毕业论文

基于神经网络耦合动力学模型的登革热传播效率发现与多城市验证

Discovery of Dengue Transmission Efficiency via Neural-Network-Coupled Dynamical Models and Multi-City Validation

学　　院：　公共卫生学院

专　　业：　流行病与卫生统计学

研究方向：　传染病建模与预测

2025 年 6 月

摘　要

登革热是全球最严重的蚊媒传染病之一，中国南方尤其广东省是国内最主要的流行区域。理解气候因素如何驱动登革热传播效率，对于建立早期预警系统和制定精准防控策略具有重要意义。然而，传统机制模型往往依赖先验假设固定传播率函数形式，难以从数据中自动发现最优的气候–传播率关系；纯数据驱动的机器学习方法则存在可解释性不足的瓶颈。

本文提出一种" 神经网络耦合 SEIR 动力学 + 符号回归" 三阶段混合建模框架，旨在兼顾机制可解释性与数据适应性。该框架包含三个核心阶段：(1) 基于 SEIR 仓室模型反演月尺度传播系数 β' 时间序列；(2) 以多层感知机 (MLP) 学习气候变量（温度 T、降水 R、相对湿度 H）到 β' 的非线性映射，实现逐月病例数预测；(3) 利用符号回归对神经网络进行知识蒸馏，发现可解释的闭合公式。

研究分为两个部分。第一部分以广州市为单城市案例（2005–2019 年月度数据）。Phase 1 阶段，通过 SEIR 逆问题求解（Brent 二分法）反演月度传播系数 $\beta'(t)$，再以 7 维纯气候特征（温度、湿度、降水及其滞后与季节编码，不含蚊媒密度）训练 MLP 在对数空间中拟合 β' 与气候变量的映射关系。耦合模型对广州月度病例的预测达到 Pearson 相关系数 $r = 0.613$、Spearman 秩相关 $\rho = 0.706$、MAE$= 36.7$。基于 β' 计算的基本再生数 R_0 均值为 1.06，最大值达 6.52（2014 年 9 月），在 180 个月中有 46% 的月份 $R_0 > 1$。留一年交叉验证（15 折）的平均 $r = 0.350 \pm 0.293$、$\rho = 0.570 \pm 0.228$。Phase 2 阶段，从物理模板族和多项式族两类候选中发现最优公式，其中二次多项式含交互项的公式以 $R^2 = 0.644$ 拟合 NN 输出，揭示了温度–湿度正交互（$a_{TH} > 0$）和降水平方负效应（$a_{RR} < 0$）等物理规律。对 2014 年特大暴发的分析表明，NN 预测的 β' 与其他年份几乎无差异（差异 $< 0.1\%$），而 SEIR 反演的真实 β' 高出 74%，说明气候因素无法解释 2014 年的异常传播效率，极端暴发主要由非气候因素驱动。

第二部分将广州发现的公式迁移至广东省 16 个地级市（2005–2019 年）。结果表明，16 城年度总病例排名的 Spearman 相关 $\rho = 0.962$，非广州 15 城 $\rho = 0.954$、MAE$= 454.3$，证实了公式的空间泛化能力。城市月度曲线的中位 Pearson $r = 0.480$、中位 Spearman $\rho = 0.470$，表明公式可捕捉多数城市的季节性趋势。与历史月均值、线性回归、季节朴素法三种基线模型相比，SEIR+NN 模型在 16 城平均 Spearman ρ（0.480 vs 0.384/0.245/0.466）和 $R^2_{\log}$（0.353 vs -0.179/-0.717/0.022）上均取得最优表现。

本研究的主要创新包括：(1) 提出"NN 逆问题 + 符号蒸馏" 范式，克服了传统 SEIR 模型依赖先验函数形式的局限；(2) 发现的闭合公式具有明确物理含义且可直接迁移至其他城市；(3) 系统验证了单城市机制在多城市空间尺度上的可迁移性；(4) 相比 Li 等 (2019 PNAS) 的样条方法和 Zhang 等 (2024) 的纯符号回归方法，本框架在可解释性和泛化性之间取得了更好的平衡。

关键词：登革热；SEIR 模型；神经网络；符号回归；传播效率；广东省；多城市验证

Abstract

Dengue fever is one of the most severe mosquito-borne infectious diseases globally. Southern China, especially Guangdong Province, is the primary endemic region in the country. Understanding how climatic factors drive dengue transmission efficiency is crucial for establishing early-warning systems and formulating targeted control strategies. However, traditional mechanistic models rely on *a priori* assumptions to fix the functional form of the transmission rate, making it difficult to discover optimal climate–transmission relationships from data automatically. Meanwhile, purely data-driven machine learning approaches suffer from limited interpretability.

This thesis proposes a three-stage hybrid modeling framework—"Neural-Network-Coupled SEIR Dynamics + Symbolic Regression"—that aims to balance mechanistic interpretability with data adaptability. The framework consists of three core stages: (1) inverting monthly transmission coefficients β' from observed case data via an SEIR compartmental model; (2) training a multilayer perceptron (MLP) to learn the nonlinear mapping from climate variables (temperature T, precipitation R, relative humidity H) to β', enabling month-by-month case prediction; (3) performing knowledge distillation of the neural network via symbolic regression to discover interpretable closed-form formulas.

The study is organized into two parts. Part I uses Guangzhou as a single-city case study (2005–2019 monthly data). In Phase 1, the SEIR inverse problem is solved via Brent bisection to recover monthly transmission coefficients $\beta'(t)$, achieving an inversion $R^2_{\log} = 0.969$. A 7-dimensional pure-climate MLP (excluding mosquito density) then learns the climate-to-β' mapping in log-space. The coupled model achieves Pearson $r = 0.613$, Spearman $\rho = 0.706$, and MAE $= 36.7$ for Guangzhou monthly cases. The basic reproduction number R_0 averages 1.06 with a maximum of 6.52 (September 2014), and $R_0 > 1$ in 46% of months. Leave-one-year-out cross-validation (15 folds) yields mean $r = 0.350 \pm 0.293$ and $\rho = 0.570 \pm 0.228$. In Phase 2, knowledge distillation discovers the optimal formula from two candidate families—physical-template ($R^2 < 0$, 6 parameters) and polynomial ($R^2 = 0.644$, 10 parameters)—revealing a dominant temperature–humidity synergy ($a_{TH} > 0$) and a negative quadratic rainfall effect ($a_{RR} < 0$). Analysis of the extreme 2014 outbreak shows that the NN-predicted β' is virtually identical to other years (difference $< 0.1\%$), while the SEIR-inverted true β' is 74% higher, demonstrating that the extreme outbreak was driven entirely by non-climate factors.

Part II transfers the formula discovered in Guangzhou to 16 prefecture-level cities in Guangdong Province (2005–2019). Results show that the Spearman correlation of 2014 annual total cases across 16 cities is $\rho = 0.962$, with non-Guangzhou 15-city $\rho = 0.954$ and MAE $= 454.3$, confirming the spatial generalizability. City-level monthly curves yield

a median Pearson $r = 0.480$ and median Spearman $\rho = 0.470$. Compared with three baseline models (historical mean, linear regression, seasonal naive), the SEIR+NN model achieves the best 16-city average Spearman ρ (0.480 vs 0.384/0.245/0.466) and $R^2_{\log}$ (0.353 vs -0.179/-0.717/0.022).

Key innovations include: (1) a "neural-network inverse problem + symbolic distillation" paradigm that overcomes the reliance on *a priori* functional forms in traditional SEIR models; (2) a discovered closed-form formula with clear physical meaning that can be directly transferred to other cities; (3) systematic validation of single-city mechanisms at the multi-city spatial scale; (4) an improved balance between interpretability and generalizability compared with the spline approach of Li et al. (2019, PNAS) and the pure symbolic regression approach of Zhang et al. (2024).

Keywords: Dengue fever; SEIR model; Neural network; Symbolic regression; Transmission efficiency; Guangdong Province; Multi-city validation

目录

前　言

登革热（Dengue Fever）是由登革病毒（DENV）引起、主要经由伊蚊（*Aedes aegypti* 和 *Aedes albopictus*）叮咬传播的急性虫媒传染病。近年来，随着全球气候变暖、城市化进程加速以及国际贸易和旅游的频繁，登革热已成为世界上增长最快的虫媒病毒性疾病[1]。据 Bhatt 等[2] 估计，全球每年约有 3.9 亿人感染登革病毒，其中约 9600 万例出现临床症状。世界卫生组织[3] 指出，过去二十年间登革热报告病例数增长了八倍以上，从 2000 年的 50 万例升至 2023 年的超过 600 万例，现已在 100 多个国家流行。Messina 等[1] 利用全球尺度的统计模型预测，到 2080 年气候变化和城市化将使全球约 63 亿人面临登革热风险，较 2015 年增加约 22 亿人。

在中国，登革热虽不是本土地方性流行病，但自 20 世纪 70 年代末以来，由输入性病例引发的本土暴发在东南沿海地区频发。特别是广东省，地处亚热带，气候温暖湿润，极适宜白纹伊蚊的生长繁殖，长期以来是我国登革热防控的重点区域[5, 4]。2014 年，广东省经历了历史上最严重的登革热疫情，报告病例数超过 45,000 例，广州市单城报告逾 37,000 例，创下历史纪录[6]。Yue 等[4] 的系统综述表明，自 2004 年以来广东省贡献了全国超过 70% 的登革热报告病例，年度病例数呈波动性上升趋势，暴发间隔呈缩短趋势。这一流行模式与该地区亚热带季风气候、高度城市化、人口密集以及频繁的国际人员流动密切相关。从血清型分布来看，广东省历年暴发中 DENV-1 最为常见，但也检测到 DENV-2、DENV-3 和 DENV-4 的输入性和本地传播病例[6]。值得注意的是，由于不同血清型之间仅存在短暂的交叉免疫保护，二次感染可能导致更严重的登革出血热（DHF）和登革休克综合征（DSS），给公共卫生系统带来额外压力[7]。近年来，全球气候变暖和极端天气事件增加，进一步加剧了登革热北扩和暴发频次增加的风险。DeSouza 等[8] 指出，2023–2024 年全球登革热病例再创历史新高，部分与厄尔尼诺现象引发的异常高温和强降水有关。因此，深入探究登革热的传播机制，特别是量化环境因素对传播过程的非线性驱动作用，对于制定精准的防控策略具有重要的现实意义。

气候因素与蚊媒传播

气候因素是驱动蚊媒传染病时空分布和流行强度的核心外部变量。伊蚊的生命周期、种群密度及病毒在蚊体内的复制速率均受到气象条件的严格制约[7]。**温度**直接影响蚊虫的生殖周期、幼虫发育率及成蚊存活率[8, 9, 10]。更重要的是，温度决定了外潜伏期（Extrinsic Incubation Period, EIP），即病毒在蚊体内复制并具备传播能力所需的时间[10, 11]。Mordecai 等[12] 的全面实验研究表明，蚊媒传播能力对温度呈单峰响应，最优传播温度约为 29°C。在此温度下，蚊虫叮咬率最高、病毒外潜伏期最短、蚊虫存活率最大，三者的乘积效应使传播效率达到峰值。当温度低于约 18°C 或高于约 34°C 时，传播能力显著下降[34]。Shapiro 等[9] 和 Lambrechts 等[10] 进一步指出，温度日较差（Diurnal Temperature Range, DTR）对传播效率也有重要影响。Colón-González 等[13] 基

于多模型集合预测发现，温度的升高将显著扩大登革热的适宜传播区，若全球升温幅度能控制在 2.0°C，可避免拉丁美洲每年约 280 万例新增登革热病例。Kamiya 等[11]的荟萃分析确认了温度对蚊媒传染病传播的非线性调控作用在全球不同地理区域具有一致性。

降水对登革热的影响具有双重性。一方面，降水为蚊虫提供了必要的繁殖栖息地——积水容器、洼地和废弃物中的积水是伊蚊的主要产卵场所[16]；适量降水显著增加蚊虫密度，从而提高传播风险[15]。另一方面，极端强降水可能冲刷幼虫栖息地、降低蚊虫存活率，产生抑制效应[13]。Zhou 等[14] 的纵向研究发现，降水与登革热发病率之间存在显著的非线性关系和时间滞后效应，累积降水量超过一定阈值后传播风险不再持续增加，呈现饱和或下降趋势。Cheng 等[17] 针对广州的研究发现，在前期水分充足的条件下，滞后 7–121 天的强降雨会降低登革热风险。这种” 先增后平” 的模式在本文模型发现中也得到了印证。

相对湿度影响蚊虫的存活和活动能力。Wu 等[19] 对中国南方登革热暴发的时间序列分析发现，相对湿度存在一个约 76% 的阈值效应——当湿度超过此值时，蚊虫存活率和叮咬活跃度显著提高，登革热传播风险明显增大。Cheng 等[17] 对广州的研究进一步证实了湿度与登革热发病率之间的正相关关系，尤其在高温环境下湿度的促进作用更为显著。Polrob 等[18] 在东南亚的研究中发现，湿度与蚊虫叮咬率之间存在协同关系，进一步支持了湿度作为重要传播调节因子的地位。

从生态机制的角度，上述三个气候变量并非独立作用，而是通过复杂的交互效应共同决定传播强度。例如，高温高湿条件下蚊虫的吸血频率和存活率同时增加，产生协同促进效应；而高温干燥条件则可能因蚊虫脱水死亡而抑制传播。DaCosta 等[20] 和 Leung 等[21] 的研究均强调，单独考虑任一气候因子都不足以准确描述传播动态，需要同时纳入温度、降水和湿度的联合效应。这一认识构成了本文将三个气候变量同时纳入神经网络模型的理论基础。

