

收敛交叉映射方法及其在生态学中的应用

王丹雨 朱媛君 杨晓晖*

(中国林业科学研究院荒漠化研究所, 北京 100091)

摘要 收敛交叉映射 (CCM) 是一种分析非线性系统中时间序列变量间因果关系的方法。其不同于传统的线性系统分析方法, 是通过变量进行状态空间重构来获取变量的历史信息, 随着时间序列不断增长, 当其估计性能呈现收敛的性质时, 可以判断因果关系的存在。本文介绍了 CCM 的发展史及其较传统的格兰杰因果检验的优点, 详细阐明了 CCM 的原理、算法过程和实现途径。CCM 作为一种针对变量间具有弱到中等强度耦合关系的系统分析方法, 可以用来有效地解决非线性生态系统多变量间复杂的因果关系问题。将该方法应用于具有空间信息的多点时间序列变量间因果分析时, 应充分考虑点位间的空间自相关性, 与去除变量及序列间空间相关性的方法相结合, 从而确保 CCM 对变量因果关系的分析更加准确, 结果也更具有信服力。

关键词 收敛交叉映射; 因果关系; 格兰杰因果检验; 非线性复杂生态系统

Convergent cross mapping method and its application in ecology. WANG Dan-yu, ZHU Yuan-jun, YANG Xiao-hui* (*Institute of Desertification Studies, Chinese Academy of Forestry, Beijing 100091, China*).

Abstract: The convergent cross mapping (CCM) is a method to analyze causality of nonlinear time series variables. Different from the traditional linear system analysis method, CCM gets historical information based on their state space reconstruction. The presence of causality can be confirmed when the estimated values perform convergent with time series extension. Here, we introduced the development history of CCM and its advantages over the traditional Granger causality test, and elaborated the principle, algorithm process, and implementation approach. As a system analysis method aiming at the coupling relationship between variables from weak to moderate, CCM can effectively solve the complex causality among nonlinear multivariable in ecosystems. When it is applied to the causality analysis of multi-point time series variables with spatial information, the spatial autocorrelation among points should be fully considered and combined with the method that can remove the spatial correlation between variables and sequences, so as to ensure more accurate causality analysis using CCM and more convincing results.

Key words: convergent cross mapping (CCM); causality; Granger causality test; nonlinear complex ecosystem.

自然系统通常是复杂的非线性系统, 当研究者利用以相关分析为基础的线性分析方法分析其中变量间的因果关系时, 常常会出现短暂拟合的问题, 并且相关性和因果关系之间也没有必然的联系, 所以,

对于存在噪声的非线性系统来说, 线性的方法并不能很好地分析系统中变量间的关系^[1-5]。目前, 越来越多的观测和模拟数据无法用传统方法分析, 使得非线性因果分析方法发展迅速, 其使用范围逐渐超越了相关分析方法。

1 收敛交叉映射 (CCM) 的产生

1956 年 Wiener 在文章中提出了关于因果关系概念的最早构想^[6], 即如果纳入时间序列 X 的信息能够改善对另一个时间序列 Y 的预测, 则 X 是 Y 的

本文由国家自然科学基金国际 (地区) 合作与交流项目 (32061123005) 和国家自然科学基金项目 (41971061) 资助 This work supported by International (Regional) Cooperation and Exchange Program of National Natural Science Foundation of China (32061123005), and National Natural Science Foundation of China (41971061).
2021-03-02 Received, 2021-09-24 Accepted.
* 通讯作者 Corresponding author. Email: yangxh@caf.ac.cn

因。随后,Granger^[7]于1969年提出了将上述概念进行实际量化的格兰杰因果检验(Granger causality test)^[8-9],认为如果将时间序列 X 的信息从时间序列 Y 中移除,会导致 Y 的可预测性下降,则 X 是 Y 的因^[9-10]。这一概念认为时间的流动决定了因果的方向,过去或现在可以预测未来,而未来不能推测过去。2003年Granger在获得诺贝尔奖的演讲中进一步说明了格兰杰因果关系的前提条件:第一原因必须发生在结果之前。这使得很多学者认为格兰杰因果检验实际上只是在判断变量发生的前后顺序,并非逻辑上的因果关系^[11-12],同时,对于非线性系统中呈“状态依赖”(系统中相互作用变量之间的关系随着系统状态的不同而发生变化的特征)关系的变量^[1,13-14],判断“时间依赖”的格兰杰因果检验并不适用;第二要求原因所包含的信息可以独立地从系统中分离^[6,15]。但在现实世界中,因果变量常常相互混杂,且很难分离。这些要求使应用格兰杰因果检验充满局限性。

分析生态系统中变量的因果关系一直是一个困扰生态学家的难题,对于以弱到中等强度耦合关系来保持群落持久性和稳定性的生态系统,针对强耦合关系变量的格兰杰因果检验并不适用^[14,16-17]。除此之外,生态系统还常常受到温度、降水等外部气候变量的强迫^[14,18-19],导致处于相似外部环境中的变量会出现如莫兰效应^[20]所描述的明显的相关性和同步性特征,使得从变量的相互作用中分辨真正的因果关系变得更加困难。基于以上原因,分析生态系统需要一种针对信息相互混杂的、弱到中度耦合关系变量的新方法,同时需要在共同驱动因子的影响下分辨出变量之间的因果关系是否真实存在^[14,21-22]。针对以上问题,Sugihara等^[14]在2012年提出了一种分析弱到中等耦合强度非线性系统中变量间因果关系的方法——收敛交叉映射(convergent cross mapping,CCM)。CCM是通过响应变量的历史信息能否准确估计触发变量的程度来判断二者之间因果关系的方法,可以应用于30个或更多的连续观测序列^[14]。

2 收敛交叉映射的原理

人类认知世界的主流观点包括决定论和混沌理论:决定论认为过去以规律的方式精确决定了未来^[23],混沌理论则认为未来是随机且不可预测的^[24]。近几十年来,随着对非线性复杂动力系统的深入认识,混沌理论的应用愈加广泛。为了更好地

说明这一理论,混沌理论之父爱德华·洛伦茨建立了一个简单模型,公式如下:

$$\begin{aligned}dx/dt &= \sigma(y-x) + q^x \\ dy/dt &= \rho x - y - xz + q^y \\ dz/dt &= xy - \beta z + q^z\end{aligned}$$

式中: x 、 y 和 z 为因变量; q^x 、 q^y 和 q^z 表示模型的未知误差; σ 是普朗特数; ρ 为相对瑞利数; β 是速度阻尼常量^[25-27]。由以上模型可以得到变量在空间内的运动轨迹(图1)^[28]:混沌系统中变量的运动并不是完全无序的,最终会在其相空间中稳定下来,仿佛被两个中心点不断地吸引着,形成一些复杂的子集。这些子集就是洛伦茨吸引子;洛伦茨吸引子运动所形成的轨迹被称为吸引子流形^[29]。自然系统中变量的运动往往也遵循着这样的规律,所以,混沌理论是CCM建立的一个重要理论基础。

