, 堤防等の設計までの機械学習システム開発を目的とする. 機械学習に用いるアルゴリズムは人間の知能をコンピュータ上に再現するものであり, 構造物の設計外力を予測する海象RAGは, 全ての港湾事業（港湾DBを構成する多数の検索用データベース）への拡張を可能にできる.

環境(生物, 生息場)調査結果のAI利用

海域環境の変化に応じた浚渫泥
の活用ができる**干潟の造成**



干潟の造成候補地における
生物生息環境調査の活用法



造成後の生物生息環境から造成後の
生態系を計画することは**困難**

造成地の選定基準

- **生物環境適合性**
海象海域の環境を機械学習し, 選
定地での干潟造成の生態系を生成
- **安定性**
波, 流れによる地形変化を検討
- **社会的ニーズ**
漁業の活性化, 環境改善, 土砂の
利用 等

Ex. (課題)

- 地形変化によって**複雑に変化する生物生息環境**を生成
- 技術者のノウハウに環境調査結果(現況)を組み込んで**造成後の生物場を提示**

本研究の目的

背景

干潟候補地での生物生息環境特性の推定するための
機械学習法の確立



課題1

生物生息環境は地形変化に
よって複雑に変化
➡場を捉えた評価

課題2

測定値を学習して場を示す特性
を生成
➡生物と生息環境の特性を推定



本研究の目的

技術者の知見をモデル化した機械学習を利用した
生物生息環境の推定手法の提案

本研究の目的達成への簡易フロー

本研究の目的

C/N比の考え方とベイズ推論を利用した
造成適地の選定のための生物生息環境の推定手法の確立



手段

C/N比の考え方に基づいて各エリアでの生物の生成する底泥特性を分類し、
ベイズ推論（学習）によって対象エリアの生物生息環境を比較・予測

ベイズ推論は新しい調査（学習）により調査海域の特性を生成する

A～Fエリアにおいて分類C/N比の出現確率を数値化す



調査地点：燧灘北沿岸域31地点

(Aエリア：6地点，Bエリア：5地点，Cエリア：7地点，
Dエリア：4地点，Eエリア：4地点，Fエリア：5地点)

調査時期：2023年7月，8月

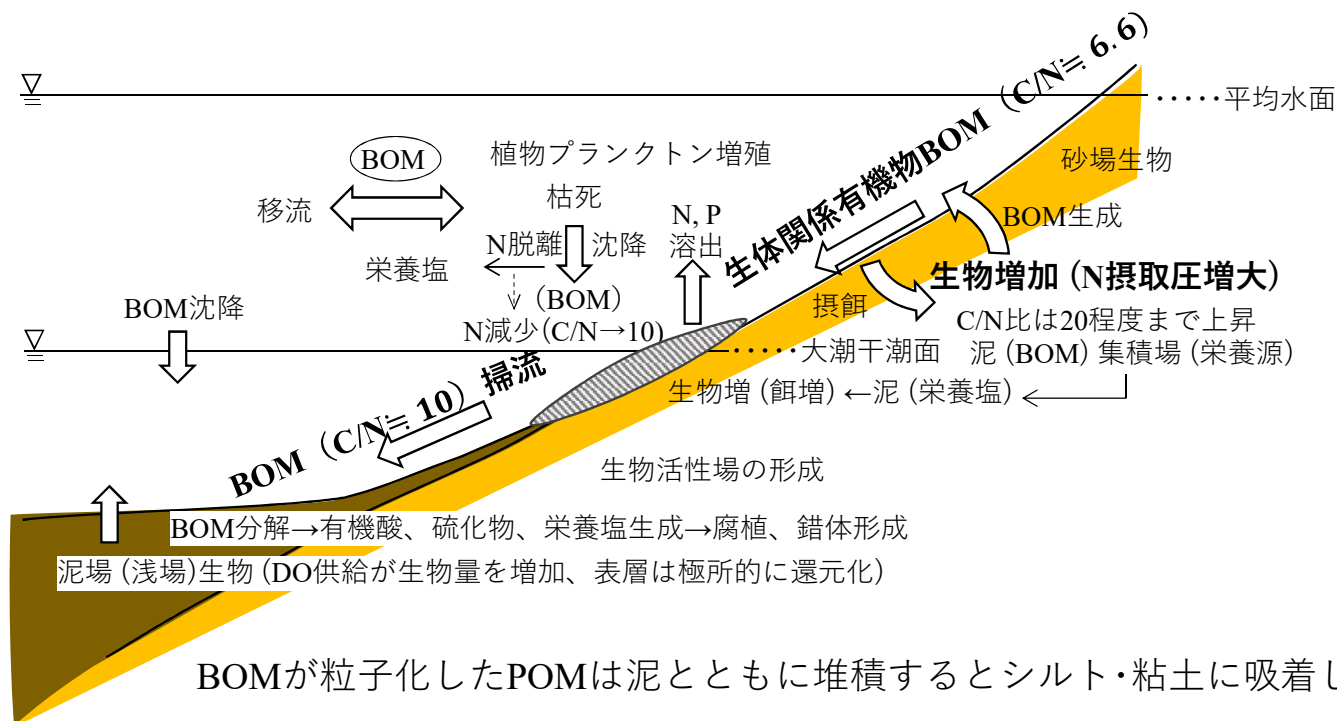
調査項目：生物、粒度組成，TOC，TN，IN，TP，IP，COD，硫化物等

生物活性の高い干潟等の浅場における有機物循環（BOMの生成、消費～堆積）

底泥（泥場）へのDO供給（生物生息）の有無により底質が大きく異なる

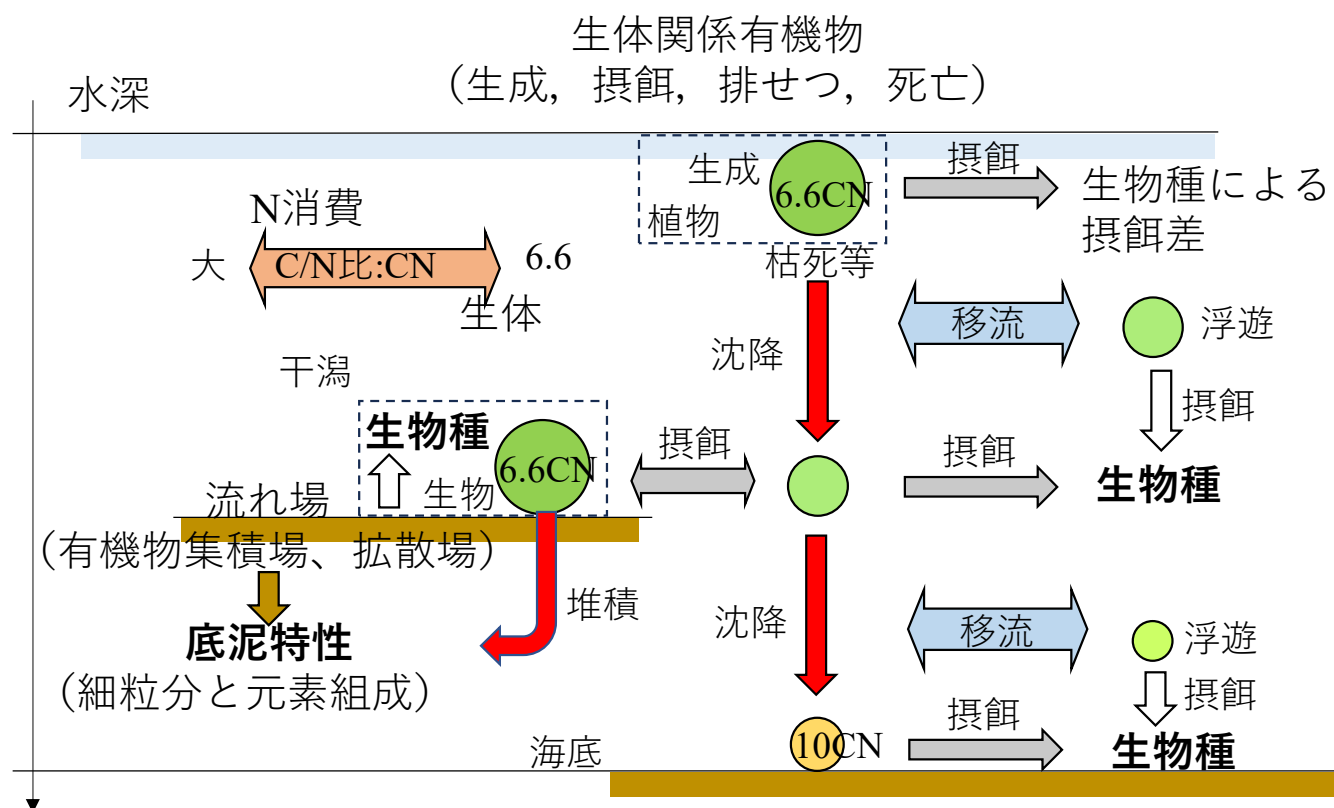
生体関係有機物 BOM：生物活動（死亡、排泄等）により生物により生成される易分解性の有機物→生物の餌となる

