

表 3-2 1 步、5 步、10 步预测下的模型性能比较

Table 3-2 Comparisons of Prediction Performances of 1, 5, 10 Steps

Steps	RMSE			MAE			MAPE		
	1	5	10	1	5	10	1	5	10
ARIMAFCN	1.446	1.773	1.988	1.299	1.452	1.651	1.308%	1.472%	1.308%
MVSTGN	0.976	1.024	1.238	0.740	0.760	0.930	0.706%	0.722%	0.884%
MVSTGN*	0.780	0.811	0.995	0.577	0.597	0.717	0.552%	0.571%	0.683%
LSTM	0.635	0.680	0.692	0.600	0.606	0.719	0.576%	0.581%	0.692%
ST-GAT	0.633	0.647	0.692	0.567	0.580	0.624	0.544%	0.558%	0.596%
GSTRGCT	0.627	0.752	0.779	0.543	0.648	0.655	0.518%	0.622%	0.626%
ASTGCN	0.609	0.612	0.668	0.541	0.531	0.594	0.518%	0.507%	0.563%
DSTGAT	0.583	0.583	0.604	0.474	0.478	0.497	0.453%	0.457%	0.475%

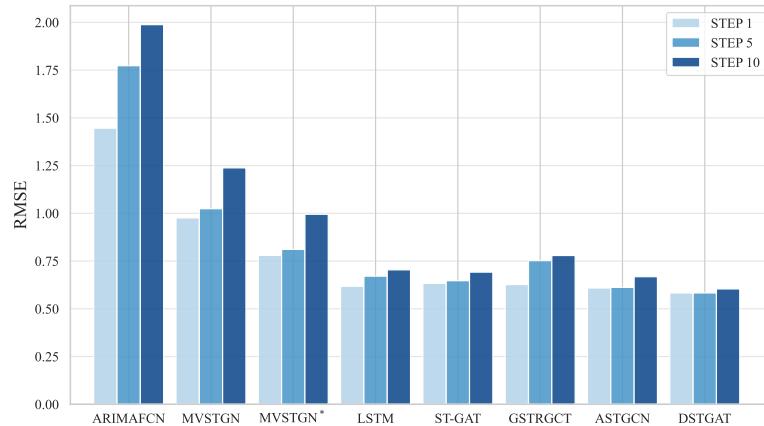


图 3-8 RMSE 横向对比

Figure 3-8 Comparative Analysis of RMSE across Different Models

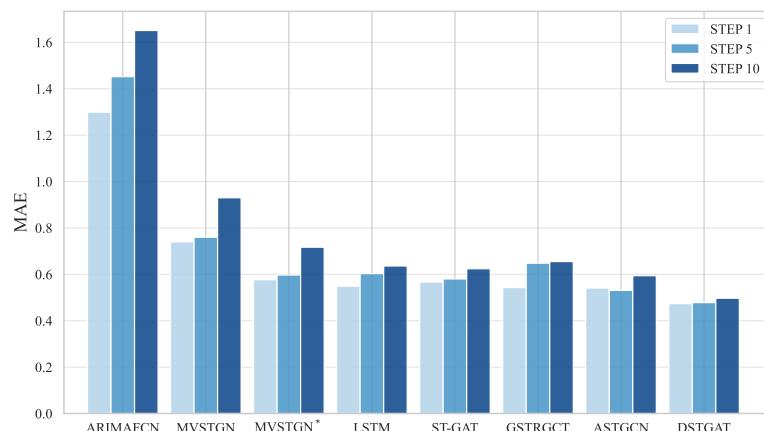


图 3-9 MAE 横向对比

Figure 3-9 Comparative Analysis of MAE across Different Models

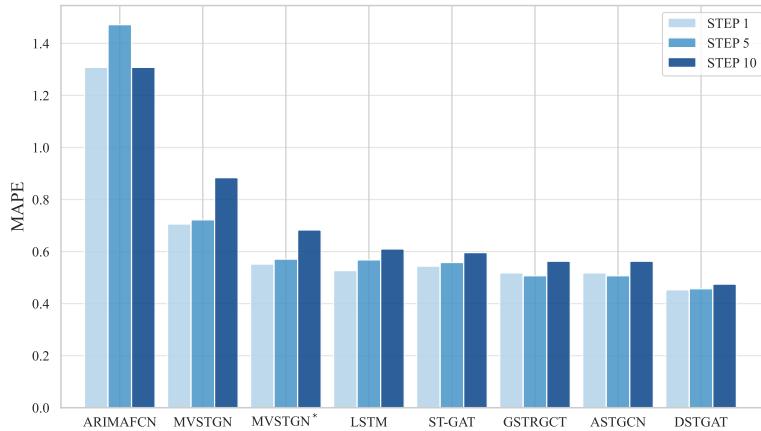


图 3-10 MAPE 横向对比

Figure 3-10 Comparative Analysis of MAPE across Different Models

任意时间步间的直接关联，有效捕捉长程依赖，规避了传统循环结构的局限，在长期预测中表现出更强的鲁棒性与稳定性。

与先进的多视图网络 MVSTGN 相比，本模型性能更优。这源于两者对“多视图”理念的诠释差异：本模型的视图是基于地理、功能等结构化先验知识构建的异构图，为模型提供了丰富的归纳偏置；而 MVSTGN 的视图则是在数据层面动态学习的相似性。对于物理关联性强的“业务-干扰”任务，结构化先验比纯数据驱动的模式相似性更为有效。

对 MVSTGN*（移除 MVSTGN 的密集卷积模块）的分析发现，更精简的模型性能反而略有提升，表明原模型的密集卷积模块可能因参数过多而对噪声产生过拟合，侧面印证了注意力机制捕捉的全局依赖是本任务的关键信号，而额外的局部精细化处理可能并非必要，甚至会干扰核心特征的表达。

与 ASTGCN 相比，本模型的空间表征更全面。ASTGCN 为每个时间分支学习单一的动态空间注意力矩阵，而本模型在时序建模前，通过四个异构邻接矩阵（地理、道路、业务、相关性）注入了多视角空间先验知识，迫使模型从不同角度学习并融合空间依赖，构建更鲁棒的特征表征。

与 GSTRGCT 相比，本模型的耦合式架构更贴合任务特性。GSTRGCT 的核心在于将时空依赖解耦，在空间和时间两个独立的超平面上分别进行建模，最后再进行融合，而“业务-干扰”关系是高度耦合的时空现象。而我们 DSTGAT 模型采用融合空间特征后再进行时序演化的级联流程，更能捕捉时空间快速、非分离的交互效应，因而超越了 GSTRGCT 的解耦框架。

图3-11可视化了各模型在 10 步预测任务上的预测曲线。可以直观地看到，DSTGAT 的预测曲线与真实值曲线（Ground Truth）的拟合程度最高，尤其在干扰的波峰和波谷处也能做出精准的预测。图3-12则展示了各模型预测绝对偏差的累积分布函数（Cumulative Distribution Function, CDF）。DSTGAT 的 CDF 曲线最陡峭，意味着其预测偏差更集中地分布在 0 附近，具有更小的预测误差和更高的稳定性。

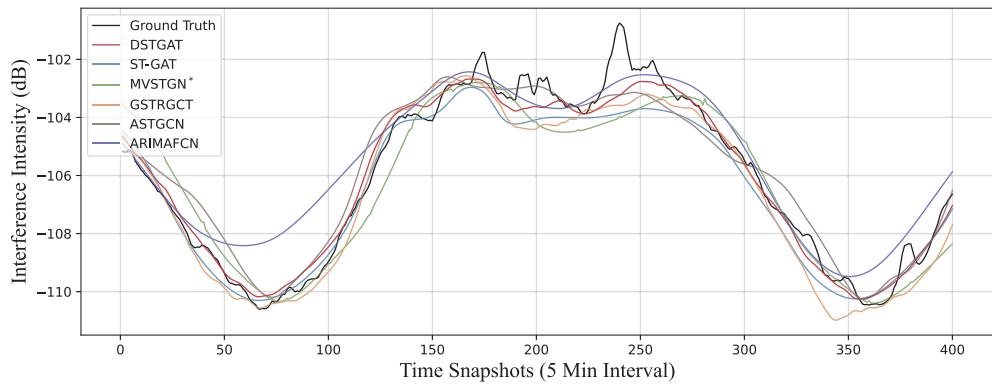


图 3-11 横向对比不同模型的预测精度可视化

Figure 3-11 Visualization of prediction across Different Models

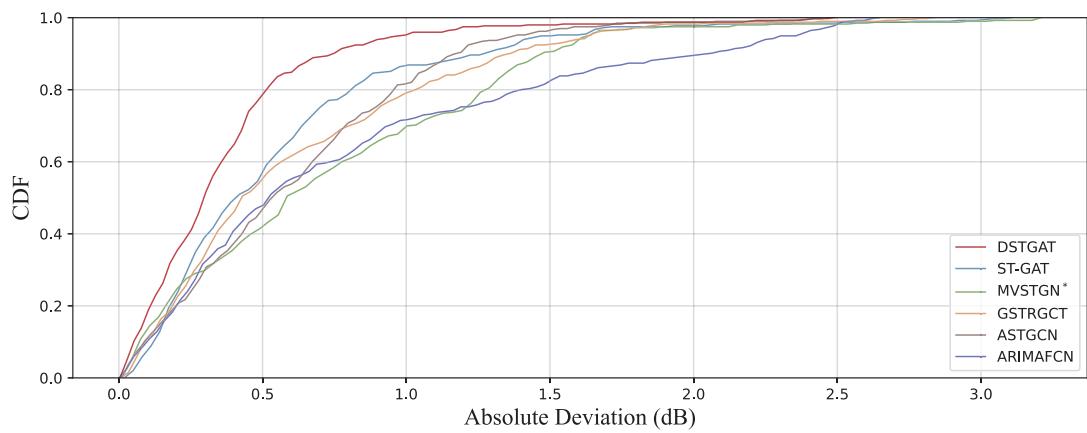


图 3-12 横向对比预测值与真实值之间绝对偏差的累积分布函数 (CDF)

Figure 3-12 CDF of the absolute deviations between the prediction and the ground truth

3.1.4.2 消融实验

为了验证 DSTGAT 模型中各个设计组件的必要性和有效性，我们进行了一系列消融实验。

- **邻接矩阵的有效性：**如表3-3所示，我们分别测试了只使用四种邻接矩阵中某一种时的模型性能。结果显示，仅使用基于道路关联的 $\mathbf{A}^{(2)}$ 时性能最佳，这印证了我们在数据分析中的判断，即主干道附近的基站业务模式确实共享相似特征，并对干扰预测有积极贡献。其次是基于数据内在特性设计的 $\mathbf{A}^{(3)}$ 和 $\mathbf{A}^{(4)}$ ，而仅使用传统欧氏距离的 $\mathbf{A}^{(1)}$ 效果最差。这表明，传统预测问题中常用的欧氏距离假设，不足以表征复杂的业务相似性。

- **时间编码的有效性：**我们还测试了移除定制化的 TE 模块后的模型性能 (DSTGAT w/o TE)。实验结果显示，移除 TE 后模型性能显著下降，证明了我们为该数据集量身打造的 TE 模块能够有效提升预测精度。

表 3-3 消融实验：1 步、5 步、10 步的预测性能

Table 3-3 Ablation Study: Prediction Performances of 1, 5, 10 Steps

Steps	RMSE			MAE			MAPE		
	1	5	10	1	5	10	1	5	10
DSTGAT	0.583	0.583	0.604	0.474	0.478	0.497	0.453%	0.457%	0.475%
DSTGAT ¹	+.084	+.064	+.091	+.048	+.051	+.044	+.045%	+.047%	+.040%
DSTGAT ²	+.020	+.036	+.029	+.011	+.016	+.003	+.000%	+.013%	+.002%
DSTGAT ³	+.035	+.083	+.035	+.020	+.051	+.011	+.017%	+.047%	+.009%
DSTGAT ⁴	+.061	+.082	+.121	+.037	+.043	+.066	+.034%	+.039%	+.074%
DSTGAT w/o TE	+.088	+.104	+.078	+.077	+.090	+.064	+.075%	+.087%	+.062%

3.1.4.3 模型效率分析

如图??所示，我们对比了各模型的训练和推理时间。可以得出结论：(1) 由于 DSTGAT 拥有更多的可学习参数（源于多特征模块），其训练时间相较于其消融版本更长。(2) 与 MVSTGN 和 GSTRGCT 等复杂的基准模型相比，DSTGAT 在训练和推理上都展现出更高的效率。这使得 DSTGAT 能够实现高效的离线训练和在线推理，有效满足实际应用中对多步预测的需求。

3.1.5 基于多重特征图的干扰预测小结

综上所述，本节提出的 DSTGAT 通过构建并融合多种异构图结构，成功地从易于获取的 PRB 利用率数据中提取了丰富的空间依赖性，实现了对频谱干扰的高精度预测。该研究验证了利用多重图视角深入刻画网络空间特性对提升预测性能的有效性。同时，通过融合定制化时间编码的自注意力机制，精确捕捉时间序列的长程依赖。

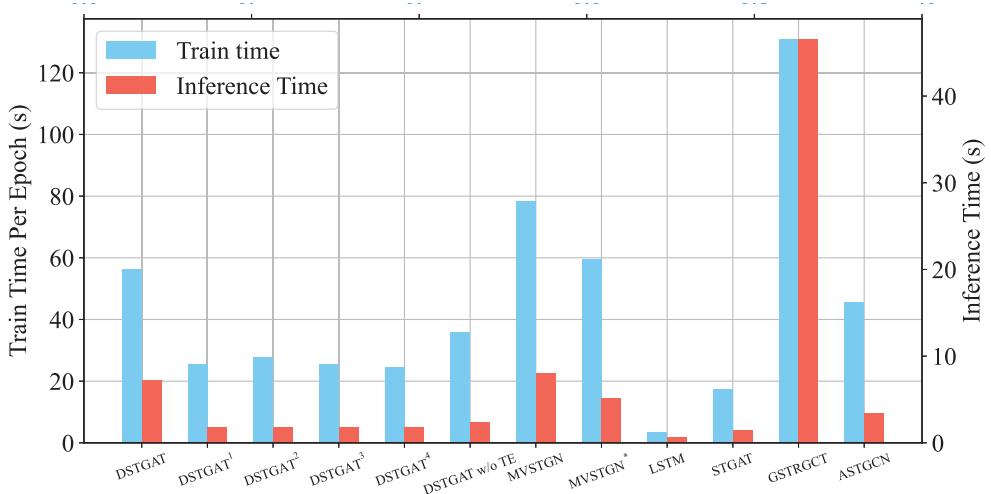


图 3-13 训练时间（每个 epoch）和推理时间比较

Figure 3-13 Train and inference time comparison

尽管 DSTGAT 模型通过融合多重特征图在预测精度上表现出色，但这一成果在一定程度上依赖于对图结构的人工设计。其根本原因在于，模型核心的图注意力机制在计算空间相关性时，其计算复杂度与网络节点数 (N) 呈二次方关系 $O(N^2)$ 。在不进行图结构预先剪枝的密集连接或全连接场景下，这种二次方复杂度的增长会带来高昂的计算成本与内存需求，严重制约了模型在真实大规模网络中的可扩展性与应用潜力。

因此，为了解决上述可扩展性挑战，本章的第二个研究重点将从“精度”转向“效率”，探索面向大规模网络的轻量化业务预测方法。接下来，我们将详细介绍融合频段先验知识与低秩注意力的时空预测模型 (FELA-Net)，该模型旨在保证预测精度的同时，显著提升大规模网络预测的可行性和效率。

3.2 面向大规模地面网络的 PRB 利用率预测

前一节的研究证实了 PRB 业务与频谱干扰的紧密关联，验证了预测干扰的可行性。为实现更主动的网络优化，本节将预测目标前移至上游的 PRB 利用率。然而，现有图注意力模型在大规模网络中面临计算效率瓶颈，限制了其部署。因此，本节聚焦于解决大规模网络中高效、精准的 PRB 利用率预测挑战，下文将首先阐述其应用场景、关键难题并进行数学建模。

3.2.1 大规模地面网络动态建模

我们将蜂窝网络抽象为一个时空图 $G = (V, E)$ ，其中 V 是一个包含 N 个基站（节点）的集合， $|V| = N$ ， E 是表示节点间连接关系的边集。在任意时间步 τ ，网络的状态由一个图信号 $\mathbf{v}_\tau \in \mathbb{R}^N$ 来表征，其中向量的每一个元素代表对应基站在该时刻的下行链路 PRB (Physical Resource Block) 利用率。

