

# LiNGAM 及び PC アルゴリズムによる 塩分測定に係る因果探索

近藤 健・久保 文雄<sup>1</sup>

## The Causal Discovery of Salinity Measurement Using LiNGAM and the PC Algorithm

Ken KONDO, and Fumio KUBO<sup>1</sup>

### Summary

In the salinity measurements of seawater conducted at this research institute, the number of repeated measurements may increase before determining the final value. To elucidate the causes of the increased number of measurements, we investigated the measurement environment related to temperature and analyzed its relationship with the number of measurements. We employed LiNGAM (Linear Non-Gaussian Acyclic Model) and the PC (Peter and Clark) algorithm, both statistical causal discovery methods, to estimate the causal relationships among factors and constructed a DAG (Directed Acyclic Graph). The results of the causal discovery revealed that the causal order is temperature → number of measurements. Additionally, it was suggested that reducing the range of water sample temperature and room temperature could be effective in minimizing the number of measurements. That is, stabilizing water sample temperature and room temperature could shorten the measurement time.

キーワード: 塩分, 因果探索, LiNGAM, PC アルゴリズム, 有向非巡回グラフ

Keywords: salinity, causal discovery, LiNGAM, PC algorithm, DAG

### I. はじめに

海水の塩分は水温とともに海水の状態を表す基本的な特性量である。塩分の測定は 1960 年代頃までは塩素量を測定し、間接的に塩分が求められ、千分率 (‰) で表されていた。それ以降は、電気伝導度から塩分を求める方法が一般的になり、1980 年にはユネスコにより採択された海水の状態方程式 (Practical Salinity Scale 1978: PSS 78) で定義される実用塩分 (practical salinity) が長く用いられ<sup>1,2)</sup>、測定方法は気象庁の海洋観測指針に示されている<sup>3)</sup>。2009 年に海水の状態方程式が改正され、新たな海水の状態方程式 (Reference Composition Salinity) で定義される絶対塩分が示されたが<sup>4,5)</sup>、同年 IOC の国際海洋データ・情報交換 (IODE: International Oceanographic Data and Information Exchange) の第 20 回会議において、引き続き実用塩分を保管するという勧告 (Recommendation IODE-XX.4) が採択されている<sup>6)</sup>。地方独立行政法人大阪府立環境農林水産総合研究所水産技術センター (水産技術センター) が実施する浅海定線調査をはじめとする各種調査、また同センター地先での定置観測においても実用塩分による測定および記録を行っているため<sup>7)</sup>、本研究は実用塩分について述べることにする。

実用塩分  $S$  は電気伝導度、水温を測定し、PSS78 の定義式により求められる。大気圧下における定義式を (1) に示す。

<sup>1</sup> 元職員

$$S = a_0 + a_1 R_t^{1/2} + a_2 R_t + a_3 R_t^{3/2} + a_4 R_t^2 + a_5 R_t^{5/2} + \Delta S$$

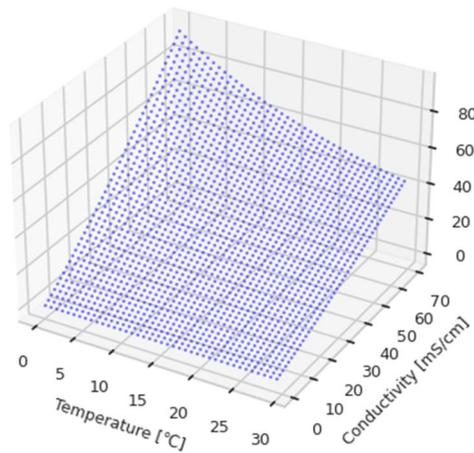
$$\Delta S = \frac{t-15}{1+k(t-15)} (b_0 + b_1 R_t^{1/2} + b_2 R_t + b_3 R_t^{3/2} + b_4 R_t^2 + b_5 R_t^{5/2}) \quad (1)$$

$R_t$ ：大気圧下で温度  $t^\circ\text{C}$  における塩化カリウム標準溶液  
(1 kg 中に 32.4356g の KCl を含んだ水溶液) に対する電気伝導度比

$$\begin{array}{lll} a_0 = 0.0080 & a_1 = -0.1692 & a_2 = 25.3851 \\ a_3 = 14.0941 & a_4 = -7.0261 & a_5 = 2.7081 \\ b_0 = 0.0005 & b_1 = -0.0056 & b_2 = -0.0066 \\ b_3 = -0.0375 & b_4 = 0.0636 & b_5 = -0.0144 \\ k = 0.0162 & & \end{array} \quad \begin{array}{l} \sum_{i=0}^5 a_i = 35.0000 \\ \sum_{i=0}^5 b_i = 0.0000 \end{array}$$

