

(一財) 港湾空港総合技術センター
令和 6 年度 研究開発助成 成果報告書

助成番号 : 令和 6 年 2 月 21 日付 第 18 — 1 号

研究開発項目 : ④港湾、海岸及び空港における技術の伝承・人材育成に関する研究に関するもの

生成 AI 時代を見据えた海象学習モデルの構築

令和 7 年 4 月 30 日

広島大学

日比野 忠史

目 次

1. はじめに	1
1.1 研究開発の目的・意義	1
1.2 研究開発の概要	1
1.3 研究開発の概要港湾領域 LM に取り込む内装 LM	2
2. 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-	2
2.1 海底堆積有機物の構成	2
2.2 方法	3
(1) 試料の採取と栄養塩分析	3
(2) C/N 比分布特性把握のためのベイズ推論	4
2.3 結果	4
(1) 各エリアでの有機物の堆積状況（調査結果）	4
(2) 観測（学習）値の頻度分布とポアソン分布	5
(3) ポアソン分布の学習と予測の方法	6
2.4 2章まとめ	9
3. 設計波高の未超過確率に関する考察-機械学習に用いる確率関数を考える-	10
3.1 未超過確率でのベイズ推定の定式化	10
(1) 高波浪の未超過確率とワイブル分布	10
3.2 最大高波出現予測の方法	12
3.3 ガウス分布を用いた機械学習の方法	12
(1) k, g をパラメータとしたワイブル分布	12
(2) 超過確率分布の学習モデルの構築	13
(3) 線形回帰モデルの設定	14
3.4 観測データの学習と更新された事後分布	16
(1) 新記録の発生確率	16
(2) 新記録の発生確率の学習による分布の変化	17
3.5 3章のまとめ	18
4. おわりに	18

1. はじめに

1.1 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 の適用法を検討する。潮位、波等の観測値（海象情報）を学習させた海象学習モデル（Learning Model）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ー上に再現するものであり、構築物の設計外力を予測する海象 LM は、全ての港湾事業（港湾 LM を構成する多数の内装 LM）への拡張を可能にできる。

1.2 研究開発の概要

内装 LM を事前学習させた領域別のローカル GPT（Generative pre-trained Transformer）を作成することで汎用 LLM を活用した精度の高い予測を可能にする。内装 LM にはランダムに学習させた技術者知識（データベース）に汎用 LLM を適用して統一した学習フォーマットを持たせる（技術者データベースの学習フォーマット統一）。多くの技術者の知見を取り入れて深い学習（Deep Learning）をさせた内装 LM を組み合わせた生産性の高いローカル GPT（港湾 GPT）を構築する。将来的には港湾 GPT が港湾計画作成支援につなげることを目指す。港湾 GPT は新米技術者の教材としての活用を可能にし、構造物の設計外力の導出等の学習ツールとしての機能も持たせる。

一般的に機械学習の設計過程はブラックボックスとなっており、設計者は個々のノウハウを汎用 LLM に反映することは難しい。本研究では海象現象を学習した海象 LM を構築して、港湾構造物に対する設計外力を対象とした機械学習のシステム開発を進める。海象変化が異常でない時代では過去の観測値から静的な確率分布を適合させることで十分であったが、現在～将来においては、異常な海象変動に対応できる安定性、確実性の高い動的な予測分布が必要となる。

海象 LM は図-1.1 に示す港湾領域 LM に内装される 1 つの LM であり、他の内装 LM と並列で用いられる。将来的には港湾領域 LM と汎用 LLM を連携して港湾 GPT（ローカル GPT）を構築する。

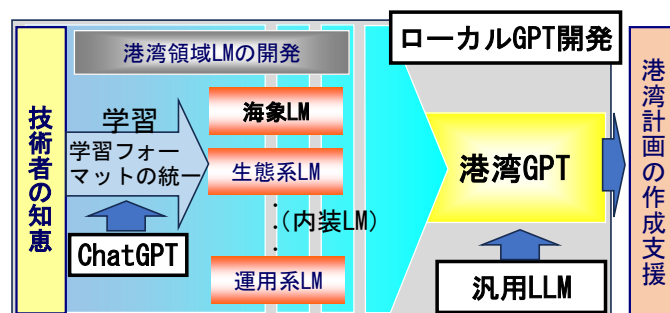


図-1.1 ローカル GPT（港湾 GPT）の概念図

（汎用 LLM と連携できる港湾領域の LM の構築、海象 LM の構築手法を他の内装 LM の構築支援のための教育資料（機械学習の手引き）を作成）

1.3 研究開発の概要港湾領域 LM に取り込む内装 LM

内装 LM として「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-」と「設計波高の未超過確率に関する考察・機械学習に用いる確率関数を考える」を検討した。これらの内装 LM は、技術者のノウハウであり、ノウハウの表現方法（活用法）を検討するためのテーマである。

2. 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-

沿岸域に堆積する底泥には、一次生産を起点として生体（植物や底生動物等の生きている生物）により生成・利用された多種の有機物（生成物や消化物質等）が含まれている。底泥中において、生体により生産される有機物（生体関係有機物）には易分解性として分画される有機物が多く含まれる¹⁾。生体を構成する主な生体関係有機物（元素）はタンパク質（C, H, O, N, S）、脂質（C, H, O, P）、炭水化物（C, H, O）であり²⁾、炭水化物の一部（リグニン等）を除いて易分解性の有機物である。生態系において生成された生体関係有機物（動植物、微生物の死骸、排泄物等）は土壌中で生物群集による分解を受け、ポリフェノール類（ベンゼン環に複数の水酸基が結合した物質）、キノン類、アミノ化合物を生成する。これらの物質は酵素、微生物の酸化酵素、無機イオン、粘土鉱物などの触媒作用により重縮合し、難分解性の無定形コロイド状高分子化合物（腐植物質）に変化していくため、海底堆積後には安定して存在している^{3), 4)}。一方、生体関係有機物からのアミノ基の脱離は速く、沈降過程で有機物内の炭素と窒素の構成比（TOC/TON比、以後C/N比と略す）は増大する^{5), 6)}。よって、土粒子とともに堆積する生体関係有機物の分解度は、C/N比で評価することができる。

この様に、生体関係有機物は、ある特性を持って元素組成を変えて蓄積されており、有機物の堆積状態から生態系をベイズ推論を用いて推定できる可能性がある。ベイズ推論では、対象とする現象が起こる周辺確率を表す事前分布と事前分布に観測値を学習させる尤度関数（条件付確率）が設定できれば、機械的（客観的）に観測値の特性を再現する事後分布、予測分布を求めることができる。

そこで本論文では、文献 7) を参考として有機物の C/N 比を表す分布特性（目的変数）を表現できる事前分布、尤度関数を現地調査（学習）データから見出し、ベイズ推論（機械学習のモデル）を設計した。本目的は、生体関係有機物の生成から海底への堆積に至る過程での C/N 比の特性をパラメータとして、流れ特性の異なる沿岸域での有機物の堆積状態を学習させて、各沿岸域での生物生息環境を推定、評価することである。

2.1 海底堆積有機物の構成

浅場（干潟等）の生物生息場に堆積する有機物の主体は、光合成による生成物（植物プランクトン）、植物プランクトンを起源とする動植物間の捕食、排泄、死亡等、生物活動から生成される生体関係有機物である。生体を構成する炭素に対する窒素の比は C/N 比で表現される。生体関係有機物の元素比（組成比）は生体の生成、死亡や生体間での栄養摂取（食物連鎖）と連動しているため、C/N 比は生物生息場の状態および生物の活性度により変化する。

生体関係有機物の生成（光合成等）から堆積に至る過程での C/N 比変化の概念図を図-2.1 に示す。水

域表層で植物プランクトンを起源として生産される生体関係有機物の元素組成はレッドフィールド比（C : N : P = 106 : 16 : 1 [mol/g]）に従えば C/N 比 = 6.6 となる。表層で生産された生体関係有機物は沈降過程でアミノ基等を脱離して C/N 比を増加させ、還元的な場（海底等）では 10 程度の C/N 比で安定して存在することが報告されている⁵⁾。

生体活動の高い浅場では生体関係有機物から N (NH_4^+) が光合成に利用されたり、選択的に底生動物に摂取されて（N の摂取圧が大きい）減少するため、C/N 比は増大し 20 以上の値をとる⁸⁾。一方、生体活性の高い浅場でも生物の死骸等の集積が多い浅場では、アミノ基の脱離が十分に起こっていない生物の死骸等の C/N は生体 C/N 比の 6.6 に近いため、C/N 比は 10 よりも小さな値をとる。