因此，蜂窝网络的 PRB 利用率预测本质上是一个时空图预测问题。形式上，

该任务旨在学习一个映射函数 \mathcal{F} , 该函数能够根据过去 P 个时间步的历史观测序列 $\mathcal{X} = (\mathbf{v}_{\tau-P+1}, \dots, \mathbf{v}_\tau) \in \mathbb{R}^{P \times N}$, 来预测未来 T_{pred} 个时间步的业务状况序列 $\mathcal{Y} = (\hat{\mathbf{v}}_{\tau+1}, \dots, \hat{\mathbf{v}}_{\tau+T_{\text{pred}}}) \in \mathbb{R}^{T_{\text{pred}} \times N}$ 。该过程可被公式化为:

$$(\hat{\mathbf{v}}_{\tau+1}, \dots, \hat{\mathbf{v}}_{\tau+T_{\text{pred}}}) = \mathcal{F}_\Theta(\mathbf{v}_{\tau-P+1}, \dots, \mathbf{v}_\tau | G) \quad (3-17)$$

其中, Θ 代表模型 \mathcal{F} 中所有待学习的参数。

然而, 传统的 GNN 方法通常依赖于一个预定义的、静态的邻接矩阵 \mathbf{A} 来描述图结构 G 。这种静态假设忽略了网络中功能性依赖关系的高度动态性——例如, 由通勤潮汐、突发事件等引起的节点间关联强度的实时变化。我们认为, 一个固定的邻接矩阵无法捕捉这些时变的复杂模式。因此, 我们模型的核心目标之一, 是摒弃静态图的约束, 转而从数据中端到端地学习一个随时间演化的 **动态空间依赖关系** $S_\tau \in \mathbb{R}^{N \times N}$, 从而更精确地对网络时空动态进行建模。

3.2.2 神经网络建模

3.2.2.1 总体架构

本节提出的时空网络模型, 我们称之为融合频段嵌入的低秩注意力网络 (Frequency-Embedded Low-rank Attention Network, FELA-Net), 其总体架构如图 3-14 所示。该模型由三个核心组件构成: 一个融合静态物理属性的输入嵌入层, 多个用于深度特征提取的堆叠式时空块, 以及一个用于直接多步预测的输出层。

更具体地说, 一个时空块包含一个低秩空间注意力模块和一个时间注意力模块, 旨在动态变化依赖关系的背景下联合学习时空特征。多个块可以堆叠形成一个深度模型, 能够捕捉更复杂和长距离的时空依赖。最后, 预测层聚合学习到的深度时空特征以生成最终的预测结果。

3.2.2.2 带有频段嵌入的输入层

为了整合基站固有的物理属性 (如其工作频段), 我们设计了一个输入模块, 将此信息与主要的流量数据相融合。该模块旨在将离散的频段类别转换为连续的、可学习的特征表示, 并将其与原始的动态时间序列特征相结合。

首先, 对于 N 个基站中的每一个, 其关联的离散频段 ID 通过一个嵌入层 $\mathcal{E}_{\text{band}}$ 映射到一个低维、密集的嵌入向量 $\mathbf{E}_{\text{band}} \in \mathbb{R}^{N \times d_{\text{band}}}$, 其中 d_{band} 是嵌入维度。

由于频段信息是静态的而流量数据是动态的, 我们将静态嵌入向量沿时间维度扩展, 以与长度为 P 的历史时间序列对齐。设原始输入时间序列特征为 $\mathbf{X}_{\text{raw}} \in \mathbb{R}^{P \times N \times F_{\text{in}}}$ 。在每个时间步 t 的融合过程公式化为:

$$\mathbf{X}_{\text{fused}}^t = \text{Concat}(\mathbf{X}_{\text{raw}}^t, \mathbf{E}_{\text{band}}) \in \mathbb{R}^{N \times (F_{\text{in}} + d_{\text{band}})} \quad (3-18)$$

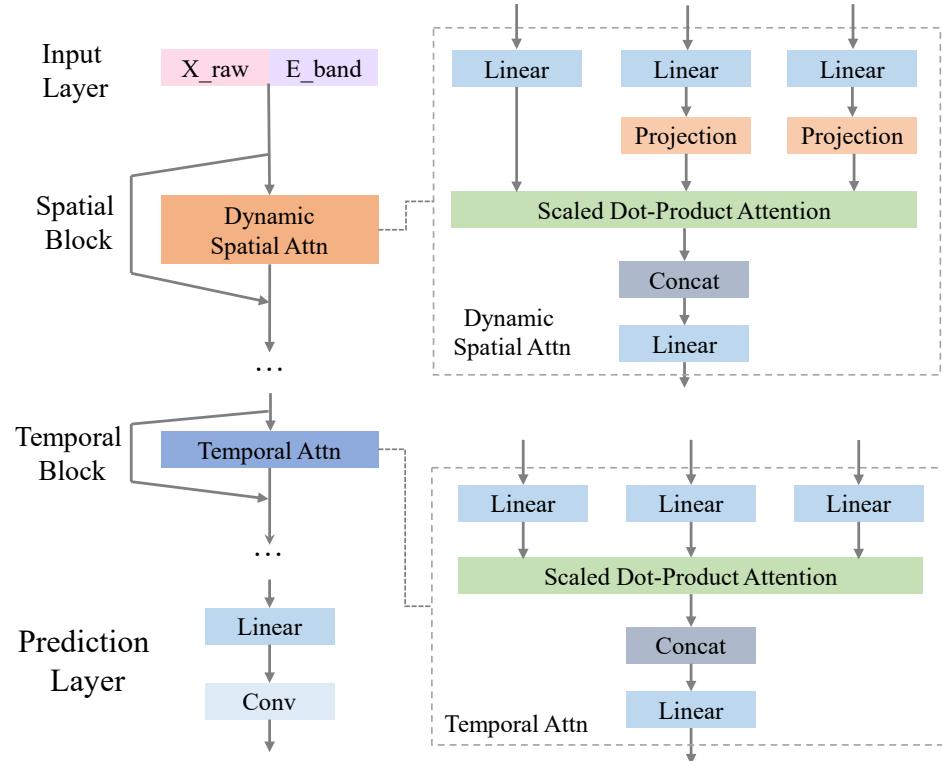


图 3-14 FELA-Net 模型架构：融合频段嵌入与分阶段低秩时空注意力

Figure 3-14 FELA-Net: Fusing Band Embedding with Staged Spatio-Temporal Attention

最后，组合后的特征序列通过一个全连接网络投影到模型的核心工作维度 D 。该输入层在每个时间步 t 的最终输出为 $\mathbf{X}_v^t \in \mathbb{R}^{N \times D}$ ，它既包含了动态的时间信息，也包含了静态的物理先验知识，为后续的时空依赖建模提供了更丰富的上下文。

3.2.2.3 时空注意力模块

(1) 带有低秩近似的动态空间注意力 该模块旨在捕捉每个时间片上节点之间动态的、瞬时的空间相关性。对于在时间步 t 的输入 $\mathbf{X}_v^t \in \mathbb{R}^{N \times D}$ ，它首先通过可学习的矩阵 $\mathbf{W}_s^q, \mathbf{W}_s^k, \mathbf{W}_s^v \in \mathbb{R}^{D \times d}$ 投影到 Query、Key 和 Value 空间。

为解决标准注意力在大型网络中 $O(N^2)$ 的复杂度问题，我们引入了一种低秩近似策略。该方法使用一个可学习的线性投影层 $\mathbf{P}_s \in \mathbb{R}^{r \times N}$ ，将 Key 和 Value 矩阵沿节点轴投影到一个更低的维度 r （其中 $r \ll N$ ）：

$$\mathbf{X}'_{v,k,s}^t = \mathbf{P}_s \mathbf{X}_{v,k,s}^t \in \mathbb{R}^{r \times d} \quad (3-19)$$

$$\mathbf{X}'_{v,v,s}^t = \mathbf{P}_s \mathbf{X}_{v,v,s}^t \in \mathbb{R}^{r \times d} \quad (3-20)$$

然后，在原始的 Query 和降采样后的 Key 之间计算注意力分数矩阵 $\mathbf{E}_s^t \in \mathbb{R}^{N \times r}$ ，从而将复杂度降低到 $O(N \cdot r)$ ：

$$\mathbf{E}_s^t = \frac{\mathbf{X}_{v,q,s}^t (\mathbf{X}'_{v,k,s}^t)^T}{\sqrt{d}} \quad (3-21)$$

每个注意力头的输出是通过将经过 softmax 归一化的分数应用于降采样后的 Value 矩阵得到的。来自多个头的结果被拼接起来，形成最终输出，表示为 $\mathbf{X}_{v,\text{dynamic}}^t \in \mathbb{R}^{N \times D}$ 。

(2) 时间注意力在空间注意力模块之后，得到的具有空间上下文的特征序列 $\mathcal{H}_{\text{spatial}} \in \mathbb{R}^{P \times N \times D}$ 被送入时间注意力模块。该模块负责捕捉不同历史时间步之间的动态相关性。

为实现这一点，输入张量被视为 N 个长度为 P 的独立时间序列。对每个序列应用标准的多头自注意力机制。序列首先被投影到 Query、Key 和 Value 空间：

$$\begin{aligned} Q_{t,s} &= \mathcal{H}_{\text{spatial}} \mathbf{W}_{t,s}^q, \\ K_{t,s} &= \mathcal{H}_{\text{spatial}} \mathbf{W}_{t,s}^k, \\ V_{t,s} &= \mathcal{H}_{\text{spatial}} \mathbf{W}_{t,s}^v \end{aligned} \quad (3-22)$$

其中 $\mathbf{W}_{t,s}^q, \mathbf{W}_{t,s}^k, \mathbf{W}_{t,s}^v \in \mathbb{R}^{D \times d}$ 是特定于时间注意力的可学习投影矩阵。然后计算注意力分数矩阵 $\mathbf{E}_{t,s} \in \mathbb{R}^{P \times P}$ ：

$$\mathbf{E}_{t,s} = \frac{Q_{t,s}(K_{t,s})^T}{\sqrt{d}} \quad (3-23)$$

单个注意力头 s 的输出是通过将经过 softmax 归一化的分数应用于值矩阵 $V_{t,s}$ 得到的：

$$\text{head}_s = \text{softmax}(\mathbf{E}_{t,s})V_{t,s} \quad (3-24)$$

来自所有 S 个头的输出被拼接并通过一个最终的线性投影层 $\mathbf{W}_{t,fc} \in \mathbb{R}^{(S \cdot d) \times D}$ 进行融合，以产生该模块的最终输出 $\mathbf{H}_{\text{temporal}} \in \mathbb{R}^{P \times N \times D}$ ：

$$\mathbf{H}_{\text{temporal}} = \text{Concat}(\text{head}_1, \dots, \text{head}_S)\mathbf{W}_{t,fc} \quad (3-25)$$

该输出动态地重新加权每个时间步，使模型能够专注于对未来预测最关键的历史时刻。

3.2.2.4 堆叠结构与预测层

为了捕捉更深层、更复杂的时空依赖关系，我们通过堆叠多个时空块来构建一个深度架构。我们的模型采用分阶段的处理流程：它首先通过一个空间注意力堆叠来深度挖掘空间相关性，然后使用一个时间注意力堆叠来建模时间动态。设第 I 个块的输入为 $\mathbf{X}^{(I-1)}$ 。特征序列首先通过 L_s 个空间块来提炼空间表示，产生一个富含空间信息的特征图 $\mathbf{H}_{\text{spatial}}^{(L_s)}$ 。随后，该特征图被 L_t 个时间块处理以建模时间动态。这种分阶段的深度堆叠结构，辅以残差连接，使我们的模型能够分层地提取复杂的时空特征，产生最终的表示 $\mathbf{H}_{\text{final}} \in \mathbb{R}^{P \times N \times D}$ 。

在深度特征提取之后，设计了一个预测层来生成多步预测。首先，一个可学习的线性投影层沿时间维度聚合信息，将长度为 P 的特征序列压缩成一个单一的、有代表性的特征向量 $\mathbf{H}_{\text{agg}} \in \mathbb{R}^{N \times D}$ ：

$$\mathbf{H}_{\text{agg}} = \mathbf{H}_{\text{final}} \mathbf{W}_{\text{proj}} \quad (3-26)$$

其中 $\mathbf{W}_{\text{proj}} \in \mathbb{R}^{P \times 1}$ 是一个可学习的投影矩阵。最后，这个聚合后的特征图被送入一个轻量级的卷积模块，该模块由一个 BatchNorm 层、一个 ReLU 激活函数和一个 1×1 卷积组成。 1×1 卷积作为一个高效的映射，将深度特征转换为期望的 T_{pred} 预测长度。最终的预测 $\hat{\mathbf{X}} \in \mathbb{R}^{N \times T_{\text{pred}}}$ 因此被直接生成，这避免了自回归模型中固有的误差累积问题。

3.2.3 负载预测精度分析

为验证 FELA-Net 模型在大规模网络 PRB 利用率预测任务上的有效性与高效性，本节将进行一系列对比与消融实验。

3.2.3.1 数据集处理与实验设置

本实验采用与前述研究相同的 PRB 利用率数据集，该数据集包含 400 个基站的 5 分钟粒度 PRB 利用率时间序列。需要强调的是，本节的任务是直接进行 PRB 利用率预测，因此不使用频谱干扰数据。数据的预处理（包括线性插值、异常值处理）与数据集划分（训练集 70%、验证集 15%、测试集 15%）策略均与章节3.1.3.1保持一致。

在预测任务中，我们同样使用过去 4 小时（48 个时间步）的历史数据，以预测未来 5、25 及 50 分钟（即 1、5、10 个时间步）的 PRB 利用率。所有输入特征使用 Min-Max 缩放归一化到 [0,1] 区间。FELA-Net 模型采用 Adam 优化器进行端到端的训练，训练周期为 100 个 epoch，批处理大小（Batch Size）为 20，初始学习率设置为 1e-3。针对本模型核心的低秩注意力机制，其关键超参数秩（Rank） r 经过在验证集上的调优，最终设置为 $r=64$ 。其他详细参数配置见表??。模型的性能评估同样采用章节3.1.3.3中定义的 RMSE、MAE 和 MAPE 指标，所有预测值在计算误差前均被反归一化至原始尺度。

3.2.3.2 基线模型

为全面评估 FELA-Net 的性能，我们选取了多种具有代表性的时间序列预测算法作为基线模型进行对比：

- 时空图卷积网络（STGCN）(2017)^[89] 开创性地建立了全卷积式的时空预测范式，其核心架构耦合了用于空间的图卷积与用于时间的门控时间卷积，避免了 RNN 的迭代计算，实现了高效的并行训练。
- 注意力时空图卷积网络（ASTGCN）(2019)^[83] 通过设计三个独立分支来并行建模近期、日、周周期性依赖，每个分支均包含时空注意力模块与时空卷积模块，最终加权融合生成预测。
- 多元时间序列图神经网络（MTGNN）(2020)^[90] 则解决了图结构未知的问题，其通过图学习层自适应地学习邻接矩阵，并结合混合跳传播图卷积与膨胀时间卷积来捕捉时空依赖。
- 时空 Transformer 网络（STTN）(2020)^[85] 创新地将 Transformer 架构同时应用于时空两个维度，其空间 Transformer 与时间 Transformer 分别用于学习节点

间的动态有向依赖和时间维度的长程双向依赖。

- 多视图时空图网络 (MVSTGN) (2023)^[87] 通过解耦方式从多个视图挖掘时空特性, 它利用全局空间与时间注意力捕捉宏观关联, 并采用密集卷积挖掘局部依赖, 代表了一种全面捕捉时空相关性的范式。
- LightST (2025)^[91] 框架则应用知识蒸馏来解决效率瓶颈, 它将一个复杂 GNN 教师模型的时空知识迁移至一个轻量级 MLP 学生模型中, 通过预测分布对齐和特征表示对齐, 使轻量模型也能捕捉复杂时空依赖。

3.2.3.3 评价指标

为全面评估模型性能, 我们采用 RMSE、MAE 与掩码平均绝对百分比误差 (Masked Mean Absolute Percentage Error, MMAPE) 三项指标。RMSE 与 MAE 已在章节3.1.3.3中定义。鉴于标准 MAPE 在处理含大量零值的 PRB 利用率数据时不稳定, 我们引入 MMAPE 以更侧重于对高业务量事件的预测评估。MMAPE 通过掩码处理, 忽略所有真实值 y_i 低于阈值 τ 的样本。令 $M = \{i \mid y_i \geq \tau\}$ 为有效样本索引集合, 其计算公式如下:

$$\text{MMAPE} = \frac{100\%}{|M|} \sum_{i \in M} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (3-27)$$