CCM基于的另一个重要理论是Takens^[30]于1981年提出的嵌入定理,其与混沌理论的应用息息相关^[31-35]。Takens嵌入定理证明了当应用混沌理论构造一个 D' 维的相空间来容纳一个无限长且无噪声的来自 D 维原系统的一维时间序列变量的运动时,要保证重构空间与原始动态系统的拓扑学意义相同,那么,空间维数需要满足 $D' > 2D + 1$ 的要求^[30,35-39]。CCM正是利用这一性质对变量进行状态空间重构(state space reconstruction,SSR),其原理如下:考虑两个长度为 L 的动态耦合的原始时间序列 X 和 Y ,随时间的变化, X 和 Y 在共同空间内的运动轨迹代表 D 维的吸引子流形 M (attractor mani-

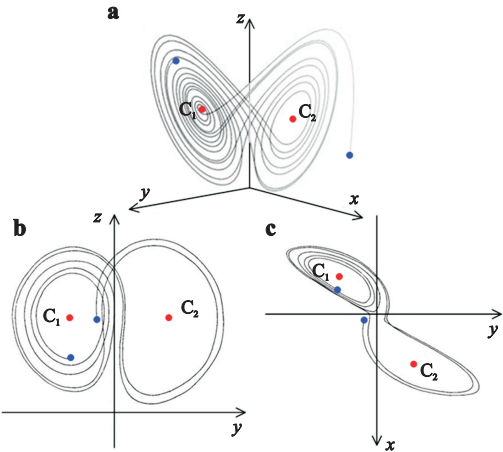


图1 洛伦茨吸引子运动轨迹图
Fig.1 Lorenz attractor trajectory diagram.

a是三维空间中的洛伦茨吸引子轨迹图;b和c是洛伦茨吸引子轨迹的二维平面图; C_1 和 C_2 是运动轨迹所围绕的中心点 a is Lorenz attractor trajectory in three-dimensional space; b and c are the two-dimensional plane of the Lorenz attractor trajectories; C_1 and C_2 are the center points around which the trajectory is moving.

fold)^[40-41],利用变量的滞后坐标分别创建 M 中 X 和 Y 的影子流形 M_X 和 M_Y ,因为变量 X 和 Y 在 M 中是动态耦合的,所以, M_X 和 M_Y 上的点在时间上是相互对应的,代表着 X 与 Y 相关;而当 L 长度增加,吸引子的轨迹被更多数据填充,使得用以估计 X 的 Y 的历史信息越多, Y 对 X 的估计也就越好,当二者的映射点越接近,因果影响也就越强。这就是 CCM 的关键——收敛性,是指预测值与观测值的相关系数会随着构建样本的时间序列长度 L 增长而增加,并在达到极值后保持平稳的性质^[14,21,42-44]。因此,CCM 并不是像格兰杰因果检验一样依靠判断触发变量对响应变量的预测能力来判断因果关系是否存在,而是从响应变量入手,认为如果因果关系存在,响应变量一定会包含触发变量的信息。当历史信息增多,响应变量对触发变量状态的估计也越准确。CCM 的优势在于:当存在多个触发变量时,观察响应变量的信息更容易^[14,21-22,45-46];同时,CCM 识别因果关系利用的是历史数据,即使在多变量的系统中也可以重复应用;而且其测量的只是基于原始吸引子流形所创建的影子流形之间的相关性,不会对系统造成扰动^[44];CCM 在耦合系数恒定的弱到中等耦合非线性系统中,CCM 可以识别单向和双向因果关系^[47]。

3 收敛交叉映射的算法和实现

CCM 检测因果关系的具体流程如图 2 所示。其算法如下。

3.1 创建影子流形

首先假设在同一个系统中有两个长度为 L 的动态耦合的时间序列变量 X 和 Y :

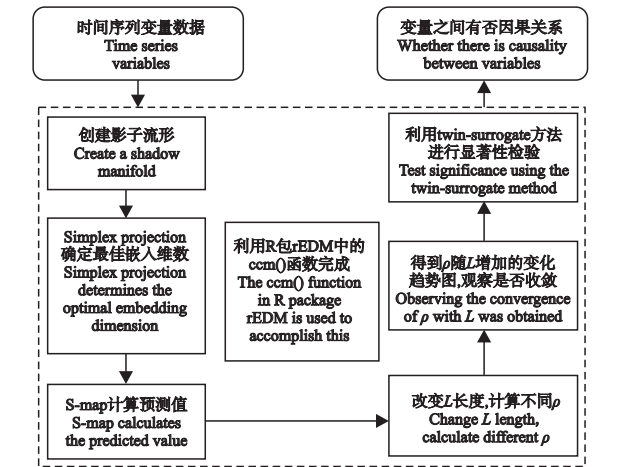


图 2 CCM 流程图
Fig.2 Flow diagram of CCM.

$$\{X\} = \{X(1), X(2), \dots, X(L)\}$$
$$\{Y\} = \{Y(1), Y(2), \dots, Y(L)\}$$

根据 Takens 嵌入定理,确定影子流形存在的空间维数 E 后,利用时滞变量 τ 创建时间序列变量 X 和 Y 的影子流形 M_X 和 M_Y ,分别满足 $x(t) = [X(t), X(t-\tau), X(t-2\tau), \dots, X(t-(E-1)\tau)]$, $y(t) = [Y(t), Y(t-\tau), Y(t-2\tau), \dots, Y(t-(E-1)\tau)]$, $[1+(E-1)\tau < t < L]$ (图 3)^[41]。

3.2 确定重构空间的最佳嵌入维数和时滞向量

重构空间的嵌入过程是一个将时间序列变量的信息进行图形化的过程,从而便于在时间序列中搜索具有相似特征的变量的动态变化趋势。这种图形化的过程也使得 CCM 并不需要一个特定的方程去判断变量之间的因果关系^[14,34,48],因此使用更加灵活。

影子流形所存在的重构空间的维数 E 受系统复杂性、时间序列长度和噪声等因素的影响^[14,49]。确定 E 非常关键。在实际应用中, E 通常会取一系列值,当 E 过大,可能会导致重构空间包含许多不必要的信息,对分析造成困难;过小则有可能无法包含原空间的所有信息,因此存在一个通过 simplex projection 的方法来确定最佳的嵌入维数值,可以使得重构空间能够完全保留原空间信息^[19,48]。Simplex projection 是一种最近邻预测算法,其原理可以通过图 4^[19,50-51] 来解释: a~e 表示要分析的动态时间序列。a 中粉色的点是需要预测的值,首先找到粉点之前的变化趋势,即 b 中红线所描绘的部分,接

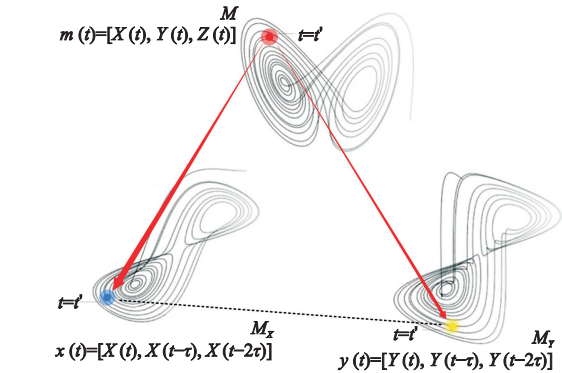


图 3 影子流形创建示意图
Fig.3 Shadow manifold creation schematic.