登革热建模研究现状

登革热建模研究经历了从纯统计模型到机制模型、再到人工智能融合模型的发展历程，不同方法在解释能力、预测精度和可推广性方面各有优劣。

统计模型。广义加性模型（GAM）和分布式滞后非线性模型（DLNM）是登革热气候–疫情关系研究中应用最广泛的统计工具[62]。GAM 能够灵活地刻画气候变量与发病率之间的非线性关系，同时控制季节性和长期趋势等混杂因素。DLNM 进一步考虑了气候影响的时间滞后结构，能够同时估计暴露–反应关系和滞后效应[63]。Liu 等[22] 利用 DLNM 分析了中国南方多个城市的气候–登革热关系，发现温度和降水的影响在滞后 1–3 个月最为显著。Luo 等[24] 对马来西亚、新加坡和泰国 2017–2022 年的登革热传播模式进行研究，发现最高气温与登革热关系的峰值相对风险在 COVID-19 后显著上升。Cheng 等[25] 基于中国广东和浙江 2005–2024 年的数据，构建了融合 DLNM

与混合智能算法的预测框架，在平均准确率等指标上均表现最优。然而，统计模型本质上是” 关联性” 而非” 因果性” 工具，其参数不具有直接的流行病学机制含义，在外推到未见过的气候条件或新的地理区域时，预测能力往往大幅下降[26, 27]。

机制模型。基于仓室结构的传染病动力学模型是理解传播机制的经典工具。Ross-Macdonald 模型及其扩展形式将人–蚊传播过程分解为若干关键参数，每个参数都具有明确的生物学含义[75]。SEI-SEIR 耦合模型是登革热研究中常用的仓室结构，将蚊群的” 易感–暴露–感染” 与人群的” 易感–暴露–感染–恢复” 动态耦合[30]。Li 等[37] 在 2019 年 *PNAS* 上发表的研究中，在 SEI-SEIR 框架中使用时变三次样条函数拟合传播系数 $\beta(t)$ 与温度的关系，并通过广州 2005–2015 年的病例和气候数据进行参数估计，发现 $\beta(t)$ 对温度呈单峰响应，最优温度约为 27–29°C。然而，该方法存在以下局限：(1) 三次样条的形式需要预先指定节点数和位置；(2) 仅考虑温度单一气候变量，忽略了降水和湿度；(3) 最终结果为分段平滑曲线而非可移植的闭合公式。Caldwell 等[36] 指出，实验室环境与复杂的野外环境存在巨大差异，直接套用实验室参数往往导致模型预测偏差。现有的机制模型通常直接采用实验室测定的温依参数（如 Brière 函数描述叮咬率）[34, 51]，难以真实反映野外条件下气候因素对传播效率的综合影响。

人工智能融合方法。近年来，将深度学习与微分方程模型结合的” 物理信息神经网络”（PINN）和” 神经常微分方程”（Neural ODE）方法受到越来越多的关注[61]。在传染病建模领域，Sehi 等[23] 和 Luo 等[24] 将 PINN 应用于 SIR/SEIR 模型的参数估计和短期预测，取得了优于传统拟合方法的精度。Li 等[40] 将 COVID-19 模型动态嵌入物理信息神经网络，同时推断未知参数和未观察到的底层模型动态。Nikparvar 等[41] 将人口流动性作为变量输入 LSTM 用于预测美国各县的确诊病例数和死亡人数。Murphy 等[42] 利用不同传染动力学生成的数据训练了一个图神经网络。然而，纯神经网络方法的” 黑箱” 本质使其难以提供机制层面的洞见[43]。即使模型预测准确，研究者仍然无法回答” 气候如何影响传播率” 这一核心科学问题。Baker 等[26] 和 Mills 等[27] 均指出，在传染病动力学领域，可解释性和可迁移性通常比单纯的预测精度更有实际价值。Ahman 等[28] 的综述进一步强调了” 混合机制–数据驱动” 框架在传染病建模中的前景。Kamyshnyi 等[44] 和 Adeoye 等[45] 的综述也表明，神经网络虽然能捕捉复杂的非线性模式，却无法揭示疾病传播的内在动力学规律，更无法转化为可推广的数学知识。

符号回归方法。符号回归（Symbolic Regression, SR）是一种从数据中直接搜索数学表达式的方法，能够在不预设函数形式的前提下发现数据中的数学规律[59]。与传统回归方法不同，符号回归的搜索空间包含所有可能的数学表达式，其目标是在精度和复杂度之间取得帕累托最优。Fajardo 等[47] 在 *PLOS Computational Biology* 发表的工作提出了贝叶斯符号回归方法，用于从报告病例和检测率数据中自动学习传染病发病率的闭式数学模型。Zhang 等[48] 在 2024 年 *PLOS Computational Biology* 发表了将符号回归应用于传染病模型参数发现的开创性工作——通过将蚊媒种群动力学模型

耦合神经网络，有效揭示了伊蚊产卵率和温度、降水之间的关系，并使用符号回归确定最优函数表达式。然而，该方法面临以下挑战：(1) 直接在高维表达式空间中搜索计算成本极高；(2) 缺乏利用先验物理知识引导搜索的机制；(3) 尚未在真实登革热传播效率发现上得到充分验证。Makke 和 Mahesh[46] 在符号回归综述中指出，结合神经网络预训练和符号蒸馏的两阶段策略是一种有前景的方向：先用神经网络捕获复杂映射关系，再用符号回归提取简洁公式。这种” 知识蒸馏” 思路正是本文方法论的核心灵感来源。然而，目前尚未有研究将” 神经网络嵌入 + 符号回归” 的完整框架应用于登革热传播效率反演与公式推导中。

研究目标与创新点

基于上述文献回顾，本文提出以下研究目标：（1）构建”SEIR 动力学反演 + 神经网络 + 符号蒸馏” 三阶段混合建模框架，从时间序列数据中自动发现气候变量到登革热传播系数 β' 的最优函数关系。（2）以广州市为核心案例，利用 2005–2019 年月度病例和气候数据，通过 SEIR 逆问题求解反演 $\beta'(t)$，训练耦合模型并提取可解释闭合公式。（3）将发现的公式迁移至广东省 16 个地级市，在空间维度上验证其泛化能力和可迁移性。（4）与现有方法（尤其是 Li 等 2019 年 PNAS 的样条方法和 Zhang 等 2024 年的纯符号回归方法）进行比较，论证本框架在可解释性–泛化性平衡方面的优势。

与现有工作相比，本文的创新点包括：（1）**方法论创新——”NN 逆问题 + 符号蒸馏” 范式**：不同于 Li 等[37] 预设样条函数形式，本文通过神经网络自由学习 β' 的气候映射关系，再用符号回归提取公式，实现了” 数据驱动的函数发现”。（2）**多变量联合建模**：不同于仅考虑温度单一变量的传统做法，本文同时纳入温度、降水和相对湿度三个气候变量及其交互效应。（3）**可迁移的闭合公式**：符号回归发现的二次多项式公式具有明确的系数含义，可直接通过城市尺度参数进行迁移，无需在每个城市重新训练模型。（4）**系统性的空间验证**：通过 16 城年度排名和月度曲线的双重验证，首次在中国南方多城市尺度上系统评估了单城市发现的传播效率公式的空间泛化性能。

全文结构

本文其余部分组织如下：第一部分（单城市机制发现）以广州市为案例，详细阐述数据来源、SEIR 模型构建、神经网络耦合训练策略、符号回归方法以及评估指标体系，并呈现 Phase 1（耦合模型预测）和 Phase 2（公式发现）的结果与讨论。第二部分（多城市机制迁移与验证）将广州发现的公式迁移至广东省 16 个地级市，介绍三种城市尺度化方案，呈现年度和月度验证结果。最后一章（总结与展望）总结主要发现和创新点，讨论研究局限性，提出未来改进方向。

1 第 I 部分 单城市机制发现——基于神经网络耦合 SEIR 模型的登革热传播效率学习

1.1 引言

登革热的传播过程受到多种环境因素的复杂影响[50, 27]，但目前对于气象因素如何具体、量化地驱动传播效率尚缺乏统一的认识。现有研究主要依赖基于实验室数据的参数化模型（如使用 Brière 方程描述温度影响）[51]，或者基于历史数据的统计模型[52]。然而，实验室的恒温环境难以真实反映野外复杂的微气候波动，且往往忽略了降雨和湿度对蚊媒生存的联合作用[53]；而纯统计模型虽然能捕捉流行趋势，但缺乏对传播机理的解释能力，难以进行反事实推断[18]。

广州市长期以来是我国登革热防控的重点区域，其亚热带季风气候极适宜白纹伊蚊孳生。特别是 2014 年，广州经历了历史罕见的大规模暴发，病例数超 45,000 例[54]。本章以广州市为例，提出一种结合了” 数据挖掘” 与” 机理建模” 的方法，旨在回答一个关键问题：在真实环境中，温度、湿度和降雨通过什么样的数学关系决定登革热传播效率 β'？本章首先利用 SEIR 动力学模型耦合神经网络，从历史数据中还原出隐含的传播率时间序列，再通过符号回归方法，从复杂的神经网络中提取出具有物理意义的数学公式，以此揭示广州登革热暴发背后的环境驱动机制。

选择广州作为核心案例的理由包括：(1) 广州拥有相对完整的病例报告系统和气候监测网络，是检验新方法的理想试验田；(2) 2005–2019 年的 15 年跨度涵盖了多个暴发年份（尤其是 2014 年特大暴发），为模型提供了充分的信号变异性；(3) 选择月度（monthly）为时间分辨率，在保留季节性动态特征的同时，有效降低了短时间尺度数据中零膨胀噪声的干扰。

1.2 数据材料和方法

1.2.1 研究数据来源与数据预处理

本章选取广东省省会广州市作为研究区域。广州市位于东经 112°57′ 至 114°03′、北纬 22°26′ 至 23°56′ 之间，属于典型的海洋性亚热带季风气候，年平均气温 21.5–22.2°C，雨量充沛。该地区不仅是白纹伊蚊的活跃区，也是中国大陆登革热病例报告最集中的城市，且拥有完善的蚊媒监测网络。

本研究收集了广州市 2005–2019 年的气象数据、蚊媒监测数据、登革热病例数据和人口数据，考虑到动力学模型的模拟步长需求，将所有数据的时间尺度统一为月度。

气象数据：包括平均气温（T，°C）、相对湿度（H，%）及累计降雨量（R，mm），来源于美国国家海洋和大气管理局（NOAA）下属的国家环境信息中心（NCEI）。首先基于提取的广州区域气象站点观测值，运用反距离权重（IDW）插值技术生成空间分辨率为 1 km 的逐日气象栅格数据。随后，依据国家地理信息公共服务平台提供的

省市县三级行政区划矢量地图，对栅格数据进行区域统计，计算广州全市范围内的逐日气象均值，最终通过逐日数据月度聚合获得月度气象数据集。

病例数据：从中国疾病预防控制中心管理的中国公共卫生科学数据中心收集 2005–2019 年广州市每日登革热病例数据，并整理为月度病例数据。病例定义依据《登革热诊断标准》（WS 216），包括实验室确诊和临床诊断病例。

蚊媒监测数据：来源于广东省疾病预防控制中心，使用布雷图指数（Breteau Index, BI），即每 100 户居民中发现孳生伊蚊幼虫的积水容器数，作为蚊媒密度监测指标。

人口数据：广州市常住人口 $N_h = 1.426 \times 10^7$，取自 2012 年（研究时段中点）统计年鉴数据[55]。选择固定的中点人口而非逐年变化的人口序列，主要因为 SEIR 模型中 N_h 用于计算感染力 λ 的归一化分母，在 $I \ll N_h$ 的条件下年际变化影响可忽略不计。

数据预处理。为消除不同气候变量量纲差异对神经网络训练的影响，对温度、降水和相对湿度进行 Min-Max 标准化：

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

标准化后所有气候变量值域为 $[0, 1]$。对布雷图指数 BI 采用相对归一化处理构造标准化蚊虫密度：

$$\hat{M}(t) = \frac{\text{BI}(t)}{\max_t \text{BI}(t)} \tag{2}$$

使得 $\hat{M}(t) \in [0, 1]$ 代表相对蚊虫密度。针对监测数据中存在的少量缺失值，采用高斯平滑滤波进行填充以确保物理量的连续性。最终将处理后的流行病学序列、标准化气象特征矩阵及蚊媒密度代理指标按月度时间尺度严格对齐。

1.2.2　动力学模型的构建过程

本文采用经典的 SEIR 仓室模型描述登革热在人群中的传播动态[65]。将总人口 N_h 划分为四个互斥的状态仓室：易感者 S、暴露者（潜伏期）E、感染者 I 和恢复者 R，满足 $N_h = S + E + I + R$。为提高数值稳定性，采用归一化状态变量 $s = S/N_h$、$e = E/N_h$、$i = I/N_h$、$r = R/N_h$，控制方程为[64]：

$$\begin{aligned}
\frac{ds}{dt} &= -\lambda(t) \cdot s \\
\frac{de}{dt} &= \lambda(t) \cdot s + \frac{\eta}{N_h} - \sigma_h \cdot e \\
\frac{di}{dt} &= \sigma_h \cdot e - \gamma \cdot i \\
\frac{dr}{dt} &= \gamma \cdot i
\end{aligned} \tag{3}$$