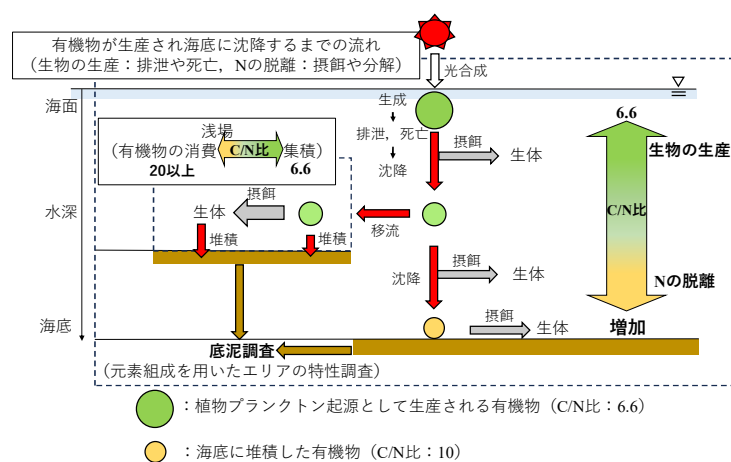


図-2.1 生体関係有機物の生成から堆積に至る過程での C/N 比変化の概念図

2.2 方法

(1) 試料の採取と栄養塩分析

図-2.2 に示される懸灘北沿岸域の A～F の 6 エリア 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）で 2023 年 7 月，8 月に底泥の採泥を行なった。31 地点での調査水深は 5 m 以浅であり，7 月，8 月の調査は同水深で行われた。さらに，F エリア付近の水深 10 m～13 m の 4 地点において 7 月に底泥の採泥を行なった。採泥では，ハンドマッキン型採泥器（採泥面積：0.05m²，22cm×22cm）を用いた。採泥した試料は試料瓶にて冷暗状態で保存し，分析を実施した。

各エリアでの海域特性（有機物の堆積状態）を把握するため，粒度組成，TOC および TN， NH_4^+ を測定し，各々のエリアの C/N 比を比較した。TOC，TN， NH_4^+ ，TP， PO_4^{3-} は底質調査法に基づいて分析された。還元状態では，DIN は NH_4^+ として存在する。本調査の底泥においても，水位 0m 以下の状態（還

元状態）であるため，TN， NH_4^+ の差を TON とした。ただし，KCl 溶液に抽出される NH_4^+ は K^+ とのカチオン交換による溶脱であるため，①Free - NH_4^+ ，②Exchangeable - NH_4^+ ，③Fixed - NH_4^+ の内，①，②の検出にとどまっている⁹⁾。

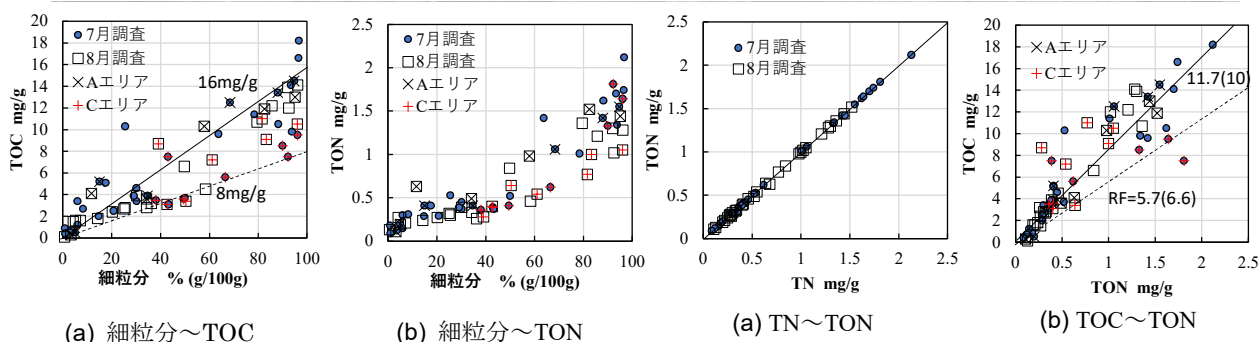


図-2.3 細粒分と TOC および TON の関係

図-2.4 TN と TON および TON の関係

図-2.2 採泥された燧灘北沿岸域の 35 地点（A～F エリア，Google Map を用いて作成）

(2) C/N 比分布特性把握のためのベイズ推論

還元的に堆積した底泥での C/N 比は 10 を中心（平均値）として分布するため⁴⁾，尤度関数として，
C/N 比 = 10

C/N 比 = 10 地点の出現頻度分布を仮定できるガウス分布が適当と考えられる．ただし，C/N 比に生体活動の特性を表現させる場合にはガウス分布の利用は困難になる．対象とする 6 つのエリアでは，生物活性が異なることが予測され，各エリア毎の C/N 比特性を表すことのできるポアソン分布，カテゴリ分布が尤度関数として適していると考えた．また，学習データ数となる各エリアでの調査地点数が 4～7 地点と少ないことから，本推論では観測値（学習データ）を適切に表す尤度関数として式(2.1)に示すポアソン分布 Poi を採用する．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」を表す確率分布である．本解析では対象となる地点数を x ，単位地点数を $\lambda = a/b$ = 全調査地点数/対象地点数として学習データ（ポアソン分布）を作成した．

$$p(x|\lambda) = \text{Poi}(x|\lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$$

$$\text{期待値}(x) = \lambda \quad (2.1)$$

2.3 結果

(1) 各エリアでの有機物の堆積状況（調査結果）

図-2.3 に細粒分（シルト・粘土含有率）と TOC および TON の関係を示す．7 月の全調査エリアの値を水色丸，8 月の全調査エリアの値を白色四角で示し，特に言及する A エリアは×印，C エリアは+印を合わせて表示している．通常，TOC は強熱減量に比例して存在している．図-2.3(a)では広島湾，福山港沖海底泥での $\text{TOC} = 0.2 \times \text{IL}$ (mg/g) 関係を基準値として実線が付記された．細泥分を主体とする海底泥では IL は細粒分の 8～10 %程度なので， $\text{TOC} = 0.016 \times \text{細粒分}[\text{g/g}]$ （細粒分が 8 %の場合 1.6 %が TOC）となる（図-2.3(a)実線）．

対象海域には細粒分に対する TOC 含有率が多いエリアと少ないエリアがあった．A エリアでは 7，8 月とも TOC が多いのに対し，C エリアでは 8 月に TOC が増加する傾向があった．細粒分に対する TOC が

多いエリアでは、細粒分 1 % の 2.4 % (0.048 g/g) 程度の TOC を含むエリアがあった (図-2.3(a)).

a) 有機物量と CN 組成

図-2.4 に TN と TON および TOC の関係、図中の実線は C/N 比 = 10 (モル比)、点線は C/N 比 = 6.6 (モル比) を示す。図-2.4(a)では、Free - NH_4^+ 、Exchangeable - NH_4^+ が 0.01 mg/g オーダー (間隙水では 1mg/L のオーダー) であり、還元的な状態で堆積する底泥では N の大部分が有機態として存在していることが確認された。細粒分に対する TOC が多いエリアでは TON においても多い傾向が見られ、細粒分に有機物が多く含まれ、生物活性が高いことが推察される (図-2.4(b)).

C/N 比は 10 を平均値として分散していた (図-2.4(b)). C/N 比の分散は TOC が 8 mg/g 以上、TON が 1 mg/g 以上で大きく、有機物量の多い場での C/N 比の分散が大きくなっていた。エリア間で C/N 比変化に差があり、C エリアでの C/N 比の変化が大きい。C エリアでは他エリアと異なり 7 月に TOC が低く TON が高い場であったが、8 月に TOC が増加、TON が減少傾向にあった。

b) C/N 比の分布

図-2.3 には水深に対する C/N 比分布が示されている。水深は TP 基準としており、TP + 0.17 m が平均海面、TP-1.28 m が大潮干潮面である。C/N 比は水深 3~5 m では C/N に 10 近く、水深 3m 以浅で分散が大きくなっていた。また、3 m 深 (大潮干潮面-1 m) 付近に細粒分、栄養塩が堆積する傾向があった。

(2) 観測 (学習) 値の頻度分布とポアソン分布

図-2.6 には(a) C/N 比の頻度分布と(b) 区分①~③の C/N 比が出現するまでの調査地点数 x の確率分布が示されている (区分① $9 < \text{C/N} < 11$ (図中では C/N=10 と記載)、区分② $\text{C/N} < 8$ 、区分③ $\text{C/N} > 12$)。図(b)中では C/N 比の大きさによって①~③に区分されたポアソン分布が示されている。学習データとして C/N 比頻度分布から算出された C/N 比区分①~③の λ_i ($i=1\sim3$) をポアソン分布に与えた。 λ_i は対象 C/N 比の出現に必要な地点数を与える x の期待値である。 i は区分①~③に対応する。図-2.6(b)は全エリアでの 7 月、8 月の調査結果に対する①~③の各区分が出現するために必要な地点数 x の確率 p_i ($i=1\sim3$) を表している。

通常、ポアソン分布では区分①を対象として頻度分布から λ を求めて学習分布を作成するが、本解析では、区分②、区分③のように頻度と大きさを与えて学習分布を作成している。例えば区分②では C/N 比 < 8 の範囲の地点数に加えて、C/N 比の大きさ (特性値) を λ_i として与えている (次節、式(2.2)).