BOMのC/N比は平均的に6.6（レッドフィールド比；生物が生成する有機物の元素比）



生物生息環境でのC/N比の変動

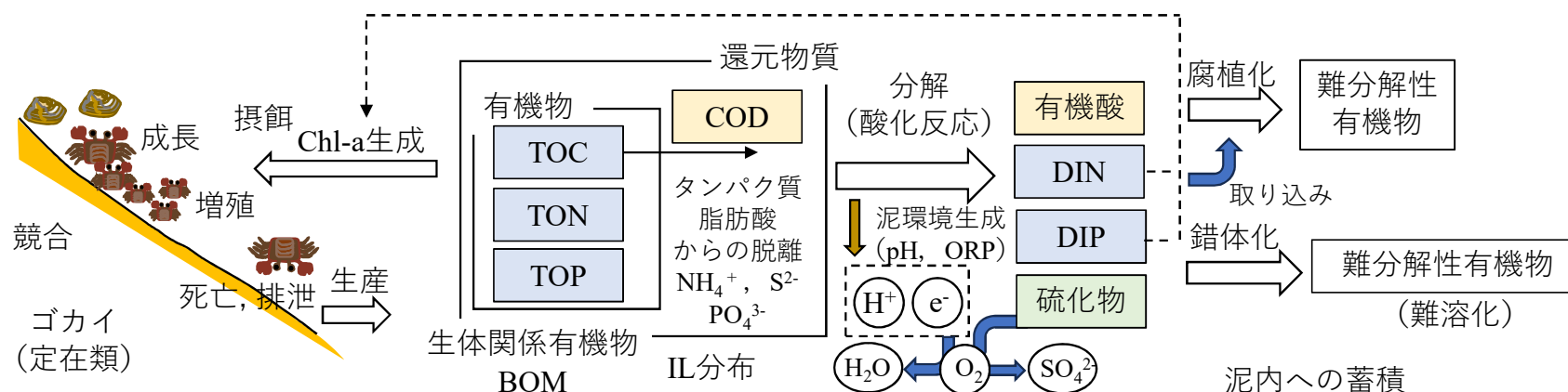
BOMの有機物組成変化から場の特性を推定する



- 1) 田口哲：レッドフィールド比：研究の歴史と現状，今後の展望，海の研究
- 2) 日比野忠史ら：有機泥の分解を考慮した沈降と堆積過程のモデル化

モニタリング項目とBOM循環の関係

底生動物が生成した有機物BOMの生物利用と泥内（浚渫泥）への蓄積



有機泥層で起きる酸化還元反応（電子 e^- が利用される反応）

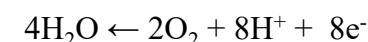
有機酸（例えば酢酸）生成反応・・・還元反応



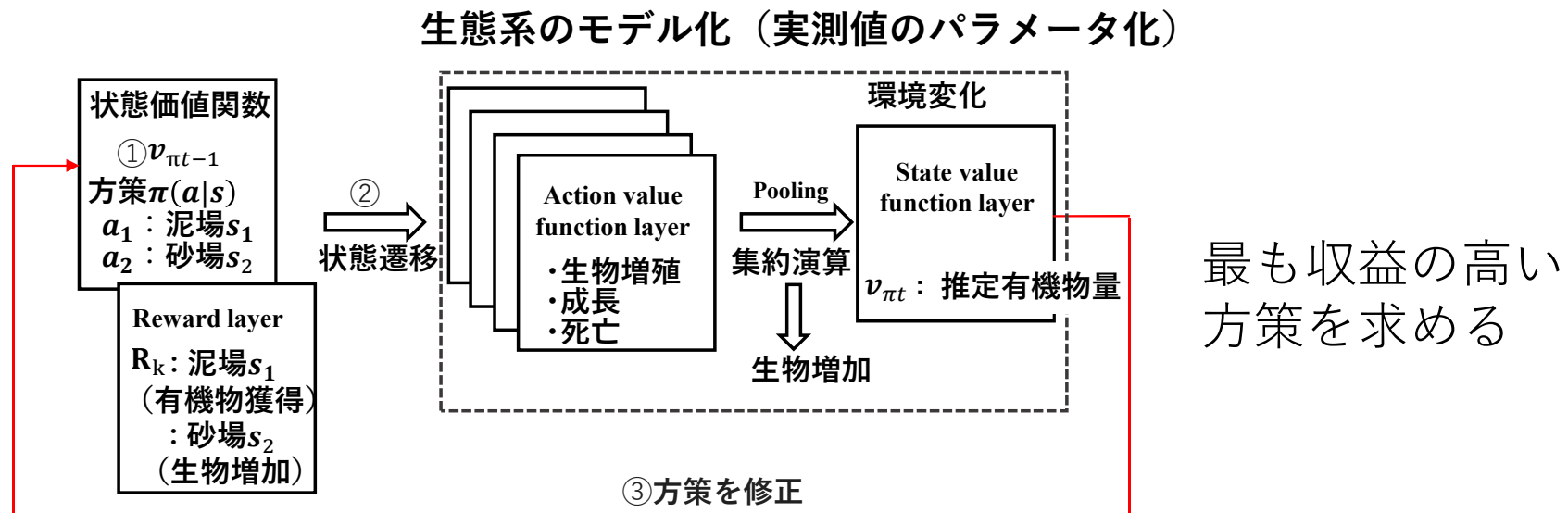
①硫化物である硫化水素（ e^- が留まると生成）の酸化反応



② O_2 による酸化反応



e^- と O_2 が反応するので海底で貧酸素化



強化学習で用いられるベルマン方程式は「状態 s の価値関数 $v_{\pi}(s')$ 」と「その次に取り得る状態 s' の価値関数」との関連性を表しており、時刻 t 現在での状態において得られる将来の収益(G_t)と $t+1$ での状態に遷移した時に得られる報酬(R_{t+1})との関係を表す方程式である。

強化学習(干潟ゲーム)

砂場 s 砂場に存在できる細粒分数 N
40個 (40%) 以上で生息阻害
死亡率+ 0.01/%

餌OM×3個/1匹がなくな
ると生物死亡

OM×3 (個/匹) 生成



生成されたOM
は砂場が保持

餌
OM×3個/匹

トータル 生成OM×1個/匹 (排せつ)

OM×-2個/匹

摂取OMは餌と排せつの差
(排せつは直接考えない)

2ステップ目から餌に変化
流れの強さによって運ばれる細粒
分の数 (0~4個) が決まる

同時に砂場にいる細粒分は N 個
 $N \geq 30$ で泥場への掃流確率上昇

↑ OM×60以上の
残存が条件

増殖時に餌
OM×30が必要

行動 y

砂場移動

泥場滞在

砂場移動 a_1

1

流れ (エージェント)

泥場へ (OM)