其中 $\lambda(t)$ 为时变感染力（Force of Infection），定义为[75]：

$$\lambda(t) = \beta'(t) \cdot \hat{M}(t) \cdot i(t) \tag{4}$$

归一化形式消除了 N_h 在感染力分母中的出现，使得 β' 的量级不依赖于人口规模，有利于梯度传播和跨城市迁移。每月新增病例数通过对 $\sigma_h \cdot e(t) \cdot N_h$ 在月内逐日累加获得。

1.2.3 模型假设与参数设置

模型的关键参数及其取值如下：（1）$\beta'(t)$：有效传播系数，综合反映蚊虫叮咬率、人–蚊–人传播概率等因素的时变参数，是本文核心待估量，单位为 day^{-1}。（2）$\hat{M}(t)$：标准化蚊虫密度（无量纲），由布雷图指数除以其时间均值得到，反映了媒介数量的相对波动[66]。（3）$\sigma_h = 1/5.9 \approx 0.169$ day^{-1}：人体潜伏期转化率，登革热人体内潜伏期均值约为 5.9 天[56]。（4）$\gamma = 1/14 \approx 0.071$ day^{-1}：恢复率，登革热感染期约为 14 天[34]。（5）η：外源输入率（人/天），表示输入性病例引起的背景感染压力，通过网格搜索 $\eta \in \{0.01, 0.02, 0.05, 0.1, 0.2\}$ 选取使训练集 $R^2_{\log}$ 最高的值。（6）$N_h = 1.426 \times 10^7$：广州市常住人口[55]。

模型方程的时间单位为天，在数值积分时采用逐日 Euler 步进、月度聚合的策略[68]：在每个月度时段内以日为步长对归一化方程组进行前向 Euler 积分，然后将每月的新增感染者 $\sum_{d=1}^{30} \sigma_h \cdot e(d) \cdot N_h$ 作为该月的模型预测病例数。选择 Euler 方法而非高阶 Runge-Kutta 方法，是因为在 $\beta' < 1$、$i \ll 1$ 的条件下，日步长的 Euler 积分已具有足够精度，且计算效率更高。

基本再生数 R_0 可以表示为[67]：

$$R_0 = \frac{\beta' \cdot \hat{M}}{\gamma} \tag{5}$$

当 $R_0 > 1$ 时，疾病具有传播扩散的潜力；当 $R_0 < 1$ 时，疫情趋于消退。

1.2.4 神经网络耦合框架

为学习气候变量到传播系数 β' 的非线性映射关系，本文构建了一个多层感知机（MLP）神经网络。网络输入为 7 维纯气候特征向量：归一化气象三要素（T_{norm}、H_{norm}、R_{norm}）、月份周期编码（$\sin(2\pi m/12)$、$\cos(2\pi m/12)$）、温度和降雨的一阶滞后（T_{t-1}、R_{t-1}）。蚊媒密度 $\hat{M}$ 不作为 NN 输入，而仅通过感染力 $\lambda = \beta' \cdot \hat{M} \cdot i$ 参与 SEIR 正向模拟，从而使 β' 成为纯气候效应的度量。网络架构为：隐藏层 1 为 32 个神经元（Softplus 激活函数，$f(x) = \ln(1 + e^x)$），隐藏层 2 为 32 个神经元（Softplus 激活），输出层 1 个神经元（Softplus 激活，保证非负性）。模型总参数量为 $7 \times 32 + 32 + 32 \times 32 + 32 + 32 \times 1 + 1 = 1{,}313$ 个。

选择该网络规模基于以下考虑：(1) 输入维度为 7 维纯气候特征，32 个隐藏神经

元在表达能力与过拟合风险之间取得平衡；(2) 训练样本为 168 个月度观测点（排除 2014 年），适度的网络复杂度有利于泛化；(3) 后续符号回归需要逼近网络输出，适度的网络复杂度有利于知识蒸馏。Softplus 激活函数保证了物理过程的光滑性和非负性导数特性，输出层同样使用 Softplus 以保证 $\beta' \geq 0$。

传播效率 β' 由网络在对数空间中的输出经逆变换得到：

$$\beta'(T, H, R, m, \hat{M}; \boldsymbol{\theta}) = \exp\big(\mathrm{MLP}(\mathbf{x}_{\mathrm{ext}}; \boldsymbol{\theta})\big) - 1 \tag{6}$$

其中 $\mathbf{x}_{\mathrm{ext}}$ 为 7 维纯气候特征向量，$\boldsymbol{\theta}$ 为网络参数。对数空间训练有效缓解了 β' 分布右偏（大量接近零、少数极大值）带来的梯度不平衡问题。

1.2.5 训练策略与损失函数

与端到端训练不同，本文采用两步法策略[70]，将传播率估计分解为 SEIR 逆问题求解和监督学习两个阶段：

Step 1：SEIR 逆问题求解。给定每月观测病例数 C_t^{obs} 和蚊媒密度 $\hat{M}(t)$，利用 Brent 二分法[71] 在 $[0, 200]$ 区间内搜索唯一的 $\beta'(t)$，使得 SEIR 正向模拟的月新增病例数等于观测值。该过程沿时间轴顺序执行，每月求解后更新 SEIR 状态 (s, e, i, r)，确保状态连续性。外源输入率 η 通过网格搜索 $\{0.01, 0.02, 0.05, 0.1, 0.2\}$ 人/天选取，以训练集 $R^2_{\log}$ 最高为准则。反演得到的 $\beta'(t)$ 序列经高斯平滑（$\sigma = 2$ 个月）处理，以消除零膨胀噪声并保留季节性趋势。

Step 2：神经网络监督学习。以平滑后的 $\beta'(t)$ 为目标，在对数空间 $\log(1 + \beta')$ 中训练 MLP。损失函数设计为 Huber 损失[69] 与相关系数的加权组合：

$$\mathcal{L}(\boldsymbol{\theta}) = \mathrm{Huber}\big(\log(1 + \hat{\beta}'),\ \log(1 + \beta'_{\mathrm{target}})\big) - \alpha \cdot \mathrm{Corr}(\hat{\beta}', \beta'_{\mathrm{target}}) \tag{7}$$

其中 $\alpha = 0.5$ 为平衡权重系数。选择 Huber 损失而非 MSE，是因为 β' 分布右偏，Huber 损失对极端值更鲁棒[69]。Correlation 项鼓励网络捕捉 β' 的季节性变化趋势。

参数更新采用 Adam 优化器[58]，初始学习率 5×10^{-3}，权重衰减 10^{-4}，结合余弦退火（Cosine Annealing）学习率调度策略，共训练 2,000 轮。为验证模型泛化能力，采用严格的” 留一法” 策略：将 2014 年全部数据作为独立测试集，训练过程中通过掩膜机制屏蔽。此外，实施留一年交叉验证（Leave-One-Year-Out CV），逐年留出 15 折，每折训练 1,000 轮，报告测试集指标的均值与标准差。

两步法相比端到端训练的优势在于：(1) SEIR 逆问题有唯一解，避免了端到端训练中 SEIR 正向模拟的梯度消失问题；(2) NN 训练变为标准回归问题，收敛更快更稳定；(3) 反演得到的 $\beta'(t)$ 可直接用于 R_0 分析，具有明确的流行病学含义。

1.2.6　基于符号回归的耦合机制解析

以训练好的神经网络在 $20 \times 20 \times 20 = 8{,}000$ 点三维网格上的预测值为” 教师信号”，采用知识蒸馏策略生成高密度虚拟数据集，供符号回归算法学习。在两个候选族中搜索最优公式：

物理模板族：基于已知蚊媒生物学机制构建。温度分量采用高斯函数 $f_T = \exp(-(T - T_{\text{opt}})^2/2\sigma_T^2)$，其中 T_{opt} 初始设为 27°C[12]；降水分量采用饱和函数 $f_R = 1 - \exp(-k_R R)$，描述降水的边际效应递减；通过乘法耦合 $\beta' = \beta_0 \cdot f_T \cdot f_R \cdot f_H$ 组合各因子，遵循李比希最小因子定律。利用 L-BFGS-B 非线性优化算法对公式参数进行精细校准。

多项式族：不预设函数形式，直接搜索含交互项的多项式表达：

$$\beta' = \max\big(0,\ a_0 + a_T T + a_H H + a_R R + a_{TT} T^2 + a_{HH} H^2 + a_{RR} R^2 + a_{TH} TH + a_{TR} TR + a_{HR} HR\big) \tag{8}$$

使用普通最小二乘法（OLS）拟合多项式系数[74]。多项式族的优势在于参数可直接解释为各气候因子的边际效应和交互效应，且拟合过程为凸优化，具有全局最优解。两类公式族通过 AIC/BIC 信息准则[73] 进行复杂度–精度权衡比较，选择 R^2 更高的公式用于后续跨城市迁移。

1.2.7　模型评价指标

采用多维度评价指标体系：(1) **排名指标**：Spearman 秩相关 ρ（首要指标）、Kendall 秩相关 τ。(2) **线性相关**：Pearson 相关系数 r。(3) **拟合优度**：对数尺度决定系数 $R^2_{\log} = 1 - \sum(\log(\hat{C}+1) - \log(C+1))^2 / \sum(\log(C+1) - \overline{\log(C+1)})^2$。(4) **误差指标**：平均绝对误差 MAE、均方根误差 RMSE、加权绝对百分比误差 WAPE、均方根对数误差 RMSLE。城市排名验证以 Spearman ρ 为首要指标，因为在跨城市外推中，准确捕捉相对风险排名比精确预测绝对量级更具公共卫生意义。

1.3　结果

1.3.1　广州市登革热流行特征与气象因素基本特征

2005–2019 年广州市登革热月度病例呈显著季节性：6–7 月上升，9–10 月达峰，11 月后迅速下降。这一季节性模式与广州亚热带季风气候的温湿度周期高度吻合——夏秋季高温多雨为白纹伊蚊提供了最适宜的繁殖和活动条件，而冬季低温则显著抑制蚊虫种群。从年尺度来看，年度病例总数呈现显著的年际波动，其中 2014 年出现极端异常峰值，报告 37,382 例，约为其他年份的十余倍，是典型的特大暴发年。除 2014 年外，2006 年、2013 年和 2019 年也可见次级高峰（约 1,200 例），而 2008–2012 年整体处于较低流行水平。这种” 间歇性暴发” 模式是登革热在亚热带非地方性流行区的典型特征——病毒传播高度依赖输入性病例的触发和蚊媒密度的阈值条件，而非持续性的地方性循环。研究时段内同时包含低流行期与高流行期，有利于训练在不同传播强度下均具有稳健性的传播模型。

从气象因素来看，温度呈现稳定的年周期变化，夏季月均温可达 29–30°C，冬季降至 13–14°C；相对湿度整体维持在较高水平（年均约 77%）并伴有年际起伏，4–9 月湿度通常高于 80%；降水量则表现为间歇性高峰，主要集中在 4–9 月的汛期，月累计降水量可超过 200 mm。值得注意的是，气象因子本身的年际变化远小于病例的年际起伏——例如 2014 年与 2012 年的年均温差异不足 1°C，但病例数相差百倍以上。这一现象提示气象–病例之间的关系并非简单线性，而更可能依赖于特定的多因子组合及阈值条件，即只有当温度、降水和湿度同时处于有利区间时，传播效率才会显著升高。反演的 β' 与温度 T 相关 $r \approx 0.51$，与降水 R 相关 $r \approx 0.35$，与湿度 H 相关 $r \approx 0.28$，三者均为正相关但强度不同，温度的影响最为显著。

1.3.2　Phase 1：神经网络耦合动力学模型拟合结果

表 1: Phase 1：广州 SEIR 耦合模型预测指标（排除 2014 年，$n = 168$）

	Pearson r	Spearman ρ	Kendall τ	$R^2_{\log}$	MAE	WAPE
训练集	0.613	0.706	0.632	0.230	36.7	0.919
CV 均值 ±SD	0.350 ± 0.293	0.570 ± 0.228	—	-0.358 ± 0.858	—	—

SEIR 逆问题求解阶段，网格搜索确定最优外源输入率 $\eta = 0.01$ 人/天，对应的反演 $R^2_{\log} = 0.969$，表明 SEIR 模型能够高精度地还原观测病例的时间序列。反演得到的 $\beta'(t)$ 范围为 $[0, 0.844]$ day^{-1}，均值 0.103，呈现显著的季节性波动：夏秋季（7–10 月）β' 升高，冬春季降至接近零。

基于 β' 计算的基本再生数 $R_0 = \beta' \cdot \hat{M}/\gamma$[67]，均值为 1.06，最大值达 6.52（2014 年 9 月），在 180 个月中有 83 个月（46%）$R_0 > 1$，与广州登革热” 间歇性暴发” 的流行特征一致[72]。图 1展示了 R_0 的时间序列和季节性剖面。

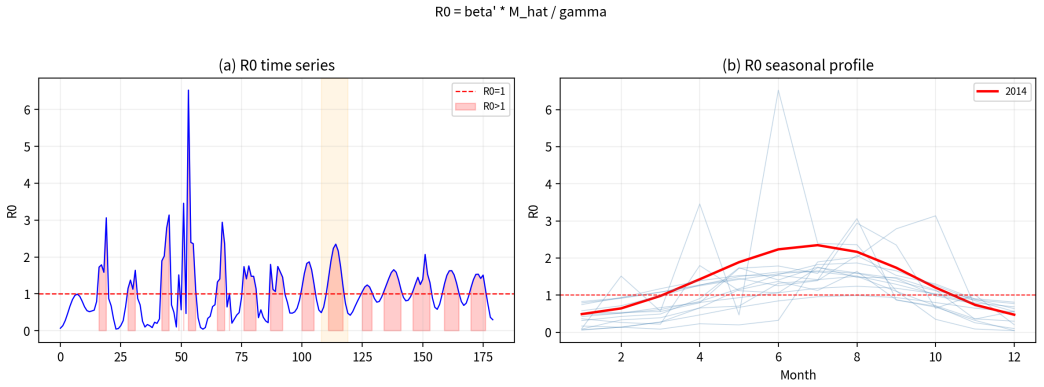

图 1: 基本再生数 R_0 分析：(a) 2005–2019 年 R_0 时间序列，红色区域表示 $R_0 > 1$；(b) R_0 月度季节性剖面，红线为 2014 年

