

# 通过多维可靠性方法预测全球未来的 艾滋病死亡率



Gaidai Oleg<sup>1</sup>, 王科霖<sup>1,\*</sup>, 白晓龙<sup>2</sup>, 邓航<sup>3</sup>

<sup>1</sup>上海海洋大学工程学院, 上海 201306

<sup>2</sup>江苏科技大学船舶与海洋工程学院, 江苏镇江 212003

<sup>3</sup>北京中科链源科技有限公司数据科学组, 北京 100000

**摘要:** 【目的】艾滋病是一种传染性疾病, 具有高传染性, 可在全球范围内传播, 发病率和死亡率都很高, 给全球公共卫生带来了巨大负担。该疾病的非平稳性和复杂性, 因此对这种现象进行建模具有挑战性。很少有数学模型可以使用, 因为流行病数据通常不是正态分布的。【方法】本文描述了一种新的生物系统可靠性方法, 特别适用于多区域环境和卫生系统, 在足够长的时间内观察, 从而对高致病性病毒爆发概率进行可靠的长期预测。传统的多区域过程时间观测统计方法不能有效地处理大区域维数和不同区域观测之间的相互相关性。在这项研究中, 选择了世界所有国家年度的记录的艾滋病患者人数。【结果】若世界各地环境和流行病学条件下的公共卫生系统管理得当, 预测 100 年重现期风险水平为 2.2%。【结论】这项工作旨在对最先进的方法进行基准测试, 这使得从动态观察到的患者人数中提取必要的信息成为可能, 同时考虑到相关的地域映射。本文提出的方法开辟了准确预测多区域生物系统流行病爆发概率的可能性。

**关键词:** 艾滋病; 可靠性; 概率预测; 动态系统; 公共卫生; 数理生物学

**DOI:** 10.57237/j.wjms.2023.02.002

## Projecting Future Global AIDS Mortality Through Multidimensional Reliability Methods

Gaidai Oleg<sup>1</sup>, Wang Kelin<sup>1,\*</sup>, Bai Xiaolong<sup>2</sup>, Deng Hang<sup>3</sup>

<sup>1</sup>College of Engineering, Shanghai Ocean University, Shanghai 201306, China

<sup>2</sup>School of Naval Architecture and Ocean Engineering, Jiangsu University of Science and Technology, Zhenjiang 212003, China

<sup>3</sup>Data Science Group, Beijing Zhongke Lianyuan Technology Co, Ltd, Peking 100000, China

**Abstract:** [Objectives] AIDS is an infectious disease that is highly contagious, can spread globally, has high morbidity and mortality, and poses a huge burden on global public health. The non-stationarity and complexity of the disease makes modeling this phenomenon challenging. Few mathematical models are available because epidemiological data are usually not normally distributed. [Methods] This paper describes a new approach to biosystem reliability that is particularly applicable to multiregional environments and health systems that are observed over a sufficiently long period of time to provide reliable long-term predictions of the probability of outbreaks of highly pathogenic viruses.

\*通信作者: 王科霖, 2763634512@qq.com

Conventional statistical methods for multiregional process time observations cannot effectively handle large regional dimensions and inter-correlations between observations in different regions. In this study, the annual number of recorded AIDS patients in all countries of the world was selected. [Results] If public health systems under environmental and epidemiologic conditions around the world are properly managed, the predicted 100-year return period risk level is 2.2%. [Conclusion] This work aims to benchmark state-of-the-art methods, which make it possible to extract the necessary information from dynamically observed patient numbers, taking into account the associated geographical mapping. The methodology proposed in this paper opens up the possibility of accurately predicting the probability of epidemic outbreaks in multi-regional biosystems.

**Keywords:** AIDS; Reliability; Probabilistic Prediction; Dynamic Systems; Public Health; Mathematical Biology

## 1 引言

艾滋病和其他近期类似流行病的统计方面在现代研究界受到广泛关注[1-11]。一般来说，使用传统的理论统计方法计算现实的生物系统可靠性因子和实际流行条件下的爆发概率是相当具有挑战性的[12-16]。后者通常是由于许多程度的系统自由度和控制动态生物系统的随机变量，分布在广阔的地形上。原则上，复杂生物系统的可靠性可以通过足够的测量或直接的蒙特卡罗模拟直接准确地估计。为更准确地预测和管理流行病爆发，作者在生物和卫生系统中引入了一种新的可靠性方法。本文的研究重点是全球的艾滋病[2, 7-9, 11]，重点关注同一气候带内不同省份之间的相互关系。选择全球是因为艾滋病发布广泛且传播复杂[17-20]，及其健康观察数据和在线相关研究较充足[21-28]。

生命周期数据或极值理论（EVT）的统计建模在医学或工程学中很普遍。例如，Gumbel 使用 EVT 来估计各种人群的人口状况[5, 10]，Aarssen [1]和 Galambos [4]的研究分别论证了关于预期寿命上限支持和反对。通常在这些领域撰写的论文假定从指数分布获得的参数双变量寿命分布以获得统计相关数据[3, 7, 11]。Romeo 等[9]提出了一种新方法，该方法使用了幂方差函数 copulas（例如，Clayton、Gumbel 和 Inverse Gaussian copulas）、条件采样和用于生存分析的数值逼近。Beisel 等[2]论证了 EVT 已被用于预测进化遗传学中的突变，并进一步从 EVT 中开发一个可能性框架，用于确定突变的适应度效应。Joyce 等[8]在此 EVT 理论的基础上，使用 Beta-Burr 分布确定适应度效应。Kristensen 等[29]讨论了使用双变量逻辑回归模型，然后将其用于访问

患有行走障碍的多发性硬化症患者以及用于视觉识别的认知实验。最后，Chen 等[30]研究使用了 EVT 来估计世界爆发流感的概率，展示了 EVT 预测流行病的潜力。由于没有太多的统计研究来预测流感或传染病爆发或传播的可能性，本文提出的多维可靠性方法将能够更好地洞察和预示疾病的传播。

