arXiv:2305.01770v1 [cs.LG] 2 May 2023

DeCom：具备非药物干预意识的后新冠肺炎时期呼吸道合胞病毒预测深度耦合张量分解机

Xinyan Li、Cheng Qian 和 Lucas Glass

IQVIA Inc. 卓越分析中心，美国

2023年5月4日

摘要

呼吸道合胞病毒（RSV）是对婴儿和幼儿而言最危险的呼吸道疾病之一。由于新冠肺炎暴发期间实施了非药物干预（NPI），北半球的呼吸道合胞病毒季节性传播模式于2020年中断，并在2021年提前数月出现转移。了解新冠肺炎如何影响RSV，并构建预测算法以预测后新冠肺炎时期RSV重新出现的时间和强度，至关重要。本文提出了一种用于后新冠肺炎时期RSV预测的深度耦合张量分解机，称为DeCom。DeCom结合了张量分解和残差建模，通过同时考虑常规的呼吸道合胞病毒季节性传播模式和NPI，使我们能够可靠地学习新冠肺炎期间受扰动的RSV传播规律。在真实RSV数据集上的实验结果表明，与当前最先进的RSV预测算法相比，DeCom的准确性更高；对于国家级预测，与基线方法相比，其均方根误差最高降低46%，平均绝对误差最高降低49%。

1 引言

呼吸道合胞病毒（RSV）是导致婴儿、幼儿和老年人患毛细支气管炎和肺炎的最常见原因 [Hall等, 2009]。根据美国疾病控制与预防中心（CDC）的数据，美国每年约有 58,000 名五岁以下儿童因感染 RSV 而住院 [CDC, 2022]。在全球范围内，这一数字每年可超过 300 万 [Shi等, 2017], 而实际负担可能要大得多，据估计，约有一半与 RSV 相关的死亡发生在医院之外 [Shi等, 2017]。

在2020年–21年季节之前，呼吸道合胞病毒（RSV）在温带和热带地区都呈现出稳定且可预测的季节性流行。这类流行通常于夏末在南北半球的热带地区开始，并于冬季和早春蔓延至大多数温带地点。例如，在美国，呼吸道合胞病毒（RSV）的活动通常于秋季从佛罗里达州开始，并逐渐由东南部向西北部扩散。然而，身体隔离、居家令、关闭学校等非药物干预（NPI）措施，

通讯作者：Xinyan Li（xinyan.li3@iqvia.com）。

旅行禁令和边境关闭已因当前新冠肺炎疫情而实施。尽管非药物干预（NPI）减少了新冠肺炎的传播，但也对呼吸道合胞病毒（RSV）的传播造成了重大影响。
自实施用于遏制新冠肺炎大流行的非药物干预以来，许多地区对呼吸道合胞病毒暴发的情况感到难以预测。例如，2020年全球呼吸道合胞病毒感染显著减少。即使在呼吸道合胞病毒流行季，感染病例数通常持续居高不下，但许多国家的病例数也接近于零。然而，2021年情况尚未恢复正常：大多数国家都经历了反季节的呼吸道合胞病毒暴发，且暴发时间比正常流行季提前了数月。例如，在北半球，法国、西班牙和美国在春季和夏季观察到呼吸道合胞病毒疫情高峰；南半球的澳大利亚和南非等国家也出现了异常的反季节疫情高峰。迄今为止，我们仍在应对异常的呼吸道合胞病毒流行。因此，了解不同因素与呼吸道合胞病毒再次流行的时间和强度之间的关联，对于后新冠肺炎时期的临床和公共卫生决策至关重要。

在考虑非药物干预的情况下预测新冠肺炎后的RSV传播是一项挑战。目前，针对这一任务的研究有限，且尚未取得令人满意的结果。在相关文献中，最相关的工作是 Baker 等人 [2020],，作者采用时间序列易感者–感染者–康复者（TSIR）模型来模拟 RSV 的未来发展轨迹。该方法基于标准 SIR 模型，并假设所有易感者最终都会感染。其局限在于，在实施 COVID-19非药物干预措施的情况下，难以估算易感人群数量和感染率，而且也无法对新冠肺炎对 RSV传播的影响进行建模。因此，TSIR 更像是一种分析工具，而非预测模型。
本文提出一种新的深度耦合张量分解方法 DeCom，用于在考虑非药物干预的情况下预测新冠肺炎后的 RSV。DeCom 由两个耦合的时空张量分解单元组成，使我们能够对 RSV 的规律性季节模式进行建模，并了解新冠肺炎引发的非药物干预措施如何影响 RSV传播。为展示 DeCom 的有效性和高准确率，我们使用真实的 RSV 数据集进行实验，并将结果与最先进的机器学习算法进行比较。

1.1 论文结构。

我们首先在第 2 节回顾相关工作。第 3 节简要介绍规范多元分解（CPD），这是我们提出的 DeCom 模型的关键组成部分。随后，在第 4 节中，我们对具体的呼吸道合胞病毒（RSV）预测问题进行形式化描述。第 5 节将深入介绍所提出的 DeCom 模型。接着，第 6 节描述数据和实验设置，第 7 节讨论实验结果，并将 DeCom 的预测能力与多个基准模型进行比较。最后，第 8 节总结全文。

2 相关工作

呼吸道合胞病毒预测与环境驱动因素。 Baker等 [2019]；Pitzer等 [2015] 表明，呼吸道合胞病毒流行呈现受气候驱动的季节性调节的极限环行为。在美国大多数地区，呼吸道合胞病毒和流感的发病率峰值出现在天气寒冷干燥的冬季月份[Baker等，2019年；Shaman等， 2010]。已有研究尝试将气候变量与呼吸道合胞病毒季节性联系起来 [Paiva等，2012年；Alonso等，2012年；du Prel等，2009年；Lapeña等，2005年；Noyola 和 Mandeville，2008年；Walton等，2010年；Yusuf等， 2007]。例如，Pitzer等 [2015] 采用线性

