

低轨多波束卫星下行业务强度时空预测方法

郑弘毅¹, 张婷婷²

(1.北京理工大学集成电路与电子学院, 北京 100081;

2.北京理工大学信息与电子学院, 北京 100081)

摘要: LEO巨型星座正在重塑全球宽带连接格局。然而, 受限于是星上功率、频谱带宽及波束跳变时隙的多维资源稀缺性, 实现高效的资源调度高度依赖对未来业务需求的先验感知。为此, 实现对逻辑波束级下行业务强度的鲁棒多步预测, 成为支持主动式资源预配置的关键。针对卫星流量的长时周期与局部突发特性, 提出一种解耦时空预测框架。具体而言, 该方法利用数据自适应正交复值基提取主导时间演化模式, 通过低维子空间投影与基外推重构实现趋势预测; 随后, 引入稀疏向量自回归模型对预测残差建模, 精准捕获波束间的瞬时空间耦合。实验结果表明, 该框架凭借紧凑表征与稀疏结构, 在显著降低计算开销的同时, 实现了优于现有基线算法的预测精度。

关键词: 卫星网络; 流量分析; 时空预测

中图分类号: TP393

文献标志码: A

doi: 10.11959/j.issn.2096-8930.2026014

Spatio-temporal prediction method for the downlink service intensity of low-orbit multi-beam satellites

Zheng Hongyi¹, Zhang Tingting²

1. School of Integrated Circuits and Electronics, Beijing Institute of Technology, Beijing 100081, China

2. School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China

Abstract: LEO mega-constellations are revolutionizing global broadband connectivity landscape. However, constrained by the multidimensional scarcity of on-board power, spectrum bandwidth, and beam hopping time slots, efficient resource scheduling critically depends on prior knowledge of future traffic demands. Consequently, achieving robust multi-step forecasting of downlink traffic intensity at the logical beam level is indispensable for enabling proactive resource provisioning. To address the long-term periodicity and local burstiness of satellite traffic, a decoupled spatio-temporal forecasting framework is proposed. Specifically, this method utilizes data-adaptive orthogonal complex-valued bases to extract dominant temporal evolution patterns, achieving trend prediction via low-dimensional subspace projection and basis extension. Subsequently, a sparse vector autoregression model is introduced to model prediction residuals, explicitly capturing transient spatial couplings among beams. Experimental results demonstrate that, leveraging its compact representation and sparse structure, the proposed framework achieves superior prediction accuracy with significantly reduced computational overhead compared to existing baselines algorithms.

Key words: satellite networks, traffic analysis, spatio-temporal prediction

0 引言

随着空天一体化网络持续演进, 低地球轨道 (low

earth orbit, LEO) 卫星多星协同凭借全球覆盖、低时延、大带宽和连续服务等优势, 正成为广域宽带接入与信息服务的重要支持^[1]。相较于中高轨系统, 低轨星座在路

收稿日期: 2026-02-27; 修回日期: 2026-05-07

基金项目: 国家自然科学基金资助项目 (No.62401047)

Foundation Item: The National Natural Science Foundation of China (No.62401047)

径损耗、部署成本和连续覆盖等方面具备显著优点，已在多国实现规模化商用部署，广域测量研究也验证了其低时延、高可用性的体系特征^[2]。面向低轨多波束宽带卫星系统，网络侧普遍采用多波束成形和按需点亮的跳波束（beam hopping, BH）策略，通过在时间、频率、功率维度灵活编排资源来动态适配地面业务负荷的时空变化，提升整体频谱效率。此类系统通常以预先规划的跳波束时序表（beam hopping time plan, BHTP）为核心，并结合业务驱动的自适应点亮策略，因此在卫星过境前获取高保真的负荷预测对调度至关重要^[3]。尽管 LEO 星座的星历与轨道参数具有严格的可预报性，但物理层拓扑的确定性并不等价于业务层需求的可知性。宽带接入业务本质上受用户行为、内容分发偏好及突发事件驱动，具有显著的随机性与局部突发性。这意味着即便存在链路状态的实时信令反馈，其固有的滞后性仍难以满足生成 BHTP 所需的前瞻性决策要求。因此，构建面向逻辑波束粒度的高保真多步需求预测模型，是支撑资源预配置、实现供需精准匹配的基础^[1,3]。

在预测对象的具体选择上，鉴于宽带互联网接入业务固有的上下行流量呈显著非对称特性，下行链路不仅承载了绝大多数的数据传输负荷，更是卫星与网关侧集中式资源调度的核心。相较于上行链路，下行需求的预测精度直接决定了全系统功率分配、带宽利用率及波束照射策略的整体收益。因此，本文聚焦更具研究挑战性及对系统性能起主导作用的下行链路业务强度预测。

考虑到 LEO 星座的几何与业务特征，一方面，主流星座多接近 Walker 构型，覆盖图样呈准周期重复模式；另一方面，互联网流量具有显著的昼夜/周周期特征以及区域差异。二者叠加使得在地球固定参考系下按日/周尺度进行波束级需求预测成为可能^[4]。有研究通过地面固定的逻辑波束（如六边形网格 H3）对覆盖区域离散化，由不同过顶卫星依次服务该固定单元，从而获得统计性质更稳定的时间序列^[5]，地面固定逻辑波束示意如图 1 所示。这种视角能够在一定程度上平滑卫星相对运动带来的非平稳性，使负荷序列体现出较强的周期模式和长期相关性，便于进行预测分析。同时，现网实测研究指出，LEO 商业系统中普遍存在与卫星切换强相关的、分钟级的可复现时延抖动与突发丢包现象。这种链路状态的波动经传输层拥塞控制机制的反馈传导会在短时尺度下显著加剧相邻覆盖区之间的流量时空耦合^[6]。这一现象进一步要求预测模型既能刻画短周期可复现的动态模式，又能精准反映空间近邻单元间的瞬时联动依赖，并严格满足在

线推理的时延与算力约束。为此，本文提出一种面向逻辑波束粒度的下行业务强度时空预测方法，该方法具有鲁棒、轻量、数学形式明确的特点，可有效支撑 BHTP 的前瞻性生成与资源预分配。

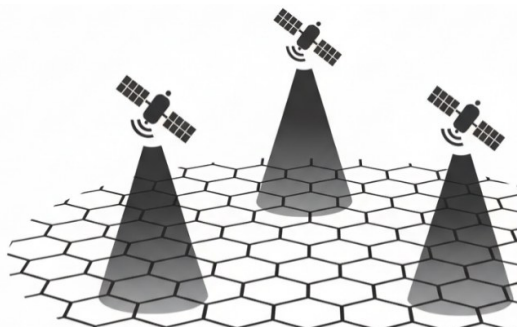


图1 地面固定逻辑波束示意

1 研究现状

早期卫星网络流量预测主要借鉴地面网络的经典时间序列模型，如马尔可夫链、自回归移动平均（autoregressive moving average, ARMA）模型、自回归积分移动平均（autoregressive integrated moving average, ARIMA）模型等^[7]。这类线性模型需要根据经验设定多种参数，只适用于短期平稳序列，难以应对卫星流量所体现的长程依赖（long-range dependence, LRD）特性。在卫星通信场景下，流量序列常呈现强自相似和非平稳性，ARMA/ARIMA 等模型因为假定序列短记忆平稳，无法准确捕捉长期相关模式。尽管分数阶 ARIMA（fractional ARIMA, FARIMA）模型通过非整数阶差分引入了长短期记忆^[8]，但其计算复杂度随序列长度显著提升，实际部署中需针对每个序列重复建模，模型阶次选择和参数估计的开销巨大，难以在卫星上实时应用。例如，有分析指出，对长度为 n 的序列执行 FARIMA 预测的时间复杂度约为 $O(n^2)$ ，且模型阶次选取额外增加了开销，难以满足星上计算约束。为提升传统方法性能，一些研究引入了信号预处理等改进手段，例如先采用经验模态分解（empirical mode decomposition, EMD）提取流量趋势和高频成分，再分别应用 ARMA/ARIMA 模型预测并重构，从而显著提高预测精度^[9]；又如经典的用于短期流量平滑预测的 Kalman 滤波，但其线性高斯假设难以应对卫星流量的非线性波动和突发变化。同时，这些统计模型通常要求时间序列平稳，而 LEO 场景下拓扑周期变动、链路频繁切换导致流量统计特性随时间推移变化，违背了平稳

性假设。传统模型难以及时自适应,往往需要频繁重新训练参数,在卫星算力有限的条件下难以满足实时性要求^[10]。综上,传统方法在卫星波束级流量预测中面临无法刻画长程依赖(精度不足)和难以适应非平稳动态(鲁棒性欠佳)的局限。

