

# 在“春季可预测壁垒”之前基于复杂方法 对大规模厄尔尼诺和南方涛动预测

Jun Meng<sup>a,1</sup>, Jingfang Fan<sup>a,b,1,2</sup>, Josef Ludescher<sup>a,1</sup>, Ankit Agarwal<sup>a,c,d</sup>, Xiaosong Chen<sup>b</sup>, Armin Bunde<sup>e</sup>, Jurgen Kurths<sup>a,f</sup>和  
Hans Joachim Schellnhuber<sup>a,2</sup>

<sup>a</sup>波茨坦气候影响研究所, 14412 德国波茨坦; <sup>b</sup>北京师范大学系统科学学院, 100875 北京; <sup>c</sup>印度鲁尔基理工学院水文系, 247667 印度鲁尔基; <sup>d</sup>第4.4  
部分: 水文学, GFZ 德国地球科学研究中心, 14473 波茨坦, 德国; <sup>e</sup>德国吉森大学物理理论研究所和 <sup>f</sup>德国柏林洪堡大学物理系, 10099.

本文由 Hans Joachim Schellnhuber 于 2019 年 11 月 5 日提供 (2019 年 10 月 1 日发送审核, 由 Zuntao Fu 和 Yoshito Hirata 审稿)

## 摘要

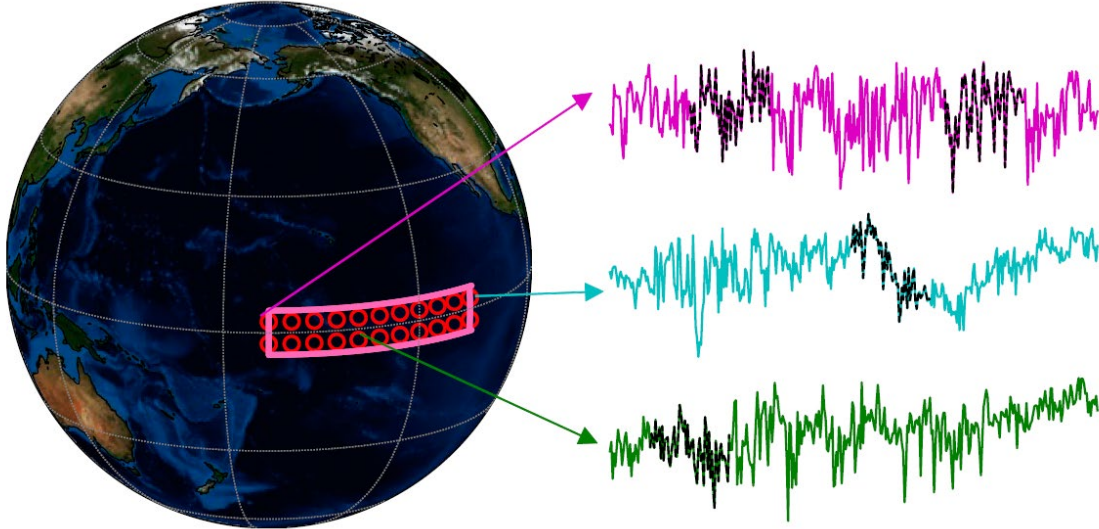
厄尔尼诺与南方涛动 (ENSO) 是最显著的国际气候现象。正因为其对经济、社会、生态有着严重的影响, 尽早以及可靠的预测 ENSO 是一个十分重要的目标。尽管动力学以及统计学模型在今年内蓬勃发展, “春季可预测壁垒”对于长时间预测 (例如超过 6 个月的预测) 仍然是一个很大的挑战。为了攻克这一壁垒, 我们提出一种分析工具: 系统样本熵 (SysSampEn), 去测量由厄尔尼诺 3.4 地区的温度异常事件序列组成的复杂 (扰乱) 的系统。当我们将此工具应用于近地面和海面空气温度数据集时, 发现所有数据集中厄尔尼诺现的强弱与上一年的 SysSampEn (复杂性) 有很强正相关关系。我们展示这种相关性能够使我们在一年内精准地预测厄尔尼诺现象的大小 (即: 每一个数据预测集的平均值的 RMSE = 0.23°C)。在 2018 年的厄尔尼诺事件中, 我们的方法预测了一个温度为:  $1.11 \pm 0.23^\circ\text{C}$  的弱厄尔尼诺现象。我们展现的这个架构不仅能够长期预测厄尔尼诺的强弱而且也可能可以用于衡量其他自然工程复杂性的指标。

**关键词:** ENSO, 系统复杂性, 熵, 春季壁垒, 预测

## 介绍

ENSO 是热带太平洋异常冷暖条件之间的年际波动, 是地球上最具影响力的海洋——大气气候现象之一<sup>[1-4]</sup>。ENSO 的暖相期的特点为: 在赤道太平洋东部每 2~7 年发生一次的异常变暖。海洋尼诺指数<sup>[5]</sup> (ONI) 是国家海洋局和大气管理局用来检测和识别 ENSO 事件的主要指标。ONI 指的是尼诺 3.4 区域 (即  $5^\circ\text{S} - 5^\circ\text{N}, 170^\circ\text{W} - 120^\circ\text{W}$ , 如图 1 所示的粉色矩形区域内的区域) 的海面温度 (SST) 异常的 3 米内的平均值。如果 ONI 连续至少 5 个月处于或高于  $0.5^\circ\text{C}$  则定义 ENSO 事件发生 (如图 2 中红色区域)。本文使用 ENSO 事件期间 ONI 最大值来量化 ENSO 事件的多少。

据报道, ENSO 会影响海洋生态系统, 商业渔业, 农业和公共安全, 甚至会导致全球许多地方出现极端天气<sup>[6-14]</sup>。因此, 对 ENSO 潜在机制的预测的理解对人类来说十分重要。多年来, 人们提出了包括动力学和统计两大类模型去模拟和预测 ENSO 事件。动力学模型<sup>[15-24]</sup>在数学上表达了海洋——大气系统的物理方程, 而相比之下, ENSO 的统计模型预测<sup>[25-26]</sup>基于数据驱动的分析。在过去的几十年里, ENSO 的预测取得了极大的进展, 并在较短的时间 (最多约 6 个月)<sup>[27-29]</sup>下可以较为精准的预测。但是, 这两种模型在北方春季前 (2-3 月) 预测水平都较低, 这就是所谓的“春季预测壁垒” (SPB)<sup>[30-33]</sup>。



**图 1.** 尼诺 3.4 区。红色圆圈表示区域的 22 个节点，空间分辨率为  $5^\circ \times 5^\circ$ 。这些曲线是一个特定年份尼诺 3.4 区域 3 个节点的温度异常时间序列示例，其子序列的几个示例用黑色标记。