其中, $|M|$ 是集合 M 中元素的数量, y_i 是真实值, \hat{y}_i 是预测值。此公式通过仅计算高于阈值的样本误差, 为模型在难以预测的高业务量事件上的性能提供一个更集中、更具代表性的评估。

3.2.3.4 模型设计的数据驱动动机

(1) 频段异质性分析

为探究蜂窝流量时间序列的内在特性, 我们采用近似熵来量化其规律性与可预测性。ApEn 通过衡量时间序列在维数从 m 增加到 $m + 1$ 时产生新模式的概率来评估数据复杂性。对于一个长度为 N 的时间序列 $\{u(1), u(2), \dots, u(N)\}$, 其 ApEn 值计算如下:

$$\text{ApEn}(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r) \quad (3-28)$$

其中, m 是嵌入维度, r 是相似性容限, N 是时间序列长度。 $\Phi^m(r)$ 和 $\Phi^{m+1}(r)$ 分别代表在维度 m 和 $m + 1$ 下, 所有子序列向量与其相似向量的条件概率的自然对数平均值。

该公式的值反映了时间序列的复杂性: 值越低, 规律性越强, 可预测性越高; 值越高, 则表示序列越随机、越复杂。

我们将 400 个基站按工作频段分为三组: N28(低频, 37 站)、N41(中频, 347 站) 与 N79(高频, 16 站)。各基站 PRB 流量序列的 ApEn 值分布如图3-15所示。图3-15揭示了不同频段间流量复杂度的显著异质性: N79 频段 ApEn 值最低且分布最集中, 表明其流量模式高度规律; 而 N28 与 N41 频段 ApEn 值更高且分布

更广，说明其流量模式更为复杂。这一发现表明，工作频段是影响流量动态的关键物理因素。将频段作为先验知识特征输入模型，有助于学习更具区分度的时空依赖关系，从而提高预测精度。

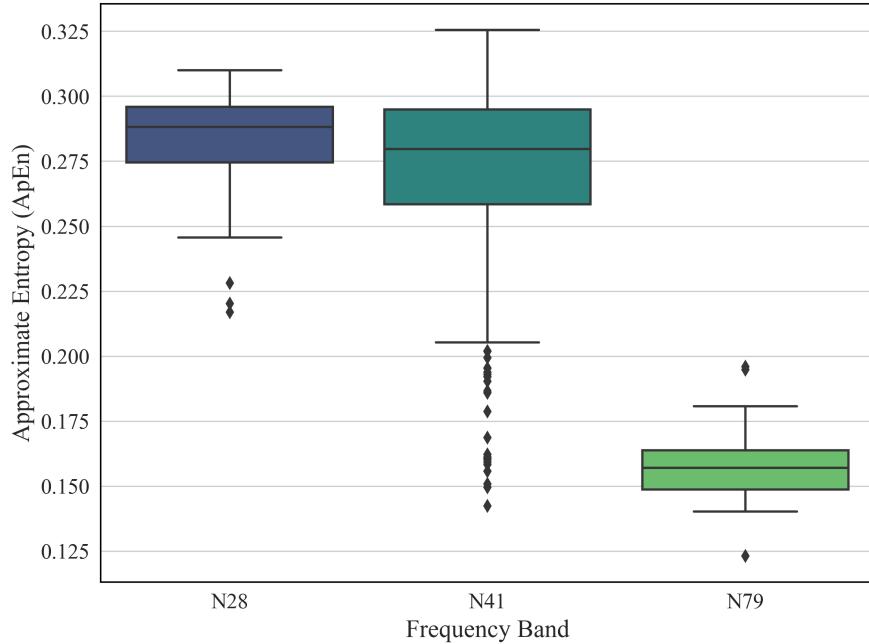


图 3-15 各频段基站流量的近似熵分布

Figure 3-15 Distribution of ApEn for Traffic Series across Different Frequency Bands

(2) 低秩近似的数据驱动动机

标准自注意力机制虽能捕捉动态空间依赖，但其 $O(N^2)$ 的计算与内存复杂度使其难以应用于大规模图。为探究空间注意力矩阵的内在属性，我们将其在不同批次下的热力图进行可视化，如图 3-16 所示。结果表明，注意力模式具有高度动态性，且分数集中于少数列，揭示了矩阵中存在显著的冗余与优化潜力。

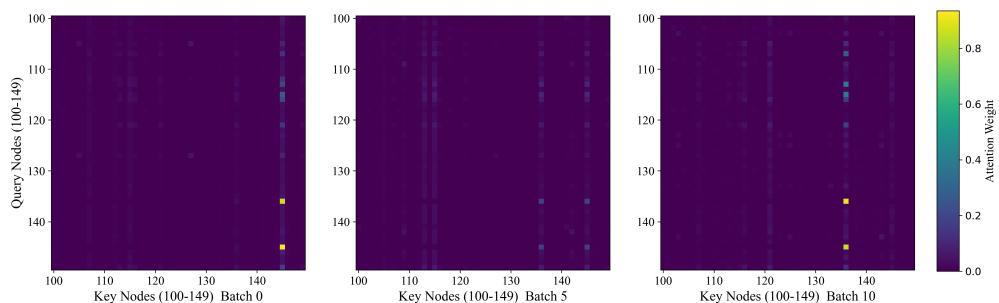


图 3-16 不同批次的空间注意力热力图

Figure 3-16 Spatial attention heatmaps for different batches

为量化该特性，我们对平均空间注意力矩阵 $\mathbf{A} \in \mathbb{R}^{N \times N}$ 进行奇异值分解 (SVD) 分析。矩阵的 SVD 分解定义为：

$$\mathbf{A} = \mathbf{U} \Sigma \mathbf{V}^T \quad (3-29)$$

式中, \mathbf{U} 和 \mathbf{V} 为正交矩阵, $\boldsymbol{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_N)$ 为对角矩阵, 其对角线元素按降序排列的奇异值。该公式将原始矩阵分解为一组正交基与奇异值谱, 从而揭示其内在秩。奇异值谱如图 3-17(a) 所示, 呈现急剧衰减趋势, 表明矩阵具有高度低秩特性。为量化此性质, 我们计算了前 k 个奇异值的累积能量贡献率 $\mathcal{E}(k)$:

$$\mathcal{E}(k) = \frac{\sum_{i=1}^k \sigma_i^2}{\sum_{i=1}^N \sigma_i^2} \quad (3-30)$$

公式中, σ_i 为第 i 个奇异值, N 为矩阵维度。该指标衡量了由前 k 个最大奇异值所捕获的矩阵信息总量。图 3-17(b) 显示, 超过 99% 的总能量集中于前 5 至 6 个奇异值。此发现证实, 信息量最丰富的空间关系仅存在于所有节点对的一个极小子集中, 为采用低秩近似方法以降低计算复杂度并实现模型正则化提供了依据。

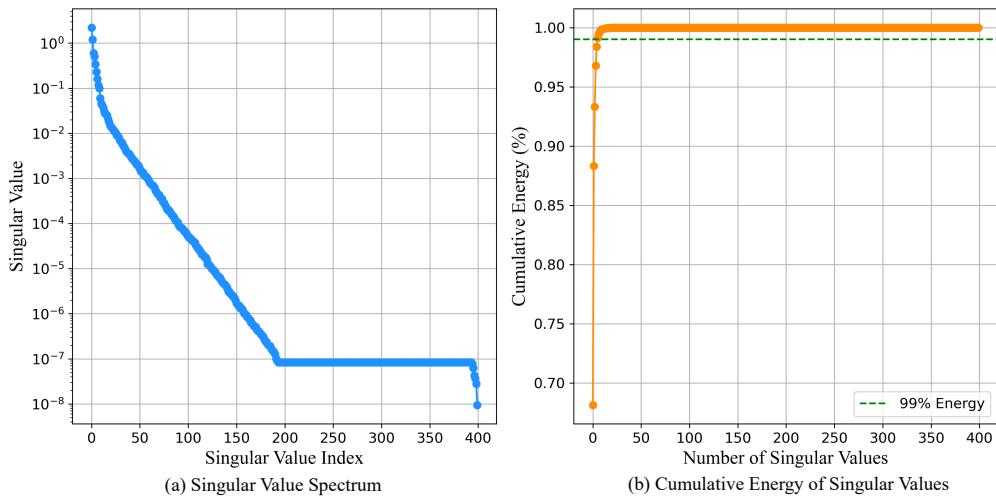


图 3-17 通过奇异值分解验证空间注意力的低秩特性

Figure 3-17 Verifying the Low-Rank Property of Spatial Attention via SVD

为探究注意力机制学习到的空间模式, 我们将注意力得分最高的 K 个连接可视化于地图, 如图 3-18 所示。图中呈现出显著的“枢纽-辐射”模式: 少数关键节点作为枢纽, 吸引了大量来自其他节点的连接。这种高度集中的分布是空间注意力矩阵低秩特性的有力证据, 因为它表明复杂的节点间关系可由少数关键节点的模式所概括。这些关键节点通常对应于地理上的交通枢纽或主干道, 其承载的高密度交通流产生了独特且信息丰富的时序模式, 从而使其与众多节点间建立了强相关性并被赋予高注意力权重。

3.2.3.5 性能对比

如表3-4所示, FELA 模型在所有指标和步长上均显著优于基线。具体而言, 在单步预测中, FELA 的 RMSE 为 0.655, MAE 为 0.461, 与表现强劲的次优模型 MVSTGN 相比, 误差降低了约 50%。随着预测步长的增加, 所有模型的误差

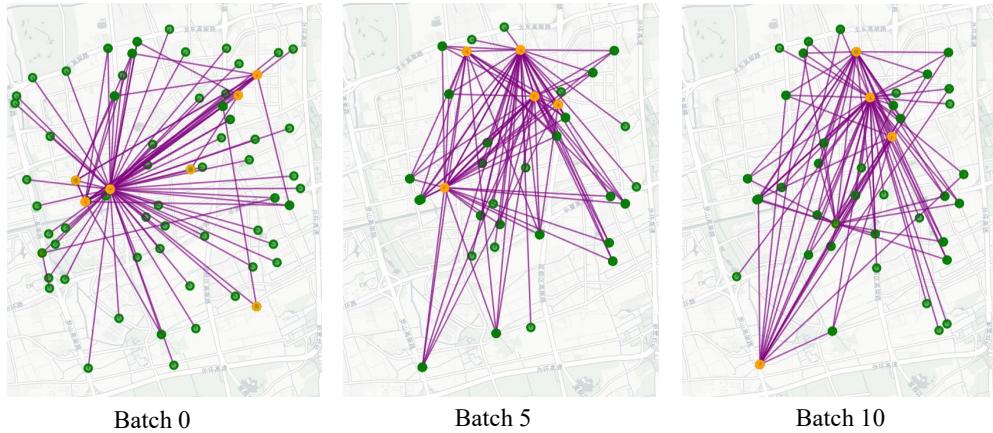


图 3-18 空间注意力机制学习到的“枢纽-辐射”模式可视化

Figure 3-18 Visualization of the Learned "Hub-and-Spoke" Pattern in Spatial Attention

都呈上升趋势，但 FELA 始终保持着最低的误差水平，证明了其在长期预测中的鲁棒性。

值得注意的是，MVSTGN 和 MTGNN 作为强大的基线模型，其优异性能分别得益于其多视角图学习和自适应依赖建模能力。然而，相比之下，它们在长期预测上的性能衰减比 FELA 更为明显。这凸显了 FELA 的架构在更有效地捕捉长期动态变化和复杂时空依赖方面的优势。

其他基线模型的表现则反映了该领域的技术演进。LightST 作为 MTGNN 的知识蒸馏版本，以微小的精度损失换取了模型复杂度的显著降低，验证了效率与性能权衡的可行性。而开创性的工作 STGCN 及其改进版 ASTGCN，由于其相对简单的静态图卷积和时间模型，性能处于中游。STTNs 的表现则表明，早期的纯 Transformer 架构若未能与图结构信息有效融合，在处理具有强空间结构性的数据时可能存在局限。综上，这些对比结果有力地证明了 FELA 模型设计的先进性。

为评估模型在整个测试集上的连续追踪性能，图 3-19 可视化了各模型在不同预测视界下的预测结果。具体而言，该曲线是通过将测试集中每个连续时间点的样本所对应的特定一步预测值连接而成。子图 (a)、(b) 和 (c) 分别对应在预测未来第 1 步、第 5 步和第 10 步时的性能对比。从图中可以直观地看出，我们提出的 FELA-Net 的预测轨迹在所有场景下都最紧密地跟随真实值的动态变化，证明了其卓越的长期稳定性和泛化能力。

3.2.3.6 消融实验

为验证我们所提模型中关键组件的有效性并探索架构变体，我们进行了一系列消融研究。我们将完整模型 FELA 与其三个变体进行比较：1) FELA (w/o FE)，消融了频段嵌入模块；2) FELA (w/o LA)，通过采用全注意力机制移除了低秩近似；以及 3) FELA (Alt-ST)，一个采用交错时空块结构的架构变体。全面的结果呈现在表 3-6 和表 3-7 中，通过分析我们得到以下三个结论：

表 3-4 多步预测任务上的预测精度对比

Table 3-4 Comparison of prediction accuracy on multi-step prediction tasks

模型	RMSE (1e-3)			MAE (1e-3)			MMAPE (%)		
	1	5	10	1	5	10	1	5	10
STTNs	7.879	11.159	10.465	5.811	8.294	7.568	2.704	5.548	5.671
STGCN	3.317	4.045	4.163	2.027	2.565	2.757	1.663	2.106	2.269
ASTGCN	2.290	2.720	3.990	1.220	1.570	2.970	0.996	1.450	2.315
LightST	2.285	2.927	3.090	1.499	1.881	1.747	1.228	1.710	1.479
MTGNN	1.872	3.243	1.757	1.072	2.115	1.169	0.853	1.781	0.937
MVSTGN	1.320	1.190	1.490	1.320	0.780	0.940	0.761	0.745	0.799
FELA (Ours)	0.655	0.941	1.421	0.461	0.662	0.935	0.431	0.545	0.862

表 3-5 与基线模型的计算效率对比

Table 3-5 Comparison of computational efficiency with baseline models

模型	训练时间 (秒/周期)	测试时间 (秒)
MVSTGN	91.69	4.22
STTNs	68.78	4.47
ASTGCN	55.29	3.30
FELA (Ours)	41.80	2.06
MTGNN	22.03	2.22
LightST	15.75	0.58
STGCN	12.42	0.81

表 3-6 FELA-Net 的消融实验以及变体

Table 3-6 Ablation experiments of FELA-Net and its variants

Model	RMSE (1e-3)			MAE (1e-3)			MMAPE (%)		
	1	5	10	1	5	10	1	5	10
FELA (w/o FE)	1.320	1.930	1.900	0.950	1.510	1.210	0.918	1.331	1.144
FELA (w/o LA)	1.323	1.580	2.272	0.951	1.199	1.719	0.918	1.018	1.408
FELA (Alt-ST)	0.783	1.004	1.566	0.559	0.707	1.017	0.531	0.634	0.988
FELA (Ours)	0.655	0.941	1.421	0.461	0.662	0.935	0.431	0.545	0.862