用于将发病模式与美国各州呼吸道合胞病毒流行病的候选环境驱动因素联系起来的模型。Walton 等人 [2010] 构建了朴素贝叶斯（NB）分类模型，用于预测犹他州盐湖县的呼吸道合胞病毒暴发。给定的NB模型使用了1985年到2008年的天气数据，仅考虑可实时获取的变量，并可提前最多3周进行预测，估计灵敏度最高可达67%，估计特异度最高可达94%至100%。然而， [White 等人，2005年，2007年；Leecaster 等人，2011年；Weber 等人， 2001] 发现呼吸道合胞病毒流行病表现出明显的非线性流行病动力学特征。很少有研究探讨不同地点、不同呼吸道合胞病毒季节性模式和气候类型中的气候关联 [Yusuf等人， 2007]。 [Altizer 等人， 2006] 改进了Pitzer 等人 [2015]的线性模型，以纳入感染动力学以及免疫力和易感性的波动，因为这些因素会影响环境因素与流行病发生时间之间的关系。

2.1 呼吸道合胞病毒的传播动力学。

数学模型已用于流行病学研究，以刻画呼吸道感染的年度流行病并估计干预策略可能产生的效果。过去，研究者利用常微分方程（ODE）构建了许多模型，用于描述呼吸道合胞病毒的传播动力学 [White等，2005年、2007年；Leecaster 等，2011年；Weber 等，2001年；Paynter 等，2014年；Arenas 等， 2009]。易感者-感染者-康复者（SIR）模型 [Kermack and McKendrick， 1927], 易感者-暴露者-感染者-康复者（SEIR）模型 [Cooke and Van Den Driessche， 1996]以及它们的变体，如 SIS、SIRS 和延迟 SIR 模型，只是此类模型中的少数几个例子。例如，Acedo 等 [2010] 引入了年龄结构化的 SIRS 模型，而 Moore 等 [2014]则开发了一种受季节性强迫的分舱式年龄结构化易感者-暴露者-传染者-康复者-易感者（SEIRS）模型，并使用 Nelder-Mead 方法拟合 RSV 阳性检测结果的季节性曲线。

最近，Baker等 [2019] 使用时间序列易感者-感染者-康复者（TSIR）模型来估计呼吸道合胞病毒的未来发展轨迹。其模型中的简单SIR组件假设每个易感病例只会感染一次。因此，这并不能反映实际情况：易感病例会先转变为暴露病例，随后一部分暴露病例才会转变为感染病例。

2.2 基于机器学习的呼吸道合胞病毒预测模型。

现代机器学习方法也已显示出益处，可作为数学模型的补充。Reis等 [2019]提出了一种超级集成方法，将三种预测方法相结合，用于预测美国不同空间分辨率下的呼吸道合胞病毒情况：全国、区域和州级。他们使用三种方法生成呼吸道合胞病毒暴发的回顾性预测，包括：1）贝叶斯加权暴发模型（BWO）；2）与动态易感者-感染者-康复者（SIR）模型相结合的集成调整卡尔曼滤波器（EAKF）；以及3）使用历史预期值的零模型 [Reis和Shaman， 2016]。Gebremedhin等 [2022] 拟合了多变量逻辑回归模型，以识别西澳大利亚州住院且接受呼吸道合胞病毒检测的五岁以下儿童中呼吸道合胞病毒阳性的预测因素（二元结局）。

3 预备知识

本节回顾规范多元分解（CPD），这是构建我们模型的关键步骤。本文使用的符号已汇总于表 1。

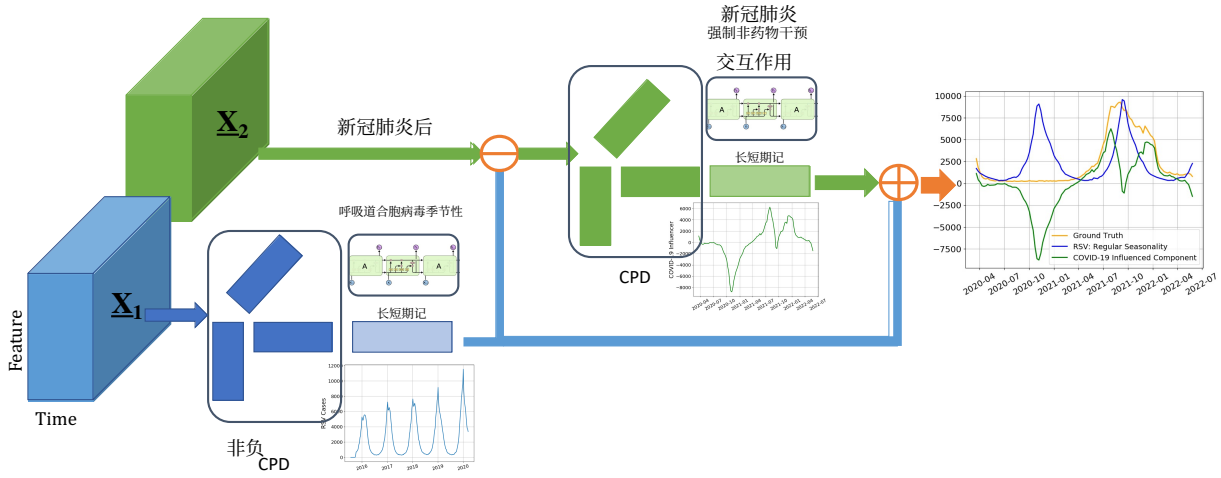


图 1：DeCom 模型的架构。

3.1 规范多元分解。

规范多元分解（CPD）用 K 个秩为 1 的分量之和近似一个三阶张量$\mathbf{X} := [\![\underline{\mathbf{A}}, \mathbf{B}, \mathbf{C}]\!] \in \mathbb{R}^{L\times M\times T}$，其中 K 为张量秩，即，

$$\underline{\mathbf{X}} \approx \sum_{k=1}^{K} \mathbf{a}_k \circ \mathbf{b}_k \circ \mathbf{c}_k, \tag{1}$$