近期，人们又提出了几种基于气候网络的方法，以提前 1 年左右预测 ENSO 的开始<sup>[34-37]</sup>。其中一种方法<sup>[34]</sup>正确的预测了自从 2012 年以来所有 ENSO 事件的发生或消失，不过这种方法无法预测 ENSO 事件的严重程度。但是，预测 ENSO 严重程度是十分重要的，因为较强的 ENSO 通常会导致更加极端的事件（例如：洪水、干旱或严重风暴），这些事件对于经济，社会和生态产生了严重的后果。特别是，始于 1997 年和 2014 年的厄尔尼诺现象表现了强厄尔尼诺现象并且对世界范围内的热带、温带地区都有很大的影响。为了填补此块空缺，我们提出了一种分析工具：SysSampEn，它可以量化厄尔尼诺 3.4 地区的时空无序性在温度差异中的程度并于 SPB 之前预测厄尔尼诺的强度。根据一年的数据，我们可以在 SPB 之前预测下一年厄尔尼诺是否会发生。一旦 SysSampEn 方法预测厄尔尼诺发生，我们就能用较好的技巧预测其强弱（即，在过去的 35 年内，预测和观测数据的厄尔尼诺现象的相关性为  $r = 0.84$ ，而 RMSE 为  $0.23^\circ\text{C}$ ）。我们惊喜的发现，我们的 SysSampEn 方法相比于当下的预测方法，可预测时长提升了两倍。我们的预测 ENSO 强弱的技巧是基于前一年的 SysSampEn，因此提前期为一年，与当下最先进的模型提前期相当（提前六个月）即，从六月开始预测，并预测同年北方冬季（11 月~1 月）的 ONI。

## SysSampEn

我们将复杂系统的 SysSampEn 定义为样本熵（SampEn）和交叉熵<sup>[41]</sup>的归一化。样本熵是改进的近似熵<sup>[42,43]</sup>，他可以测量与马尔科夫熵和由单个时间序列表示的过程的信息生产率相关的复杂性。而我们使用交叉熵测量两个相关时间序列之间的差异程度<sup>[41,45]</sup>。样本熵和交叉熵都已经广泛的应用于生理领域，例如通过分析心率变异性在新生儿败血症的临床体征之前进行更早的诊断<sup>[46]</sup>，利用癫痫脑电图的自动诊断<sup>[47]</sup>，以及通过分析人类姿势摇摆数据来区分不同感觉状态如何<sup>[48]</sup>。

然而，一个像气候系统这样的复杂的系统，通常由几个相关的时间序列组成（例如图 1 的曲线）。因此，在这里，我们介绍 SysSampEn 作为系统复杂性的一种度量，以同时量化复杂系

统的所有时间序列的平均时间无序程度以及他们之间的异步性。具体而言，它近似等于  $m$  个连续数据点的 2 个相似子序列（在一定允许误差范围内）对于后  $p$  个点保持相似条件概率的负自然对数，其中子序列可以来自相同或不同的时间序列（例如，中的黑色曲线），即：

$$\text{SysSampEn}(m, p, l_{\text{eff}}, \gamma) = -\log\left(\frac{A}{B}\right) \quad (1)$$

其中  $A$  是长度为  $(m + p)$  的相似子序列的对数， $B$  是长度为  $m$  的相似子序列的对数， $l_{\text{eff}}$  是长度为  $l$  的每个时间序列计算中使用的数据点数量（ $\leq l$ ）， $\gamma$  是一个常数，它决定了允许误差的范围。数据和方法中描述了由  $N$  个时间序列组成的任意复杂系统的 SysSampEn 的详细定义。当  $N = 1, p = 1, l_{\text{eff}} = 1$  时，我们的定义等同于经典的样本熵<sup>[41]</sup>。与样本熵和交叉熵的情况一样，在将 SysSampEn 用作有效工具之前，必须确定适当的参数值，因为只有某些值组合可用于相当准确地估计系统的复杂性。为了更好地证明我们的系统抽样的数学意义，我们在 SI 中展示了一个应用系统抽样估计系统复杂性并将其与 Lyapunov 指数进行比较的示例（逻辑图）。我们发现 SysSampEn 的高（低）值与高（低）Lyapunov 指数密切相关，这表明 SysSampEn 能够很好地捕捉系统的复杂性。然而，我们注意到，在不同的复杂系统中，有效参数组合可能不同。对于我们为厄尔尼诺预测而研究的气候系统，我们选择  $m$  为 30 天或 60 天， $p$  为 15 天或 30 天，因为厄尔尼诺是一种年际现象。

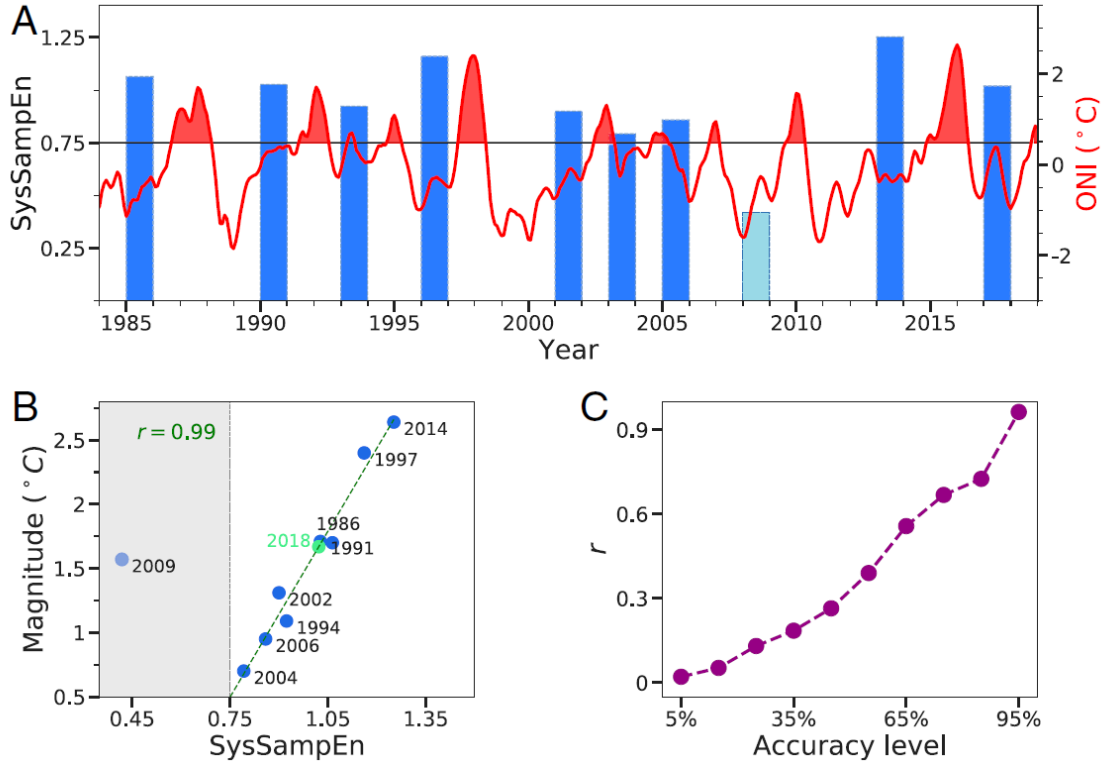
### 厄尔尼诺震级与其上一日历年的 SysSampEn 之间存在强正相关

我们计算了厄尔尼诺 3.4 区域由近地表空气或 SST 异常时间序列组成的气候系统的系统采样点，并发现厄尔尼诺量级与其上一日历年的系统采样点之间存在强正相关（图 2 中的 A 和 B）。这种正相关性在所有分析数据集中都是显著的（平均  $r = 0.90$ ）和稳健的（见 SI）。