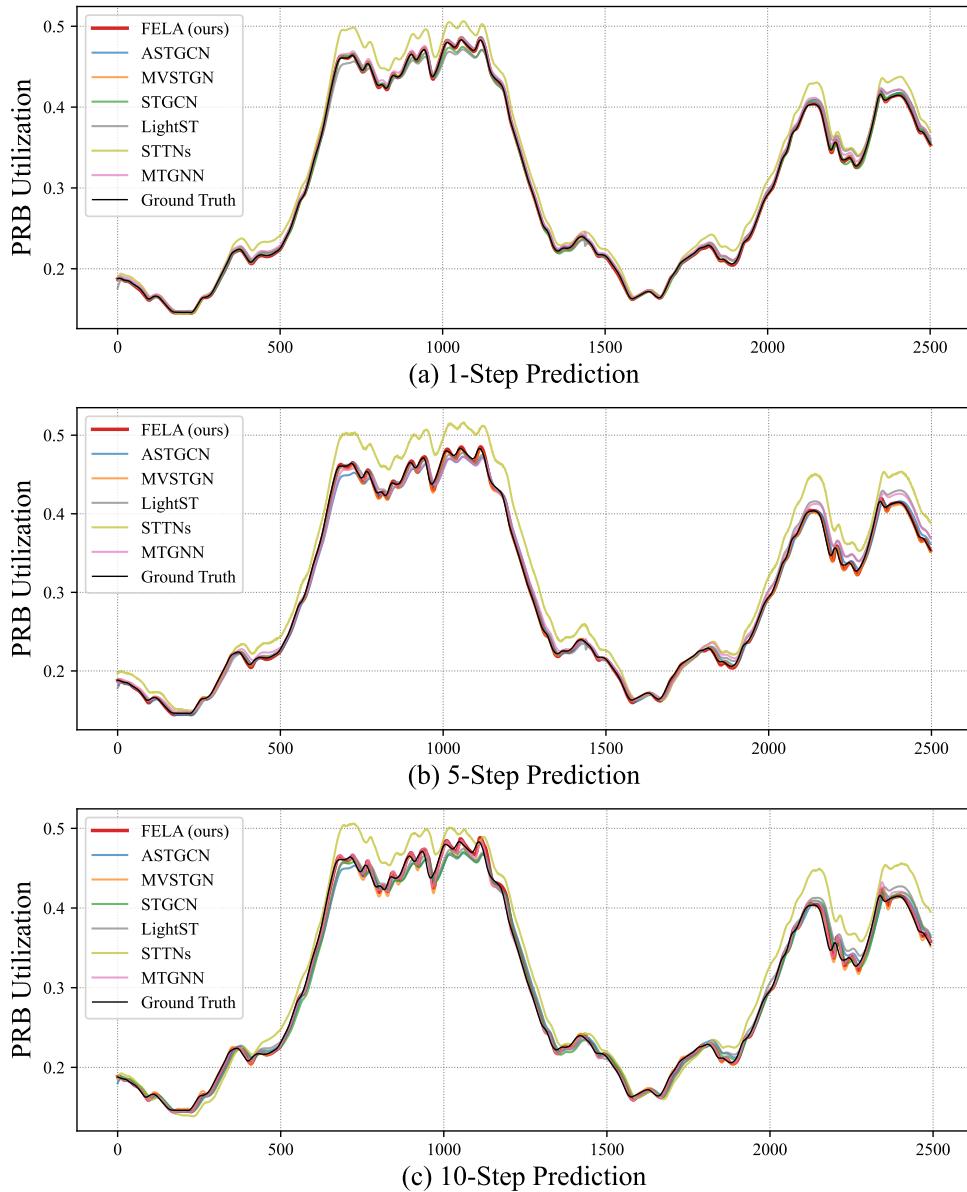


图 3-19 不同预测步数下的预测值可视化

Figure 3-19 Visualization of predicted values at different prediction steps

表 3-7 FELA-Net 的消融模型以及变体的计算效率

Table 3-7 Ablation model of FELA-Net and computational efficiency of its variants

模型	训练时间	测试时间
FELA (w/o FE)	41.58	2.03
FELA (w/o LA)	78.88	4.00
FELA (Alt-ST)	44.01	3.09
FELA (Ours)	41.80	2.06

为验证频段嵌入 (FE) 模块的有效性，我们将 FELA 与移除该模块的变体 FELA (w/o FE) 进行比较。实验结果表明，尽管 FELA (w/o FE) 因架构更简洁而获得了最高的推理效率，但这种效率提升是以预测精度的显著下降为代价的。在 5 步预测任务中，移除 FE 模块导致 RMSE 指标上升约 25.3%。该结果证实，融合物理先验知识至关重要，频段嵌入为模型提供了关键的上下文信息，使其能够学习更具辨识度的空间模式，从而提升预测精度。

为评估低秩近似 (LA) 策略的效果，我们对比了 FELA 与采用全秩注意力的变体 FELA (w/o LA)。在计算效率方面，FELA (w/o LA) 的训练时间几乎翻倍，验证了本方法的计算高效性。值得注意的是，全秩模型的预测精度反而全面下降，这一现象印证了 SVD 分析的先验假设，即空间注意力矩阵具有内在的低秩属性。尽管全秩矩阵表达能力更强，但也更容易对数据噪声及伪相关产生过拟合。因此，低秩近似不仅是一种计算优化手段，更是一种有效的正则化器，它迫使模型聚焦于主导性空间模式，从而提升泛化能力。

为验证所提时空块 (ST-block) 设计的合理性，我们测试了采用空间和时间层交错结构的变体 FELA (Alt-ST)。实验结果表明，该交错变体的计算成本与预测误差均略有上升。此性能下降表明，原始的堆叠式块设计更为有效，这可能是因为堆叠结构能够更清晰地构建分层式时空特征表示，而交错结构可能干扰这一层次化学习过程。该对比验证了所提堆叠式架构在本预测任务中的优越性。

3.2.3.7 超参数敏感性分析：最优秩数的确定

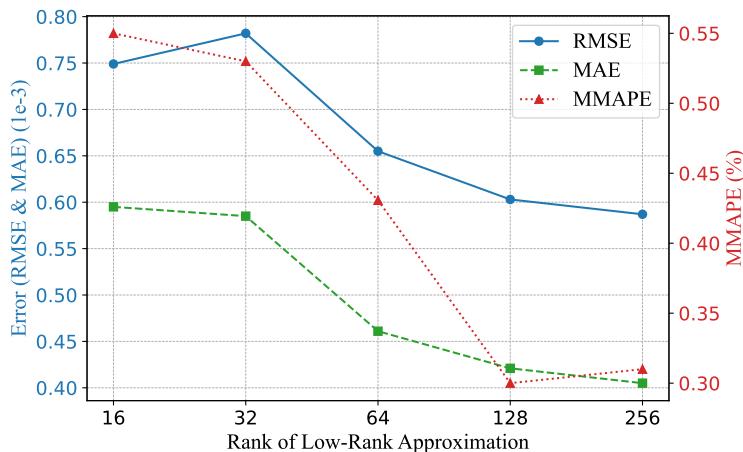


图 3-20 模型预测精度对秩超参数的敏感性分析

Figure 3-20 Sensitivity Analysis of Model Prediction Accuracy to the Rank

为确定低秩近似策略的最优秩，我们进行了超参数敏感性分析。如图 3-20 所示，当秩从 16 增至 64 时，模型预测误差显著下降；此后，性能趋于饱和，精度增益微乎其微。同时，图 3-21 表明，模型计算成本（包括训练与推理时间）与秩正相关，且在秩超过 64 后增长更为迅速。

综合权衡预测精度与计算效率，所有主实验均采用秩为 64 的设定。该设定

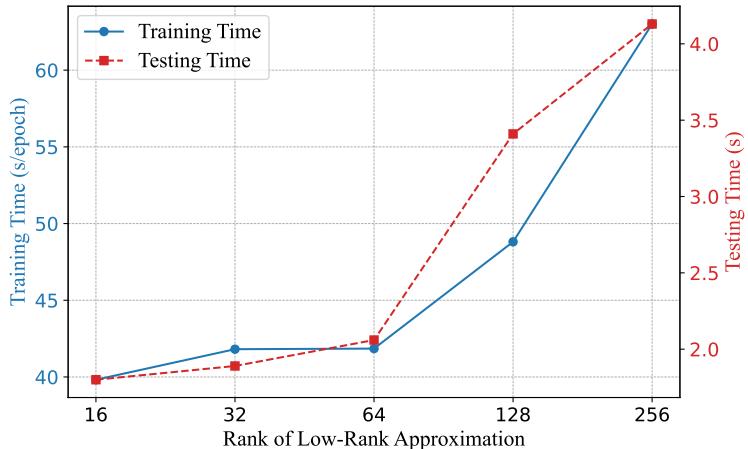


图 3-21 模型计算效率对秩超参数的敏感性分析

Figure 3-21 Sensitivity Analysis of Model Computational Efficiency to the Rank

可在可控计算开销下捕捉绝大多数主导性空间依赖模式，并有效避免过高秩带来的资源浪费与潜在过拟合风险，从而在性能与效率间达成最佳平衡。

3.2.4 本节小结

本节提出时空预测模型 FELA-Net，通过融合频段先验知识与低秩空间注意力机制，应对大规模网络预测中精度与效率的挑战。真实世界数据集实验表明，FELA-Net 在短中期预测中性能优越，关键误差指标降低超过 21.2%。该模型同时实现了计算效率的显著提升，体现了低秩设计作为计算加速器与有效正则化器的双重优势。

3.3 本章总结

本章聚焦于地面蜂窝网络的时空状态预测技术，旨在构建对频谱资源态势的全面认知。我们围绕网络业务负载与电磁环境状态这两个核心组成部分，分别从预测精度和模型效率两个维度展开研究，提出了两种互补的深度学习模型，为后续的智能资源调度与优化决策提供数据支撑。

首先，针对电磁环境状态的精准感知，本章提出基于多重图特征的干扰预测框架，并设计了多样化特征时空图注意力网络（DSTGAT）。该研究通过分析证实了从易获取的 PRB 流量预测高成本干扰数据的可行性，并揭示了两者间复杂的、受多重因素影响的空间关联性。为此，DSTGAT 并行利用四种异构邻接矩阵（基于距离、道路、谱聚类和相关性）构建多维度空间表征，并结合定制化时间编码，实现了对干扰态势的高精度、精细化预测。

其次，为解决业务负载预测在大规模网络中的效率瓶颈与节点同质性假设的局限，本章提出频段嵌入的低秩注意力网络（FELA-Net）。该研究基于近似熵（ApEn）和奇异值分解（SVD）的先验分析，证实了不同频段业务的异构性与空

间注意力的低秩特性。基于此, FELA-Net 设计了频段嵌入层以捕捉异构模式, 并引入低秩近似注意力机制, 将空间依赖建模的复杂度从二次方降至近线性级别, 为实现广域、高效的业务态势感知提供了可行方案。

综上所述, DSTGAT 与 FELA-Net 分别从“精细化高精度预测”和“规模化高效预测”两个层面, 共同构建了地面网络的频谱态势感知能力。本章建立的感知能力为后续研究奠定了基础, 其深远价值在于, 这些精准、高效的预测结果将作为关键输入, 为后续章节探讨的星地协同频谱共享策略提供决策依据与动态优化基础。

第4章 融合频谱态势预测的星地联合资源管理

4.1 引言

在地面蜂窝网络的资源调度领域，3GPP 标准本身并未规定 gNB 的具体调度算法^[92]。标准主要定义了接口、信令流程和资源块的结构，而将资源分配功能留给了设备制造商。传统的资源分配方法在理论和实践上分化为两个主要流派：

首先是以商用实现为代表的、在单个 gNB 上运行的分布式启发式算法。实际的商用策略由设备商实现，其中最常见的是最大吞吐量、轮询或比例公平。这类算法的优势是实现简单、开销低。然而，其根本缺陷在于决策是“本地利己”的：它们仅根据本小区的信道和队列信息来优化本地目标（如吞吐量或公平性），完全无法感知和处理由卫星引入的、复杂的“跨域干扰”，因此在星地融合场景下自然是失效的。

其次是以学术研究为代表的、追求理论最优解的集中式优化方法。集中式优化核心在于一个拥有全局信息的中央控制器，根据精确的系统模型求解目标函数以获得最优解。尽管该方法论能通过连续凸近似^[93] 和多目标演化算法^[94–96] 等技术处理复杂的非凸和多目标问题，但在星地融合网络这类大规模动态系统中，其局限性显著。首先，它面临“维度灾难”，可扩展性极差，求解时间随网络规模呈指数级增长。其次，它依赖静态或准静态环境假设，无法实时适应低地球轨道 (Low Earth Orbit, LEO) 卫星的高速移动和业务的非均匀变化。最后，它对精确数学模型的依赖使其在面对复杂现实环境时性能大打折扣。

无论是追求局部效率的分布式启发式算法，还是追求理论最优但缺乏现实适应性的集中式优化方法，这种技术与应用场景上的失配，催生了对能够自我学习和适应的智能决策方法的需求。研究人员转向单智能体强化学习 (Single Agent Reinforcement Learning, SARL)^[97,98]，将网络资源分配建模为一个马尔可夫决策过程 (Markov Decision Process, MDP)^[99–101]。SARL 的优势在于它是一种无模型学习方法，能够通过与环境的交互，从试错中学习最优策略，从而适应环境的随机性和不确定性。在多智能体系统中，SARL 的环境平稳性假设不再成立，因为其他智能体的动态决策行为会导致个体智能体所处的状态转移与奖励函数持续变化，从而违背 MDP 的马尔可夫条件。这种非平稳性使过往经验迅速失效，引发学习过程不稳定，难以收敛至有效协同策略，并可能导致因错误信誉分配而陷入低效竞争的困境^[102]。

为了解决单智能体强化学习面临的非平稳性问题，多智能体强化学习 (Multi-Agent Reinforcement Learning, MARL) 应运而生。它将系统建模为多个相互作用的智能体，最主流的范式是集中式训练与分布式执行 (Centralized Training with Distributed Execution, CTDE)。在训练阶段，一个能够访问全局信息的集中式评论家为每个智能体提供稳定的梯度信号，有效解决了非平稳性挑战^[103]。在执行阶段，该评论家被丢弃，每个智能体仅根据自身局部观测独立做出决策。这一框

架天然契合分布式网络的架构，提供了高执行效率、容错性与可扩展性。然而，MARL 也面临新的挑战，例如部分可观测性以及训练可扩展性问题。CTDE 框架中集中式评论家的输入维度会随智能体数量线性增长，导致“输入爆炸”问题，限制了其在大规模网络中的应用^[104]。

为此，本章的主要贡献在于：

分层博弈 MARL 框架设计：本文提出一种分层博弈 MARL 框架，将复杂的联合资源分配问题分解为智能体间的上层非合作博弈与智能体内的下层资源优化。在 CTDE 范式下，全局 Critic 网络通过评估联合动作价值来隐式求解上层博弈问题，而本地 Actor 网络则专注于下层优化，从而降低了决策复杂度并提升了系统可扩展性。

面向混合动作空间的 Actor-Critic 结构：为应对离散信道选择与连续功率控制的混合动作空间，我们设计了特定的 Actor-Critic 结构，以解耦并高效输出混合动作。

基于未来负载预测的前瞻性决策机制：本框架的核心创新在于引入了前瞻性决策机制，通过将带噪声的未来 PRB 利用率预测作为状态维度，使 Actor 网络能学习资源预留等前瞻性策略，而掌握全局未来信息的 Critic 网络则能更精准地评估决策的长期价值。

4.2 预测驱动的星地联合资源管理问题建模

本章旨在构建一个高动态性的星地融合网络频谱共享仿真环境。该模型包含了网络场景、星地异构信道、信号传输与多维干扰以及优化问题的数学建模，为后续提出的多智能体强化学习资源分配算法提供基础。

系统模型如图4-1所示，我们考虑一个遵循 5G 非地面网络 (Non-Terrestrial Network, NTN) 标准的 LEO 卫星与多个采用非正交多址接入 (Non-Orthogonal Multiple Access, NOMA) 技术的 gNB 之间的多信道下行协作频谱共享场景。在该场景中，一颗轨道高度为 600km 的 LEO 卫星采用固定波束技术，在可见时间内持续服务于一个中心的 NTN 站。卫星的下行波束覆盖了 NTN 站周围的多个地面 gNB，这些 gNB 的集合记为 $\mathcal{M} = \{1, \dots, m, \dots, M\}$ 。为实现广域覆盖，这 M 个 gNB 环形部署在 NTN 站周围。

在所考虑的星地融合场景中，地面系统面临海量高速率终端连接的需求，而其自有频谱资源紧张；相比之下，LEO 卫星系统当前服务负载较轻，导致其授权频谱资源未得到充分利用。因此，每个 gNB 下的用户被分为两类：常规用户和频谱共享用户。前者使用蜂窝网络自有频段，而后者则机会式地共享 LEO 卫星的下行频谱资源。共享的频谱总带宽为 B_{total} ，被划分为 N 个正交子信道，其集合记为 $\mathcal{N} = \{1, \dots, n, \dots, N\}$ 。在地面 5G 系统中，物理资源块 (Physical Resource Block, PRB) 是资源调度的基本单元，因此我们假设每个子信道 n 恰好对应一个 PRB，其带宽为 B_n 。我们将 gNB m 中的地面共享用户集合记为 $\mathcal{K}_m = \{1, \dots, k, \dots, K_m\}$ ，

