

Convergent cross mappingの紹介

生態学における時系列間の因果関係推定法

誌名	日本生態學會誌
ISSN	00215007
著者名	中山,新一朗 阿部,真人 岡村,寛
発行元	日本生態学会暫定事務局
巻/号	65巻3号
巻号補足	
掲載ページ	p. 241-253
発行年月	2015年11月

農林水産省 農林水産技術会議事務局筑波産学連携支援センター

Tsukuba Business-Academia Cooperation Support Center, Agriculture, Forestry and Fisheries Research Council
Secretariat



総説

Convergent cross mapping の紹介：生態学における時系列間の因果関係推定法

中山 新一朗^{1,*}・阿部 真人^{2,**}・岡村 寛¹

¹ 水産総合研究センター中央水産研究所・² 東京大学大学院広域システム科学系

An introduction to convergent cross mapping: a novel method for causality detection in ecological time series

Shin-Ichiro Nakayama^{1,*}, Masato S. Abe^{2,**} and Hiroshi Okamura¹

¹National Research Institute of Fisheries Science, Fisheries Research Agency,

²Department of General Systems Studies, Graduate School of Arts and Sciences, University of Tokyo

要旨: 生態学者はしばしば、複数の時系列データからそれらの間に存在する因果関係を推定する必要に迫られる。しかし、生物学的事象は決定論的かつ非線形で複雑な過程を背景に持つのが一般的であり、そのような応答から生じた時系列データ間の因果関係を推定するのは非常に困難である。本稿では、このような状況において有効である因果関係推定法である Convergent cross mapping について、その仕組み、使い方、将来の課題等を解説する。

キーワード: 因果関係、Granger 因果、決定論的力学系、非線形、時系列解析

Abstract: Ecologists frequently need to detect causalities among events from time series data. It can be difficult to detect causality from time series data created from complex, deterministic, nonlinear systems, which are universal. Convergent cross mapping is a novel method for detecting causality in such situations. This review explains the mechanism, how to use the method, and future issues with this method.

Keywords: Causality, deterministic dynamical systems, Granger causality, non-linearity, time series analyses

1. はじめに

天候は作物の出来不出来を決定し、経済を動かす。網膜に結ばれた異性の像はニューロン発火の連鎖を引き起こし、人は恋に落ちる。この世の事象は因果で結ばれている。事象間の因果を一つ一つ解明していくことで人類は世界の姿を描いてきた。事象間に存在する因果の姿を知ることは、現代においてもなお基礎科学に携わる者の大きな目的の一つである。

「因果」、あるいは「原因」と「結果」という言葉の厳密な定義はアリストテレスの時代から議論されてきたが、

一般に、「A は B の原因である」あるいは「B は A の結果である」といった場合には、「A の挙動は B の挙動に影響する」と言い換えても差し支えないだろう。本稿では、以後「因果」や「原因」、「結果」という言葉をこの意味で用いることとする。このとき事象 A、B 間の関係は 4 パターン、すなわち① A が B の原因（の一つ）であり、かつ B は A に影響しない、② B が A の原因（の一つ）であり、かつ B は A に影響しない、③それぞれが他方の原因、及び結果である、④両者の間には因果が存在しない、に分類できる。目的とする事象が上記のどのパターンに属するのかを確かめるための有効な手段の一つは、それぞれの事象に実験的に変化を加えることである。例えば、ある植物の開花数と気温に注目するときには、実験的に気温を操作し、開花数に変化が見られるかどうかを観察することによって気温が開花数の原因となっているかを確かめることができる。もちろんその逆、開花数を（花

2015 年 1 月 27 日受付、2015 年 7 月 6 日受理

第一・第二著者は本総説の執筆に等しく貢献した。

* 責任著者: 中山新一朗 e-mail: shinichironak@affrc.go.jp

** 現所属: 国立情報学研究所

Present address: JST, ERATO, 河原林巨大グラフプロジェクト

をちょん切る事によって) 変化させて気温への影響を見ることが可能であるが、開花数が気温に影響しないことは明白であろう。とにかく、一方の事象に実験的に変化を加え、他方への影響を見ることによって事象間の因果の有無を推定することができる。

一方で、多くの事象は人間が実験的に操作を加える事が不可能である。特にスケールの大きな事象、地球規模で起こる気象現象や自然界における生物の個体数の増減などに実験的变化を加える事は非常に困難である。このような場合、我々は過去の時系列データからそれらの間に存在する因果関係を推定しなければならない。

ある時刻の系の状態が決定論的な微分方程式や差分方程式で表されるとき、その系を力学系とよぶ。例えば2つの変数 X, Y が $X_t = f(X_{t-1})$, $Y_t = g(X_{t-1}, Y_{t-1})$ に従うとき、これらは X が Y に一方的に影響を与える力学系である。現在、時系列間の因果関係を推定する方法として広く使われているのは Granger 因果性テスト (Granger 1969) である。これは片方の時系列の情報が他方の時系列の予測精度に寄与するかどうかで因果性を判断する方法であり、確率過程に強く支配される系については有効であるが、決定論的過程に強く支配される力学系においてはうまく因果を推定できないことがある (Granger 1969)。厄介なことに、我々の興味の対象となる生物学的事象の多くは (程度の差はあれ) 決定論的過程を背景にもつ。細胞内における酵素反応速度のようなミクロスケールのものから個体群成長における密度効果のようなマクロスケールのものまで、生物学に携わるものであれば誰でも身近なところから決定論的過程に強く支配される事象を見つけ出すことができるだろう。また、環境や刺激に対して非線形な応答を示すのも生物の特徴である (例えば微生物の増殖速度は温度に対して直線的に増加するわけではない)。非線形な力学系は確率過程によく似た挙動を示す事があり、両者を見分けるためにはそれなりの手続き (4.2. 節) を踏む必要がある。これらの理由から、生物学的データに盲目的に Granger 因果性テストを適用するのは危険である。決定論的過程に強く支配される力学系に対しても適用可能な因果関係推定法がないわけではない。移動エントロピー (transfer entropy, Schreiber 2000) に代表される情報理論的方法がそれである。しかしこれらの方法は Granger 因果性テストに比べて多くのデータを必要とする傾向があり (Pereda et al. 2005)、生態学データはデータ収集に時間と労力がかかることを考えると、適用困難となる場合が多いだろう。

本稿で紹介する Convergent cross mapping (CCM、

Sugihara et al. 2012) は、Granger 因果性テストが苦手とする決定論的で、なおかつ非線形な力学系において有効で、かつ比較的少ないデータに対して適用可能な因果性検定法として Science 誌上で発表された。本稿ではそのアルゴリズム、適用対象の解説、検定法についての提案を試みる。第2章では CCM の理論的背景と具体的なアルゴリズムについて解説する。2.1. 節は CCM の基本的な理論の解説であり、これを理解することで 2.2. 節以降の数理的記述を理解することが容易になると思われる。そのため、数理的な記述に馴染みの薄い読者はまず 2.1. 節を重点的に理解することが重要だろう。第3章で結果の出力と検定法について解説する。第4章では時系列データの準備と CCM の適用可能性のチェックといった、実際の解析手順について解説する。正しい手順を踏まずに CCM を用いると誤った結果をもたらすことがあるため、この章は重要である。第5章では先に述べた Granger 因果性テストと CCM の理論的な違いについて解説する。第6章では現段階での CCM の適用例、課題、CCM をベースとした発展的手法について紹介する。本稿が多くの人々の理解の助けとなれば幸いである。

2. Convergent cross mapping

CCM で2つの事象の時系列データからそれらの間に存在する因果関係を検出する際、2つの事象間には決定論的で非線形な力学系が存在することが仮定される。ある変数の状態を表す式中に他の変数が現れる時、後者は前者の原因の一つである可能性があるが、観察された時系列データからその背後にどのような力学系があるのかを知ることは難しい。CCM は model-free (背後にある力学モデルに仮定を置く必要がない) な手法であるため、力学系の形が予めわかっている必要はない。本章では CCM の理論的背景と具体的なアルゴリズムについて解説する。