泥場 s

報酬

OM×1個/1ステップ

OM1個は0.02mg保持

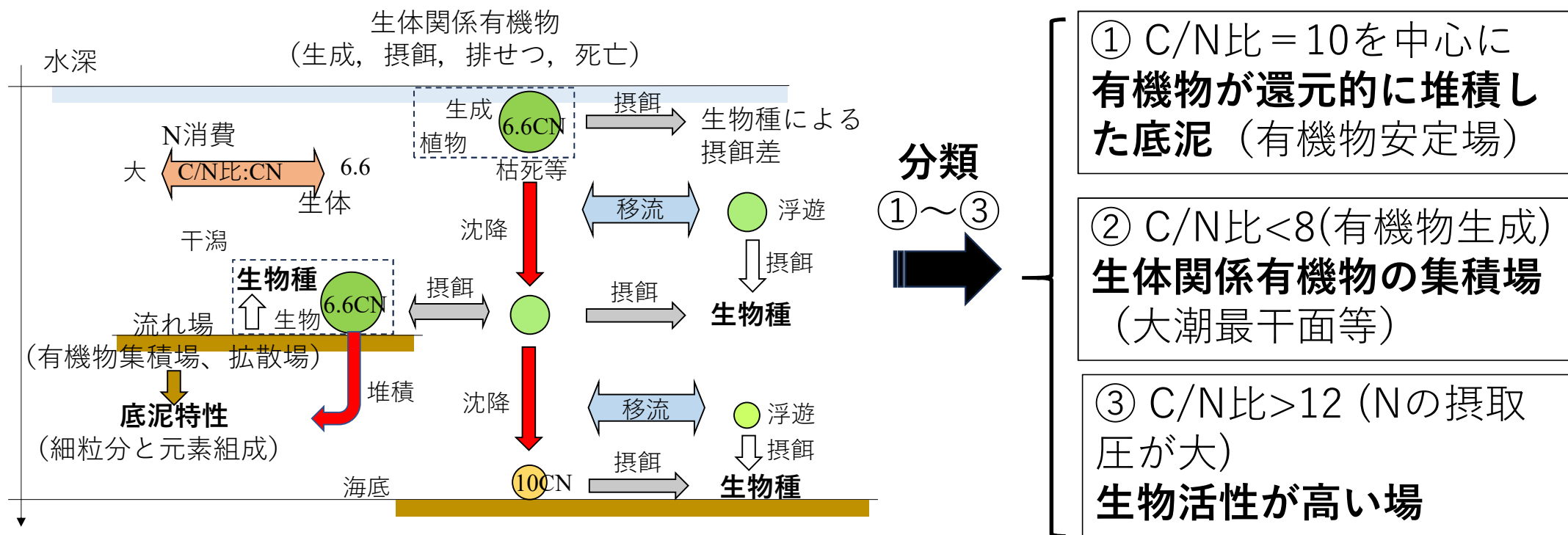
可能携带有機物数
 $L = 3$

泥場滞在 a_2

パラメータの決定

メナシピンノ成体(0.25g/個体) 1匹= 4個体 =1g...成体行動単位
OC (OM) 1個=0.02mg, 3個 (0.06mg) でメナシピンノ1匹が生息可能
増殖 未成体30個体, 15個を基準
未成体500個体重量 = 0.03 g/個体 × 500個体 = 15g (成体15匹相当)
成長1匹 (1g) を基準

C/N比から生物場を生成(ベイズ推論による機械学習)



C/N比の考え方に基づいて生物生息環境を分類



ベイズ推論により場 (エリア) の特性を生成する

ベイズ推論による学習と予測

モデルの基本形

学習モデル $p(D, \theta) = p(D|\theta)p(\theta)$ $p_{\theta D}(D, \theta) = p_{D|\theta}(D|\theta)p_{\theta}(\theta)$

とも書く

パラメータ θ の値を観測データ D （尤度関数）から学習して求めたい事象の精度を上げる

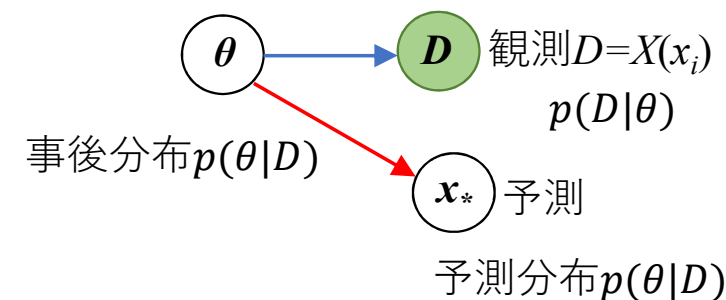
予測モデルも同じ形式

尤度関数 $p(D|\theta)$ と事前分布 $p(\theta)$ の積が新しい確立密度関数 $p(\theta|D)$ （事後分布、予測分布）になる

乗法定理・線形識別器は $\ln$ (確率密度関数)

赤白玉問題では「赤の出る確率 $p(\theta, D)$ 」は「取り出した箱での赤玉の出る確立 $p(\theta)$ 」と「取り出した箱に何個の赤玉が入っているか $p(\theta|D)$ 」の積で決まる。箱の中身 D を学習すれば赤の出る確率を精度よく見積ることができる。

事前分布 $p(\theta)$



乗法定理 $p_{D|\theta}(D|\theta)p_{\theta}(\theta)$ を使って精度を上げた事後分布 $p_{\theta D}(D, \theta)$ を算出

乗法定理 2種類の確率変数 ($i = 1, 2$)

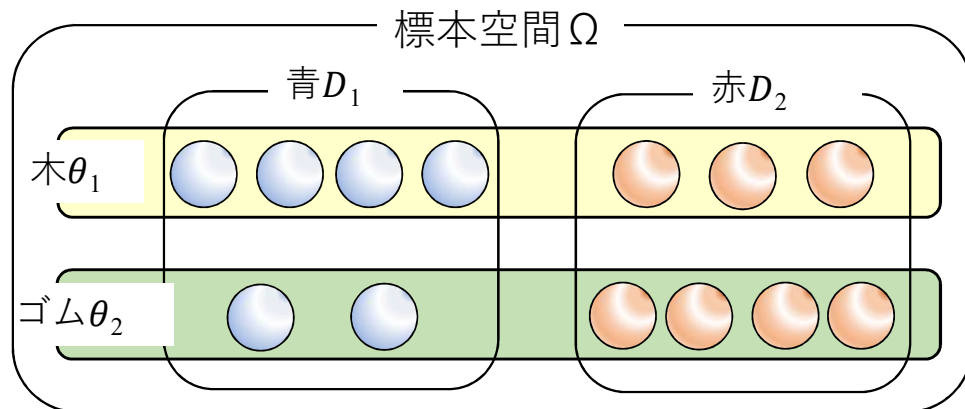


図 箱からボールを取り出す試行の標本空間 Ω のベン図

木と青が同時に起こる確率（同時確率）

$$p_{\theta D}(\theta = \text{木}, D = \text{青}) = p_{\theta D}(\text{木}, \text{青}) = \frac{N(\text{木} \cap \text{青})}{N(\Omega)} = \frac{4}{13}$$

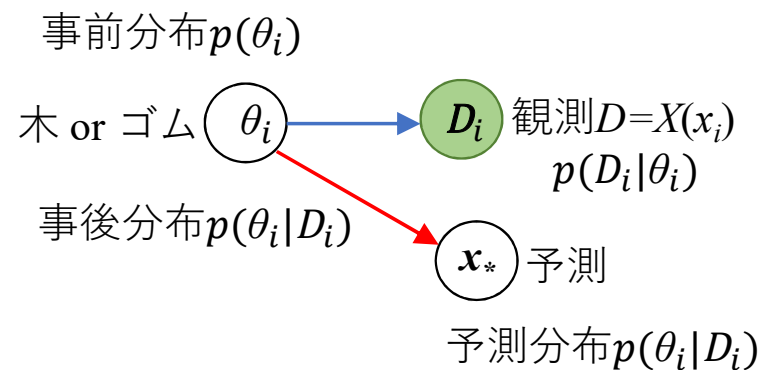
取り出すボールが木で青

木であるボール（条件）で、かつ青（条件付確率）

$$p_{D|\theta}(D = \text{青} | \theta = \text{木}) = p_{D|\theta}(\text{青} | \text{木}) = \frac{N(\text{木} \cap \text{青})}{N(\text{木})} = \frac{4}{7}$$

周辺化（周辺確率）：同時確率 $p_{\theta D}$ がわかっている状況で x の確率を求める方法

$$p_D(\text{青}) = \sum_{\theta \in \{\text{木}, \text{ゴム}\}} p_{\theta D}(\text{木}, \text{青}) + p_{\theta D}(\text{ゴム}, \text{青}) = \frac{4}{13} + \frac{2}{13}$$



乗法定理とベイズの定理

同時確率 $p_{\theta D}$ の定義、条件付き確率 $p_{D|\theta}$ と周辺確率 p_{θ} の積

$$p_{\theta D}(\text{木}, \text{青}) = \frac{N(\text{木} \cap \text{青})}{N(\Omega)} = \frac{N(\text{木} \cap \text{青})}{N(\text{木})} \frac{N(\text{木})}{N(\Omega)} = p_{D|\theta}(\text{青}|\text{木})p_{\theta}(\text{木}) = \frac{4}{7} \times \frac{7}{13}$$