## 2 模型的建立

在本文中，流行病的爆发被视为可能在任何时间发生在特定国家的任何地区的意外事件，因此空间传播被考虑在内。此外，引入特定的无量纲因子来预测后者在任何时间、任何地点的流行病风险。

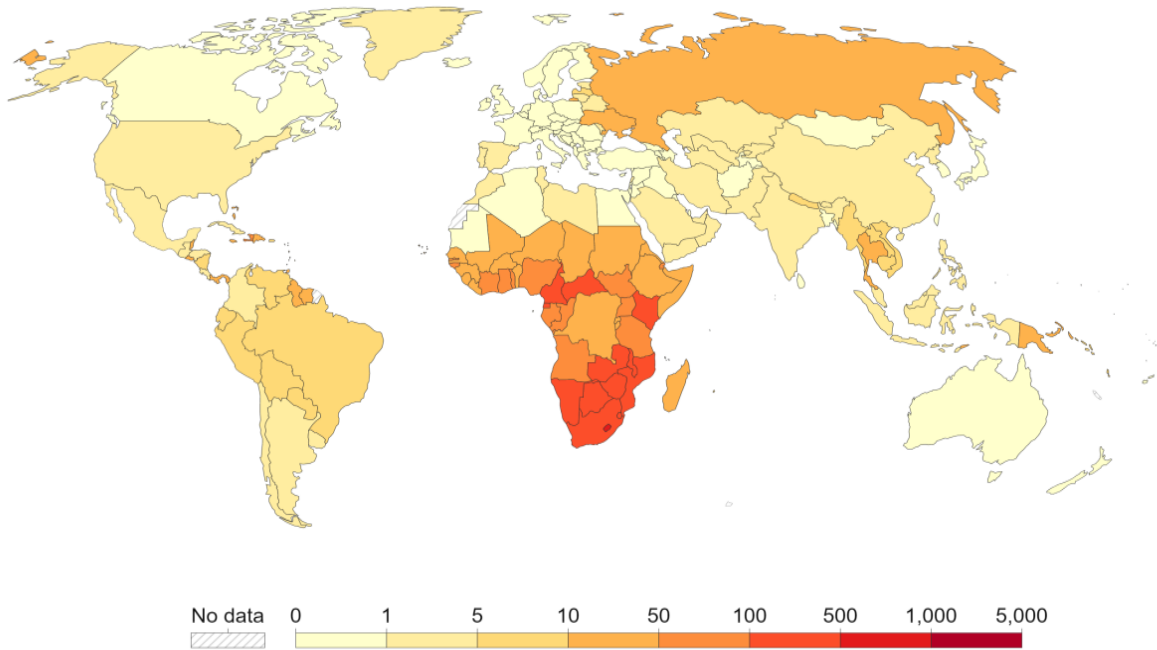
生物系统受到遍历性环境的影响。另一种选择是将过程视为依赖于特定的环境参数，其时间变化可以单独建模为遍历过程。

从公共网站检索 2020 年 2 月至今世界所有省份的艾滋病发病率数据[10]。由于这个有价值的数据集是每个世界地区的，所考虑的生物系统可以被视为具有高度相互关联的区域成分/维度的多自由度（MDOF）动态系统。最近的一些研究已经使用统计工具来预测艾滋病的发展，线性对数模型见[10]。

请注意，虽然这项研究的目的是通过预测来减少未来流行病爆发的风险，但它只关注每日登记的患者人数，而不是症状本身。图 1 介绍了 2019 年世界各国的艾滋病死亡率。

# Death rate from HIV/AIDS, 2019

The number of deaths from HIV/AIDS<sup>1</sup> per 100,000 people.



Source: IHME, Global Burden of Disease (2019)

OurWorldInData.org/hiv-aids • CC BY

Note: To allow comparisons between countries and over time this metric is age-standardized.

图 1 艾滋病毒/艾滋病死亡人数的全球分布

Figure 1 The global distribution of deaths from HIV/AIDS

在足够长的时间间隔(0, T), 测量的 MDOF 健康响应向量过程为  $R(t) = (X(t), Y(t), Z(t), \dots)$ 。整个时间跨度(0, T)上的一维全局最大值表示为:  $X_T^{\max} = \max_{0 \leq t \leq T} X(t)$ ,  $Y_T^{\max} = \max_{0 \leq t \leq T} Y(t)$ ,  $Z_T^{\max} = \max_{0 \leq t \leq T} Z(t), \dots$

足够长的时间 T 主要是指相对于动态系统自相关

时间的较大值。令  $X_1, \dots, X_{N_X}$  为过程  $X(t)$  离散单调递增时刻  $t_1^X < \dots < t_{N_X}^X$  在 (0, T) 的时间局部最大值

对于其他 MDOF 响应分量, 类似的定义如下。为简单起见, 所有分量, 因此它的最大值被假定为非负数。

$$TTX_1, \dots, X_{N_X} X(t) t_1^X < \dots < t_{N_X}^X (0, T) Y(t), Z(t), \dots Y_1, \dots, Y_{N_Y}; Z_1, \dots, Z_{N_Z} R(t)$$

目标是准确估计系统故障概率, 即超越概率:

$$1 - P = \text{Prob}(X_T^{\max} > \eta_X \cup Y_T^{\max} > \eta_Y \cup Z_T^{\max} > \eta_Z \cup \dots) \tag{1}$$

$$P = \iiint_{(0,0,0,\dots)}^{(\eta_X, \eta_Y, \eta_Z, \dots)} p_{X_T^{\max}, Y_T^{\max}, Z_T^{\max}, \dots}(X_T^{\max}, Y_T^{\max}, Z_T^{\max}, \dots) dX_T^{\max} dY_T^{\max} dZ_T^{\max} \dots \tag{2}$$