M 是由 X 、 Y 和 Z 3 个时间序列变量组成的吸引子流形, M_X 和 M_Y 分别是变量 X 和 Y 利用滞后坐标所重构的影子流形。因为变量 X 和 Y 在 M 中是动态耦合的,在任意的 $t=t'$ 时刻, M_X 和 M_Y 是相互对应的 M is an attractor manifold composed of three time series variables X , Y and Z , and M_X and M_Y are shadow manifolds reconstructed by the lagged-coordinate vector of X and Y respectively. Because the variables X and Y are dynamically coupled in M , at any time $t=t'$ prime, M_X and M_Y correspond to each other.

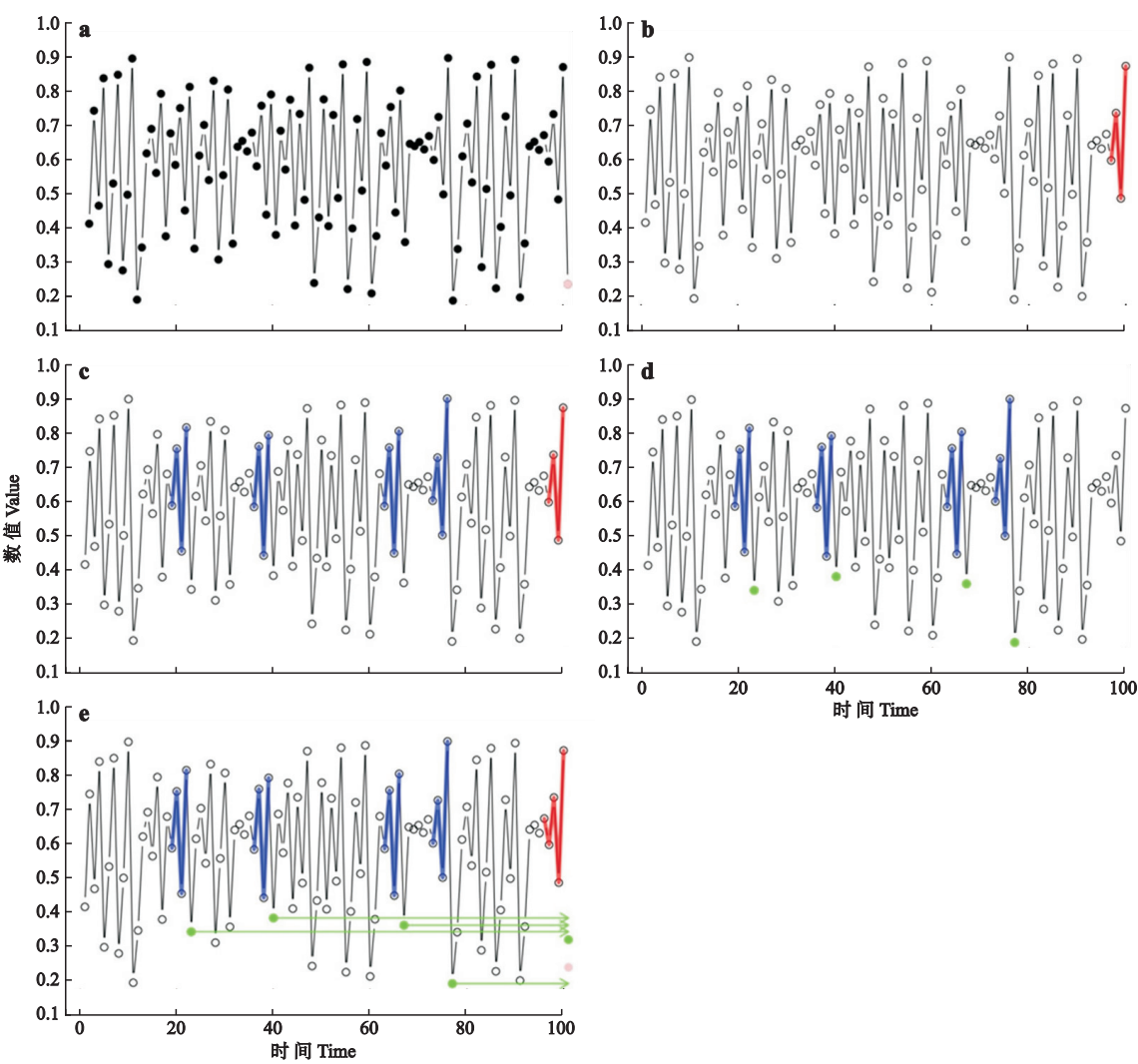


图 4 Simplex projection 原理示意图
Fig.4 Schematic diagram of simplex projection.

a~e) 要分析的部分时间序列,时间为相对值 The partial time series to be analyzed. The time was relative value. a) 粉点表示要预测的值 The pink dot represents the predicted values; b) 红线表示预测值发生前的变化趋势 The red line represents the trend of change before the predicted value occurs; c) 蓝线部分表示时间序列中与红线描绘趋势类似的事件 The blue lines represent events in the time series that are similar to the red line in depicting the trend; d) 绿点表示蓝线趋势引发的值 The green dots represent the values caused by the trend of the blue lines; e) 将这些值取平均,得到与粉点最为接近的绿点,所用的类似趋势的事件个数就是最佳的嵌入维数 These values are averaged to get the green dot that is closest to the pink dot. The number of similar trend events used is the optimal embedding dimension.