$S$  は無次元の値のため、数値のみで表示されるが、psu (Practical Salinity Units) をつけて表示されることもある。なお、式 (1) から求められる塩分は、 $-2 \leq t \leq 35 (^{\circ}\text{C})$ 、 $2 \leq S \leq 42$  の範囲内で有効である。オープンソースのプログラミング言語である Python のライブラリ gsw を用い、式 (1) における実用塩分、電気伝導度及び水温の関係を第 1 図に示す<sup>8)</sup>。電気伝導度が高く水温が低いほど塩分が高くなる非線形の関係である。電気伝導度は検水の塩分以外に水温にも密接に関係していることがわかる。

水産技術センターでは、鶴見精機社製のサリノメーター T.S.K. DIGI-AUTOMODEL.5 (以下 DA5) で塩分を測定している。DA5 は国際標準の実用塩分の定義に従って測定し、精度は  $\pm 0.005$ 、分解能は 0.001 である。同センターにおける DA5 による測定では、標準海水の付与値  $\pm 0.005$  以内と測定値の真度は高いものの、測定値を決定するまでに指示値にばらつきが生じ、繰り返し測定回数が増えることがある。繰り返し測定回数が増えると、測定に時間を要することになるほか、測定を繰り返すうちに試水が不足し、欠測を招く事態が懸念される。そこで本研究では、温度に関する項目として、分析室の室温、検水温度、DA5 の恒温槽水温及び環境温度 (DA5 が観測し恒温槽水温を決定するために用いられる)、測定回数との関係を相関分析した。また、海水中の懸濁物や植物プランクトンの量による測定回数への影響の有無を確認するため、濁度に加え、植物プランクトンの指標であるクロロフィル  $a$  濃度との関係も相関分析した。さらにこれらの観測変数間の因果関係を明らかにするため、統計的因果探索の手法である線形非ガウス非巡回モデル (LiNGAM : Linear Non-Gaussian Acyclic Model)<sup>9,10)</sup> や PC (Peter and Clark) アルゴリズム<sup>11,12)</sup> を用いての因果探索も行い、因果構造を示す有向非巡回グラフ (DAG : Directed Acyclic Graph) を作成し、その因果構造を明らかにした。



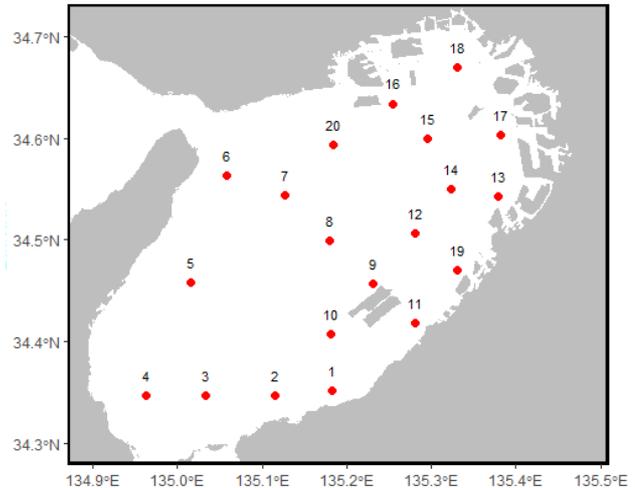
第 1 図. 実用塩分、水温および電気伝導度との関係

電気伝導度を 0.5~70 mS/cm の範囲において 1.5 mS/cm 間隔で、水温を 0~30 $^{\circ}\text{C}$  の範囲において 0.75 $^{\circ}\text{C}$  間隔で、それぞれから実用塩分を求めプロットした (青点)。

## II. 材料および方法

### 1. 材料

塩分測定時の分析室室温の測定には、温度記録計 TR71A (ティアンドディ社) を用いた。検水の水温の測定には、隔測式プローブデジタル温度計 (シンワ測定社) を用いた。塩分の測定には先述の DA5 を用いた。DA5 の標準化には標準海水として IAPSO (International Association for Physical Sciences of the Ocean) Standard Seawater P (Ocean Scientific International 社) シリーズを用いた。海水は 2023 年 11 月から 2024 年 3 月までの間に大阪湾 20 点 (第 2 図) の表層水および底層水を用いた。表層水はバケツ、底層水は北原式採水器あるいはニスキン採水器で採水し、いずれも 250 mL ポリ瓶で密閉の上、分析室内で常温保存した。解析には、温度測定に欠損値のある 2 月の定点 11 の底層及び 3 月の定点 12~20 の底層を除外した 190 検体の測定データを用いた。



第 2 図. 大阪湾における採水地点

図中の赤点は採水地点を、数字は地点番号を表す。

### 2. 方法

#### DA5 による塩分測定と測定環境の観測

検水を DA5 で塩分測定し、塩分が  $\pm 0.001$  に収まる値が 3 回観測された場合、それら 3 つの値における最頻値あるいは最頻値がない場合は中央値を測定値として採用した。測定値が確定されるまでの測定回数、測定値が得られた際の室温、検水の温度、DA5 内の恒温槽の水温及び DA5 が測定した環境温度を記録した。温度の測定値については塩分の測定開始から終了までの平均値、標準偏差及び範囲 (最大値-最小値) を算出し、測定回数との相関を Pearson の積率相関係数により確認した。