P 是响应分量  $\eta_X, \eta_Y, \eta_Z, \dots$  的临界值不超过的概率;  $\cup$  表示逻辑统一运算 «OR»; 和  $p_{X_T^{\max}, Y_T^{\max}, Z_T^{\max}, \dots}$  是整个时间跨度(0, T)上全局最大值的联合概率密度。

然而, 由于后者的高维度和可用数据集的限制, 直接估计后者的联合概率分布是不可行的。更具体地说, 当  $X(t)$  超过  $\eta_X$ , 或  $Y(t)$  超过  $\eta_Y$  或  $Z(t)$  超过  $\eta_Z$  等时刻,

系统被视为立即失效。固定失效水平  $\eta_X, \eta_Y, \eta_Z, \dots$  当然, 当然对于  $R(t)$  的每个一维响应分量都是独立的, 如  $X_{N_X}^{\max} = \max\{X_j; j = 1, \dots, N_X\} = X_T^{\max}$ ,  $Y_{N_Y}^{\max} = \max\{Y_j; j = 1, \dots, N_Y\} = Y_T^{\max}$ ,  $Z_{N_Z}^{\max} = \max\{Z_j; j = 1, \dots, N_Z\} = Z_T^{\max}$

现在, 局部最大值时刻  $[t_1^X < \dots < t_{N_X}^X; t_1^Y < \dots < t_{N_Y}^Y; t_1^Z < \dots < t_{N_Z}^Z]$  以单调非递减的顺序排列成一个单

一的合并时间向量  $t_1 \leq \dots \leq t_N$ 。注意  $t_N = \max\{t_{N_X}^X, t_{N_Y}^Y, t_{N_Z}^Z, \dots\}$ ,  $N = N_X + N_Y + N_Z + \dots$ 。在这种情况下,  $t_j$  表示 MDOF 结构响应分量的局部最大值  $X(t)$  或  $Y(t)$ , 或  $Z(t)$  等。这意味着具有  $R(t)$  的时间记录, 只需连续并同时筛选一维响应分量局部最大值, 并记录其超过 MDOF 极限向量  $(\eta_X, \eta_Y, \eta_Z, \dots)$  的任何分量  $X, Y, Z, \dots$ 。在合并的时间向量  $t_1 \leq \dots \leq t_N$  之后, 将局部一维响应分量最大值合并为一个时间非递减向量  $\vec{R} = (R_1, R_2, \dots, R_N)$ 。即, 每个局部最大值  $R_j$ , 是实际遇到的局部最大值对应于  $X(t)$  或  $Y(t)$ , 或  $Z(t)$  等。最后, 引入统一的极限向量

$(\eta_1, \dots, \eta_N)$ , 每个分量  $\eta_j$  都是  $\eta_X, \eta_Y$  或  $\eta_Z$  等, 取决于  $X(t)$  或  $Y(t)$ , 或  $Z(t)$  等中的哪一个对应于当前运行指数  $j$  的局部最大值。

现在, 引入缩放参数  $0 < \lambda \leq 1$ , 人为地同时降低所有响应分量的极限值, 即新的 MDOF 极限向量  $(\eta_1^\lambda, \eta_2^\lambda, \eta_3^\lambda, \dots)$ , 其中  $\eta_X^\lambda \equiv \lambda \cdot \eta_X, \eta_Y^\lambda \equiv \lambda \cdot \eta_Y, \eta_Z^\lambda \equiv \lambda \cdot \eta_Z, \dots$  [31]。引入统一的极限向量  $(\eta_1^\lambda, \dots, \eta_N^\lambda)$ , 每个分量为  $\eta_j^\lambda, \eta_X^\lambda, \eta_Y^\lambda$  或  $\eta_Z^\lambda$ , 以此类推。后者自动将概率  $P(\lambda)$  定义为  $\lambda$  的函数, 从公式(1)中可以看出  $P \equiv P(1)$ 。不超过概率  $P(\lambda)$  可以估计如下

$$\begin{aligned} P(\lambda) &= \text{Prob}\{R_N \leq \eta_N^\lambda, \dots, R_1 \leq \eta_1^\lambda\} \\ &= \text{Prob}\{R_N \leq \eta_N^\lambda \mid R_{N-1} \leq \eta_{N-1}^\lambda, \dots, R_1 \leq \eta_1^\lambda\} \cdot \text{Prob}\{R_{N-1} \leq \eta_{N-1}^\lambda, \dots, R_1 \leq \eta_1^\lambda\} \\ &= \prod_{j=2}^N \text{Prob}\{R_j \leq \eta_j^\lambda \mid R_{j-1} \leq \eta_{j-1}^\lambda, \dots, R_1 \leq \eta_1^\lambda\} \cdot \text{Prob}\{R_1 \leq \eta_1^\lambda\} \end{aligned} \quad (3)$$

在下文中, 概述了基于条件的串联近似背后的原理, 第一个近似是一步记忆近似, 因此与马尔可夫链近似有一些相似之处。然而, 需要强调的是, 这种第一近似并不等同于这样的近似[20]。在实践中, 相邻的  $R_j$  之间的依赖性是不可忽视的; 因此, 引入了下一步 (将被称为条件水平  $k = 1$ ) 记忆近似法

$$\text{Prob}\{R_j \leq \eta_j^\lambda \mid R_{j-1} \leq \eta_{j-1}^\lambda, \dots, R_1 \leq \eta_1^\lambda\} \approx \text{Prob}\{R_j \leq \eta_j^\lambda \mid R_{j-1} \leq \eta_{j-1}^\lambda\} \quad (4)$$

对于  $2 \leq j \leq N$  (条件等级  $k = 2$ )。公式(4)引入的近似值可以进一步表示为

$$\text{Prob}\{R_j \leq \eta_j^\lambda \mid R_{j-1} \leq \eta_{j-1}^\lambda, \dots, R_1 \leq \eta_1^\lambda\} \approx \text{Prob}\{R_j \leq \eta_j^\lambda \mid R_{j-1} \leq \eta_{j-1}^\lambda, R_{j-2} \leq \eta_{j-2}^\lambda\} \quad (5)$$