随着人工智能的发展,机器学习和深度学习逐渐被应用于卫星网络流量预测。早期机器学习模型(如人工神经网络、支持向量机、最小二乘支持向量机、极限学习机等)被引入流量预测,凭借其优越的非线性映射与泛化性能,在一定程度上提升了预测精度^[11]。然而,此类模型大多将流量序列视作独立样本进行孤立的静态回归,未显式建模时间依赖结构,难以有效捕捉卫星流量中隐含的长时相关模式,因而精度提升有限。为此,研究者开始采用循环神经网络(recurrent neural network, RNN)等深度时序模型以更有效提取挖掘深层时间依赖。长短期记忆(long short-term memory, LSTM)网络通过引入输入门、遗忘门、输出门机制缓解了传统 RNN 的梯度消失问题,能够学习时间序列中的长期依赖关系^[12]。LSTM 网络在处理时序数据方面表现出良好的稳定性和鲁棒性,能够同时刻画短期和长期模式,但其参数规模大、计算开销高,在训练数据不足时易产生过拟合。门控循环单元(gated recurrent unit, GRU)对 LSTM 网络结构进行了简化,将遗忘门和输入门合并为更新门,并将细胞状态与隐状态合并,在降低复杂度的同时保留了长程记忆能力^[13]。在依赖较强的短期预测任务上,GRU 的性能可媲美甚至优于 LSTM 网络,因而更适合在卫星算力有限环境中部署。

卫星流量具有显著的空间相关性,不同波束/卫星间的流量状态并非独立。为建模空间依赖关系,图神经网络(graph neural network, GNN)近年来受到广泛关注。一些研究将卫星或波束视作图模型的节点,引入图卷积网络(graph convolution network, GCN)从拓扑结构中提取复杂的空间特征^[14]。GCN 的优势在于能够将拓扑关系融入模型学习,通过对图结构数据的卷积操作,挖掘邻近节点间的依赖关系和特征传播模式,从更全面的视角学习流量数据的内在联系。例如,有研究将 GCN 与时间序列模型结合用于卫星网络流量预测,充分利用了网络拓扑的复杂关系,不仅提高了预测精度,也增强了对突发流量和复杂网络环境的适应性^[15]。另一项研究构建了静态和动态相结合的空间相关模型,考虑卫星之间固定的邻接关系和随时间变化的连通性因素,仿真结果表明其预测误差显著小于传统方法^[16]。然而,现有的时空图

神经网络大多面向地面交通或蜂窝网络等拓扑静态场景设计,若将其直接迁移,LEO 星座的高动态性会导致空间相邻关系随时间剧烈漂移,使得基于静态图的特征提取失效。因此,研发能够规避拓扑演化干扰并有效解耦长短时模式的专用预测架构,已成为该领域的关键命题。

作为近年来流行的序列预测框架,Transformer 也开始应用于卫星流量预测领域。Transformer 基于自注意力机制,善于捕获长距离的依赖关系且易于并行计算,非常适合处理高粒度、长序列的预测任务。已有研究提出面向 LEO 卫星用户需求的 Transformer 模型,用于替代传统 FARIMA 模型实现多步预测,解决了后者需对每条序列重新拟合的问题^[17]。通过引入概率稀疏自注意力等改进机制(如 Informer^[18]),Transformer 模型在长序列预测中的效率大幅提升。相比于常规 RNN 方法,基于 Transformer 的方案能够在保持预测精度的同时降低推理开销,展现出在卫星流量预测中的应用潜力。除上述主要方向外,研究者还探索了多种混合模型与优化技术来提高预测性能。例如,奇异谱分析(singular spectrum analysis, SSA)等方法也被用于提取隐藏周期特征^[19]。此外,遗传算法、粒子群优化(particle swarm optimization, PSO)算法等元优化技术被用于神经网络参数调优,借助全局搜索提高模型的收敛速度和精度^[20]。总体而言,这些基于机器学习和深度学习的方法克服了传统模型无法捕捉非线性长相关的缺陷,大幅提升了卫星波束级流量预测的准确性和鲁棒性。

虽然上述方法在精度上取得了进展,但不少深度模型依赖庞大的参数规模或复杂的注意力机制,造成推理开销较高,且模型的可解释性相对不足,这在星上实时部署时可能受到算力和时延限制,并降低运营方对模型预测结果的信任度。因此,亟须一种轻量级且具备可解释性的时空预测方法,在有限计算资源和严格时延约束下,既能有效捕获多尺度时间模式又能刻画短时空邻域关联。针对上述需求,本文提出的预测方案在时间域采用紧凑的正交基表示,在空间域通过稀疏向量自回归进行邻域校正,旨在以较小的参数规模实现对 LEO 卫星下行业务流量的稳定多步预测,从而为卫星过境前的资源调度提供可靠支撑。

2 算法架构与实现

2.1 问题建模与区域图构建

本文聚焦低轨多星座宽带系统的下行用户链路。由

于LEO卫星相对于地面进行高速运动，若直接以卫星本体为视角对其服务流量进行建模，将面临两大难题：一是星地拓扑关系持续高速时变；二是流量数据的统计特征随之呈现非平稳性。为有效应对上述挑战，本文采用业界广泛采用的地球固定参考系策略^[21]，将建模视角从卫星侧转换至地面侧。具体而言，该策略将地面服务区域预先划分为固定的地理单元（即逻辑波束或网格），过顶卫星则以离散时间切片为单位，依次为这些固定单元提供通信服务。由此，动态的星地链路观测问题被转化为对固定地面单元流量序列的建模问题。

从业务来源看，LEO卫星宽带流量主要由地面用户行为产生，因此在时空层面往往呈现与地面移动互联网相似的周期性与偏态分布特征。基于此，相关研究常借鉴蜂窝网络或互联网流量的周期与自相似建模思路用于卫星波束需求预测与资源调度^[22]。因此本文将动态的卫星网络流量预测问题规约为固定拓扑下的时空序列预测问题。

具体而言，将覆盖面离散成 N 个逻辑波束（亦可称“区域”），用整数 $i \in \{1, 2, \dots, N\}$ 标识第 i 个区域；第 i 个区域在球面上的几何中心用 $\mathbf{g}_i$ 表示，其面积用 $\mathcal{A}_i$ 表示。为方便业务量统计以及时序数据构建，将下行业务按照离散时隙划分，离散时隙用正整数 $t = 1, 2, \dots$ 表示，时间粒度初步设定为10 min。在时刻 t ，第 i 个区域的原始下行业务强度（按照平均吞吐量或单位时间内的比特数/请求数进行量化记录）用实数 $x_i(t) \in \mathbb{R}_{\geq 0}$ 表示。为进行多步预测，取一个长度为 L 的回看时间窗，将第 i 个区域在时刻 t 的历史时序片段写成列向量 $\mathbf{x}_i^t = [x_i(t-L+1), \dots, x_i(t)]^T \in \mathbb{R}^L$ ，并计划预测未来 H 步（horizon）的时序向量 $\mathbf{y}_i^t = [x_i(t+1), \dots, x_i(t+H)]^T \in \mathbb{R}^H$ ， $\mathbb{R}$ 表示实数集合，上标 T 代表矩阵转置。本文的直接目标就是：在给定所有区域历史时序集合 $\{\mathbf{x}_i^t\}_{i=1}^N$ 的条件下，输出所有区域的预测集合 $\hat{\mathbf{Y}}^t = \{\hat{\mathbf{y}}_i^t\}_{i=1}^N$ 。

为了在模型中表征空间邻近关系，构造一个仅由几何信息确定的静态邻接图，记为 $\mathcal{G}_{\text{geo}} = (\mathcal{V}, \mathcal{E}_{\text{geo}}, \mathbf{W})$ 。其中 $\mathcal{V} = \{1, 2, \dots, N\}$ 是节点集合，对应各区域； $\mathcal{E}_{\text{geo}} \subseteq \mathcal{V} \times \mathcal{V}$ 是无向边集合，表示区域之间的几何相邻； $\mathbf{W} = [w_{ij}] \in \mathbb{R}^{N \times N}$ 是边权矩阵。为了定义相邻关系与边权取值，引入球面大圆距离 d_{ij} ，即球面上两点间的最短路径长度，可由经纬度通过标准公式计算。图的边集通过两种等价方式确定：其一，若两个区域在网格剖分结构上共享边界，则视为相邻，满足 $(i, j) \in \mathcal{E}_{\text{geo}}$ 。边权 w_{ij} 为非

负实数，用以刻画距离越近耦合越强的特性，采用径向基函数形式，即：

$$w_{ij} = \exp(-d_{ij}^2/\rho^2) \cdot 1[(i, j) \in \mathcal{E}_{\text{geo}}] \quad (1)$$

其中， $\rho > 0$ 是衰减尺度的超参数， $1[\cdot]$ 是指示函数（条件为真取1，否则取0）。为更拟合真实环境，若已有估计到的波束足迹重叠度（如两区域覆盖多边形的交并比 $\text{IoU}(\mathcal{A}_i, \mathcal{A}_j)$ ），可在指数权重上进一步线性叠加一项以强调重叠带来的实际联动效应。如此得到的 $\mathcal{G}_{\text{geo}}$ 只取决于固定的空间几何，因此在卫星交接与链路切换频繁的场景中，时间序列仍然绑定于固定地面区域 i ，避免了卫星编号变化导致的统计漂移，从而更有利于模型建模与结果解释。

