Topology-Aware vs Feature-Based QoT Estimation in Multiband Optical Networks

Jana Eid[1,*], Franco Galante[1], Omran Ayoub[2], Giovanni Simone Sticca[3],
Mëmëdhe Ibrahimi[3], Andrea Bianco[1], Massimo Tornatore[3], Cristina Rottondi[1]
[1]Politecnico di Torino, Turin, Italy; [2]University of Applied Sciences of Southern Switzerland, Lugano, Switzerland
[3]Politecnico di Milano, Milan, Italy

***Abstract*—Accurate estimation of the Signal-to-Noise Ratio (SNR) of prospective lightpaths is a key requirement for the efficient operation of dynamic multiband optical networks. This paper compares a topology-aware approach based on Graph Neural Networks (GNNs) to a traditional feature-based eXtreme Gradient Boosting (XGBoost) regressor for SNR prediction. The GNN exploits knowledge of the complete network topology and per-link spectral occupation, while XGBoost relies on a compact set of features describing the closest spectral environment along the selected path.**

The two approaches are evaluated under an incremental traffic-loading scenario, in which the C band is filled firstly, and the L band is later activated when needed. Results show that the GNN provides higher accuracy and more balanced prediction errors during early network operation, when spectral occupation is sparse and limited training data are available. As the network evolves toward full multiband utilization, XGBoost achieves lower mean absolute and mean squared errors, but exhibits systematic overestimation of SNR for long-haul lightpaths. In contrast, the GNN maintains better-calibrated predictions, especially for long lightpaths, as reflected by lower mean and median errors.

***Index Terms*—QoT Estimation, Multiband Optical Networks, Graph Neural Networks**

I. Introduction

The continuous growth of traffic demands and service heterogeneity is driving optical transport networks toward new flexible and spectrum-efficient technologies. In this context, multiband optical networks, extending transmission beyond the conventional C band to additional spectral regions such as the L band, have emerged as a promising approach to significantly expand capacity without disruptive infrastructure upgrades [1]. However, extending the spectral range increases the complexity of Quality of Transmission (QoT) estimation for prospective lightpaths, as band-dependent physical impairments, heterogeneous amplification characteristics, and inter-band spectral interactions must be considered.

Traditional QoT estimation relies on analytical models or Machine Learning (ML) approaches that use as input features describing the lightpath and its spectral environment [2], [3]. Tree-based models such as eXtreme Gradient Boosting (XGBoost) have demonstrated competitive performance under diverse network conditions [4]. However, their reliance on compact, predefined feature sets limits awareness of the global network state, particularly in multiband scenarios where cross-band interactions become relevant.

Recently, *Graph Neural Networks* (GNNs) have emerged as a powerful alternative for learning from structured network data by explicitly encoding graph topology and relational dependencies through message passing mechanisms [5]. By representing optical networks as graphs, GNNs can exploit both topological information and per-link attributes, such as spectral occupation. This makes the model able to capture long-range dependencies and wide-range spectral context that are typically inaccessible to traditional feature-based ML models.

While these properties suggest that GNNs are well suited to address the growing complexity of QoT estimation in dynamic multiband optical networks, their practical advantages over feature-based approaches remains insufficiently investigated. In particular, it is necessary to quantify whether topology-aware learning improves prediction accuracy and generalization under evolving spectral conditions, and whether such benefits persist across both single-band and multiband operation.

To address these questions, we design a GNN-based QoT estimator and compare it with an XGBoost-based approach in an optical network operating over the C+L bands under incremental traffic loading. The GNN directly ingests the complete network topology and per-link spectral occupation, while XGBoost relies on a limited set of predefined features characterizing the spectral neighbors and aggregate occupation of the traversed links. Both models are benchmarked against a baseline XGBoost regressor that does not leverage any knowledge of the spectral context, and are retrained and evaluated across multiple time epochs to assess adaptability under evolving network conditions.