乗法定理・・・同時確率、条件付き確率、周辺確率間の関係

$$\begin{aligned} p_{\theta D}(\theta, D) &= p_{D|\theta}(D|\theta)p_{\theta}(\theta) \\ &= p_{\theta|D}(\theta|D)p_D(D) \end{aligned}$$

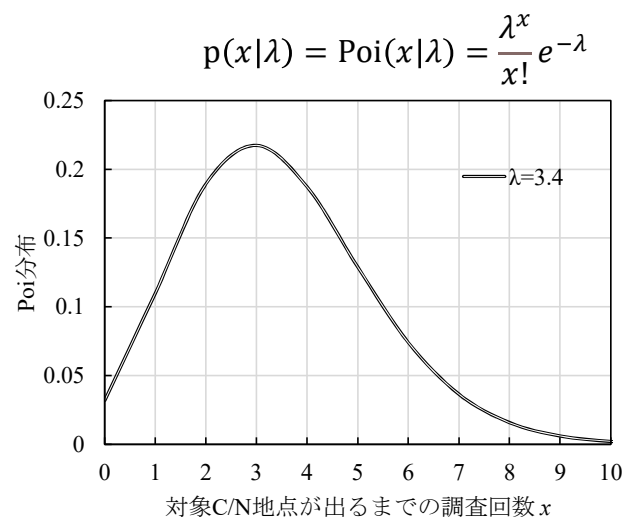
ベイズの定理

$$p_{\theta|D}(\theta|D) = \frac{p_{\theta D}(\theta, D)}{p_D(D)} = \frac{p_{D|\theta}(D|\theta)p_{\theta}(\theta)}{p_D(D)}$$

$$p_{\theta|D}(\text{木}, \text{青}) = p_{\theta D}(\text{木}, \text{青})/p_D(\text{青}) = p_{D|\theta}(\text{青}|\text{木})p_{\theta}(\text{木})/p_D(\text{青}) = \frac{4}{7} \times \frac{7}{13} \times \frac{13}{6} = \frac{4}{6}$$

生態系と地盤環境を結びつけるC/N比の利用(ノウハウの定式化)

ある期間内にランダムに発生する事象の回数とその確率の関係を表す分布
C/N比は生物活動と流れにより形成される場での有機物の生成履歴を表す指標である
C/N比はランダムに決定される事象であり、C/N比の出現にポアソン分布を仮定できれば①～③に分類されたC/N比が出現する調査地点により期待値 λ が求められる



ポアソンPoi分布はC/N比 = 10となる地点が λ の期待値 = 3.4 (= 31地点/9地点) 回目に出現する確率を表す

ピークが3.4からずれるのはPoi分布の特性

- ・ a : 全観測点数
- ・ b : 対象C/N比の出現数
- ・ $\frac{a}{b}$: λ の期待値 (出現する地点数)

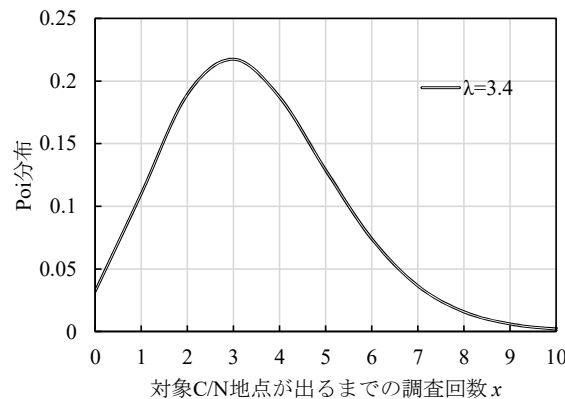
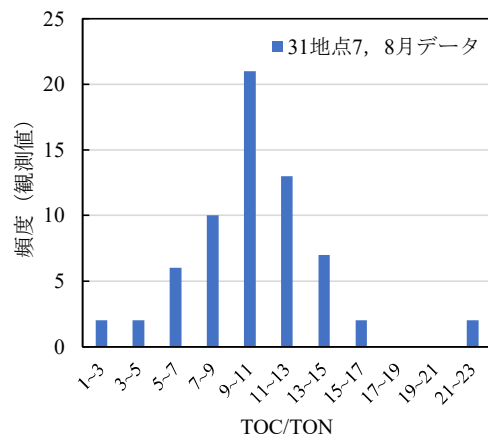
ポアソン分布を仮定したC/N比出現確率分布

環境調査を行った海域において全ての調査地点の内、対象のC/N比が現れる地点数

の期待値を λ として現れた回数を x として**ポアソン分布**を決定する

本解析では対象となる地点数を x ，必要調査地点数（期待値）を $\lambda = a/b = \text{全調査地点数} / \text{対象出現地点数}$ として学習データ（ポアソン分布）を作成した

本解析ではA～Fエリアの31地点で調査された結果から対象のC/N比が全調査地点の内、対象C/N比が出現した地点数から期待値 λ を求めた



頻度分布は $9 < C/N < 11$ ($C/N = 10$)の出現する回数を示しているが、ポアソン分布は C/N 比 = 10の出方を表している

C/N 比の出現の必要地点類 λ を変化することで対象エリアの生態特性が比較できる

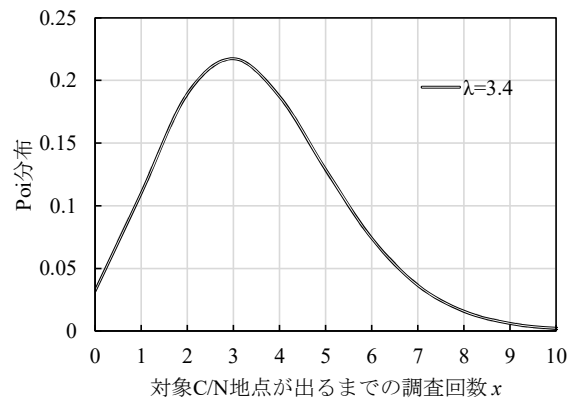
ガンマ分布

$$p(\lambda) = \text{Gam}(\lambda_i | a, b_i) = C_G(a, b_i) \lambda_i^{a-1} e^{-b_i \lambda}$$

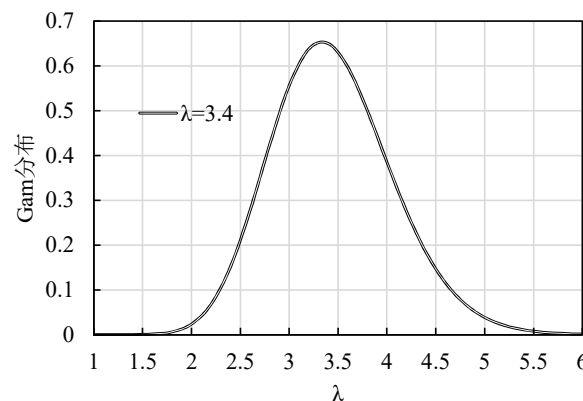
ポアソン分布の共役事前分布であるガンマ分布は「任意の単位時間や単位面積 λ の出現が1回起きることが期待できる出来事が実際に起きるまでの時間分布や面積分布」を表す確率分布

本解析では、ガンマ分布のパラメータ λ_i の期待値が a/b_i である対象C/N比地点の出現率を表す確率密度関数である。本ベイズ推論では事前分布（ガンマ分布）のパラメータを全エリアでの全観測データとすることにより事前分布は個々のエリアの特性を表す基準値（平均値）として用いられる