2.1. 理論的背景

本節では、CCM の基本的な考え方の直感的な説明を与える。例として、ネズミの個体数と桶の売上について考えよう。ある年のネズミの個体数は、前年の個体数のみによって決まる (ここでは決定論的過程を対象としているため、ノイズは考えない)。 t 年におけるネズミの個体数を X_t と表すと、 $X_t = f(X_{t-1})$ と書ける。ネズミの個体数は、桶の売上価格には影響されないとする。また、桶の売上は前年度の売上 (たとえば、よく売れた年の翌年は売れない、など) とネズミの個体数 (桶を齧ってしまう) に

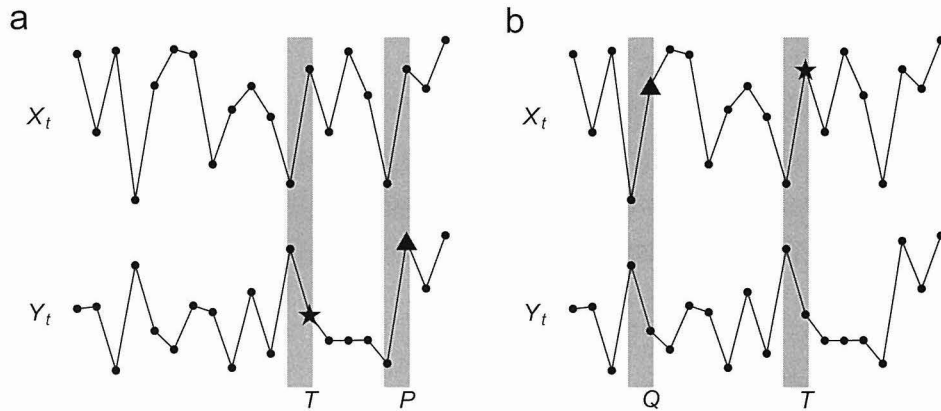


図 1. CCM の仕組みの模式図。(a) Y_T を類推する場合。(b) X_T を類推する場合。

よって決定する。 t 年における桶の売上を Y_t と表すと、 $Y_t = g(X_{t-1}, Y_{t-1})$ と書ける。ここでネズミの個体数は原因、桶の売上は結果となっている。ネズミの個体数は密度効果などによって、また桶の売上はその相場や需要が絡み合うことによって、ともに非線形な（そして複雑な）挙動を示すだろう。そのため、桶屋はネズミの個体数と桶の売上がどのような関係にあるのかを直接知ることはできない。桶屋にとっては、ネズミと桶は無関係かもしれないし、両者が互いに影響を及ぼしているかもしれない。もしネズミの個体数が桶の売上に影響することが分かれば、前年度の桶の売上と前年度のネズミの個体数から今年度の売上を予想することができるかもしれない。あるいは、ネズミの数を操作して売上を増やすことさえも。

ここでの目的は、ネズミの個体数および桶の売上の時系列から、両者の間の因果関係を（前者は後者に影響し、その逆はないと）正しく推定することである。そのため、まず、 T 年の桶の売上（ Y_T 、図中★）が観測されていないと仮定し、それを以下の方法で見積もる場合を考える（図 1a）。

1. ネズミの個体数が T および $T-1$ 年（図中左のグレーの年度）と一番近い値を示す連続した 2 年（ P および $P-1$ 年とする）を探す（図中右のグレーの年度）。
2. P 年の桶の売上（ Y_P 、図中▲）を Y_T の見積もりとする。

このように、類似したパターンを参照することによる予測を類推とよぶことにする。この場合は原因（ネズミの個体数）から結果（桶の売上）の類推であるため、直感的に正確に類推することができると予想する読者も多いだろう。しかし桶の売上は前年の売上からも影響を受ける（ $Y_t = g(X_{t-1}, Y_{t-1})$ ）ため、たとえば $T-1$ 年と $P-1$ 年でネズミの個体数が近い値だったとしても T 年と P 年の

桶の売上が近い値となるとは限らない。すなわち、（直感に反して）この類推はうまくいかないのである。

では逆に、 T 年のネズミの個体数（ X_T 、図中★）が観測されていないと仮定し、それを以下の方法で見積もる場合を考えよう（図 1b）。

1. 桶の売上が T および $T-1$ 年（図中右のグレーの年度）と一番近い値を示す連続した 2 年（ Q および $Q-1$ 年とする）を探す（図中左のグレーの年度）。
2. P 年のネズミの個体数（ X_Q 、図中▲）を X_T の見積もりとする。

桶の売上は前年度の売上と前年度のネズミの個体数によって決定する（ $Y_t = g(X_{t-1}, Y_{t-1})$ ）ため、 T 年と Q 年および $T-1$ 年と $Q-1$ 年で桶の売上が近い値となった時には $T-1$ 年と $Q-1$ 年のネズミの個体数も近い値であったはずである。そして $T-1$ 年と $Q-1$ 年のネズミの個体数が近い値であった時、 T 年と Q 年のネズミの個体数も近い値となる。ネズミの個体数は自身の前年度の個体数のみによって決まる（ $X_t = f(X_{t-1})$ ）からである。このため、この類推はある程度高い正確性で行うことができると考えられる。

この例からわかるように、決定論的な力学系から生じた時系列においては結果から原因の類推は正確に行うことができるが、逆の類推は不正確である。CCM はこの原理を利用し、2 本の時系列を互いに類推した時の正確性を測ることで因果関係の有無を判定する方法である。類推の正確性は、類推した時系列と実際の時系列の平均二乗誤差や相関係数で評価する。本稿では一貫して相関係数を採用している。事象 A と B の間の 4 つのパターンの関係（第 1 章）に対応して得られる結果は以下のとおりである。

1. A が B の原因（の一つ）であり、かつ B は A に影響しない場合：A の時系列から B の時系列を類推した場合には無相関となるが、逆の場合には正の相関係数が期待される。
2. B が A の原因（の一つ）であり、かつ A は B に影響しない場合：A の時系列から B の時系列を類推した場合には正の相関係数が期待されるが、逆の場合には無相関となる。
3. それぞれが他方の原因、及び結果である場合：双方向の類推において正の相関係数が期待される。
4. 両者の間には因果が存在しない場合：双方向の類推において無相関となる。

ただし、AB 間に擬似相関がある場合には正の相関が観察される場合があり、この場合の対処については第 3 章で取り扱う。

具体的な作業は大きく 2 つに分けられる。1 つは類推を行うための下準備である埋め込みである。これは時系列の情報を空間内の図形に反映させる過程である。この埋め込みによって時系列中の似通った挙動を図形的に検索できるようになり、これが CCM の model-free 性をもたしている。もう一つは、双方の時系列からの埋め込みによって得られた図形を互いに参照することで類推を行う cross-mapping である。以下の 2.2. および 2.3. 節ではこれらのアルゴリズムをそれぞれ解説するが、根本的な原理は本節で解説したものと同一であることに留意されたい。

2.2. 埋め込み (embedding)

力学系の時間発展を観察・分析するには、関与する変数を軸とする空間に各時刻における値をプロットするのが一般的である。例えば X, Y の 2 変数からなる力学系であれば、時刻 t におけるそれぞれの変数の値を 2 次元平面上にプロットすることで系の時間発展を追うことができる。時間発展が「落ち着いた」状態をアトラクタと呼ぶ。アトラクタの形は系によって様々であるが、不動点、リミットサイクル、ストレンジアトラクタに分類できる（合原ほか 2000）。例えば、密度効果の入った個体群動態は不動点に収束し、約 24 時間の周期をもつ概日リズムはリミットサイクルとなる。また Lorenz 方程式 (Lorenz 1963) に代表されるカオス時系列は複雑で美しい軌道であるストレンジアトラクタを描くことが知られている（図 2a）。CCM はアトラクタの情報を元に幾何学的に因果関係を推定するため、特定の力学系関数を仮定する必要がない。