(3) ポアソン分布の学習と予測の方法

a) 尤度関数 (学習データ) によると事前分布

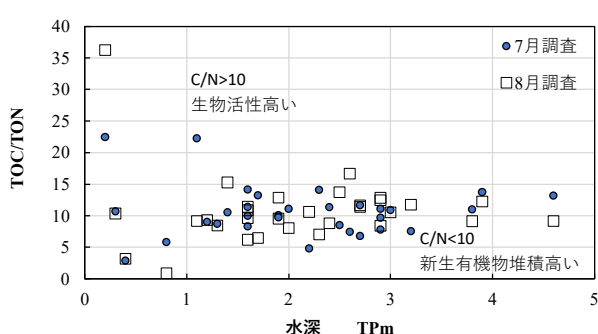


図-2.5 水深に対する C/N 比分布

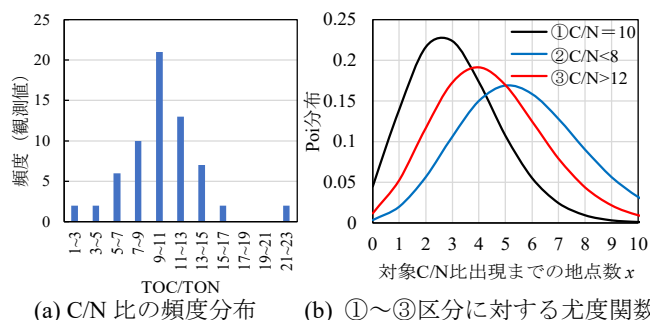


図-2.6 C/N 比の頻度分布とポアソン分布 (尤度関数) の関係

1) 学習データの作成

A～Fの各々のエリアでの全調査地点数 a に対して、対象C/N比である b_i 地点を表す λ_i が学習される．各エリア調査点での対象C/N比の出現の有無（区分①），および出現とエリアの特性値（区分②，③）を学習データ x_m （観測値）とし，区分①～③に対して各エリアの λ_i の平均値が全エリアの λ_i になるよう x_m を補正して学習データ $X = (x_1, x_2, \dots, x_M)$ を作成した．C/N比 = 10である地点の出現の有無を学習させる区分①では，常に $\sum x_m < M$ となるため（式(2.4)）， x_m に C_1 を乗じて $(x_m = C_1, x_m)$ ， $\sum x_m \geq M$ 条件を満たすように学習データを補正，設定した．

対象地点の出現の有無に特性値（各地点のC/N比）を加えた区分②，③では，C/N比 < 8，C/N比 > 12の出現に対して境界値からの差を重みとして x_m を式(2)で与えた．

$$\begin{aligned} \text{②} : x_m &= C_2(8 - (C/N)_m) \\ \text{③} : x_m &= C_3((C/N)_m - 12) \end{aligned} \quad (2.2)$$

C_2, C_3 は， $\hat{a}, \hat{b}_i$ ($i=2, 3$)，式(4)を補正するための係数である．

2) 事前分布（ガンマ分布）

事前分布には式(2.3)に示すガンマ分布 Gam を用いた．

$$\begin{aligned} \text{Gam}(\lambda_i, a, b_i) &= C_G(a, b_i) \lambda_i^{a-1} e^{-b_i \lambda_i} \\ \text{期待値 } \langle \lambda_i \rangle &= \frac{a}{b_i} \end{aligned} \quad (2.3)$$

ガンマ分布は期待値が $\lambda_i (= a/b_i)$ 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．通常，ベイズ推論では事前分布（ガンマ分布）のパラメータは学習に対して影響の小さな値を適当に用いるが，全エリアでの全観測データを用いることで事前分布は個々のエリアの特性を表す基準値（平均値）として用いることができる．したがって事前分布には7月の全観測値に対する区分①～③の対象 C/N 比地点の出現確率 $1/\lambda_i$ が与えられた．

本解析では $a=31$ の調査地点に対する7月の観測値を用いて， λ_i ($i=1\sim3$) は① $9 < C/N < 11$ （以降 C/N 比 = 10 と記述， $i=1$ ），② $C/N < 8$ ($i=2$)，③ $C/N > 12$ ($i=3$) である地点の出現期待値としており，全観測 31 点のうち，各々 b_i = 区分①に該当する地点数（14 地点），区分②に該当する地点数（6 地点），区分③に該当する地点数（7 地点）の期待値を $\lambda_1 = 2.2 \left(= \frac{31}{14} \right), \lambda_2 = 5.2 \left(= \frac{31}{6} \right), \lambda_3 = 4.4 \left(= \frac{31}{7} \right)$ として事前分布に与えた．

b) 学習と事後分布（ガンマ分布）

各エリアに対し，①～③区分の尤度関数（式(2.1)）により事前分布（式(2.3)）を学習させると，式(4)で表される事後分布（ガンマ分布）が得られる．

$$\begin{aligned} p(\lambda_i | X) &= \text{Gam}(\lambda_i | \hat{a}, \hat{b}_i) \\ \hat{a} &= \sum_{m=1}^M x_m + a \quad \hat{b}_i = M + b_i \end{aligned} \quad (2.4)$$

M は各エリアでの調査（採泥）地点数、 $\hat{a}$, $\hat{b}_i$ は学習された a , b_i である。 $\hat{a}$ は調査地点数 a に $\sum x_m$ （各エリアでの特性値）、 $\hat{b}_i$ は b_i に調査地点数 M が加えられる。すなわち、各エリアでの調査地点数 a と対象 C/N 比の出現数 b_i が与えられる学習では 2 つのパラメータに各々エリア毎の特性値を加えた $\hat{a}$ と調査地点を加えた b_i により λ_i が推定される（式(2.4)）。事後分布では λ_i は各区分の特性を表すため、特性値の大きいエリアでの期待値を大きくさせている。ポアソン分布による学習では $\sum x_m > M$ が満たされない時には、事後分布の λ_i は事前分布の λ_i よりも小さくなり、事前分布との比較による対象 C/N 地点の増加を評価できないことに注意を要する。各区分とも事前分布と同じ期待値がある調査エリアでの期待値が等しくなるように $C_1 = 12$, $C_2 = 8$, $C_3 = 3$ を与えた。

この操作により、特性値の大きさは期待値の増加として現れており、ポアソン分布とは逆に事後分布で学習された $\lambda_i (i = 2, 3)$ は各エリアで期待される特性値を表している。例えば区分③の特性は生物活性を表しており、 $\lambda_3 (i = 2, 3)$ は生物活性の高さを表す期待値である。

c) 負の二項分布が表す予測分布

ガンマ分布をポアソン分布で学習させると式(2.5)が得られる。式(2.5)を変形すると予測分布として式(2.6)で表される負の二項分布 NB が得られる。パラメータ a , b_i はポアソン分布、ガンマ分布と共通である。ただし、ガンマ分布に従うポアソン分布が負の二項分布であり、事後分布で得られた λ に基づいた確率分布（式(2.4)と同様に特性の期待値）が求められる。

$$p(\lambda_i|X) \propto p(X|\lambda_i)p(\lambda_i) = \{\prod_{m=1}^M \text{Poi}(x_m|\lambda_i)\}\text{Gam}(\lambda_i|a, b_i) \quad (2.5)$$

$$p(x_*) = \text{NB}(x_{*i}|r, p_i) = \frac{\Gamma(x_{*i}+r)}{x_{*i}!\Gamma(r)} (1-p_i)^r p_i^{x_{*i}} \quad (2.6)$$

$$\text{期待値 } \langle x_* \rangle = \lambda = \frac{a}{b_i} = \frac{p_i r}{1-p_i}$$

$$r = a, \quad p_i = \frac{1}{b_i + 1}, \quad b_i = \frac{p_i}{1-p_i}$$

式(2.6)では $1/b_i$ がオッズを表しており、 λ_i は a の内、対象 C/N 地点になり易い地点数を表すことになる。オッズは p_i が調査地点数 M （学習データ）の増加により出現確率の低下を表す（式(2.4)）が、 a には対象 C/N 比出現地点の特性値の和 $\sum x_m$ が加わることで、出現地点の特性の期待値を増加させている。本解析では区分①～③での x_i （期待値）を比較することで各エリアでの C/N 比特性（生物生息環境）が把握される。