在下表中，我们展示了基于 ERA-Missional 数据集的结果，该数据集的相关性最高。对于 1984 年至 2018 年之间的给定日历年，我们构建了一个系统，该系统由厄尔尼诺 3.4 区域（图 1）的温度异常时间序列组成，空间分辨率为  $5^\circ \times 5^\circ$ 。

首先，我们确定 SysSampEn 的参数组合，从而能够准确估计系统的复杂性。我们通过执行 2 个测试（详情见数据和方法）来实现这一点，对于给定的参数组合，这些测试确定 SysSampEn 区分无序系统能力的高低。在时间无序测试中，我们将随机数添加到真实温度数据中，而在空间异步测试中，我们比较了两个系统，一个是由地球上的相邻点构造的，另一个是由地球上随机选择的点构造的。准确的复杂性度量应该能够识别更随机系统中的更高无序度，从而为其分配更高的 SysSampEn 值。我们将准确度定义为正确任务的百分比。因此，使用合适的系统参数组合，可以量化系统中的时间和空间无序。

令人惊讶的是，我们发现上一日历年的 SysSampEn 与厄尔尼诺量呈强正相关，因而 SysSampEn 的参数组合能够很好地量化系统复杂性。在图 2.C 的一个示例中显示， $m = 60\text{d}$  和  $p = 15\text{d}$ ，随着式(1.1)中  $l_{\text{eff}}$  和  $\gamma$  值的改变，厄尔尼诺震级和上一日历年的系统采样点（如图 2.A 中的蓝色矩形）之间的皮尔逊相关性（ $r$ ）随精度水平显著增加。值得注意的是，精度的计算完全独立于任何厄尔尼诺震级分析或预测。因此，SysSampEn 和厄尔尼诺量级之间的强相关性是从参数组合中自然获得的，这使得 SysSampEn 能够以高精度量化系统复杂性。换句话说，SPB 前厄尔尼诺量级的高度可预测性并非过拟合的结果，而是源于系统复杂性和厄尔尼诺量级之间强相关性和鲁棒性。



**图 2.** SysSampEn 和厄尔尼诺震级之间的相关性。(A) 蓝色矩形的高度表示 SysSampEn 的值 (左刻度) 对于厄尔尼诺事件之前的日历年, 使用一组参数 ( $m = 60d, p = 15d, l_{eff} = 345d, \gamma = 9$ ) 计算厄尔尼诺事件, 这些参数对应于厄尔尼诺震级的最高相关性  $r$ 。红色曲线为 ONI, 红色阴影表示厄尔尼诺周期 (右刻度)。(B) 最大厄尔尼诺震级与上一日历年 SysSampEn 的散点图 (A 中的蓝色矩形)。灰色区域表示 SysSampEn 的值, 该值预测最大 ONI 小于  $0.5^\circ$ 。因此, 根据定义, 非厄尔尼诺事件。绿色虚线显示最佳最小二乘拟合线。(C) 每个紫色点的  $y$  坐标是参数组合的平均相关  $r$ , 在空间异步和时间无序测试中, 精度不低于某一水平 (即其  $x$  坐标)。SysSampEn 与厄尔尼诺量级之间的相关性  $r$  随着精度水平的提高而单调增加。准确度水平的计算与任何厄尔尼诺事件无关, 因此系统采样笔和厄尔尼诺震级之间的强相关性自然出现, 无需拟合。

我们还发现, 1984 年至 2018 年期间, SysSampEn 的模式与数据分辨率无关, 并且对于提供高精度的不同参数组合高度一致 (SI 附录, 图 S3 和表 S1)。特别是厄尔尼诺震级与上一日历年不同影响因子系统的相关性 ( $\geq 95\%$  的准确度水平) 参数均显著较高 (平均  $r$  为  $0.83 \pm 0.12$ ), 而对于  $m = 60d, p = 15d, l_{eff} = 345d, \gamma = 9$ , 则获得了最佳相关性  $r = 0.99$ 。(图 2.B)

我们对其他数据集进行了相同的分析, 并获得了类似的结果 (SI 附录, 图 S4-S6)。我们还在 SI 附录图 S2 中展示了厄尔尼诺量与上一日历年系统采样点的散点图, 该散点图给出了其他 3 个数据集的最高  $r$ 。其他 3 个数据集的相关性  $r$  也显著较高, 使用 4 个数据集 (SI 附录, 表 S1) 的所有高精度参数组合时的平均  $r$  为  $0.79 \pm 0.11$ 。请注意, 2009 年厄尔尼诺是发病预测 (下文讨论) 中唯一遗漏的事件, 是线性关系中的一个例外。



**表 1.** 1984 年至 2017 年期间，表明厄尔尼诺震级与其上一日历年 SysSampEn 之间相关性最高的参数值

Data			Parameter				<i>r</i>	2018, °C
Type	Name	Resolution	<i>m</i> , d	<i>p</i> (= <i>q</i> ), d	$\gamma$	<i>l</i> <sub>eff</sub> , d		
T 1,000 hPa	ERA-Interim	5°	60	15	9	345	0.99	1.67
	ERA5	5°	30	30	8	330	0.87	0.58
SST	ERA5	5°	30	30	5	330	0.86	1.09
	JRA-do	4°	30	30	5	360	0.87	1.09
Average							0.90	1.11

为了获得最佳的预测性能，我们通过首先进行精度测试并仅接受在空间异步和时间无序测试中都具有较高精度（空气温度准确度 $\geq 95\%$ 且 SST $\geq 85\%$ ）的参数组合来选择 SysSampEn 参数。从这些高精度参数组合中，我们在第二步中选择了与过去厄尔尼诺事件的震级具有最高相关性 *r* 的参数组合。我们对所有数据集重复此操作。**表 1** 显示了不同数据集中 2018 年之前厄尔尼诺事件的最高相关性。