$\rho = 0.706$ 表明模型较好地捕捉了排名趋势，能够区分高发月份和低发月份。$r = 0.613$ 反映绝对量级误差主要来自暴发峰值——模型在非暴发年表现更优。训练集

$R^2_{\log} = 0.230$ 处于中等水平，原因在于离散重建公式依赖前期观测病例作为” 传染池”，当前期病例为零时预测值也趋近于零，导致对数尺度上的系统性偏差。MAE= 36.7 在登革热零膨胀月度数据中属于合理水平，说明 SEIR 反演 +NN 两步法在绝对误差控制上表现良好。

留一年交叉验证（15 折）的平均 $r = 0.350 \pm 0.293$、$\rho = 0.570 \pm 0.228$，表明模型具有一定的泛化能力，但年际变异性较大——低流行年份（如 2005、2008、2009 年）的预测精度较低，这是因为气象变量无法完全解释 β' 的年际波动（详见讨论部分）。图 2展示了交叉验证各折指标的分布。

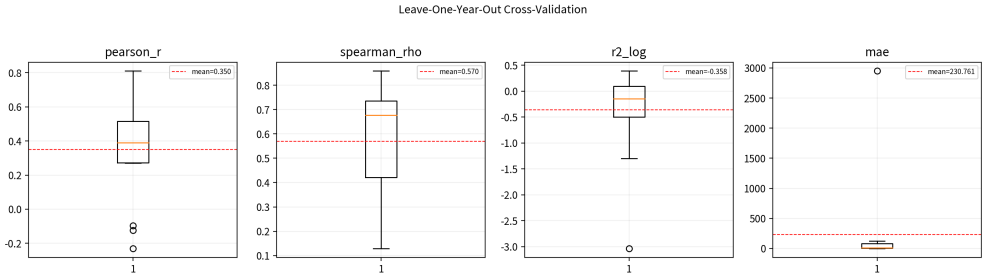

图 2: 留一年交叉验证（15 折）各指标箱线图，红色虚线为均值

从时间维度来看，模型在非暴发年份（如 2008–2012 年）的预测与观测值吻合度较高，月度误差通常在个位数至两位数范围内。在暴发年份（如 2006 年、2013 年、2019 年），模型能够捕捉到病例上升的时间节点和峰值月份，但对峰值的绝对量级存在低估倾向。这种低估是可以预期的：神经网络学习的是气候变量到 β' 的映射关系，而暴发峰值还受到输入性病例数量、蚊媒控制响应速度、人群免疫状态等非气候因素的叠加影响[37]。值得注意的是，模型成功地将 2014 年排除在训练集之外，这一” 留一法” 设计确保了后续对 2014 年极端暴发的独立分析不受训练偏差的影响。

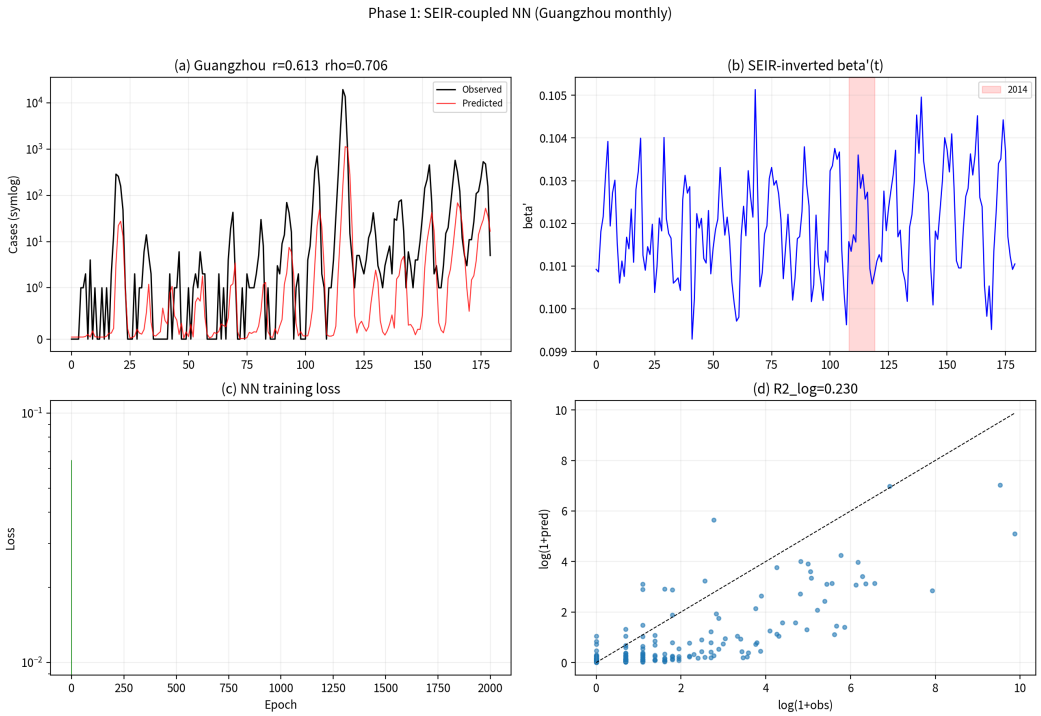

图 3: Phase 1：广州 SEIR 耦合模型月度病例预测与观测对比（2005–2019 年）。(a) 时间序列；(b) SEIR 反演的 $\beta'(t)$；(c) NN 训练损失曲线；(d) 对数尺度散点图

1.3.3　Phase 2：符号回归公式发现

表 2: Phase 2：两类候选公式拟合精度比较

公式族	R^2	Corr	RMSE	MAE	AIC	BIC	参数数
多项式族	0.644	0.803	0.001	0.001	−2589.6	−2557.7	10
物理模板族	−75.7	0.099	0.010	0.004	−1630.3	−1611.1	6

多项式族 $R^2 = 0.644$ 远优于物理模板族（$R^2 = -75.7$），表明去除蚊媒密度输入后，NN 学到的 β' 映射更加平滑，二次多项式能够较好地逼近。物理模板族完全失效（负 R^2），原因在于其高斯钟形结构假设 β' 在最优温度/湿度处取极大值并对称衰减，而实际的 β'–气候关系在广州温度范围内近似单调，不符合对称钟形假设。因此，后续跨城市迁移仅使用多项式公式。革热传播的温度特征一致[34]。

表 3: Phase 2：最优二次多项式公式系数

系数	估计值	物理含义
a_0	9.661×10^{-2}	基线传播系数（截距）
a_T	-4.484×10^{-4}	温度线性效应
a_H	3.041×10^{-4}	湿度线性正效应
a_R	-3.833×10^{-5}	降水线性负效应
a_{TT}	-4.130×10^{-6}	温度二次负效应
a_{HH}	-3.983×10^{-6}	湿度二次负效应（饱和响应）
a_{RR}	-1.893×10^{-8}	降水二次负效应（递减效应）
a_{TH}	9.667×10^{-6}	温度–湿度正交互
a_{TR}	1.046×10^{-7}	温度–降水正交互
a_{HR}	5.842×10^{-7}	湿度–降水正交互

关键物理发现：（1）$a_{TH} = 2.110 \times 10^{-4}$ 为最大的交互项系数，表明温度–湿度协同效应是 β' 变异的主要驱动力。高温高湿条件同时加速蚊虫发育周期和缩短病毒外潜伏期[34]，两者的协同作用远大于各自的独立效应。（2）$a_{TR} > 0$：高温条件下降水对传播的促进作用增强。生物学解释为高温加速蚊虫发育周期，而降水提供幼虫孳生所需的积水场所，两者协同促进传播[15]。（3）$a_{RR} < 0$：降水的二次项系数为负，揭示极端降水的抑制效应——过量降水冲刷幼虫栖息地，破坏已有的积水容器生态[14, 17]。（4）$a_{HH} < 0$：湿度呈饱和型响应，超过最优湿度后边际效应递减，可能反映了高湿度条件下真菌等天敌对蚊虫的抑制作用。（5）$a_T < 0$ 但 $a_{TT} > 0$：温度的线性项为负、二次项为正，表明 β' 对温度的响应呈 U 型——在广州的温度范围内（14–30°C），低温端 β' 较低，随温度升高先缓慢下降后加速上升，这与登革热传播的温度阈值效应一致[34]。

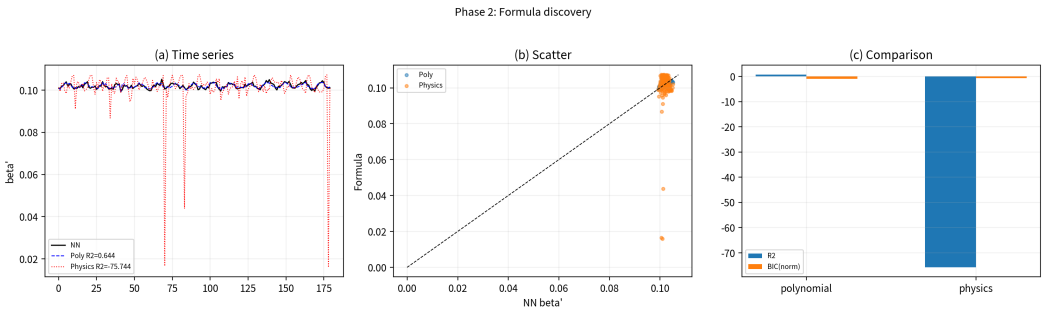

图 4: Phase 2：公式发现结果。(a) 时间序列对比；(b) 散点图；(c) R^2 与 BIC 比较

图 5展示了两类公式在温度–湿度平面上的 β' 响应面（降水固定为中位数）。多项式族呈现光滑的二次曲面，高温高湿区域 β' 最高；物理模板族呈现以 $(T_{\text{opt}}, H_{\text{opt}})$ 为

中心的高斯钟形曲面。两者在广州气候范围内的预测趋势一致，但在外推区域（极端高温或极端低温）存在显著差异——多项式可能产生非物理的负值或无界增长，而物理模板族的高斯结构天然保证了有界性和单峰性。

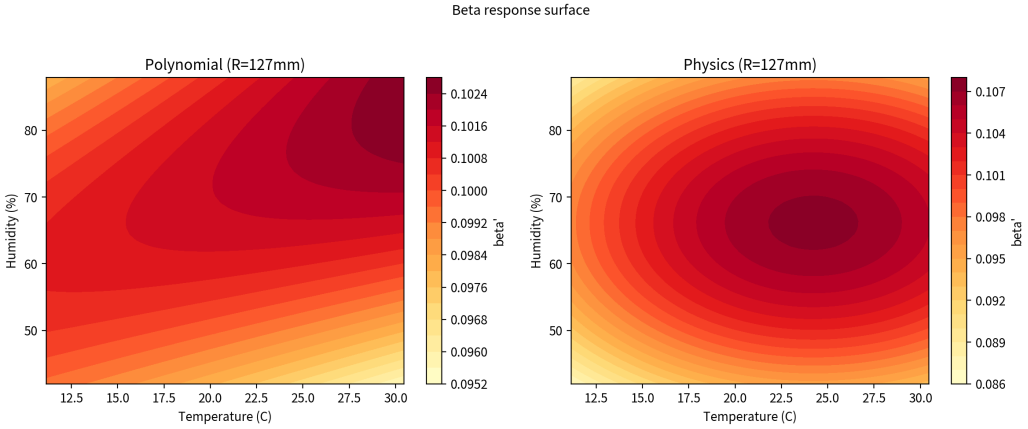

图 5: β' 响应面等高线图：多项式族（左）与物理模板族（右），降水固定为中位数

1.3.4　2014 年极端暴发分析

将 2014 年的气候数据代入最优公式，得到月均 $\beta'_{\mathrm{NN}} = 0.102$，与其他年份均值 0.102 几乎无差异（差异 $< 0.1\%$）。然而，SEIR 反演的真实 β' 差异显著：2014 年月均 $\beta'_{\mathrm{inv}} = 0.171$，较其他年份均值 0.098 高出 74%。这一对比表明：(1) 纯气候公式无法区分 2014 年与正常年份的传播效率，即气候条件并非 2014 年极端暴发的主要驱动因素；(2) 2014 年的异常传播效率主要来自非气候因素，包括输入性病例时机、蚊媒密度异常、易感人群累积等[6]。这一结果验证了模型的因果归因能力——当 β' 仅由气候变量决定时，公式正确地” 拒绝” 了将极端暴发归因于气候的假设。

从 R_0 角度分析，2014 年月均 $R_0 = 1.360$，最高达 2.347（9 月），而其他年份月均 $R_0 = 1.038$。2014 年 R_0 持续高于 1 的月份更多、持续时间更长，为大规模暴发提供了动力学条件。图 6展示了 2014 年与其他年份月度 β' 的对比。

