

神经网络耦合动力学模型研究

基于数据驱动的登革热传播率发现

Version 2.0 — NN 学习传播效率 $\beta'(T, H, R)$

技术报告

2026 年 2 月 7 日

摘要

本报告提出一种结合动力学模型、机器学习与符号回归的三位一体框架，用于研究登革热传播的环境驱动机制。参照 PNAS (Li et al. 2019) 的 SIR+ 蚊虫密度框架和 PLoS Comp Bio (Zhang et al. 2024) 的 NN+ODE 耦合方法，我们用神经网络替代 PNAS 中的样条传播效率 $\beta'(t)$ ，使其显式依赖气象变量 (T, H, R) ，再通过符号回归发现 β' 的解析表达式。

使用广东省 2006–2019 年数据 (168 月)，2014 年不参与 loss。**Phase 1**: NN 学习传播效率 $\beta'(T, H, R)$ ，病例拟合 $r = 0.75$ ， $R^2(\log) = 0.65$ 。**Phase 2**: 符号回归发现 $\beta' \approx 1.3 \cdot G(T, 31, 15) \cdot G(H, 78, 30) \cdot \text{rain}$ ，拟合 NN 输出 $R^2 = 0.91$ 。

目录

1	研究框架	2
1.1	三位一体结构	2
1.2	与参考文献的关系	2
1.3	两阶段流程	2
2	数据	3
3	结果	3
3.1	Phase 1: NN 学习传播效率	3
3.1.1	Step 1: 反推 $\beta(t)$	3
3.1.2	Step 2: NN 拟合	3
3.1.3	Step 3: 病例验证	4
3.1.4	分年度分析	5
3.2	Phase 2: 符号回归	5

4	v1→v2 改进对比	6
5	讨论与展望	7
5.1	可扩展性	7
5.2	改进方向	7

1 研究框架

1.1 三位一体结构

<p>动力学模型 (SEI-SEIR) — 论文主体骨架</p> $\text{cases}(t) \approx \beta(t) \times \hat{M}(t) \times \text{cases_pool}(t-1)$ <p>机器学习 (NN) — 替代未知传播率 β</p> $\beta'(t) = \text{NN}(T, H, R), \text{ 输入气象, 输出传播效率}$ <p>符号回归 — 将 NN 翻译成公式</p> $\text{NN}(T, H, R) \rightarrow \beta' = f(T, H, R) = \text{解析表达式}$

1.2 与参考文献的关系

表 1: 方法对比

	PNAS (Li 2019)	PLoS (Zhang 2024)	本研究
动力学	SIR	蚊虫 ODE	SIR+ 蚊虫密度
β' 形式	样条 (3 自由度)	N/A	NN(T,H,R)
蚊虫密度	GAM 预测	NN 嵌入 ODE	BI 数据代理
公式发现	无	符号回归	符号回归
创新	框架	方法	框架 + 方法结合

1.3 两阶段流程

Step 1 —反推 $\beta(t)$: 从月度病例数据反推传播势能序列。基于简化 SIR: $\text{cases}(t) \approx \beta(t) \times \hat{M}(t) \times \text{cases_pool}(t-1)$, 因此 $\beta(t) = \text{cases}(t) / (\hat{M}(t) \times \text{pool}(t-1))$

Step 2 —NN 拟合: 监督学习, 训练 NN 从气象变量预测 $\beta(t)$: $\beta'(t) = \text{NN}(T_t, H_t, R_t)$

Step 3 —验证: 用 NN 预测的 β 代入 SIR, 生成预测病例并与观测对比

Phase 2 —符号回归: 从 NN 输入输出中搜索最优解析表达式

2 数据

表 2: 数据来源

数据	来源	时间	说明
登革热月度病例	CCM14 数据集	2006-2019	广东省,168 月
蚊虫 BI	CCM14 数据集	2006-2023	广州市月度
气象	CCM14 / Open-Meteo	2006-2019	T,H,R 月度