其中  $3 \leq j \leq N$  (将称为条件水平  $k = 3$ ), 以此类推不同条件水平  $k$  的取值[21]。目的是及时监测本地首先发生的每个独立故障, 从而避免串联的本地相互关联的超出[16, 31]。

公式(5)提出了统计独立性假设的后续改进。后一种近似以更高的精度捕获相邻最大值之间的统计依赖性效应。由于初始 MDOF 过程  $R(t)$  被假定为遍历的, 因此是平稳的, 概率  $p_k(\lambda) := \text{Prob}\{R_j > \eta_j^\lambda \mid R_{j-1} \leq \eta_{j-1}^\lambda, R_{j-k+1} \leq \eta_{j-k+1}^\lambda\}$ , 对于  $j \geq k$  将独立于  $j$  但仅取决于条件  $k$ 。因此, 不超过概率可以近似为平均条件超过率方法[32]。

$$P_k(\lambda) \approx \exp(-N \cdot p_k(\lambda)), k \geq 1. \quad (6)$$

注意, 公式(6)来自于公式(1), 通过忽略  $\text{Prob}(R_1 \leq \eta_1^\lambda) \approx 1$  得到的, 由于设计失效概率必须很小, 也假设  $N \cdot k$ 。公式(6)类似于众所周知的平均上交率方程超过概率[31, 32]。条件参数  $k$  有明显的收敛性

$$P = \lim_{k \rightarrow \infty} P_k(1); p(\lambda) = \lim_{k \rightarrow \infty} p_k(\lambda) \quad (7)$$

注意, 对于  $k = 1$  的公式(6)变成一个众所周知的非超限概率关系, 其平均上行速率函数为

$$P(\lambda) \approx \exp(-v^+(\lambda)T); v^+(\lambda) = \int_0^\infty \zeta p_{RR}(\lambda, \zeta) d\zeta \quad (8)$$

其中  $v^+(\lambda)$  表示上述由缩放的 MDOF 系统响应  $(\frac{X}{\eta_X}, \frac{Y}{\eta_Y}, \frac{Z}{\eta_Z}, \dots)$  组合而成的非维向量  $R(t)$  的响应水平  $\lambda$  的平均上交率。公式(8)中平均上交率由 Rice 公式给出[33],  $p_{RR}$  是  $(R, \dot{R})$  的联合概率密度,  $\dot{R}$  是时间导数  $R'(t)$ 。公式(8)依赖于泊松假设, 即高  $\lambda$  水平 (在本文中是  $\lambda \geq 1$ )

的上交事件可以被假定为是独立的。后者可能不适用于窄带响应和更高级别的动力系统, 这些系统在不同的维度上表现出串联故障[34-37]。后者是由极端事件之间的内在相互依赖性引起的, 表现为在集合向量  $\vec{R} = (R_1, R_2, \dots, R_N)$  内出现高度相关的局部最大值集群

[38-41]。

上文中，已经使用了平稳性假设。然而，所提出的方法也可以处理非平稳情况。对于非平稳情况，在  $m = 1, \dots, M$  个季节性流行情况的散点图，每个短期季节性状态的概率为  $q_m$ ，所以  $\sum_{m=1}^M q_m = 1$ 。接下来，让我们介绍一下长期方程

$$p_k(\lambda) \equiv \sum_{m=1}^M p_k(\lambda, m) q_m \quad (9)$$

$p_k(\lambda, m)$  与公式(7)中的函数相同，但对应于一个特定的短期季节性流行状态，数字为  $m$ 。

请注意，在前期的工作中作者已经成功验证了针对各种一维动态系统的建议方法的准确性[16, 31]。

下面简单介绍下外推法[31, 32]，作为失效概率分布尾部外推的基础，渐近为 Gumbel 分布类型。后一种方法假设在一般情况下外推所需的参数函数类可以建模为类似于 Gumbel 分布和一般极值 (GEV) 分布之间

$$F(a_k, b_k, c_k, p_k, q_k) = \int_{\lambda_0}^{\lambda_1} \omega(\lambda) \{ \ln(p_k(\lambda)) - d_k + (a_k \lambda + b_k)^{c_k} \}^2 d\lambda, \lambda \geq \lambda_0 \quad (11)$$

其中  $\lambda_1$  为合适的分布尾部截断值，即最显著的波高值，置信区间宽度仍可接受。参数  $a_k, b_k, c_k, p_k, q_k$  的最佳值也可以使用 NAG 数值库[32]中包含的顺序二次规划 (SQP) 方法来确定[49-51]。

权重函数  $\omega$  可以定义为  $\omega(\lambda) = \{ \ln CI^+(\lambda) - \ln CI^-(\lambda) \}^{-2}$ ，其中  $(CI^-(\lambda), CI^+(\lambda))$  是一个置信区间 (CI)，由模拟或测量数据集的经验估计[16, 31]。当参数  $c = \lim_{k \rightarrow \infty} c_k$  等于 1 或接近于 1 时，分布就接近于 Gumbel 分布。

对于任何一般的遍历波高或风速过程，可以假设超过阈值  $\lambda$  的条件超标序列构成泊松过程；但是，通常是非同质的。因此，对于接近 1 的  $\lambda$  水平， $p_k(\lambda)$  的  $p$ -% 置信区间 (CI) 的近似极限可以如下给出

$$CI^\pm(\lambda) = p_k(\lambda) \left( 1 \pm \frac{f(p)}{\sqrt{(N-k+1)p_k(\lambda)}} \right). \quad (12)$$

其中  $f(p)$  是根据逆正态分布估计的，例如  $f(90\%) = 1.65$ ， $f(95\%) = 1.96$ 。其中  $N$  是在分析向量  $\vec{R}$  中组装的局部最大值的总数。