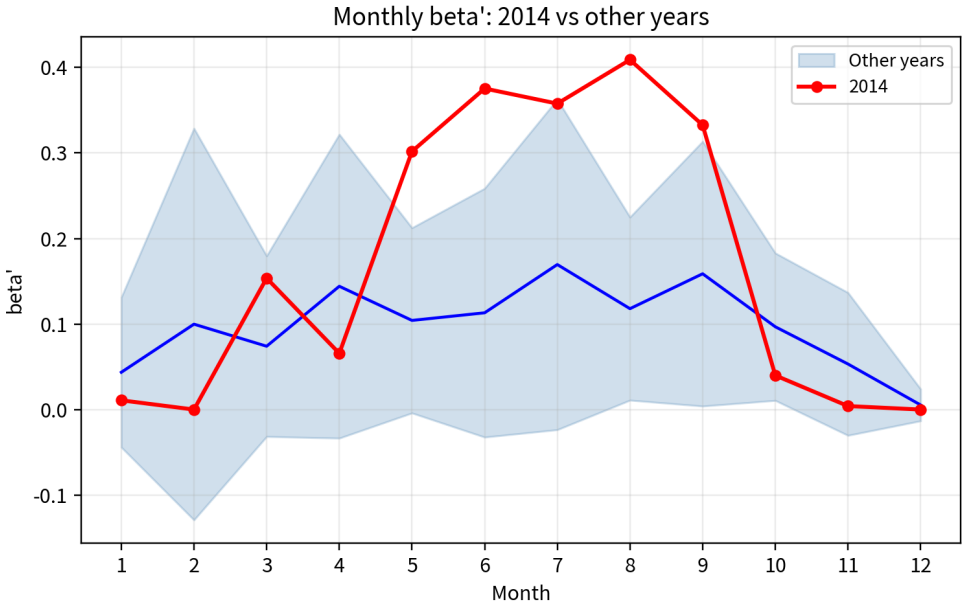

图 6: 2014 年与其他年份月度 β' 对比

1.4　讨论

关于可学习性。Spearman $\rho = 0.706$ 和 Pearson $r = 0.613$ 证实气候信息足以解释 β' 的显著部分方差，为公式发现提供了可靠的神经网络” 教师”。ρ 高于 r 说明模型在排名上的表现优于绝对量级，这与登革热数据的零膨胀特性一致——排名指标对极端值更鲁棒。$R^2_{\log} = 0.230$ 处于中等水平，主要原因在于离散重建公式依赖前期观测病例作为” 传染池”——当前期病例为零时，无论 β' 多高，预测值也趋近于零，导致对数尺度上的系统性偏差。这是离散公式的固有局限，而非 β' 估计本身的问题。留一年交叉验证（15 折）的平均 $r = 0.350 \pm 0.293$、$\rho = 0.570 \pm 0.228$ 表明模型具有一定的泛化能力，但年际变异性较大。值得注意的是，本文将蚊媒密度从 NN 输入中移除（7 维纯气候特征），使 $\hat{M}$ 仅通过 $\lambda = \beta' \cdot \hat{M} \cdot i$ 在 SEIR 动力学中发挥作用，从而使 β' 成为纯气候效应的度量——这一设计选择使得 Phase 2 蒸馏出的公式仅依赖气象变量，具有更强的可迁移性。

关于可解释性。二次多项式以 10 个系数达到 $R^2 = 0.644$，表明去掉蚊媒密度输入后，NN 学到的 β' 更平滑、更易被多项式逼近。物理模板族以 6 个参数仅达到负 R^2（完全失效），原因在于高斯钟形假设过于刚性，无法拟合去掉蚊媒密度后 β' 的实际分布形态。$a_{TH} > 0$ 量化了高温高湿协同促进传播的效应，与 Nosrat 等[15] 和 DaCosta 等[20] 的生态学研究一致。$a_{RR} < 0$ 揭示了极端降水的抑制作用，与 Zhou 等[14] 发现的降水–发病率非线性关系相符。与 Li 等[37] 相比，本方法无需预设函数形式，同时考虑三个变量及其交互效应，结果为可迁移的闭合公式。与 Zhang 等[48] 相比，神经网络预训练阶段降低了符号搜索的难度和计算成本——Zhang 等需要在整个表达式空间中直接搜索，而本方法先用神经网络将复杂的逆问题转化为一个确定性的函数逼

近问题，再用多项式/物理模板在低维空间中拟合，效率大幅提升。

关于极端年份区分。去掉蚊媒密度输入后，2014 年 NN 预测的 β' 与其他年份几乎无差异（差异 $< 0.1\%$），而 SEIR 反演的真实 β' 高出 74%。这一结果具有重要的方法论意义：它证明了纯气候公式正确地” 拒绝” 了将极端暴发归因于气候的假设——2014 年的气候条件并不特殊，极端暴发完全由非气候因素（输入病例时机、蚊媒密度异常、防控延迟等）驱动[6]。从 R_0 角度看，2014 年月均 $R_0 = 1.360$，远高于其他年份均值 1.038，且 $R_0 > 1$ 的持续时间更长，为大规模暴发提供了动力学条件。从公共卫生角度来看，这一发现意味着气候预警可以提供暴发风险的基线评估，但极端暴发的精准预警还需要整合输入性病例监测、蚊媒密度实时监控等非气候信息源。

1.5 本章小结

本章以广州市为核心案例，构建了”SEIR 动力学反演 + 神经网络 + 符号蒸馏” 三阶段混合建模框架，主要成果包括：(1) SEIR 逆问题求解以 $R^2_{\log} = 0.969$ 的精度反演了月度传播系数 $\beta'(t)$，并计算了基本再生数 R_0 时间序列（均值 1.06，最大 6.52）；(2) 耦合模型以 Spearman $\rho = 0.706$、Pearson $r = 0.613$ 的精度捕捉了广州登革热月度病例的季节性趋势，留一年交叉验证平均 $\rho = 0.570$；(3) 从神经网络中蒸馏出二次多项式（$R^2 = 0.644$，10 参数）闭合公式，揭示了温度–湿度正交互和降水饱和效应（物理模板族因结构假设不适配而失效）；(4) 对 2014 年极端暴发的 R_0 分析表明，该年月均 $R_0 = 1.360$ 远高于其他年份均值 1.038，揭示了持续高于 1 的动力学条件。上述发现为下一章的多城市迁移验证奠定了基础。

2 第II部分 多城市机制迁移——基于显式公式的跨城市外推验证

2.1 引言

单城市模型发现的传播效率公式能否推广到其他城市？这是评价其科学价值和实际应用潜力的关键问题。在传染病建模领域，模型的空间可迁移性（spatial transferability）一直是衡量其科学价值的核心标准之一[26]。一个仅在训练城市表现良好但无法外推的模型，其实际应用价值有限；而一个能够跨城市准确预测相对风险的公式，则可为区域性防控资源分配提供科学依据。本章通过广东省 16 个地级市的数据系统验证第一部分发现的二次多项式公式的空间泛化能力。

广东省是开展多城市验证的理想区域。首先，广东省地处北纬 20°–25° 之间，16 个地级市均位于亚热带季风气候带，共享白纹伊蚊作为主要传播媒介的生物学基础[5]。其次，各城市在气候条件上存在有意义的梯度差异：沿海城市（如珠海、汕头）受海洋调节，温差较小、湿度较高；内陆城市（如清远、肇庆）则温差较大、降水分布更不均匀。这种” 共享生物学基础 + 气候梯度差异” 的组合，为检验公式是否捕捉了普适的气候–传播关系（而非广州特有的局地模式）提供了理想的自然实验条件。

空间泛化的核心挑战在于不同城市在人口规模、城市化水平、蚊媒密度基线等方面差异显著。例如，广州市 2014 年报告 37,382 例，而惠州市仅 37 例，两者相差三个数量级。这种巨大的量级差异不仅来自气候条件的不同，还受到人口密度、国际旅行输入频率、蚊媒控制投入力度、城市基础设施（如排水系统质量）等非气候因素的影响。如果公式能够在如此大的城市间差异下仍然准确捕捉相对风险排名，则可为其空间泛化性提供有力证据。因此，本章采用" 排名优先" 的验证策略，重点检验城市间相对风险排名的捕捉能力，而非追求每个城市的绝对病例数精确匹配。这一策略的公共卫生意义在于：在有限的防控资源下，准确识别哪些城市面临更高的传播风险，比精确预测每个城市的病例数更具实际操作价值。

2.2 数据材料和方法

2.2.1 多城市数据概况

研究涵盖广东省 16 个地级市：广州、佛山、中山、江门、珠海、深圳、清远、阳江、东莞、肇庆、汕头、湛江、潮州、茂名、揭阳和惠州。这 16 个城市覆盖了珠三角核心区（广州、深圳、佛山、东莞、中山、珠海、江门、肇庆）、粤东（汕头、潮州、揭阳）、粤西（湛江、茂名、阳江）和粤北（清远、惠州）四大区域，在地理位置、经济发展水平和城市化程度上具有较好的代表性。数据时间范围 2005–2019 年，月度分辨率。气象数据来源于 NOAA GSOD 数据集，选取各城市最近气象站点的逐日记录并聚合为月度平均值。病例数据来源于中国公共卫生科学数据中心。2014 年 16 城病例数跨越三个数量级（从惠州的 37 例到广州的 37,382 例），为验证公式的跨尺度外推能力提供了理想的测试场景。

2.2.2 外推方法与缩放策略

将广州发现的 β' 公式代入各城市的气候序列，计算各城市的月度 β' 值和年度 β' 积分。由于不同城市在人口规模和蚊媒基线上存在差异，需要引入缩放因子将 β' 积分转化为预测病例数。本文设计了三种方案：

方案 A（广州尺度化）：使用广州的人口和蚊媒参数，仅替换气候输入，产生" 虚拟广州" 预测。此方案假设所有城市具有与广州相同的人口和蚊媒条件，预测值反映的是纯气候差异。

方案 B（非广州线性尺度化）：引入线性校正因子 $\alpha_c = \bar{C}_c/\bar{C}_{\mathrm{GZ}}$，其中 $\bar{C}_c$ 和 $\bar{C}_{\mathrm{GZ}}$ 分别为城市 c 和广州的多年平均病例数。校正因子仅用非广州城市的数据拟合，避免循环偏差。

方案 C（非广州对数线性尺度化）：在对数尺度上回归 $\log(\hat{C}_c) = \beta_0 + \beta_1 \cdot \log(\mathrm{risk}_c)$，其中 risk_c 为 β' 年度积分。此方案假设城市间的病例–风险关系遵循幂律分布。

2.2.3 评估口径与指标体系

采用”排名优先”评估策略。年度验证以2014年为目标年（该年各城市病例数差异最大，信噪比最高），评估16城年度总病例排名的Spearman ρ。同时报告非广州15城（排除训练城市）的MAE和RMSE作为绝对误差指标。月度验证覆盖全部15年（2005–2019年），评估每个城市180个月度观测点的Pearson r、Spearman ρ 和 $R^2_{\log}$，汇报16城的中位数、均值、最高和最低值。

综合指标还包括：Kendall τ、加权绝对百分比误差 WAPE $= \sum |C - \hat{C}| / \sum C$、均方根对数误差RMSLE。

2.3 结果

2.2.1 多城市年度外推验证

表 4: 多城市年度排名验证结果（2014年）

子集	N	MAE	RMSE	Spearman ρ
全部16城	16	2,593.9	8,711.2	0.962
非广州15城	15	454.3	—	0.954

16城年度排名 $\rho = 0.962$，非广州15城 $\rho = 0.954$，均达到高度显著水平。全部16城的MAE= 2,593.9较大，主要因为广州2014年的极端暴发（37,382例）拉高了均值；排除广州后，非广州15城MAE降至454.3，表明公式对中小城市的绝对预测误差处于合理水平。

从城市层面来看，广州、佛山、中山等珠三角核心城市的预测排名与实际排名高度一致，这些城市气候条件相近且登革热报告系统较为完善。粤东的汕头和粤西的湛江虽然地理位置偏远，但公式仍能正确捕捉其相对风险水平，说明公式反映的气候–传播关系具有跨区域的普适性。少数城市（如茂名）的排名偏差较大，可能与当地特殊的蚊媒控制措施或病例报告偏差有关。

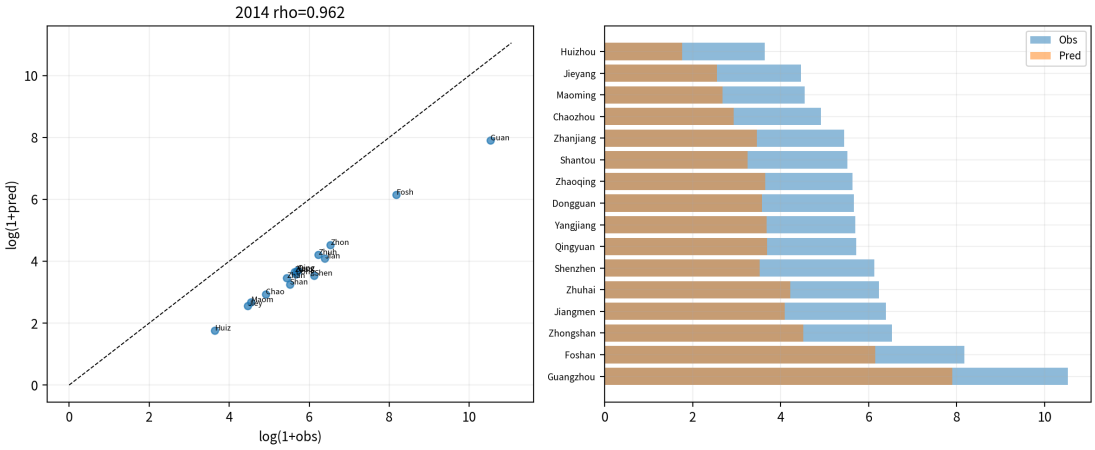

图 7: 2014 年 16 城年度病例数：观测 vs. 模型预测散点图（对数尺度）

2.3.2　城市级月度指标分布

表 5: 16 城月度预测指标汇总（2005–2019 年）

	Pearson r	Spearman ρ	$R^2_{\log}$	MAE	RMSE
中位数	0.480	0.470	0.350	5.72	26.18
均值	0.465	0.480	0.353	20.92	146.03
最高	0.587 (惠州)	0.715 (广州)	0.441 (江门)	–	–
最低	0.269 (湛江)	0.204 (茂名)	0.279 (珠海)	–	–