ポアソン分布



ガンマ分布



ポアソン分布は λ である事象が発生するまでの回数と確率、ガンマ分布は λ が発生する確率を示している

- a : 全観測点数
- b_i : 対象C/N比①～③の出現数
- $\frac{a}{b_i}$: λ_i の期待値（①～③が出現する地点数）

ベイズ推論を用いた各エリアの生物生息環境の評価

ベイズ推論

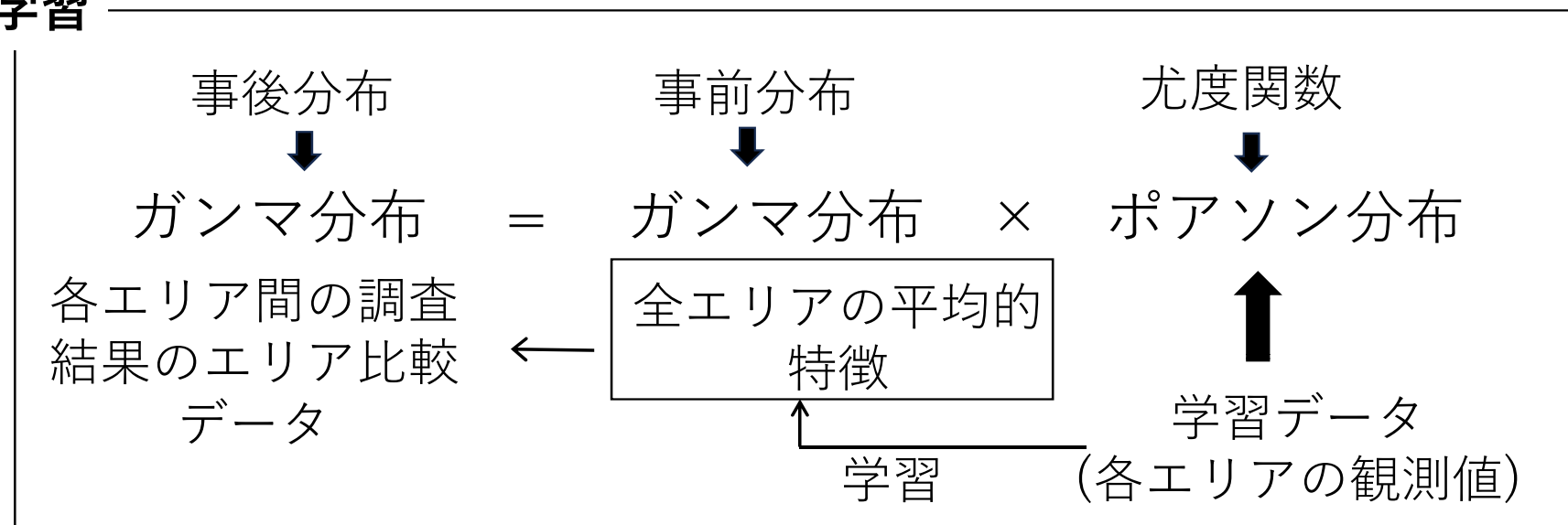
$$\begin{array}{ccccc} \text{事後分布} & = & \text{事前分布} & \times & \text{尤度関数} \\ \text{各エリアの} & & \text{全エリアの} & & \text{各エリアの} \\ \text{特性評価分布} & & \text{平均的特性} & & \text{特性} \end{array}$$

1. 各候補地（エリア）分布間の特性分布尤度関数を比較して各エリアでの生息環境の特徴を評価することが可能
各エリアの候補地としてのポテンシャルを比較
2. 調査結果を学習させることにより全エリアの平均的特性（分布）との差を調査の特異性を排除する

ポアソン分布を用いたベイズ推論について

尤度関数（学習データ）としてポアソン分布を採用・・・観測値がポアソン分布に従うことが条件、共役事前分布は**ガンマ分布**

学習

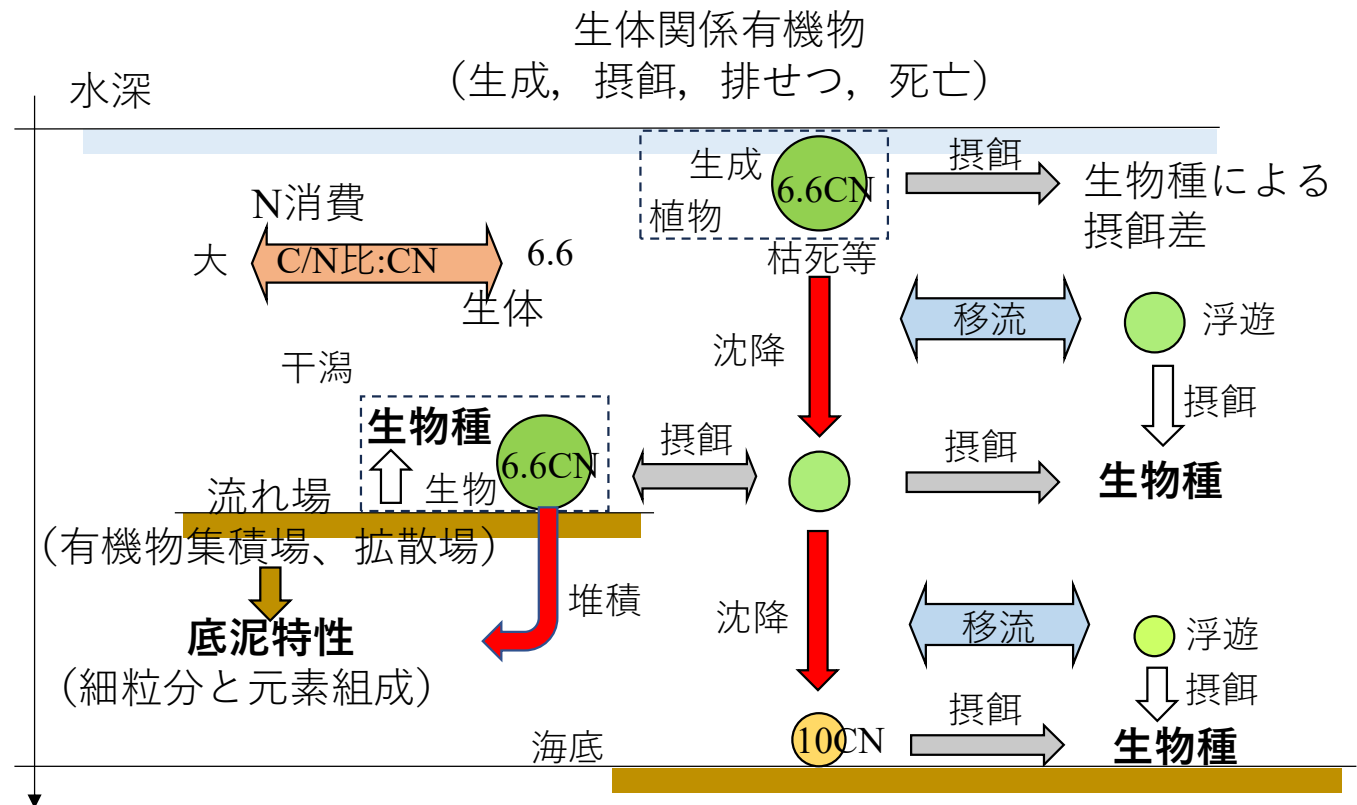
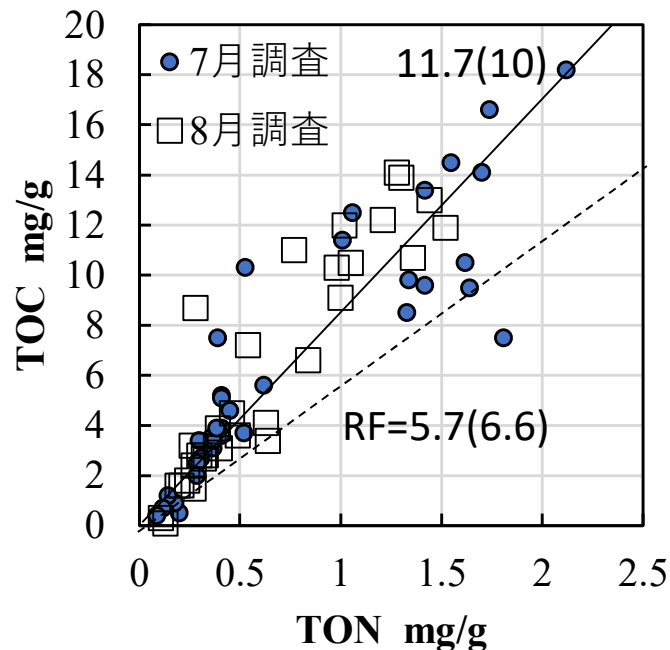


事前分布は調査データとともに増強させる

事前分布（調査データの特徴）に新しい対象海域データを学習させることにより対象海域の特性を持つ事後分布が生成される

生物活動に対応したC/N比の変化

全調査地点でのC/N比の散布図



※RF=レッドフィールド比：破線 C/N比=6.6 mol/mol (=5.7 g/g)、実線C/N比 = 10 mol/mol(= 11.7g/g)