Results show that while XGBoost achieves strong generalization performance in terms of Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) as the network evolves toward multiband utilization, it is more prone to systematic bias, particularly for long lightpaths. In contrast, the GNN-based model yields more balanced error distributions, with lower mean and median errors, indicating that topology and spectrum awareness primarily improve bias-related indicators rather than aggregate accuracy metrics.

The remainder of the manuscript is organized as fol-

This work has been partially supported by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, partnership on “Telecommunications of the Future” (PE00000001 - program “RESTART”) and by the Swiss Innovation Agency Innosuisse through the innovation project SUSTAINET (No. 119.588 INT-ICT), carried out within the EUREKA Cluster CELTIC-NEXT SUSTAINET-Advance project.

lows: Sec. II briefly overviews the related literature, whereas Sec. III presents the architecture of the XGBoost- and GNN-based SNR predictors. Simulation scenarios are described in Sec. IV-A and the performance comparison of the considered models is discussed in Sec. V. Finally, Sec. VI ends the paper.

II. Related Work

Traditional QoT estimation relies on analytical models that compute physical impairments like amplified spontaneous emission and nonlinear losses from lightpath parameters [6]. While accurate, these approaches scale poorly with network size [7], [8]. ML-based techniques, and particularly Deep Neural Networks (DNNs), advanced the field further by learning hierarchical feature representations, but they cannot ingest all the information on spectral loading and neighboring channels, since they rely on a limited set of predefined features characterizing the spectral vicinity of prospective lightpaths, typically accounting only for the spectrally-nearest left/right neighbors [9]. GNNs have emerged as a promising paradigm for QoT estimation by explicitly modeling network topology and lightpath interactions. Panayiotou et al. [7] first applied Deep Graph Convolutional Neural (DGCN) networks to estimate QoT for unseen network states, demonstrating the value of topology-aware learning. The work addressed in [10] developed a supervised DGCN for Optical SNR (OSNR) estimation in optical mesh networks, showing that GNN architectures operating directly on graph representations outperform traditional DNNs. More recently, Savva et al. [11] applied deep graph learning to capture crosstalk impacts on in-service lightpaths, achieving over 92% accuracy compared to 77% for per-lightpath DNN approaches. GNN-based meta-learning techniques have also been explored for QoT estimation, as demonstrated in [12], where rapid model adaptation is enabled to support dynamic virtual network topology reconfiguration under changing network conditions.

Despite these advances, existing GNN-based QoT estimation studies have focused primarily on C-band-only scenarios, leaving multiband transmission largely unexplored. Moreover, the temporal generalization capability of GNNs, in comparison with traditional feature-engineered methods such as XGBoost, remains insufficiently investigated under evolving network conditions. This study aims to address these gaps by deriving operational insights that can guide model selection and deployment decisions across different stages of network loading.

III. Problem Formulation and GNN Architecture

A. Problem Formulation

Let the multiband Wavelength-Division Multiplexed (WDM) optical network be modeled as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of ROADMs and $\mathcal{E}$ is the set of fiber links connecting them (each fiber is represented as two directed edges $e \in \mathcal{E}$, one per direction), characterized by physical parameters such as length, attenuation, dispersion, and current spectral occupation in both the C and L bands, assuming a 50GHz-spaced grid. A lightpath request r_i is defined by a source-destination node pair (s_i, d_i), an assigned path $\mathcal{P}_i \subseteq \mathcal{E}$, and an assigned slot f_i.

The goal is to learn a function $\hat{y}_i = f(\mathcal{X}_i)$, where $\hat{y}_i$ denotes the predicted SNR of the candidate lightpath r_i, and $\mathcal{X}_i$ represents the information available at provisioning time. The learning model $f(\cdot)$ is trained to minimize the MSE between predicted and actual (post-deployment) SNR values.