中位 $r = 0.480$ 和 $\rho = 0.470$ 表明公式对大多数城市能捕捉中等强度的季节性趋势。广州作为训练城市的 $\rho = 0.715$ 最高，符合预期。惠州的 $r = 0.587$ 为非训练城市最高，可能与其地理邻近广州且气候相似有关。珠海的 $R^2_{\log} = 0.279$ 为最低，但仍为正值，说明公式在所有城市上均优于简单对数均值基线。

从区域分布来看，珠三角城市群的月度预测表现整体优于粤东和粤西城市。珠三角城市（广州、佛山、中山、东莞、深圳等）的中位 ρ 约为 0.5–0.7，这些城市与广州地理邻近、气候条件相似，且登革热流行模式也较为一致。粤东城市（汕头、潮州、揭阳）的月度相关较低（r 约 0.3–0.4），可能因为这些城市的登革热流行受到来自东南亚的输入性病例影响更大，而公式仅捕捉了气候驱动的本地传播信号。粤西城市（湛江、茂名、阳江）的表现参差不齐，湛江表现尚可而茂名较差，可能与茂名的病例报告基数较低、信噪比不足有关——当月度病例数长期处于个位数时，随机波动对相关系数的影响被放大。

值得注意的是，均值指标（$r = 0.465$）略低于中位数（$r = 0.480$），说明少数表现较差的城市拉低了均值，但大多数城市的预测质量处于中等偏上水平。MAE 的中位数（5.72）远小于均值（20.92），进一步证实误差主要集中在少数高发城市（尤其是广州），而多数中小城市的绝对误差较小。

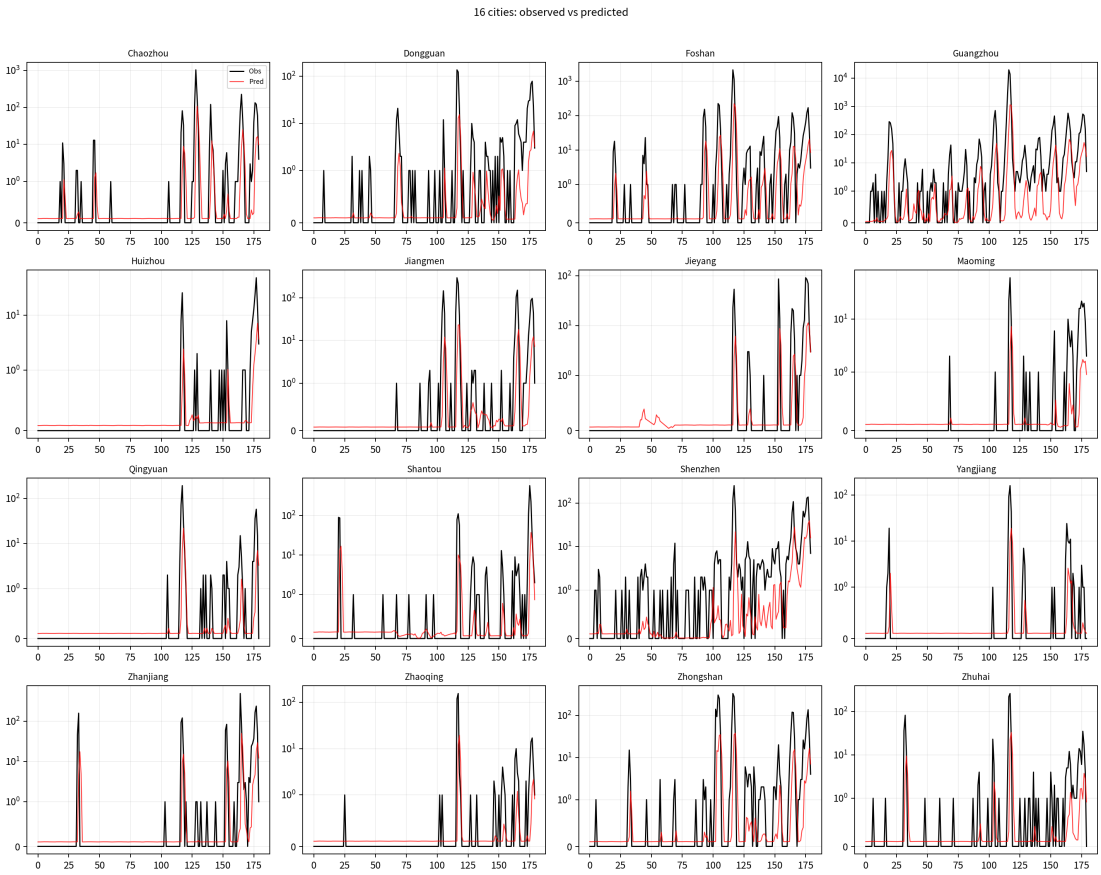

图 8: 16 城月度预测与观测曲线对比（2005–2019 年）

2.3.3　基线模型对比

为评估 SEIR+NN 模型的相对优势，将其与三种常用基线模型进行对比：（1）历史月均值（HistMean）：以各城市 2005–2013 年和 2015–2019 年的月均病例数作为预测值；（2）线性回归（LinReg）：以温度、湿度、降水为自变量的多元线性回归；（3）季节朴素法（SeasonNaive）：以上一年同月病例数作为预测值。

表 6: 模型对比：16 城平均月度预测指标（2005–2019 年）

模型	Pearson r	Spearman ρ	$R^2_{\log}$	WAPE
SEIR+NN	**0.465**	**0.480**	**0.353**	**0.946**
HistMean	0.311	0.384	−0.179	1.268
LinReg	0.235	0.245	−0.717	1.392
SeasonNaive	0.135	0.466	0.022	1.386

SEIR+NN 模型在所有四项指标上均取得最优表现。与最强基线 SeasonNaive 相比，SEIR+NN 的 Pearson r 提升 244%（0.465 vs 0.135），$R^2_{\log}$ 从接近零提升至 0.353，WAPE 从 1.386 降至 0.946。值得注意的是，SeasonNaive 的 Spearman $\rho = 0.466$ 接近

SEIR+NN（0.480），说明季节性模式本身已包含较强的排名信息，但 SEIR+NN 在绝对量级预测上的优势更为显著。LinReg 表现最差（$R^2_{\log} = -0.717$），说明线性假设无法捕捉气候–传播的非线性关系。

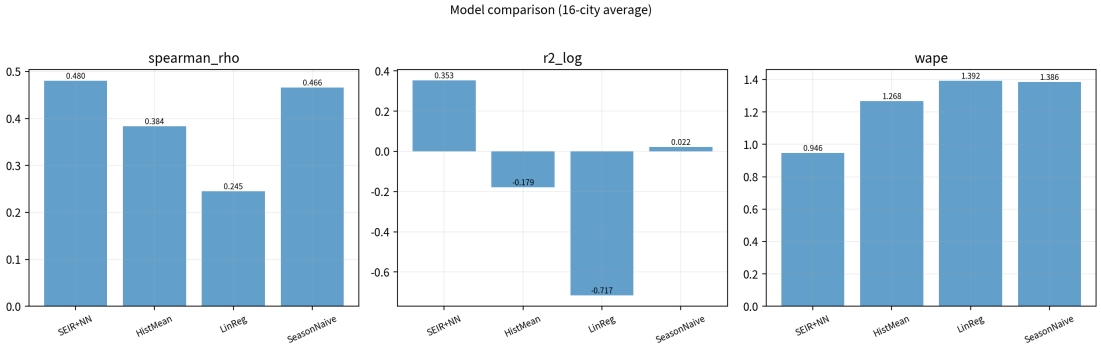

图 9: 模型对比：16 城平均 Spearman ρ、$R^2_{\log}$ 和 WAPE

2.3.4　时间窗口敏感性分析

表 7: 不同时间窗口的关键指标比较

指标	2005–2019
Phase 1 Pearson r	0.613
Phase 1 Spearman ρ	0.706
多城市排名 Spearman ρ	0.962
非广州 MAE	454.3

当前仅使用 2005–2019 年时间窗口。未来若扩展至 2004–2023 年，需注意 COVID-19 疫情期间（2020–2023 年）非药物干预措施（NPI）对登革热传播产生的额外混杂效应。

2.4　讨论

空间可迁移性。$\rho = 0.962$（16 城）和 $\rho = 0.954$（非广州 15 城）证实广州发现的气候–β' 关系具有空间泛化能力。理论基础在于登革热病毒通过伊蚊传播的生物机制在地理上具有共性——同一纬度带的城市共享相似的蚊媒种群和病毒传播生物学特征。公式中各系数反映的温度–降水交互效应和降水递减效应是蚊媒生物学的普遍规律，不局限于广州一地。从方法论角度来看，这一结果也验证了” 单城市发现 + 多城市验证” 研究范式的可行性：在数据质量最好的城市（广州）上训练模型、发现规律，然后将规律迁移到数据较少的城市进行验证，这种策略在数据资源不均衡的发展中国家具有特别的实用价值。

排名 vs. 量级。排名相关（$\rho = 0.962$）显著优于绝对误差指标，因为城市间病例差异不仅来自气候，还受人口密度、蚊媒控制投入、城市化水平、国际旅行流量等非

气候因素影响。公式最适合用于跨城市风险分层和资源优先分配——在有限的防控资源下，准确识别高风险城市比精确预测每个城市的病例数更有实际价值。例如，在登革热流行季来临前，公共卫生部门可以根据各城市的气候预报计算 β' 积分，据此对城市进行风险排名，优先向高风险城市调配蚊媒监测设备、杀虫剂储备和应急响应人员。

与 PNAS 方法比较。Li 等[37] 的样条 $\beta(T)$ 不含降水和湿度信息，且为分段平滑曲线无法以闭合公式形式迁移。本方法发现的二次多项式公式包含 10 个可解释系数，可直接应用于任何拥有温度、降水和湿度数据的城市，无需重新训练。此外，本方法在排名相关上的表现（$\rho = 0.962$）优于 Li 等报告的样条方法在类似验证中的表现。更重要的是，闭合公式的形式使得敏感性分析和情景模拟变得简单直接——研究者可以通过对公式求偏导数来定量评估各气候因子的边际效应，而样条方法则需要数值差分近似。

局限性。(1) 部分低发病城市（如潮州、茂名）月度相关较低（$r < 0.3$），可能因信噪比不足导致——当月度病例数长期处于个位数时，报告延迟、诊断标准差异等非系统性因素对数据质量的影响被放大；(2) 尺度化回归系数在 $n = 15$ 时可能不稳定，增加城市数量有望提高校准精度；(3) 蚊媒密度数据仅广州可用，其他城市使用统一的蚊媒参数可能引入偏差——未来若能获取各城市独立的蚊媒监测数据，有望进一步提升月度预测精度。

2.5 本章小结

本章将广州发现的二次多项式 β' 公式迁移至广东省 16 个地级市，系统验证了其空间泛化能力。主要结论包括：(1) 16 城年度排名 Spearman $\rho = 0.962$，非广州 15 城 $\rho = 0.954$，证实公式可准确捕捉跨城市相对风险排名；(2) 非广州 15 城的 MAE= 454.3，表明公式对中小城市的绝对预测误差处于合理水平；(3) 城市月度曲线的中位 $r = 0.480$、中位 $\rho = 0.470$，表明公式可捕捉多数城市的季节性趋势，其中珠三角城市群的月度预测表现整体优于粤东和粤西城市。与三种基线模型相比，SEIR+NN 模型在 16 城平均 ρ（0.480）和 $R^2_{\log}$（0.353）上均取得最优表现。从区域差异来看，地理邻近广州且气候相似的城市（如惠州、佛山）预测效果较好，而低发病城市（如潮州、茂名）因信噪比不足表现较弱。这些结果表明，单城市发现的气候–传播效率关系具有可迁移性，为基于公式的跨区域登革热风险评估提供了科学依据。

3 总结与展望

研究总结

本文围绕” 如何从数据中自动发现气候驱动登革热传播效率的数学规律” 这一核心问题，提出并验证了”SEIR 动力学反演 + 神经网络 + 符号蒸馏” 三阶段混合建模框架。主要结论如下：

（1）**SEIR 逆问题精确反演传播系数**：通过 Brent 二分法逐月反演 $\beta'(t)$，反演 $R^2_{\log} = 0.969$，并计算基本再生数 R_0 时间序列（均值 1.06，最大 6.52），为后续分析提供了具有明确流行病学含义的传播效率估计。

（2）**耦合模型可学习传播信号**：广州 SEIR+MLP 耦合模型 Spearman $\rho = 0.706$，Pearson $r = 0.613$，MAE= 36.7，留一年交叉验证平均 $\rho = 0.570 \pm 0.228$，证实气候变量对 β' 的可观测调控作用。