2.2 基于自适应正交基的时序模式提取

本节描述的时间域表示与预测模块遵循“自适应基+线性重构”的思想，即在复杂向量空间中学习一组长度一致、近似正交、带有“外推段”的复值基向量，通过对历史片段进行内积投影得到系数，再仅用同一组基在未来预测区间内的外推段进行线性组合以生成未来多步预测结果。首先定义一个复值基字典矩阵 $\mathbf{B} \in \mathbb{C}^{K \times (L+H)}$ ，其中， K 是基的个数（ $K \ll L$ ），矩阵的第 k 行 $\mathbf{b}_k \in \mathbb{C}^{1 \times (L+H)}$ 表示第 k 个复值基随时间的取值。将矩阵 $\mathbf{B}$ 按列切分成两部分，即：

$$\mathbf{B} = [\mathbf{B}^{(L)} \mid \mathbf{B}^{(H)}] \quad (2)$$

其中， $\mathbf{B}^{(L)} \in \mathbb{C}^{K \times L}$ 是每个基向量在回看窗内的取值（记作“窗内段”）， $\mathbf{B}^{(H)} \in \mathbb{C}^{K \times H}$ 是每个基向量在未来时刻的取值（记作“外推段”）。给定区域 i 的历史片段 $\mathbf{x}_i^t \in \mathbb{R}^L$ ，将其与每个基向量的窗内段做标准内积，得到初始复系数向量，即：

$$\begin{aligned} c_{i,k} &= \langle \mathbf{x}_i^t, (\mathbf{b}_k)_{1:L} \rangle \\ \mathbf{c}_i &= [c_{i,1}, \dots, c_{i,K}]^T \in \mathbb{C}^K \end{aligned} \quad (3)$$

其中， $(\mathbf{b}_k)_{1:L}$ 表示第 k 个基向量的前 L 个元素。由于 $\mathbf{x}_i^t$ 是实向量而基为复向量，本文采用复内积定义 $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^\dagger \mathbf{v}$ 进行计算，上标 $\dagger$ 表示复共轭转置，即厄米转置。这样 $c_{i,k}$ 一般是复数，其模长可以视为模式强度，辐角可以视为相位， $\mathbb{C}$ 表示复数集合。理想情况下，如果 $\mathbf{B}^{(L)}$ 的各行严格两两正交且模长一致，则 $\mathbf{c}_i$ 可视为该信号在一组复正弦样模式上的正交展开系数；但由于 $\mathbf{B}^{(L)}$ 是数据驱动学习得到的，仅能实现近似正交，不同基之间仍可能存在串扰。

为消除非正交造成的系数串扰，本文在训练过程中构造一个线性系数补偿映射，将初始系数 $\mathbf{c}_i$ 映射为真实系数 $\mathbf{d}_i \in \mathbb{C}^K$ 。记该线性映射矩阵为 $\mathbf{W}_{\text{comp}} \in \mathbb{C}^{K \times K}$ ，

则有:

$$\mathbf{d}_i = \mathbf{W}_{\text{comp}} \mathbf{c}_i \quad (4)$$

从解析角度看, 若令 $\mathbf{G} = \mathbf{B}^{(L)} (\mathbf{B}^{(L)})^\dagger \in \mathbb{C}^{K \times K}$ 表示基在窗内段的格拉姆矩阵 (格拉姆矩阵的第 p 、 q 个元素是第 p 个与第 q 个基向量在窗内段的内积), 在 $\mathbf{G}$ 可逆时有 $\mathbf{W}_{\text{comp}} \approx \mathbf{G}^{-1}$ 。但直接求逆在数值上可能不稳定, 且在训练时代价较高, 因此将 $\mathbf{W}_{\text{comp}}$ 作为可学习的线性层参数, 随训练过程端到端学习, 并在损失函数中加入正交性正则约束以推动 $\mathbf{B}^{(L)}$ 接近酉矩阵 (即满足 $\mathbf{B}^{(L)} (\mathbf{B}^{(L)})^\dagger \approx \mathbf{I}_K$, $\mathbf{I}_K$ 是 $K \times K$ 的单位矩阵), 从源头减轻串扰。

获得补偿后的系数 $\mathbf{d}_i$ 后, 只使用同一字典的外推段 $\mathbf{B}^{(H)}$ 进行未来 H 步的线性重构, 得到区域 i 在时刻 t 的时序预测向量, 即:

$$\hat{\mathbf{y}}'_i = \mathbf{R} \{ (\mathbf{B}^{(H)})^\dagger \mathbf{d}_i \} \in \mathbb{R}^H \quad (5)$$

之所以取实部 $\mathbf{R} \{ \cdot \}$, 是因为目标量 $x_i(\cdot)$ 为实数, 而复基展开在理论上对应幅相可分的复正弦表征, 取实部即可得到与现实量纲一致的实值预测。同理, 可用同一系数 $\mathbf{d}_i$ 在窗内段进行线性重构, 得到回看窗的重构序列 $\hat{\mathbf{x}}'_i = \mathbf{R} \{ (\mathbf{B}^{(L)})^\dagger \mathbf{d}_i \}$, 它用于训练时的一致性约束与后续空间残差建模。

训练目标采用 3 部分损失函数的加权求和。

第一部分是预测均方误差 (mean square error, MSE), 对所有区域、所有未来时间的误差平方求平均, 形式为:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N \times H} \sum_{i=1}^N \|\hat{\mathbf{y}}'_i - \mathbf{y}'_i\|_2^2 \quad (6)$$

其中, $\|\cdot\|_2$ 表示欧氏范数。

第二部分是正交性约束损失 $\mathcal{L}_{\text{Orth}}$, 用于引导 $\mathbf{B}^{(L)}$ 在窗内段趋近酉正交, 写作:

$$\mathcal{L}_{\text{Orth}} = \|\mathbf{B}^{(L)} (\mathbf{B}^{(L)})^\dagger - \mathbf{I}_K\|_F^2 \quad (7)$$

其中, $\mathbf{I}_K$ 是 $K \times K$ 的单位矩阵, $\|\cdot\|_F$ 表示矩阵的弗罗贝尼乌斯范数。

第三部分是窗内重构损失 $\mathcal{L}_{\text{Rec}}$, 它要求利用补偿后的系数在窗内段的重构序列与真实历史时序一致, 同样采用均方误差刻画, 写作:

$$\mathcal{L}_{\text{Rec}} = \frac{1}{N \times L} \sum_{i=1}^N \|\hat{\mathbf{x}}'_i - \mathbf{x}_i\|_2^2 \quad (8)$$

总损失是三者的线性组合, 即:

$$\mathcal{L} = \mathcal{L}_{\text{MSE}} + \lambda_1 \mathcal{L}_{\text{Orth}} + \lambda_2 \mathcal{L}_{\text{Rec}} \quad (9)$$

其中, $\lambda_1 > 0$, $\lambda_2 > 0$ 是权衡系数, 由验证集调节。该训练目标可使模型在学习一组数据自适应的复正弦/缓变调制基的同时, 保证模式系数的可恢复性与稳定性, 而纯线性的外推—重构路径又天然具备可解释性与低推理开

销。上述训练过程结束后, 字典的窗内段与外推段分别固定为 $\mathbf{B}^{(L)} \in \mathbb{C}^{K \times L}$ 与 $\mathbf{B}^{(H)} \in \mathbb{C}^{K \times H}$ 。而在在线推理阶段, 对任一区域在时刻 t 的历史片段 $\mathbf{x}_i^t \in \mathbb{R}^L$, 首先按复内积计算初始系数 $\mathbf{c} = (\mathbf{B}^{(L)})^\dagger \mathbf{x}_i^t \in \mathbb{C}^K$ 。在线推理阶段, 本文不再显式调用训练阶段的可学习补偿矩阵 $\mathbf{W}_{\text{comp}}$, 而是基于已固定的窗内基矩阵构造格拉姆矩阵 $\mathbf{G}$, 并通过楚列斯基 (Cholesky) 分解与两次三角回代获得补偿后的系数, 从而实现对非正交串扰的数值稳定校正。该方式避免了显式矩阵求逆, 适合在线重复推理。预先离线构造格拉姆矩阵 $\mathbf{G} = \mathbf{B}^{(L)} (\mathbf{B}^{(L)})^\dagger \in \mathbb{C}^{K \times K}$, 并对 $\mathbf{G}$ 进行 Cholesky 分解 $\mathbf{G} = \mathbf{L}\mathbf{L}^\dagger$, 其中 $\mathbf{L} \in \mathbb{C}^{K \times K}$ 为对角线元素为正实数的下三角矩阵。推理时, 将系数补偿等价于两次三角求解: 先求解 $\mathbf{L}\mathbf{u} = \mathbf{c}$ 得到中间向量 $\mathbf{u} \in \mathbb{C}^K$, 再求解 $\mathbf{L}^\dagger \mathbf{d} = \mathbf{u}$ 得到真实系数 $\mathbf{d} \in \mathbb{C}^K$ 。这种分解一次、回代多次的做法等价于显式乘以 $\mathbf{G}^{-1}$, 且数值更稳定。随后用外推段进行线性重构并取实部, 得到多步预测: $\hat{\mathbf{y}} = \mathbf{R} \{ (\mathbf{B}^{(H)})^\dagger \mathbf{d} \} \in \mathbb{R}^H$ 。