(4) ポアソン分布学習結果の考察

図-2.7 には区分① (C/N 比=10), 図-2.8 には区分② (C/N 比<8), 図-2.9 には区分③ (C/N 比>12) に対して(a) 期待される地点数 (ポアソン分布) を 学習させた時の A~F エリアでの(b) 事後分布と(c) 予測分布が示されている. (a)学習データに記述されていないエリアが存在するのは, そのエリア内で対象 C/N 比の地点が存在していないためである. 図-2.7~2.9 の結果 (各図(b), (c)) は7月の調査を学習させて推定された各エリアの区分①~③の特性が現れる地点数の期待値 (図(b)) と調査地点数の内, 区分特性が現れる地点数の確率分布 (図(c)) を示している. ポアソン分布学習では区分①~③特性を持つ地点

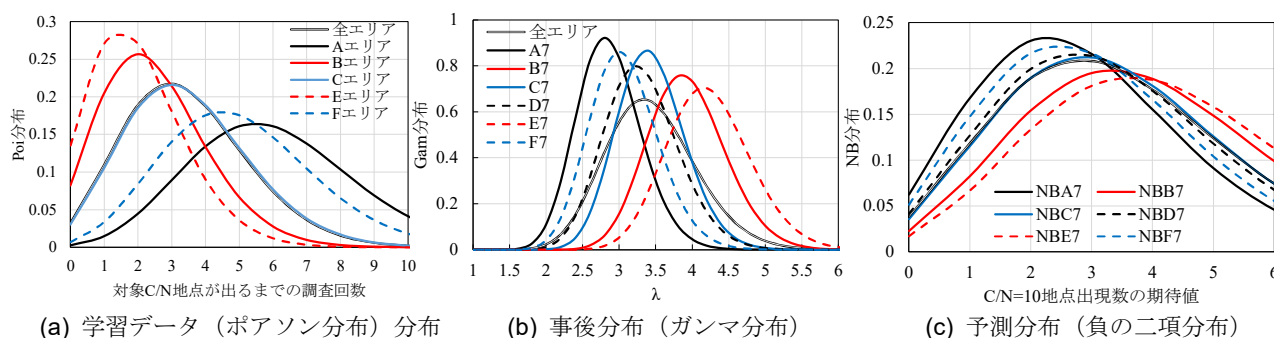


図-2.7 ① (C/N 比=10) 区分 となる地点数 (ポアソン分布) を学習させた時の A~F エリアでの事後分布と予測

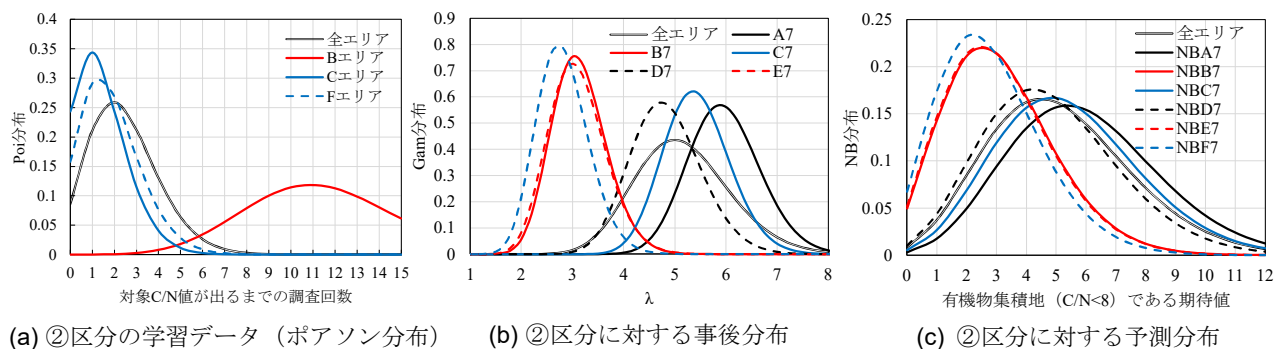


図-2.8 ② (C/N < 8) 区分に対する学習データ, 事後分布と予測分布

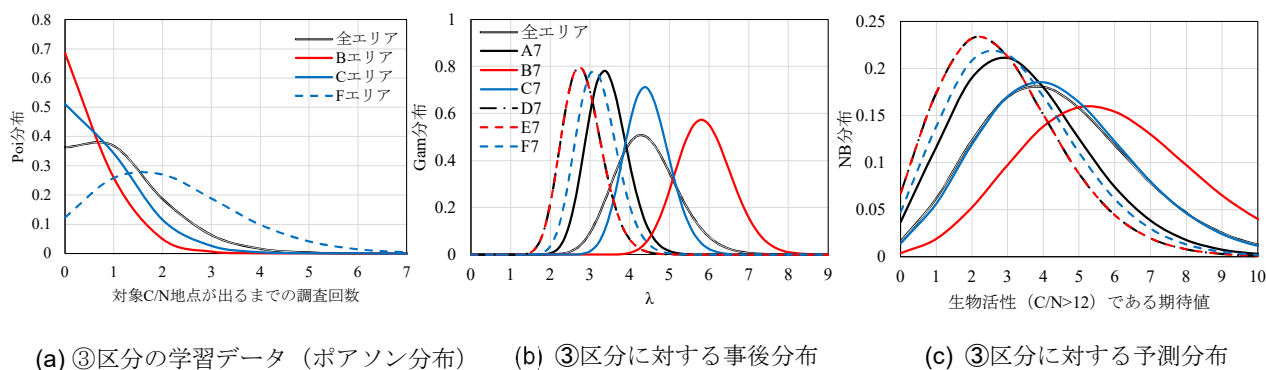


図-2.9 ③ (C/N > 12) 区分に対する学習データ, 事後分布と予測分布

数の出現頻度を学習データとしており, 区分①~③に対して各々のエリアで異なる学習データになっている. ポアソン分布では調査回数 (期待値) が小さいエリアが区分①~③の特性を期待できる場であり, 全エリアの分布よりも調査回数の小さいエリアが平均以上の期待値があることを示している. したがっ

て、C/N 比（実測値）が 10 に近い場合は E、B エリアであり、新生有機物の堆積が多い場（C/N>8）は A、C エリア、生物活性の高い場（C/N>12）では B エリアであることが学習される。

図-2.10 には予測分布から求められた各エリアでの区分②、③特性が出現する地点の期待値が示されている。区分②、③の期待値は式(6)から求められている。B エリアは区分①、③の期待値が高く生物活動の高い場として評価（図-2.7、図-2.10）。E エリアは B エリアとともに区分①の期待値が高いが、区分③の期待値が低く、B エリアよりも生物活性が低いと評価されている。C エリアは区分②、③が混在しており、エリア内に生物の生息環境の異なる場が形成されている。これらの特性は図-2.3～2.6 に示す観測値から読み取れるが、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を得ることができる。ただし、観測値の特性を表す学習データ（尤度関数、図-2.6(a)）が見つけれない場合には適切な機械学習は実施できない。

2.4 2章まとめ

沿岸域に堆積する底泥の有機物組成（C/N比）を説明変数としてポアソン分布学習により底泥質を評価した。C/N比=10を基準として未分解有機物の堆積場、生物活性場を定義して調査エリアの生物生息環境を評価した。

- (1) 沿岸域底泥のC/N比は10を平均値とした分布で堆積している。C/N比は水深4～5 mでは分散が小さい（C/N=10に近く）が、水深3m以浅で分散が大きく、生物活性が高い特性があった。TOCが8 mg/g以上、TONが1 mg/g以上でC/N比の分散が大きく、有機物量の多い場でのC/N比の分散が大きくなっている。
- (2) 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を行った。ポアソン分布学習法を用いて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ることが確認された。
- (3) 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 λ_i ($i=1\sim3$) をポアソン分布に与えて学習分布を作成した。C/N 比に頻度と大きさを与えることによって各エリアの生物生息場の特性を学習させた。
- (4) 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を得ることができる。ただし、観測値の特性を表す学習データ（尤度関数）が見つけれない場合には適切な機械学習は実施できない。