力学系中のある変数の挙動には過去の他の変数の挙動の情報が含まれているため、一つの変数の時間遅れベクトルを座標としてプロットすること（埋め込み）ですべての変数を用いたアトラクタと一対一に対応し、各点の相対的な位置関係が同等であるアトラクタを再構成することができる（Takens の定理、Takens 1981；Sauer et al. 1991、図 2b）。この方法はたとえ観測されていない変数があってもその力学系のアトラクタを描くことができるため非常に有用であり、系の未来予測などにも用いられる（Sugihara 1994）。埋め込みを用いた時系列解析手法は合原（2000）に詳しい。CCM の最初の作業は、因果関係を推定したい 2 本の時系列データ ($X = \{X_1, X_2, \dots, X_T\}$ と $Y = \{Y_1, Y_2, \dots, Y_T\}$ とする) から埋め込みによって 2 つのアトラクタ（それぞれ M_X と M_Y とする）を描くことである。アトラクタを描く空間の次元を埋め込み次元 (embedding dimension, E) という。適切な埋め込み次元で描かれたアトラクタは、すべての変数を用いたアトラクタと同等の情報を保持している。 X からアトラクタ M_X を再構成する場合、時間遅れを τ として時間遅れベクトル

$$x(t) = (X_t, X_{t-\tau}, X_{t-2\tau}, \dots, X_{t-(E-1)\tau}) \quad (1)$$

を作る。このベクトルは E 次元空間の中の一点として表され、その点の集まりが再構成されたアトラクタ M_X である。時刻 $t_1 - (E-1)\tau$ から t_1 までの X の挙動と時刻 $t_2 - (E-1)\tau$ から t_2 までの X の挙動が似通っている時、 $x(t_1)$ と $x(t_2)$ の座標は近くなる。そのため、点同士の距離をその時系列における時間 $(E-1)\tau$ の間の時間発展の類似性を表す指標とすることができる。例えば前節の例で桶の売上からネズミの時系列を類推する場合には連続した 2 年の桶の売上の類似性に着目するため、適切な埋め込み次元は $E = 2$ 、時間遅れは $\tau = 1$ となる。埋め込み次元の決定方法については後述する（2.4. 節）。

2.3. Cross mapping

X と Y の 2 変数からなる力学系を例にして解説する。 X が Y に一方的に影響している状況（つまり X が原因で Y が結果）を考える。この場合、 Y は過去の自身の状態と X によって決定されるため、時系列 Y 中の類似した時間発展のパターンを示す 2 つの時間帯 ($m - (E-1)\tau$ から m と $n - (E-1)\tau$ から n とする) においては、対応する時間帯の X も類似した時間発展のパターンを示すことが期待される。時系列 Y 中の類似した時間発展のパターンは M_Y 上の近傍点として表されるため、2 つのベクトル $y(m)$

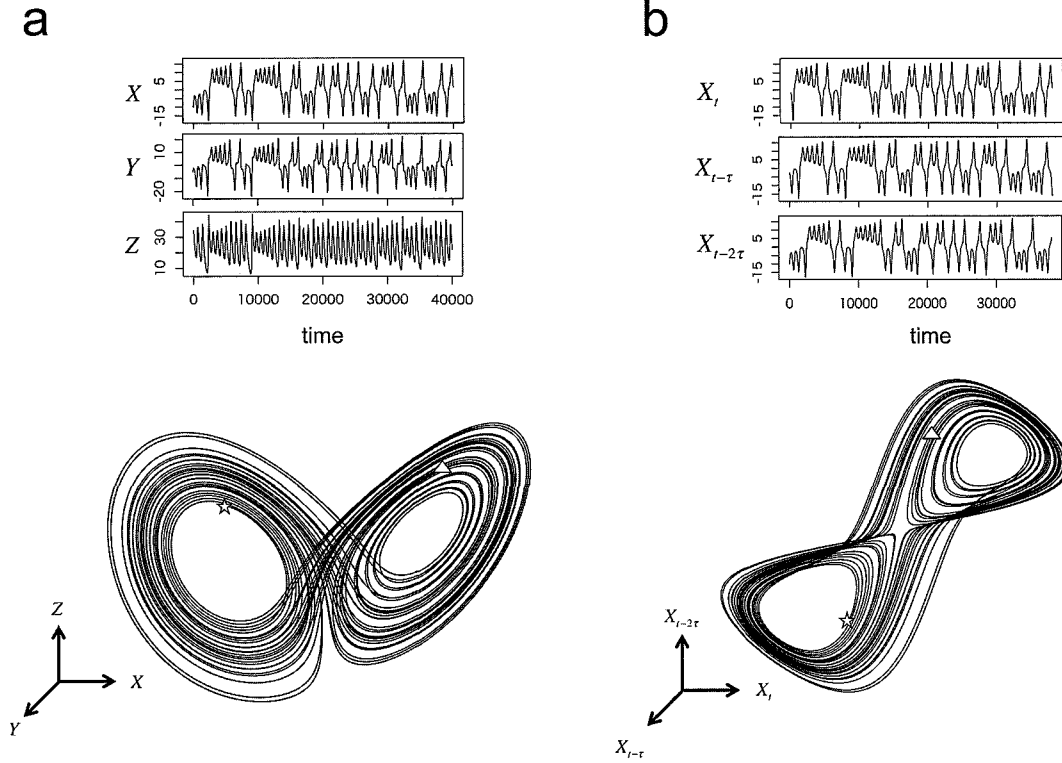


図2. (a) ストレンジアトラクタの例。 $dX/dt = -10X + 10Y$, $dY/dt = -XZ + 28X - Y$, $dZ/dt = XY - 8/3Z$ から得られた三本の時系列を X , Y , Z の三次元空間にプロットした。星マークは軌道の始点、三角マークは終点を表す。(b) X 時系列から $\tau = 0.1$, $E = 3$ で埋め込んだベクトルを用いて再構成したアトラクタ。同様に星マークは軌道の始点、三角マークは終点を表す。上の時系列は見やすくするため、 $\tau = 1$ で埋め込んだ場合を図示した。

と $y(n)$ が近い座標を表す時、対応する $x(m)$ と $x(n)$ も近い位置にあるはずである。逆に、 M_X 上の2点が近傍にあるからといって時間的に対応する M_Y 上の2点が近傍にあるとは限らない。 X は Y から影響を受けないため、 X の時間発展のパターンが似ているからといって Y の時間発展のパターンが似たようなものになるとは限らないからである。よって、ある時間帯における Y の挙動から対応する時間帯の X の挙動を (Y が似た挙動を示す時間帯を参照することによって) 知ることはできるが、その逆は一般的に不可能である。この性質を利用し、CCM では M_X の情報から Y を、また M_Y の情報から X を互いに類推し (cross mapping)、その正確性を評価することで時系列間の因果関係を推測する。上記の例で言えば、 $M_Y \rightarrow X$ の類推は高い正確性が期待されるが、 $M_X \rightarrow Y$ の類推の正確性は低いはずである。ただし、 X と Y が強く同調する場合や、系が線形な場合には X の挙動から Y の挙動をうかがい知ることが出来てしまうため、この方法で正しく因果を検出することはできない。CCM が適用可能なのは X と Y の間に非線形な関係があり、かつ一方の挙動が他方のみに

よって決まってしまうような場合である (Sugihara et al. 2012)。上記の例で、もし桶の売上がネズミの個体数のみによって決まってしまうのであれば、CCM は双方向の因果を (誤って) 検出してしまうだろう。具体的な mapping の方法を $M_X \rightarrow Y$ の類推を例として以下に示す。

1. M_X から Y_t を類推する場合、まずは $x(t)$ の $E+1$ 個の最近傍点 (近い順に $x(t_1), x(t_2), \dots, x(t_{E+1})$ とする) を探す。ベクトル $x(t)$ は、これら $E+1$ 個の点の位置ベクトルの加重平均で表されるはずである (例えば2次元平面では3個の点の位置ベクトルの加重平均で平面上のいかなる点も表現することができる)。
2. $x(t_1), x(t_2), \dots, x(t_{E+1})$ と時間的に対応する M_Y 上の点 $y(t_1), y(t_2), \dots, y(t_{E+1})$ を用いて Y_t を類推する。 Y_t を第一成分とするベクトル $y(t)$ は、やはりこれら $E+1$ 個の点の位置ベクトルの加重平均で表される。 M_X による Y_t の類推を $\hat{Y}_t$ とすると、

$$\hat{Y}_t = \sum_{i=1}^{E+1} w_{t,i} Y_{t_i} \quad (2)$$

で、ここで $w_{i,j}$ は重み付けパラメタであり、

$$w_{i,i} = \frac{u_{i,i}}{\sum_{j=1}^{E+1} u_{i,j}} \quad (3)$$

である。ただし $u_{i,i} = \exp\{-d[x(t), x(t_i)]/d[x(t), x(t_i)]\}$ であり、 $d[x(a), x(b)]$ は 2 つの座標 $x(a)$, $x(b)$ のユークリッド距離である。