其中，$\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_K] \in \mathbb{R}^{L\times K}$、$\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_K] \in \mathbb{R}^{M\times K}$、$\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_K] \in \mathbb{R}^{T\times K}$，且 $\circ$ 表示外积。

张量的 CPD 可以用多种不同方式表示。例如，我们可以将$\mathbf{X}^{(1)} = (\mathbf{C} \odot \mathbf{B})\mathbf{A}^T$展开，这可以看作是利用张量的模 1 纤维构造矩阵。利用“角色对称性”，模 2 和模 3 的矩阵展开分别为 $\mathbf{X}^{(2)} = (\mathbf{C} \odot \mathbf{A})\,\mathbf{B}^T$、$\mathbf{X}^{(3)} = (\mathbf{B} \odot \mathbf{A})\,\mathbf{C}^T$。CPD 仅需 $(L + M + T) \times K$ 个参数，即可简洁地表示大小为 $L \times M \times T$的张量。CPD 模型是数据分析中尤为强大的工具，这归因于两个关键性质：1）每个张量都存在有限秩的 CPD，表明该模型具有普适性；2）在适度条件下，CPD 还具有唯一性，即能够从合成的$\mathbf{X}$中提取真实的潜在因子。

4 问题形式化

考虑一个地点 $l \in L$，我们在此监测与 RSV 随时间演变相关的 M 特征/信号，例如新增 RSV 感染人数、温度、湿度、新增新冠肺炎感染人数等。在时间 $t \in T$，地点 l 处第 m 个信号的值记为 x_{lmt}。已知共有 L 个地点、M 个特征和 T 个时间步，该数据集可表示为一个三阶时空张量 $\mathbf{X} \in \mathbb{R}^{L\times M\times T}$，其中 $\mathbf{X}(l, m, t) := x_{lmt}$。从 $t = 1$ 到 $t = T$的所有特征和区域的演变由张量 $\mathbf{X} \in \mathbb{R}^{L\times\overline{M}\times T}$表示。目标是通过以下方式估计从时间 $t = T + 1$ 到 $t = T + T_0$ 的 RSV病例：

表 1：符号说明。

符号	描述
$\underline{\mathbf{X}} \in \mathbb{R}^{L\times M\times T}$	时空张量
$\mathbf{A} \in \mathbb{R}^{L\times K}$	位置因子矩阵
$\mathbf{B} \in \mathbb{R}^{M\times K}$	信号/特征因子矩阵
$\mathbf{C} \in \mathbb{R}^{T\times K}$	时间因子矩阵
$\underline{\mathbf{X}}^{(i)}$	$\mathbf{X}$ 的模-i展开
K	隐含成分数量
L	地点数量
M	特征数量
T	时间步数
$\circ$	外积
$\odot$	Khatri-Rao积

预测未来 T_0 个时间步的前向切片 $\mathbf{X}(:,:t)$，在我们的案例中， T_0 通常设为 52 周。

表2：RSV病例的州级RMSE和平均绝对误差性能比较。结果显示了3个不同的预测区间：最长12周、24周和52周。

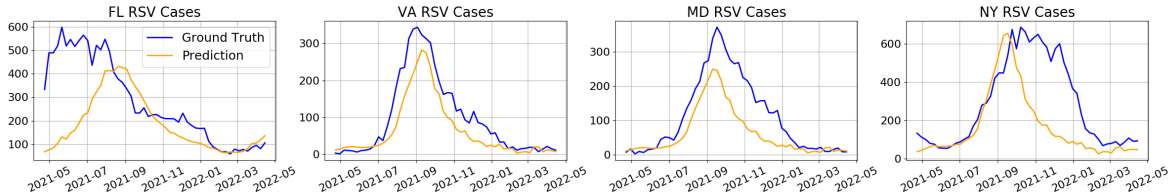
模型	均方根误差（12周）	平均绝对误差（12周）	均方根误差（24周）	平均绝对误差（24周）	均方根误差（52 周）	平均绝对误差（52 周）
DeCom	**12.63(4.86)**	**11.01(3.53)**	**41.26(39.18)**	**30.64(17.35)**	**86.75(56.71)**	**41.67(25.81)**
DeTensor	25.48(29.00)	16.19(26.15)	56.69(70.48)	32.07(49.11)	103.09(56.74)	55.83(34.41)
LSTM	42.72(108.57)	37.08(91.87)	54.53(62.40)	45.00(51.65)	99.19(122.49)	71.14(86.91)
ARIMA	18.47(24.81)	16.68(24.02)	78.37(92.44)	71.20(85.05)	100.11(115.40)	69.34(78.36)
SARIMA	15.91（14.26）	14.83（12.14）	56.24(68.17)	50.15(61.51)	101.76(120.21)	71.57(85.91)

5 所提方法

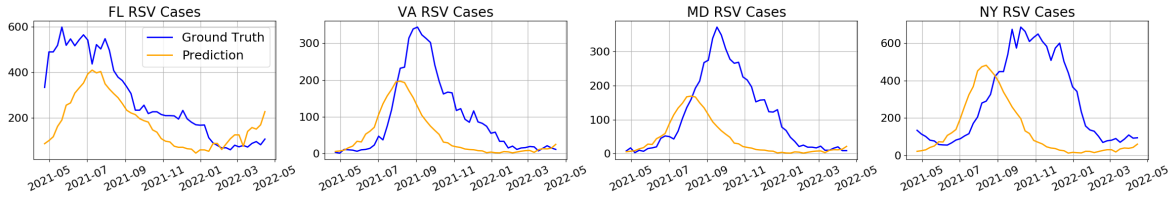
DeCom 是一种具有非药物干预感知能力的后新冠肺炎时期呼吸道合胞病毒预测模型。我们的研究动机源于这样一个事实：2020年实施国家特定的非药物干预措施后，呼吸道合胞病毒季节性传播模式随即中断。此后，随着新冠肺炎病例数降至某个阈值以下，各地重新开放，呼吸道合胞病毒再次出现，这标志着非药物干预控制期结束。因此，非药物干预控制对打破呼吸道合胞病毒的正常季节性有着重大影响。