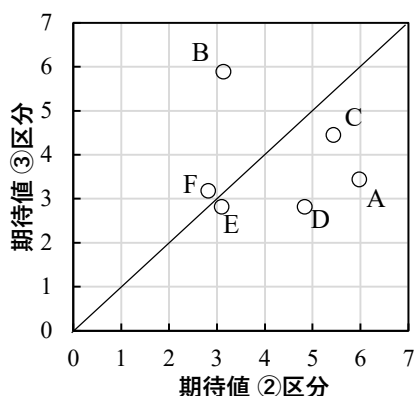


図-2.10 予測分布から求められた各エリアでの②、③区分特性出現地点の期待値

3. 設計波高の未超過確率に関する考察-機械学習に用いる確率関数を考える-

異常化する波浪に対して、設計外力となる沖波高の設定には気象条件に合わせた物理モデルによる波浪計算と観測値が用いられている。波浪計算により適切な異常波浪が設定できれば、極値統計による最大波高に気象変動の影響が考慮されることになる。本解析では機械学習を用いて極値統計により推定される最大波高発生確率の考え方を検討した。機械学習にはベイズ推論が用いられている。ベイズ推論は観測値である最尤関数を事前分布に乗ずることによって事後分布の発生精度を向上させる学習法である。ベイズ推論を設計波高の推定に用いるために適切な最尤関数と事前関数を設定した。本解析では波高の超過確率分布はワイブル分布に従うとして、ワイブルのパラメータの設定、学習方法について検討している。

異常波浪（波高）の極値は、例えば 1 年を区切りとして気象擾乱による年最大波高は各年毎に測定された高波が独立して発生することを前提としており、個々に発生した高波相互の関連性を認めていない。本論文では極値統計論に用いられる確率関数の物理的な意味を検討するとともに、ワイブル分布を仮定して求められる最大波高の発生確率を算定することを目的としてベイズ推論を行った。時系列で変化する高波浪の発生率を最尤関数、ガウス分布を事前関数として最大波高発生モデルとして（ベイズ推論モデル）構築した。異常気象下において未知の現象が不確実な状態にある沿岸域において新記録として出現する最大波高の発生率がワイブル分布に従うとして、機械学習により経年的な最大波高の発生傾向を組み込んだ高波浪の発生確率を推定した。

3.1 未超過確率でのベイズ推定の定式化

(1) 高波浪の未超過確率とワイブル分布

異常波浪の極値統計解析では高波の発生率を表す確率分布の 1 つとしてワイブル分布が用いられている。(3.1)式で表わされる $F(x)$ は未超過確率を表すワイブル分布であり、パラメータは k, α, β ($k > 0, \alpha > 0, \beta \geq 0$) の 3 つがある。 k は形状、 α は尺度、 β は位置を表すパラメータである。

$$F(x) = 1 - \exp\left(-\left(\frac{x-\beta}{\alpha}\right)^k\right) \quad (3.1)$$

$$P[H < x_m, N] = P(m) = 1 - \frac{m-\alpha}{N+\beta} \quad (3.2)$$

x_m, N は N 個の極値中で m 番目に大きい波高を意味する。(3.2)式の α, β は極値の分布関数として想定したパラメータ、 $P(m)$ は観測期間に観測された m 番目に大きい波高の未超過確率である。(3.2)式ではある母集団から抽出された標本のあてはめ誤差が最小になるようにワイブル分布に対して α, β を選定している²⁾。 N 個の波から m 番目に大きい波高 x_m より大きな波が出現しない確率（未超過確率）は昇順に並べ替えた波高 x に対して(3.2)式で与えられる。(3.1)式は(3.2)式を用いて γ_v を(3.3)式としてワイブル分布が求められる。

$$\gamma_v = \frac{x-\beta}{\alpha} = [-\ln\{1 - P(H \leq x)\}]^{1/k} \quad (3.4)$$

観測値を $x = A\gamma_v + B$ とおき、最適な A, B を最小二乗法で求める．求められた A, B を(3.4)式に代入すると実測値に対する(3.4)式のワイブル分布が求まる．

$$F(x) = P(m)[H \leq x] = 1 - \exp\left[-\left\{\frac{(x-B)}{A}\right\}^k\right] \quad (3.4)$$

極値統計では k, a, β を変え算出される A, B の内、相関係数の最も高い値が用いられるが、本標本（観測波高 x_m ）では $k=1.5, \alpha=0.42, \beta=0.34$ を与えた時の相関係数は 1.008 であった．図-3.1 に示される Case 1 が標本に用いられた観測波高の時経列である．観測波高に超過確率（(2)式）を与えて、最小二乗法により求められた A, B を(3.4)式に代入して x_m と $P(m)$ の関係求めると図-3.2 のようになる． $P(m)$ と実測値 x_m との関係はワイブル分布上にプロットされており、対象とした観測値はワイブル分布として機械学習を行った．

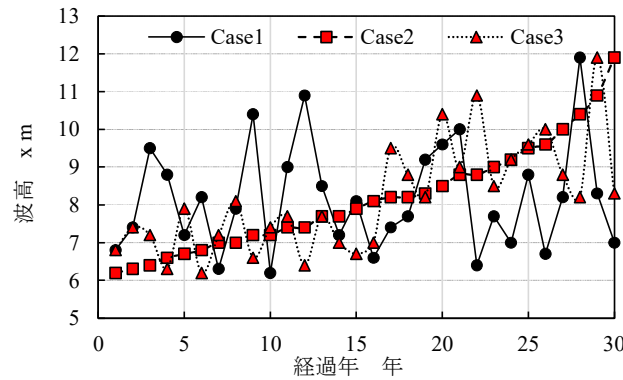


図-3.1 3つのケースでの観測年と波高の関係
(Case 1 は実測値, Case 2, 3 は並び変えデータ)

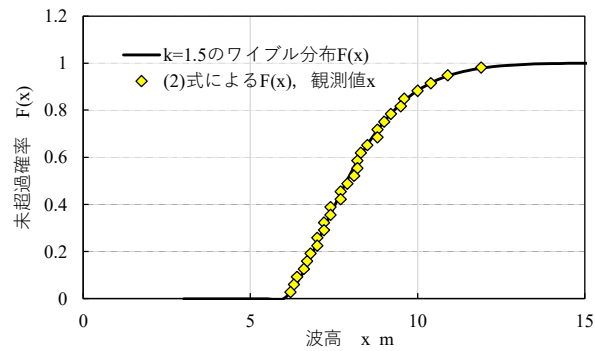


図-3.2 波浪の極値統計によるワイブル分布 ($F(x), k=1.5$)

3.2 最大高波出現予測の方法

$N-1$ 個までに現れる全ての H が x_m よりも小さい確率をとすると、 N 番目に出現した波高 H_N の波が x_m より大きくなる確率は $1-P(x_m)$ であり、次式で表わされる最大波高 H_N の出現確率 $F(N)$ は N 年目に x_m より大きい波が出現する確率を表す。

$$F(N) = \{P(x_m)\}^{N-1}\{1 - P(x_m)\} \quad (3.5)$$

N の期待値 R_p は再現期間（年）であり次式で表わされる。

$$R_p = \sum_{N=1}^{\infty} NF(N)^{N-1} = \frac{1}{1-P(x_m)}$$

R_p は高波の平均発生率 $\rho = N/K$ （個/年、 K 年間の高波の総個数 N ）の場合には極限値の発生時間単位が個 ρR_p 個含まれるとして次式で定義される。毎年最大値資料に対しては $\rho = 1$ が用いられる。

$$R_p = \frac{1}{\rho} \frac{1}{1-P(x_m)} \rightarrow P(x_m) = 1 - \frac{1}{\rho R_p} \quad (3.6)$$

図-3.2 の関係から $R_p = 50$ 年、 $\rho = 1$ ($F(x) = 0.98$) として確率波高を求めると $H_N = 12.0\text{m}$ が求まる。

3.3 ガウス分布を用いた機械学習の方法

(1) k, g をパラメータとしたワイブル分布

本解析では、毎年得られる新記録情報を用いるため、初期観測値 ($n=1$) が新記録となる波を基準 ($x=0$) とし、波高軸の原点を移動する位置パラメータをゼロ ($\beta=0$) とした。これにより、パラメータを k, a の2つに減らし、(3.1)式において $\alpha^k = a(g)$ とおいた(3.6)式を未超過関数として用いる。

$$F(x) = 1 - \exp\left(-\frac{x^k}{a(g)}\right) \quad (3.7)$$

$k, a(g)$ は(3.5)式から以下の手順で求められた。

$$1 - F(x) = \exp\left(-\frac{x^k}{a(g)}\right)$$

両辺の対数を2度とって、

$$\ln(1 - F(x)) = -\frac{x^k}{a} \rightarrow \ln\left(\frac{1}{1-F(x)}\right) = \frac{x^k}{a}$$

$$\ln\left(\ln\left(\frac{1}{1-F(x)}\right)\right) = k \times \ln(x) - \ln(a)$$

$\ln\left(\ln\left(\frac{1}{1-F(x)}\right)\right) = y, \ln(x) = t, -\ln(a) = g$ とおけば、次式の直線式を得る。

$$y = k \times t + g \quad (3.8)$$

波高 x での $F(x)$ のデータをプロットし最小二乗法により、傾き k と切片 g を求める． k 、 g を用いて(3.7)式を表すと(3.9)式が得られる．

$$F(x) = 1 - \exp\left(-\frac{x^k}{e^{-g}}\right) \quad (3.9)$$

(2) 超過確率分布の学習モデルの構築

ワイブル分布のパラメータ(k, g)を求める学習モデル $P(\mathbf{X}, \mathbf{w}) = P(\mathbf{X}|\mathbf{w})P(\mathbf{w})$ を定式化する．まず、ベイズ推論で用いる尤度関数（観測データを得た際のパラメータの確率分布 $P(\mathbf{w}|\mathbf{X})$ ）、事前分布（観測データを得る前のパラメータ確率分布 $P(\mathbf{w})$ ）、事後分布（尤度と事前分布によって、更新されたパラメータの確率分布 $P(\mathbf{X}|\mathbf{w})$ ）を以下のように設定する．図-3.3に未超過確率分布 $F(x)$ を予測するためのグラフィックモデルの組み立て手順が示されている．