下来需要寻找类似的变化趋势, c 中用蓝色标示了 4 个最接近的匹配, d 中的 4 个绿点代表蓝色趋势获得的值,随后对这些值取平均,得到 e 所示,与粉点十分接近的绿点,二者越接近,预测误差就越小。也就是说, simplex projection 是利用过去的类似事件对未来进行预测。而所谓的嵌入维数也就是在时间序列中有多少类似的过去事件可以用来预测未来,最佳的嵌入维数是能够使观测值和预测值之间的平均绝对误差最小的维数。

时滞向量 τ 一般通过经验值来确定,如以降水量和蒸发量为研究变量时, τ 可以设置为一个月^[52],在温度影响油棕同化物的储存研究中, τ 被

设置成 7 个月^[53],研究鱼类时,在明确知道隔年产卵的情况下, τ 会设置为 1 年^[54];也可以通过生成 Bootstrapping 库的方法,迭代计算每个数据点在不同时间滞后情况下的交叉收敛性能,从而确定最佳的时间滞后向量值^[55]。

3.3 利用 S-map 方法计算预测值

S-map 是 Sugihara^[22] 于 1994 年提出的一种标准线性变量自回归模型的非线性扩展分析方法。类似于线性回归中的回归系数, S-map 赋予空间中不同位置变量以不同权重值^[56-57],并通过生成相关的 Jacobian 矩阵元素来定义吸引子中各点的相互作用强度^[42, 56-57]。在非线性系统中, S-map 相当于对线

性方程 C 进行奇异值分解 (Singular value decomposition, SVD), 计算方程^[22,43,57-58] 如下:

$$B=A\times C$$

式中: $B_k=w_kx_i(t_k+1)=w_kY_i; A_{kj}=w_kX_j(t_k), A$ 的维度为 $n\times(E+1)$;

$$w_k=\exp\frac{-\theta\parallel x(t_k)-x(t^*)\parallel}{\bar{d}}$$

$$\bar{d}=\frac{1}{n}\sum_i\parallel x(t_i)-x(t^*)\parallel$$

式中: $x(t^*)$ 是预测值; \bar{d} 是空间中邻近向量距离预测值点的平均距离。权重值 w 依赖于调优参数 θ ($\theta\geq0$)。线性系统中 $\theta=0$, 空间中所有向量的权重值都相等, 空间中局部的变化趋势可以代表全局变化趋势; 非线性系统中 $\theta>0$, 越靠近预测值的向量权重会越大, 强调局部的信息^[22,43,56-58]。在实践中, 为了平衡偏倚和不确定性, 并避免 θ 的微小变化所造成巨大的误差, 常常会选择一个中间值来达到最佳状态^[57-58]。

3.4 计算预测值和观测值之间的相关系数

计算相关系数(ρ)的公式^[57-58] 如下:

$$\rho=cov(x,\hat{x})/\sigma_x\sigma_{\hat{x}}$$

式中: 分子是观测值和预测值之间的协方差; 分母是观测值和预测值标准差的乘积。

ρ 的值越高, 代表对预测值的估计越好, 但 ρ 的存在仅代表相关性, 因果关系需要通过收敛性来判断。收敛性是指随着预测信息增加, 预测性能(即相关系数 ρ) 改善并达到最大值的特性^[59], 是确定因果关系的关键。根据 Sugihara 等^[14] 的说明, 如果 X 和 Y 受到同一个系统外变量 Z 的影响, 会检测到二者之间存在相关性, 但二者的相关系数并不会随着时间序列长度的增加而收敛。也就是说, CCM 能够区分变量之间的相关性是由于二者的因果关系还是系统外部变量的强迫。需要注意的是, 在所有条件都相同的情况下, 相关系数 ρ 能够代表因果关系的强度^[14,60]。

3.5 检验因果关系的显著性

有时变量在没有因果关系的情况下, 也会因为时间序列的结构而呈现出比较高的相关性。因此, 需要用 twin surrogate 方法对 ρ 进行显著性检验, 进一步判断因果关系是否可信。Twin surrogate 是一种基于递归性质生成替代数据的检验方法。其原假设是时间序列变量 X 和 Y 之间并没有因果关系, 通过相位随机化的方法将 X 和 Y 打碎重组, 形成与原本时间序列变量具有相似结构的替代时间序列变量,

并再次进行 CCM 过程, 得到替代时间序列的相关系数 ρ' ^[61-62]。当原本时间序列变量得到的相关系数 ρ 大于 ρ' 的 95% 时, 可以认为原本时间序列变量之间的因果关系是确实可信的^[54,63]。

3.6 CCM 的实现

Park 等^[64] 于 2020 年发布的 R 语言中的 rEDM 包(<https://github.com/SugiharaLab/rEDM/>) 可实现以上算法, 其中包括多种 CCM 相关函数: 函数 simplex() (利用 simplex projection 确定最佳嵌入维数)、函数 s_map() (利用 S-map 计算预测值)、函数 ccm() (完成 CCM 全部过程)、函数 make_surrogate_data() (生成替代时间序列) 等。

函数 ccm() 使用单纯形投影执行收敛交叉映射过程, 主要参数包括 block (代表作为时间序列的数据)、 E (重构空间的嵌入维数)、tau (时滞向量)、lib_sizes (库的长度) 等; 输出结果包括 L (库的长度)、num_pred (预测结果的数量)、rho (相关系数 ρ)、mae (平均绝对误差)、rmse (均方根误差)。以 rEDM 包中的 sardine_anchovy_sst 数据集 (加州当前鳀鱼-沙丁鱼-海表温度系统的时间序列) 为例, 函数设置为 ccm(sardine_anchovy_sst, E=3, lib_column="anchovy", target_column="np_sst", lib_sizes=seq(10,80,by=10), num_samples=100), 其中嵌入维数为 3, 交叉映射列的名称为 "anchovy", 需要预测的列名称为 "np_sst", 库为一个长度为 10~80、间隔为 10 的序列, 库中不同长度时间序列所用的随机样本数为 100。Kawatsu 等^[65] 在对斯洛伐克共和国同域叶食森林鳞翅目物种的种间同步潜在机制的研究中, 函数设置为 ccm(cbind(block[,env], block[,sp]), lib=seg, E=E, lib_sizes=c(E+1,28), silent=TRUE), 其中库设置为长度为 $E+1\sim28$ 的数列, 得到不同物种的嵌入维数范围为 1~15, 相关系数数值则在 0~0.4 之间; Masayuki 等^[62] 在自然鱼类群落交互作用网络的波动与时变稳定性的分析中, 函数设置为 ccm(block, E=BestErEDM(x, E=config\$kBest-E.Range), lib_sizes=c(11,15,24,40,80,120,160,200,240,m), silent=T), 研究中使用了预先算好的最佳维数来进行状态空间重构, 库的长度分别为 11、15、24、40、80、120、160、200、240 和所使用的物种数据的列数, 结果为交互作用强度的动力稳定性与交互作用强度标准差和相互作用分布的偏态之间的相关系数在库长度为 100 左右时开始收敛。