#### 濁度、クロロフィル a 濃度の測定

濁度、クロロフィル a 濃度は、検水を採水した地点の深度毎において多項目水質計 RINKO-Profiler (JFE アドバンテック社) を用いて測定した。

#### 非ガウス性検定

データの非ガウス性を確認するために、測定回数、室温、検水の温度、恒温槽水温、環境温度、濁度及びクロロフィル a 濃度のデータに対してシャピロ・ウィルク検定<sup>13)</sup>を行った (有意水準 0.05)。

#### LiNGAM 及び PC アルゴリズムによる因果探索

因果探索手法として用いた LiNGAM とは、一定の仮定の下において、連続値データ間の因果構造を推定する手法である。その構造方程式モデルを式 (2) に示す。

##### 構造方程式モデル

$$x_i = \sum_{j \neq i} b_{ij} x_j + e_i \quad (i=1, \dots, p), (j=1, \dots, p; j \neq i) \quad (2)$$

$x$  : 観測変数  $b_{ij}$  : 係数  $e_i$  : 誤差変数

観測変数  $x_i$  はその変数以外の観測変数  $x_j$  とその誤差変数  $e_i$  の線形和からなる。これを行列表現したものを式 (3) に示す。

##### 行列表現

$$\begin{aligned} x &= Bx + e \\ \downarrow \\ x &= Ae \\ A &= (I - B)^{-1} \end{aligned} \quad \begin{array}{l} \mathbf{x}: \text{観測変数ベクトル} \\ \mathbf{B}: \text{係数行列} \\ \mathbf{e}: \text{誤差変数ベクトル} \\ \mathbf{A}: \text{混合行列} \end{array} \quad (3)$$

B は  $i \times j$  の係数行列である。  $x = Bx + e$  を  $x$  について解くと得られる行列 A は、独立成分分析を用いて混合行列として識別可能であり、混合行列 A から係数行列 B を導くことができる。係数行列が明らかになれば因果構造を一意に推定することが可能となる。観測変数は、塩分の測定回数、環境温度範囲、恒温槽水温範囲、室温範囲、検水温度範囲、塩分、濁度及びクロロフィル  $a$  濃度の実測値をそれぞれ標準化（平均 0、標準偏差 1 になるようスケール）した値を用いた。標準化にはプログラミング言語 Python 及び機械学習用のオープンソースライブラリ scikit-learn<sup>14)</sup> を用いた。因果探索は Python の機械学習用オープンソースライブラリ lingam を用い、ICA-LiNGAM (Independent Component Analysis)<sup>15)</sup> により計算した。モデルの混合行列 A は統計的手法である独立成分分析<sup>16,17)</sup>を用いて推定され、A を置換・正規化し、求められた係数行列 B の成分のゼロ・非ゼロパターンから DAG を作成した。なお、本研究における LiNGAM では、観測変数から得られた構造方程式が線形であること、誤差変数が非ガウス分布であること、塩分は電気伝導度と水温の 2 変数に依存することから未観測変数はないことと仮定した。

LiNGAM は、関数形にも外生変数の分布にも仮定を置くセミパラメトリックな手法として知られているが、本研究ではさらに関数形にも外生変数の分布にも仮定を置かないノンパラメトリックな手法として代表的な PC アルゴリズムを用いての因果探索も行い、LiNGAM との因果構造を比較した。PC アルゴリズムは主に離散値を対象としたベイジアンネットワークによる因果探索手法の一つである。すべてのノード間に無向エッジを接続した完全グラフから開始し、2 変数が互いに独立、または他の変数で条件付けしたときに独立となる場合、エッジを除外する。まだ隣接しているすべての変数の組に対して、1 つの変数で条件付けして独立性を検定する。このプロセスをすべての変数のペアで実施し、無向グラフを得る。得られた無向グラフはオリエンテーションルール<sup>18,19)</sup>による方向づけを行い DAG を構築する手法である。PC アルゴリズムは、Python のオープンソースライブラリ gCastle<sup>20)</sup> により因果探索を行った。観測変数を離散値として扱うため、データの前処理として連続値データをそれぞれ等幅分割法（順序尺度）により 5 つのビンに区切り、1～5 の離散値とした。条件付き独立性の検定にはカイ二乗検定を用いた（有意水準 0.05）。さらに対象とする観測変数は、測定回数を除いていずれも本来は連続値のため、ビン分割は行わず独立性の検定に偏相関係数による Fisher の Z 検定を用いることで連続値を対象とした PC アルゴリズムによる因果探索も行った。その際、観測変数は LiNGAM 同様に標準化した値を用いた。