全調査地点におけるC/N比の頻度分布

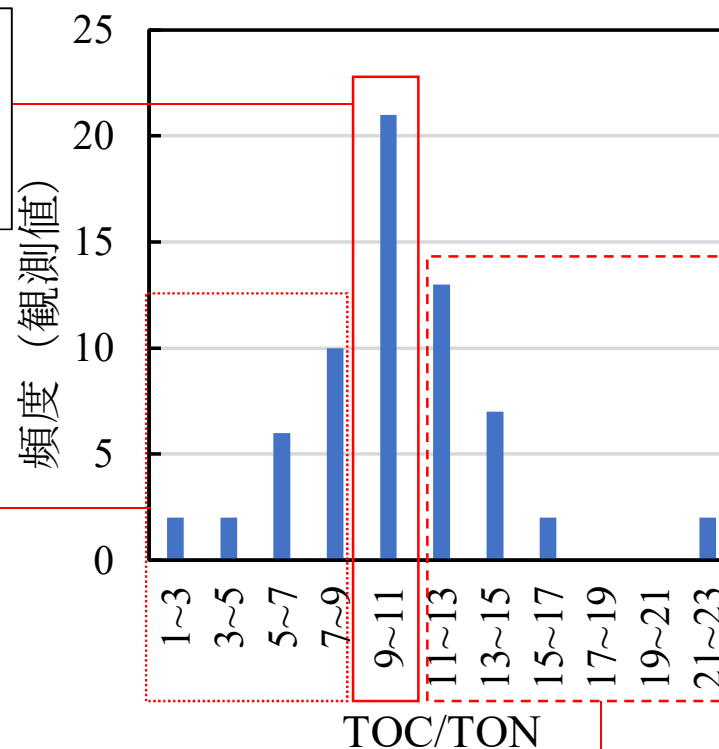
ベイズ推論の準備

対象C/N比区分①～③が
出現するまでの回数は
各エリアA～Fの特性を表し
ていると仮定

① 有機物が堆積した底泥生
物に変化の少ない安定場
(還元的な場では変化小)