图1展示了 DeCom 的总体示意图。如图所示，DeCom 需要两个时空张量作为输入，每个张量包含三个属性：位置 l、特征 m 和时间 t。第一个张量记为$\mathbf{X}_1 \in \mathbb{R}^{L\times M_1\times T_1}$，是新冠肺炎疫情前的张量，包含2020年3月15日之前出现的所有特征值；在此之前，呼吸道合胞病毒季节性传播尚未受到新冠肺炎的影响。第二个张量$\mathbf{X}_2 \in \mathbb{R}^{L\times M_2\times T-T_1}$表示新冠肺炎疫情后的张量，涵盖 T_1至 T之间的时间范围。

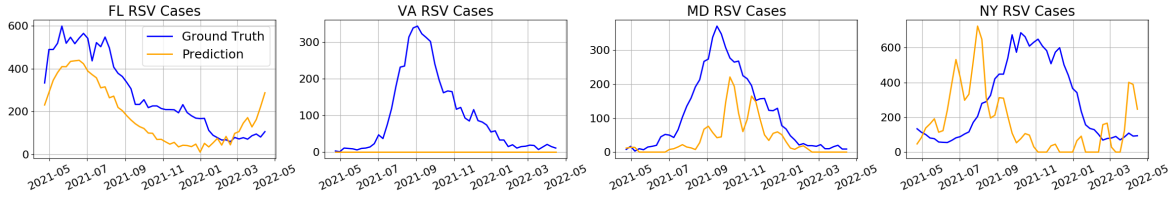
DeCom首先对$\mathbf{X}_1$ 进行非负 CPD，以学习$\mathbf{A}_1 \in \mathbb{R}^{L\times K_1}$、$\mathbf{B}_1 \in \mathbb{R}^{M_1\times K_1}$和$\mathbf{C}_1 \in \mathbb{R}^{T_1\times K_1}$ 这些子空间，它们分别表示典型的 COVID 前 RSV 流行季中位置、特征和时间的潜在子空间。其中， K_1 表示$\mathbf{X}_1$的秩。具体通过求解以下



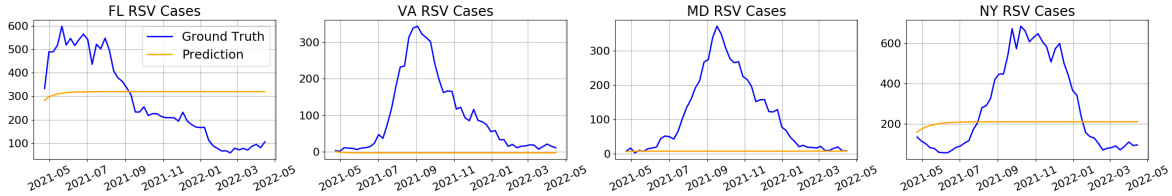
(a) DeCom



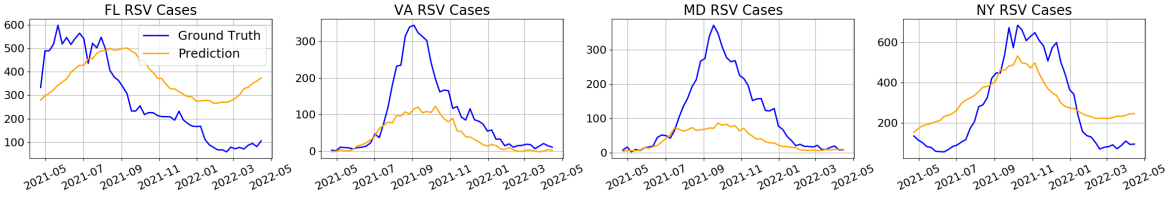
(b) DeTensor



(c) 长短期记忆网络



(d) ARIMA



(e) SARIMA

图 2：州级预测性能比较。

问题：

$$\min_{\mathbf{A}_1,\mathbf{B}_1,\mathbf{C}_1} ||\underline{\mathbf{X}}_1 - [\![\mathbf{A}_1, \mathbf{B}_1, \mathbf{C}_1]\!]||_F^2 \quad \text{s.t. } \mathbf{A}_1 \geq \mathbf{0},\ \mathbf{B}_1 \geq \mathbf{0},\ \mathbf{C}_1 \geq \mathbf{0}. \tag{2}$$

随后，采用第一个长短期记忆网络（LSTM） [Hochreiter and Schmidhuber, 1997] network$f_1(\cdot)$, 从时间子空间 $\mathbf{C}_1$, 中捕捉正常的呼吸道合胞病毒季节性，从而可以估计未来的变化

$\mathbf{C}_1$ 从 $t = T_1$ 到 $t = T$ 可估计为

$$\tilde{\mathbf{C}}_1 = f_1(\mathbf{C}_1) \tag{3}$$

˜遵循常规季节性（即没有新冠肺炎影响）时， $t = T_1$ 至 $t = T$ 期间的RSV病例$\mathbf{X}_1$ 可近似表示为

$$\tilde{\underline{\mathbf{X}}}_1 = [\![\mathbf{A}_1, \mathbf{B}_1, \tilde{\mathbf{C}}_1]\!] = \sum_{k=1}^{K_1} \mathbf{a}_{k,1} \circ \mathbf{b}_{k,1} \circ \tilde{\mathbf{c}}_{k,1} \tag{4}$$