3. 全ての時刻について $\hat{y}_t$ を求め、類推された時系列 $\hat{Y}$ と実際の時系列 Y との相関係数 ρ によって mapping の正確性を評価する。

2.4. 埋め込み次元 E と時間遅れ τ の決め方

CCM を行う際には、埋め込み次元 E と埋め込みの時間遅れ τ を予め決めておかなければならない。これらは CCM の結果に大きく影響する場合があるため、慎重に選択する必要がある。例えば E が過小であれば因果推定に必要な情報を十分に保ったアトラクタを再構成することができないかもしれないし、過大であれば不必要な遠い過去の情報によって有用なパターンを近傍点として見つけ出すのを邪魔されてしまうことになる。これらのパラメタの決定法に関して議論がなされてきたが、未だ統一の見解がない。一つの方法として、CCM と同様の方法で未来予測を行う (simplex projection, Sugihara and May 1990) ことで決定する方法が考えられる (Clark et al. 2015)。すなわち、

1. $A = \{X_1, X_2, \dots, X_{t-1}\}$ に対し $B = \{X_2, X_3, \dots, X_t\}$ とおく。このとき、 B は A の 1 タイムステップ先の時系列となっている。
2. 様々な E, τ のもとで M_A から B の類推を試みる。
3. 最もよい B の類推 (すなわち未来予測) を与えるパラメタセットを CCM に適用する。

未来予測の精度も因果関係推定精度と同様にアトラクタ再構成の精度に大きく依存するため、この方法は E と τ を決定するのに有効であると考えられる。 τ に関しては経験的な証拠によって決定するのが適切な場合もあるだろう。例えばある生物の個体数の時系列を分析する時、その生物が隔年で産卵することがはっきりとわかっているのであれば $\tau=2$ とすることは合理的であろう。

3. 結果の解釈と有意性の検定

因果関係推定における厄介な問題の一つが擬似相関である。2 つの事象 X, Y の間に直接の因果関係がなくとも、

それらが共通の要因から影響をうける場合には相関が生じ、ある程度正確な cross mapping を実現してしまう場合がある。このため、cross mapping の正確性だけから因果関係を判断するのは危険である。また、cross mapping の正確性は類推された時系列と実際の時系列の間の相関係数で評価されるため、その値の有意性の検定が重要となる。本章では CCM 結果の出力、解釈の仕方と、有意性の検定について説明する。

3.1. 結果の解釈

CCM では、実際の因果関係と擬似相関を見分けるため、アトラクタ再構成に用いる埋め込まれたベクトルの数 (L) に対する時系列類推の正確性の変化を観察する。例えば $L=10$ で $M_Y \rightarrow X$ の類推を行う場合、ベクトルの集合 $\{y(t), y(t+1), \dots, y(t+9)\}$ を用いたアトラクタ M_Y の再構成と時系列 X の類推を全ての t について行い、それぞれの t における X の類推 $\hat{x}$ と X との相関係数 ρ の平均 $\bar{\rho}$ を類推の正確性とする。図 3 は、 $L=5, 10, 15, \dots, 200$ として $M_Y \rightarrow X$ の類推を行った時の $\bar{\rho}$ を示している。 X が Y の原因である時、 L が大きくなるにつれて M_Y が高い解像度で描かれるため、 X の類推の正確性は上昇、飽和する (図 3a、これを convergence という)。CCM では、この convergence をもって因果が存在すると判断する。一方で X, Y 間に因果関係は存在せず、第三の要因 Z によって両者が (非線形に) 制御された結果として擬似相関が生じている場合、 M_Y から X を、また M_X から Y をある程度正確に類推することはできるが、 M_X と M_Y は互いの情報を含んでいないため、 L が大きくなっても類推がより正確になることはない (図 3b)。つまり、因果関係が存在すると断じるためには L とともに有意に上昇する必要がある。ただし、例えば Z と X が単純な線形関係にあるような場合には M_Z と M_X が実質的に同じ図形となるため、 M_Y から X の類推で convergence を観察することができる。この性質は、多くの場合において真の動態の観測値を用いて因果関係を推定することを可能とする。2.1. 節の例でいえば、ネズミの個体数の真値ではなく、それと比例する (と仮定できる) 観測値と桶の売上のデータを用いてネズミの個体数そのものと桶の売上の因果関係を推定することができる。しかし一方で、対象となる事象と線形関係にある第 3 の事象が存在する場合には、推定された因果関係は見かけ上のものであり、第 3 の事象との因果関係が真である可能性を考慮せねばならないだろう。ネズミの個体数とネコの個体数が比例している時には、ネズミではなくネコの個体数が桶の売上に影響している

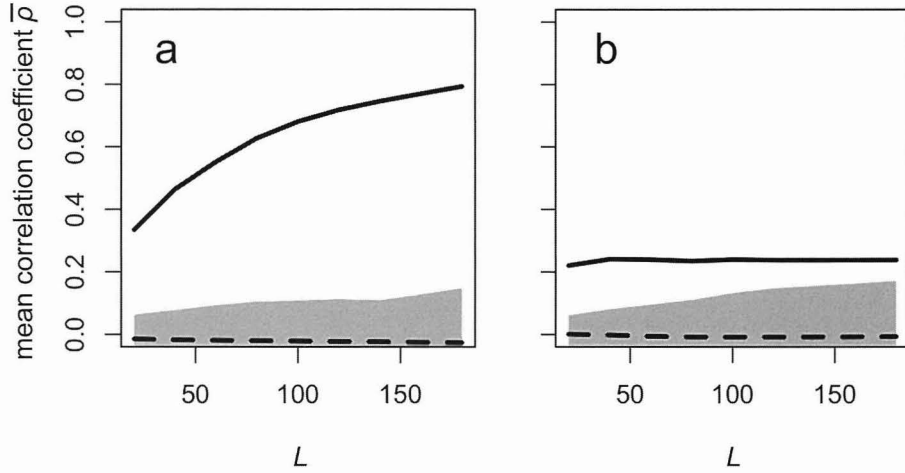


図3. CCMの結果出力。横軸が類推に用いるデータ点の長さ L 、縦軸が類推された値と真の値の間の相関係数 ρ の平均値 $\bar{\rho}$ を表す。実線が観測データに対して CCM を行った結果。点線はサロゲートデータ、すなわち2つの時系列間に因果がないと仮定した場合に CCM を行った結果の平均値を表している。影はサロゲートデータに CCM を適用した場合の真値との平均相関係数の片側 95% 信頼区間を示している。(a) 2つの時系列の間に因果が存在する場合。(b) 擬似相関の場合。平均 0、標準偏差 1.5 の正規分布に従うノイズを $\eta(t)$ として $r_x(t) = 3.1x(t-1)(1-x(t-1))e^{-0.3\eta(t)}$, $r_y(t) = 2.9y(t-1)(1-y(t-1))e^{-0.36\eta(t)}$, $x(t) = 0.4x(t-1) + \max(r_x(t), 0)$, $y(t) = 0.35y(t-1) + \max(r_y(t), 0)$ から得られた2つの時系列を用いて CCM を行った結果。ノイズ $\eta(t)$ が共通の原因になっている。 r_x と r_y には指数関数の部分によって相関が生じるため、 L が小さくともある程度正確に x と y の cross mapping を行うことができる。しかしそれぞれの変数から再構築したアトラクタ上の点に対応関係があるわけではないので、 L が大きくなっても cross mapping の正確性が向上することはない。

可能性も考えなければならない。加えて、因果関係が存在すると言明するためには、得られた類推が因果関係のない場合に比べて有意に高い正確性をもっている必要がある。次節以降で有意性の検定法について紹介する。