B. Graph Neural Network-based Model

Our GNN-based approach takes as input the network graph $\mathcal{G}$, enriched with per-link attributes capturing the spectral occupation in each band. For lightpath r_i, the model input is $\mathcal{X}^{\text{GNN}} = \{\mathcal{G}_{\text{line}}, \mathcal{P}_i, f_i\}$. where $\mathcal{G}_{\text{line}}$ is a line graph derived from $\mathcal{G}$. Such transformation is needed because GNNs require per-node attributes, instead of link-associated features. The GNN learns to propagate and aggregate information across the input network topology, yielding the predicted SNR $\hat{y}_i$.

Line Graph: The line graph $\mathcal{G}_{\text{line}} = (\mathcal{V}_{\text{line}}, \mathcal{E}_{\text{line}})$ is built by setting $\mathcal{V}_{\text{line}} = \mathcal{E}$, i.e., each node in the line graph corresponds to a directed physical link. Two nodes $u, v \in \mathcal{V}_{\text{line}}$ are connected if:

$$(u, v) \in \mathcal{E}_{\text{line}} \iff \text{destination}(u) = \text{source}(v) \text{ in } \mathcal{G} \tag{1}$$

This representation enables the GNN to propagate information along physically realizable lightpaths, as edges in $\mathcal{G}_{\text{line}}$ connect exclusively those nodes representing physical links that can be traversed consecutively.

Node Features: Each node $v \in \mathcal{V}_{\text{line}}$ encodes the spectral state of the corresponding physical link. We represent the spectrum using N frequency slots, where each slot is described by a 7-dimensional feature vector capturing the occupying lightpath's characteristics (number of links, path length, longest link, frequency, SNR, traffic, modulation) or set to zero if empty. Concatenating these with the normalized link length yields $\mathbf{x}_v \in \mathbb{R}^d$, with $d = 7 \times N + 1$. For prospective lightpath r_i at insertion step i (where i captures the arrival order of traffic requests), node features $\mathbf{x}_v$ encode the spectral state at step $(i-1)$, representing the network environment before establishment.

Lightpath Context: The route of the prospective lightpath is specified via a binary mask $\mathbf{m}_i \in \{0,1\}^{|\mathcal{V}_{\text{line}}|}$ where $m_i(v) = 1$ if $v \in \mathcal{P}_i$, whereas the feature vector $\mathbf{l}_i \in \mathbb{R}^{n_l}$ contains its per-link slot assignments and attributes (path length, number of links, longest link, frequency, traffic, modulation).

GNN Architecture: The model consists of three stages. *1) Message passing:* L GCN layers propagate information across the topology:

$$\mathbf{h}_v^{(\ell+1)} = \text{ReLU}\left(\sum_{u \in \mathcal{N}(v) \cup \{v\}} c_{uv} \mathbf{W}^{(\ell)} \mathbf{h}_u^{(\ell)} + \mathbf{b}^{(\ell)}\right) \tag{2}$$

where $\mathbf{h}_u^{(\ell)}$ is the hidden representation of node u at layer ℓ, $c_{uv} = \frac{1}{\deg(u) \cdot \deg(v)}$, $\mathbf{h}_v^{(0)} = \mathbf{x}_v$, $\mathbf{W}^{(\ell)} \in \mathbb{R}^{d_h \times d_h}$ and $\mathbf{b}^{(\ell)} \in \mathbb{R}^{d_h}$ are learnable weight and bias parameters.

2) Graph pooling: Masked aggregation is performed over traversed links as follows: $\mathbf{h}_{\text{path}} = \text{AGG}(\{\mathbf{h}_v^{(L)} : m_i(v) = 1\})$, where AGG $\in \{\text{mean}, \max, \text{sum}\}$.

3) Readout: A Multi-Layer Perceptron (MLP) combines graph and lightpath features: $\hat{y}_i = \text{MLP}([\mathbf{h}_{\text{path}} \| \mathbf{l}_i])$.