˜其中， $\mathbf{a}_{k,1}$、$\mathbf{b}_{k,1}$ 和 ˜$\mathbf{c}_{k,1}$ 分别表示$\mathbf{A}_1, \mathbf{B}_1$ 和$\mathbf{C}_1$, 的第 k列。
COVID-19非药物干预措施对RSV传播的影响通常体现在RSV确诊病例数上。一种直接的方法是计算正常RSV流行季与受新冠肺炎影响的流行季之间的RSV病例数差异。请注意，在新冠肺炎疫情暴发期间，我们观测到的是受新冠肺炎影响的RSV病例数，而假设没有新冠肺炎时的病例数则只能通过公式（4）˜近似估算。因此，我们首先使用$\mathbf{X}_1$ 生成残差张量

$$\Delta\underline{\mathbf{X}}_2 = \underline{\mathbf{X}}_2 - \tilde{\underline{\mathbf{X}}}_1 \tag{5}$$

该残差张量近似表示COVID-19非药物干预措施对正常RSV传播造成的扰动。随后，我们对 $\Delta\mathbf{X}_2$ 执行第二次CPD，以学习残差子空间$\mathbf{A}_2 \in \mathbb{R}^{L\times K_2}$、$\mathbf{B}_2 \in \mathbb{R}^{M_2\times K_2}$ 和$\mathbf{C}_2 \in \mathbb{R}^{(T-T_1)\times K_2}$，其中$\mathbf{C}_2$ 捕捉COVID-19非药物干预措施如何在潜在空间中影响RSV传播， K_2 是 $\Delta\mathbf{X}_2$的张量秩。我们在$\mathbf{C}_2$ 上训练第二个长短期记忆网络$f_2(\cdot)$，以学习潜在空间$\mathbf{C}_2$中的残差传播模式。

由于$\mathbf{C}_2$ 的最后一行对应时间 T，因此将$\mathbf{C}_2$ 输入 $f_2(\cdot)$，即可预测从 $(T + 1)$ 到 $(T + T_0)$ 的残差传播模式，即

$$\tilde{\mathbf{C}}_2 = f_2(\mathbf{C}_2). \tag{6}$$

因此，从 $(T + 1)$ 到 $(T + T_0)$ 的残差张量计算如下

$$\Delta\tilde{\underline{\mathbf{X}}}_2 = [\![\mathbf{A}_2, \mathbf{B}_2, f_2(\mathbf{C}_2)]\!] = \sum_{k=1}^{K_2} \mathbf{a}_{k,2} \circ \mathbf{b}_{k,2} \circ \tilde{\mathbf{c}}_{k,2}. \tag{7}$$

由 (4) 和 (5) 可得，从 $(T + 1)$ 到 $(T + T_0)$ 的 RSV病例通过以下方式预测

$$\hat{\underline{\mathbf{X}}}_2 \approx [\![\mathbf{A}_1, \mathbf{B}_1, f_1(\tilde{\mathbf{C}}_1)]\!] + [\![\mathbf{A}_2, \mathbf{B}_2, f_2(\mathbf{C}_2)]\!] \tag{8}$$

wh ˜ere $[\![\mathbf{A}_1,\mathbf{B}_1, f_1(\mathbf{C}_1)]\!]$ 表示不受新冠肺炎影响时 RSV病例的预测值。

备注1。 我们的方法具有两方面优势：*1*）能够纳入来自其他来源的数据，以提升预测能力。*2*）能够轻松推广，用于预测新冠肺炎影响下的其他呼吸道疾病暴发，例如流感。

6 实验设置

数据。 实验所使用的数据集是一个美国州级RSV数据集，包含2015年至2022年间美国52个州（包括波多黎各和华盛顿特区）的每周RSV病例数据。该数据集由三个数据源组成：1）来自大规模医疗理赔数据库的RSV特征，我们从中提取了州级RSV确诊病例，以及与RSV相关的ICD-10诊断特征，例如B79.4（RSV）、J21.0（由RSV引起的急性细支气管炎）、J20.5（由RSV引起的急性支气管炎）和J21.1（由RSV引起的肺炎） [世界卫生组织， 2004]。2）来自约翰斯·霍普金斯大学的与新冠肺炎相关的特征 [Dong 等， 2020]。3）来自大气研究再分析数据集的气候特征，包括湿度、降水量和温度 [Kalnay 等， 1996]。研究发现，这些特征与RSV流行病发病时间的平均时点显著相关 [Baker等， 2019]。

6.1 评估流程。

训练数据范围为2015年11月6日至2021年4月16日，而测试数据范围为2021年4月23日至2022年4月15日。值得一提的是，2020年没有发生RSV大流行，尽管测试数据显示，所有州都出现了非季节性的呼吸道合胞病毒暴发，且大多数州的峰值水平超过了此前的最高值。采用这种训练集—测试集划分的目的是检验算法能否学习非药物干预对RSV传播的影响，并在前一年未观测到RSV激增的情况下，预测下一轮RSV波次的发生时间和强度。事实上，这种实验设计为所有机器学习算法都提供了极具挑战性的场景。通过这种方式，我们能够验证DeCom的非药物干预感知机制是否能在新冠肺炎影响下准确预测呼吸道合胞病毒暴发。

6.2 评估指标与基线。

我们在美国州级和国家级RSV数据上报告预测性能，其中后者由52个州的数据汇总而成，以评估模型的性能。我们重点估计RSV峰值日期和每周确诊病例数。我们计算均方根误差（RMSE）和平均绝对误差（MAE）作为性能指标。