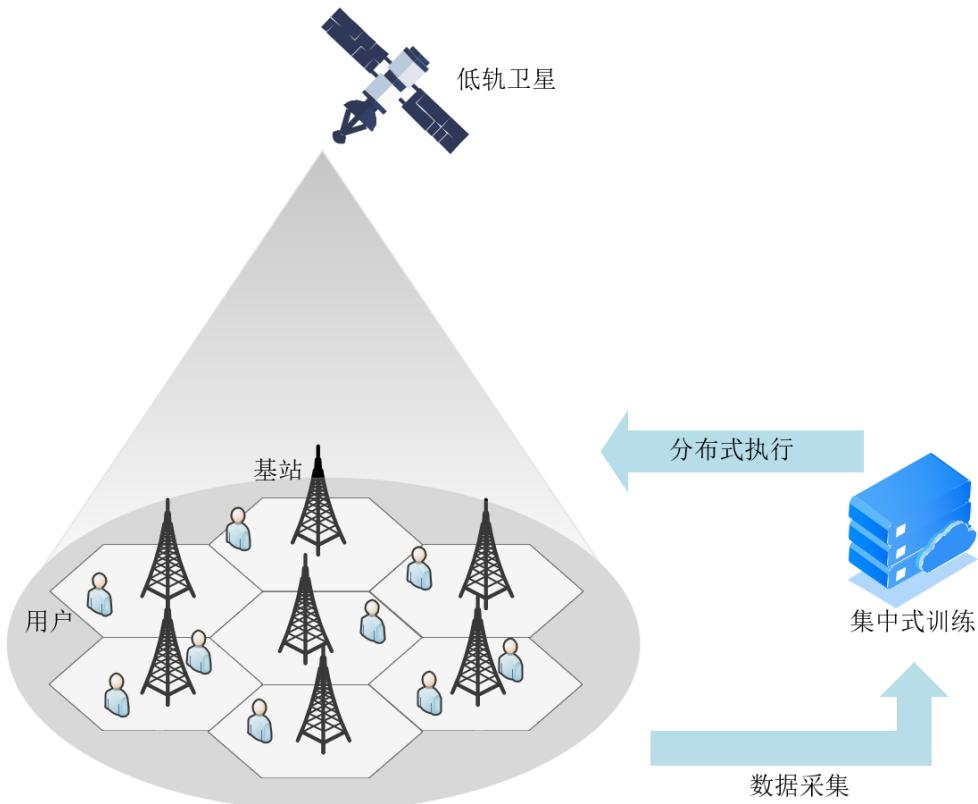


图 4-1 星地融合网络场景示意图

Figure 4-1 Schematic diagram of the Star-Earth Integration Network scenario

其中 K_m 是该基站服务的共享用户数。信道分配决策由二进制指示符表示，若卫星将信道 n 分配给 NTN 站 e ，则 $\chi_{\text{sat},e,n} = 1$ ，否则为 0；若 gNB m 将信道 n 分配给其用户 k ，则 $l_{m,k,n} = 1$ ，否则为 0。

为了提升频谱效率并支持海量接入，地面系统采用 NOMA 技术。然而，这种星地频谱共享模式也引入了复杂的多维同信道干扰：每个地面共享用户不仅会受到来自卫星下行链路的系统间干扰，还会受到来自其他 gNB 的小区间干扰，以及来自本小区内同一信道上其他 NOMA 用户的区内干扰。为了实现有效的协同，NTN 站与地面 gNB 之间通过 Xn 接口进行信息交互，gNB 之间则采用 X2 接口。这使得所有共享用户的速率需求、信道状态信息（CSI）以及资源分配策略等关键信息可以在网络实体间共享与反馈。

4.2.1 星地动态异构网络信道模型

为精确建模星地融合网络的高动态与异构特性，本节分别构建卫星与地面链路的信道模型。

4.2.1.1 星地链路模型

星地链路模型综合考量了由卫星高速运动引起的几何时变性、路径损耗、阴影衰落及多普勒效应。在离散时间步 Δt , 卫星轨道角位置 ψ 的更新如下:

$$\psi(t + \Delta t) = \psi(t) + \frac{v \cdot \Delta t}{R_a + h} \quad (4-1)$$

该公式用于更新卫星在轨道上的角位置, 它将卫星的线速度 v 转换为在极短时间 Δt 内扫过的角度, 从而实现对卫星轨道运动的离散化模拟。其中, $\psi(t)$ 代表在时刻 t 的卫星轨道角位置 (弧度), Δt 为仿真的时间步长 (秒), v 是卫星的轨道速度 (米/秒), R_a 为地球平均半径 (米), h 是卫星轨道高度 (米)。

基于 $\psi(t)$, 实时仰角 $\phi(t)$ 可计算为:

$$\phi(t) = \arctan \left(\frac{\cos \psi(t) - 1/\lambda}{\sin \psi(t)} \right) \quad (4-2)$$

此公式计算地面终端观测卫星的仰角, 即卫星与地面终端的连线与地平线的夹角。仰角是决定链路质量、是否被遮挡以及路径损耗的关键几何参数。其中, $\phi(t)$ 为在时刻 t 的卫星仰角 (弧度), $\psi(t)$ 是在时刻 t 的卫星轨道角位置 (弧度), λ 是轨道半径与地球半径的比值, 定义为 $\lambda = (R_a + h)/R_a$)。

进而, 动态斜距 $d(t)$ 由下式确定:

$$d(t) = \sqrt{R_a^2 \sin^2 \phi(t) + h^2 + 2R_a h - R_a \sin \phi(t)} \quad (4-3)$$

该公式计算卫星与地面终端之间的直线距离, 即斜距。斜距是计算自由空间路径损耗的直接输入, 其时变性是星地信道时变特性的主要来源。其中, $d(t)$ 是在时刻 t 的星地斜距 (米), R_a 为地球平均半径 (米), $\phi(t)$ 是在时刻 t 的卫星仰角 (弧度), h 为卫星轨道高度 (米)。

大尺度衰落采用自由空间路径损耗 (Free Space Path Loss, FSPL) 模型, 其 dB 形式为:

$$L_{\text{FSPL}}(\text{dB}) = 32.45 + 20 \log_{10}(f_{\text{GHz}}) + 20 \log_{10}(d_{\text{km}}) \quad (4-4)$$

此公式计算电磁波在自由空间传播时因能量扩散造成的信号功率衰减, 描述了最基本的大尺度路径损耗。其中, $L_{\text{FSPL}}(\text{dB})$ 是自由空间路径损耗值 (分贝), f_{GHz} 为载波频率 (吉赫兹), d_{km} 是传播距离, 即星地斜距 (千米)。

小尺度衰落采用阴影莱斯衰落模型。其信道功率 $|g_{\text{sat}}|^2$ 的概率密度函数 (PDF) 为:

$$f_{|g_{\text{sat}}|^2}(x) = \alpha \exp(-\beta x) {}_1F_1(m_s; 1; \delta x) \quad (4-5)$$

该概率密度函数描述了星地信道中信号功率的统计分布, 适用于存在较强视距信号但其强度受阴影效应随机影响的环境。其中, $f_{|g_{\text{sat}}|^2}(x)$ 是星地信道功率的概率密度函数, x 代表信道功率的随机变量。 m_s 为衰落严重性参数, $2b_s$ 为多径散射分量的平均功率, Ω_s 为视距分量的平均功率, ${}_1F_1(\cdot; \cdot; \cdot)$ 是合流超几何函数。参数 α, β, δ 由 $\alpha = \left(\frac{2b_s m_s}{2b_s m_s + \Omega_s} \right)^{m_s} / (2b_s)$, $\beta = \frac{1}{2b_s}$, $\delta = \frac{\Omega_s}{2b_s(2b_s m_s + \Omega_s)}$ 定义。

4.2.1.2 地面链路模型

地面 gNB 至用户链路的路径损耗采用幂律模型：

$$L_{\text{TERR}}(d) = d^{-\alpha} \quad (4-6)$$

此公式描述了地面信号功率随传播距离增加而衰减的规律，是一个常用经验模型。其中， $L_{\text{TERR}}(d)$ 是地面链路的路径损耗（线性值）， d 是 gNB 与用户间的距离（米）， α 为路径损耗指数（无量纲），其大小反映了传播环境的复杂程度。

其快速衰落遵循 Nakagami-m 分布。信道功率 $|g_{\text{gNB}}|^2$ 服从伽马分布，其 PDF 为：

$$f_{|g_{\text{gNB}}|^2}(x) = \frac{(m/\Omega)^m x^{m-1}}{\Gamma(m)} \exp\left(-\frac{m}{\Omega}x\right) \quad (4-7)$$

该概率密度函数描述了地面信道中信号功率的统计分布，其灵活性使其能拟合从瑞利衰落到莱斯衰落的多种场景。其中， $f_{|g_{\text{gNB}}|^2}(x)$ 是地面信道功率的概率密度函数， x 代表信道功率的随机变量。 m 为塑形因子，决定衰落的严重程度。 Ω 是信道平均功率。 $\Gamma(\cdot)$ 是伽马函数。

4.2.1.3 综合信道增益

最终，星地与地面信道的增益 h 由其路径损耗 L 和小尺度衰落分量 g 共同决定：

$$h = L^{-1} \cdot g \quad (4-8)$$

该公式表示信道的总增益是路径损耗（大尺度衰落）和快衰落（小尺度衰落）共同作用的结果。路径损耗决定了信号的平均衰减水平，而快衰落则导致了信号功率围绕该平均值的快速波动。其中， h 是总的信道增益（线性值）， L 是路径损耗（线性值， $L > 1$ ）， g 是小尺度衰落分量（复数），其均值为 0，方差由上述 PDF 描述。

4.2.2 星地动态异构网络信号传输与干扰模型

本系统的性能通过各链路的信干噪比（Signal to Interference plus Noise Ratio, SINR）来评估。对于 NTN 关口站 e ，其在信道 n 上的接收 SINR $\gamma_{\text{sat},e,n}$ 受到所有共享信道的地面 gNB 的干扰，其表达式为：

$$\gamma_{\text{sat},e,n} = \frac{P_{\text{sat}} G_{\text{sat},e} |h_{\text{sat} \rightarrow e,n}|^2}{\sum_{m=1}^M \sum_{k=1}^{K_m} l_{m,k,n} P_{m,k} G_{m,e} |h_{m \rightarrow e,n}|^2 + \sigma^2} \quad (4-9)$$

该公式计算了卫星接收端的信干噪比，其分子为期望信号的功率，分母为所有地面干扰信号的总功率与噪声功率之和。其中， P_{sat} 是卫星发射功率， $G_{\text{sat},e}$ 是卫星到接收站的天线增益， $|h_{\text{sat} \rightarrow e,n}|^2$ 是星地信道增益。在分母中， $P_{m,k}$ 是地面 gNB m 的用户 k 的发射功率， $G_{m,e}$ 是该用户到卫星接收站的天线增益， $|h_{m \rightarrow e,n}|^2$ 是

相应的信道增益， $l_{m,k,n}$ 是一个二进制指示符（当用户 k 在 gNB m 上占用信道 n 时为 1，否则为 0）， σ^2 是接收端的加性高斯白噪声功率。对于 gNB m 下的用户 k ，其在信道 n 上的接收 SINR $\gamma_{m,k,n}$ 同时受到卫星、其他 gNB 及本 gNB 内其他用户的干扰，其表达式为：

$$\gamma_{m,k,n} = \frac{P_{m,k} G_{m,k} |h_{m \rightarrow k,n}|^2}{I_{\text{sat}} + I_{\text{inter}} + I_{\text{intra}} + \sigma^2} \quad (4-10)$$

该公式计算了地面用户的信干噪比，其分子为期望信号功率，分母为所有干扰源（卫星、其他小区、同小区）的干扰功率与噪声功率之和。其中， $P_{m,k}$ 是用户 k 的发射功率， $G_{m,k}$ 是从 gNB m 到用户 k 的天线增益， $|h_{m \rightarrow k,n}|^2$ 是该链路的信道增益。分母中的干扰项 I_{sat} 代表来自卫星的干扰， I_{inter} 代表来自其他地面小区的干扰， I_{intra} 代表来自同一小区内其他用户的 NOMA 干扰， σ^2 是噪声功率。

4.2.3 预测增强的联合优化问题

本研究的目标是在保障所有共享用户（包括卫星和地面）服务质量 QoS 的前提下，通过联合优化所有地面 gNB 的信道分配矩阵 \mathbf{L} 和功率分配矩阵 \mathbf{P} ，来最大化整个星地融合系统的全局频谱效率（Global Spectrum Efficiency, SE）。该优化问题可以建模为：

$$\max_{\mathbf{L}, \mathbf{P}} \quad \frac{1}{B_{\text{total}}} \left(\sum_{n=1}^N C_{\text{sat},e,n} + \sum_{m=1}^M \sum_{k=1}^{K_m} \sum_{n=1}^N l_{m,k,n} C_{m,k,n} \right) \quad (4-11)$$

$$\text{s.t.} \quad \sum_{n=1}^N C_{\text{sat},e,n} \geq Q_{\text{sat}} \quad (4-12)$$

$$\sum_{n=1}^N l_{m,k,n} C_{m,k,n} \geq Q_{\text{user}}, \quad \forall m, k \quad (4-13)$$

$$P_{\min} \leq P_{m,k} \leq P_{\max}, \quad \forall m, k \quad (4-14)$$

$$\sum_{n=1}^N l_{m,k,n} = 1, \quad l_{m,k,n} \in \{0, 1\}, \quad \forall m, k \quad (4-15)$$

此公式体系旨在通过联合优化信道与功率分配来最大化系统的全局频谱效率，同时满足所有用户的服务质量要求。公式 (4-11) 是目标函数，旨在最大化系统总吞吐量与总带宽的比值。其中， \mathbf{L} 和 \mathbf{P} 分别是信道分配和功率分配的决策矩阵， B_{total} 是系统总带宽， $C_{\text{sat},e,n}$ 和 $C_{m,k,n}$ 分别是卫星链路和地面链路在信道 n 上的香农容量，其值为 $B_{\text{ch}} \log_2(1 + \gamma)$ 。公式 (4-12) 和 (4-13) 分别是卫星和地面用户的最低服务质量约束，确保其总速率不低于门限 Q_{sat} 和 Q_{user} 。公式 (4-14) 限制了用户的发射功率在 $[P_{\min}, P_{\max}]$ 范围内。公式 (4-15) 确保每个用户只能被分配到一个信道。该问题是一个非凸混合整数非线性规划问题，传统方法难以求解，因此采用多智能体强化学习方法寻找其动态最优策略。

4.2.4 预测增强的多智能体联合决策算法：PE-MADDPG

为应用多智能体强化学习，该动态频谱共享问题被建模为一个分布式部分可观察马尔可夫决策过程（Decentralized Partially Observable Markov Decision Process, Dec-POMDP）。该框架能够精确地刻画多智能体在信息不完全环境下的序贯决策过程，其模型由元组 $\langle \mathcal{M}, \mathcal{S}, \{\mathcal{O}_m\}, \{\mathcal{A}_m\}, \mathcal{R} \rangle$ 定义。

4.2.4.1 强化学习要素定义

(1) 智能体

我们将星地联合频谱共享问题构建为一个多智能体系统。在该系统中，智能体集合 \mathcal{M} 被定义为地面蜂窝网络基站的集合：

$$\mathcal{M} = 1, 2, \dots, M \quad (4-16)$$

该公式中， \mathcal{M} 代表所有智能体的集合， M 是系统中智能体的总数，而每个索引 $m \in \mathcal{M}$ 唯一标识一个作为决策实体的 gNB。这种设计的合理性在于 gNB 是实际执行本地频谱与功率控制的物理单元，这种直接的映射方式确保了算法学习的策略 π_m 能够应用于实际的网络操作中，确保了方案的可部署性。

(2) 状态空间

本研究遵循中心化训练与分布式执行（CTDE）范式，因此我们定义全局状态 $S_t \in \mathcal{S}$ 。全局状态 S_t 是 t 时刻整个星地融合系统的完整信息快照，它仅在中心化训练阶段对全局 Critic 可见，用于指导所有智能体学习协同策略。

为全面表征系统的高动态特性， S_t 被设计为一个集成了瞬时信息、历史动作和未来预测的向量：

$$S_t = [\underbrace{\mathcal{H}(t)}_{\text{瞬时 CSI}}, \underbrace{\mathcal{S}_{\text{sat}}(t)}_{\text{卫星系统状态}}, \underbrace{\mathcal{A}(t-1)}_{\text{历史联合动作}}, \underbrace{\hat{\mathbf{U}}(t+1)}_{\text{预测信息}}] \quad (4-17)$$

该公式中， $\mathcal{H}(t)$ 代表瞬时 CSI， $\mathcal{S}_{\text{sat}}(t)$ 是卫星系统状态， $\mathcal{A}(t-1)$ 是历史联合动作， $\hat{\mathbf{U}}(t+1)$ 是前瞻性预测信息。该公式定义了一个全面捕捉系统动态的复合状态空间。下面是详细介绍：