我们要注意的是，通过对尼诺 3.4 区域中每个节点的平均 SampEn 或每对节点的平均交叉熵重复的计算，我们得到的相关性（平均 *r* = 0.42）比在 SysSampEn 方法中不显著。

### 厄尔尼诺开始时间和强度预测

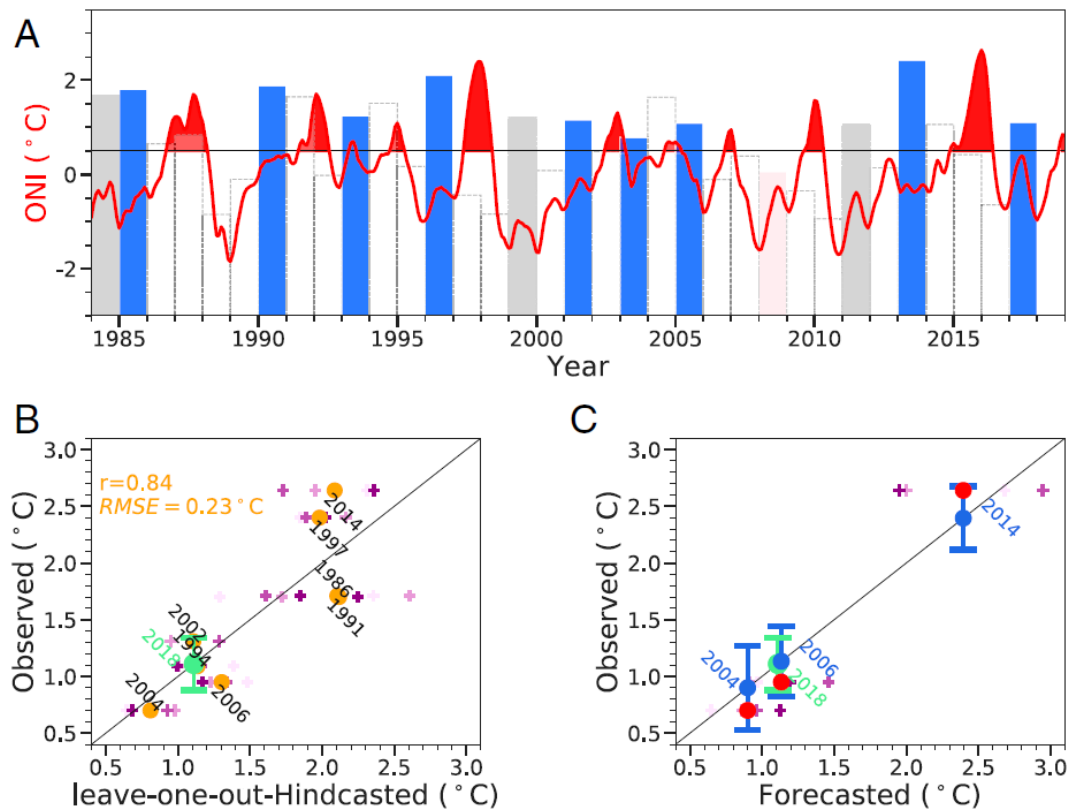
基于 SysSampEn 和厄尔尼诺强度之间的实质性相关性，我们开发了有效的厄尔尼诺开始时间和强度的预测方法（在**数据和方法**这一章中介绍）。

为了显示 SPB 前厄尔尼诺现象的高度可预测性，我们通过将每个历年年的 SysSampEn 值替换为最佳拟合的线性函数（**图 2.B** 和 SI 附录中的绿色虚线，**图 S2**），然后取所有 4 个数据集的平均值，组成了一个新的指数（**图 3.A** 中的矩形），新指数的单位是摄氏度。我们发现，一个特定历年年的该指数值可用于预测下一年是否出现厄尔尼诺现象，且具有很高的准确性：即厄尔尼诺现象的 10 个正确预测中有 9 个（深蓝色矩形），只有一个遗漏（粉红色矩形）24 个厄尔尼诺无发病缺勤年的正确预测中有 21 个（透明矩形），3 个（灰色矩形）缺失。在数据和方法上介绍了详细的算法。

为了证明 SPB 前厄尔尼诺磁事件的高度可预测性，我们首先对 1984 年至 2017 年期间所有厄尔尼诺事件的强度进行了剔除后预测（如**数据和方法**所述）。对于每个数据集，我们在 SysSampEn 函数中使用参数组合，该参数组合给出了 SysSampEn 与 2018 年前厄尔尼诺事件量级之间的最高相关性 *r*（**表 1**）。观察到的厄尔尼诺强度和后报强度如**图 3.B** 所示。与实际数据相比，我们发现我们的后预测方法非常有效，具有相当高的精度，即均方根误差（RMSE）=0.23°C.这表明 SysSampEn 方法有可能进行熟练的厄尔尼诺震级预测，预测范围为 1 年。

其次，我们仅使用事件过去的数据（数据和方法）对 2004 年、2006 年和 2014 年厄尔尼诺事件进行强度预测，发现观测值和预测值之间的 RMSE 在正负 1 以内（**图 3.C**）。这些结果表明，RMSE 上下一度可以视为一个误差条。RMSE 是通过仅适用于所考虑事件的过去（例如，对于 2004 年的厄尔尼诺事件，它仅取决于 1984 年至 2003 年期间）的遗漏后预测获得的。类似地，SysSampEn 参数也仅取决于所考虑事件的过去，并在 SI 附录**表 S2-S4** 中给出。注意，对于后来的厄尔尼诺事件，随着我们的方法有更多的数据可用，估计的 RMSE 变得更

小（图 3.C）。最近 3 次厄尔尼诺事件的预测性能证明了我们的方法预测厄尔尼诺震级以及提供正确误差估计的能力。



**图 3.** 预测厄尔尼诺事件和震级。(A) 发病预测指数的值（4 个数据集的平均预测值）显示为矩形高度，用于预测下一年是否出现厄尔尼诺发病。如果索引值为  $\geq 0.5^{\circ}C$ ，12 月的观测值低于  $0.5^{\circ}C$ ，我们预测厄尔尼诺现象将在下一年出现。蓝色矩形表示正确预测的厄尔尼诺事件，粉色矩形表示错过厄尔尼诺事件，灰色矩形表示假警报，透明矩形表示何时正确预测没有厄尔尼诺事件。(B) 观测温度与 2018 年前厄尔尼诺震级（橙色点）后测温度的对比。得到的 RMSE 为  $0.23^{\circ}C$ 。预测震级（ $1.11^{\circ}C$ ）2018 年厄尔尼诺事件中，没有事件被绘制为带有  $1 \times RMSE$  误差条的浅绿点。“+” 符号表示通过使用 4 个数据集集中的每个数据集获得的后预测（预测）值。(C) 对 2004 年、2006 年和 2014 年厄尔尼诺的预测仅仅基于过去的信息。每个预测厄尔尼诺事件（蓝点）的误差条等于  $1 \times RMSE$ （即  $0.37^{\circ}C$ ， $0.31^{\circ}C$  和  $0.28^{\circ}C$  分别适用于 2004 年、2006 年和 2014 年的事件），并根据过去事件中的遗漏事后预测进行计算。因此，预测值及其误差条（即  $1 \times RMSE$ ）仅基于事件的过去信息。红点表示观测到的震级，并且在误差条内。预测的 2018 年震级及其误差条显示为浅绿色。