### Ⅲ. 結果および考察

#### 1. 測定回数と各観測項目との相関

190 検体の塩分の測定回数における最小値は 3 回、最大値は 29 回、平均値は 10 回、中央値は 9 回であった。塩分の測定回数と、測定値を確定するまでの環境温度、恒温槽水温、室温、検水温の観測値の平均値、標準偏差及び範囲との相関係数、そして、塩分の測定回数と、塩分、濁度及びクロロフィル  $a$  濃度との相関係数を第 1 表に示す。検水温の範囲の相関係数が 0.6443 と最も高く、次に検水温の標準偏差との相関係数が高く 0.5783 であった。恒温槽水温を除いて、標準偏差よりも範囲との相関が高い傾向が見られた。なお、恒温槽水温の標準偏差と範囲との相関係数は、それぞれ 0.2024 と 0.1386 といずれも低くほとんど差がなかった。これらの結果から、因果探索に用いる温度の変数として、より相関係数が高かった範囲を解析に用いることとした。

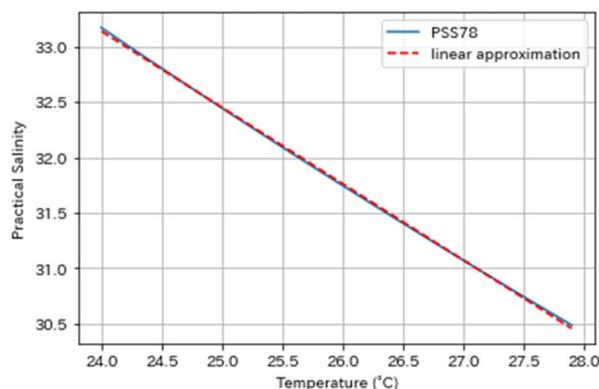
第 1 表. 測定回数と各観測項目との相関係数

観測項目	相関係数	
環境温度	平均値	0.0427
	標準偏差	0.1538
	範囲	0.3143
恒温槽水温	平均値	-0.0102
	標準偏差	0.2024
	範囲	0.1386
室温	平均値	0.0654
	標準偏差	0.1889
	範囲	0.2852
検水温	平均値	-0.0323
	標準偏差	0.5783
	範囲	0.6443
塩分	0.0292	
濁度	-0.0206	
クロロフィルa	-0.1079	

## 2. データの線形性と非ガウス性の確認

LiNGAM において、一般的に非線形な関係の変数は適用できないため、塩分と温度間における線形性の確認を行った。観測した室温は最小値が 24.92°C、最大値が 26.20°C、最も範囲の大きかった試水の水温は最小値が 24.8°C、最大値が 27.7°C だった。いずれも温度は調節されているため、その範囲は大きくない。PSS78 の定義式から、電気伝導度 50 mS/cm、1 気圧での塩分と温度 (24~28°C) の関係を第 3 図に示す。この温度範囲においては、塩分と温度の関係は線形近似し、LiNGAM の前提である線形性は満たされると考えられる。なお、その他の変数間の関係は線形性が未知であるが、線形性があるものと仮定した。

次に LiNGAM に用いるデータの非ガウス性を確認するため、測定回数、室温範囲、検水温度範囲、恒温槽水温範囲、環境温度範囲、塩分、濁度及びクロロフィル a 濃度についてシャピロ・ウィルク検定を行った。その結果、いずれも p 値が 0.05 未満であり、データがガウス分布であるという仮説は棄却された。LiNGAM の前提である非ガウス性は満たされると考えられる。



第 3 図. 一定条件下 (電気伝導度 50 mS/cm, 1 気圧) における電気伝導度と温度の関係

青線は PSS78 の定義式、赤破線は PSS78 の定義式を線形近似させたもの。縦軸は実用塩分、横軸は温度。

### 3. LiNGAM による因果探索

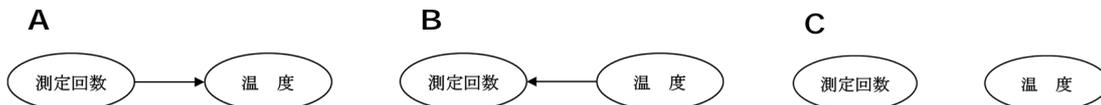
相関分析は変数間の関係を示すためによく用いられるが、ある変数間で相関関係が認められたとしても、影響の方向は判明できず、変数間の因果関係や因果順序といった因果構造を同定することはできない。そこで、測定回数と温度等の因果関係の有無、因果関係がある場合はその因果順序、すなわち測定回数が多いから時間を要しその間に温度変化が生じたのか、温度変化が生じたから測定値が安定せず測定回数が多くなったのか（第 4 図）を明らかにするため、LiNGAM を用いて変数間の因果関係を推定した。

測定回数、塩分、室温範囲、検水温度範囲、恒温槽水温範囲、環境温度範囲、濁度及びクロロフィル *a* 濃度について LiNGAM による因果探索を行った結果、行列 **B** が得られた。得られた行列 **B** を式 (3) に代入したものが式 (4) である。

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0.200 & 0 & 0 & 0.607 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.746 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -0.217 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \end{bmatrix} \quad (4)$$