3.2. サロゲートデータを用いた類推の正確性の検定

得られた類推の正確性が因果関係のない場合に比べて有意に高いことを示すためには、因果性が存在しない場合にその類推の正確性がどの程度の確率で生じるかを推定する必要がある。因果関係がない場合に $\bar{\rho}$ が必ず 0 になるかということそうではない。データ時系列の構造によっては、因果関係がなくてもある程度高い $\bar{\rho}$ が実現する場合があります。得られた $\bar{\rho}$ がどの程度信頼できるのかを検定するためには、「2つの時系列間に因果性がない」という帰無仮説を設定し、この仮説の下でもとの時系列データと同様の構造を持つ新たな時系列を多数作り、CCM で相互類推の正確性を求め、元のデータのそれと比較すればよい。元の埋め込まれた時系列 $x(t)$ にシャッフルリング等を施すことによって得られる、ある特定の性質を除いた時系列データ $x^s(t)$ をサロゲートデータといい、時系列解析の仮説検定でしばしば使われる。

Sugihara et al. (2012) はサロゲートデータを作成する際に、パワースペクトルを保存する手法 (Fourier Transform や Iterative Amplitude Adjusted Fourier Transform など) を用いたが、その場合、得られた時系列は非線形力学系の性質も失う。先に設定した帰無仮説に対しては非線形力学系であることも保存しなければならないため、パワースペクトルを保存する手法は適切ではないだろう。ここでは Thiel et al. (2006) によって提案された twin surrogate 法を採用するのが有効だろう。twin surrogate 法は、2つの時系列の非線形力学系の性質を保ちつつ、それらの間の因果性を断ち切ったサロゲートデータを作ることができる。この手法は時系列から埋め込まれたベクトル列を一定の規則に基づいてシャッフルするものであり、具体的には以下の4つのステップから成る。

1. 元の時系列データを埋め込むことによって作られた N 個のベクトルに対し、 $i, j = 1, \dots, N$ として recurrence matrix

$$R_{i,j} = \Theta(\delta - \|x(i) - x(j)\|) \quad (4)$$

を作る。 $\Theta(\cdot)$ はヘヴィサイド関数 ($\cdot < 0$ で $\Theta = 0$ 、 $\cdot \geq 0$

で $\Theta = 1$)、 $\|\cdot\|$ は最大ノルム ($x(i) - x(j)$ の各成分の絶対値の最大値) を表す。最大ノルムはユークリッド距離で代用可能である (Thiel, 私信)。この recurrence matrix は任意の 2 つのベクトル $x(i)$ と $x(j)$ が閾値 δ より離れているかどうかを 0、1 で表している。そのため、(解像度が落ちてはいるが) recurrence matrix には元の非線形力学系のアトラクタと同等の情報が含まれている (Thiel et al., 2004)。

2. $k = 1, 2, \dots, N$ なる全ての k に対し $R_{ik} = R_{jk}$ なるペア $x(i)$ と $x(j)$ を twin と呼ぶ。すべてのペアに関して twin かどうかを調べる。0 と 1 の recurrence matrix 上では、互いに twin であるベクトル同士を区別することはできない。
3. i 番目のサロゲートデータを $x^s(i)$ と呼ぶことにし、その 1 つ目として元のベクトル $x(i)$ をランダムに選び $x^s(1) = x(i)$ とする。
4. $x^s(j) = x(m)$ としたとき、 $x(m)$ が twin を持たないのであれば $x^s(j+1) = x(m+1)$ とする。 $x(m)$ が twin として $x(n)$ を持つのであれば、等確率で $x^s(j+1) = x(m+1)$ または $x^s(j+1) = x(n+1)$ とする。twin が 3 つ以上ある場合でも同様に、候補の数だけ等確率で選ぶ。twin 同士は recurrence matrix 上で見分けがつかないので、入れ替えても時系列の性質 (自己相関関数や相互情報量など) は保存される。しかしアトラクタを構成するベクトルがシャッフルされているので、他方の変数を用いて再構成したアトラクタとの時間的対応関係は断ち切られている。

Step4 を繰り返すことで元と同じ長さを持ったサロゲートデータが作られる。元と同じ長さに達する前に元の軌道ベクトルの末端に達してしまい、なおかつ末端の点が twin を持たなかった場合には、そのサロゲートデータは不採用とする。

この手法ではパラメタ δ を決定する必要があるが、 δ への依存性は小さいことが示されており (Thiel et al. 2006)、 $\delta = 0.05 \sim 0.2$ で最も良好な結果が得られるようである。本稿中の解析では $\delta = 0.125$ を採用している。この手法によって作られたサロゲートデータは元の時系列の性質を保存していることがわかっている。特に非線形な性質である相互情報量に関しては他のサロゲート法では保存されないが twin surrogate では保存されることが知られている (Thiel et al. 2006)。

2 つの時系列間の因果関係を推定するとき、一般的には原因と仮定される時系列のサロゲートデータを用いて検定を行う (Verdes 2005; Lizier et al. 2011)。例えば X_t が

原因で Y_t が結果であると仮定する場合、 $x(t)$ のサロゲートデータと $y(t)$ のオリジナルデータを用いて CCM を行う。それを多数 (例えば 1000 本) のサロゲートデータに対して行い、実際の $\bar{\rho}$ がサロゲートデータから得られた $\bar{\rho}^s$ の分布の中で上位何番目に位置するかで P 値を計算することができる。図 3 では影として片側 95% 信頼区間を表示している。実際の $\bar{\rho}$ がこれよりも大きければ、統計的に有意に高い正確性で時系列の類推ができたと判断できる。

3.3. convergence の検定

“convergence” の検定と書くと「飽和していること」を検定するかのような印象を受けてしまうが、CCM の結果の解釈で重要になるのは (飽和しているかどうかではなく) $\bar{\rho}$ が L とともに上昇しているかどうかである (3.1. 節)。従って、本節では「convergence の検定」といった場合には $\bar{\rho}$ の上昇の検定を指していると理解していただきたい。CCM が発表された論文 (Sugihara et al. 2012) 中では、convergence の検定は行われていない。しかし、真の因果関係と擬似相関を見分けるためには convergence の有意性を統計的に検定することが必要である。現在、convergence の検定を行った唯一の例は Clark et al. (2015) であるが、これは CCM を用いた発展的手法である multispatial CCM (6.3. 節) のための方法である。そのため、本節では Clark et al. (2015) の方法を CCM で使える形に改変したものを提案する。

Sugihara et al. (2012) ではアトラクタを構成する際に L 個の時間的に連続したベクトルを選択した (3.1. を参照) が、代わりに全てのベクトルから重複を許してランダムに選択する bootstrapping によって L 個のベクトルを選択することで convergence の検定ができると考えられる。bootstrapping によって L 個のベクトルを含むアトラクタを多数作り出し、それぞれに対し CCM を行うことで L に対する類推の正確性 $\bar{\rho}$ の信頼区間を得ることができる。Clark et al. (2015) では、最大の L に対応する $\bar{\rho}$ が最小の L に対応する $\bar{\rho}$ より有意に大きい場合に convergent であるとしている。

4. 実際の運用手順

第 3 章までで CCM の仕組みについては解説したが、CCM を適用するには対象となる時系列が一定の基準を満たす必要があり、これを確認しないまま CCM を適用すると間違った結論を導き出してしまう可能性がある。本章において、CCM を運用するにあたっての実際の手順を

まとめておく。実装としては Ye et al. (2015) において、埋め込み次元決定のための Simplex projection、非線形性チェックのための S-map、因果関係検出のための CCM のコード (R と C++) がセットになったパッケージ (rEDM、<https://github.com/ha0ye/rEDM>) が公開されている。