接下来，我们仅根据截至 2017 年的数据，应用 SysSampEn 方法预测 2018 年厄尔尼诺事件的强度。表 1 给出了使用的 SysSampEn 参数，并获得了其强度为  $1.11^{\circ}C$ 、误差为  $0.23^{\circ}C$ 。

## 讨论

我们介绍了复杂系统的 SysSampEn, 并将其应用于估算厄尔尼诺 3.4 区域温度变化的时空无序度。我们发现, 海温或近地表气温的低水平同步度和高水平随机时间变化是强厄尔尼诺现象的前兆。我们的结果显示, 厄尔尼诺爆发年的北半球春季之前, 厄尔尼诺事件和强度都具有很高的可预测性。1984 年至 2018 年间, 我们的方法正确预测了 10 例厄尔尼诺发作中的 9 例, 而 24 例中的 21 例正确预测了没有厄尔尼诺发作。对于正确预测的厄尔尼诺的量级, 我们获得了  $0.23^{\circ}\text{C}$  的预测 RMSE。特别是, 对于 2018 年开始的最后一次厄尔尼诺, 我们的方法仅基于 2017 年的数据预测厄尔尼诺较弱, 量级为  $1.11 \pm 0.23^{\circ}\text{C}$ 。

我们注意到, 对于接近半年的较短  $I_{\text{eff}}$ , 在某些参数范围内, 厄尔尼诺震级和 SysSampEn 之间的相关性仍然很高 (SI 附录, 图 S3-S6 C)。这表明, 即使是更早的厄尔尼诺现象, 也可以提前 18 个月进行预测, 但预测技巧较低。

作为我们方法的可能机制, 我们从近地表海洋湍流与 SST 变化之间的关系中找到了一些线索<sup>[52]</sup>。最近, 我们发现, 强烈的厄尔尼诺与强烈的海洋湍流有关, 其特点是横向扩散率大<sup>[53-55]</sup>。厄尔尼诺过程中横向扩散率的增强导致水平温度梯度变弱和水平混合变高, 从而导致厄尔尼诺 3.4 区域的系统采样率降低。我们的发现支持了这一推测, 即厄尔尼诺年的系统采样点与厄尔尼诺量级成反比, 也就是说, 对于包含强烈厄尔尼诺的年, 系统采样点往往较低 (详情见 SI 中的图 S7)。此外, 我们还观察到 SysSampEn 动态演化中的记忆效应, 即: 较小的 SysSampEn 后面更有可能跟着较大的 SysSampEn, 较大的 SysSampEn 后面跟着较小的 SysSampEn, 如 SI 图 S8 所示。因此, 具有较强厄尔尼诺效应和较低系统采样点的日历年很可能在具有较高系统采样点的日历年之前。我们假设, 这一较高的 SysSampEn 可能是由于厄尔尼诺爆发前一日历年内地表横向扩散率较弱所致, 但需要根据气候模型和观测数据进行进一步分析。我们注意到, 刚刚发现某些区域海洋表面中尺度湍流的扩散率与厄尔尼诺 3.4 指数<sup>[55]</sup>相关, 这支持了我们的假设。我们还注意到, 发展渗流相变分析是为了研究 ENSO 对气候的影响, 并进一步帮助我们更好地预测 ENSO 触发的后续事件<sup>[56, 57]</sup>。

这里开发的理论框架不仅提高了长前置时间厄尔尼诺预测能力, 而且在应用于其他非线性和复杂动力系统时, 还可能带来改进的预测或新的见解<sup>[58, 59]</sup>。

## 数据和方法

### 数据:

欧洲中期天气预报中心 (ECMWF) 的全球再分析资料使用手册 (ERA-Interim) 可在 <https://apps.ecmwf.int/datasets/> 找到。全球再分析资料使用手册是从 1979 年开始的全球大气再分析, 并定期更新。在目前的工作中, 我们使用了每天零点的近地表 (1000 hPa) 温度, 空间 (纬向和经向) 分辨率为  $2.5^{\circ} \times 2.5^{\circ}$ 。我们使用的是 1979 年至 2017 年的数据, 于 2018 年 10 月 4 日下载, 上一年 2018 年的数据于 2019 年 1 月 29 日更新。

ERA5 (<https://climate.copernicus.eu/climate-reanalysis>) 是通过哥白尼气候变化服务 (C3S) 开发的气候再分析数据集。它目前可用数据时间跨度为自 1979 年以来的到目前的前 3 个月。与 ERA 相比, ERA5 的分析场具有 31 km 的更高空间分辨率和 1 h 的更高时间分辨率。本研究中使用的数据为 2019 年 1 月 25 日下载的每日零点近地表 (1000 hPa) 温度和 2019 年 1 月 30 日下载的 SST, 空间 (纬向和经向) 分辨率为  $2.5^{\circ} \times 2.5^{\circ}$ 。

JRA55 的来源于: <https://esgf-node.llnl.gov/search/input4mips/>, 该数据从 1958 年到 2018 年, 预计每年更新一次 (每年 4 月左右)。SST 场的空间分辨率为  $1^{\circ} \times 1^{\circ}$ , 时间分辨率为 1 天。

本研究中使用的数据为 2018 年 11 月 8 日下载的日平均 SST，空间（纬向和经向）分辨率为  $1^\circ \times 1^\circ$ 。

### 数据处理:

对于自 1984 年以来的每个日历年(y)（使用数据集 ERA-Missional、ERA5 和 ERA5 SST 的 1979 年至 1983 年的前 5 年来计算 1984 年的第一个异常值），在厄尔尼诺 3.4 区域的每个网格点  $\alpha$  处，我们通过从实际温度中减去气候平均值，然后除以气候 SD 来计算异常。我们在每个日历日  $t$  都这样做。为简单起见，不包括闰日。气候平均值和 SD 的计算仅基于截至年  $y$  的过去数据。

### 用于任意复杂系统的 SysSampEn:

我们首先为任意系统定义 SysSampEn。假设我们有  $N$  个相互依赖的时间序列  $x_\alpha(t)$  ( $\alpha = 1, 2, \dots, N$ )，长度为  $l$ ，组成系统。