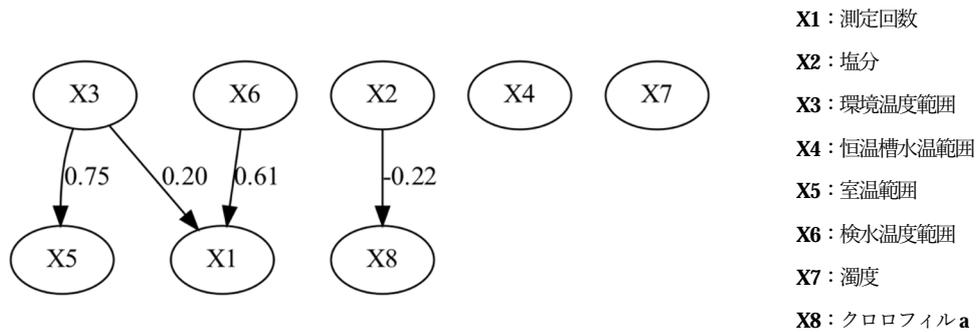
$x_1$  は測定回数、 $x_2$  は塩分、 $x_3$  は環境温度範囲、 $x_4$  は恒温槽水温範囲、 $x_5$  は室温範囲、 $x_6$  は検水温度範囲、 $x_7$  は濁度、 $x_8$  はクロロフィル *a* 濃度を示す。 $e$  はそれぞれの誤差項を示す。式 (4) の係数行列を因果的順番が早い（行成分に 0 が多い）順、対角成分が 0、下三角行列になるよう並べ替え、行成分と列成分がいずれも 0 の変数は全ての行列から削除し（他の変数と因果関係を持たないため）、係数行列の並べ替えに伴い目的変数、説明変数、及び誤差項のベクトルを並び替えたものが式 (4)' である。

$$\begin{bmatrix} x_2 \\ x_3 \\ x_5 \\ x_6 \\ x_1 \\ x_8 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.746 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.200 & 0 & 0.607 & 0 & 0 \\ -0.217 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_2 \\ x_3 \\ x_5 \\ x_6 \\ x_1 \\ x_8 \end{bmatrix} + \begin{bmatrix} e_2 \\ e_3 \\ e_5 \\ e_6 \\ e_1 \\ e_8 \end{bmatrix} \quad (4)'$$



第 4 図. 測定回数と温度の想定される因果関係

図中の楕円は要因（ノード）を、矢印は因果の方向（エッジ）を示す。図はそれぞれ、測定回数が温度に影響を与えている(A)、温度が測定回数に影響を与えている(B)、及び測定回数と温度が独立した状態(C)を示す。



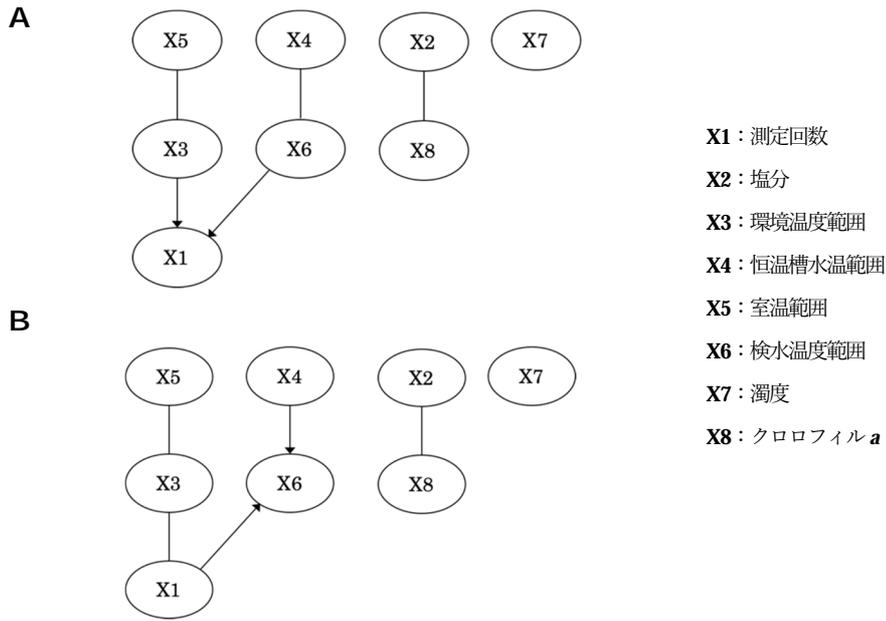
第 5 図. LiNGAM による測定回数, 塩分, 環境温度範囲, 恒温槽水温範囲, 室温範囲, 検水温度範囲, 濁度及びクロロフィル a 濃度の因果関係を示した有向非巡回グラフ

楕円はノード, 矢印は因果順序を示した有向エッジ, 数値はノード間の係数を表す.