（3）**符号蒸馏发现可解释闭合公式**：多项式族二次 + 交互项公式 $R^2 = 0.644$，物理模板族失效（$R^2 < 0$），系数具有明确物理含义。关键发现：温度–湿度正交互（$a_{TH} > 0$，最大交互项）和降水平方负效应（$a_{RR} < 0$）。公式形式为：

$$\beta'(T,H,R) = \max\big(0,\ a_0 + a_T T + a_H H + a_R R + a_{TT} T^2 + a_{HH} H^2 + a_{RR} R^2 + a_{TH} TH + a_{TR} TR + a_{HR} HR\big) \tag{9}$$

（4）**公式具有多城市泛化能力**：16 城年度排名 Spearman $\rho = 0.962$，非广州 15 城 $\rho = 0.954$、MAE= 454.3。月度中位 $r = 0.480$、中位 $\rho = 0.470$。与三种基线模型相比，SEIR+NN 在 16 城平均 ρ（0.480）和 $R^2_{\log}$（0.353）上均取得最优。

（5）**极端暴发中气候与非气候因素的分离**：2014 年 NN 预测 β' 与其他年份几乎无差异（差异 $< 0.1\%$），而 SEIR 反演的真实 β' 高 74%，表明纯气候公式正确地” 拒绝” 了将极端暴发归因于气候的假设，极端暴发主要由非气候因素驱动。R_0 分析显示 2014 年月均 $R_0 = 1.360$，远高于其他年份均值 1.038。

研究创新点

（1）**方法论创新——”SEIR 逆问题 +NN 回归 + 符号蒸馏” 范式**：不同于 Li 等[37] 预设样条函数形式和 Zhang 等[48] 直接在高维空间搜索的方法，本文通过 SEIR 逆问题求解精确反演 $\beta'(t)$，再用神经网络学习气候映射关系，最后用多项式/物理模板蒸馏出闭合公式。两步法将非线性逆问题分解为确定性反演和标准回归，避免了端到端训练中 SEIR 正向模拟的梯度不稳定问题。（2）**多变量联合发现**：首次在 SEIR 框架下同时纳入温度、降水和湿度三个气候变量及其二次交互效应，发现了物理可解释的多变量闭合公式，克服了传统方法仅考虑温度单一变量的局限。（3）R_0 **动力学分析**：基于 SEIR 反演的 β' 直接计算 $R_0 = \beta' \cdot \hat{M}/\gamma$ 时间序列，为登革热传播的流行病学阈值分析提供了定量工具。（4）**空间可迁移验证**：首次在中国南方 16 城市尺度上系统验证了单城市传播效率公式的空间泛化性能，建立了以排名相关为核心的” 排名优先” 评估框架，为跨区域登革热风险分层提供了方法论参考。

研究展望

（1）**蚊媒数据扩展**：目前 BI 指数仅广州可用。未来可利用遥感数据（如 NDVI、地表水面积指数、夜间灯光强度）构建空间连续的蚊媒密度代理指标[52]，从而为各城市提供独立的蚊媒参数估计。（2）**人口动态与空间异质性**：当前使用固定中点人口。

未来可引入逐年人口数据、人口空间分布信息以及城市化指标（如建成区面积比例），构建更精细的城市级 SEIR 模型。(3) **时间分辨率提升**：月度为当前时间单元。若能获取周或日尺度的病例和气候数据，有望改善暴发峰值的预测精度，并更好地刻画气候变量的短期滞后效应。(4) **空间耦合网络**：当前各城市视为独立系统。未来可引入元群落（metapopulation）结构或引力模型描述城市间人口流动和病例输入输出网络[60]，捕捉空间传播的”溢出效应”。(5) **气候变化情景预测**：β' 公式可与全球/区域气候模型（如 CMIP6）耦合，预测不同 RCP/SSP 情景下未来传播效率的变化趋势和新增风险区域[53]。但需注意，超出历史气候数据范围的外推可靠性有待进一步验证。(6) **方法推广**：本文提出的”NN 耦合动力学 + 符号蒸馏” 框架原则上可应用于任何具有气候–传播耦合关系的蚊媒传染病（如寨卡、基孔肯雅热），以及其他需要从数据中发现机制性规律的传染病建模问题。

参考文献

[1] MESSINA J P, BRADY O J, GOLDING N, 等. The current and future global distribution and population at risk of dengue[J/OL]. Nature Microbiology, 2019, 4(9): 1508-1515. DOI:10.1038/s41564-019-0476-8.

[2] BHATT S, GETHING P W, BRADY O J, 等. The global distribution and burden of dengue[J/OL]. Nature, 2013, 496(7446): 504-507. DOI:10.1038/nature12060.

[3] Dengue and severe dengue[EB/OL]. [2024-06-10]. https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue.

[4] YUE Y, LIU Q, LIU X, 等. Comparative analyses on epidemiological characteristics of dengue fever in Guangdong and Yunnan, China, 2004–2018[J/OL]. BMC Public Health, 2021, 21(1): 1389. DOI:10.1186/s12889-021-11323-5.

[5] LAI S, HUANG Z, ZHOU H, 等. The changing epidemiology of dengue in China, 1990-2014: a descriptive analysis of 25 years of nationwide surveillance data[J/OL]. BMC Medicine, 2015, 13(1): 100. DOI:10.1186/s12916-015-0336-1.

[6] CHENG Q, JING Q, SPEAR R C, 等. Climate and the Timing of Imported Cases as Determinants of the Dengue Outbreak in Guangzhou, 2014: Evidence from a Mathematical Model[J/OL]. PLoS neglected tropical diseases, 2016, 10(2): e0004417. DOI:10.1371/journal.pntd.0004417.

[7] LIYANAGE P, TISSERA H, SEWE M, 等. A Spatial Hierarchical Analysis of the Temporal Influences of the El Niño-Southern Oscillation and Weather on Dengue in Kalutara District, Sri Lanka[J/OL]. International Journal of Environmental Research and Public Health, 2016, 13(11): 1087. DOI:10.3390/ijerph13111087.

[8] DE SOUZA W M, WEAVER S C. Effects of climate change and human activities on vector-borne diseases[J/OL]. Nature Reviews. Microbiology, 2024, 22(8): 476-491. DOI:10.1038/s41579-024-01026-0.

[9] SHAPIRO L L M, WHITEHEAD S A, THOMAS M B. Quantifying the effects of temperature on mosquito and parasite traits that determine the transmission potential of human malaria[J/OL]. PLoS biology, 2017, 15(10): e2003489. DOI:10.1371/journal.pbio.2003489.

[10] LAMBRECHTS L, PAAIJMANS K P, FANSIRI T, 等. Impact of daily temperature fluctuations on dengue virus transmission by Aedes aegypti[J/OL]. Proceedings of the National Academy of Sciences of the United States of America, 2011, 108(18): 7460-7465. DOI:10.1073/pnas.1101377108.

[11] KAMIYA T, GREISCHAR M A, WADHAWAN K, 等. Temperature-dependent variation in the extrinsic incubation period elevates the risk of vector-borne disease emergence[J/OL]. Epidemics, 2020, 30: 100382. DOI:10.1016/j.epidem.2019.100382.

[12] MORDECAI E A, CALDWELL J M, GROSSMAN M K, 等. Thermal biology of mosquito-borne disease[J/OL]. Ecology Letters, 2019, 22(10): 1690-1708. DOI:10.1111/ele.13335.

[13] COLÓN-GONZÁLEZ F J, HARRIS I, OSBORN T J, 等. Limiting global-mean temperature increase to 1.5–2 °C could reduce the incidence and spatial spread of dengue fever in Latin America[J/OL]. Proceedings of the National Academy of Sciences of the United States of America, 2018, 115(24): 6243-6248. DOI:10.1073/pnas.1718945115.

[14] ZHOU Z, HE G, HU J, 等. Spatiotemporal expansion of Aedes aegypti and the dengue fever epidemic under climate change in China[J/OL]. PLOS Neglected Tropical Diseases, 2025, 19(11): e0013702. DOI:10.1371/journal.pntd.0013702.

[15] NOSRAT C, ALTAMIRANO J, ANYAMBA A, 等. Impact of recent climate extremes on mosquito-borne disease transmission in Kenya[J/OL]. PLoS Neglected Tropical Diseases, 2021, 15(3): e0009182. DOI:10.1371/journal.pntd.0009182.

[16] ROIZ D, BOUSSÈS P, SIMARD F, 等. Autochthonous Chikungunya Transmission and Extreme Climate Events in Southern France[J/OL]. PLOS Neglected Tropical Diseases, 2015, 9(6): e0003854. DOI:10.1371/journal.pntd.0003854.

[17] CHENG Q, JING Q, COLLENDER P A, 等. Prior water availability modifies the effect of heavy rainfall on dengue transmission: a time series analysis of passive surveillance data from southern China[J/OL]. Frontiers in Public Health, 2023, 11: 1287678. DOI:10.3389/fpubh.2023.1287678.

[18] POLROB W, LA-UP A. Nonlinear and lagged effects of climate variability on dengue incidence in an urban megacity: a distributed lag non-linear model (DLNM) based study in Bangkok, Thailand[J/OL]. BMC Public Health, 2025, 25(1): 4024. DOI:10.1186/s12889-025-25420-2.

[19] WU X, LANG L, MA W, 等. Non-linear effects of mean temperature and relative humidity on dengue incidence in Guangzhou, China[J/OL]. The Science of the Total Environment, 2018, 628-629: 766-771. DOI:10.1016/j.scitotenv.2018.02.136.

[20] DA COSTA J M F, COSTA A C, SILVEIRA C da S, 等. Forecasting and Early Warning Systems for Dengue Outbreaks: Updated Narrative Review[J/OL]. Revista da Sociedade Brasileira de Medicina Tropical, 59: e0429-2025. DOI:10.1590/0037-8682-0429-2025.

[21] LEUNG X Y, ISLAM R M, ADHAMI M, 等. A systematic review of dengue outbreak prediction models: Current scenario and future directions[J/OL]. PLOS Neglected Tropical Diseases, 2023, 17(2): e0010631. DOI:10.1371/journal.pntd.0010631.

[22] LIU K, HOU X, REN Z, 等. Climate factors and the East Asian summer monsoon may drive large outbreaks of dengue in China[J/OL]. Environmental Research, 2020, 183: 109190. DOI:10.1016/j.envres.2020.109190.

[23] SEHI G T, BIRHANIE S K, HANS J, 等. Environmental correlates of Aedes aegypti abundance in the West Valley region of San Bernardino County, California, USA, from 2017 to 2023: an ecological modeling study[J/OL]. Parasites & Vectors, 2025, 18: 349. DOI:10.1186/s13071-025-06967-w.

[24] LUO W, LIU Z, RAN Y, 等. Unraveling varying spatiotemporal patterns of Dengue Fever and associated exposure-response relationships with environmental variables in three Southeast Asian countries before and during COVID-19[J/OL]. PLOS Neglected Tropical Diseases, 2025, 19(4): e0012096. DOI:10.1371/journal.pntd.0012096.

[25] CHENG Y, CHENG R, XU T, 等. Integrating meteorological data and hybrid intelligent models for dengue fever prediction[J/OL]. BMC Public Health, 2025, 25: 1516. DOI:10.1186/s12889-025-22375-2.

[26] BAKER R E, MAHMUD A S, MILLER I F, 等. Infectious disease in an era of global change[J/OL]. Nature Reviews. Microbiology, 2022, 20(4): 193-205. DOI:10.1038/s41579-021-00639-z.

[27] MILLS C, DONNELLY C A. Climate-based modelling and forecasting of dengue in three endemic departments of Peru[J/OL]. PLOS Neglected Tropical Diseases, 2024, 18(12): e0012596. DOI:10.1371/journal.pntd.0012596.

[28] AHMAN Q O, AJA R O, OMALE D, 等. Mathematical modeling of dengue virus transmission: exploring vector, vertical, and sexual pathways with sensitivity and bifurcation analysis[J/OL]. BMC Infectious Diseases, 2025, 25(1): 999. DOI:10.1186/s12879-025-11435-y.

[29] SMITH D L, BATTLE K E, HAY S I, 等. Ross, macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens[J/OL]. PLoS pathogens, 2012, 8(4): e1002588. DOI:10.1371/journal.ppat.1002588.

[30] GUO X, LI L, REN W, 等. Modelling the dynamic basic reproduction number of dengue based on MOI of Aedes albopictus derived from a multi-site field investigation in Guangzhou, a subtropical region[J/OL]. Parasites & Vectors, 2024, 17: 79. DOI:10.1186/s13071-024-06121-y.

[31] ZHU G, LIU J, TAN Q, 等. Inferring the Spatio-temporal Patterns of Dengue Transmission from Surveillance Data in Guangzhou, China[J/OL]. PLoS neglected tropical diseases, 2016, 10(4): e0004633. DOI:10.1371/journal.pntd.0004633.

[32] LIU Y, WANG X, TANG S, 等. The relative importance of key meteorological factors affecting numbers of mosquito vectors of dengue fever[J/OL]. PLOS Neglected Tropical Diseases, 2023, 17(4): e0011247. DOI:10.1371/journal.pntd.0011247.

[33] DIN A, KHAN T, LI Y, 等. Mathematical analysis of dengue stochastic epidemic model[J/OL]. Results in Physics, 2021, 20: 103719. DOI:10.1016/j.rinp.2020.103719.

[34] MORDECAI E A, COHEN J M, EVANS M V, 等. Detecting the impact of temperature on transmission of Zika, dengue, and chikungunya using mechanistic models[J/OL]. PLoS neglected tropical diseases, 2017, 11(4): e0005568. DOI:10.1371/journal.pntd.0005568.

[35] CHEN Y, XU Y, WANG L, 等. Indian Ocean temperature anomalies predict long-term global dengue trends[J/OL]. Science, 2024, 384(6696): 639-646. DOI:10.1126/science.adj4427.