上述实现具有清晰的复杂度上界。对于单个样本, 内积计算 $\mathbf{c} = (\mathbf{B}^{(L)})^\dagger \mathbf{x}^t$ 的代价为 $O(KL)$ 次复乘加; 两次三角回代的总代价为 $O(K^2)$ 次复乘加; 线性重构的代价为 $O(KH)$ 。因此单个样本的总复杂度为:

$$O(KL) + O(K^2) + O(KH) = O(K(L + H) + K^2) \quad (10)$$

其中, $K \ll L$ 是典型配置, 二次项 K^2 的常数因子很小且与 L 、 H 无关。在有 N 个并行区域时, 总复杂度线性叠加为 $O(N[K(L + H) + K^2])$, 这与训练阶段常数相分离, 便于在星上或边缘设备开展吞吐性能评估。为了进一步提升数值稳健性与消除偶发的病态条件数, 可在格拉姆矩阵上施加对角加载得到 $\mathbf{G}_\delta = \mathbf{G} + \delta \mathbf{I}_K$ ($\delta > 0$ 为极小实数, $\mathbf{I}_K$ 为 $K \times K$ 单位矩阵), 并对 $\mathbf{G}_\delta$ 进行楚列斯基分解, 此举等价于对补偿引入 Tikhonov 正则, 仅有极小的偏差代价而显著提升回代稳定性。若遇到 $\mathbf{G}$ 秩亏或近秩亏 (如个别基几乎共线), 亦可离线采用特征分解或奇异值分解 (singular value decomposition, SVD) 计算广义逆: 若 $\mathbf{G} = \mathbf{U}\mathbf{A}\mathbf{U}^\dagger$ 且对过小特征值做阈值截断, 则 $\mathbf{d} \approx \mathbf{U}\mathbf{A}^* \mathbf{U}^\dagger \mathbf{c}$, 本文采用符号 $*$ 作为上标表示 Moore-Penrose 广义逆或截断广义逆。由于 SVD 的离线分解复杂度为 $O(K^3)$, 高于 Cholesky 分解, 且在线阶段仍需 $O(K^2)$ 级别的矩阵回乘, 因此 SVD 仅在格拉姆矩阵明显病态时作为稳健替代方案。

2.3 基于稀疏残差图的瞬时空间修正

仅依靠时间端的分解与外推即可为每个区域给出一条初始预测, 但这一步的设计初衷是通过少量、可解释

的复值近似正交基去吸收可重复的主导时间结构，例如昼夜周期、周周期以及缓变的趋势项。在实际网络场景中，许多短时尺度的扰动并不完全可由这些可重复模式解释：例如，邻接区域之间由物理几何邻近与覆盖重叠引起的用户切换与相互挤占；星间链路或信关站的瞬时拥塞反馈带来的速率联动；GNSS/天气/活动等因素触发的局部外生冲击，以及卫星切换导致的瞬时可用性波动。这些因素常常表现为局地、弱平稳、短记忆的噪声样结构，会在相邻或重叠区域之间以极短时延传播，进而形成具有方向性或对称性的短期联动特征。如果直接在原始序列上进行空间建模，长周期与短周期联动会叠加混杂，一方面破坏模型的可解释性，另一方面也会因为参数冗余而增大估计方差。基于此，本文首先在时间端利用字典 $\mathbf{B}$ 的窗内段 $\mathbf{B}^{(L)}$ 与补偿系数 $\mathbf{d}_i$ 对第 i 个区域的回看窗进行线性重构，拟合可重复、可解释的长周期/趋势成分；然后仅对剩余的、未被解释的那一部分进行空间统计建模，从而将时序可重复性和空间的短期联动在建模上解耦。由时域模型给出同一窗口内的线性重构 $\hat{\mathbf{x}}_i^t = \mathbf{R}\{(\mathbf{B}^{(L)})^\top \mathbf{d}_i\} \in \mathbb{R}^L$ ，定义时间端残差向量为：

$$\mathbf{e}_i^t = \mathbf{x}_i^t - \hat{\mathbf{x}}_i^t \in \mathbb{R}^L \quad (11)$$

为便于跨区域统计，本文把窗口内每一个时间位点的瞬时残差标量写成：

$$\mathbf{e}_i(s) = \mathbf{x}_i(s) - \hat{\mathbf{x}}_i(s), s \in \{t-L+1, t-L+2, \dots, t\} \quad (12)$$

并把 L 个时刻的残差按列堆叠成一个残差数据矩阵，即：

$$\mathbf{R} = [\mathbf{e}(t-L+1), \mathbf{e}(t-L+2), \dots, \mathbf{e}(t)] \in \mathbb{R}^{N \times L} \quad (13)$$

$$\mathbf{e}(s) = [\mathbf{e}_1(s), \dots, \mathbf{e}_N(s)]^\top \in \mathbb{R}^N$$

其中， $\mathbf{R}$ 的第 i 行是区域 i 在窗口内的残差时间序列，第 i 列是所有区域在某一时刻的残差横截面。时间端已经把可重复的长周期、趋势、相位推进等强全局成分解耦，残差序列因此更接近弱平稳（均值近零、方差近常数、相关性随滞后衰减），这恰好是经典空间—时间统计模型（如向量自回归（vector autoregression, VAR）模型）的工作区间。同时，残差把跨区域共同驱动（如全局昼夜）大幅削弱，剩下的相关性更可能对应局地、直接的相互作用，适合作为条件相关图的建模对象。从优化角度看，在残差域，协方差矩阵的谱半径更小、条件数更好，后续的精度矩阵估计与 VAR 回归在数值上更稳定。为消除可能存在的常数偏移，对矩阵 $\mathbf{R}$ 按行中心化：定义第 i 行的样本均值 $\bar{e}_i = \frac{1}{L} \sum_{l=1}^L R_{il}$ ，并令 $\mathbf{R}_c = \mathbf{R} - \bar{\mathbf{e}} \mathbf{1}_L^\top$ ，其中 $\bar{\mathbf{e}} = [\bar{e}_1, \dots, \bar{e}_N]^\top$ ，且 $\mathbf{1}_L$ 为长度 L 的全 1 向量。据此构造无

偏样本协方差矩阵 $\mathbf{S} = \frac{1}{L-1} \mathbf{R}_c \mathbf{R}_c^\top \in \mathbb{R}^{N \times N}$ 。当窗口长度 L 与节点规模 N 量级相近时，可采用收缩估计 $\mathbf{S}_\gamma = (1-\gamma)\mathbf{S} + \gamma \text{diag}(\mathbf{S})$ ($\gamma \in [0,1]$) 以提升稳健性，后续仍以 $\mathbf{S}$ 代表用于建模的协方差估计。本文旨在剔除其他所有节点影响之后，识别两个节点之间是否仍存在直接线性依赖的结构，这对应于高斯图模型中的条件相关关系。与皮尔逊相关不同，条件相关能有效避免共同驱动或长周期残留导致的虚假边。为此，在几何先验图 $\mathcal{G}_{\text{geo}} = (\mathcal{V}, \mathcal{E}_{\text{geo}})$ 的掩码约束下，估计残差的精度矩阵（协方差矩阵的逆） $\boldsymbol{\Theta} \in \mathbb{R}^{N \times N}$ 。优化问题写为：

$$\boldsymbol{\Theta}^* = \arg \min_{\boldsymbol{\Theta} \succ 0} \text{tr}(\mathbf{S}\boldsymbol{\Theta}) - \log \det \boldsymbol{\Theta} + \alpha \|\boldsymbol{\Theta}\|_{\text{L,off}} \text{ 且 } \boldsymbol{\Theta}_{ij} = 0 \text{ 当 } (ij) \notin \mathcal{E}_{\text{geo}} \quad (14)$$

其中， $\text{tr}(\cdot)$ 为迹，即矩阵对角线元素之和； $\det(\cdot)$ 为行列式； $\boldsymbol{\Theta} \succ 0$ 表示 $\boldsymbol{\Theta}$ 对称正定； $\text{tr}(\mathbf{S}\boldsymbol{\Theta}) - \log \det \boldsymbol{\Theta}$ 可识别为高斯图模型的负对数似然项，其中 $\log \det \boldsymbol{\Theta}$ 用于保证精度矩阵的正定性并提高估计问题的良定性； $\|\boldsymbol{\Theta}\|_{\text{L,off}} = \sum_{i \neq j} |\boldsymbol{\Theta}_{ij}|$ 表示仅对非对角元素施加 L1 范数惩罚，避免对角方差（对角线）被稀疏化； $\alpha > 0$ 为稀疏超参数，用于权衡边数量与稳健性； $\boldsymbol{\Theta}_{ij} = 0$ 强制所有不在几何邻接集合 $\mathcal{E}_{\text{geo}}$ 中的远距节点对不连边，从而将不可能的条件依赖先验裁剪，既提升可解释性，又显著降低统计方差与计算负担。该问题可用块坐标下降的图形 Lasso 算法求解，算法在每一次迭代中求解一组带软阈值的凸子问题；由于掩码将大多数条目恒置零，子问题维度远小于 N ，可高效实现。求得最优精度矩阵 $\boldsymbol{\Theta}^*$ 后，以其非零的非对角元素对应节点对作为边，即得到条件相关图 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ ，满足 $(ij) \in \mathcal{E}$ 当且仅当 $(\boldsymbol{\Theta}^*)_{ij} \neq 0$ 。需要强调的是，这里的边表示在控制其余 $N-2$ 个节点后，节点 i 与 j 之间仍存在的直接线性依赖，因此相比于简单相关更能揭示真正的邻域联动。然而， $\boldsymbol{\Theta}^*$ 刻画的是同一时刻 s 的瞬时相互作用，现实中拥塞控制反馈、用户移动与调度策略还会带来跨时刻的滞后传导。为捕捉这一动态，本文在图 $\mathcal{G}$ 的稀疏结构先验下，拟合一个低阶的向量自回归模型 $\text{VAR}(p)$ ，即：

$$\mathbf{e}(s) = \sum_{\ell=1}^p \mathbf{W}_\ell \mathbf{e}(s-\ell) + \boldsymbol{\varepsilon}(s), s = t-L+1, \dots, t \quad (15)$$