### 3 全球的艾滋病死亡率

流感样流行病的预测长期以来一直是流行病学和数学生物学关注的焦点。众所周知，公共卫生动态是

的关系。

从采样时间序列中观察到的可用极值并不一定构成渐近分布，或者至少证明它们是真正渐近的是一项不平凡的任务。这意味着将我们感兴趣的领域扩展到亚渐近水平的相关性。上面介绍的函数通常在尾部是规则的，特别是对于  $p_k(\lambda)$   $\lambda$  接近和超越 1。更准确地说，对于  $\lambda \geq \lambda_0$ ，分布尾部的行为类似于  $\exp\{-(a\lambda + b)^c + d\}$  其中  $a, b, c, d$  是适当拟合的常数，用于适当的尾部截断  $\lambda_0$  值。因此，可以得到

$$p_k(\lambda) \approx \exp\{-(a_k \lambda + b_k)^{c_k} + d_k\}, \lambda \geq \lambda_0 \quad (10)$$

接下来，通过绘制  $\ln\{\ln(p_k(\lambda)) - d_k\}$  与  $\ln(a_k \lambda + b_k)$ ，通常可以观察到几乎完美的线性尾部行为[42-48]。

通过最小化以下误差函数  $F$ ，在对数水平上进行优化是很有帮助的，相对于四个参数  $a_k, b_k, c_k, p_k, q_k$  而言

一个高度非线性的多维和空间互相关动态系统，分析起来总是具有挑战性。以前的研究使用了多种方法来模拟流感样病例。本节说明了上述方法的效率，该方法使用应用于现实生活中的艾滋病数据集的新方法，呈现为新的每日记录的感染患者时间序列，分布在大范围内。世界有 195 个国家，本节介绍  $X, Y, Z, \dots$  从而构成了一个 195 维 (195D) 动态生物系统的例子。为了统一所有 195 个测量的时间序列  $X, Y, Z, \dots$ ，进行了以下的缩放：

$$X \rightarrow \frac{X}{\eta_X}, Y \rightarrow \frac{Y}{\eta_Y}, Z \rightarrow \frac{Z}{\eta_Z}, \dots \quad (13)$$

使所有 195 个响应都是无量纲的，并且具有相同的失效限制等于 1。失效限制  $\eta_X, \eta_Y, \eta_Z, \dots$ ，或者换句话说，流行病阈值，不是明显的选择。最简单的选择是，不同的国家设定的失效限制等于相应国家人口占当地人口的百分比，基本上使  $X, Y, Z, \dots$  等于每个国家年度的感染的百分比。

接下来，将 195 个测量时间序列中的所有局部最大值合并为一个时间序列，方法是保持时间非递减顺序： $\vec{R} = (\max\{X_1, Y_1, Z_1, \dots\}, \dots, \max\{X_N, Y_N, Z_N, \dots\})$ ，其中每个集合  $\max\{X_j, Y_j, Z_j, \dots\}$  是根据这些局部最大值出现的非递减次数进行排序。

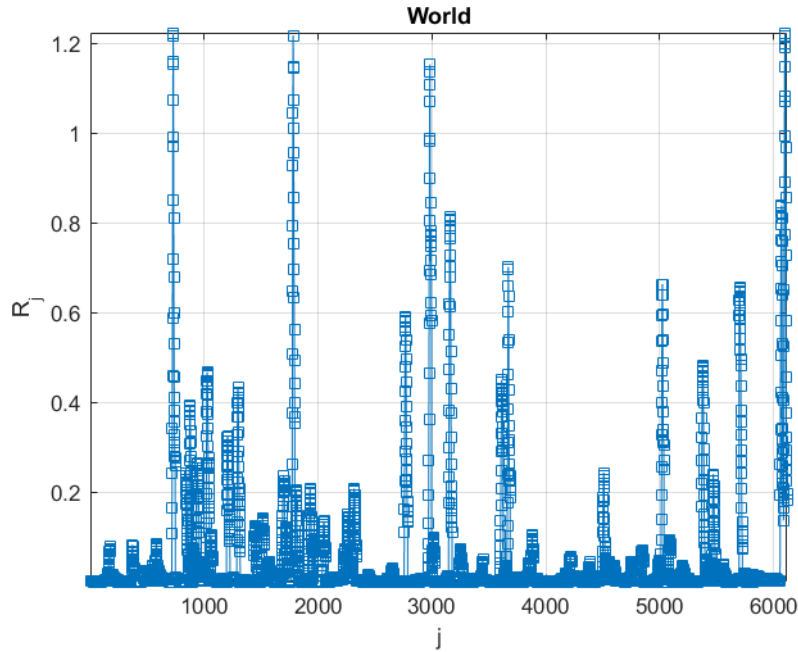


图 2 艾滋病 3 年记录的死亡率，以 195D 向量表示的百分比人口的百分比

Figure 2 Annual recorded death rate in percent as 195D vector  $\vec{R}$  in percent of corresponding regional population.

图 2 以 195D 向量 $\vec{R}$ ，表示年度的死亡率数量，由相应区域人口中每百万人口中的每日新增患者人数组成，按公式(13)进行缩放。请注意，向量 $\vec{R}$ 是由具有不同流行病背景的不同区域成分组成的。索引 $j$ 只是在非递减的时间序列中遇到的局部最大值的运行索引。