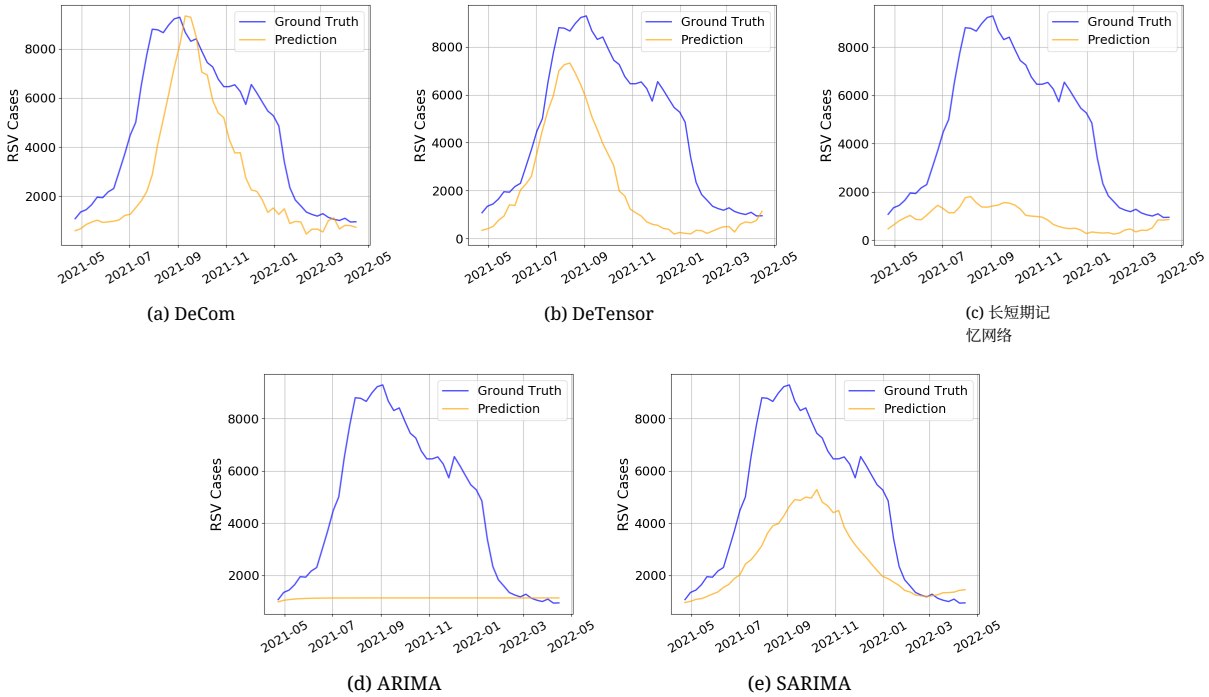
$$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t}(X_t - \hat{X}_t)^2} \tag{9}$$

$$\mathrm{MAE} = \frac{1}{T}\sum_{t}|X_t - \hat{X}_t| \tag{10}$$

我们比较 DeCom 与深度张量分解模型（DeTensor） [Kargas等人、 2021], 长短期记忆网络 [Hochreiter 和 Schmidhuber、 1997], ARIMA 以及季节性 ARIMA（SARIMA）[Hyndman andAthanasopoulos, 2018] 模型在各州分别训练后的预测性能。

DeTensor。

无残差连接的张量分解模型。与提出的 DeCom 模型不同，DeTensor 仅执行一次非负 CPD，以学习潜在子空间，该子空间表示正常的呼吸道合胞病毒季节性与新冠肺炎影响的混合。通过与 DeTensor 进行比较，我们能够评估在模型中使用第二个张量来感知非药物干预的有效性。



(a) DeCom (b) DeTensor (c) 长短期记忆网络

(d) ARIMA (e) SARIMA

图3：(a) DeCom、(b) DeTensor、(c) 长短期记忆网络、(d) ARIMA 和 (e) SARIMA 的国家级预测表现。DeCom 能够准确预测峰值时间（第±1 周），并生成幅度显著更大的预测值。

长短期记忆网络。

长短期记忆网络（LSTM）是一种特殊的循环神经网络（RNN），能够学习时间序列数据中的长期依赖关系。对于数据集中的每个州，我们都使用该州的历史特征（包括新冠肺炎特征）作为输入，训练其专属的 LSTM 模型，并独立预测该州呼吸道合胞病毒的演变。

ARIMA 和 SARIMA。

ARIMA（差分自回归移动平均）是一种广泛使用的统计方法，可用于识别时间序列数据中的一系列不同标准时间结构。SARIMA 是 ARIMA 的扩展，支持对时间序列的季节性成分进行建模。ARIMA 和 SARIMA 模型使用 Python 中的 statmodel 软件包 Seabold 和 Perktold [2010], 实现，并通过对（p, q, d）进行网格搜索来确定最优参数。对于 SARIMA，我们每个周期使用 $s = 48$ 个步骤。

备注 2. 我们注意到，存在一些用于季节性呼吸道疾病传播的标准流行病学模型，例如基于 *SIR* 的方法 *[Baker* 等， *2020]*。标准做法是嵌入一个周期为 *52* 周的余弦函数，以捕捉呼吸道合胞病毒的季节性。为了实现非药物干预，基于 *SIR* 的方法会手动选择一个降低比例，并将其乘以预先定义的传播率，但这种方法可能无法捕捉不同地点和不同时间窗口下非药物干预的异质性 *[Baker* 等， *2020]*。因此，用于预测下一次感染高峰到达时间的基于 *SIR* 的模型高度依赖于先前的

表3：RSV病例峰值时间（周）估计、均方根误差和平均绝对误差的国家级性能比较。

模型	峰值差异。	RMSE	MAE
DeCom	**1**	**2461**	**1891**
DeTensor	3	3175	2485
LSTM	4	4555	3711
ARIMA	8	4523	3490
SARIMA	5	2688	2102

学习到的相位偏移、传播率的选择以及非药物干预终止的时间，而后者通常难以预测。然而，像 *DeCom* 这样的数据驱动模型以及其他深度学习模型不受此类限制，因为非药物干预的影响会反映在新冠肺炎感染病例中，并被模型学习。这是我们不将此类流行病学方法纳入比较的主要原因。

7 实验结果

我们首先评估 DeCom、DeTensor、长短期记忆网络、ARIMA 和 SARIMA 的州级 RSV 预测性能。图2展示了州级 RSV 预测结果，所选州包括佛罗里达州、弗吉尼亚州、马里兰州和纽约州。选择这四个州的原因是，它们从南到北分布，且呼吸道合胞病毒暴发的传播呈现出变化。因此，我们可以检验算法能否捕捉由地理差异导致的传播差异。