Training: The model's learnable parameters Θ [1] are optimized by minimizing the Huber loss (Smooth L1): $\mathcal{L}(\Theta) = \frac{1}{N}\sum_{i=1}^{N} h_\beta(\hat{y}_i - y_i)$, where the robust loss function h_β is defined as:

$$h_\beta(e) = \begin{cases} \frac{1}{2}e^2 & \text{if } |e| \leq \beta \\ \beta(|e| - \frac{\beta}{2}) & \text{otherwise} \end{cases} \quad (3)$$

with $\beta = 1.72$. This loss function provides robustness to outliers while maintaining quadratic behavior for small errors. All continuous features undergo min-max normalization to $[0, 1]$, with statistics computed over the entire dataset.

C. XGBoost-based Model

This approach employs an XGBoost regressor trained on a set of handcrafted features. We evaluate two variants that differ in their treatment of spectral neighbor information.

XGB-NN (No Neighbors): This baseline variant uses a compact 6-dimensional feature vector: $\mathcal{X}_i^{\text{XGB-NN}} = [\ell_i, |\mathcal{P}_i|, \max_{e \in \mathcal{P}_i} \ell_e, f_i, t_i, m_i]^\top \in \mathbb{R}^6$ where ℓ_i is path length, $|\mathcal{P}_i|$ is number of links, $\max_{e \in \mathcal{P}_i} \ell_e$ is longest link length, f_i is frequency, t_i is traffic demand, and m_i is modulation format.

XGB-WN (With Neighbors): This ensemble variant employs two complementary models. The **0N+1N** model uses the same features as XGB-NN and predicts the QoT of lightpaths with 0 or 1 spectral neighbors. The **2N** model augments these with 14 neighbor descriptors: $\mathcal{X}_i^{\text{XGB-2N}} = [\mathcal{X}_i^{\text{XGB-NN}} \| \mathbf{n}_i^{\text{left}} \| \mathbf{n}_i^{\text{right}}] \in \mathbb{R}^{20}$ where each $\mathbf{n}_i^{\text{left/right}} \in \mathbb{R}^7$ encodes the adjacent lightpath's characteristics (number of links, path length, longest link length, frequency slot, SNR, traffic volume, modulation format). The 2N model processes only lightpaths with both neighbors present. At inference, predictions from both models are combined across their respective test samples. Both variants use squared error loss function and share identical hyperparameters.

IV. Simulation Settings

A. Simulated Scenarios

To address the research objectives outlined in Sec. I, we consider an incremental network-loading process, in which lightpaths are sequentially established over time. This allows us to analyze whether the potential advantages of GNN-based models persist as the network transitions from single-band to multiband utilization, and to identify the operational phases in which their additional modeling complexity is justified when compared to XGBoost-based approaches.

We consider a continental European topology (We consider the 14-node Japan network topology (22 bidirectional links, see Fig. 4 in [13]), and generate traffic demands between randomly chosen (s_i, d_i) node pairs, with data rates in the range [100, 200, 300, 400] Gbps, and modulation formats [QPSK, 16-QAM, 32-QAM, and 64-QAM]. We perform the Routing and Spectrum Assignment following a shortest-path, first-fit approach [14], [15]. Each generated lightpath is characterized by an SNR at the receiver, calculated considering the linear and the non-linear noise (as in [16]), which depends on the lightpath data rate and modulation format. Overall, the dataset is composed of 16,678 lightpaths.

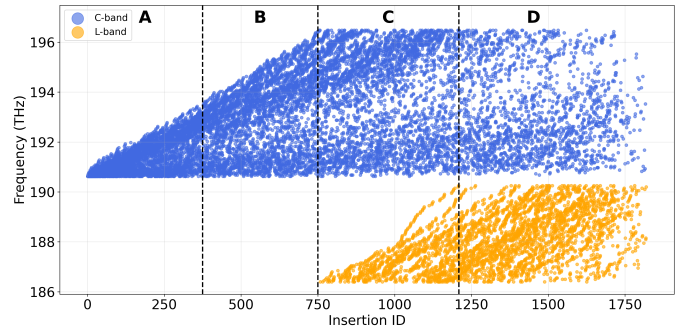

Fig. 1. Temporal evolution of spectral allocation of C-band and L-band lightpaths across epochs A-D.