1) 从每个时间序列中，我们选择长度为  $m < l$  的子记录  $k$ ，从每个第  $q$  个数据点开始，即从  $t = k \times q + 1 = 0 \times q + 1, 1 \times q + 1, 2 \times q + 1 \dots$  只要  $k \times q + m \leq l$ 。因此，一个特定的子序列是  $X_\alpha^k(m, q) = \{x_\alpha(k \times q + 1), x_\alpha(k \times q + 2), \dots, x_\alpha(k \times q + m)\}$ 。然后，我们从每个时间序列中选择  $n$  个子序列，构造  $N \times n$  模板向量： $\Theta(m, q, n) = \{X_\alpha^k(m, q) : 0 \leq k \leq n-1, 1 \leq \alpha \leq N\}$ 。我们假设，如果两个向量的欧氏距离  $d(X_\alpha^i(m, q), X_\alpha^j(m, q)) < \gamma \times \max\{\sigma_\alpha, \sigma_\beta\}$  ( $\alpha = \beta$ ，则  $i \neq j$ )，其中  $\sigma_\alpha, \sigma_\beta$  分别是时间序列  $x_\alpha(t), x_\beta(t)$  的 SDs。 $\gamma$  定义了相似性标准，是一个非零常数。

2) 为了检验在  $m$  个数据点接近的 2 个时间序列在下一个  $p$  个数据点仍然接近的概率，我们通过选择长度为  $(m + p)$  的子记录来类似地构造另一个集合  $\Theta(m + p, q, n)$ 。为了使长度为  $m$  的模板向量的数量等于长度为  $(m + p)$  个模板向量。我们选择  $n \leq (l - m - p) / q + 1$ 。为了减少参数自由度和节省计算时间，我们取  $p = q$ ，然后  $n \leq (l - m) / q$ 。我们假设集合  $\Theta(m + p, q, n)$  中的 2 个模板向量是闭合的，如果  $d(X_\alpha^i(m, q), X_\alpha^j(m, q)) < \gamma \times \max\{\sigma_\alpha, \sigma_\beta\}$  ( $\alpha = \beta$ ，则  $i \neq j$ )。

3) 系统的 SysSampEn 定义为： $\text{SysSampEn}(m, p, l_{\text{eff}}, \gamma) = -\log(A / B)$ ，其中  $A$  是来自集合  $\Theta(m + p, q, n)$  的闭合向量对的数量， $B$  是来自集合  $\Theta(m, q, n)$  的闭合向量对的数量， $l_{\text{eff}}(n) = n \times p + m$  是自每年 1 月 1 日起的天数，用于计算 SysSampEn。

### SysSampEn 的参数确定:

在这里，我们演示了如何使用全球再分析资料使用手册确定我们的厄尔尼诺地区 3.4 气候系统中的 SysSampEn( $m, p, l_{\text{eff}}, \gamma$ ) 参数。对于每个日历年，我们定义一个由  $N = 22$ （图 1 中的红色圆圈）温度异常时间序列  $T_\alpha(t)$  ( $1 \leq \alpha \leq N$ ) 组成的系统，长度  $l = 365d$ 。

1) 我们选择向量长度  $m$  为 30d 或 60d，长度  $p$  增加为 15d 或 30d。由于厄尔尼诺是一种年际现象，我们将重点放在（两个月）的时间尺度上。

2) SysSampEn 的目的是量化给定系统的空间和时间无序。这意味着，如果我们有一个时空相关的复杂系统，由时间序列表示，并向每个时间序列中添加随机项（例如白噪声），那么新系统的 SysSampEn 应该以高概率大于原始 SysSampEn。类似地，如果我们用不相关的时间序列替换空间高度相关系统中的时间序列，SysSampEn 应该增加。我们使用这些性质作为 2 个



测试的基础，以确定给定  $m$  和  $p$  的  $l_{eff}$  和  $\gamma$  值，从而能够可靠地区分更多和更少的有序系统。为简单起见，我们假设  $\gamma$  为整数。

a) 空间异步测试：在随机选择的一年中，我们随机选择全球  $T_\beta, T_{\beta+1}$  和  $T_{\beta+2}$  上的 3 个相邻点。为了构造一个高度耦合的系统，我们随机选择这 3 个  $N = 22$  的节点中的乘以 1。因此，我们得到了一个具有 22 个节点的系统  $G_1$ ，其中  $T_\beta, T_{\beta+1}$  和  $T_{\beta+2}$  可能存在于具有不同频率的系统中。相比之下，我们从地球上随机选择 22 个不相关的节点来创建一个系统  $G_2$ 。我们执行此程序  $M$  次。精度定义为：

$$\text{accuracy} = \frac{1}{M} \sum_{i=1}^M S_i \quad (2)$$

其中：

$$S_i = \begin{cases} 1, & \text{SysSampEn}_{G_2} > \text{SysSampEn}_{G_1} \\ 0, & \text{Otherwise} \end{cases} \quad (3)$$

在本研究中，我们使用  $M = 100$ 。在 SI 的图 S3.A 中，对于  $m = 60, p = 15$ ，精度显示为  $l_{eff}$  和  $\gamma$  的函数。

b) 时间无序测试：我们将未受干扰的气候系统  $G_1$  的系统样本与新系统  $G_2$  进行比较，新系统  $G_2$  将随机数添加到原始时间序列中。新系统正在运行由  $N = 22$  个时间序列  $\tilde{T}_\alpha(t) = T_\alpha(t) + R_\alpha(t)$  组成。 $R_\alpha(t)$  是范围内独立均匀随机数的不相关序列  $[-0.5\sigma, 0.5\sigma]$ 。这里， $\sigma$  是 1984 年 1 月 1 日至 2018 年 12 月 31 日期间  $N = 22$  个单独时间序列 SDs 的平均值。我们执行此程序  $M$  次。精度定义如公式 2 所示，并在 SI 的图 S3.B 中显示为  $m = 60, q = 15$  的  $l_{eff}$  和  $\gamma$  的函数。

### 厄尔尼诺事件的预测算法：

如果预测指数（4 个数据集的平均预测）不小于  $0.5^\circ\text{C}$ ，12 月的观测值低于  $0.5^\circ\text{C}$ 。另外，我们预测厄尔尼诺不会发作。预测指数如图 3.A 所示，为矩形的高度。

请注意，本工作中使用的预测指数是根据 SysSampEn 与 1984 年至 2017 年期间发生的厄尔尼诺事件震级之间的显著线性关系计算的。为了预测 2018 年后厄尔尼诺事件的发生或不发生，应在厄尔尼诺事件终止后，通过选择 SysSampEn 函数，不断更新预测指数，该函数与所有终止厄尔尼诺事件的震级具有最高的相关性  $r$ 。