表2比较了不同模型在不同时间间隔（分别为12、24和52周）上的州级预测性能。我们给出了均方根误差和平均绝对误差的均值及其标准差（括号内）。在最具挑战性的任务——预测长达52周的RSV病例——中，DeCom的均方根误差和平均绝对误差显著更低。同时，在24周预测中，DeCom以明显优势胜过其他模型，并在最长12周的预测中表现最佳。然而，随着问题变得更容易，即预测区间缩短，DeTensor、ARIMA和SARIMA的表现也更具竞争力。

结果清楚地表明，在峰值时间和强度估计方面，DeCom 优于其他四种方法。例如，在马里兰州，DeCom 的估计结果与真实值一致，而 DeTensor、长短期记忆网络、ARIMA 和 SARIMA 的估计峰值则相对于真实峰值向左或向右偏移。

我们还利用美国国家级的 RSV 预测结果评估 DeCom 的有效性。具体而言，我们首先使用全部五种模型估算美国 52 个州的 RSV病例，然后将州级病例汇总至国家级。RSV 估算结果见图3，性能指标汇总于表3。当实际峰值与预测峰值之间的差异小于± 1 周时，DeCom 能够准确预测峰值时间。此外，与其他基线模型相比，DeCom 对 RSV病例的预测结果更接近真实值，因此能够更准确地捕捉其强度。总体而言，与 DeTensor 相比，DeCom 的均方根误差降低了 23%，平均绝对误差降低了 24%；与分州LSTM 相比，均方根误差和平均绝对误差分别降低了 46% 和 49%；与 ARIMA 相比，均方根误差和平均绝对误差均降低了 46%；与 SARIMA 相比，性能提升约 10%。

8 结论

本文介绍了 DeCom，这是一种深度耦合张量分解机，能够同时捕捉 RSV 的季节性以及 COVID-19非药物干预措施对 RSV传播的影响。计算结果表明，在 COVID-19 的影响下，DeCom 在预测 RSV暴发的时间和规模方面优于多种时间序列预测基线方法。因此，我们认为 DeCom 可作为一种有效模型，用于预测和评估新冠肺炎之后的 RSV 暴发。

参考文献

L Acedo、J Diez-Domingo、J-A Morano 和 R-J Villanueva。呼吸道合胞病毒（呼吸道合胞病毒）的数学建模：疫苗接种策略和预算应用。《流行病学与感染》，138(6):853–860，2010年。

Wladimir J Alonso、Bruno J Laranjeira、Samuel AR Pereira、Caroline MGD Florencio、Eduardo C Moreno、Mark A Miller、Ricardo Giglio、Cynthia Schuck-Paim 和 Fernanda EA Moura。赤道城市儿童病毒性呼吸道感染的流行动态、发病率和死亡率负担比较。《儿科传染病杂志》，31(1):e9，2012年。

Sonia Altizer、Andrew Dobson、Parviez Hosseini、Peter Hudson、Mercedes Pascual 和 Pejman Rohani。传染病的季节性和动态。《生态学通讯》，9(4):467–484，2006年。

Abraham J Arenas、Gilberto González-Parra 和 Jose-A Moraño。西班牙瓦伦西亚地区呼吸道合胞病毒（呼吸道合胞病毒）的传播随机建模。《生物系统》，96(3):206–212，2009年。

Rachel E Baker、Ayesha S Mahmud、Caroline E Wagner、Wenchang Yang、Virginia E Pitzer、Cecile Viboud、Gabriel A Vecchi、C Jessica E Metcalf 和 Bryan T Grenfell。当前和未来气候下呼吸道合胞病毒的流行病动力学。《自然·通讯》，10(1):1–8，2019年。

Rachel E. Baker、Sang Woo Park、Wenchang Yang、Gabriel A. Vecchi、C. Jessica E. Metcalf 和 Bryan T. Grenfell。新冠肺炎非药物干预措施对地方性传染病未来动态的影响。《美国国家科学院院刊》，117(48):30547–30553，2020年。

美国疾病控制与预防中心。呼吸道合胞病毒趋势与监测，2022年。

Kenneth L Cooke 和 P Van Den Driessche。含两个时滞的 SEIRS 流行病模型分析。《数学生物学杂志》，35(2):240–260，1996年。

Ensheng Dong、Hongru Du 和 Lauren Gardner。用于实时追踪新冠肺炎的交互式基于网络的仪表板。《柳叶刀·传染病》，20(5):533–534，2020年。

Jean-Baptist du Prel、Wolfram Puppe、Britta Gröndahl、Markus Knuf、Franziska Weigl、Franziska Schaaff、Franziska Schaaff 和 Heinz-Josef Schmitt。气象参数是否与急性呼吸道感染有关？《临床传染病》，49(6):861–868，2009年。

Amanuel Tesfay Gebremedhin、Alexandra B Hogan、Christopher C Blyth、Kathryn Glass 和 Hannah C Moore。开发一种预测模型，用于估算西澳大利亚住院儿童中呼吸道合胞病毒（呼吸道合胞病毒）的实际疾病负担。科学报告，12(1):1–12，2022年。