式 (4) をもとに作成した DAG を第 5 図に示す. 測定回数の子ノードとして, 環境温度範囲及び検水温度範囲の 2 つの親ノードを上流とするエッジが示された. すなわち, 測定回数は環境温度範囲及び検水温度範囲の影響を受ける関係にあることを示している. 検水温度範囲の係数は 0.607 であり, 検水温度範囲が上昇すると測定回数が上昇する関係にあることを示している. 環境温度範囲の係数は 0.200 であり, 環境温度範囲が上昇すると測定回数が上昇することが示された. 室温範囲の親ノードとして環境温度範囲が示されたが, 環境温度範囲は DA5 が計測する室温であるため, 任意の値に設定することはできず, 室温に依存するという外部知識により, このエッジは適切でないと判断できる. また, 塩分, 恒温槽水温範囲, 濁度及びクロロフィル a 濃度は測定回数とはエッジがなく, 独立していることがわかった. これらの項目は測定回数の変動に寄与しないといえる. 恒温槽水温範囲は測定に影響を与えるように思われるが測定回数とは独立していた. 恒温槽は DA5 の内部にあり, 検水温度を一定に保持するための装置である. 恒温槽水温は DA5 の観測した環境温度から決定され, 6 L の水が入った恒温槽が加温されるため, 室温に変化が生じた際, 恒温槽水温の変化までには時間を要する. 他の温度に関する項目に比べ, 恒温槽温度変化が測定に影響を与えるまでの時間差が比較的大きく, 測定回数との間にエッジが検出されにくかったことが理由として考えられる. なお, 塩分を親ノード, クロロフィル a 濃度を子ノードとして係数が -0.217 であった. 湾口部や底層では塩分が高くなり, そのような検水ではクロロフィル a 濃度 (植物プランクトン) が低くなることを示したものと考えられる. 以上のことから LiNGAM を用いた因果探索では, 因果の順序としては温度→測定回数の順であることがわかった. また, 測定回数を少なくするには, 環境温度範囲及び検水温度範囲を小さくすることが有効であり, 特に検水温度範囲の寄与が最も大きかったことが示された.

#### 4. PC アルゴリズムによる因果探索

標準化した測定回数, 塩分, 室温範囲, 検水温度範囲, 恒温槽水温範囲, 環境温度範囲, 濁度及びクロロフィル a 濃度について, PC アルゴリズムによる因果探索を行い, 得られた因果グラフを第 6 図に示す. なお, 変数を離散値として扱った際には表 2 の通り 5 つのビンに分割した. 第 6 図の A は連続値を用いて得られたグラフ, B は離散値を用いて得られたグラフである. A, B ともにスケルトン構造は共通していた. いずれの結果も測定回数と環境温度範囲及び検水温度範囲ノード間, 環境温度範囲と室温範囲ノード間及び恒温槽水温範囲と検水温度範囲ノード間でエッジが検出された. また, 塩分, 濁度, クロロフィル a 濃度の 3 つのノードは測定回数ノードとは独立した関係であったが, 塩分とクロロフィル a 濃度のノードはエッジが検出された. A は環境温度範囲と検水温度範囲から測定回数へ向かう 2 本の有向エッジが検出された. B は A と共通する構造であったが, 測定回数と恒温槽水温範囲から検水温度範囲へ向かう 2 本の有向エッジが検出された. 測定回数と検水温度範囲の有向エッジが A と B で逆を示した. 逆方向のエッジが検出された理由としては, 変数を分割する前処理を加え離散値として扱ったことが影響したと考えられる. また, 連続値と離散値では独立性検定が異なるので, これらの影響により A と B では因果の方向が異なると考えられる. 他にも A では環境温度範囲から測定回数ノードへの有向エッジが B では無向エッジになっており, B では恒温槽水温範囲から検水温度範囲ノードへの有向エッジが A では無向エッジになっていた. いずれも先述の前処理や検定の影響によるものと考えられる.



**第6図. PC アルゴリズムによる測定回数, 塩分, 環境温度範囲, 恒温槽水温範囲, 室温範囲, 検水温度範囲, 濁度及びクロロフィル a 濃度の因果関係を示した因果グラフ**

楕円はノード, 矢印は因果の方向, 線は因果関係を示したエッジを表す. 変数に連続値を用いて作成したグラフ(A)と離散値を用いて作成したグラフ(B).