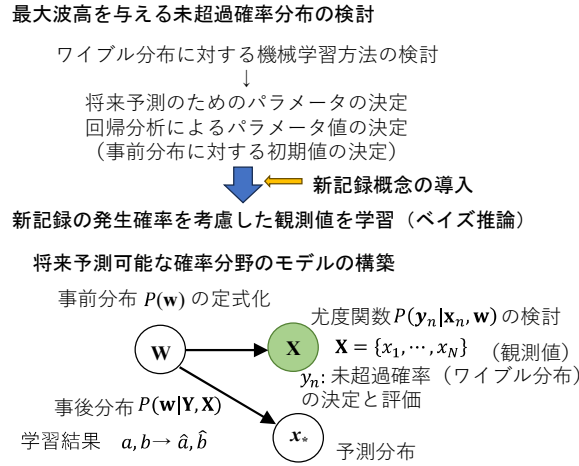


図-3.3 $\mathbf{w}^T = (k, g)$ によるワイブル分布形状の推定のためのグラフィカルモデルの組み立て

a) 観測モデル（尤度関数）の定式化

(3.8)式の誤差 ε を(3.10)式のように組み込む．

$$y_m = \mathbf{w}^T \mathbf{x}_m + \varepsilon_m \quad (3.10)$$

ここに、 $\mathbf{x}_m = \begin{pmatrix} t_1 \\ \vdots \\ t_M \\ 1 \end{pmatrix}$, $\mathbf{w} = \begin{pmatrix} k \\ g \end{pmatrix}$ である．

誤差 $\varepsilon_m = (\varepsilon_1, \dots, \varepsilon_M)$ が、平均値 0、精度 $\lambda^{-1} (= 1/\sigma^2)$ の正規分布に従うと仮定して、事前分布を次式の正規分布で表わす．未知の平均値 $\mathbf{w}(k, g)$ と既知の精度 λ^{-1} を条件として、 y_m の確率分布が $F(x)$ 観測モデ

ルとなる.

$$\begin{aligned} P(y_m|\mathbf{x}_m, \mathbf{w}) &= \mathcal{N}(y_m|\mathbf{w}^T \mathbf{m}, \lambda^{-1}) \\ &= \frac{\sqrt{\lambda}}{\sqrt{2\pi}} \exp\left\{-\frac{\lambda}{2} (y_m - \mathbf{w}^T \mathbf{x}_m)^2\right\} \end{aligned} \quad (3.11)$$

ここに, $y_m = (y_1, \dots, y_M)$ である.

b) 超過確率分布とワイブル分布との関係

図-3.4 には図-3.1, Case1 に用いた実測波高 $x_m \sim P(m)$ に対してパラメータ k, g として求められた関係が波高 $x_m \sim P(m)$ と比較されている. 図(a)には 観測期間を変えて(3.10)式により最小2乗法で求められた回帰直線, (b) には k, g をパラメータとしたワイブル分布 ((3.9)式) と実測値 ($x_m \sim P(m)$ 関係) が比較されている. なお, 図-3.2 で求められた未超過確率の設定方法 ((3.2)式による $P(m)$ 設定) では Case1~3 のいずれに対しても図(b)の結果が得られる. k, g をパラメータとしたワイブル分布においても観測値を近似できている (相関 0.9) が, 最大値を過小評価する (x_m を超える H_N の発生確率が高くなる). 3 つの観測期間での誤差を比較すると, (3.2)式で $P(m)$ を与え, かつ k をパラメータとして回帰曲線を求めた場合には最大値の過小評価が拡大されることになる. 図-3.4(b)の関係は本解析に用いる初期のワイブル分布とした. (3.6)式を用いて R_p を求めると図-3.3 の関係から $H_N = 10.8\text{m}$ なる. 観測値の出現確率を検討すること (学習) により予測精度を向上させるを試みる.

(3) 線形回帰モデルの設定

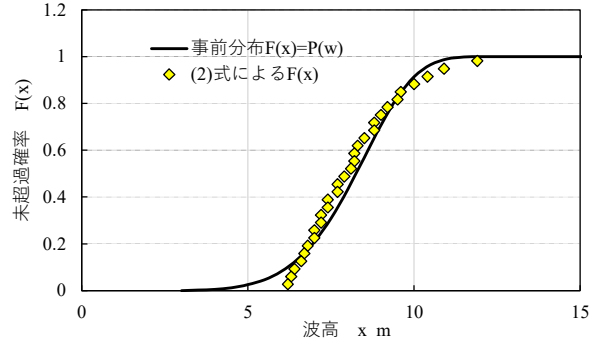
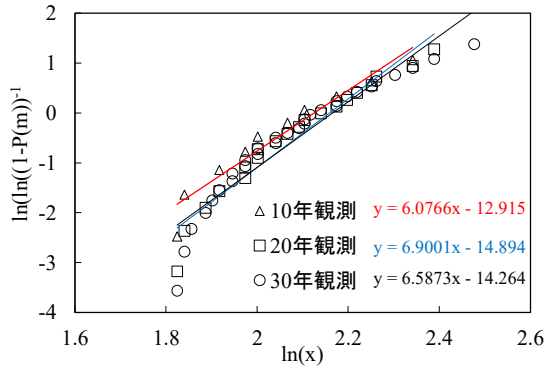
a) 事前分布の定式化

パラメータ k, g の事前分布 $P(\mathbf{w})$ が正規分布に従うとすると, $P(\mathbf{w})$ は次式となる.

$$\begin{aligned} P(\mathbf{w}) &= \mathcal{N}(\mathbf{w}|\mathbf{m}, \Lambda^{-1}) \\ &= \frac{\sqrt{|\Lambda|}}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} (\mathbf{w} - \mathbf{m})^T \Lambda (\mathbf{w} - \mathbf{m})\right\} \end{aligned} \quad (3.12)$$

ここに, $\mathbf{m} = \begin{pmatrix} \mu_k \\ \mu_g \end{pmatrix}, \Lambda = \begin{pmatrix} \lambda_k^{-1} & 0 \\ 0 & \lambda_g^{-1} \end{pmatrix}$ である. $\mathbf{m} \in \mathbb{R}^M$ は平均パラメータ、正定値行列 $\Lambda \in \mathbb{R}^{M \times M}$ は精度行列パラメータである. $\mathbf{m}$ や Λ は超パラメータであり, 事前にある値で固定する必要がある. 本解析では μ_k, μ_g として最小2乗法で求められた k, g を用いる (図-3.4(b), 実線). 初期値となるワイブル分布は極値統計にならない, (3.2)式が用いられている. したがって事前分布に与える観測値 x_m に対する誤差は図-3.2(b)での観測値とワイブル分布との差としている.

この事前分布 (線形回帰モデル) を使って, データを観測したあとの事後分布と予測分布が求められた. 学習データには異常高波の発生に対して新記録という概念を導入した最大波高の未超過確率 $P(m)$ が用いられる.