### 厄尔尼诺震级的预测算法：

为了预测从  $y$  年开始的厄尔尼诺事件的规模：

1) 对于一个数据集，我们通过使用与预测事件  $y$  之前发生的厄尔尼诺事件的强度有最高相关性  $r$  的数据集来确定 SysSampEn 的参数。我们只考虑能够提供高精度水平的参数组合。

2) 对于一个数据集，我们使用最小二乘回归法计算厄尔尼诺震级和上一日历年系统采样点之间的最小二乘拟合  $Y = aX + b$ 。这里  $Y$  代表厄尔尼诺事件的大小， $X$  代表对应的上一年的系统采样点。在计算最小二乘拟合时，仅使用预测事件  $y$  的过去事件。

3) 我们计算  $(y-1)$  年的 SysSampEn 并将其替换为最佳拟合线的功能。然后，我们获得从日历年  $y$  开始的厄尔尼诺事件的预期量级。

4) 对其他数据集重复步骤 1 和 2。通过取 4 个预期强度（图 3.C 中的“+”符号）的平均值，获得预测强度（图 3.C 中的蓝点）。

5) 为了确定我们预测的误差条，我们对预测的厄尔尼诺事件  $y$  的每一个过去事件执行以下“留一”后预测：

a) 与 1) 相同。

b) 为了获得每个过去事件  $\tilde{y} < y$  的遗漏后测震级，我们使用  $y$  之前发生的所有事件（后测事件除外）来计算最佳拟合线。

c) 我们计算  $(\tilde{y}-1)$  年的 SysSampEn，并将其细分为最佳拟合线的函数。然后，我们获得从日历年  $y$  开始的厄尔尼诺事件的预期量级。

d) 对其他数据集重复步骤 1 和 2。通过取 4 个预期强度（图 3.A 中的“+”符号）的平均值，获得后测试件（图 3.B 中的橙色点）。

为了预测 2018 年厄尔尼诺事件的规模，我们将每个数据集 2017 年的 SysSampEn 值替换为相应的最佳拟合线性函数，该函数由过去所有厄尔尼诺事件（2009 年事件除外）确定。因此，我们有 4 个单独的预测，我们平均得到最终预测。

### 开源数据：

支持本研究结果的数据/再分析如下：可在线公开获取： ERA-Interim<sup>[49]</sup>, <https://apps.ecmwf.int/datasets/>; ERA5<sup>[50]</sup>, <https://climate.copernicus.eu/climateanalysis>; 以及 JRA55-do<sup>[51]</sup>, <https://esgf-node.llnl.gov/search/input4mips/>.

### 致谢

我们感谢 M. J. McPhaden, S. Havlin, Y. Ashkenazy 和 N. Marwan 提出的有益建议；以及东非——秘鲁——印度气候能力项目，该项目是国际气候倡议的一部分。联邦环境、自然保护和核安全部根据德国联邦议院通过的一项决定支持这项倡议。波茨坦气候影响研究所（Potsdam Institute for Climate Impact Research）与其项目合作伙伴能源与资源研究所（Energy and Resources Institute）和德国维特迪安（Deutscher Wetterdienst）共同领导该项目的实施。

### 参考文献

1. H. A. Dijkstra, Nonlinear Physical Oceanography: A Dynamical Systems Approach to the Large Scale Ocean Circulation and El Niño (Atmospheric and Oceanographic Sciences Library, Springer, ed. 2, 2005).
2. M. J. McPhaden, S. E. Zebiak, M. H. Glantz, ENSO as an integrating concept in Earth science. *Science* 314, 1740–1745 (2006).
3. A. J. Clarke, An Introduction to the Dynamics of El Niño & the Southern Oscillation (Academic Press, Cambridge, 2008).
4. E. S. Sarachik, M. A. Cane, The El Niño-Southern Oscillation Phenomenon (Cambridge University Press, Cambridge, 2010).
5. NOAA, Oceanic Niño Index. <https://esrl.noaa.gov/psd/data/correlation/oni.data>. Accessed 20 March 2019.

6. C. F. Ropelewski, M. S. Halpert, Global and regional scale precipitation patterns associated with the El Niño/southern oscillation. *Mon. Weather Rev.* 115, 1606–1626 (1987).
7. G. N. Kiladis, H. F. Diaz, Global climatic anomalies associated with extremes in the southern oscillation. *J. Clim.* 2, 1069–1090 (1989).
8. M. S. Halpert, C. F. Ropelewski, Surface temperature patterns associated with the southern oscillation. *J. Clim.* 5, 577–593 (1992).
9. H. F. Diaz, M. P. Hoerling, J. K. Eischeid, ENSO variability, teleconnections and climate change. *Int. J. Climatol.* 21, 1845–1862 (2001).
10. K. K. Kumar, B. Rajagopalan, M. Hoerling, G. Bates, M. Cane, Unraveling the mystery of Indian monsoon failure during El Niño. *Science* 314, 115–119 (2006).
11. S. M. Hsiang, K. C. Meng, M. A. Cane, Civil conflicts are associated with the global climate. *Nature* 476, 438–441 (2011).
12. M. Burke, E. Gong, K. Jones, Income shocks and HIV in Africa. *Econ. J.* 125, 1157–1189 (2015).
13. C. F. Schleussner, J. F. Donges, R. V. Donner, H. J. Schellnhuber, Armed-conflict risks enhanced by climate-related disasters in ethnically fractionalized countries. *Proc. Natl. Acad. Sci. U.S.A.* 113, 9216–9221 (2016).
14. J. Fan, J. Meng, Y. Ashkenazy, S. Havlin, H. J. Schellnhuber, Network analysis reveals strongly localized impacts of El Niño. *Proc. Natl. Acad. Sci. U.S.A.* 114, 7543–7548 (2017).
15. S. E. Zebiak, M. A. Cane, A model El Niño-Southern Oscillation. *Mon. Weather Rev.* 115, 2262–2278 (1987).
16. J. P. McCreary, D. L. T. Anderson, An overview of coupled ocean-atmosphere models of El Niño and the Southern Oscillation. *J. Geophys. Res.: Oceans* 96, 3125–3150 (1991).
17. R. Kleeman, On the dependence of hindcast skill on ocean thermodynamics in a coupled ocean-atmosphere model. *J. Clim.* 6, 2012–2033 (1993).
18. R. Kleeman, A. M. Moore, N. R. Smith, Assimilation of subsurface thermal data into a simple ocean model for the initialization of an intermediate tropical coupled ocean-atmosphere forecast model. *Mon. Weather Rev.* 123, 3103–3113 (1995).
19. B. Wang, Z. Fang, Chaotic oscillations of tropical climate: A dynamic system theory for ENSO. *J. Atmos. Sci.* 53, 2786–2802 (1996).
20. F. F. Jin, An equatorial ocean recharge paradigm for ENSO. Part I: Conceptual model. *J. Atmos. Sci.* 54, 811–829 (1997).
21. F. F. Jin, An equatorial ocean recharge paradigm for ENSO. Part II: A stripped-down coupled model. *J. Atmos. Sci.* 54, 830–847 (1997).
22. B. Wang, A. Barcilon, Z. Fang, Stochastic dynamics of El Niño–Southern Oscillation. *J. Atmos. Sci.* 56, 5–23 (1999).
23. T. N. Palmer et al., Development of a European multimodel ensemble system for seasonal-to-interannual prediction (demeter). *Bull. Am. Meteorol. Soc.* 85, 853–872 (2004).
24. S. Saha et al., The NCEP climate forecast system. *J. Clim.* 19, 3483–3517 (2006).
25. J. S. Xu, H. Von Storch, Predicting the state of the southern oscillation using principal oscillation pattern analysis. *J. Clim.* 3, 1316–1329 (1990).