**第2表. 各変数のビン分割による離散値化**

ビン番号	x1: 測定回数	x2: 塩分	x3: 環境温度範囲	x4: 恒温槽水温範囲	x5: 室温範囲	x6: 検水温度範囲	x7: 濁度	x8: クロロフィル a
1	$x1 \leq 8$ (92)	$23.2700 < x2 \leq 25.2736$ (3)	$0 \leq x3 \leq 0.4$ (92)	$0 \leq x4 \leq 2$ (74)	$0 \leq x5 \leq 0.18$ (188)	$0 \leq x6 \leq 0.32$ (146)	$0.12 < x7 \leq 2.07$ (166)	$-3.522 < x8 \leq 0.882$ (160)
2	$9 \leq x1 \leq 13$ (62)	$25.2736 < x2 \leq 27.2672$ (1)	$0.4 < x3 \leq 0.8$ (65)	$2 < x4 \leq 4$ (55)	$0.18 < x5 \leq 0.36$ (1)	$0.32 < x6 \leq 0.64$ (32)	$2.07 < x7 \leq 4.01$ (18)	$0.882 < x8 \leq 5.264$ (22)
3	$14 \leq x1 \leq 18$ (19)	$27.2672 < x2 \leq 29.2608$ (6)	$0.8 < x3 \leq 1.2$ (24)	$4 < x4 \leq 6$ (54)	$0.36 < x5 \leq 0.54$ (0)	$0.64 < x6 \leq 0.96$ (6)	$4.01 < x7 \leq 5.95$ (3)	$5.264 < x8 \leq 9.646$ (6)
4	$19 \leq x1 \leq 23$ (11)	$29.2608 < x2 \leq 31.2544$ (11)	$1.2 < x3 \leq 1.6$ (7)	$6 < x4 \leq 8$ (4)	$0.54 < x5 \leq 0.72$ (0)	$0.96 < x6 \leq 1.28$ (3)	$5.95 < x7 \leq 7.89$ (2)	$9.646 < x8 \leq 14.028$ (1)
5	$24 \leq x1 \leq 29$ (6)	$31.2544 < x2 \leq 33.2480$ (169)	$1.6 < x3 \leq 2.0$ (2)	$8 < x4 \leq 10$ (3)	$0.72 < x5 \leq 0.9$ (1)	$1.28 < x6 \leq 1.60$ (3)	$7.89 < x7 \leq 9.83$ (1)	$14.028 < x8 \leq 18.410$ (1)

表中の ( ) 内の数値はビンに含まれる標本数を示す.

LiNGAM で得られた DAG と比較すると, 測定回数, 環境温度範囲, 室温範囲及び検水温度範囲ノード間を結ぶエッジのパターンは同じであり, 類似した構造を示している. しかしながら, LiNGAM ではそれらの構造から恒温槽水温範囲は独立しているほか, ノード間の方向もそれぞれ異なっている. 測定回数ノードの2本のエッジに着目すると, 連続値を用いる LiNGAM と連続値を用いた PC アルゴリズムのエッジの方向は2本とも測定回数への有向エッジが示されているが, 離散値を用いた PC アルゴリズムでのみ環境温度範囲と測定回数ノード間に無向エッジが, 測定回数から検水温度範囲への有向エッジが示された. 連続値を用いた際の有向エッジが共通していることから, 連続値を無理に離散化することで, 元のデータの特徴を損ねたことが考えられる. 例えば第2表の室温範囲や検水温度範囲では, 標本数の多くがビン番号1に含まれており, ビン番号1に含まれる標本のもつ情報が欠落してしまっていると考えられる. 従って, LiNGAM と連続値を用いた PC アルゴリズムにおいて示された DAG が妥当と考えられる. これらの DAG で共通して示された有向エッジは, 測定回数を子ノードとして, 環境温度範囲及び検水温度範囲の2つの親ノードを上流とするエッジであり, 因果の順序としては, 温度→測定回数の順であることを示し, 測定回数を少なくするには, 環境温度範囲及び検水温度範囲を小さくすることが有効であることが示された.

## 5. まとめ

LiNGAM 及び PC アルゴリズムの解析結果から、塩分の測定回数は環境温度範囲や検水温度範囲といった温度範囲が影響を与えていることが、その因果順序とともに示すことができた。

LiNGAM では測定回数に影響を与える主要な要因として、検水温度範囲の影響の大きさを定量的に示すことができた。検水温度範囲は電気伝導度に密接に関連し、直接的に塩分測定に影響することが考えられるので、合理的で理解しやすい結果であった。室温範囲が環境温度範囲の子ノードとなっていたが、外部知識によりその方向を排除した。DAG では示されなかったが、環境温度範囲は室温を計測して決定されるので、環境温度範囲を小さくするためには室温を一定に保つことが重要であると考えられる。塩分、濁度及びクロロフィル *a* 濃度も測定回数とは独立しており、検体の水質が測定回数に影響を与えている可能性は低いと考えられた。PC アルゴリズムは、ビン分割せずに連続値を用いることで LiNGAM により示された DAG と多くの点で共通した因果構造を示した。