To analyze the evolution of the learning process and the relative performance of the investigated models, the dataset was divided into four distinct time epochs, denoted as A, B, C, and D, as illustrated in Fig. 1. The horizontal axis represents the insertion step i, while the vertical axis represents the spectrum range. Blue (above) and orange (below) points correspond to the lightpaths frequency allocation in the C-band and L-band respectively. The four epochs were defined as follows. **Epoch A** corresponds to the initial phase, with low network load, including the first half of the C-band allocations (3,745 samples). **Epoch B** contains the second half of the C-band samples (3,745 samples), up to the moment when the first L-band lightpath is established. **Epoch C** starts with the insertion of the first L-band lightpath and includes the first half of the coexisting C- and L-band samples (3,541 C-band samples, 2,565 L-band samples). **Epoch D** includes the second half of the coexisting band samples (2,029 C-band samples, 2,565 L-band samples).

We define a set of scenarios that consider both single-band and multiband QoT predictors, depending on the stage of the network evolution, with distinct combinations of training and testing datasets, summarized in Tab. I. For each configuration, the training and testing epochs are specified, as well as the spectral bands involved. More in detail, epochs A and B represent scenarios where only C-band samples are available. In these stages, both the GNN and the XGB models are trained exclusively on C-band lightpaths, first on epoch A alone, and then on the combined set A+B, after which they are tested on unseen C-band samples belonging to the subsequent epoch. Then, in epoch C the first L-band lightpaths are introduced: these are leveraged for training i) a mixed-band model, trained on both C and L-band lightpaths gathered during epochs A, B,

[1] Parameters Θ comprise all the GNN's weights $\mathbf{W}^{(\ell)}$ and biases $\mathbf{b}^{(\ell)}$, and the readout MLP parameters

TABLE I
SIMULATION MATRIX ACROSS EPOCHS FOR GNN, XGB-NN, AND XGB-NN/XGB-WN MODELS.

Model	Epoch A	Epoch B			Epoch C			Epoch D		
		Train-Test Band	Train Set	Test Set	Train-TestBand	Train Set	Test Set	Train-Test Band	Train Set	Test Set
GNN	Collect training data	C/C	A	B	C/C	A+B	C	C+L/C+L	A+B+C	D
								C/C	A+B+C	
								L/L	C	
XGB-NN	Collect training data	C/C	A	B	C/C	A+B	C	C+L/C+L	A+B+C	D
								C/C	A+B+C	
								L/L	C	
XGB-NN **XGB-WN**	Collect training data	C/C	A $A_{(2N)}$	$B_{(0N+1N)}$ $B_{(2N)}$	C/C	A+B $A+B_{(2N)}$	$C_{(0N+1N)}$ $C_{(2N)}$	C+L/C+L	A+B+C $A+B+C_{(2N)}$	$D_{(0N+1N)}$ $D_{(2N)}$
								C/C	A+B+C $A+B+C_{(2N)}$	
								L/L	C $C_{(2N)}$	

Subscript 2N indicates lightpaths exhibiting both left and right spectral neighbors, subscript 0N+1N indicates lightpaths missing one or both neighbors.

TABLE II
HYPERPARAMETER SEARCH SPACE FOR GNN MODEL OPTIMIZATION

Category	Hyperparameter	Search Range
Architecture	Hidden dimension d_h	64 – 512
	Number of GCN layers	2 – 8
	Number of MLP layers	1 – 5
	MLP layer size	32 – 512
	Graph-level aggregation	{mean, max, sum}
	Dropout rate p_{drop}	0.0 – 0.5 (step: 0.1)
Optimizer	Learning rate α	5×10^{-5} – 5×10^{-3} (log)
	Weight decay λ	1×10^{-6} – 3×10^{-3} (log)
	LR scheduler*	8 scheduler types
Training	Batch size	16 – 64
	Maximum epochs	50 – 200
Regularization	Gradient clipping norm	{None, 0.5, 1.0, 2.0}
Loss Function	Loss type	{MSE, SmoothL1}
	Huber $\beta^\dagger$	0.1 – 2.0
Early Stopping	Patience (epochs)	5 – 15
	Min. improvement δ	1×10^{-5} – 1×10^{-3} (log)