② 生体関係有機物
(生物の死骸等新生有機
物)の集積場

③ Nの摂取圧が大きい
生物活性が高い場

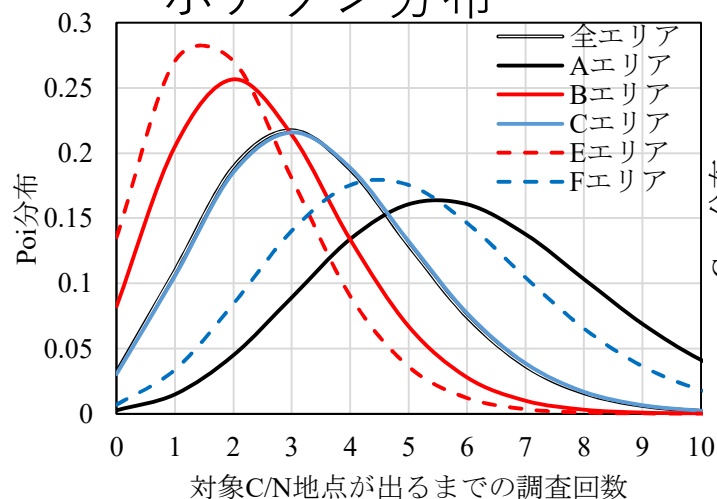


算出結果 区分① C/N比 = 10

① C/N比=10 の安定場（生物量の変化量）

尤度関数（学習データ）

ポアソン分布

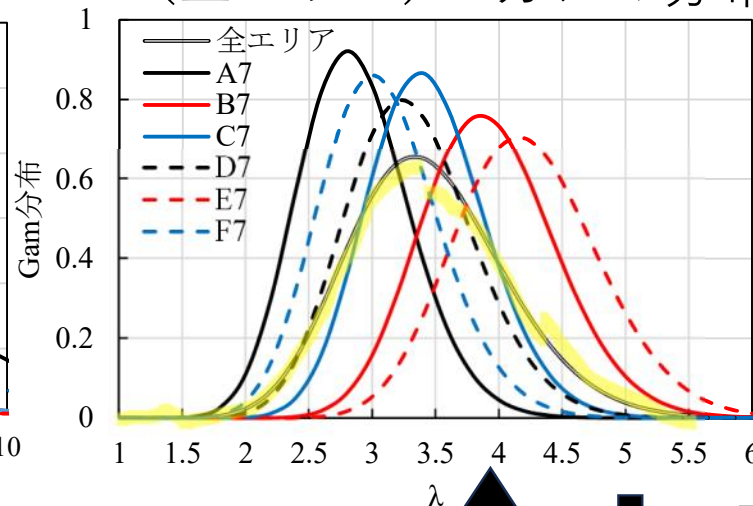


事前分布

（全エリア）

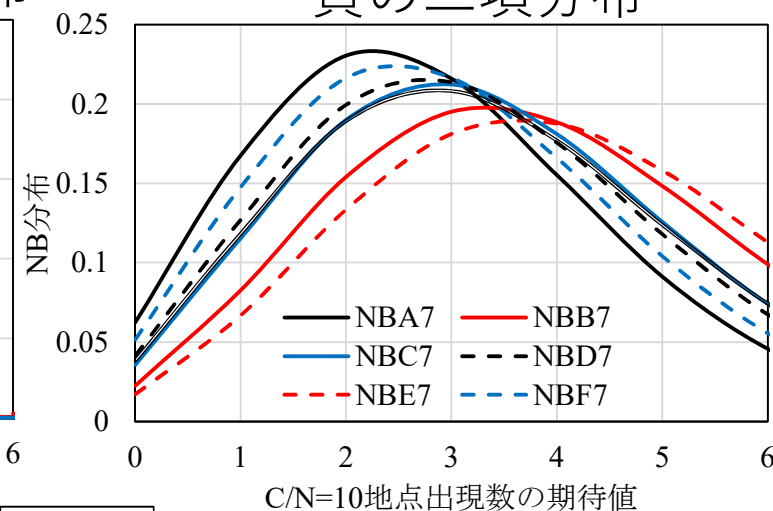
事後分布

ガンマ分布



予測分布

負の二項分布



A、FエリアはC/N比
（生物量）の変動が
大きい

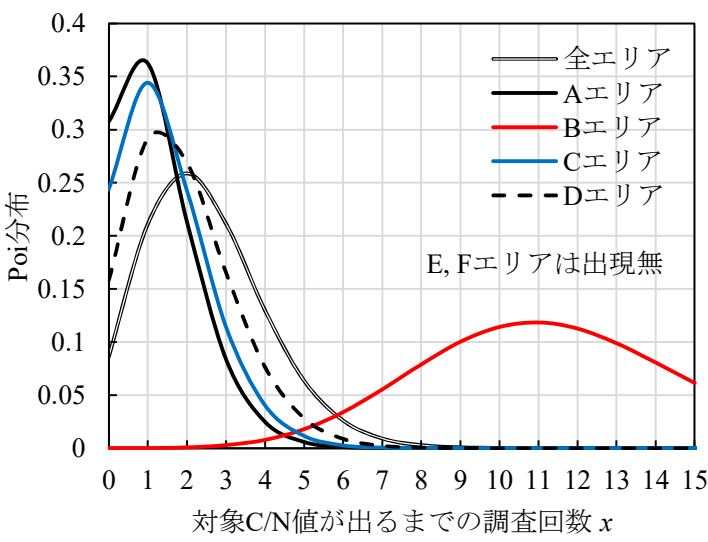
学習

予測

算出結果 区分② C/N比<8

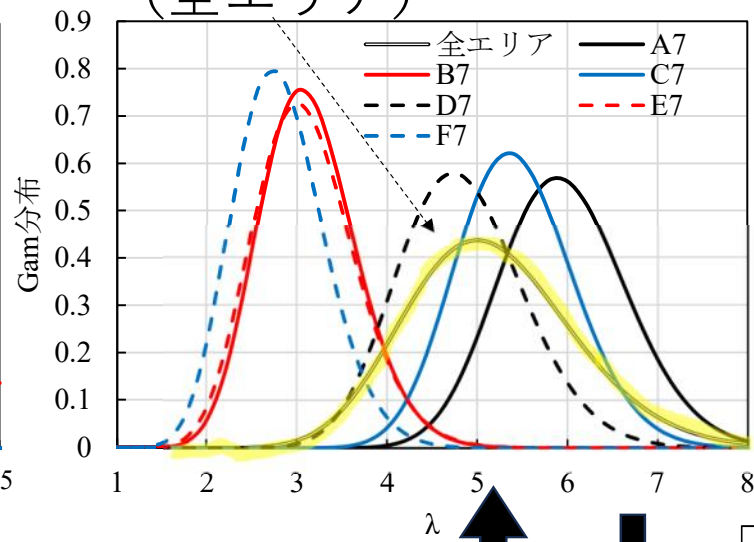
区分② C/N比<8の生物の死骸等の新生有機物の集積場

尤度関数



A, Cエリアは
新生有機物の
堆積が多い

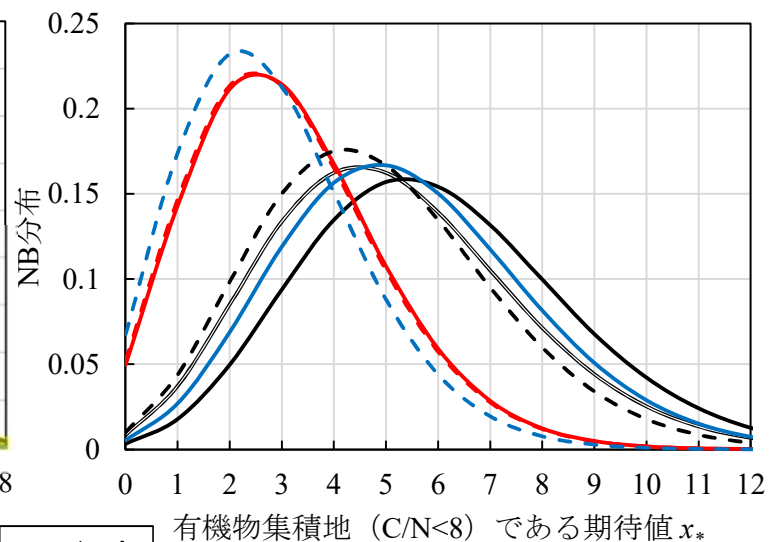
事前分布
(全エリア)



学習

事後分布

予測分布

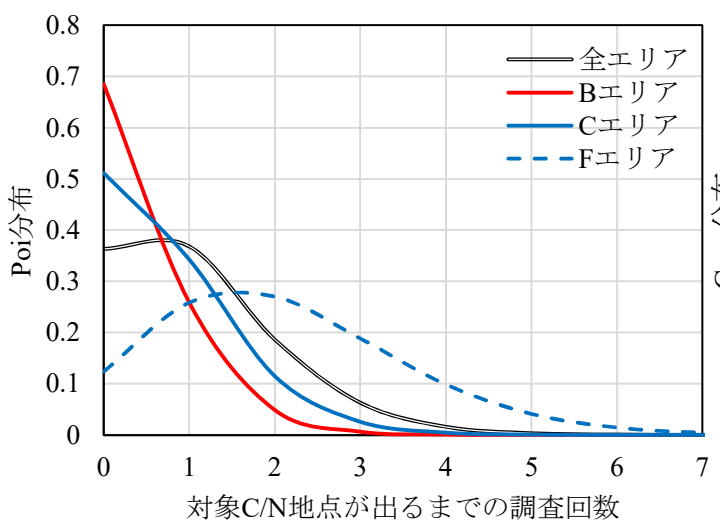


予測

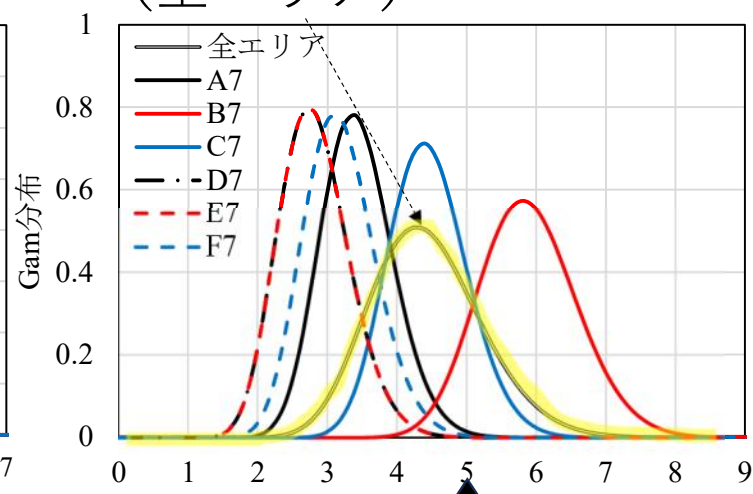
算出結果 区分③ C/N比>12

区分③ C/N比>12の生物活性が高い場

尤度関数

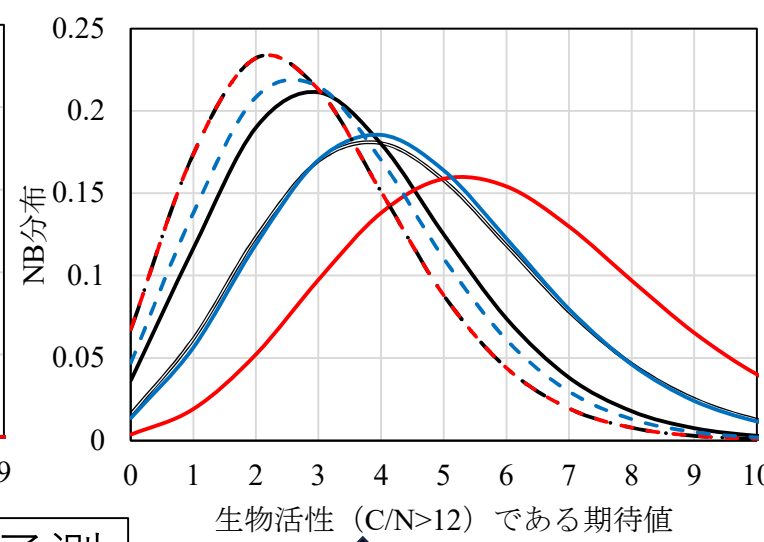


事前分布
(全エリア)