- **瞬时 CSI $\mathcal{H}(t)$** : 包含 t 时刻所有关键链路的信道增益、所有 gNB 到其用户的地面通信链路 $\mathbf{H}_m(t)$ 、卫星对所有地面用户的干扰链路 $\mathbf{G}_m(t)$ 以及小区间干扰链路。

- **卫星系统状态 $\mathcal{S}_{\text{sat}}(t)$** : 它描述了非地面组件的状态，由卫星的实时轨道状态向量 $\mathbf{v}_{\text{sat}}(t)$ 和指示卫星信道占用的二元向量 $\chi_{\text{sat}}(t) \in \{0, 1\}^N$ 构成，即 $\mathcal{S}_{\text{sat}}(t) = (\mathbf{v}_{\text{sat}}(t), \chi_{\text{sat}}(t))$ 。

- **历史联合动作 $\mathcal{A}(t-1)$** : 其定义为 $\mathcal{A}(t-1) = (\mathbf{L}(t-1), \mathbf{P}(t-1))$ ，其中 $\mathbf{L}(t-1)$ 与 $\mathbf{P}(t-1)$ 分别是全局信道与功率分配矩阵。在中心化训练中引入历史动作，旨在帮助评判器网络捕捉智能体间的策略演化，从而部分缓解环境多智能体并发学习导致的非平稳性问题。

- 预测信息 $\hat{\mathbf{U}}(t+1) \in [0, 1]^N$: 是由外部模块提供的、对下一时刻地面网络 PRB 利用率的客观预测值。将未来信息纳入当前状态，赋予了学习框架超越当前观测的预判能力，使智能体能够学习到更具前瞻性的长期策略。

(3) 部分可观测信息

在分布式执行阶段，每个智能体 m 在一个部分可观测马尔可夫决策过程 (POMDP) 环境中运行，它无法访问全局状态 $s(t)$ ，而是依赖其局部的观测向量 $\mathcal{O}_m(t) \in \mathcal{O}_m$ 。该向量被设计为四个关键部分的拼接，旨在为智能体提供最大化的决策信息，同时反映真实网络中的信息约束。令 K_m 为 gNB m 下属的用户集合， N 为正交子信道的集合， $\mathcal{O}_m(t)$ 的结构如下：

$$\underbrace{\mathcal{O}_m(t)}_{\text{总观测向量}} = [\underbrace{\mathcal{O}_{m,\text{csi}}(t)}_{\text{CSI 部分}}, \underbrace{\mathcal{O}_{m,\text{comp}}(t)}_{\text{复合特征部分}}, \underbrace{\mathcal{O}_{m,\text{shared}}(t)}_{\text{共享信息部分}}, \underbrace{\mathcal{O}_{m,\text{hist}}(t)}_{\text{历史动作部分}}] \quad (4-18)$$

在该公式中， $\mathcal{O}_m(t)$ 是智能体 m 总的局部观测向量，它由信道状态信息 $\mathcal{O}_{m,\text{csi}}(t)$ 、预测性复合特征 $\mathcal{O}_{m,\text{comp}}(t)$ 、共享全局信息 $\mathcal{O}_{m,\text{shared}}(t)$ 和本地动作历史 $\mathcal{O}_{m,\text{hist}}(t)$ 四个部分级联构成。

- 本地与干扰信道状态信息 $\mathcal{O}_{m,\text{csi}}(t)$: 是决策的基础物理信息，由本地链路 CSI 和系统间干扰 CSI 两部分信息拼接而成。本地链路信道状态信息矩阵 $\mathbf{H}_m(t) \in \mathbb{R}^{K_m \times N}$ 表示 gNB m 到其用户的瞬时增益，其元素为 $h_{m,k}^n(t)$ 。系统间干扰信道状态信息矩阵 $\mathbf{G}_m(t) \in \mathbb{R}^{K_m \times N}$ 表示来自卫星的下行链路对 gNB m 的用户造成的干扰信道增益，其元素为 $g_{\text{sat},k}^n(t)$ 。然后将上述两个矩阵展平并拼接得到 $\mathcal{O}_{m,\text{csi}}(t)$ ，最终的总维度为 $2K_m N$ 。

- 预测性复合特征矩阵 $\mathcal{O}_{m,\text{comp}}(t)$: 为实现“预测驱动”的决策，我们通过特征工程构建了一个复合特征矩阵 $\mathbf{F}_m(t) \in \mathbb{R}^{K_m \times N}$ ，旨在通过特征工程将当前信道质量与未来网络负载进行深度融合。其每个元素的定义如下：

$$\mathbf{F}_{m,k}^n(t) = f(h_{m,k}^n(t), \hat{u}_n(t+1)) \quad (4-19)$$

其中， $h_{m,k}^n(t)$ 是用户 k 在信道 n 上的瞬时增益， $\hat{u}_n(t+1) \in [0, 1]$ 是信道 n 在下一时刻的 PRB 利用率预测值， $f(\cdot, \cdot)$ 是一个特征融合函数。接下来将 $\mathbf{F}_m(t) \in \mathbb{R}^{K_m \times N}$ 展平，得到向量 $\mathcal{O}_m(t) \in K_m N$ 。该设计构建了一个拥塞调整后的信道质量指标，使得即使信道当前质量高，若预测未来拥塞，其评价价值也会降低。

- 共享全局信息 $\mathcal{O}_{m,\text{shared}}(t)$: 包含卫星信道占用状态 $\chi_{\text{sat}}(t) \in \{0, 1\}^N$ 以及全局 PRB 利用率预测向量 $\hat{\mathbf{U}}(t+1) = (\hat{u}_n(t+1))_{n \in N} \in [0, 1]^N$ ，拼接后得到共享全局向量 $\mathcal{O}_{m,\text{shared}}(t) \in 2N$ 。该部分包含对所有智能体都至关重要、但维度较低的全局信息，并通过 Xn/X2 接口的低开销信令进行交互和广播。

- 本地动作历史 $\mathcal{O}_{m,\text{hist}}(t)$: 包含智能体 m 在上一时间步的信道分配动作 $\mathcal{A}_{m,\text{ch}}(t-1) \in N^{K_m}$ 。引入历史动作有助于智能体构建对环境状态的更完整信念 (belief)。

最终，总的局部观测向量 $\mathcal{O}_m(t)$ 通过拼接上述四个部分构成，维度为 $D_m = 3K_m N + 2N + K_m$ 。

(4) 动作空间

在 t 时刻，每个智能体 m 必须为其下属的 K_m 个用户确定一个联合动作 $\mathcal{A}_m(t)$ 。该动作空间是一个由离散的信道分配和连续的功率分配共同构成的混合动作空间：

$$\mathcal{A}_m(t) = (\mathcal{A}_{m,ch}(t), \mathcal{A}_{m,p}(t)) \in \mathcal{A}_{ch} \times \mathcal{A}_p \quad (4-20)$$

在该公式中， $\mathcal{A}_m(t)$ 是智能体 m 的总动作， $\mathcal{A}_{m,ch}(t)$ 和 $\mathcal{A}_{m,p}(t)$ 分别为其离散信道分配动作和连续功率分配动作。 \mathcal{A}_{ch} 和 \mathcal{A}_p 是对应的动作空间。该定义将一个复杂的决策问题分解为两个子问题。智能体需要解决的首要挑战是离散信道分配。每个用户需从 N 个可用子信道中选择一个，因此完整的信道分配动作 $\mathcal{A}_{m,ch}(t)$ 是一个 K_m 维向量：

$$\mathcal{A}_{m,ch}(t) = (c_1, c_2, \dots, c_{K_m}) \in \mathcal{A}_{ch} \quad (4-21)$$

在此公式中， $c_i \in \{1, 2, \dots, N\}$ 是分配给用户 i 的信道索引。此动作空间的基数是 $|\mathcal{A}_{ch}| = N^{K_m}$ ，意味着其维度随用户数 K_m 呈指数级增长，即组合爆炸问题。这使得传统算法难以进行有效的探索与收敛。为规避组合爆炸问题，Actor 网络的策略采用了多分枝架构。该设计将 N^{K_m} 维的联合动作空间分解为 K_m 个并行的 N 维决策子空间。具体而言，Actor 网络的信道分配输出头包含 K_m 个并行分支，每个分支 i 对应用户 i 并输出一个 N 维的分类概率分布：

$$\pi_{ch,i}(\cdot | o_m) = P(c_i = \cdot | o_m) \in \mathbb{R}^N \quad (4-22)$$

该公式中， $\pi_{ch,i}(\cdot | o_m)$ 表示在局部观测 o_m 下，为用户 i 选择各信道的概率分布。在给定局部观测 o_m 时，该架构假设各用户信道选择动作之间满足条件独立性，因此完整的信道策略可表示为各分支策略的乘积：

$$\pi_{ch}(\mathcal{A}_{m,ch} | o_m) = \prod_{i=1}^{K_m} \pi_{ch,i}(c_i | o_m) \quad (4-23)$$

尽管输出层假设了条件独立性，但所有 K_m 个分支共享底层的神经网络，且智能体由全局奖励和中心化 Critic 网络引导，因此算法仍能学习到动作之间必需的隐性协同。

动作的第二个组成部分是连续的功率分配 $\mathcal{A}_{m,p}(t)$ ，它是一个 K_m 维的连续向量：

$$\mathcal{A}_{m,p}(t) = (p_1, p_2, \dots, p_{K_m}) \in \mathcal{A}_p \subset \mathbb{R}^{K_m} \quad (4-24)$$

此公式中， $p_i \in [P_{\min}, P_{\max}]$ 是分配给用户 i 的发射功率。该向量由 Actor 网络的一个独立输出头确定。关键在于，该功率分配决策以信道分配结果为条件。在实现中，决策 $\mathcal{A}_{m,p}(t)$ 的输入不仅包含观测 o_m ，还包含了已采样的信道动作 $\mathcal{A}_{m,ch}(t)$ ，从而学习一种信道感知的功率控制策略：

$$\pi_p(\mathcal{A}_{m,p} | o_m, \mathcal{A}_{m,ch}) \quad (4-25)$$

该条件策略公式表明，功率分配是根据实时观测和已确定的信道分配方案来动态调整的，实现了精细化的资源管理。

(5) 奖励函数为实现多智能体系统的高效协同，设计了一个综合性的全局奖励函数 $\mathcal{R}(t)$ 。该奖励函数在 t 时刻被所有 gNB 智能体共享，旨在最大化即时系统性能，并集成对未来网络状态的预测以惩罚“短视”决策。 $\mathcal{R}(t)$ 由即时频谱效率、服务质量惩罚和前瞻性拥塞惩罚三部分构成，其结构定义如下：

$$\underbrace{\mathcal{R}(t)}_{\text{全局奖励信号}} = \underbrace{\mathcal{R}_{\text{efficiency}}(t)}_{\text{即时频谱效率奖励}} - \underbrace{\mathcal{P}_{\text{qos}}(t)}_{\text{服务质量惩罚项}} - \underbrace{\mathcal{P}_{\text{congestion}}(t)}_{\text{预测拥塞惩罚项}} \quad (4-26)$$

在该公式中， $\mathcal{R}(t)$ 表示 t 时刻的全局奖励信号， $\mathcal{R}_{\text{efficiency}}(t)$ 是即时频谱效率奖励， $\mathcal{P}_{\text{qos}}(t)$ 是服务质量惩罚项， $\mathcal{P}_{\text{congestion}}(t)$ 为预测拥塞惩罚项。该奖励函数的整体设计目标是通过奖励高效行为并惩罚违规与拥塞风险，引导智能体学习最优协同策略。下面将详细介绍各个组成部分：

- **即时频谱效率奖励 $\mathcal{R}_{\text{efficiency}}(t)$** : 该项为系统的主要优化目标，定义为经缩放的全局频谱效率：

$$\mathcal{R}_{\text{efficiency}}(t) = \alpha \cdot \frac{R_S(t) + \sum_{n \in \mathcal{N}} \sum_{m \in \mathcal{M}_n} R_{n,m}(t)}{B_{\text{total}}} \quad (4-27)$$

在此公式中， α 是一个用于平衡奖励函数各分量数值尺度的缩放超参数。 $R_S(t)$ 是 t 时刻的卫星链路速率， $R_{n,m}(t)$ 是 gNB n 下用户 m 的速率， \mathcal{N} 和 \mathcal{M}_n 分别是 gNB 集合和每个 gNB 下的用户集合， B_{total} 是总共享带宽。此项将系统总速率与带宽的比值作为主要激励信号。

- **QoS 违规惩罚项 $\mathcal{P}_{\text{qos}}(t)$** : 该项用于确保所有用户的基本服务质量。为定义此项，首先引入一个通用的二元违规函数 $\mathcal{V}(R, R^{\min})$:

$$\mathcal{V}(R, R^{\min}) = \begin{cases} 1 & \text{if } R < R^{\min} \\ 0 & \text{if } R \geq R^{\min} \end{cases} \quad (4-28)$$

该函数中， R 为实际速率， R^{\min} 为最低速率要求，函数在发生 QoS 违规时返回 1，否则返回 0。基于此函数，总的 QoS 惩罚被设计为对卫星和地面用户的混合惩罚：

$$\mathcal{P}_{\text{qos}}(t) = \beta \cdot \mathcal{V}(R_S(t), R_S^{\min}) + \delta \cdot \sum_{n \in \mathcal{N}} \sum_{m \in \mathcal{M}_n} \mathcal{V}(R_{n,m}(t), R_U^{\min}) \quad (4-29)$$

此公式中， β 是对卫星链路 QoS 违规施加的固定惩罚权重， δ 是对每一个违规地面用户施加的惩罚权重， R_U^{\min} 是地面用户最低速率要求。该惩罚项通过差异化的策略对 QoS 违规行为进行量化。

- **前瞻性拥塞惩罚项 $\mathcal{P}_{\text{congestion}}(t)$** : 此项将当前时刻 t 的动作与对 $t+1$ 时刻的网络状态预测相关联。该惩罚基于 gNB n 在 $t+1$ 时刻的 PRB 利用率预

测值 $f_n \in [0, 1]$, 以及其在 t 时刻的信道占用率 $u_n(t) = \frac{|\mathcal{K}_n(t)|}{K}$, 其中 $\mathcal{K}_n(t)$ 是已分配的子信道集合, K 是总信道数。为便于表达, 再定义一个拥塞触发函数 $\mathcal{V}_{\text{congestion}}(f_n, \tau_{\text{congestion}})$:

$$\mathcal{V}_{\text{congestion}}(f_n, \tau) = \begin{cases} 1 & \text{if } f_n > \tau \\ 0 & \text{if } f_n \leq \tau \end{cases} \quad (4-30)$$

该函数在预测拥塞度 f_n 超过高拥塞预警阈值 $\tau_{\text{congestion}}$ 时被激活。总的拥塞惩罚由软惩罚和硬惩罚共同构成:

$$\mathcal{P}_{\text{congestion}}(t) = \sum_{n \in \mathcal{N}} u_n(t) \cdot (\gamma_1 \cdot f_n + \gamma_2 \cdot \mathcal{V}_{\text{congestion}}(f_n, \tau_{\text{congestion}})) \quad (4-31)$$

该公式中, γ_1 和 γ_2 分别是软惩罚权重和硬惩罚权重。这种设计明确激励智能体: 若预测到未来拥塞, 应在当前主动减少信道占用, 从而学习到具备长期收益的资源分配策略。

4.2.4.2 PE-MADDPG 的架构与训练过程

为求解前述 Dec-POMDP 模型, 我们提出了一种专为星地融合网络设计的动态协作频谱共享算法, 命名为 PE-MADDPG。该算法以 MADDPG 为基础, 并针对问题特性在 Actor 网络结构、观测空间和训练策略上进行了深度定制。

(1) 算法整体架构

PE-MADDPG 的核心架构遵循中心化训练与分布式执行 (CTDE) 范式。在分布式执行阶段, 每个 gNB 智能体 m 仅根据其局部观测 $\mathcal{O}_m(t)$, 利用本地 Actor 网络 $\mu_m(\cdot | \theta^{\mu_m})$ 快速计算出确定性动作 $\mathcal{A}_m(t)$, 保证了决策的低延迟和系统的高可扩展性。在中心化训练阶段, 所有智能体将各自的经验元组上传至一个中心化的训练服务器。全局 Critic 网络通过访问所有智能体的观测和动作, 学习一个中心化的值函数, 为每个 Actor 的更新提供稳定的梯度指导。为解决混合动作空间决策难题, 我们设计了一个统一的 Actor 网络, 其包含两个并行的输出头。离散动作头负责信道选择, 为规避组合爆炸问题, 该输出头采用多分枝结构, 将一个巨大的联合动作空间分解为 K_m 个并行的、维度为 N 的独立分类任务。连续动作头负责功率分配, 其输入以信道分配结果为条件, 为每个用户确定一个连续的功率值。该设计通过降维简化了学习难度。