本研究の結果、DA5 を用いた塩分測定を速やかに行うためには、検水温度及び測定室温の変動を抑制し、温度範囲を低く維持することが重要であると考えられる。DA5 のマニュアルには試水を室温に十分 ( $\pm 2^{\circ}\text{C}$  以内) なじませておく旨が記載されており、これまでの塩分測定においても、すでに温度管理には一定の注意が払われていたが、本研究により測定回数と温度との関係が定量化、視覚化されたことで、検水と室温の温度管理がより一層求められることが示された。本研究とは別に温度と塩分に関わる研究を進めているが、検体を保管する室内の位置によって検水の温度差に影響を与えるデータが出ているところである。例えば、冬季では室内の高さによって温度差があること、暖房の送風の向きによる温度差の影響などである。これらを考慮した検水と室温の管理が品質改善につながると考えられ、本研究から今後の精度管理及び測定時間の短縮につながる情報が得られたといえる。LiNGAM を使用する際には、セミパラメトリック手法としていくつかの条件を満たす必要があるが、それを満たしていれば強力な因果探索ツールになることが示唆された。連続値を用いた PC アルゴリズムも同様である。本研究で用いた検水は 11 月から 3 月の冬季のものであり、赤潮が発生しやすい夏季の検水を使用していないため、比較的クロロフィル *a* 濃度の低い検水を使用していたといえる。一年を通した検水を用いてデータを蓄積し因果探索を行うと、より確からしい因果関係が明らかになるかもしれない。

## IV. 摘要

塩分測定における測定回数増加の原因を明らかにするため、温度に関連する測定環境を調査し、測定回数との関係を解析した。統計的因果探索の手法である LiNGAM と PC アルゴリズムを用いて要因間の因果関係を推定した結果、因果順序は温度→測定回数の順であることがわかった。また、測定回数を少なくするには、検水温度範囲及び測定機器の環境温度範囲を小さくすることが有効であることが示唆された。

## V. 引用文献

- 1) Fofonoff, P. and Millard, R.C. (1983). Algorithms for computation of fundamental properties of seawater. *UNESCO Tech. Pap. in Mar. Sci.*, 44, 53.
- 2) Lewis E. L. (1980). The Practical Salinity Scale 1978 and its antecedents. *J. Ocean. Eng.*, 5(1). 3-8.
- 3) 気象庁 (1999). 第 5 章 水温・塩分の観測. 海洋観測指針, 第 1 部, 38-48.
- 4) McDougall, T. J., Jackett, D. R. and Millero, F. J. (2009). An algorithm for estimating Absolute Salinity in the global ocean. *Ocean. Sci. Discuss* 6, 215-242.
- 5) 河野健 (2010). 新しい海水の状態方程式と新しい塩分 (Reference Composition Salinity) について. *海の研究*, 19(2), 127-137.
- 6) The IOC Committee on International Oceanographic Data and Information Exchange (2009). Recommendation IODE-XX.4. [https://iode.org/index.php?option=com\\_content&view=article&id=155&Itemid=0](https://iode.org/index.php?option=com_content&view=article&id=155&Itemid=0) (2025 年 2 月アクセス確認)

- 7) 地方独立行政法人大阪府立環境農林水産総合研究所水産技術センター事業資料集. 気象・海象の定置観測. [https://www.knsk-osaka.jp/publication/suisan\\_shiryo/index.html](https://www.knsk-osaka.jp/publication/suisan_shiryo/index.html). (2025 年 2 月アクセス確認)
- 8) TEOS-10 developers. GSW-Python. <https://teos-10.github.io/GSW-Python/intro.html>. (2025 年 2 月アクセス確認)
- 9) Shimizu, S., Hoyer, P. O., Hyvärinen, A. and Kerminen, A. J. (2006). A linear non-gaussian acyclic model for causal discovery. *J. Mach. Learn. Res.*, 7, 2003-2030.
- 10) 清水昌平 (2017). 統計的因果探索. 講談社.
- 11) Spirtes, P., Glymour, C. N., Scheines, R., and Heckerman, D. (2000). *Causation, prediction, and search*. MIT press.
- 12) 植野真臣 (2013). ベイジアンネットワーク. コロナ社.
- 13) Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality. *Biometrika*, 52, 3 and 4, 591-611.
- 14) Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *J. Mach Learn. Res.*, 12, 2825-2830.
- 15) Ikeuchi, T., Haraoka, G., Ide, M., Kurebayashi, W. and Shimizu, S. Lingam's documentation. [https://lingam.readthedocs.io/en/latest/reference/ica\\_lingam.html](https://lingam.readthedocs.io/en/latest/reference/ica_lingam.html). (2025 年 2 月アクセス確認)
- 16) Comon, P. (1994). Independent component analysis – a new concept? *Signal Processing*, 36, 287-314.
- 17) Hyvarinen, A., Karhunen, J. and Oja, E. (2001). *Independent Component Analysis*. Wiley Interscience.
- 18) Verma, T. and Pearl, J. (1992). An algorithm for deciding if a set of observed independencies has a Causal Explanation. *Proc. Conf. on Uncertainty in Artif. Intell.*, 323-330.
- 19) Meek, C. (1995). Causal Inference and Causal Explanation with Background Knowledge. *Proc. Conf. on Uncertainty in Artif. Intell.*, 403-410.
- 20) Kalisch, M. and Bühlmann, P. (2005). Estimating high-dimensional directed acyclic graphs with the PC-algorithm. ar Xiv. <https://arxiv.org/abs/math/0510436>. (2025 年 2 月アクセス確認)