(a) (10)式による $x_m \sim P(m)$ 関係の近似

(b) $x_m \sim P(m)$ 関係とワイブル分布 ((9)式)

図-3.4 $\mathbf{w}^T = (k, g)$ によるワイブル分布形状の推定のためのグラフィカルモデルの組み立て

b) 事後分布の計算

事前分布に $P(\mathbf{w})$ を $P(\mathbf{y}_n | \mathbf{x}_n, \mathbf{w})$ 学習させると、事後分布 $P(\mathbf{w} | \mathbf{Y}, \mathbf{X})$ が計算される。ベイズ推論では $\mathbf{w}$ を含まない $P(\mathbf{Y} | \mathbf{X})$ は当初無視して、後で正規化される。

$$P(\mathbf{w} | \mathbf{Y}, \mathbf{X}) = \frac{P(\mathbf{w}) \prod_{n=1}^N P(\mathbf{y}_n | \mathbf{x}_n, \mathbf{w})}{P(\mathbf{Y} | \mathbf{X})} \propto P(\mathbf{w}) \prod_{n=1}^N P(\mathbf{y}_n | \mathbf{x}_n, \mathbf{w}) \quad (3.13)$$

(3.13)式に(3.11), (3.12)式を代入して, $\mathbf{w}$ について整理すると, (3.13)式が得られる。

$$P(\mathbf{w} | \mathbf{Y}, \mathbf{X}) = \mathcal{N}(\mathbf{w} | \mathbf{m}, \hat{\Lambda}^{-1}) \prod_{n=1}^N \mathcal{N}(y_n | \mathbf{w}^T \mathbf{x}_n, \lambda^{-1})$$

$$(\mathbf{w} | \mathbf{Y}, \mathbf{X}) = \mathcal{N}(\mathbf{w} | \hat{\mathbf{m}}, \hat{\Lambda}^{-1}) \quad (3.14)$$

$\hat{\Lambda} = \lambda \sum_{n=1}^N \mathbf{x}_n \mathbf{x}_n^T + \Lambda$, $\hat{\mathbf{m}} = \hat{\Lambda}^{-1} (\lambda \sum_{n=1}^N y_n \mathbf{x}_n + \Lambda \mathbf{m})$ である。上式でのハット ($\hat{\cdot}$) は学習された値を示している。

c) 予測分布

新規入力値 $\mathbf{x}_*$ が与えられたときの出力値 y_* の予測分布 $P(y_* | \mathbf{x}_*, \mathbf{Y}, \mathbf{X})$ を求める。パラメータの事後分布は事前分布と同じガウス分布である。事前分布を使った場合の予測分布を計算し, 事前分布を事後分布に置き換えて予測分布 $P(y_* | \mathbf{x}_*, \mathbf{Y}, \mathbf{X})$ を求める。

$$P(\mathbf{w} | \mathbf{y}_*, \mathbf{x}_*) = \frac{P(\mathbf{w}) p(\mathbf{y}_* | \mathbf{x}_*, \mathbf{w})}{P(\mathbf{y}_* | \mathbf{x}_*)} \quad (3.14)$$

(3.14)式は(3.15)式に示される1次元のガウス分布としてまとめることができる。

$$P(y_* | \mathbf{x}_*) = \mathcal{N}(y_* | \mu_*, \lambda_*^{-1})$$

$$\mu_* = \hat{\mathbf{m}}^T \mathbf{x}_* \quad \lambda_*^{-1} = \lambda^{-1} + \mathbf{x}_*^T \hat{\Lambda}^{-1} \mathbf{x}_* \quad (3.15)$$

$\hat{\mathbf{m}} = (\hat{k}, \hat{g})$ は学習結果であり, 回帰直線の傾きと切片を表している。回帰分析では予測値 μ_* は回帰直線上の点になるため, 本解析では得られる予測分布は使用できない。

3.4 観測データの学習と更新された事後分布

(1) 新記録の発生確率

a) 新記録の発生トレンドの異なる学習データの生成

図-3.1 に用いた波高データ（観測データ）を用いて作成した仮想の2つの波高データを作成して、3つのケースでのN年目の最大波高（新記録）の出現率を検討した。仮想の2つのケースに用いた波高データはCase 1での観測値であり、観測年を並べ替えて新記録の出現が異なるトレンドにしている。図-3.1に3つのケースでの観測年と波高の関係が示されている。Case 1は観測波高実測値（学習データ）、Case 2は観測値を昇順に並び替えた毎年新記録が出現する学習データ、Case 3は観測期間前半に平均値以下、後半に平均値以上の観測値に並び替えたデータ学習である。

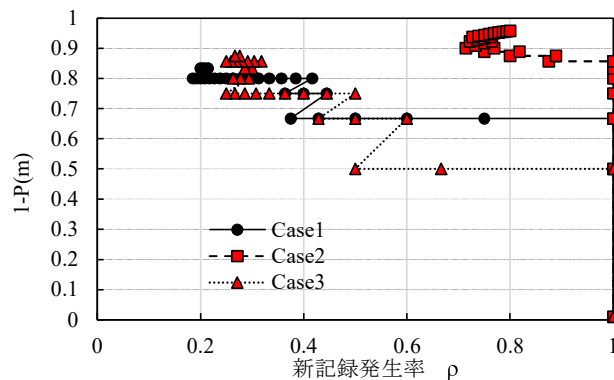


図-3.5 新記録の発生率 ρ と発生確率 $1-P(m)$ の関係

b) 新記録発生確率の考え方

波浪の極値統計では、 $P(m)$ は(3.2)式によって与えられているが、本解析では学生データ (y_n, x_n) として毎年発生し得る新記録 ($H_N > x_m$) の未発生確率 $P(m)$ を与えた。 $P(m)$ は(3.5)式で与えられる最大波高出現確率 $F(N)$ に用いられる $P(x_m)$ と同様の考え方であり、 k_{n-1} 年 k_n は $H_{n=1} < x_m$ の発生前)まで新記録が発生しない確率であり、 n は新記録の発生回数である。 k_n が予測年 K に達した時 $n=N$ となる。 $P(m)$ は K 年間に生じた最大高波（新記録）の更新回数 N を用いて(6)式の $P(x_m) = P(m)$ として表わしている。 ρ は新記録発生率であり、(3.6)式と同様に $\rho = N/K$ で与えられるが、極値統計の考え方とは異なる。極値統計では毎年発生する最大波高を考えているため、 $\rho = 1$ を用いるが、本解析では K 年間に発生した新記録数 N を数えるため、 k_n 年目に発生する最大値の出現率 n/k_n を考慮している。新記録となる波高 x_m に対応する $P(m)$ は(3.6)式により与えられる。

また、 R_p は N の期待値（本解析では N 個目に新記録が発生するために必要な期間）であることから、対象期間を K とした場合には、新記録の発生回数 R_p の再現期間 K が R_p を表わすことになる。したがって R_p は k_{m-1} 年目まで新記録が出現しない確率を表すことになり $R_p = K$ を(3.6)式に代入すれば新記録の発生確率として $1-P(m)$ が求まる。

c) 新記録の発生率と発生確率

図-3.5 には図-3.1 に示した各ケースでの新記録の発生率 ρ と発生確率 $1-P(m)$ の関係が示されている。いずれのケースにおいても ρ は新記録の発生とともに上昇するのに対し、新記録が発生しない期間では、

ρ は低下し $1-P(m)$ は一定値を示す．すなわち，新記録が継続して発生する（発生率が上昇）期間には，新記録の発生確率は高くなることがわかる．

Case 1～3を比較すると，継続的に新記録が出現するCase 2，Case 3の順に $1-P(m)$ が高くなっており，新記録の発生確率は新記録の出現傾向に依存している．すなわち同期間内に新記録の発生回数の増加に伴い $1-P(m)$ は上昇しており，新記録の発生回数が新記録の発生確率を高めることを $1-P(m)$ は表現している．

(2) 新記録の発生確率の学習による分布の変化

図-3.6 には Case1～3 の観測値，図-3.7 には異なる期間の観測値を各々学習データとした時の新記録発生未超過確率 $P(m)$ を学習させた事前分布，学習データ，事後分布が示されている．図-3.7(c)には図-3.6，図-3.7(a)，(b)に示された事後分布がまとめられている．図中の $P(m)$ は R_p =観測期間 $K+1$ 年とした時の新記録の未発生確率であり，30 年観測では 31 年目に新記録が出ない確率が計算されている．事前分布には，観測期間に測定された波高データの K 年までの新記録の高波 x_w を対象として新記録波高とその未超過確率を学習させている．例えば，30 年の観測期間では 30 年間に出現した最大波高を超える新記録に対する未超過確率が学習データであり，再現期間 R_p を $K+1=31$ 年とし， K 年までに発生した新記録の個数 N から $P(m)$ が求められている．

(3) 観測（学習）期間の高波未超過確率との関係

a) 新記録発生傾向の影響

$P(m)$ を学習させた $F(x) = P(w|Y, X)F(x) = P(w)$ は最大波高の出現確率を表わす尤もらしいワイブル分布が生成されている．昇順に並び替えたCase 2では学習された事後分布 $P(w|Y, X)$ が(3.2)式による $F(x)$ （図-