4 收敛交叉映射在生态学中的应用

通过对所引用文献的整理分析, 获得了 CCM 在

生态学中不同领域的应用情况(表1):在植物生理生态学和植物生态学领域,CCM 被用以隔离植物的扰动源或预测藻华的发生^[13,15],并被 Matin 等^[52]用以分析甘肃省两大内河流域两岸植被指数增强与水循环的季节动态之间的双向因果关系。在动物生态学领域,CCM 常被用以解决不同物种种间同步机制的问题,如 Kawatsu 等^[65]不仅利用 CCM 发现外源气候条件可能是斯洛伐克共和国同域叶食森林鳞翅目昆虫种间同步的原因,还发现在两种不同的豆甲虫竞争试验中,胜种和败种之间存在显著的因果关系。CCM 在群落生态学、水域生态学领域也发挥着极大的作用。大量的试验已证实,CCM 在海洋生态系统中具有很强的适用性^[48]。如 CCM 可以量化和预测美国威斯康星州波光湖生态系统中的种间相互作用^[58];Nakayama 等^[54]使用 CCM 方法确定北太平洋西部海洋表面温度影响了南部沿海和近海区域的沙丁鱼(*Sardinops* spp.)种群和北部沿海地凤尾鱼(*Engraulis* spp.)种群的自然增长,且不同季节的影响范围不同;Masayuki 等^[62]利用日本 Maizuru 湾天然海洋鱼类群落的观测数据,通过 CCM 分析发现种间相互作用、群落网络结构和群落稳定性之间是动态耦合的关系。在物理生态学、污染生态学等新兴的生态学领域,Tsonis 等^[47]利用 CCM 分析表明在较短的年际时间尺度上,银河宇宙射线和全球温度的短期年变率之间存在显著的因果效应;Zhang 等^[66]用 CCM 发现北半球环流模式对东北亚冬季地表气温有单向的因果关系;Yao^[67]和 Chen 等^[68]利用 CCM 分析发现 PM_{2.5}浓度对不同地区的气象都有因果影响。以 Naoki 等^[53]的试验为例来详细说明 CCM 在生态学领域中的应用。实验地为马来西亚

檳城的马来西亚大学(5°08'37" N,100°29'39" E),试验对象为 4 棵经雌性生殖器官修剪和 4 棵未作修剪的油棕(*Elaeis guineensis*),2012 年 10 月 23 日至 2016 年 3 月 31 日记录雌性生殖器官的宽度、深度和高度等相关数据,采集茎芯以获取茎中的淀粉和可溶性糖含量,计算茎生长速率,同时,从马来西亚气象局获取观察时段内的气象数据。通过以上观察和测量获取了茎中淀粉和可溶性糖的时间序列数据、雌性生殖器官生长和体积的时间序列数据,以及日积温和累积降雨量的时间序列数据用以 CCM 分析。首先使用固定的库长度进行 simplex projection 确定最佳维数,通过 S-map 获得不同库长度下的预测值,并判断收敛性,最终通过 twin surrogate 方法检验因果关系的显著性,结果表明,茎中所含的非结构性碳水化合物(淀粉含量和可溶性糖)与雌性生殖器官生长之间存在相互因果关系,同时受到长期累积温度的影响,时间滞后为 7 个月,而与长期累积降雨量之间的动态存在较短的时滞。

虽然 CCM 已经逐步在生态学的多个领域发挥作用,但结构方程模型(structural equation model, SEM)在生态学领域的应用更加广泛^[69-71]。SEM 融合了回归分析、相关分析和因子分析等多种分析方法,CCM 也以相关分析为基础,但二者有很多不同之处,最根本的区别在于分析对象不同:CCM 的分析对象是具有一定时间长度的非线性时间序列变量,分析结果包括变量间的因果方向和因果影响的时间滞后效应;SEM 只针对线性因果关系,一般不对时间序列变量进行分析,只有在具有时间约束条件的情况下(原因发生在结果之前),才能够估计具有瞬时效应的线性时滞因果关系。其次,CCM 解决

表 1 CCM 应用领域统计
Table 1 CCM application fields statistics

应用领域分类 Application field classification		解决的问题 Problem solved	参考文献 Reference
分类依据: 生物 Classification basis: Creature	植物生理生态学 Plant physioecology	探讨植物与环境的相互作用、物质代谢和能量传递规律,以及植物对不同环境条件适应性的生理机制	[12,53,74]
	植物生态学 Plant ecology	分析植物与植物、植物与环境之间的关系	[13,52,75]
	动物生态学 Animal ecology	分析动物与动物、动物与环境之间的关系	[55,65,75]
分类依据: 生态系统结构层次 Classification basis: Hierarchy of ecosystem structure	群落生态学 Community ecology	分析群落内部组成及群落与环境之间相互关系	[54,62,76]
分类依据: 生物栖居的环境 Classification basis: The environment in which the organism lives	水域生态学 Aquatic ecology	研究水生生物生存条件、生物及其群体与环境相互作用的过程及其规律	[48,58,77]
分类依据: 应用的问题 Classification basis: Application problem	污染生态学 Pollution ecology	研究生物系统与被污染的环境系统之间的相互作用和机理	[67,68,78]
与其他非生命科学相结合 Combined with other non-life sciences	物理生态学 Physical ecology	研究有机体与物理环境之间的相互作用,包括能量传递、气体交换等	[47,66,79]

的是变量之间因果方向的问题,更倾向于变量间一对一的分析,并且其相关性必须要在其他环境条件都相同的情况下才能作为表征因果影响强度的指标进行互相比;SEM 中则需要一定的知识基础对路径的指向性进行预先判断,在预判下分析单变量或多变量之间在因果解释下的相关性,也就是变量间关系的强度^[14,72-73]。CCM 与 SEM 虽然原理类似,但针对的对象、解决的问题以及结果都不相同。SEM 不仅包括偏理论性的概念模型,同时也关注变量间相互关联的强度,还涉及更具复杂性的混合模型;CCM 分析的变量关系更加简单,但针对性也更强。CCM 的出现填补了 SEM 在分析非线性变量间因果关系的空白,使得人们可以更加全面地了解生态系统中的因果关系。

5 展 望

虽然 CCM 相较于格兰杰因果关系检验更具优势,但其仍不足以适应纯粹的随机系统或是由噪声主导的系统^[80]。研究者对 CCM 是否能够从噪声中分辨出真正的因果关系进行了相关测试,结果表明,在噪声水平分别为 2.5%、5%和 25%的情况下,CCM 能够分辨出正确的因果交互方向;当噪声增加,相关系数 ρ 会变小,并且在噪声存在的情况下,嵌入维数较高时 CCM 的性能更好^[63];但在噪声非常小的情况下,有可能检测到并不存在的因果关系^[49]。

基于此,许多研究者对 CCM 进行了扩展。Cao 等^[81]于 1998 年提出了确定多元时间序列变量系统嵌入维数的方法,并对比了使用单时间序列变量和多时间序列变量的预测结果,发现使用多元时间序列变量预测结果明显更优,利用多元时间序列变量的嵌入方法,结合多个观测函数生成影子流形,这种多时间序列变量的使用让单变量时间序列变量系统中噪声的限制大大减少^[34,81]。除此之外,针对由于外部变量异常强的单向强迫导致的广义强迫,使得 CCM 可能无法正确地区分因果关系方向的问题。Ye 等^[82]考虑到交叉映射的不同时间滞后对 CCM 进行了扩展,以解决这一问题。罗磊等^[83]在 2016 年也对 CCM 进行改进,开发了扰动过滤交叉映射 (DFCM) 方法,来解决化工过程中遇到的外部扰动问题。CCM 面临的另一个关键问题是时间序列过短。Heish 等^[84]在 2007 年开发了一种将生态系统中多个较短的时间序列结合到一个长时间序列中的方法。这种方法避免了数据上的限制,同时也减少了采样误差,对仅包含 20 个数据点的单个生态时间