4.1. 時系列の準備

CCM を適用するデータは定量的であり、時間的に等間隔でサンプリングされた一組の時系列データでなければならない。欠測値がある場合にはその前後の値の平均値などで補間する必要があるだろう。Sugihara et al. (2012) によるとデータ点数は 35—40 点あれば適用可能である。データにトレンドが見られる場合には差分をとる等して定常化する。ノイズが大きいと思われる場合には移動平均等のスムージング手法が有効である場合があるが、これは同時にデータの構造を変化させ、誤った結論を導き出してしまう危険をはらんでいるかもしれないため慎重に検討して利用すべきである。Sugihara et al. (2012) で紹介されているマイワシとカタクチイワシの漁獲量と水温の間の CCM では、漁獲量データは差分をとって定常化し、水温データは 3 年移動平均をとってスムージングを行っている。

4.2. 埋め込み次元・時間遅れの決定と非線形性のチェック

時系列の準備ができたら、2.4. 節に記した方法で埋め込み次元 E と時間遅れ τ を決定する。CCM は結果側の時系列のアトラクタの構造をもとに原因側の時系列を類推するので、正しく因果関係を推定するためには決定した E と τ のもとで結果側の時系列の非線形構造が確率過程 (ノイズ) によって失われていないことが必要である。結果側の時系列が非線形構造を保持しているかどうかは、未来予測手法によって確認できる。もしある時系列が確率過程によってその非線形構造を (完全に) 失っているのであれば、データの平均値を与えるのが最良の予測となるだろう。平均値は全ての点の情報を (等しく) 用いた時に一番正確に与えられるため、これは線形予測において傾きが 0 の場合であると考えられる。一方でその時系列が非線形性を保持しているのなら、線形予測ではなく非線形予測が最良の予測を与えるはずである。非線形性のチェックは S-map を用いて行うのが簡便である (Tsonis et al. 2015)。これは時系列の一部 (library) を用いて描かれたアトラクタの情報をもとにして残りの部分の未来予測を行い、その正確性を観察することで非線形性を確認する方法である。もし、library を用いて描か

れたアトラクタ上の X_τ の近傍点に重みをおいた線形予測 (これは実質的な非線形予測である) が、library 全てを等しい重みで用いた線形予測 (AR モデル) よりも良い精度を示すのであればその系は非線形であるということができる。具体的な予測法と、それをを用いた非線形性の確認法は Sugihara (1994) と Hsieh et al. (2005) でそれぞれ詳しく説明されている。

4.3. CCMによる解析・結果の解釈・検定

2 章および 3.1. 節に記された手順に従って埋め込みと cross-mapping を行う。アトラクタ再構成に用いるデータ点の数 (L) を変え、類推された原因側の時系列とものデータとの平均相関係数 ($\bar{\rho}$) の、 L への依存性を出力する。得られた $\bar{\rho}$ が、因果関係がない場合と比べて有意に高いかどうかを twin surrogate 法で確認する (3.2. 節)。また $\bar{\rho}$ の convergence を 3.3. 節の方法で確認する。

5. Granger 因果性テストとの比較

因果性の推定に現在最も一般的に使われているのが Granger 因果性テストである。CCM と Granger 因果性テストは適用対象が異なるため、それぞれの原理を理解し、正しく使い分けることが重要である。本章ではまずこの Granger 因果性テストについて解説した後、CCM との原理的な違いを解説していく。

5.1. Granger因果性テスト

2 種の時系列 X と Y を考える。使える全ての情報の集合を U とした時、 U から Y を差し引いた集合を用いて予測した時の X の予測精度が U を使って予測したときの精度に劣る時、 Y は X に対して Granger 因果性を持つと定義される。このため、Granger 因果性は CCM で仮定されている因果性の定義と必ずしも一致しない。次のような、プロセス誤差を含む 2 本の時系列を考える。

$$\begin{aligned} X_t &= aX_{t-1} + bY_{t-1} + \varepsilon_{X,t-1} \\ Y_t &= cX_{t-1} + dY_{t-1} + \varepsilon_{Y,t-1} \\ \varepsilon_X &\sim N(0, \sigma_X^2) \\ \varepsilon_Y &\sim N(0, \sigma_Y^2) \end{aligned} \quad (5)$$

ここで $N(\mu, \sigma^2)$ は平均 μ 、分散 σ^2 の正規分布を表す。 X および Y は、係数 b と c を通じて互いに因果をもつ構造になっている。この場合、過去の X と Y の状態からそれぞれの現在の状態が決定される。また、これらは、以下のように書き換える事ができる。

$$\begin{aligned} X_t &= (a+d)X_{t-1} + (bc-ad)X_{t-2} + \varepsilon_{X,t-1} - d\varepsilon_{X,t-2} + b\varepsilon_{Y,t-2} \\ Y_t &= (a+d)Y_{t-1} + (bc-ad)Y_{t-2} + \varepsilon_{Y,t-1} + c\varepsilon_{X,t-2} - a\varepsilon_{Y,t-2} \end{aligned} \quad (6)$$

これは、 X の過去の情報のみから現在の X を、 Y の過去の情報のみから現在の Y を、それぞれ予測できることを表している。しかしそれぞれの誤差項の分散に注目すると、 X 、 Y 両方の過去の情報から現在を予測する場合の誤差項の分散は σ_X^2 、 σ_Y^2 であるのに対し、 X 、 Y それぞれの過去の情報のみから現在の状態を予測する場合の誤差項の分散はそれぞれ $(1+d^2)\sigma_X^2 + b^2\sigma_Y^2$ 、 $c^2\sigma_X^2 + (1+a^2)\sigma_Y^2$ となり、前者の場合よりも大きくなる。つまり、 X と Y の情報を使って X を予測した場合に比べて、 X のみから X を予測した場合のほうが粗精度となるため、 Y は X に対してGranger因果を持つ、ということが出来る。この方法はプロセス誤差、つまり確率的過程が主な系の駆動力になっている場合には有効であるが、確率過程の影響が小さい場合、つまり決定論的な力学系においては検出力が落ちてしまう。

5.2. CCM

一方でCCMは、決定論的な力学系において威力を発揮する方法である。ここでは完全に決定論的である系を考える。上記の線形系は誤差項がないと固定点に収束（あるいは発散）してしまうので、ここではロジスティック方程式によるカオスを考える。Sugihara et al. (2012)に倣い、2本の時系列を

$$\begin{aligned} X_t &= 3.9X_{t-1}(1-X_{t-1}-bY_{t-1}) \\ Y_t &= 3.7Y_{t-1}(1-Y_{t-1}-0.2X_{t-1}) \end{aligned} \quad (7)$$

としよう。これを変形することで

$$\begin{aligned} X_t &= \frac{3.9}{0.2} \left\{ \left(1 - bY_{t-1} \right) \left(1 - Y_{t-1} - \frac{Y_t}{3.7Y_{t-1}} \right) - \frac{1}{0.2} \left(1 - Y_{t-1} - \frac{Y_t}{3.7Y_{t-1}} \right)^2 \right\} \\ Y_t &= \frac{3.7}{b} \left\{ \left(1 - 0.2X_{t-1} \right) \left(1 - X_{t-1} - \frac{X_t}{3.9X_{t-1}} \right) - \frac{1}{b} \left(1 - X_{t-1} - \frac{X_t}{3.9X_{t-1}} \right)^2 \right\} \end{aligned} \quad (8)$$

が導かれる。これは Y の現在と過去の情報から X の現在の状態を推定できることを示している。仮に Y は X にほとんど影響を及ぼさないとする（ $b \approx 0$ ）。この場合、(8)下側の式では b が分母にあるため、 X_t や X_{t-1} の小さな誤差が Y_t の大きな誤差を生むことになる。それは M_X 上での近傍点と同時刻の Y が必ずしも M_Y 上の近傍点とはならないことを意味しており、 M_X から Y を類推する際の正

確性が低くなってしまふ。つまり、 X が Y の原因となっている場合、 M_Y から X の類推は正確であるが、その逆は正確でないという結果が得られる事となる。一方で、上式をGranger因果性テストの場合と同様に变形し、 X_t を X_{t-1} と X_{t-2} で表すこともできる。これは、 X の過去の時系列のみから現在の X の状態を、また Y の過去の時系列のみから現在の Y の状態をそれぞれ予測できることを示している。このような状況でGranger因果性テストを用いてしまうと、 Y の情報を使っても使わなくても同じ精度で X の予測が出来てしまうため、因果なしと判断されてしまふ。