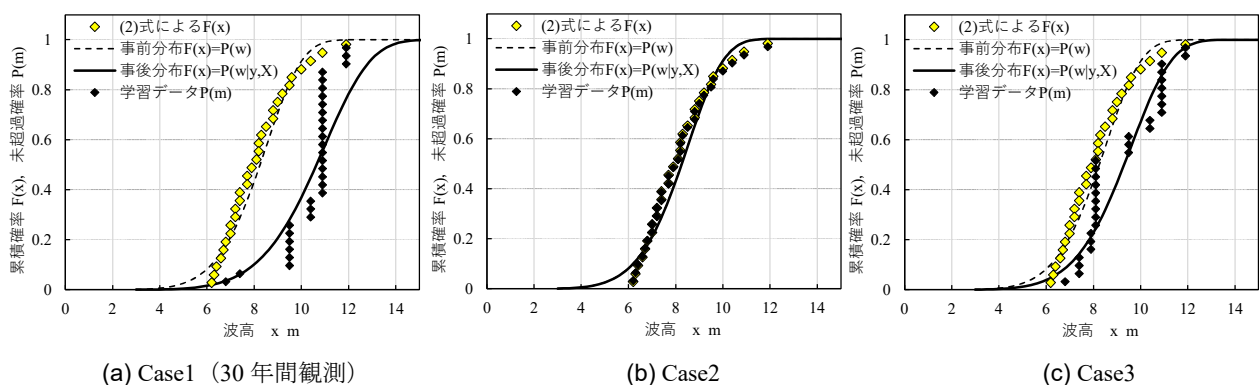


図-3.6 事前分布，新記録発生未超過確率 $1-P(m)$ を用いた学習データ，事後分布（Case1～3 での未超過確率の比較）

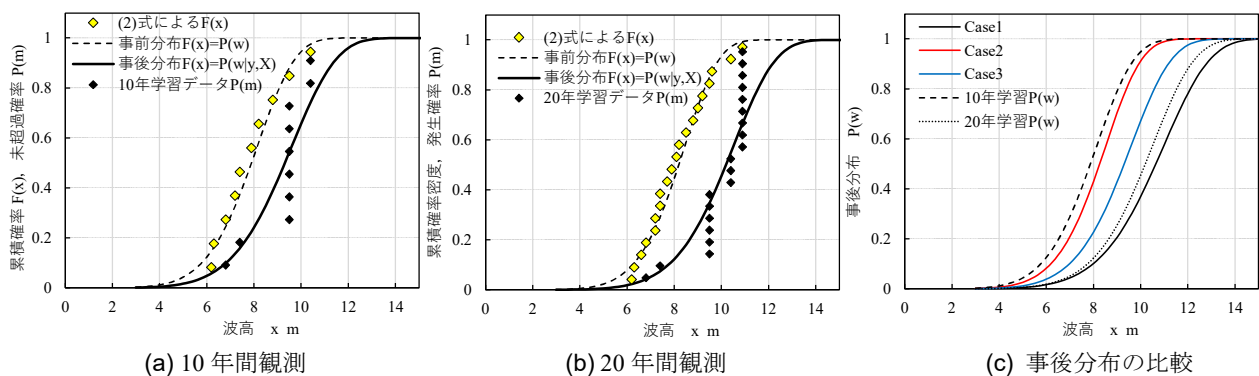


図-3.7 事前分布，新記録発生未超過確率 $1-P(m)$ を用いた学習データ，事後分布（異なる観測期間での比較）

3.1) と等しくなっており、未超過確率の設定、学習方法の妥当性が確認できる。(3.2)式と(3.6)式は異なる関数であるが、各々で計算した未超過確率が概ね等しくなっている。この結果は、(3.2)式は試行錯誤によって求められた関数であるが、特殊な場合を仮定したCase2では(3.2)式が x_m よりも大きい波が出現する各年の生起確率関数を表していることがわかる。

Case3では観測期間後半に新記録が出現し易い傾向を観測値に与えたため、新記録の発生確率がCase1に比較して低下している。この推論では、新記録の発生率が高くなるほど、高波浪の出現確率は低下する傾向を推定している。この特性は機械学習の特性であり、 $F(N)$ ((3.5)式) に基づいており、平均発生率を考慮した $F(N)$ ((3.6)式) を用いれば、「近年、新記録が出ていないので、新記録が出易くなるはず」という結果が推定される。「これまで新記録が出る傾向にあるので、新記録が出易くなるはず」という傾向を示すためには、(3.6)式とは異なる確率関数を用いる必要がある。

b) 観測（学習）期間の影響

学習期間（観測期間）が長くなるにつれ、高波の出現率は高くなることは自明であるが、本解析においても観測期間が30年のCase1での新記録の発生確率が高くなるように学習されている。新記録の発生は初めの10年間で2回、20年で3回、30年で4回発生している（Case1, 図-3.1）が、発生回数よりも高い新記録の発生が高波高の波の出現確率を上げていることがわかる。

3.5 3章のまとめ

波高の極値統計にベイズ推論を適用した。過去の情報を学習して将来予測する手法は機械学習に適している。地球温暖化問題がある近未来では、ベイズ推論による将来予測は有効な手段となる。

4. おわりに

本研究で構築する海象LMが汎用LLMと連携するアルゴリズムが開発を目的として「堆積泥のCN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-」、「設計波高の未超過確率に関する考察・機械学習に用いる確率関数を考える-」を提案した。これらに用いたベイズ推論は、今後開発される特定領域の港湾LMと汎用LLMとの連携を可能にするツールである。

自然環境が急激に変化する近年では、造成干潟、防波堤の設計等に必要となる外力や形状の設定における専門的な知識の必要性が高くなっており、技術の継承するためのツールでもある。

REFERENCES

- 1) 米林甲陽, 久馬一剛, 川口桂三郎: 土壌腐植の形態別画分と易分解性有機物との関係, 易分解性有機物の存在形態(その2), 日本土壌肥料科学雑誌, 第44巻, 第10号, pp.367-371, (1973). [Yonebayashi, K., Kyuma, Kawaguchi, K.: Relationship between morphological fractions of soil humus and biodegradable organic matter, and the forms of presence of biodegradable organic matter (Part 2), Japanese society of soil science and Plant Nutrition, Vol. 44, No. 10, pp. 367-371, 1973.]
- 2) 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, Vol. 25, No. 4, pp. 123-132, 2016. [Taguchi, A.: Redfield ratio: History and current situation of research and future outlook, Oceanography in Japan, Vol. 25, No. 4, pp. 123-132, 2016.]
- 3) 米林甲陽: 我が国の腐植物質研究とその展望, 1.腐植物質研究の成果と問題点, 日本土壌肥料科学雑誌, 第73巻, 第5号. pp.549-554, (2002). [Yonebayashi, K.: Humic substances research in Japan and its prospects, 1. Achievements and problems

- in humic substances research, Japanese society of soil science and Plant Nutrition, Vol. 73, No. 5, pp.549-554, 2002.]
- 4) 平舘俊太郎：土壤腐植物質の化学構造とその機能，土壤の物理性，第105号，pp.23-30，(2007). [Hiradate, S.: Soil Humic Substances : their Chemical Structures and Functions, Japanese society of soil science and Plant Nutrition, No.105, pp.23-30, 2007.]
 - 5) 日比野忠史，永尾謙太郎，松永康司：有機泥の分解を考慮した沈降と堆積過程のモデル化，土木学会論文集B2（海岸工学），Vol. 64, No. 3, pp.202-213, 2008. [Hibino, T., Nagao, K., Matunaga, K.: Modeling of settling and sedimentary processes considering decomposition of organic mud, *Journal of Japan Society of Civil Engineers, Ser. B2 (Coastal Engineering)*, Vol. 78, No. 2, pp.202-213, 2008.]
 - 6) Westrich, J.T. and Berner, R.A.: The role of sedimentary organic matter in bacterial sulfate reduction: The G model tested, *Limnology and Oceanography*, Vol. 29, pp. 236-249, 1984.
 - 7) 須山敦志：ベイズ推論による機械学習入門，講談社，2017. [Suyama, Y. : Introduction to machine learning with Bayesian inference, KODANSHA, 2017.]
 - 8) 坂井友亮，田多一史，広兼元，日比野忠史：底泥に炭素を固定する難分解性物質の形成，土木学会論文集 B1（海洋開発），Vol. 78, No. 2, pp. I_631-I_636, 2022. [Sakai, Y., Tada, Y., Hirokane, H., Hibino, T.: Modeling of settling and sedimentary processes considering decomposition of organic mud, *Journal of Japan Society of Civil Engineers, Ser. B1 (Marine Development)*, Vol. 78, No. 2, pp. I_631-I_636, 2022.]
 - 9) 小池勲夫：海底堆積物における有機窒素化合物の形態とその分解過程，日仏海洋学会，Vol.24, No.2, pp.104-110, 1986. [Koike, I.: Chemical composition and mineralization of organic nitrogen in marine sediment, *Societe Franco-Japonaise d'Océanographie*, Vol.24, No.2, pp.104-110, 1986.]