事後分布

予測分布



Bエリアは
生物活性が高い

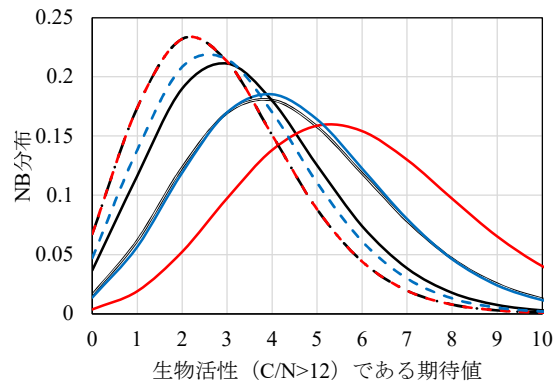
学習

予測

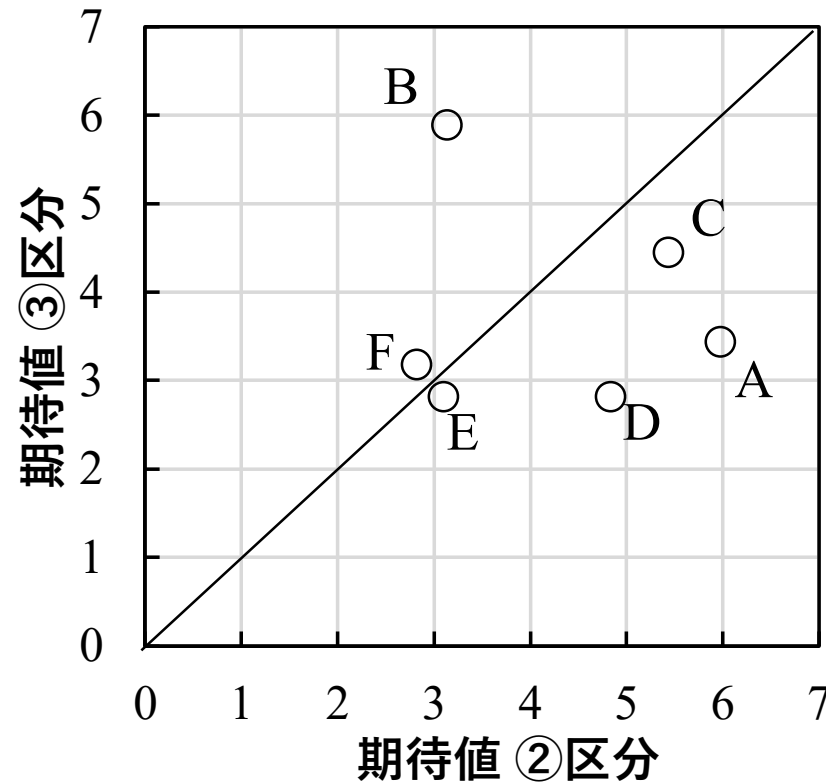
予測結果(負の二項分布)の比較

生物活性が高い場

③C/N比>12 の予測分布



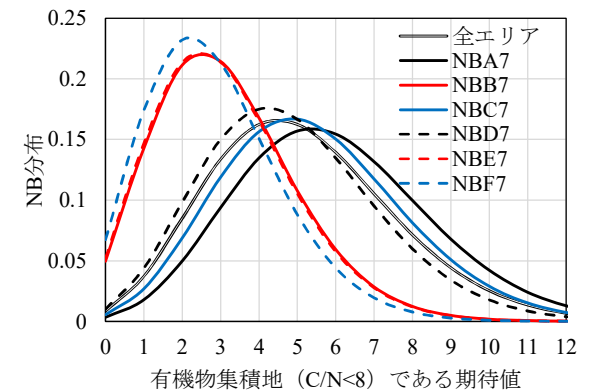
C/N比 > 12の
出現数の期待値



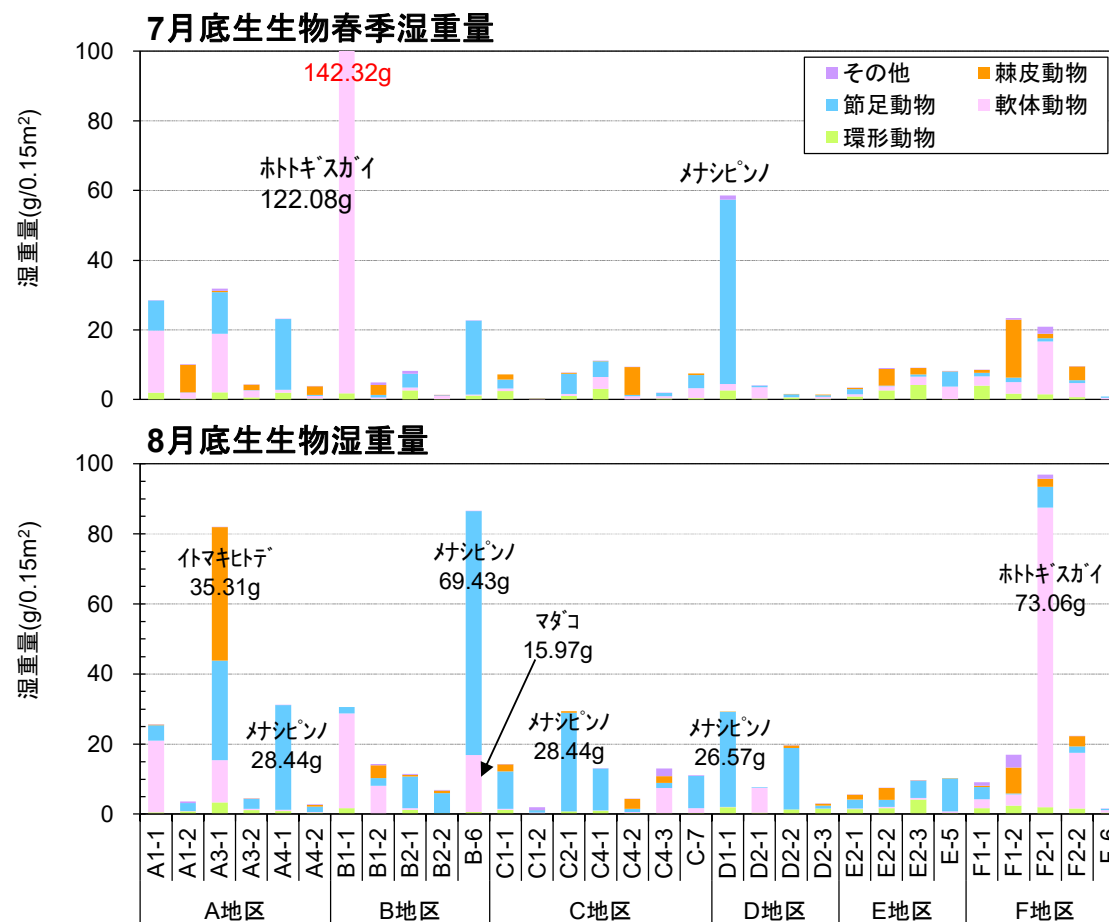
C/N比 < 8の
出現数の期待値

新生有機物の集積場

②C/N比 < 8 の予測分布



候補地（各エリア）での底生生物湿重量



A地区（エリア）

有機物集積場



生物が成長し易い
環境

B地区

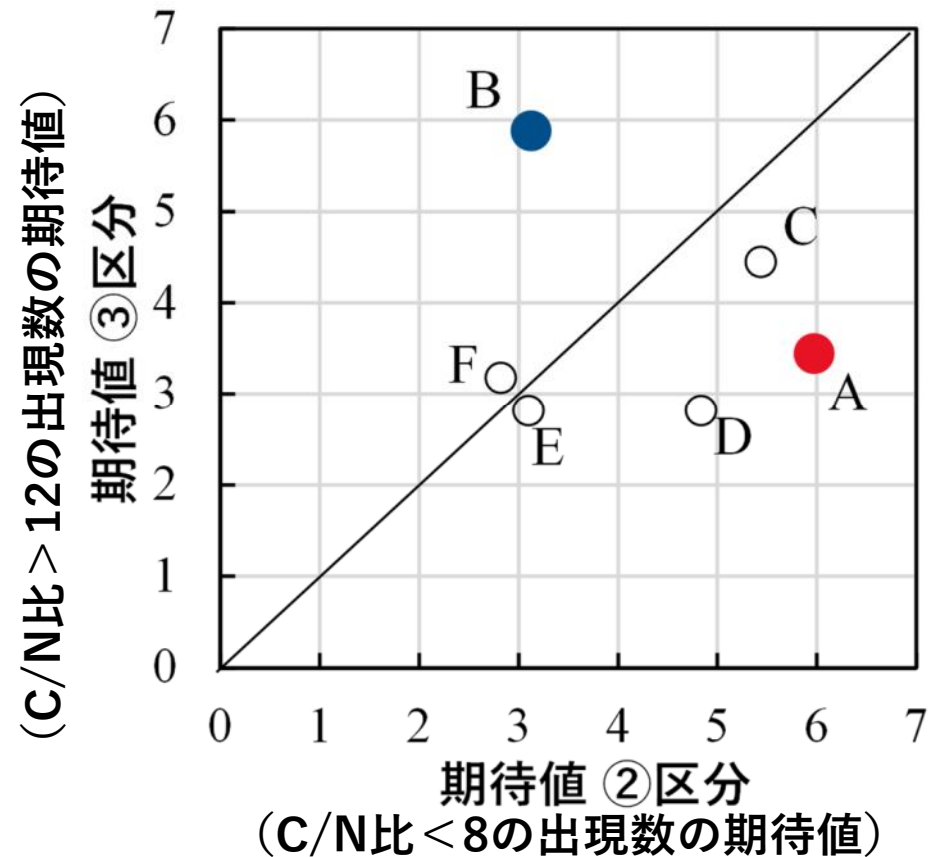
生物活性が高い



生物が増加し易い
環境

生物生息に適したエリア(候補地)の選定

- ②期待値 $A > B$
→ AはBに比べて新生有機物の集積が多く、生物活性が増大する可能性が高い
- ③期待値 $B > A$
→ BはAに比べてNの摂取圧が大きく、生物活性がある
- Cエリアは死新生有機物の集積地であると同時に、Nの摂取圧が大きい
→ 生物活性があり、これからも増大するポテンシャルが高い



まとめ

エリア間でC/N比変化に差があり、各エリアでのC/N比を説明変数として機械学習の一つであるベイズ推論を実行した。

1. ポアソン分布を尤度関数として、C/N比頻度と大きさの分布を学習させることにより、エリア毎のC/N比が示す生物生息場の特性を評価できた。
2. 学習データとしてC/N比頻度分布から算出されたC/N比の区分①～③の期待値をポアソン分布に与えることにより、各エリアの生物生息場の特性を学習できた。
3. 観測値を適切にパラメータ化して機械学習させることにより、観測値の特性を促えた予測分布が作成でき、対象エリアが持つ生物環境を客観的に表す期待値が得られた(候補地の選定ができた)。