序列也有较好的预测能力。Clark 等^[85]在 2014 年将 CCM 方法与露珠回归 (dewdrop regression) 技术结合,构建了多空间 CCM (multispatial convergent cross mapping) 的方法。这种方法注重于对研究变量进行整体的空间复制,从而将时间信息一并包含进去,补偿时间序列过于简短所带来的信息缺失。经过测试,这种方法即使在试验难以执行且时间序列较短的系统中也能很好地使用。

尽管 CCM 已经做了多方面的扩展,但其并未考虑地理空间中不同点位变量间的相互影响。Neeti 等^[86]在 2011 年提出了一种分析时间序列数据的非参数的方法——CMK 检验 (Contextual Mann-Kendall Test)。这种方法利用空间自相关原理来分析地理现象,认为在地理位置上相近的地点会具有类似的变化趋势。如果一个位置表现出与周围其他地点截然不同的变化趋势,可以认为该地点的趋势变化并不真实。CMK 检验首先会对需要分析的时间序列数据进行预白处理 (Prewhitening),去除数据的自相关性,随后对经过处理的数据进行空间交叉相关矫正,也就是考虑空间中不同时间序列之间的互相关性,从而添加相关项来调整计算的方差,判断趋势是否显著^[85]。对于 CCM 来说,可以借鉴 CMK 检验的原理,通过对数据进行预白处理,去除数据自相关性带来的影响,同时,利用空间交叉相关矫正,得到互相关性的计算结果,与 CCM 计算的变量之间的相关性结合,从而消除空间中不同点位空间相关性的影响,使 CCM 对变量之间的因果关系分析更为准确。通过结合不同的分析方法对 CCM 加以改进,CCM 在生态学领域中将会获得更广阔的应用空间。

参考文献

- [1] Chang CW, Masayuki U, Hsieh CH. Empirical dynamic modeling for beginners. *Ecological Research*, 2017, **32**: 1-12
- [2] Ye H, Beamish RJ, Glaser SM, et al. Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. *Proceedings of the National Academy of Sciences of the United States of America*, 2015, **112**: 1569-1576
- [3] Perretti CT, Sugihara G, Munch SB. Non-parametric forecasting outperforms parametric methods for a simulated multispecies system. *Ecology*, 2013, **94**: 794-800
- [4] Cenci S, Saavedra S. Non-parametric estimation of the structural stability of non-equilibrium community dynamics. *Nature Ecology & Evolution*, 2019, **3**: 912-918
- [5] Wang YQ, Yang J, Chen YN, et al. Detecting the causal effect of soil moisture on precipitation using convergent cross mapping. *Scientific Reports*, 2018, **8**: 12171 -

- 12179
- [6] Duan P, Chen TW, Shah SL, *et al.* Methods for root cause diagnosis of plant-wide oscillations. *Aiche Journal*, 2014, **60**: 2019–2034
 - [7] Granger CWJ. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 1969, **37**: 424–438
 - [8] Runge J, Bathiany S, Bollt E, *et al.* Inferring causation from time series in earth system sciences. *Nature Communications*, 2019, **10**: 2553–2566
 - [9] Craciunescu T, Murari A, Gelfusa M. Causality detection methods applied to the investigation of malaria epidemics. *Entropy*, 2019, **21**: 784–801
 - [10] Ding MZ, Chen YH, Bressler SL. Granger causality: Basic theory and application to neuroscience// Schelter B, eds. *Handbook of Time Series Analysis*. Berlin, Germany: Wiley-Verlag, 2008: 437–460
 - [11] 陈雄兵, 张宗成. 再议 Granger 因果检验. 数量经济技术经济研究, 2008 (1): 155–161 [Chen X-B, Zhang Z-C. Re-examination of Granger causality test. *Journal of Quantitative & Technical Economics*, 2008 (1): 155–161]
 - [12] Papagiannopoulou C, Miralles DG, Decubber S, *et al.* A non-linear granger-causality framework to investigate climate-vegetation dynamics. *Geoscientific Model Development*, 2017, **10**: 1945–1960
 - [13] McGowan JA, Deyle ER, Ye H, *et al.* Predicting coastal algal blooms in southern California. *Ecology*, 2017, **98**: 1419–1433
 - [14] Sugihara G, May R, Ye H, *et al.* Detecting causality in complex ecosystems. *Science*, 2012, **338**: 496–500
 - [15] Aftab MF, Hovd M, Sivalingam S. Convergent cross mapping (CCM) based approach for isolating the source of plant-wide disturbances. 2017 IEEE Conference on Control Technology and Applications (CCTA), Hawaii, USA, 2017: 1492–1498
 - [16] Mccann K, Hastings A, Huxel GR. Weak trophic interactions and the balance of nature. *Nature*, 1998, **395**: 794–798
 - [17] Chen ZY, Xie XM, Cai J, *et al.* Understanding meteorological influences on PM_{2.5} concentrations across China: A temporal and spatial perspective. *Atmospheric Chemistry and Physics*, 2018, **18**: 5343–5358
 - [18] Hurrell JW. Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, 2009, **269**: 676–679
 - [19] Post E, Stenseth PNC. Climatic variability, plant phenology, and northern ungulates. *Ecology*, 1999, **80**: 1322–1339
 - [20] Moran PAP. The statistical analysis of the Canadian Lynx cycle. *Australian Journal of Zoology*, 1953, **1**: 291–298
 - [21] Sugihara G, May RM. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 1990, **344**: 734–741
 - [22] Sugihara G. Nonlinear forecasting for the classification of natural time series. *Philosophical Transactions of the Royal Society A-Mathematical, Physical and Engineering Sciences*, 1994, **348**: 477–495
 - [23] 张华夏. 决定论究竟是什么? 中国社会科学, 1993 (6): 30–45 [Zhang H-X. What exactly is determinism? *Social Science in China*, 1993 (6): 30–45]
 - [24] 吴彤. 非线性动力学混沌理论方法及其意义. 清华大学学报: 哲学社会科学版, 2000, **15** (3): 72–79 [Wu T. Chaos theory method of nonlinear dynamics and its significance. *Journal of Tsinghua University: Philosophy and Social Sciences*, 2000, **15** (3): 72–79]
 - [25] 晏志武. 洛伦兹吸引子相图的简易实现. 物理通报, 2012 (12): 64–65 [Yan Z-W. Simple realization of Lorenz attractor phase diagram. *Physics Bulletin*, 2012 (12): 64–65]
 - [26] 沈浙奇, 章向明, 唐佑民. EAKF 和 SIR-PF 在贝叶斯滤波框架下的比较和结合. 海洋学报, 2015, **35** (3): 69–78 [Shen Z-Q, Zhang X-M, Tang Y-M. Comparison and combination of EAKF and SIR-PF in the Bayesian filter framework. *Acta Oceanologica Sinica*, 2015, **35** (3): 69–78]
 - [27] Lorenz EN. Maximum simplification of the dynamic equations. *Tellus*, 2010, **12**: 243–254
 - [28] Lorenz EN. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 1963, **20**: 130–141
 - [29] Asciti FA, Beltrami E, Carroll TO, *et al.* Is there chaos in plankton dynamics? *Journal of Plankton Research*, 1993, **15**: 603–617
 - [30] Takens F. Detecting strange attractors in turbulence// Rand D, Young LS, eds. *Dynamical Systems and Turbulence*. Berlin, Germany: Springer-Verlag, 1981: 366–381
 - [31] Noakes L. The Takens embedding theorem. *International Journal of Bifurcation and Chaos*, 1991, **1**: 867–872
 - [32] Sauer T, Yorke JA, Casdagli M. Embedology. *Journal of Statistical Physics*, 1991, **65**: 579–616
 - [33] Stark J, Broomhead DS, Davies ME, *et al.* Takens embedding theorems for forced and stochastic systems. *Nonlinear Analysis Theory Methods & Applications*, 1997, **30**: 5303–5314
 - [34] Deyle ER, Sugihara G. Generalized theorems for nonlinear state space reconstruction. *PLoS One*, 2011, **6** (3): e18295
 - [35] Stark J, Broomhead DS, Davies ME, *et al.* Delay embeddings for forced systems. II. stochastic forcing. *Journal of Nonlinear Science*, 2003, **13**: 519–577
 - [36] Huke JP. Embedding nonlinear dynamical systems: A guide to Takens' theorem [EB/OL]. (2006-03-09) [2021-02-26]. <http://eprints.maths.manchester.ac.uk/id/eprint/175>
 - [37] Bush K, Pineau J. Manifold embeddings for model-based reinforcement learning under partial observability. *Advances in Neural Information Processing Systems*, 2009, **21**: 189–197
 - [38] 张淑清, 贾健, 高敏, 等. 混沌时间序列重构相空间参数选取研究. 物理学报, 2010, **59** (3): 1576–1582 [Shen S-Q, Jia J, Gao M, *et al.* Study on the parameter determination for reconstructing phase-space in chaos