(2) 网络训练过程

PE-MADDPG 的训练过程是一个交替优化全局 Critic 和各个本地 Actor 的过程。在数据预处理阶段, 输入的信道状态信息 (CSI) 会进行对数归一化处理以稳定训练。在探索策略上, 训练初期在确定性动作上加入噪声, 具体为对信道选择采用 ϵ -greedy 策略, 对功率分配加入高斯噪声, 并随训练进程逐渐衰减噪声, 以平衡探索与利用。全局 Critic 网络 $V(S; \theta^V)$ 的更新旨在最小化时序差分误差 (Temporal Difference Error, TDE), 其损失函数定义为:

$$\mathcal{L}(\theta^V) = \mathbb{E}_{\mathcal{D}} [(y - V(S; \theta^V))^2] \quad (4-32)$$

其中, $\mathcal{L}(\theta^V)$ 是 Critic 网络的损失函数, $V(S; \theta^V)$ 是 Critic 网络对全局状态 S 的价值估计, y 是目标值, $\mathbb{E}_{\mathcal{D}}$ 表示从经验回放池 \mathcal{D} 中采样计算期望。目标值 y 通过目标网络计算得出, 即 $y = R + \gamma V'(S'; \theta^{V'})$, 其中 R 是奖励, γ 是折扣因子, V' 是目标 Critic 网络, S' 是下一时刻的全局状态。每个本地 Actor 网络 $\mu_m(\mathcal{O}_m; \theta^{\mu_m})$ 的更新目标是最大化长期回报, 其策略梯度由全局 Critic 提供的优势函数 $A(S, \mathcal{A}) \approx y - V(S)$ 指导。通过这种方式, 仅依赖局部信息的 Actor 在学习过程中被注入了全局视角, 从而学会了与其他智能体进行隐性协同。

(3) 网络结构与复杂度分析每个 gNB 的 Actor 网络是一个多层前馈神经网络, 输入为局部观测向量, 输出为信道选择和功率分配。全局 Critic 网络同样是一个多层前馈神经网络, 输入为所有局部观测拼接而成的全局状态向量, 输出为标量的状态价值。在线执行时, 每个 gNB 仅需执行一次本地 Actor 网络的前向传播, 计算开销低, 满足实时决策要求。离线训练时, 计算复杂度与批次大小、智能体数量和网络规模成正比。这种复杂度分配完美体现了 CTDE 框架的优势, 将高昂计算置于离线阶段, 保持在线执行的轻量级, 在理论性能和实际部署可行性之间取得了良好平衡。

4.3 仿真实验与性能评估

本节通过仿真实验, 系统性地验证 PE-MADDPG 算法的有效性, 并通过与基线算法的对比和消融实验, 深入分析其性能优势来源。

4.3.1 仿真环境与参数设置

为确保仿真环境的真实性和实验结果的可复现性, 我们参考了 3GPP 非地面网络 (NTN) 标准和相关文献, 并结合我们的研究设定了关键参数。

在我们的仿真场景中, 一颗轨道高度为 600km 的 LEO 卫星以 7.5 km/s 的速度高速移动。地面蜂窝系统由多个 gNB 构成, 环形部署在一个中心 NTN 站周围。系统共享的频谱位于 Ku 频段, 中心频率为 17.5 GHz, 总带宽为 200 MHz, 该带宽被平均划分为 10 个正交子信道。每个仿真时间步长 (Δt) 设置为 20 ms。

信道模型方面, 星地链路遵循阴影莱斯衰落分布, 而地面链路则遵循动态变化的 Nakagami-m 衰落分布, 其路径损耗指数 α 设为 3.5。为了体现算法的泛化能力, 我们将在不同的网络拓扑下进行测试, 其中 gNB 数量 M 的测试范围为 3, 5, 7, 每个 gNB 下的共享用户数 K_m 的测试范围为 4, 8, 12, 16。所有智能体均在基于 $\text{max_gnbs}=7$ 和 $\text{max_users_per_gnb}=16$ 构建的“全尺寸”网络上进行一次训练, 继而利用输入填充 (Padding) 与输出掩码 (Masking) 技术, 在所有小于等于该全尺寸的网络拓扑 (即不同的 gNB 数量与用户密度组合) 下进行零样本 (Zero-shot) 泛化能力测试。

其他关键的系统超参数总结在下表 4-1 中。

表 4-1 关键仿真与 PE-MADDPG 超参数
Table 4-1 Key Simulation and PE-MADDPG Hyperparameters

参数	值
gNB 发射功率范围	15 dBm – 45 dBm
卫星发射功率	46 dBm
地面用户 QoS 门限	10 Mbps
卫星用户 QoS 门限	20 Mbps
噪声功率	-110 dBm
Actor/Critic 隐藏层维度	128 / 256
经验回放池大小	500,000
批次大小	256
折扣因子 (γ)	0.99
学习率 (Actor / Critic)	3e-4 / 1e-4
Epsilon 衰减率 (ϵ -decay)	0.999
目标网络软更新因子 (τ)	0.005

4.3.2 评估指标与基准算法

4.3.2.1 评估指标

为全面、多维度地评估所提算法及基准算法的性能，我们设定了以下三个关键评价指标。

奖励 (Reward)。收敛性是衡量深度强化学习算法训练效果和稳定性的首要指标，在我们的实验中，我们通过绘制平均回合奖励随训练回合数变化的曲线来对其进行评估。一个理想的收敛过程应表现为：奖励曲线在训练初期快速上升，随后增速放缓，并最终在一个较高的奖励水平上趋于平稳，仅有小范围的随机波动，这证明智能体已经学到了一个稳定的、性能优越的策略。

平均频谱效率 (Spectrum Efficiency, SE)。平均频谱效率是衡量资源分配策略优劣的核心性能指标，它直接反映了算法对稀缺频谱资源的利用效率。根据我们的系统模型，全局频谱效率被定义为在整个星地融合系统中，所有通信链路的总数据速率与系统所共享的总带宽之比。在性能对比中，我们将计算在多个测试回合中所有时间步的频谱效率的平均值。该指标越高，表明算法的整体吞吐量性能越好。

平均每步 QoS 违规数 (QoS violation)。用于评估算法在保障用户服务质量方面的公平性和可靠性。一次 QoS 违规被定义为任何一个地面共享用户或卫星用户的实际通信速率低于预设的最低速率门限。我们将统计在所有测试回合的所有时间步中平均每一步发生的 QoS 违规用户数量。这个指标越低，表明算法在追求高效率的同时能更好地兼顾所有用户的基本服务体验，策略的公平性也更优。

4.3.2.2 基线算法

为全面评估 PE-MADDPG 的性能，我们选取了以下基线算法进行对比：

- **GD-FP(Greedy Channel Assignment and Fractional Programming based Power Allocation):** GD-FP 是一种基于传统优化思想的启发式算法。该算法将联合优化问题解耦为两个阶段：首先，在信道分配阶段采用贪婪策略，为每个用户分配当前瞬时信道增益最佳的信道；随后，在功率分配阶段，采用分数阶规划的近似算法来求解局部最优功率解。作为一类经典的非学习式方法，GD-FP 可有效衡量深度强化学习方法相较于传统专家算法的性能优势。

• **独立 Double DQN (Independent Double DQN, IDDQN):** IDDQN 将每个 gNB 建模为一个独立的 DQN 智能体。为处理混合动作空间，该算法将连续功率域离散化为 P 个等级，构建了大小为 $N \times P$ 的联合动作空间。训练中，每个智能体仅依据局部观测和标量奖励进行学习，并将其他智能体的行为视为环境的一部分，缺乏直接的协同机制。该算法是 MARL 领域一个重要的基础学习型基准，通过与它对比，可以突显本文所提 CTDE 框架在处理多智能体协同与克服环境非平稳性方面的优势。

• **QMIX:** QMIX 是一种基于值函数分解的先进 MARL 算法，它遵循 CTDE 范式。该算法中，每个智能体根据局部观测学习一个独立的效用函数 Q_i ，一个混合网络在中心化训练阶段将所有 Q_i 组合成全局联合动作价值函数 Q_{tot} 。该混合网络被施加了单调性约束，即任何 Q_i 的提升都会保证 Q_{tot} 的提升，这一设计确保了算法满足个体-全局最大化原则。然而，正是这种严格的单调性假设，使得 QMIX 难以表示需要智能体通过“自我牺牲”来避免对他人造成“负外部性”的复杂协同策略。作为值函数分解方法的代表性前沿算法，QMIX 提供了一个强大的协同多智能体学习基准，用于衡量本文所提 Actor-Critic 框架的性能与特性。

4.3.3 前瞻性决策的性能增益分析

4.3.3.1 奖励收敛曲线

为评估所提 PE-MADDPG 算法的训练效率与收敛性能，我们将其与 IDDQN (Independent Double DQN) 和 QMIX 两种基于值函数的 DRL 基准算法进行了对比。图4-2展示了三种算法在 8000 个训练回合内的平均回合奖励的变化情况。

从图中可以清晰地观察到三种算法在收敛动态上的显著差异。PE-MADDPG 算法展现出最优的收敛性能，其奖励曲线从一个较高的正值（约 +2）起步，在训练初期经历了快速的性能爬升，并最终稳定收敛在约 +4.5 的最高奖励水平。其较高的起点主要归因于 Actor-Critic 架构的特性：一个随机初始化的 Actor 网络能够输出比“完全随机”更有结构的初始策略（例如，倾向于选择中等功率），从而有效避免了最差的初始性能。而其快速的收敛速度和最高的最终性能，则有力地证明了全局 Critic 网络在指导分布式 Actor 学习复杂协同策略方面的有效性，使得算法能够高效地探索并收敛到高质量的解。

IDDQN 算法的奖励曲线从一个极低的负值（约 -6）开始，这与 ϵ -greedy 探

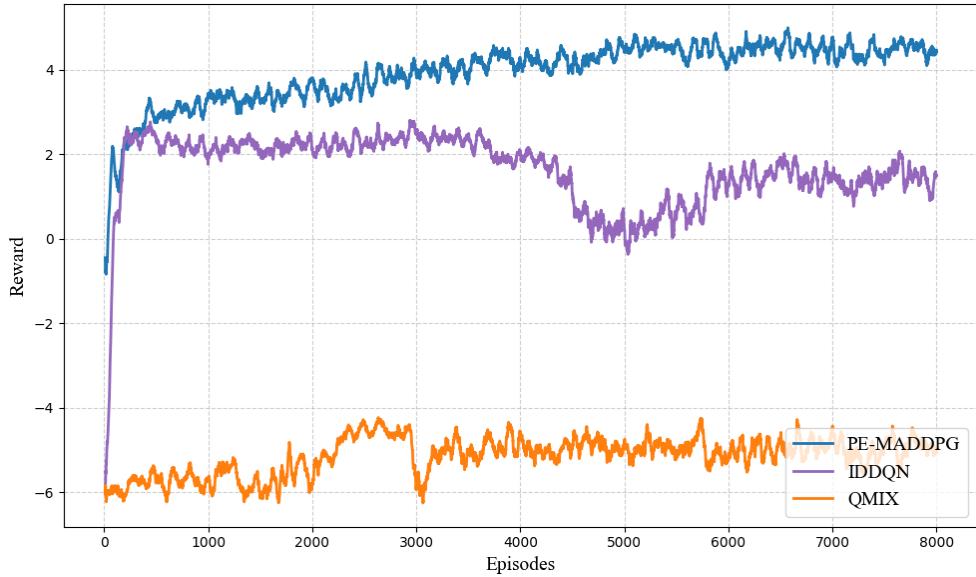


图 4-2 强化学习算法 Reward 收敛曲线

Figure 4-2 The Reward convergence curves for various reinforcement learning algorithms

索策略在训练初期等同于纯随机策略的行为相符，大量的 QoS 违规导致了巨大的初始惩罚。该算法在训练前期同样表现出快速的学习能力，奖励值迅速由负转正并稳定在 +2 左右。然而，在训练的后期（约 4500 回合后），其性能出现了明显的下降。这一现象揭示了，尽管 Double DQN 机制在一定程度上缓解了 Q 值过高估计问题，但在复杂的多智能体独立学习场景中，由环境非平稳性引起的学习不稳定问题依然存在，可能导致算法收敛到次优策略或出现性能波动。

QMIX 算法在本场景下未能实现有效收敛。其奖励曲线虽然也从约-6 的低点开始有缓慢的上升趋势，但始终在负值区域内剧烈震荡，远未达到一个有效的策略水平。这一结果深刻地揭示了 QMIX 算法的核心假设——即个体-全局最大化 (IGM) 原则——与本频谱共享问题的内在特性之间存在根本性冲突。频谱共享是一个充满负外部性的竞争环境，最优的全局策略往往需要某个智能体做出利他式的、对自己非最优的决策。而 QMIX 严格的单调性约束使其难以表示此类复杂的协同行为，导致其在训练中无法找到有效的梯度方向，最终收敛失败。

综上所述，收敛性分析不仅证明了本文所提 PE-MADDPG 算法在收敛速度和最终性能上的显著优势，也通过与 QMIX 的对比，从侧面验证了在处理具有强负外部性的多智能体竞争问题时，灵活的 Actor-Critic 框架相比于受单调性约束的值函数分解方法具有更强的适用性和优越性。

4.3.3.2 不同基站数量对 QoS 性能的影响

图4-3展示了在共享用户密度固定的条件下，随地面网络 gNB 数量从 3 增至 7，本文所提 PE-MADDPG 算法与三种基准算法在平均每步 QoS 违规数上的性能对比。该指标衡量用户服务质量保障能力，其值越低性能越优。

对于所有算法，gNB 数量的增加均导致 QoS 违规数单调递增。这源于固定

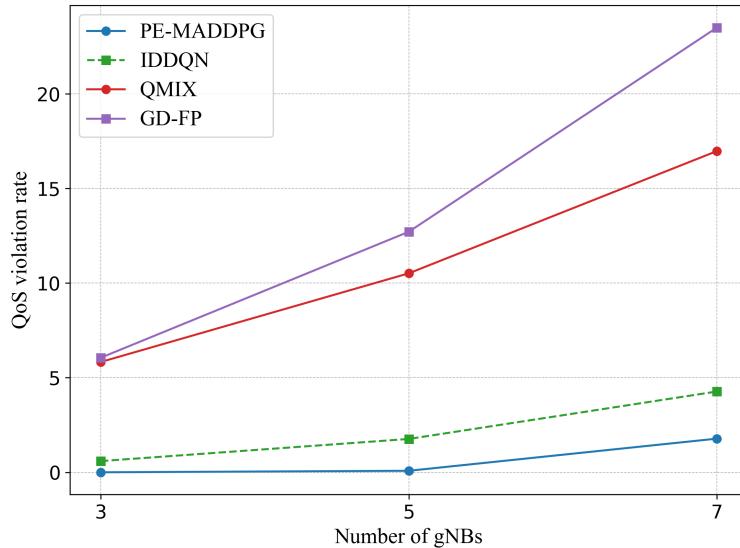


图 4-3 不同 gNB 数量对于平均每步 QoS 违规数的影响

Figure 4-3 The impact of numbers of gNBs on the average number of QoS violations per step

频谱资源下网络密度增加所引起的显著增强的小区间干扰，这使得满足所有用户 QoS 的资源配置更为困难。

在 QoS 保障能力上，PE-MADDPG 与 IDDQN 展现了最优性能，二者曲线几乎重合且 QoS 违规数显著低于其他算法。这表明端到端学习范式能让智能体有效学会规避严重干扰的策略。尤其值得注意的是，PE-MADDPG 在追求高频率效率的同时，其 QoS 保障能力达到了与 IDDQN 相当的水平，未出现预期中的性能牺牲。

QMIX 与 GD-FP 的性能较差，且随 gNB 数量增加急剧恶化。QMIX 的性能瓶颈源于频谱共享中强烈的“负外部性”与其严格的单调性约束相冲突，使其难以学习需要“自我牺牲”的复杂协同。GD-FP 的贪婪决策因忽略了跨 gNB 干扰，在高密度下必然导致严重的“信道堆叠”和大规模小区间干扰。

综上，在保障服务质量和平等性方面，基于深度学习的方法显著优于传统启发式算法和受单调性约束的算法，这凸显了端到端学习在适应复杂高动态干扰环境中的强大能力。PE-MADDPG 在 QoS 保障上达到了与 IDDQN 基准相当的水平，展现了其在效率与公平性之间取得良好平衡的潜力。