其中， $\mathbf{e}(s) \in \mathbb{R}^N$ 为时刻 s 的残差横截面， $\boldsymbol{\varepsilon}(s)$ 为零均值噪声， $\mathbf{W}_\ell \in \mathbb{R}^{N \times N}$ 为第 ℓ 阶的跨区域线性影响矩阵。为将结构先验注入并控制参数规模，对各阶滞后矩阵施加掩码稀疏约束：当 $(ij) \notin \mathcal{E}$ 时令 $[\mathbf{W}_\ell]_{ij} = 0$ ，仅允许条件相关图相连节点间存在直接的滞后影响。参数估计可构造

矩阵最小化目标: $\min_{\{W_\ell\}} \sum_s \|e(s) - \sum_{\ell=1}^p W_\ell e(s - \ell)\|_2^2 + \sum_\ell \lambda_1 \|W_\ell\|_{1,\text{mask}}$, 其中 $\|\cdot\|_{1,\text{mask}}$ 仅对允许连边位置惩罚, $\lambda_1 \geq 0$ 控制各滞后阶的稀疏度。为保证系统稳定性, 可对对应的伴随矩阵实施谱半径约束 (如在每次迭代后进行轻微缩放使谱半径小于 1), 或在目标中加入小幅岭正则以抑制极端系数并间接限制谱半径。滞后阶 p 采用交叉验证或信息准则 (AIC/BIC) 在小整数集合中选择 (通常 $p \in \{1, 2, 3\}$ 已足够), 以平衡刻画能力与方差。

得到 $\{W_\ell\}_{\ell=1}^p$ 后, 进行多步残差外推与空间校正。设时间端在第 2.2 节获得的未来 H 步初始预测为 $\hat{y}_i' = [\hat{x}_i(t+1), \dots, \hat{x}_i(t+H)]^T$, 又设本文从窗内重构中保留了最近 p 个时刻的残差历史 $e(t), e(t-1), \dots, e(t-p+1)$ 。通过 VAR 递推得到未来残差 $\hat{e}(t+1), \dots, \hat{e}(t+H)$, 并逐步叠加到时间端初始预测, 形成本区域的空间校正结果 $\hat{y}_{i,\text{final}} = \hat{y}_i' + [\hat{e}_i(t+1), \dots, \hat{e}_i(t+H)]^T$ 。

算法整体流程如图 2 所示。

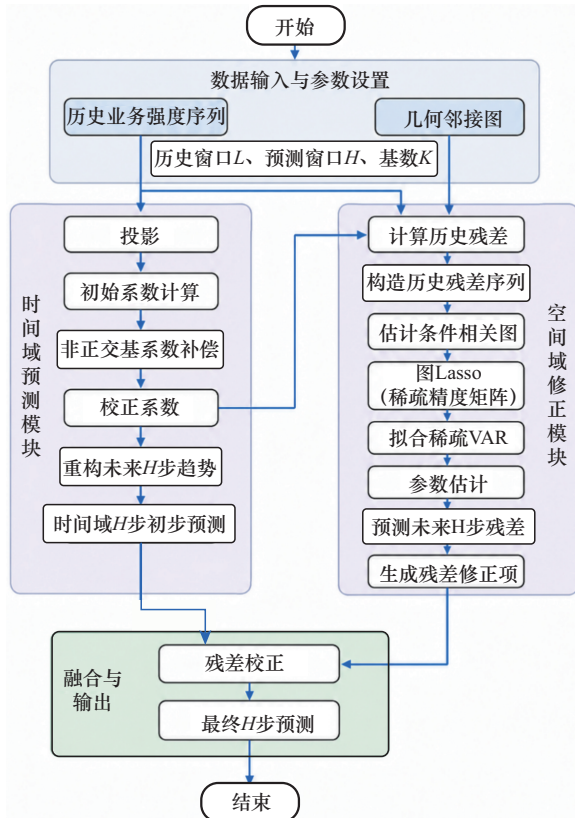


图2 算法整体流程

由于时间端主干是基字典上的线性外推, 空间端是图约束下的稀疏线性滤波, 整体推理复杂度低且路径固

定, 适合在边缘或星上设备部署。在工程实现上, 时间端的基字典与补偿 (或白化) 矩阵可离线训练并周期性刷新; 精度矩阵 Θ 与滞后矩阵 $\{W_\ell\}$ 在滑动窗内采用热启动与块坐标下降法在线更新即可, 计算代价主要与图的边数 $|\mathcal{E}|$ 与滞后阶 p 呈正比, 满足低时延与低功耗约束。

3 结果与分析

商业 LEO 星座的波束级流量数据通常涉及运营隐私, 目前鲜有公开数据集可供学术研究。考虑到前述逻辑波束与地面蜂窝小区在业务生成机制上的相似性——即二者均由该区域内的用户密度与应用行为决定, 而非取决于接入基站 (或卫星) 的物理位置, 本文选用广泛使用的米兰通信数据集 (telecom italia milan grid) 作为 LEO 下行业务强度的代理数据进行概念验证^[23]。该策略在近期多项卫星网络流量分析研究中被证实是合理且必要的^[24]。需要明确的是, 本文使用该数据集的核心目的在于验证所提出的预测模型对需求侧业务强度本身所蕴含的时空演化规律的建模与多步预测能力, 而与物理层信道特性、星上链路状态等卫星系统特有因素相关的误差源不在本文讨论范围内。未来工作将重点聚焦跨数据集的泛化能力验证, 例如, 在运营商提供的波束级统计数据或高保真仿真平台生成的波束需求轨迹上, 对所提方法进行进一步评估与优化。

3.1 数据集、区域定义与预处理

米兰通信数据集按固定地面网格离散城市区域 (常为 $100 \text{ m} \times 100 \text{ m}$ 规则栅格单元), 每天一个文本文件, 时间粒度为固定等间隔 (常为 10 min)。本文将每个非空栅格单元视为一个区域, 以索引 $i \in \{1, \dots, N\}$ 标识, N 为有效单元数 (去除全零或长期缺失单元后得到)。在离散时刻 t (每 10 min 一采样), 区域 i 的原始下行业务强度记为 $x_i(t) \in \mathbb{R}_{\geq 0}$, 具体取自数据文件的 `internet` 列。为了与卫星逻辑波束的固定地面参考系保持一致, 本文不绑定卫星编号, 而是将序列始终绑定到固定区域 ID。时间窗记为 L (回看) 与 H (预测步数), 例如 $L \in \{144, 288\}$ (1 天与 2 天历史, 10 min 粒度下分别对应 24 h 与 48 h), $H \in \{72, 144, 288\}$ (分别对应 12 h 、1 天、2 天)。

首先, 由于米兰通信数据集存在一些位置流量收集不完备的现象, 需要先对数据集进行数据质量检查。按天为每个网格单元统计缺失率和极端值占比, 其中, 极端值指超过 P99.9 或出现长时间连续为 0 的情况。如果一段时间内该单元全为 0, 但其邻域普遍活跃, 则将该时间段视为缺

失而非真实0。修补规则定义为连续缺失不超过3个点，用线性插值替换（复制前值）；连续缺失时间更长或该单元整体缺失率大于20%的样本将从评测集中剔除。由于internet序列通常服从重尾分布，本文对每个单元先在训练集上进行log1p变换，再做Z-score标准化；验证/测试集统一使用训练集的均值和方差。本文用格子中心间的欧氏距离来定义关联权重：当两格中心距离为 d_{ij} ($i \neq j$) 时取 $w_{ij} = 1/d_{ij}$ ；为限制影响范围与降噪，当 $d_{ij} > 1000$ m时直接置 $w_{ij} = 0$ （不建立连边）。这体现了距离越近影响越强、距离越远影响越弱的空间衰减规律，同时避免长距离微弱关联引入不必要的噪声与计算开销。该几何先验图仅用于图形Lasso的掩码与VAR模型的结构稀疏约束，不作为学习目标的一部分。为避免城市边界稀疏区域对图估计产生高方差，本文可仅保留活动密度高的连通子图，即过去一周内非零比例大于10%的单元。