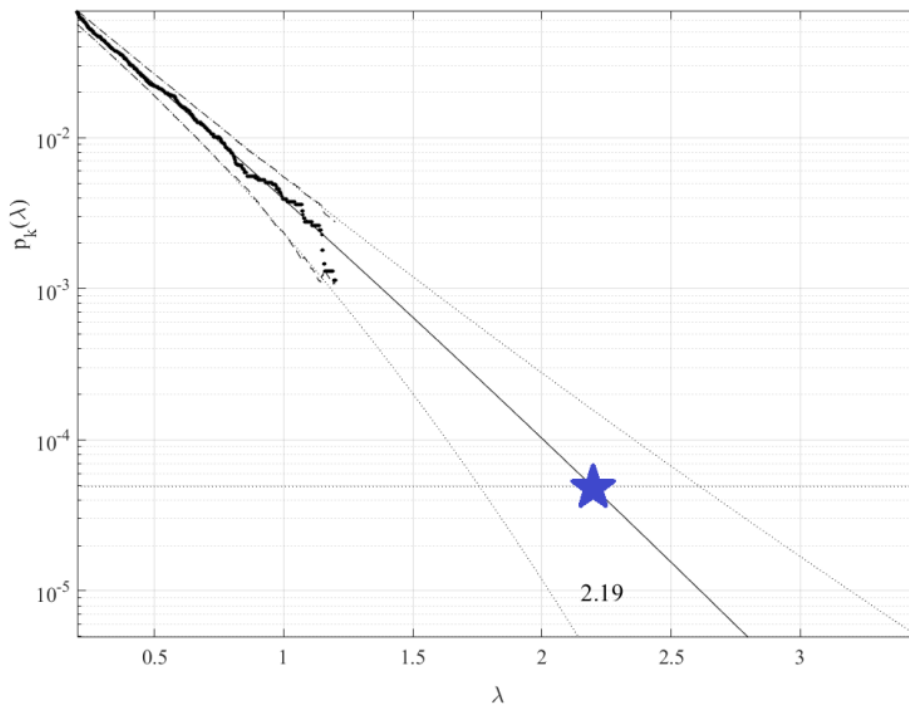


图 3 向临界水平（用星号表示）的 10 年回报水平外推占当地人口的百分比，虚线表示外推的 95%CI，横轴为当地人口的百分比  
Figure 3 100 years return level extrapolation of  $p_k(\lambda)$  towards critical level (indicated by star) in percent to local population. Extrapolated 95% CI indicated by dotted lines. Percentage of local population on horizontal axis.

图 3 显示了根据公式(10)对流行病爆发的 100 年回归期的推断, 由水平虚线表示, 并在一定程度上超越了, 使用  $\lambda = 0.2\%$  临界值, 横轴为当地人口百分比。虚线表示根据公式(11)推断的 95% 置信区间。根据公式(6)中的  $p(\lambda)$  与公式(1)中的目标失效概率  $1 - P$  直接相关。因此, 根据公式(6)可以估算出系统失效概率  $1 - P \approx 1 - P_k(1)$ 。注意, 在公式(6)中,  $N$  对应于统一响应向量  $\vec{R}$  中局部极大值的总数。由于出现了与  $k$  有关

的收敛, 故条件参数  $k = 5$  已经足够, 见公式(7)。图 3 显示了合理狭窄的 95%CI, 后者是所提方法的一个优点。

请注意, 上述方法新颖, 但由于其能够对卫生系统多维度进行处理, 并根据非常有限的数据集进行准确的外推, 故可以有效地利用可用的测量数据集。需要注意的是, 图 3 中的预测无量纲  $\lambda$  水平(星号表示), 代表未来几年世界任何地区发生疫情的概率。

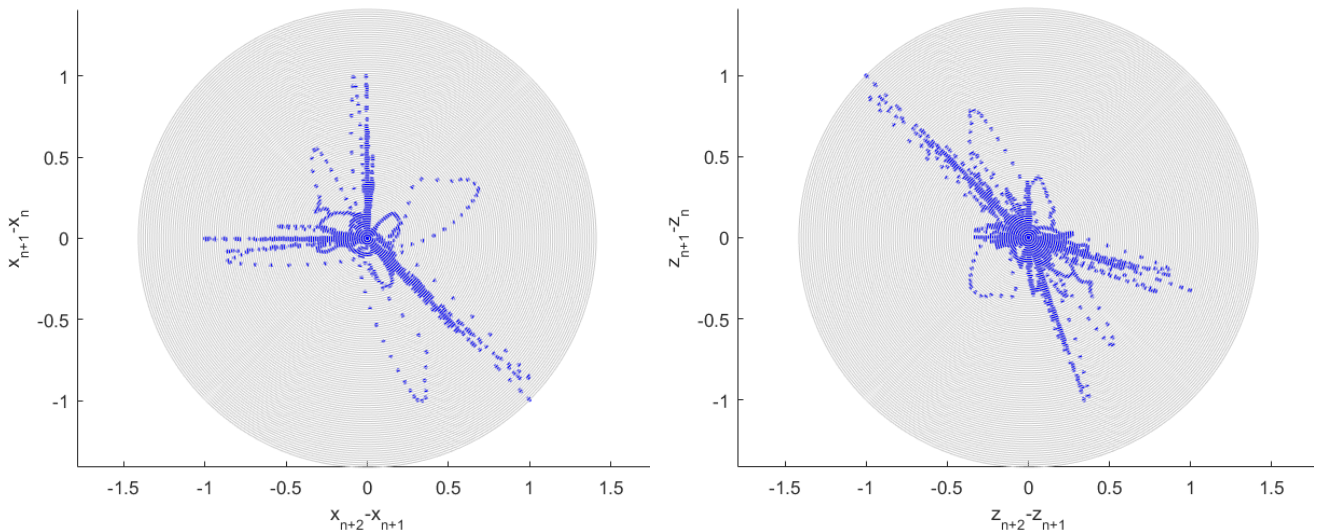


图 4 全球艾滋病死亡统计数据。左图: SODP 图; 右图: FODP 图

Figure 4 HIV death global statistics. Left: SODP plot; Right: FODP plot.