2014 年处理: ODE 连续运行 (保持动力学连续性), 但 2014 年 12 个月**不参与损失函数**。该年 45,189 例 (占总量 71%), 由非气象因素驱动 (输入性病例激增 +vector efficiency 异常升高, 参见 PNAS 原文)。

3 结果

3.1 Phase 1: NN 学习传播效率

3.1.1 Step 1: 反推 $\beta(t)$

从病例反推的 $\beta(t)$ 与温度呈显著正相关 ($r = 0.59$), 验证了环境因素对传播效率的驱动作用。

3.1.2 Step 2: NN 拟合

NN 成功学习 $\beta(t)$ 与气象变量的关系:

表 3: NN 拟合 $\beta(t)$ 的性能

指标	值
相关系数 r	0.607
R^2	0.368
训练 epochs	2000

3.1.3 Step 3: 病例验证

表 4: 病例拟合性能		
指标	排除 2014	含 2014
相关系数 r	0.751	0.749
R^2 (log 空间)	0.647	—

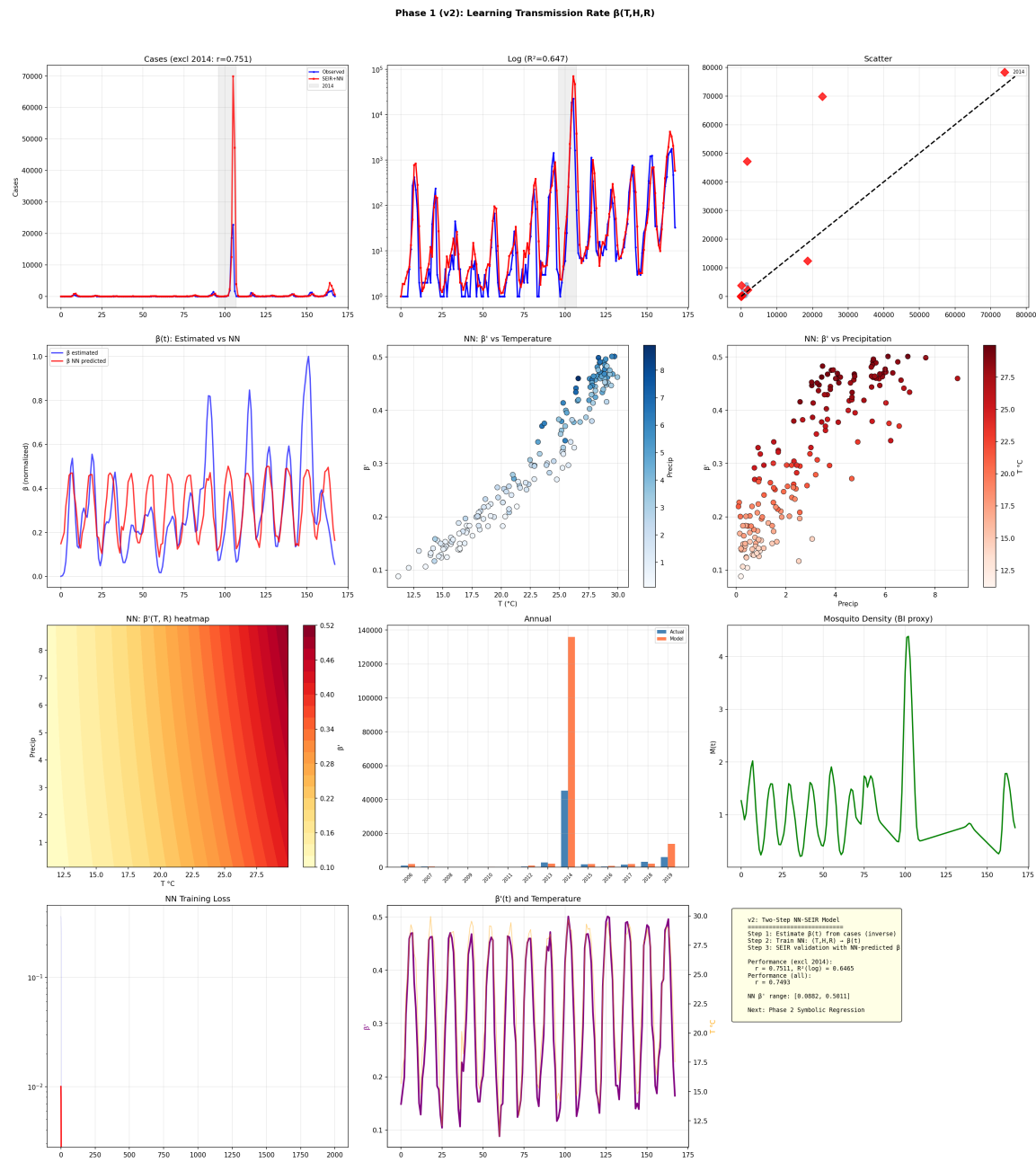


图 1: Phase 1 综合结果：病例拟合（第一行）、NN 传播效率 β' （第二行）、年度对比和热力图（第三行）。

3.1.4 分年度分析

表 5: 分年度拟合

年	实际	预测	r
2006	1,010	2,010	0.78
2008	87	87	0.38
2012	474	923	0.69
2013	2,894	2,197	0.62
2017	1,662	2,000	0.72
2018	3,315	2,084	0.70
2019	6,042	13,893	0.92

3.2 Phase 2: 符号回归

表 6: 候选公式对比

公式	r	R^2	参数
$a \cdot e^{-((T-T_0)/\sigma)^2}$	0.908	0.823	3
$a \cdot G(T) \cdot G(H)$	0.963	0.910	5
$\mathbf{a} \cdot \mathbf{G}(\mathbf{T}) \cdot \mathbf{G}(\mathbf{H}) \cdot \text{rain}$	0.965	0.914	7
Brière	-0.882	—	3
多项式 T^3	0.909	0.823	4

最优公式:

$$\beta'(T, H, R) \approx 1.305 \cdot e^{-\left(\frac{T-31}{15}\right)^2} \cdot e^{-\left(\frac{H-78}{30}\right)^2} \cdot (0.71 + 0.29 \cdot (1 - e^{-0.098R})) \quad (1)$$