- time series. *Acta Physica Sinica*, 2010, **59**(3): 1576–1582]
- [39] 张雨, 任成龙. 确定重构相空间维数的方法. 国防科技大学学报, 2005(6): 105–109 [Zhang Y, Ren C-L. The methods to confirm the dimension of re-constructed phase space. *Journal of National University of Defense Technology*, 2005(6): 105–109]
- [40] Günther B, Segal J. Every attractor of a flow on a manifold has the shape of a finite polyhedron. *Proceedings of the American Mathematical Society*, 1993, **119**: 321–329
- [41] Günther B. A compactum that cannot be an attractor of a self-map on a manifold. *Proceedings of the American Mathematical Society*, 1994, **120**: 653–655
- [42] Bozorg MAE, Motesharrei S, Penny SG, *et al.* Causality analysis: Identifying the leading element in a coupled dynamical system. *PLoS One*, 2015, **10**(6): e0131226
- [43] Packard NH, Crutchfield JP, Shaw RS. Geometry from a time series. *Physical Review Letters*, 1980, **45**: 712–716
- [44] Van Nes EH, Scheffer M, Brovkin V, *et al.* Causal feedbacks in climate change. *Nature Climate Change*, 2015, **5**: 445–448
- [45] Tsonis AA, Deyle ER, Ye H, *et al.* Convergent cross mapping: Theory and an example// Tsonis A, eds. *Advances in Nonlinear Geosciences*. Cham, Germany: Springer, 2018: 587–600
- [46] Liu HJ, Lei MY, Zhang NX, *et al.* The causal nexus between energy consumption, carbon emissions and economic growth: New evidence from China, India and G7 countries using convergent cross mapping. *PLoS One*, 2019, **14**(5): e0217319
- [47] Tsonis AA, Deyle ER, May RM, *et al.* Dynamical evidence for causality between galactic cosmic rays and interannual variation in global temperature. *Proceedings of the National Academy of Sciences of the United States of America*, 2015, **112**: 3253–3260
- [48] Nakayama SI, Masato SA, Hiroshi O. An introduction to convergent cross mapping: A novel method for causality detection in ecological time series. *Japanese Journal of Ecology*, 2015, **65**: 241–253
- [49] Ye H. *Nonlinear Tools for a Nonlinear World: Applications of Empirical Dynamic Modeling to Marine Ecosystems*. PhD Thesis. San Diego, CA, USA: University of California, 2015
- [50] Vadicamo L, Mic V, Falchi F, *et al.* Metric embedding into the hamming space with the n-simplex projection. International Conference on Similarity Search and Applications, Newark, DE, USA, 2019: 265–272
- [51] Petchey OL. Simplex Projection Walkthrough [EB/OL]. (2016-12-22) [2021-02-26]. <https://zenodo.org/record/57081>
- [52] Matin MA, Bourque CPA. Relating seasonal dynamics of enhanced vegetation index to the recycling of water in two endorheic river basins in north-west China. *Hydrology and Earth System Sciences*, 2015, **19**: 3387–3403
- [53] Naoki T, Zubaidah AAH, Natra J, *et al.* Small temperature variations are a key regulator of reproductive growth and assimilate storage in oil palm (*Elaeis guineensis*). *Scientific Reports*, 2020, **10**: 1–11
- [54] Nakayama S, Takasuka A, Ichinokawa M, *et al.* Climate change and interspecific interactions drive species alternations between anchovy and sardine in the western North Pacific: Detection of causality by convergent cross mapping. *Fisheries Oceanography*, 2018, **27**: 312–322
- [55] Kawatsu K, Kishi S. Identifying critical interactions in complex competition dynamics between bean beetles. *Oikos*, 2017, **127**: 553–560
- [56] Hsieh C, Glaser S, Lucas A, *et al.* Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean. *Nature*, 2005, **435**: 336–340
- [57] Cenci S, Sugihara G, Saavedra S. Regularized S-map for inference and forecasting with noisy ecological time series. *Methods in Ecology and Evolution*, 2019, **10**: 650–660
- [58] Deyle ER, May RM, Munch SB, *et al.* Tracking and forecasting ecosystem interactions in real time. *Proceedings of the Royal Society B: Biological Sciences*, 2016, **283**: 2015–2258
- [59] Vannitsem S, Ekemans P. Causal dependences between the coupled ocean-atmosphere dynamics over the tropical Pacific, the North Pacific and the North Atlantic. *Earth System Dynamics*, 2018, **9**: 1063–1083
- [60] Deyle ER, Maher MC, Hernandez RD, *et al.* From the cover: Global environmental drivers of influenza. *Proceedings of the National Academy of Sciences of the United States of America*, 2016, **113**: 13081–13086
- [61] Thiel M, Romano MC, Kurths J, *et al.* Twin surrogates to test for complex synchronisation. *Europhysics Letters*, 2006, **75**: 535–541
- [62] Masayuki U, Heish CH, Masuda R, *et al.* Fluctuating interaction network and time-varying stability of a natural fish community. *Nature*, 2018, **554**: 360–363
- [63] Schiecke K, Pester B, Feucht M, *et al.* Convergent cross mapping: Basic concept, influence of estimation parameters and practical application. International Conference of the IEEE Engineering in Medicine and Biology Society, Milan, 2015: 7418–7421
- [64] Park J, Smith C, Sugihara G, *et al.* rEDM: Empirical Dynamic Modeling (‘EDM’) [EB/OL]. <https://CRAN.R-project.org/package=rEDM>
- [65] Kawatsu K, Yamanaka T, Patoëka J, *et al.* Nonlinear time series analysis unravels underlying mechanisms of interspecific synchrony among foliage-feeding forest *Lepidoptera* species. *Population Ecology*, 2020, **62**: 5–14
- [66] Zhang NN, Wang GL, Tsonis AA. Dynamical evidence for causality between Northern Hemisphere annular mode and winter surface air temperature over Northeast Asia. *Climate Dynamics*, 2019, **52**: 3175–3182
- [67] Yao L. Causative impact of air pollution on evapotranspiration in the North China Plain. *Environmental Research*, 2017, **158**: 436–442
- [68] Chen ZY, Cai J, Gao BB, *et al.* Detecting the causality