二阶差分图 (SODP) 源自 Poincare 图。SODP 可以观察时间序列数据中连续差值的统计情况。图 4 展示了二阶差分图 (SODP) 和四阶差分图 (FODP)。这类曲线图可用于数据模式识别以及与其他数据集的比较, 例如用于人工智能 (AI) 识别方法[52]。请注意, EVT 是渐近和 IDOF 方法, 而本研究引入了 MDOF 和次渐近方法。

### 4 结论

传统的时间序列卫生系统可靠性分析方法并不具备有效处理拥有高维度和不同系统响应之间交叉相关的系统的优势。该方法的主要优点是能够研究高维非线性动力系统的可靠性。

尽管算例简单, 但本研究成功地提供了一种新颖的多维建模策略和理论方法来实施流行病的预测。已经讨论了如何正确设置每地区的流行病学警报限值 (故障限值)。

本文以世界各省记录的艾滋病病例数据集, 构建了一个全维度 (195D) 模型。将该方法作为一个多维实时系统应用于每日新增患者人数。文中给出了该方法背后的理论推理。注意, 直接使用测量或蒙特卡罗模拟进行动态生物系统可靠性分析是有吸引力的, 然而, 动态系统的复杂性及其高维性要求开发具有强鲁棒性和准确的技术, 以处理有限的数据集, 尽可能有效地利用可用数据。

主要结论为: 如果世界当地环境和流行病学条件下的公共卫生系统管理得当。预测 100 年重现期风险水平为 2.2%, 因为每日新增患者占当地流行病爆发人群的百分比较低。

本研究旨在进一步开发一种通用、稳定且直观的多维可靠性方法。本文介绍的方法之前已通过应用到广泛的仿真模型进行验证, 适用于一维系统响应, 通常可以获得非常准确的预测, 测量和数值模拟的时间序列响应都可以进行分析。结果表明, 所提出的方法

产生了合理的置信区间。因此, 所提出的方法可能成为各种非线性动态生物系统可靠性研究的合适工具。最后, 此方法可应用许多公共卫生系统, 所呈现的艾滋病示例不会以任何方式限制新方法的适用范围。

## 参考文献

- [1] AARSSSEN K, DE HAAN L. On the maximal life span of humans [J]. *Mathematical Population Studies*, 1994, 4 (4): 259-81.
- [2] BEISEL C J, ROKYTA D R, WICHMAN H A, et al. Testing the extreme value domain of attraction for distributions of beneficial fitness effects [J]. *Genetics*, 2007, 176(4): 2441-9.
- [3] BLOCK H W, BASU A. A continuous, bivariate exponential extension [J]. *Journal of the American Statistical Association*, 1974, 69 (348): 1031-7.
- [4] GALAMBOS J, MACRI N. The life length of humans does not have a limit [J]. *Journal of Applied Statistical Science*, 2000, 9(4): 253-64.
- [5] GUMBEL E J. *La durée extrême de la vie humaine* [M]. Hermann, 1937.
- [6] GUMBEL E J. *Statistics of extremes* [M]. Columbia university press, 1958.
- [7] GUPTA R D, KUNDU D. Theory & methods: Generalized exponential distributions [J]. *Australian & New Zealand Journal of Statistics*, 1999, 41(2): 173-88.
- [8] JOYCE P, ABDO Z. Determining the distribution of fitness effects using a generalized Beta-Burr distribution [J]. *Theoretical Population Biology*, 2018, 122: 88-96.
- [9] ROMEO J S, MEYER R, GALLARDO D I. Bayesian bivariate survival analysis using the power variance function copula [J]. *Lifetime data analysis*, 2018, 24: 355-83.
- [10] ROSER M, RITCHIE H. Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/hiv-aids' [Online Resource] [Z]. 2018.
- [11] SARKAR S K. A continuous bivariate exponential distribution [J]. *Journal of the American Statistical Association*, 1987, 82(398): 667-75.
- [12] CHOI S-K, CANFIELD R A, GRANDHI R V. Probabilistic Analysis [J]. *Reliability-based Structural Design*, 2007: 51-80.
- [13] DITLEVSEN O, MADSEN H O. *Structural reliability methods* [M]. Wiley New York, 1996.
- [14] MADSEN H O, KRENK S, LIND N C. *Methods of structural safety* [M]. Courier Corporation, 2006.
- [15] MELCHERS R E, BECK A T. *Structural reliability analysis and prediction* [M]. John wiley & sons, 2018.
- [16] Næss A, GAIDAI O. Estimation of extreme values from sampled time series [J]. *Structural safety*, 2009, 31(4): 325-34.
- [17] ANDINET W. *Pattern and Determinants of Survival in Adult HIV patients on ART, Ethiopia* [J]. Umea International School of Public Health, Umea University, 2009.
- [18] BIRTUKAN T. *Predictors of Early Death in HIV Positive Individuals Enrolled for Chronic HIV Care in Jimma Zone, South Western Ethiopia* [D]; Masters Thesis, Addis Abeba University. Addis Abeba, 2010.
- [19] ETARD J-F, NDIAYE I, THIERRY-MIEG M, et al. Mortality and causes of death in adults receiving highly active antiretroviral therapy in Senegal: a 7-year cohort study [J]. *Aids*, 2006, 20(8): 1181-9.
- [20] JOHANNESSEN A, NAMAN E, NGOWI B J, et al. Predictors of mortality in HIV-infected patients starting antiretroviral therapy in a rural hospital in Tanzania [J]. *BMC infectious diseases*, 2008, 8: 1-10.
- [21] ASSEFA T. *Survival Analysis on HIV/AIDS Patients at Tekur Ambesa Hospital* [D]; M. Sc. Thesis, Addis Ababa, Ethiopia, 2007.
- [22] DI BIASE G, D'AMICO G, DI GIROLAMO A, et al. A stochastic model for the HIV/AIDS dynamic evolution [J]. *Mathematical Problems in Engineering*, 2007, 2007.
- [23] GETAHUN Z. *Statistical Modelling of HIV/AIDS Progression and Survival of AIDS Patients. A Case Study of Bahir-Dar, Feleg-Hiwot Referral Hospital* [D]; M. Sc. Thesis. Hawassa University, Hawassa, Ethiopia, 2010.
- [24] IBRAHIM N. *Education of factor affecting the chance of survival/death status among hiv positive people under the antiretroviral treatment program: The Case of Adama Referral Hospital* [D]; M. Sc. Thesis, Addis Ababa University, Ethiopia, 2007.
- [25] LAWN S D, MYER L, HARLING G, et al. Determinants of mortality and nondeath losses from an antiretroviral treatment service in South Africa: implications for program evaluation [J]. *Clinical infectious diseases*, 2006, 43(6): 770-6.
- [26] MZILENI M O, LONGO-MBENZA B, CHEPHE T J. Mortality and causes of death in HIV-positive patients receiving antiretroviral therapy at Tshepang Clinic in Doctor George Mukhari Hospital [J]. *Polskie Archiwum Medycyny Wewnetrznej*, 2008, 118(10): 548-54.