4.3.3.3 不同用户密度对 QoS 违规数的影响

图4-4展示了固定共享用户密度 $K_m = 16$ 时，不同 gNB 数量下各算法的平均每步 QoS 违规数。一个共同趋势是，随着 gNB 数量从 3 增至 7，所有算法的平均 QoS 违规数均单调递增。该趋势符合无线通信的基本物理规律：在固定频谱资源下，网络密度增加导致小区间干扰增强，使得满足所有用户 QoS 需求的难度加大。

图中算法性能呈现明显分层。本文提出的 PE-MADDPG 与 IDDQN 算法性

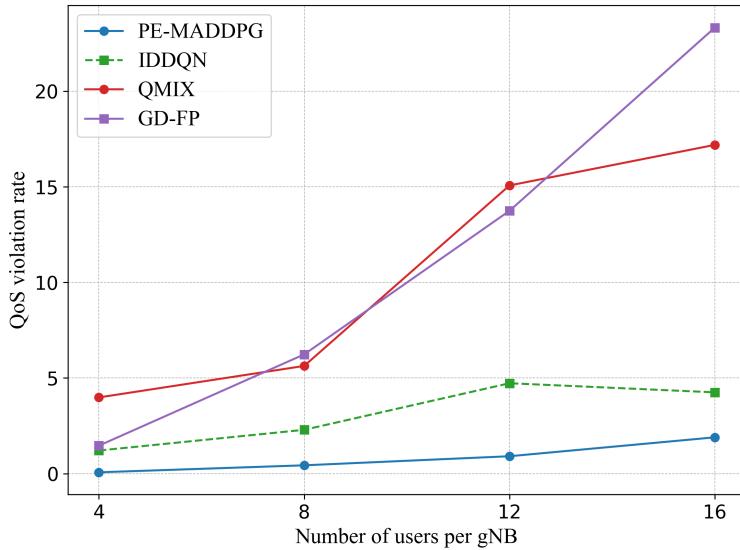


图 4-4 用户密度对每步服务质量 (QoS) 违规次数的平均影响

Figure 4-4 The impact of user densities on the average number of QoS violations per step

能最优，二者的曲线几乎重合，在所有 gNB 配置下均保持了极低的 QoS 违规数，显著优于其他算法。这证明了端到端的学习范式能让智能体有效规避严重干扰、保障用户服务质量。值得注意的是，PE-MADDPG 在追求高频谱效率的同时，其 QoS 保障能力并未被牺牲，达到了与 IDDQN 相当的领先水平。

QMIX 与 GD-FP 算法的性能较差，且随 gNB 数量增加而急剧恶化。QMIX 性能受限的原因在于，强烈的“负外部性”与其严格的单调性约束产生冲突，使其难以学习需要“自我牺牲”的协同策略。GD-FP 作为一种贪婪算法，决策时仅考虑本地信道增益，忽略了对其他 gNB 产生的干扰。随着 gNB 增多，这种“自私”决策必然导致“信道堆叠”与严重干扰，使大量用户 QoS 无法满足。

综上所述，图4-4的结果证明，基于深度学习的资源分配方法在保障用户服务质量与公平性方面，显著优于传统启发式算法与受单调性约束的 MARL 算法。这凸显了端到端学习框架在适应复杂高动态干扰环境方面的强大能力。本文所提 PE-MADDPG 算法在 QoS 保障方面达到了与领先基准相当的水平，展现了其在效率与公平间取得良好平衡的潜力。

4.3.3.4 不同基站数量对频谱效率的影响

图4-5展示了固定用户密度下，不同 gNB 数量时各算法的平均频谱效率对比。一个共同趋势是，所有算法的平均频谱效率均随 gNB 数量增加而单调递增。这表明，尽管增加 gNB 会加剧小区间干扰，但同时也极大地增加了空间复用机会。只要能有效管理干扰，系统便能通过复用相同频段来显著提升总吞吐量。

图中算法的性能排序从优至劣依次为：PE-MADDPG > IDDQN > QMIX ≈ GD-FP，这凸显了各算法在最大化系统吞吐量上的能力差异。本文提出的 PE-MADDPG 算法在所有场景下均取得了最优频谱效率，且其优势随 gNB 数量增加

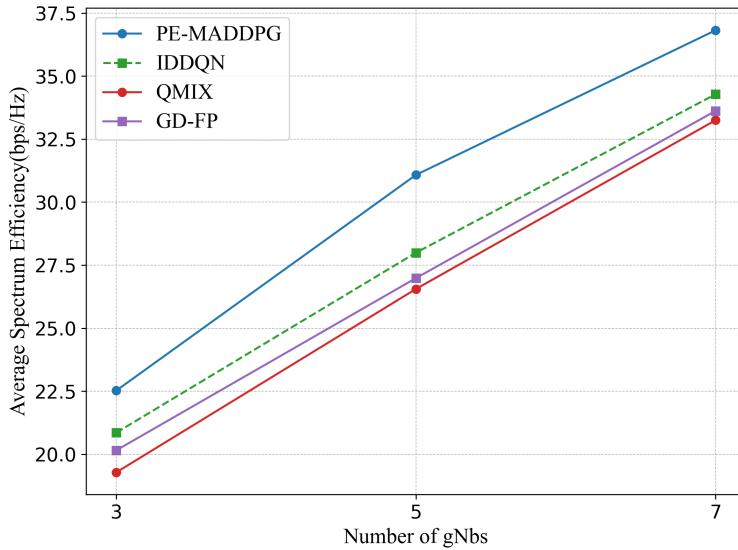


图 4-5 不同小区下 gNB 数量对于平均 SE 的影响

Figure 4-5 The impact of numbers of gNBs on the average SE

而愈发显著。这证明了其基于 Actor-Critic 的 CTDE 框架的强大能力，全局 Critic 能够评估联合动作的长期影响，引导各 Actor 学习到一种能高效平衡信道质量与干扰规避的协同策略。

IDDQN 作为一种非协同学习算法，其性能优于 QMIX 和 GD-FP，表明其端到端学习范式能让智能体在一定程度上适应复杂干扰环境。QMIX 的性能与 GD-FP 相近，但显著落后于 PE-MADDPG 和 IDDQN。这进一步证实了 QMIX 的局限性，其严格的单调性约束使其难以学习到需要“自我牺牲”以换取全局最优的复杂策略。GD-FP 作为一种贪婪启发式方法，其短视性使其无法应对复杂、连锁的干扰效应，导致在网络复杂度增加时性能增长乏力。

综上所述，图4-5的结果证明了本文所提 PE-MADDPG 算法在最大化系统全局频谱效率方面的卓越性能。通过其灵活的 Actor-Critic 协同框架，它成功克服了高密度网络的复杂干扰问题，学习到的资源分配策略优于其他基准算法，在大规模、高干扰场景下优势尤为明显。

4.3.3.5 不同用户密度对频谱效率的影响

图4-6展示了固定拓扑下，各算法的平均频谱效率随每个 gNB 下属共享用户数从 4 增加到 16 的变化。一个共同趋势是，所有算法的平均频谱效率均随用户密度增加而单调递增。这主要得益于多用户分集与 NOMA 增益。用户越多，出现信道条件极佳用户的概率越大，算法可优先为其服务以提升整体速率。用户数超过信道数时，系统采用 NOMA 技术复用用户，通过串行干扰消除提升频谱效率，图中从用户数 8 到 16 的性能陡增即为 NOMA 增益驱动的体现。

图中算法的性能排序随用户密度变化，揭示了其在处理系统复杂性上的能力差异。本文提出的 PE-MADDPG 算法在所有用户密度下均取得了最优频谱效

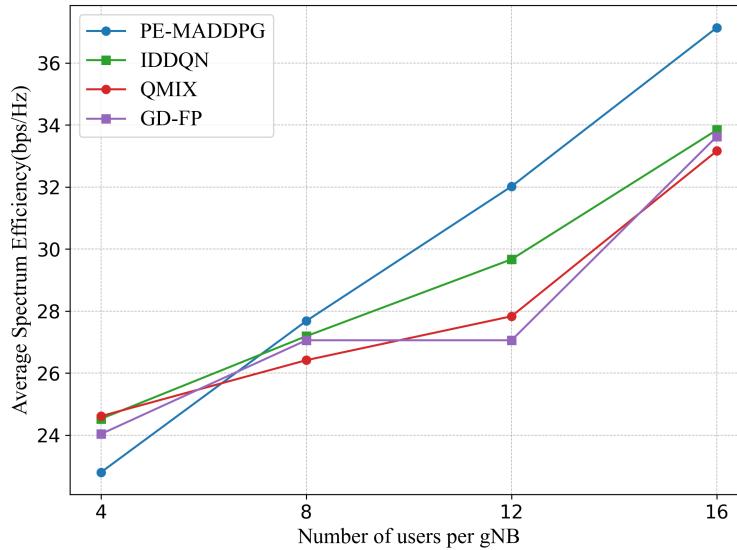


图 4-6 不同小区下的用户密度对于平均 SE 的影响
Figure 4-6 The impact of user densities on the average SE

率，并在高密度场景下优势进一步扩大。这证明了其灵活的 Actor-Critic 协同框架在处理高密度、高干扰复杂场景时的卓越能力。全局 Critic 能够引导各 Actor 学习精妙的协同策略，实现精细化的连续功率控制，以最大化 NOMA 增益并抑制小区间干扰。

IDDQN 与 QMIX 在中低密度时表现尚可，但在高密度下性能增长放缓，与 PE-MADDPG 的差距逐渐显著。这主要归因于其采用的功率离散化技术限制了它们充分利用 NOMA 潜力的能力。GD-FP 算法在低密度时表现不佳，但在高密度下性能出现“反常”陡增。当用户远超信道数时，其“小区内最优匹配”逻辑能为每个信道找到条件较好的用户，奠定了性能基础。然而，其贪婪本质无法处理小区间干扰，导致性能上限依然受限。

综上所述，图4-6的结果证明，用户密度的增加为系统带来了显著的多用户分集和 NOMA 增益。本文所提 PE-MADDPG 算法凭借其连续功率控制和高效协同学习的优势，在高密度、高复杂度场景下展现出最强性能，能最充分地挖掘 NOMA 技术潜力，实现远超其他基准算法的频谱效率。

4.3.3.6 预测敏感性分析

为量化评估 P-MADDPG 算法对前瞻性预测信息的依赖性与鲁棒性，本节通过系统性改变 PRB 利用率预测的准确率，分析算法关键性能指标的变化。为模拟不同精度的外部预测模块，本文采用注入可控高斯白噪声的方法。在每个时间步 t ，首先通过启发式策略生成一个客观的未来 PRB 利用率基准值 $\hat{U}_{ideal}(t+1)$ 。随后，生成均值为 0、标准差为 σ_{noise} 的高斯白噪声。最终提供给智能体的预测值 $\hat{U}(t+1)$ 通过对理想值注入噪声并裁剪至有效范围获得。

$$\hat{U}(t+1) = \text{clip}(\hat{U}_{ideal}(t+1) + N(0, \sigma_{noise}^2)) \quad (4-33)$$

其中, $\hat{U}(t+1)$ 为最终提供给智能体的预测值, $\hat{U}_{ideal}(t+1)$ 是客观的未来 PRB 利用率基准值, $N(0, \sigma_{noise}^2)$ 代表均值为 0、方差为 σ_{noise}^2 的高斯白噪声。该公式通过对理想值添加可控噪声并裁剪至有效范围 $[0, 1]$, 用以模拟具有不同准确率的预测信息。

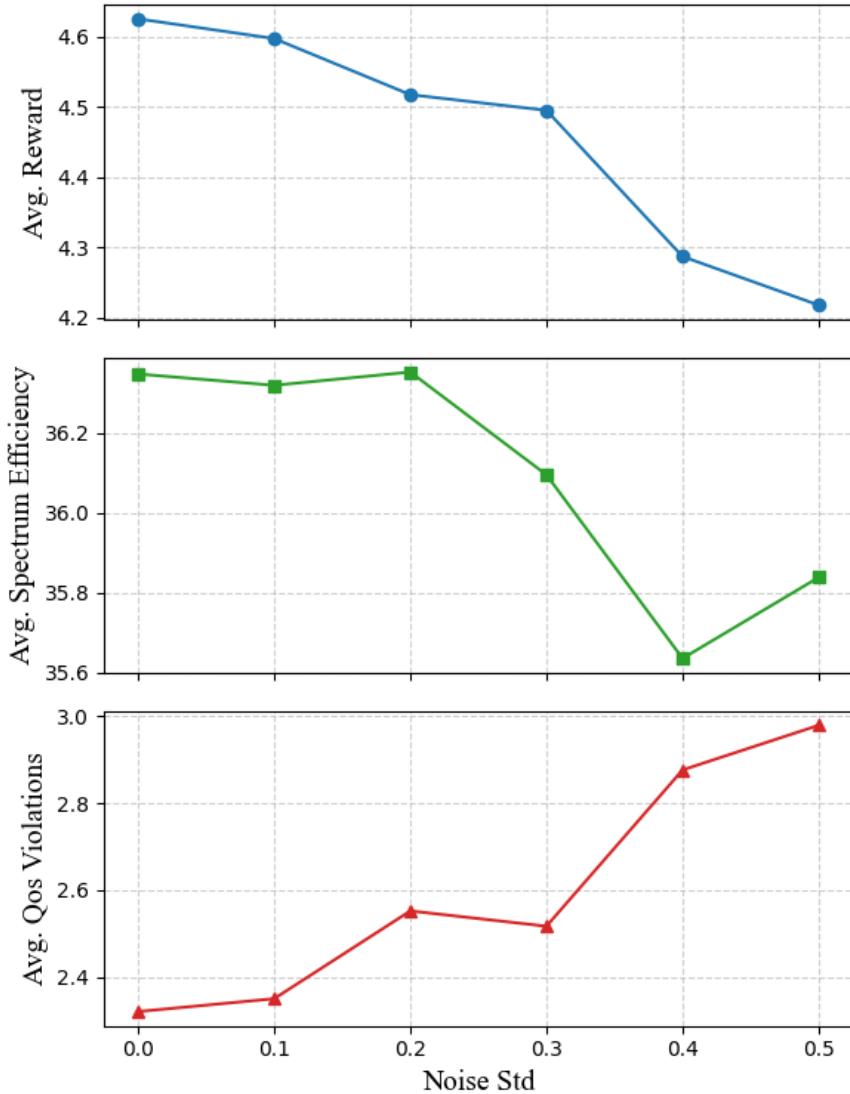


图 4-7 PRB 利用率预测准确度对算法性能的敏感性分析

Figure 4-7 Sensitivity analysis under the accuracy of PRB utilization prediction

图4-7展示了在不同预测噪声标准差下, PE-MADDPG 算法关键性能指标的变化。如图所示, 算法性能均对预测信息准确度表现出明显敏感性。平均奖励随噪声标准差增加呈单调下降趋势, 表明智能体已学会利用准确未来信息优化决策以获取更高长期回报。平均频谱效率亦随噪声增大而下降, 说明准确预测有助于智能体做出更高效的资源分配决策, 以提升系统总吞吐量。从服务质量角度看, 平均 QoS 违规数随噪声增加而显著上升, 证明智能体利用前瞻性信息进行主动干扰管理和资源预留, 以保障用户服务质量。当预测准确性降低时, 其保障

QoS 的可靠性也随之下降。综上所述，这三组结果共同表明，准确的未来预测信息使智能体能做出更优决策，其优越性同时体现在更高的系统效率、更好的用户体验以及更高的综合回报上，有力证实了引入前瞻性预测信息的创新价值。

4.4 本章小结

本章对所提出的 PE-MADDPG 算法进行了全面的仿真评估，旨在解决星地融合网络中的动态协作频谱共享问题。实验通过将 PE-MADDPG 与多种基准算法（包括 QMIX、IDDQN、GD-FP 等）在不同网络规模（变化的 gNB 数量与用户密度）和外部预测信息准确度的条件下进行性能对比，以验证其有效性和鲁棒性。实验结论明确指出：本文所提出的 PE-MADDPG 算法在最大化全局频谱效率方面展现出卓越性能，尤其在高密度、高干扰的复杂场景下，其优势愈发显著，全面超越了所有基准算法。此外，实验不仅验证了所提模型具备在单一训练后适应不同网络拓扑的出色泛化能力，还通过敏感性分析量化了前瞻性预测信息对决策的积极影响，为算法设计的有效性提供了坚实的数据支撑。