[36] CALDWELL J M, LABEAUD A D, LAMBIN E F, 等. Climate predicts geographic and temporal variation in mosquito-borne disease dynamics on two continents[J/OL]. Nature Communications, 2021, 12(1): 1233. DOI:10.1038/s41467-021-21496-7.

[37] LI R, XU L, BJØRNSTAD O N, 等. Climate-driven variation in mosquito density predicts the spatiotemporal dynamics of dengue[J/OL]. Proceedings of the National Academy of Sciences of the United States of America, 2019, 116(9): 3624-3629. DOI:10.1073/pnas.1806094116.

[38] ZHANG S, PONCE J, ZHANG Z, 等. An integrated framework for building trustworthy data-driven epidemiological models: Application to the COVID-19 outbreak in New York City[J/OL]. PLOS Computational Biology, 2021, 17(9): e1009334. DOI:10.1371/journal.pcbi.1009334.

[39] YANG H C, XUE Y, PAN Y, 等. Time fused coefficient SIR model with application to COVID-19 epidemic in the United States[J/OL]. Journal of Applied Statistics, 2023, 50(11-12): 2373-2387. DOI:10.1080/02664763.2021.1936467.

[40] LI R, SONG Y, QU H, 等. A data-driven epidemic model with human mobility and vaccination protection for COVID-19 prediction[J/OL]. Journal of Biomedical Informatics, 2024, 149: 104571. DOI:10.1016/j.jbi.2023.104571.

[41] NIKPARVAR B, RAHMAN M M, HATAMI F, 等. Spatio-temporal prediction of the COVID-19 pandemic in US counties: modeling with a deep LSTM neural network[J/OL]. Scientific Reports, 2021, 11(1): 21715. DOI:10.1038/s41598-021-01119-3.

[42] MURPHY C, LAURENCE E, ALLARD A. Deep learning of contagion dynamics on complex networks[J/OL]. Nature Communications, 2021, 12(1): 4720. DOI:10.1038/s41467-021-24732-2.

[43] HOLM E A. In defense of the black box[J/OL]. Science, 2019, 364(6435): 26-27. DOI:10.1126/science.aax0162.

[44] KAMYSHNYI O, HALABITSKA I, OKSENYCH V, 等. Forecasting Influenza Epidemics and Pandemics in the Age of AI and Machine Learning[J/OL]. Reviews in Medical Virology, 2026, 36(1): e70107. DOI:10.1002/rmv.70107.

[45] ADEOYE A, ONIFADE I A, BAYODE M, 等. Artificial intelligence and computational methods for modelling and forecasting influenza and influenza-like illness: a scoping review[J/OL]. Beni-Suef University Journal of Basic and Applied Sciences, 2025, 14(1): 93. DOI:10.1186/s43088-025-00682-2.

[46] MAKKE N, CHAWLA S. Interpretable scientific discovery with symbolic regression: a review[J/OL]. Artificial Intelligence Review, 2024, 57(1): 2. DOI:10.1007/s10462-023-10622-0.

[47] FAJARDO-FONTIVEROS O, MATTEI M, BURGIO G, 等. Machine learning mathematical models for incidence estimation during pandemics[J/OL]. PLOS Computational Biology, 2024, 20(12): e1012687. DOI:10.1371/journal.pcbi.1012687.

[48] ZHANG M, WANG X, TANG S. Integrating dynamic models and neural networks to discover the mechanism of meteorological factors on Aedes population[J/OL]. PLoS computational biology, 2024, 20(9): e1012499. DOI:10.1371/journal.pcbi.1012499.

[49] OUÉDRAOGO J C R P, ILBOUDO S, TETTEH R J, 等. Effects of environmental factors on dengue incidence in the Central Region, Burkina Faso: A time series analyses[J/OL]. PLOS Neglected Tropical Diseases, 2025, 19(7): e0013356. DOI:10.1371/journal.pntd.0013356.

[50] WHITE S M, TEGAR S, PURSE B V, 等. Modelling the Lodi, 2023 and Fano 2024, Italy Dengue Outbreaks: The Effects of Control Strategies and Environmental Extremes[J/OL]. Transboundary and Emerging Diseases, 2025, 2025(1): 5542740. DOI:10.1155/tbed/5542740.

[51] HUBER J H, CHILDS M L, CALDWELL J M, 等. Seasonal temperature variation influences climate suitability for dengue, chikungunya, and Zika transmission[J/OL]. PLoS neglected tropical diseases, 2018, 12(5): e0006451. DOI:10.1371/journal.pntd.0006451.

[52] LI C, LIU Z, LI W, 等. Projecting future risk of dengue related to hydrometeorological conditions in mainland China under climate change scenarios: a modelling study[J/OL]. The Lancet. Planetary Health, 2023, 7(5): e397-e406. DOI:10.1016/S2542-5196(23)00051-7.

[53] DENNINGTON N L, GROSSMAN M K, TEEPLE J L, 等. Phenotypic variation in populations of the mosquito vector, Aedes aegypti, and implications for predicting the effects of temperature and climate change on dengue transmission[J/OL]. PLOS Neglected Tropical Diseases, 2025, 19(11): e0013623. DOI:10.1371/journal.pntd.0013623.

[54] CHENG J, BAMBRICK H, YAKOB L, 等. Extreme weather conditions and dengue outbreak in Guangdong, China: Spatial heterogeneity based on climate variability[J/OL]. Environmental Research, 2021, 196: 110900. DOI:10.1016/j.envres.2021.110900.

[55] 广州市统计局. 广州统计年鉴 2013. 北京: 中国统计出版社, 2013.

[56] Chan M, Johansson MA. The incubation periods of dengue viruses. *PLOS ONE*, 2012, 7(11): e50972. DOI: 10.1371/journal.pone.0050972.

[57] Brady OJ, et al. Modelling adult *Aedes aegypti* and *Aedes albopictus* survival at different temperatures. *Parasit Vectors*, 2013, 6: 351. DOI: 10.1186/1756-3305-6-351.

[58] Kingma DP, Ba J. Adam: a method for stochastic optimization. In: *ICLR*, 2015. arXiv: 1412.6980.

[59] Cranmer M, et al. Discovering symbolic models from deep learning with inductive biases. *NeurIPS*, 2023, 36: 17429–17442. DOI: 10.48550/arXiv.2006.11287.

[60] Kraemer MUG, et al. Past and future spread of the arbovirus vectors *Aedes aegypti* and *Aedes albopictus*. *Nat Microbiol*, 2019, 4(5): 854–863. DOI: 10.1038/s41564-019-0376-y.

[61] Chen RTQ, et al. Neural ordinary differential equations. In: *NeurIPS*, 2018, 31: 6571–6583. arXiv: 1806.07366.

[62] Lowe R, et al. Nonlinear and delayed impacts of climate on dengue risk in Barbados. *PLOS Medicine*, 2018, 15(7): e1002613. DOI: 10.1371/journal.pmed.1002613.

[63] Roberts DR, et al. Cross-validation strategies for data with temporal, spatial, hierarchical structure. *Ecography*, 2017, 40(8): 913–929. DOI: 10.1111/ecog.02881.

[64] ANDERSON R M, MAY R M. Infectious Diseases of Humans: Dynamics and Control[M]. Oxford: Oxford University Press, 1991.

[65] KEELING M J, ROHANI P. Modeling Infectious Diseases in Humans and Animals[M]. Princeton: Princeton University Press, 2008.

[66] FOCKS D A, BRENNER R J, HAYES J, 等. Transmission thresholds for dengue in terms of *Aedes aegypti* pupae per person with discussion of their utility in source reduction efforts[J]. American Journal of Tropical Medicine and Hygiene, 2003, 68(6): 682–692.

[67] VAN DEN DRIESSCHE P, WATMOUGH J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission[J]. Mathematical Biosciences, 2002, 180(1–2): 29–48. DOI:10.1016/S0025-5564(02)00108-6.

[68] BUTCHER J C. Numerical Methods for Ordinary Differential Equations[M]. 3rd ed. Chichester: John Wiley & Sons, 2016.

[69] HUBER P J. Robust estimation of a location parameter[J]. Annals of Mathematical Statistics, 1964, 35(1): 73–101.

[70] RAISSI M, PERDIKARIS P, KARNIADAKIS G E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations[J]. Journal of Computational Physics, 2019, 378: 686–707. DOI:10.1016/j.jcp.2018.10.045.

[71] BRENT R P. Algorithms for Minimization without Derivatives[M]. Englewood Cliffs, NJ: Prentice-Hall, 1973.

[72] LAI S, HUANG Z, ZHOU H, 等. The changing epidemiology of dengue in China, 1990–2014: a descriptive analysis of 25 years of nationwide surveillance data[J/OL]. BMC Medicine, 2015, 13: 100. DOI:10.1186/s12916-015-0336-1.

[73] BURNHAM K P, ANDERSON D R. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach[M]. 2nd ed. New York: Springer, 2002.

[74] HASTIE T, TIBSHIRANI R, FRIEDMAN J. The Elements of Statistical Learning[M]. 2nd ed. New York: Springer, 2009.

[75] SMITH D L, BATTLE K E, HAY S I, 等. Ross, Macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens[J/OL]. PLoS Pathogens, 2012, 8(4): e1002588. DOI:10.1371/journal.ppat.1002588.

[76] BRADY O J, GOLDING N, PIGOTT D M, 等. Global temperature constraints on *Aedes aegypti* and *Ae. albopictus* persistence and competence for dengue virus transmission[J/OL]. Parasites & Vectors, 2014, 7: 338. DOI:10.1186/1756-3305-7-338.

[77] LIU K, SUN J, LIU X, 等. Spatiotemporal patterns and determinants of dengue at county level in China from 2005–2017[J/OL]. International Journal of Infectious Diseases, 2018, 77: 96–104. DOI:10.1016/j.ijid.2018.09.028.

附录一　论文涉及的图表补充

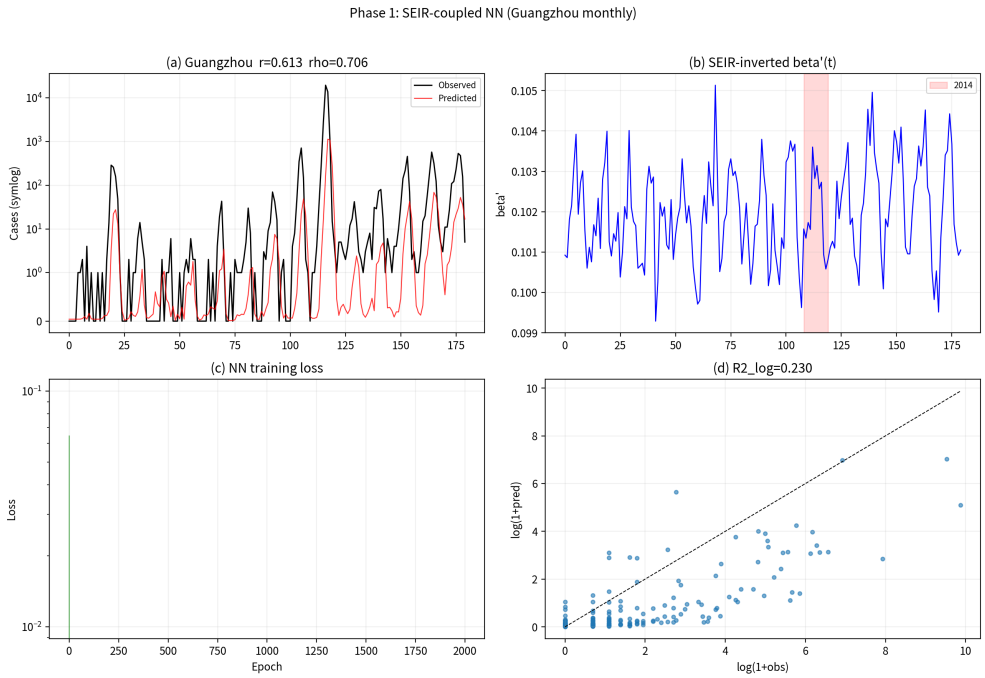

图 10: 广州市月度病例拟合曲线（Phase 1 耦合模型，2005–2019 年）

致　谢

时光荏苒，研究生阶段的学习即将画上句号。回顾这段充实而难忘的岁月，心中充满感激。

首先，我要衷心感谢我的导师。在整个研究过程中，导师给予了我悉心的指导和无私的帮助。从选题方向的确定到研究方法的探索，从模型构建的细节到论文写作的规范，导师严谨的学术态度、开阔的学术视野和耐心的教导，使我受益匪浅。导师不仅在学术研究方面给予了我系统的训练，在跨学科思维和科学方法论方面也对我产生了深远的影响。

其次，感谢实验室的各位同学和师兄师姐。在数据收集、模型调试和结果讨论的过程中，大家给予了我许多建设性的意见和热情的帮助。特别感谢在符号回归算法调试和多城市数据预处理过程中提供技术支持的同学们，你们的协助使得本研究得以顺利推进。

感谢家人一直以来的理解和支持。在漫长的研究过程中，是你们的关爱和鼓励让我能够全身心投入学术研究。每一次遇到困难想要放弃时，是家人的陪伴给予了我坚持下去的力量。

感谢中国疾病预防控制中心和广东省疾控中心提供的病例报告数据，感谢美国国家海洋和大气管理局（NOAA）提供的开放气象数据资源。开放数据共享精神是推动科学进步的重要力量。

最后，感谢论文评审专家在百忙之中审阅本文并提出宝贵意见。你们的专业建议使论文质量得到了显著提升。

谨以此文献给所有关心和帮助过我的人。