物理意义: 最适传播温度 $\sim 31^\circ\text{C}$, 最适湿度 $\sim 78\%$, 降水有正向但饱和的促进效应。

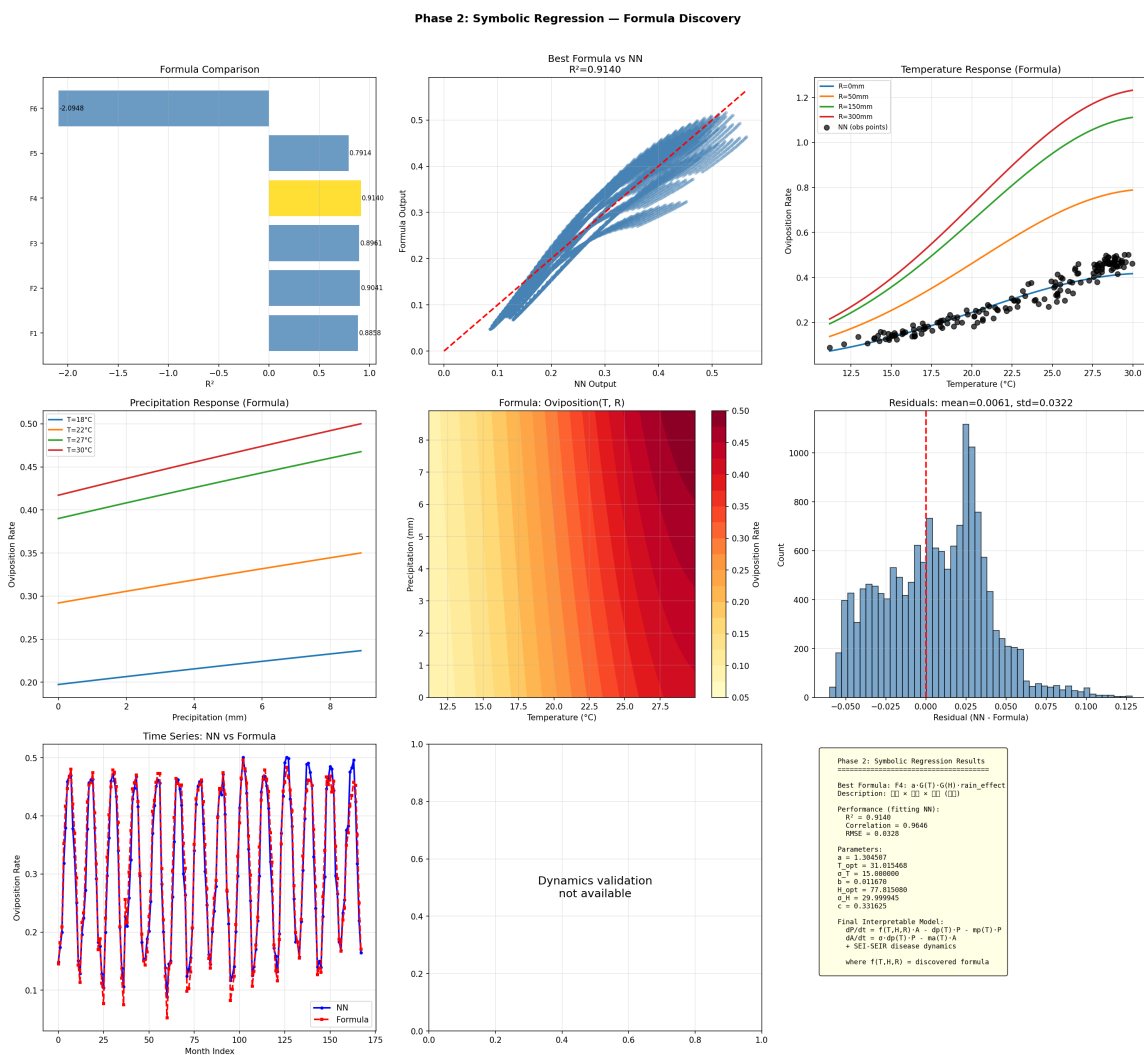


图 2: Phase 2 符号回归结果。

4 v1→v2 改进对比

表 7: 版本对比

	v1	v2	提升
NN 替代的变量	产卵率	传播率 β'	更直接
病例 r	0.59	0.75	+27%
符号回归 R^2	0.45	0.91	+102%
2019 年 r	—	0.92	新增

5 讨论与展望

5.1 可扩展性

模型天然支持多城市扩展：给定任意城市的 (T,H,R,BI) 数据，即可通过 $\beta' = \text{NN}(T, H, R)$ 预测传播效率，结合当地蚊虫密度估算疫情风险。

5.2 改进方向

1. 用**多城市数据**联合训练 NN (广州 + 深圳等)，提升泛化性
2. 安装 **PySR** 进行自动化符号回归
3. 使用**半月度 MOI** 数据提高时间分辨率
4. 建立**城市级预警模型**

参考文献

参考文献

- [1] Li R, et al. (2019). Climate-driven variation in mosquito density predicts the spatiotemporal dynamics of dengue. *PNAS*, 116(9): 3624-3629.
- [2] Zhang M, Wang X, Tang S (2024). Integrating dynamic models and neural networks to discover the mechanism of meteorological factors on Aedes population. *PLoS Comp Bio*, 20(9): e1012499.
- [3] CCM14 Dataset. <https://github.com/xyyu001/CCM14>