时间切分遵循时间顺序并留出独立验证集：以数据日期区间 $[T_{\min}, T_{\max}]$ 划分为训练集/验证集/测试集，比例约为7:1:2。例如：训练集为2013-11-01至2013-12-12，验证集为2013-12-13至2013-12-19，测试集为2013-12-20至2013-12-31。所有模型的超参数均在验证集上选定，测试集只评测一次。

3.2 模型与对比设置

本文方法的时间端学习4条复值近似正交基（通过实验分析，使用4条正交基可以在计算开销与预测精度之间达到最佳平衡），并使用线性补偿消除非正交串扰，仅以外推段线性重构生成多步预测；空间端在时间端窗内重构残差基础上估计条件相关图，并在图上拟合稀疏VAR(p)模型，实现短时联动校正。优化器采用Adam（学习率 $\eta \in [1 \times 10^{-4}, 3 \times 10^{-3}]$ ），早停策略以验证集S-MAPE为准，正交性权重 λ_1 与窗内重构权重 λ_2 通过网格搜索确定。图形Lasso的稀疏系数 α 在集合 $\{1 \times 10^{-3}, 3 \times 10^{-3}, 1 \times 10^{-2}, 3 \times 10^{-2}\}$ 上选，VAR滞后阶数 $p \in \{1, 2, 3\}$ 。

为全面评估长时序预测性能，本文选取以下7种代表性方法作为对比基线。由于卫星下行业务强度预测（特别是面向时空多步预测）端到端方法的公开研究成果较少，且往往缺少可复现的代码，因此，本文参考长时序预测领域的常见做法，选取经典、成熟、已被广泛验证的主流方法作为对比基线，具体如下。

- ARIMA^[7]：经典单变量统计模型，通过自回归—差分—滑动平均对平稳序列进行建模，作为统计学对比基线。

- LSTM^[25]：典型循环神经网络方法，通过门控机制在隐状态中有选择地保留长期依赖信息、抑制梯度消失，从而建模非线性时序依赖。在长序列多步预测任务中，其通常采用编码器—解码器或序列到序列结构，将历史序列映射为未来序列，是深度学习中的经典基线模型之一。
- NLinear^[26]：将标准化后的时间序列直接映射为未来步长的单层线性预测，被广泛采用为最小可用的深度基线。来源与实现细节可参考对Transformer有效性进行系统再评估的工作。
- DLinear^[26]：先将序列分解为趋势项与季节项，再分别以线性层拟合相加，结构简洁，但在长时序预测任务上表现稳健。
- PatchTST^[27]：将原始序列切分为子序列patch作为token，并采用通道独立的Transformer编码以降低注意力计算复杂度，同时保留局部语义特征。
- CycleNet^[28]：显式学习可复用的周期成分并对残差进行预测，在存在显著周期的长序列任务中具有明显优势。
- Informer^[29]：提出ProbSparse注意力与蒸馏机制，提升长序列场景下的效率与可扩展性。

其中，上述方法的各对比数据预处理保持一致。

3.3 评价指标

为全面评估模型的预测性能与计算效率，本文采用多维度评价指标。设 $y_{i,t}$ 为区域 i 在时刻 t 的真实业务强度值， $\hat{y}_{i,t}$ 为对应的预测值。测试集包含所有地面小区的 N_{test} 个样本，每个样本预测未来 H 个时间步，覆盖 N 个区域。

本文首先在经过Z-score标准化的数据尺度上计算误差，这更能公平地反映不同模型对数据内在模式的学习能力，避免数值尺度差异带来的评价偏差。主要指标包括以下几个。

- (1) MSE：衡量预测误差的二阶矩，对异常值敏感。表示为：

$$\text{MSE} = \frac{1}{N_{\text{test}} \cdot N \cdot H} \sum_{k=1}^{N_{\text{test}}} \sum_{i=1}^N \sum_{h=1}^H (y_{i,t_k+h} - \hat{y}_{i,t_k+h})^2 \quad (16)$$

- (2) 平均绝对误差（mean absolute error, MAE）：衡量预测偏差的平均量级，具有较好的鲁棒性。表示为：

$$\text{MAE} = \frac{1}{N_{\text{test}} \cdot N \cdot H} \sum_{k=1}^{N_{\text{test}}} \sum_{i=1}^N \sum_{h=1}^H |y_{i,t_k+h} - \hat{y}_{i,t_k+h}| \quad (17)$$

- (3) 相对平方误差（relative squared error, RSE）：将

模型预测误差归一化至简单均值预测基线, $RSE < 1$ 表明模型有效。RSE 表示为:

$$RSE = \frac{\sum_{k,i,h} (y_{i,t_k+h} - \hat{y}_{i,t_k+h})^2}{\sum_{k,i,h} (y_{i,t_k+h} - \bar{y})^2} \quad (18)$$

其中, $\bar{y}$ 是训练集上所有真实值的均值。

为了评估模型在卫星上受限资源环境下的可行性, 本文对比了各模型的参数量以及理论计算复杂度。

3.4 结果分析

为直观对比不同模型在真实业务场景下的多步预测性能, 图 3 展示了 4 个代表性测试样本 (索引 #0、#40、#880、#1 500) 在 4 种代表性预测方法下的预测结果。其中, 蓝色实线代表真实业务强度, 橙色曲线为模型预测值, 灰色虚线标识历史观测与未来预测的分界。可视化结果揭示了各模型显著的性能差异: LSTM 模型 (如图 3 (b) 所示) 虽能跟随基本周期趋势, 但存在典型的过平滑现象与相位滞后, 在突变点响应迟缓且峰值处欠拟合; 基于 Transformer 架构的 PatchTST 模型 (如图 3 (c) 所示) 在处理此类序列时, 其预测曲线过于平缓, 在测试样本

#40 与 #880 的波峰、波谷处丢失了大量高频细节; 线性分解模型 DLinear (如图 3 (d) 所示) 在趋势捕获上优于前两者, 但在非线性波动强烈的区域, 其拟合紧密度仍显不足。相比之下, 本文所提方法 (如图 3 (a) 所示) 在所有测试样本中均表现出更优的拟合性能, 该方法通过自适应正交基精准刻画主导周期, 并结合稀疏 VAR 模型对局部残差进行有效修正, 不仅消除了预测滞后, 更在关键极值点与幅值变化上实现了精准重构。这种对长程趋势与瞬时突变的双重捕获能力, 直观验证了本文解耦框架处理高动态流量时所具备的鲁棒性。

为考察不同历史窗口 (L) 与预测跨度 (H) 对模型性能的影响, 不同预测情况下各模型在测试集上的性能对比见表 1。

- 场景 1 (长窗—中期): $L = 288, H = 144$ 。
- 场景 2 (长窗—长期): $L = 288, H = 288$ 。
- 场景 3 (短窗—中期): $L = 144, H = 144$ 。
- 场景 4 (短窗—长期): $L = 144, H = 288$ 。