このように、決定論的な系においては、結果である時系列の中に原因となる時系列の情報が完全に保存されているため、未来予測精度を指標とした因果性推定はうまく機能しなくなってしまう。CCMの骨子は未来予測性ではなく、互いの時系列の相互類推の正確性を指標として因果性推定を行うことで決定論的な力学系から生じた時系列間の因果性推定が可能となっていることである。これはつまり、片方の時系列からもう一方の状態を類推する方法は別の方法であっても機能する可能性があることを示している。実際、Hirata and Aihara (2010)は埋め込みとrecurrence plotを用いて同様の因果性推定を行う方法を開発している。時系列の相互類推を行う方法は、今後の発展が期待される部分であろう。

逆に、確率的過程が系の主な駆動力になっている場合にCCMを適用してしまうと間違った結論に至ってしまう可能性がある。そのため、CCMを適用する前に結果側の時系列が非線形な構造を保持している（ノイズによる動態ではない）ことを確認しておく必要がある（4.2.節）。

6. 適用例・今後の課題・発展的手法

6.1. 適用例

Sugihara et al. (2012)では、2つの生態学データに対してCCMを適用してその因果関係を解析している。1つは実験室内で得られたもので、捕食者（*Didinium nasutum*）と被食者（*Paramecium aurelia*）がLotka-Volterra方程式様の振動を示すデータである。CCMを行った結果、被食者から捕食者、また捕食者から被食者の双方向の因果が検出された。捕食者の時系列を類推した場合の方が逆の場合と比べて正確であり、捕食者によるトップダウンコントロールが被食者によるボトムアップコントロールに比べて個体群動態へ強く寄与していることが示された。もう一つはカリフォルニアにおけるマイワシ（*Sardinops*

sagax) とカタクチイワシ (*Engraulis mordax*) の漁獲量および水温データである。この2種は生物量の変動が非常に激しく、またマイワシの生物量が多い時にはカタクチイワシは少なく、カタクチイワシが多い時にはマイワシは少ないことが知られている (魚種交替現象)。魚種交替の原因としては種間競争 (Matsuda et al. 1992) や環境の影響 (Chavez et al. 2003) がこれまでに考えられてきたが、CCM による解析の結果、2 種の間には因果関係が検出されなかったが、水温は両者に影響していることがわかった。

2014 年 12 月の段階で Sugihara et al. (2012) の被引用数は 49 件 (Web of Science) であり、様々な分野から注目されている。しかし、ほとんどの論文において、時系列データから因果関係を検出することは重要であり、新しい手法が提案されている、などの言及に留まっている。おそらく研究例が少ないことだけでなく、メカニズムの難解さやコーディングの手間がハードルを高くしていると考えられる。実際に CCM を用いて解析した論文としては、温室効果ガスと気温の関係を解析した Wang et al. (2014) および van Nes et al. (2015)、砂塵嵐と植物の成長の関係を解析した Fan et al. (2014) がある。また、Web of Science の検索結果には出てこないが、動脈の血圧と脳の血流速の関係を解析した Heskamp et al. (2014)、ビールの売上と販促を解析した Huffaker and Fearn (2014) もある。

6.2. 今後の課題：ノイズに対する頑健性について

実際のデータに CCM を適用する際に留意しなければならないのがプロセス誤差および観測誤差の存在であろう。CCM は決定論的力学系を対象としているが、現実のデータにおいてノイズは不可避である。このため、CCM がノイズに対してどの程度の頑健性を持っているのか評価することは非常に重要であると考えられる。CCM の頑健性評価を、確率過程を含む力学系で行った例は今までに存在しない (次節で紹介する multispatial CCM の頑健性評価は Clark et al. (2015) 中で行われている)。ノイズに対する頑健性の評価は、対象となる系において想像される力学系関数を用いたシミュレーションで行うのが妥当であろう。以下に、自己回帰 (AR) モデルで表される環境要因の生物の個体群動態に対する影響を解析する際の頑健性評価例を紹介する。

環境要因 a_t は、AR(1) モデル

$$a_t = 0.9a_{t-1} + \varepsilon_{t-1}, \quad \varepsilon \sim N(0, 0.15) \quad (9)$$

で表される。環境要因は内的自然増加率 r と環境収容力 K に影響し、時刻 t における r, K はそれぞれ

$$\begin{aligned} r_t &= ra_t^{e_r} \exp(\varepsilon_{r,t}), \quad \varepsilon_{r,t} \sim N(0, \sigma_p^2) \\ K_t &= Ka_t^{e_K} \exp(\varepsilon_{K,t}), \quad \varepsilon_{K,t} \sim N(0, \sigma_p^2) \end{aligned} \quad (10)$$

で表される。ここでは $r = 0.5, K = 1000, e_r = e_K = 0$ (因果性なし) または $e_r = e_K = 2$ としている。個体群動態は Ricker モデル

$$x_t = x_{t-1} \exp \left\{ r_{t-1} \left(1 - \frac{x_{t-1}}{K_{t-1}} \right) \right\} \quad (11)$$

である。個体群動態として観測される時系列 o_t はこれに観測誤差を乗じて

$$o_t = x_t \varepsilon_o, \quad \varepsilon_o \sim \exp \left[N(0, \sigma_o^2) \right] \quad (12)$$

とした。様々な σ_p と σ_o の組み合わせのもと、それぞれ 100 組の a_t と o_t について CCM で因果関係推定を行った。時系列の長さは 101、埋め込み次元 $E = 2$ 、時間遅れ $\tau = 1$ としているため、アトラクタを構成するベクトル数 L の最大値は 100 である。 $L = 100$ における類推の正確性が、twin surrogate 法で作られた 1000 データを用いて類推した場合の正確性 $\bar{p}^s$ の上位 5% 以内に入った場合に因果ありと判定した。 $e_r = e_K = 2$ の場合に a_t から x_t および x_t から a_t の因果の有無について、(前者は因果あり、後者は因果なし) 正しく推定できた割合を図 4 に示す。 a_t から x_t の因果が正しく推定された (true positive) 割合はプロセス誤差や観測誤差が大きくなるにつれて減少したのに対し、 x_t から a_t の因果が正しく推定された (true negative) 割合は誤差の大きさによらずほぼ 95% であった。また $e_r = e_K = 0$ の時、観測誤差の大きさにかかわらず a_t から x_t および x_t から a_t の因果の有無が正しく (因果なし) 推定された (true negative) 割合はほぼ 95% であった。この結果は、実際の時系列を CCM で解析した際に、①検出された因果は実際に存在するものとして信用してよい、②仮に検出されなかったとしても因果が存在しないと断定はできない、ということを示している。例外的に、プロセス誤差と観測誤差が非常に小さい場合 (例えば $e_r = e_K = 2$ かつ $\sigma_p = \sigma_o = 0$)、本来は存在しない x_t から a_t の因果が検出されてしまうようだ。2 つの系が非常に強くリンクしている場合に同様の現象が見られることは Sugihara et al. (2012) でも報告されている。因果関係が非常に強く、片方の挙動がほとんど他方の挙動のみによって決まってしまう場合には、 M_x 上の近傍点群と同時刻の M_y 上の点