26. C. Penland, T. Magorian, Prediction of Niño 3 sea surface temperatures using linear inverse modeling. *J. Clim.* 6, 1067–1076 (1993).
27. B. P. Kirtman et al., Current status of ENSO forecast skill (Tech. Rep.56, CLIVAR Working Group on Seasonal to Interannual Prediction, 2001), p. 26.
28. D. Chen, M. A. Cane, El Niño prediction and predictability. *J. Comput. Phys.* 227, 3625–3640 (2008).
29. A. Gavrilov et al., Linear dynamical modes as new variables for data-driven ENSO forecast. *Clim. Dyn.* 52, 2199–2216 (2019).
30. P. J. Webster, S. Yang, Monsoon and ENSO: Selectively interactive systems. *Q. J. R. Meteorol. Soc.* 118, 877–926 (1992).
31. K. M. Lau, S. Yang, The Asian monsoon and predictability of the tropical ocean–atmosphere system. *Q. J. R. Meteorol. Soc.* 122, 945–957 (1996).
32. M. J. McPhaden, Tropical Pacific Ocean heat content variations and ENSO persistence barriers. *Geophys. Res. Lett.* 30, 33-1–33-4 (2003).
33. M. J. McPhaden, A 21st century shift in the relationship between ENSO SST and warm water volume anomalies. *Geophys. Res. Lett.* 39, L09706 (2012).
34. J. Ludescher et al., Improved El Niño forecasting by cooperativity detection. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11742–11745 (2013).
35. J. Meng, J. Fan, Y. Ashkenazy, S. Havlin, Percolation framework to describe El Niño conditions. *Chaos* 27, 035807 (2017).
36. J. Meng, J. Fan, Y. Ashkenazy, A. Bunde, S. Havlin, Forecasting the magnitude and onset of El Niño based on climate network. *New J. Phys.* 20, 043036 (2018).
37. P. D. Nooteboom, Q. Y. Feng, C. López, E. Hernández-García, H. A. Dijkstra, Using network theory and machine learning to predict El Niño. *Earth System Dynamics* 9, 969–983 (2018).
38. T. P. Hughes et al., Coral reefs in the anthropocene. *Nature* 546, 82–90 (2017).
39. A. G. Barnston, M. K. Tippett, M. L. L’Heureux, S. Li, D. G. DeWitt, Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing?. *Bull. Am. Meteorol. Soc.* 93, 631–651 (2011).
40. A. Wang-Chun Lai, M. Herzog, H. F. Graf, ENSO forecasts near the spring predictability barrier and possible reasons for the recently reduced predictability. *J. Clim.* 31, 815–838 (2017).
41. J. S. Richman, J. R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* 278, H2039–2049 (2000).
42. S. Pincus, Approximate entropy (ApEn) as a complexity measure. *Chaos* 5, 110–117 (1995).
43. S. M. Pincus, Quantifying complexity and regularity of neurobiological systems. *Methods Neurosci.* 28, 336–363 (1995).
44. A. N. Kolmogorov, New metric invariant of transitive dynamical systems and automorphisms of lebesgue spaces. *Dokl. Akad. Nauk SSSR* 119, 861–864 (1958).
45. S. Pincus, B. H. Singer, Randomness and degrees of irregularity. *Proc. Natl. Acad. Sci. U.S.A.* 93, 2083–2088 (1996).
46. D. E. Lake, J. S. Richman, M. P. Griffin, J. R. Moorman, Sample entropy analysis of neonatal



- heart rate variability. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 283, R789–797 (2002).
47. U. R. Acharya et al., Automated diagnosis of epileptic EEG using entropies. *Biomed. Signal Process. Control* 7, 401–408 (2012).
48. S. Ramdani, B. Seigle, J. Lagarde, F. Bouchara, P. L. Bernard, On the use of sample entropy to analyze human postural sway data. *Med. Eng. Phys.* 31, 1023–1031 (2009).
49. D. P. Dee et al., The ERA-interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* 137, 553–597 (2011).
50. Copernicus Climate Change Service (C3S), ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS). (2017) <https://cds.climate.copernicus.eu/cdsapp#/home>. Accessed 25 January 2019.
51. H. Tsujino et al., JRA-55 based surface dataset for driving ocean–sea-ice models (JRA55-do). *Ocean Model.* 130, 79–139 (2018).
52. S. A. Thorpe, *An Introduction to Ocean Turbulence* (Cambridge University Press, Cambridge, 2007).
53. Q. Schiermeier, Hunting the Godzilla El Niño. *Nature* 526, 490–491 (2015).
54. A. O. Gnanadesikan, A. O. Russell, M. A. O. Pradal, R. O. Abernathey, Impact of lateral mixing in the ocean on el nino in a suite of fully coupled climate models. *J. Adv. Model. Earth Syst.* 9, 2493–2513 (2017).
55. J. J. M. Busecke, R. P. Abernathey, Ocean mesoscale mixing linked to climate variability. *Sci. Adv.* 5, eaav5014 (2019).
56. Z. Lu, N. Yuan, Z. Fu, Percolation phase transition of surface air temperature networks under attacks of El Niño/La Niña. *Sci. Rep.* 6, 26779 (2016).
57. Z. Lu, Z. Fu, L. Hua, N. Yuan, L. Chen, Evaluation of ENSO simulations in cmip5 models: A new perspective based on percolation phase transition in complex networks. *Sci. Rep.* 8, 14912 (2018).
58. Y. C. Lai, C. Grebogi, Quasiperiodicity and suppression of multistability in nonlinear dynamical systems. *Eur. Phys. J. Spec. Top.* 226, 1703–1719 (2017).
59. W. X. Wang, Y. C. Lai, C. Grebogi, Data based identification and prediction of nonlinear and complex dynamical systems. *Phys. Rep.* 644, 1–76 (2016).