从表 1 的实验结果中可以清晰地观察到, 本文所提方法在绝大多数预测场景下均取得最优的性能, 展现出卓

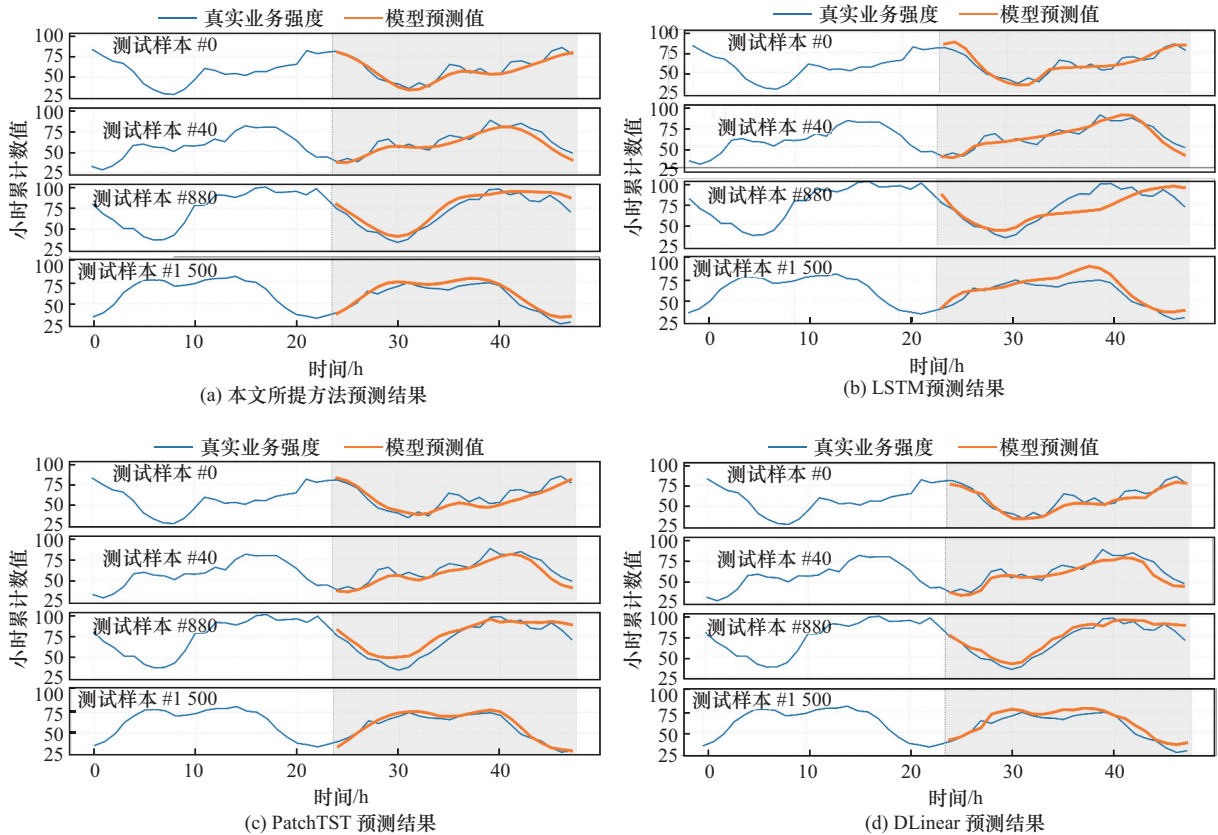


图 3 4 个代表性测试样本 (索引 #0、#40、#880、#1 500) 在 4 种代表性预测方法下的预测结果

表 1 不同预测情况下各模型在测试集上的性能对比

模型	预测情况	MSE (归一化后)	MAE (归一化后)	RSE	MSE	MAE
ARIMA	1. $L=288, H=144$	0.665 7	0.635 2	0.885 0	8.121 6	2.218 7
	2. $L=288, H=288$	0.722 0	0.732	0.915 0	8.808 4	2.556 8
	3. $L=144, H=144$	0.667 9	0.643	0.886 3	8.148 4	2.245 9
	4. $L=144, H=288$	0.724 5	0.740 2	0.916 7	8.838 9	2.585 4
LSTM	1. $L=288, H=144$	0.263 4	0.409 8	0.556 7	3.213 5	1.431 4
	2. $L=288, H=288$	0.389 8	0.500 3	0.672 3	4.755 6	1.747 5
	3. $L=144, H=144$	0.256 2	0.399 8	0.548 9	3.125 7	1.396 5
	4. $L=144, H=288$	0.400 7	0.510 3	0.681 7	4.888 7	1.782 3
NLinear	1. $L=288, H=144$	0.221 9	0.370 7	0.511 0	2.707 2	1.295 2
	2. $L=288, H=288$	0.266 6	0.403 9	0.556 1	3.252 5	1.410 7
	3. $L=144, H=144$	0.219 6	0.370 2	0.508 3	2.679 2	1.293 4
	4. $L=144, H=288$	0.288 5	0.422 7	0.578 5	3.519 6	1.476 9
DLinear	1. $L=288, H=144$	0.217 1	0.359 9	0.505 4	2.648 9	1.257 4
	2. $L=288, H=288$	0.263 6	0.399 5	0.552 9	3.216 0	1.395 4
	3. $L=144, H=144$	0.209 5	0.353 3	0.496 4	2.555 8	1.234 3
	4. $L=144, H=288$	0.278 0	0.410 6	0.567 8	3.391 6	1.434 6
PatchTST	1. $L=288, H=144$	0.381 1	0.477 2	0.669 6	4.649 0	1.666 4
	2. $L=288, H=288$	0.269 5	0.399 9	0.559 1	3.287 9	1.396 8
	3. $L=144, H=144$	0.316 2	0.436 1	0.610 0	3.857 4	1.523 1
	4. $L=144, H=288$	0.380 3	0.474 3	0.664 1	4.639 2	1.656 3
CycleNet	1. $L=288, H=144$	0.271 0	0.408 5	0.564 7	3.306 2	1.426 8
	2. $L=288, H=288$	0.322 3	0.446 7	0.611 4	3.931 9	1.560 2
	3. $L=144, H=144$	0.229 5	0.373 2	0.519 6	2.799 4	1.303 9
	4. $L=144, H=288$	0.318 3	0.440 3	0.607 6	3.883 1	1.538 0
Informer	1. $L=288, H=144$	0.353 6	0.444 8	0.645 0	4.313 5	1.553 6
	2. $L=288, H=288$	0.371 3	0.458 3	0.656 3	4.529 4	1.600 8
	3. $L=144, H=144$	0.338 5	0.419 3	0.631 0	4.129 3	1.464 5
	4. $L=144, H=288$	0.334 6	0.432 4	0.623 0	4.081 7	1.510 2
本文所提方法	1. $L=288, H=144$	0.204 3	0.356 7	0.490 3	2.492 6	1.245 7
	2. $L=288, H=288$	0.258 8	0.398 5	0.547 8	3.157 7	1.391 7
	3. $L=144, H=144$	0.203 8	0.354 8	0.489 6	2.486 5	1.239 1
	4. $L=144, H=288$	0.277 1	0.404 9	0.566 9	3.380 6	1.414 3

越的预测精度和稳定性。其中，ARIMA 作为经典的统计学模型，在所有测试场景中表现最差，其 RSE 指标均高于 0.79。这再次验证了传统线性模型难以有效捕捉卫星业务流量中固有的非线性、非平稳以及多尺度周期等复杂动态特性，尤其在长时预测任务中误差会迅速累积。相较于 ARIMA，所有基于深度学习的基线模型均具备显著优势。其中，以 DLinear 和 NLinear 为代表的纯线性分

解模型展现出强大的竞争力，其性能在多数情况下优于基于 Transformer 架构的 Informer 和 PatchTST 模型。这表明对于当前数据集，通过解耦趋势项和季节项的线性模型能够非常有效地拟合其主要变化模式。在与所有基线模型的横向对比中，本文所提方法在 4 种预测场景中的 3 种（场景 1、2、4）中均取得了所有指标的全面领先。特别是在最具挑战性的长历史窗、长跨度预测任务（ $L=$

288, $H=288$) 中, 本方法的 MSE (0.258 8) 和 RSE (0.547 8) 均是最低的, 证明了其在长程预测上的鲁棒性。在 $L=144$, $H=144$ 的短历史输入场景下, 本文所提方法的性能与表现最佳的 DLinear 模型也极为接近, RSE 指标 (0.489 6 vs 0.496 4) 依然保持最优。对比不同 L/H 设置可以发现, 所有模型的预测误差都随着预测步长 H 的增加而上升, 这符合预测任务的普遍规律。然而, 本文所提方法在不同输入长度 L 下均能保持稳定的高性能。例如, 无论使用 288 步还是 144 步的历史数据进行 144 步的预测 (场景 1 vs 场景 3), 本文所提方法均能取得约 0.49 的 RSE, 展现出对输入信息长度的良好适应能力。这主要得益于本文所提方法通过学习数据自适应的紧凑基底来表征核心时序模式, 而非依赖固定长度的滑动窗口或复杂的注意力机制, 因此能够更鲁棒地从不同长度的历史数据中提取出一致的、可用于外推的关键信息。

为验证模型核心组件的必要性, 在场景 3 ($L = 144, H = 144$) 设置下进行了消融实验。

为了验证本文所提方法中各个核心组件的重要性, 本文在场景 3 ($L = 144, H = 144$) 设置下进行了消融实验 (在此实验中, VAR 阶数设定为 $p = 3$), 分别评估时间域正交基学习 (Ortho) 与空间域残差校正 (Spatial) 模块的贡献, 模型组件消融实验结果见表 2。

表 2 模型组件消融实验结果

模型	MSE(归一化后)	MAE(归一化后)	RSE(归一化后)
本文所提方法 (不含空间校正模块)	0.209 9	0.364 3	0.496 9
本文所提方法 (不含正交约束)	0.211 5	0.369 7	0.498 8
本文所提方法 (完整)	0.203 8	0.354 8	0.489 6

- 本文所提方法 (不含空间校正模块): 仅使用时间域的可学习复值基进行预测, 不进行任何空间校正。

- 本文所提方法 (不含正交约束): 在时间域模块的训练中, 移除了对基底矩阵的正交性约束损失 (即 $\lambda_1 = 0$)。

消融实验结果清晰地表明, 去除空间校正模块后 (w/o Spatial), 模型所有精度指标均出现明显下降。这证明了业务强度在空间上并非孤立演变, 邻近区域间的短时联动是预测任务中不可忽视的重要信息。本文提出的基于残差条件相关图和稀疏 VAR 的校正方法能够有效捕获这类联动, 从而对纯时间预测的结果进行精细修正。去除正交性约束后 (w/o Ortho), 模型性能同样出现下降。正交性约束促使模型学习一组近似解耦的基向量, 每个基向量对应一种相对独立的时序模式 (如不同频率的周期), 这不仅增强了模型的可解释性, 更重要的是, 它通过减少基向量间的冗余和串扰, 使得信号的投影表示 (即系数) 更稳定和高效, 从而提升了线性重构外推的鲁棒性和准确性。