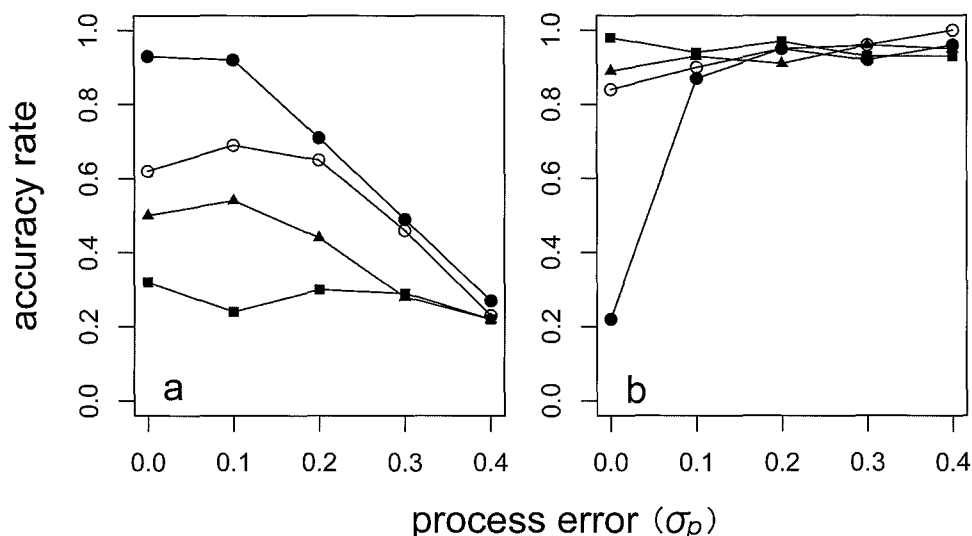


図4. ノイズに対する頑健性評価の結果。横軸はプロセス誤差の大きさの指標 (σ_p)、縦軸は正解率。各マーカーは観測誤差の大きさの指標 (σ_o) を表している。 $\sigma_o=0$ (黒丸)、 $\sigma_o=0.1$ (白丸)、 $\sigma_o=0.2$ (三角)、 $\sigma_o=0.4$ (四角)。(a) a_i から x_i への因果ありと正しく判断されたケースの割合。(b) x_i から a_i への因果なしと正しく判断されたケースの割合。

群も必ず近傍となるため、両方向の類推が正確に行われてしまうのだろう。

本節で紹介したようなシミュレーションを様々な系に対して行うことで、CCMの一般的な傾向、強み、弱点等を知ることができるだろう。背後に存在する力学系の形が全くわからない系に対してCCMを適用するためには、今後こうした知見を蓄積し、適切な解析を行っていくことが重要である。

6.3. CCMを用いた発展的手法：multispatial CCM

CCMを生態学データに用いる場合の利点の一つは、比較的短い時系列に対して適用できる点である。Sugihara et al. (2012)によればデータ点数は35—40点あれば適用可能である。とはいえ、例えば年一回行われる調査のデータを30年以上にわたって蓄えるのは至難の業である。この問題に対し、Clark et al. (2015)は短い時系列を多数用いることでCCMを適用できるようにする方法(multispatial CCM)を提案した。各々の時系列が n 個のデータ点からなるとき、一本の時系列から E 次元の埋め込みによって生じる点は $n-E+1$ 個である。multispatial CCMは、各時系列(m 本とする)から生じた $m(n-E+1)$ 個の点をプールすることによって精度の高いアトラクタを描き、CCMを行う方法である。彼らは同じ力学系を共有するが初期値の異なる100本のシミュレーションデータを用い、各々が5つのデータ点しか持たない場合でも

(さらにはプロセス誤差と観測誤差が加わっても)正しく因果関係を推定できることを示している。

multispatial CCMを実際の生態学データに用いるためには、各々の時系列が同じ力学系を共有していることが必要である。このことが担保されるのであれば、時系列の長さというCCMを用いる上での最大の困難は大幅に克服されることになるだろう。

謝 辞

本稿の執筆は科学技術振興機構の戦略的創造研究推進事業(CREST)「海洋生態学と機械学習法の融合によるデータ不足下の生態系評価手法の開発」の予算で行われた。本稿を執筆するにあたり、台湾大学のChih-hao Hsieh教授にはCCMの細かな点についてたくさんの解説をいただいた。中央水産研究所の市野川桃子博士と東京大学大学院総合文化研究科の笠田実博士には論文の構成や内容に関して多数のアドバイスをいただいた。東京大学生産技術研究所の平田祥人特任准教授にはGranger因果性テストとCCMの比較、およびtwin surrogate法の導入について多大なアドバイスをいただいた。さらに本原稿を精読して頂き、非常に有用なコメントをいただいた。論文の査読にあたって、潮雅之博士と二名の匿名の査読者には有益なコメントをいただいた。ここに謹んでお礼申し上げる。

引用文献

- 合原 一幸, 池口 徹, 山田 泰司, 小室 元政 (2000) カオス時系列解析の基礎と応用. 産業図書, 東京
- Chavez FP, Ryan J, Lluch-Cota SE, Niquen MC (2003) From anchovies to sardines and back: multidecadal change in the Pacific Ocean. *Science*, 299:217-221
- Clark AT, Ye H, Isbell F, Deyle ER, Cowles JM, Tilman D, Sugihara G (2015) Spatial 'convergent cross mapping' to detect causal relationships from short time-series. *Ecology*, 96:1174-1181
- Fan B, Guo L, Li N, Chen J, Lin H, Zhang X, Ma L (2014) Earlier vegetation green-up has reduced spring dust storms. *Scientific reports*, 4:6746
- Granger CWJ (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424-438
- Heskamp L, Abeelen ASSM, Lagro J, Claassen JAHR (2014) Convergent cross mapping: a promising technique for cerebral autoregulation estimation. *International journal of clinical neuroscience and mental health*, 1:S20
- Hirata Y, Aihara K (2010) Identifying hidden common causes from bivariate time series: A method using recurrence plots. *Physical Review E*, 81:016203
- Hsieh CH, Glaser SM, Lucas AJ, Sugihara G (2005) Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean. *Nature*, 435:336-340
- Huffaker R, Fearn A (2014) Empirically testing for dynamic causality between promotions and sales beer promotions and sales in England. *Proceedings in Food System Dynamics*, 270-274
- Lizier JT, Heinze J, Horstmann A, Haynes JD, Prokopenko M (2011) Multivariate information-theoretic measures reveal directed information structure and task relevant changes in fMRI connectivity. *Journal of Computational Neuroscience*, 30:85-107
- Lorenz, EN (1963) Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, 20:130-141
- Matsuda H, Wada T, Takeuchi Y, Matsumiya Y (1992) Model analysis of the effect of environmental fluctuation on the species replacement pattern of pelagic fishes under interspecific competition. *Researches on Population Ecology*, 34:309-319
- Perada E, Quiroga RQ, Bhattacharya (2005) Nonlinear multivariate analysis of neurophysiological signals. *Progress in Neurobiology*, 77:1-37
- Sauer T, Yorke JA, Casdagli M (1991) Embedology. *Journal of Statistical Physics*, 65:579-616
- Schreiber T (2000) Measuring information transfer. *Physical Review Letters*, 85:461-464
- Sugihara G, May RM (1990) Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 344:734-741
- Sugihara G (1994) Nonlinear forecasting for the classification of natural time series. *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences*, 348:477-495
- Sugihara G, May R, Ye H, Hsieh CH, Deyle E, Fogarty M, Munch S (2012) Detecting causality in complex ecosystems. *Science*, 338:496-500
- Takens F (1981) Detecting strange attractors in turbulence. In: Rand D A and Young L S, *Lecture Notes in Mathematics* 898, 366-381. Springer-Verlag, Berlin Heidelberg
- Thiel M, Romano MC, Kurths J (2004) How much information is contained in a recurrence plot? *Physcis Letters A*, 330:343-349
- Thiel M, Romano MC, Kurths J, Rolf M, Kliegl R (2006) Twin surrogates to test for complex synchronization. *Europhysics Letters*, 75:535-541
- Tsonis AA, Deyle ER, May RM, Sugihara G, Swanson K, Verbeten JD, Wang G (2015) Dynamical evidence for causality between galactic cosmic rays and interannual variation in global temperature. *Proceedings of the National Academy of Sciences*, 112:3253-3256
- van Nes EH, Scheffer M, Brovkin V, Lenton TM, Ye H, Deyle E, Sugihara G (2015) Causal feedbacks in climate change. *Nature Climate Change*, 5:445-448
- Verdes PF (2005) Assessing causality from multivariate time series. *Physical Review E*, 72:026222
- Wang X, Piao S, Ciais P, Friedlingstein P, Myneni RB, Cox P, Heimann M, Miller J, Peng S, Wang T, Yang H, Chen A (2014) A two-fold increase of carbon cycle sensitivity to tropical temperature variations. *Nature*, 506:212-215
- Ye H, Beamish RJ, Glaser SM, Grant SCH, Hsieh CH, Richards LJ, Schnute JT, Sugihara G (2015) Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. *Proceedings of the National Academy of Sciences*, 112:E1569-E1576