- [27] SIELEUNOU I, SOULEYMANOU M, SCHÄNENBERGER A M, et al. Determinants of survival in AIDS patients on antiretroviral therapy in a rural centre in the Far - North Province, Cameroon [J]. *Tropical Medicine & International Health*, 2009, 14(1): 36-43.
- [28] STRINGER J S, ZULU I, LEVY J, et al. Rapid scale-up of antiretroviral therapy at primary care sites in Zambia: feasibility and early outcomes [J]. *Jama*, 2006, 296(7): 782-93.
- [29] KRISTENSEN S B, BIBBY B M. A bivariate logistic regression model based on latent variables [J]. *Statistics in Medicine*, 2020, 39(22): 2962-79.
- [30] CHEN J, LEI X, ZHANG L, et al. Using extreme value theory approaches to forecast the probability of outbreak of highly pathogenic influenza in Zhejiang, China [J]. *PloS one*, 2015, 10(2): e0118521.
- [31] NAESS A, MOAN T. *Stochastic dynamics of marine structures* [M]. Cambridge University Press, 2013.
- [32] GAUL J. *Experts in numerical software and HPC services* [J]. 2010.
- [33] RICE S O. *Mathematical analysis of random noise* [J]. *The Bell System Technical Journal*, 1944, 23(3): 282-332.
- [34] CHENG Y, GAIDAI O, YURCHENKO D, et al. Study on the dynamics of a payload influence in the polar ship; proceedings of the ISOPE International Ocean and Polar Engineering Conference, F, 2022 [C]. ISOPE.
- [35] GAIDAI O, WANG F, WU Y, et al. Offshore renewable energy site correlated wind-wave statistics [J]. *Probabilistic Engineering Mechanics*, 2022, 68: 103207.
- [36] GAIDAI O, XING Y. A novel bio-system reliability approach for multi-state COVID-19 epidemic forecast [J]. *Engineered Science*, 2022, 21: 797.
- [37] GAIDAI O, XING Y, BALAKRISHNA R. Improving extreme response prediction of a subsea shuttle tanker hovering in ocean current using an alternative highly correlated response signal [J]. *Results in Engineering*, 2022, 15: 100593.
- [38] GAIDAI O, XING Y, XU X. COVID-19 epidemic forecast in USA East coast by novel reliability approach [J]. 2022.
- [39] SUN J, GAIDAI O, WANG F, et al. Extreme riser experimental loads caused by sea currents in the Gulf of Eilat [J]. *Probabilistic Engineering Mechanics*, 2022, 68: 103243.
- [40] XING Y, GAIDAI O, MA Y, et al. A novel design approach for estimation of extreme responses of a subsea shuttle tanker hovering in ocean current considering aft thruster failure [J]. *Applied Ocean Research*, 2022, 123: 103179.
- [41] XU X, WANG F, GAIDAI O, et al. Bivariate statistics of floating offshore wind turbine dynamic response under operational conditions [J]. *Ocean Engineering*, 2022, 257: 111657.
- [42] BALAKRISHNA R, GAIDAI O, WANG F, et al. A novel design approach for estimation of extreme load responses of a 10-MW floating semi-submersible type wind turbine [J]. *Ocean Engineering*, 2022, 261: 112007.
- [43] GAIDAI O, FU S, XING Y. Novel reliability method for multidimensional nonlinear dynamic systems [J]. *Marine Structures*, 2022, 86: 103278.
- [44] GAIDAI O, XU X, NAESS A, et al. Bivariate statistics of wind farm support vessel motions while docking [J]. *Ships and offshore structures*, 2021, 16(2): 135-43.
- [45] GAIDAI O, YAN P, XING Y. A novel method for prediction of extreme wind speeds across parts of Southern Norway [J]. *Frontiers in Environmental Science*, 2022, 10: 997216.
- [46] GAIDAI O, YAN P, XING Y. Prediction of extreme cargo ship panel stresses by using deconvolution [J]. *Frontiers in Mechanical Engineering*, 2022, 8: 992177.
- [47] GAIDAI O, YAN P, XING Y, et al. A novel statistical method for long-term coronavirus modelling [J]. *F1000 Research*, 2023, 11: 1282.
- [48] GAIDAI O, YAN P, XING Y, et al. Oil tanker under ice loadings [J]. *Scientific Reports*, 2023, 13(1): 8670.
- [49] GAIDAI O, WU Y, YEGOROV I, et al. Improving performance of a nonlinear absorber applied to a variable length pendulum using surrogate optimization [J]. *Journal of Vibration and Control*, 2022: 10775463221142663.
- [50] GAIDAI O, XING Y. Novel reliability method validation for offshore structural dynamic response [J]. *Ocean Engineering*, 2022, 266: 113016.
- [51] GAIDAI O, XU J, YAN P, et al. Novel methods for wind speeds prediction across multiple locations [J]. *Scientific reports*, 2022, 12(1): 19614.
- [52] YAYIK A, KUTLU Y, ALTAN G. Regularized HessELM and inclined entropy measurement for congestive heart failure prediction [J]. *arXiv preprint arXiv: 190705888*, 2019.