Caroline Breese Hall、Geoffrey A Weinberg、Marika K Iwane、Aaron K Blumkin、Kathryn M Edwards、Mary A Staat、Peggy Auinger、Marie R Griffin、Katherine A Poehling、Dean Erdman 等。呼吸道合胞病毒感染给幼儿带来的疾病负担。新英格兰医学杂志，360(6):588–598，2009年。Sepp Hochreiter 和 Jürgen Schmidhuber。长短期记忆。神经计算，9(8):1735–1780，1997年。Rob J Hyndman 和 George Athanasopoulos。预测：原理与实践。OTexts，2018年。E. Kalnay、M. Kanamitsu、R. Kistler、W. Collins、D. Deaven、L. Gandin、M. Iredell、S. Saha、G. White、J. Woollen、Y. Zhu、M. Chelliah、W. Ebisuzaki、W. Higgins、J. Janowiak、K. C. Mo、C. Ropelewski、J. Wang、A. Leetmaa、R. Reynolds、Roy Jenne 和 Dennis Joseph。NCEP/NCAR 40年再分析项目。美国气象学会公报，77(3):437–472，1996年。Nikos Kargas、Cheng Qian、Nicholas D Sidiropoulos、Cao Xiao、Lucas M Glass 和 Jimeng Sun。Stelar：结合潜在流行病学正则化的时空张量分解。载于*AAAI*人工智能会议论文集，第35卷，第4830–4837页，2021年。William Ogilvy Kermack 和 Anderson G McKendrick。对流行病数学理论的贡献。伦敦皇家学会会刊：*A*辑，数学与物理学论文，115(772):700–721，1927年。Santiago Lapeña、María Belén Robles、Leticia Castañón、Juan Pablo Martínez、Sofiá Reguero、María Paz Alonso 和 Isabel Fernández。气候因素与西班牙北部住院婴儿的呼吸道合胞病毒所致下呼吸道感染。欧洲流行病学杂志，20(3):271–276，2005年。

Molly Leecaster、Per Gesteland、Tom Greene、Nephi Walton、Adi Gundlapalli、Robert Rolfs、Carrie Byington 和 Matthew Samore。儿童呼吸道合胞病毒季节性流行的变化建模。*BMC* 传染病，11(1):1–9，2011年。

Hannah C Moore、Peter Jacoby、Alexandra B Hogan、Christopher C Blyth 和 Geoffry N Mercer。幼儿呼吸道合胞病毒季节性流行的建模研究。*PloS one*，9(6):e100422，2014年。DE Noyola 和 PB Mandeville。气候因素对呼吸道合胞病毒流行的影响。《流行病学与感染》，136(10):1328–1332，2008年。Terezinha M Paiva、Maria A Ishida、Margarete A Benega、Clóvis RA Constantino、Daniela BB Silva、Kátia CO Santos、Maria I Oliveira、Helena A Barbosa、Telma RMP Carvalhanas、Cynthia Schuck-Paim 等。亚热带特大城市中呼吸道合胞病毒传播时间的变化：对免疫预防的影响。*Journal of medicalvirology*，84(11):1825–1830，2012年。Stuart Paynter、Laith Yakob、Eric AF Simoes、Marilla G Lucero、Veronica Tallo、Hanna Nohynek、Robert S Ware、Philip Weinstein、Gail Williams 和 Peter D Sly。运用数学传播模型探究菲律宾儿童呼吸道合胞病毒季节性的驱动因素。*PLoS One*，9(2):e90094，2014年。Virginia E Pitzer、Cécile Viboud、Wladimir J Alonso、Tanya Wilcox、C Jessica Metcalf、Claudia A Steiner、Amber K Haynes 和 Bryan T Grenfell。美国呼吸道合胞病毒时空动态的环境驱动因素。*PLoS pathogens*，11(1):e1004591，2015年。

Julia Reis 和 Jeffrey Shaman。美国呼吸道合胞病毒的回顾性参数估计与预测。*PLoS* 计算生物学，12(10):e1005133，2016年。Julia Reis、Teresa Yamana、Sasikiran Kandula 和 Jeffrey Shaman。美国国家、区域和州级呼吸道合胞病毒暴发的超级集合预测。《流行病》，26:1–8，2019年。Skipper Seabold 和 Josef Perktold。statsmodels：使用 Python 进行计量经济学与统计建模。在*Python*科学会议，2010年。Jeffrey Shaman、Virginia E Pitzer、Cécile Viboud、Bryan T Grenfell 和 Marc Lipsitch。美国本土绝对湿度与流感季节性起始。*PLoS* 生物学，8(2):e1000316，2010年。Ting Shi、David A McAllister、Katherine L O'Brien、Eric AF Simoes、Shabir A Madhi、Bradford D Gessner、Fernando P Polack、Evelyn Balsells、Sozinho Acacio、Claudia Aguayo 等。2015年幼儿因呼吸道合胞病毒导致的急性下呼吸道感染全球、区域和国家疾病负担估计：系统综述与建模研究。《柳叶刀》，390(10098):946–958，2017年。

Nephi A Walton、Mollie R Poynton、Per H Gesteland、Chris Maloney、Catherine Staes 和 Julio C Facelli。利用实时气象变量预测呼吸道合胞病毒暴发的起始周。*BMC*医学信息学与决策制定，10(1):1–9，2005年。

Andreas Weber、Martin Weber 和 Paul Milligan。呼吸道合胞病毒（rsv）引起的流行病建模。《数学生物科学》，172(2):95–113，2001年。

LJ White、M Waris、PA Cane、DJ Nokes 和 GF Medley。A组和B组人呼吸道合胞病毒（hrsv）在英格兰和威尔士及芬兰的传播动力学：季节性与交叉保护。《流行病学与感染》，133(2):279–289，2005年。

LJ White、JN Mandl、MGM Gomes、AT Bodley-Tickell、PA Cane、P Perez-Brena、JC Aguilar、MM Siqueira、SA Portes、SM Straliotto 等。利用多重时间序列和嵌套模型理解呼吸道合胞病毒的传播动力学。《数学生物科学》，209(1):222–239，2007年。世界卫生组织。国际疾病和相关健康问题统计分类第十版：第十次修订。第*1*卷：类目表。国际疾病和相关健康问题统计分类。世界卫生组织，2004年。S Yusuf、G Piedimonte、A Auais、G Demmler、S Krishnan、Pâ Van Caeseele、R Singleton、S Broor、S Parveen、L Avendano 等。气象条件与呼吸道合胞病毒流行活动的关系。《流行病学与感染》，135(7):1077–1090，2007年。