本文所提方法在预测精度之外, 其模型轻量化特性尤为突出, 这对于资源受限的实际应用场景至关重要。为了直观展示本文所提方法的参数效率, 表 3 对比了本文所提方法与主流深度学习模型 (LSTM、GRU) 及对比算法 Informer 及 PatchTST 在参数量和计算复杂度上的差异。在单变量预测实验中, 本文所提方法仅需学习 $K = 4$ 个复值基向量与极小规模线性补偿矩阵。相比百万参数 (~100 万) 的 Informer 或 PatchTST, 本方法的总参数量被控制在数万级别 (~4 万), 存储开销降低了 1~2 个数量级。Transformer 类模型的推理涉及复杂的注意力矩阵计算, 且对内存带宽要求极高。而本文所提方法的推理过程主要由线性投影 (基底内积) 和稀疏矩阵乘法 (VAR 校正) 组成, 其计算复杂度相对于输入序列长度几乎是线性的。

4 结束语

本文针对低轨多波束卫星网络动态资源调度对下行

表 3 各模型参数量与计算复杂度对比 (输入窗长 L , 预测步长 H , 隐藏层维度 d)

模型	核心架构	参数量级	训练复杂度	推理复杂度
LSTM	循环神经网络	$\sim 4(d^2 + dL)$ (15 万以上)	$O(L \cdot d^2)$ (串行计算, 慢)	$O(H \cdot d^2)$
Informer	ProbSparse Attention	$\sim 12d^2$ (编码器/解码器) (120 万以上)	$O(L \log L \cdot d)$	$O(L \log L \cdot d)$
PatchTST	Transformer + Patching	$\sim 12d^2 \times \text{Patches}$ (80 万以上)	$O(L^2 \cdot d)$	$O(L^2 \cdot d)$
本文所提方法	正交基 + 稀疏 VAR	$2K(L + H) + p \epsilon_{\text{geo}} $ (4 万以上)	$O(K(L + H))$	$O(K \cdot H)$

业务强度进行精准、高效预测的迫切需求,提出一种兼具轻量化与可解释性的时空预测方法。本文所提方法创新性地构建了一个两阶段预测框架:首先在时间域,通过学习数据自适应的复值正交基,以线性重构的方式捕捉并外推序列中主导的多尺度周期与趋势模式;随后在空间域,对时间预测的残差进行建模,通过估计稀疏条件相关图并拟合图约束的向量自回归模型,精确修正由邻近区域短时联动引起的局部波动。在公开数据集上的综合实验结果表明,本文所提方法不仅在多种预测配置下均取得了优于现有主流基线模型的预测精度,展现出卓越的稳定性和鲁棒性;更重要的是,其“分解—修正”架构赋予了模型良好的可解释性,同时极低的参数量和线性计算复杂度使它在部署效率上具备显著优势。这使得本文所提方法能够为卫星网络的BHTP等任务提供可靠的决策支持,对于提升频谱资源利用率和保障用户服务质量具有重要的实际应用价值。

参考文献

- [1] Han C, Liu A J, Huo L Y, et al. A prediction-based resource matching scheme for rentable LEO satellite communication network[J]. IEEE Communications Letters, 2020, 24(2): 414-417.
- [2] Pachler N, Del portillo I, Crawley E F, et al. An updated comparison of four low earth orbit satellite constellation systems to provide global broadband[C]//Proceedings of the 2021 IEEE International Conference on Communications Workshops (ICC Workshops). Piscataway: IEEE Press, 2021: 1-7.
- [3] Kassing S, Bhattacharjee D, Águas A B, et al. Exploring the “Internet from space” with hypatia[C]//Proceedings of the ACM Internet Measurement Conference. New York: ACM, 2020: 214-229.
- [4] Choi C S, Baccelli F. Stochastic geometry and dynamical system analysis of walker satellite constellations[J]. IEEE Transactions on Vehicular Technology, 2025: 1-6.
- [5] Fang H, Zhao H Y, Shi J X, et al. Robust live streaming over LEO satellite constellations: measurement, analysis, and handover-aware adaptation[C]//Proceedings of the 32nd ACM International Conference on Multimedia. New York: ACM, 2024: 5958-5966.
- [6] Mohan N, Ferguson A E, Cech H, et al. A multifaceted look at starlink performance[C]//Proceedings of the ACM Web Conference 2024. New York: ACM, 2024: 2723-2734.
- [7] Aasim, Singh S N, Mohapatra A. Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting[J]. Renewable Energy, 2019, 136: 758-768.
- [8] Zhu Y, Sheng M, Li J D, et al. Traffic modeling and performance analysis for remote sensing satellite networks[C]//Proceedings of the 2018 IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE Press, 2018: 1-6.
- [9] Han C, Xiong W, Yu R H. A hybrid forecasting model for self-similar traffic in LEO mega-constellation networks[J]. Aerospace, 2024, 11(3): 191.
- [10] Demirci Y, Mantelet G, Martel S, et al. Forecasting self-similar user traffic demand using transformers in LEO satellite networks[PP]. arXiv preprint, 2025, arXiv:2509.10917.
- [11] Chen Q, Zhang Y, Guo J M, et al. Traffic prediction based on surrogate model in satellite constellation networks[C]//Proceedings of the 2019 12th IFIP Wireless and Mobile Networking Conference (WMNC). Piscataway: IEEE Press, 2019: 126-130.
- [12] Cai J Y, Song S T, Zhang H P, et al. Satellite network traffic prediction based on LSTM and GAN[C]//Proceedings of the 2023 IEEE 3rd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA). Piscataway: IEEE Press, 2023: 175-178.
- [13] Liu Z G, Li W J, Feng J X, et al. Research on satellite network traffic prediction based on improved GRU neural network[J]. Sensors, 2022, 22(22): 8678.
- [14] Wang C Z, Wang L, Wei S W, et al. STN-GCN: spatial and temporal normalization graph convolutional neural networks for traffic flow forecasting[J]. Electronics, 2023, 12(14): 3158.
- [15] Vinchoff C, Chung N, Gordon T, et al. Traffic prediction in optical networks using graph convolutional generative adversarial networks[C]//Proceedings of the 2020 22nd International Conference on Transparent Optical Networks (ICTON). Piscataway: IEEE Press, 2020: 1-4.
- [16] Liu K X, Zhang Y S, Lu S. A dynamic spatio-temporal traffic prediction model applicable to low earth orbit satellite constellations[J]. Electronics, 2025, 14(5): 1052.
- [17] Li M, Wang Y W, Wang Z W, et al. A deep learning method based on an attention mechanism for wireless network traffic prediction[J]. Ad Hoc Networks, 2020, 107: 102258.
- [18] Fournier Q, Caron G M, Aloise D. A practical survey on faster and lighter transformers[J]. ACM Computing Surveys, 2023, 55(14s): 1-40.
- [19] Zhou W, Wang W, Zhao D. Passenger flow forecasting in metro transfer station based on the combination of singular spectrum analysis and AdaBoost-weighted extreme learning machine[J]. Sensors, 2020, 20(12): 3555.
- [20] Mirjalili S, Mirjalili S M, Lewis A. Grey wolf optimizer[J]. Advances in Engineering Software, 2014, 69: 46-61.
- [21] Liu C T, Pan J Y. Optimal beamwidth for maximizing uplink coverage probability in quasi earth-fixed LEO satellite communication system[J]. Electronics, 2024, 13(7): 1349.

- [22] Zhang Y, Wang Y, Cao H T, et al. Self-similar traffic prediction for LEO satellite networks based on LSTM[J]. IET Communications, 2025, 19: (1): 1-10.
- [23] Fan D H, Zhou S L, Luo J H, et al. Joint traffic prediction and handover design for LEO satellite networks with LSTM and attention-enhanced rainbow DQN[J]. Electronics, 2025, 14(15): 3040.
- [24] 马千里, 赵辉, 邵思远, 等. 一种基于业务量预测的低轨卫星波束动态开关算法[J]. 无线通信, 2023, 13(4): 35-44.
- Ma Q L, Zhao H, Shao S Y, et al. A dynamic beam switching algorithm for LEO satellite based on communication traffic volume prediction[J]. Hans Journal of Wireless Communications, 2023, 13 (4): 35-44.
- [25] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [26] Zeng A L, Chen M X, Zhang L, et al. Are transformers effective for time series forecasting? [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(9): 11121-11128.
- [27] Nie Y, Nguyen N H, Sinthong P, et al. A time series is worth 64 words: long-term forecasting with transformers[C]//International Conference on Learning Representations (ICLR). 2023.
- [28] Hu X Y, Lin S S, Lin W W, et al. CycleNet: enhancing time series forecasting through modeling periodic patterns[C]//Proceedings of the Advances in Neural Information Processing Systems 37. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2024.
- [29] Zhou H Y, Zhang S H, Peng J Q, et al. Informer: beyond efficient transformer for long sequence time-series forecasting[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35 (12): 11106-11115.

[作者简介]



郑弘毅 (2002-), 男, 北京理工大学集成电路与电子学院硕士生, 主要研究方向为智能网络和时空预测。



张婷婷 (1989-), 女, 博士, 北京理工大学信息与电子学院助理研究员, 主要研究方向为天地一体化网络、软件定义网络 and 智能网络。