- influence of individual meteorological factors on local $PM_{2.5}$ concentration in the Jing-Jin-Ji region. *Scientific Reports*, 2017, **7**: 40735–40746
- [69] 王世林, 曹文侠, 王小军, 等. 河西走廊荒漠盐碱地人工柽柳林土壤水盐分布. *应用生态学报*, 2019, **30**(8): 2531–2540 [Wang S-L, Cao W-X, Wang X-J, *et al.* Distribution of soil moisture and salt of *Tamarix ramosissima* plantation in desert saline-alkali land of Hexi Corridor Region, China. *Chinese Journal of Applied Ecology*, 2019, **30**(8): 2531–2540]
- [70] 车盈, 金光泽. 物种多样性和系统发育多样性对阔叶红松林生产力的影响. *应用生态学报*, 2019, **30**(7): 2241–2248 [Che Y, Jin G-Z. Effects of species diversity and phylogenetic diversity on productivity of a mixed broadleaved-Korean pine forest. *Chinese Journal of Applied Ecology*, 2019, **30**(7): 2241–2248]
- [71] 石芳忠, 李小雁, 吴秀臣, 等. 青藏高原植被生长对PDO响应的季节分异. *应用生态学报*, 2018, **29**(4): 1107–1116 [Shi F-Z, Li X-Y, Wu X-C, *et al.* Seasonal divergence in the response of vegetation growth to PDO in Tibetan Plateau, China. *Chinese Journal of Applied Ecology*, 2018, **29**(4): 1107–1116]
- [72] Spirtes P, Zhang K. Causal discovery and inference: Concepts and recent methodological advances. *Applied Informatics*, 2016, **3**: 1–28
- [73] Grace JB. Structural Equation Modeling Natural Systems. Cambridge, UK: Cambridge University Press, 2006: 3–37
- [74] 温腾, 徐德琳, 徐驰, 等. 全球变化背景下的现代生态学——第六届现代生态学讲座纪要. *生态学报*, 2012, **32**(11): 3606–3612 [Wen T, Xu D-L, Xu C, *et al.* An overview on the 6th international symposium on modern ecology series of 2011. *Acta Ecologica Sinica*, 2012, **32**(11): 3606–3612]
- [75] 杨晓. 生态学分支学科介绍(2). *干旱区研究*, 1999, **16**(3): 73–75 [Yang X. Introduction to the subdisciplines of ecology (2). *Arid Zone Research*, 1999, **16**(3): 73–75]
- [76] 杨晓. 生态学分支学科介绍(1). *干旱区研究*, 1999, **16**(2): 71–73 [Yang X. Introduction to the subdisciplines of ecology (1). *Arid Zone Research*, 1999, **16**(2): 71–73]
- [77] 罗秉征. 海洋生态学. *地球科学进展*, 1993, **8**(5): 94–96 [Luo B-Z. Marine ecology. *Advance in Earth Sciences*, 1993, **8**(5): 94–96]
- [78] 陈宜菲. 污染生态学的学科分支进展. *资源与人居环境*, 2008(20): 62–64 [Chen Y-F. Advances in the subdiscipline of pollution ecology. *Resource and Habitat Environment*, 2008(20): 64–66]
- [79] Gates DM. Biophysical Ecology. New York: Springer, 1980: 3–10
- [80] Yang AC, Peng CK, Huang NE. Causal decomposition in the mutual causation system. *Nature Communications*, 2018, **9**: 3378–3388
- [81] Cao L, Mees A, Judd K. Dynamics from multivariate time series. *Physica D: Nonlinear Phenomena*, 1998, **121**: 75–88
- [82] Ye H, Deyle ER, Gilarranz LJ, *et al.* Distinguishing time-delayed causal interactions using convergent cross mapping. *Scientific Reports*, 2015, **5**: 1–9
- [83] 罗磊, 程非凡, 邱彤, 等. 改进CCM算法检测外部扰动下系统变量间的时滞和因果关系. *化工学报*, 2016, **67**(12): 5122–5130 [Luo L, Cheng F-F, Qiu T, *et al.* An improved convergent cross mapping algorithm for causality identification and time delay analysis between systemic variables under external disturbance. *Journal of Chemical Industry and Engineering*, 2016, **67**(12): 5122–5130]
- [84] Hsieh C, Anderson C, Sugihara G. Extending nonlinear analysis to short ecological time series. *American Naturalist*, 2008, **171**: 71–80
- [85] Clark AT, Ye H, Isbell F, *et al.* Spatial ‘convergent cross mapping’ to detect causal relationships from short time-series. *Ecology*, 2015, **96**: 1174–1195
- [86] Neeti N, Eastman JR. A contextual Mann-Kendall approach for the assessment of trend significance in image time series. *Transactions in GIS*, 2011, **15**: 599–611

作者简介 王丹雨, 女, 1996年生, 博士研究生. 主要从事荒漠生态学研究. E-mail: dy_w9655@163.com

责任编辑 肖红