*Eight scheduler types: ReduceLROnPlateau, StepLR, CosineAnnealingLR, OneCycleLR, CosineAnnealingWarmRestarts, ExponentialLR, LinearLR, and SequentialLR. Each has 2–4 specific hyperparameters that were also optimized (e.g., decay factors, periods, patience values). †Only when loss type is SmoothL1.

and C; ii) two models exclusively dedicated either to the C or the L-band. Both models are then tested in epoch D, after being trained using all the respectively available lightpaths deployed during epochs A, B, and C.

B. Hyperparameter Settings

We performed hyperparameter optimization for both the GNN and XGBoost models using Optuna [17] with the Tree-structured Parzen Estimator (TPE) sampler and MedianPruner. To ensure robust parameter selection independent of temporal dynamics, we employed 5-fold cross-validation across 3 random seeds.For the GNN, the learning rate was critical for performance; hence, we evaluated 9 distinct scenarios: eight scheduler types and a baseline without a scheduler. The complete search space is detailed in Tab. II. The resulting GNN configuration is a 4-layer GCN (hidden size 176) followed by a bottleneck-shaped MLP with hidden layers [224, 192, 96, 256], trained using the Huber loss (SmoothL1) function ($\beta = 1.72$) and the CosineAnnealingWarmRestarts [18] scheduler, with graph pooling using mean aggregation. For XGBoost, optimization yielded the following configuration: 800 estimators, learning rate 0.06, max depth 5, minimum child weight 3, subsample ratio 0.9, and L2 regularization $\lambda = 8.0$. For each model and epoch, we compare the performance of the three different predictors presented in Sec III.

V. NUMERICAL ASSESSMENT

We compare the GNN model to the baseline XGBoost models in terms of Mean Absolute Error (MAE) and Mean Squared Error (MSE), to quantify the overall accuracy. Additionally, Mean Error (ME) and Median Error (MedE) are computed to capture systematic bias and to characterize the central tendency of the error distribution, which is less influenced by outliers. Fig. 2 summarizes the performance comparison across the three different testing epochs (B, C, and D) and error metrics. Epochs B and C report results for single-band configurations, while Epoch D additionally distinguishes performance across frequency bands (C, L, and C+L), indicated by different hatch patterns.

A. Per-Epoch and Band-Specific Accuracy Performance

1) Epoch B: With limited training data available from Epoch A and sparse spectral occupation, the availability of full spectral context proves particularly beneficial. In this phase, the GNN achieves a MAE of 0.358 dB, outperforming XGB-NN (0.455 dB) and XGB-WN (0.453 dB) by approximately 21%. Moreover, the GNN exhibits mean errors closer to zero and lower median errors, indicating reduced systematic bias and more balanced prediction errors.

Takeaway: In early deployment phases characterized by sparse spectrum usage and limited training data, topology- and spectrum-aware learning provides clear benefits in terms of bias mitigation, yielding more balanced and accurate QoT predictions for the majority of lightpaths and reduced sensitivity to outliers.

2) Epoch C: As additional training samples become available, performance differences across models narrow. While XGB-NN achieves slightly lower MAE and MSE than both XGB-WN and GNN, the latter continues to exhibit lower mean and median errors, indicating reduced bias and improved accuracy for the central portion of the error distribution.

Takeaway: As training data accumulate under stable single-band operation, feature-based models match or even outperform topology-aware approaches in terms of aggregate accuracy. However, full spectral awareness remains beneficial

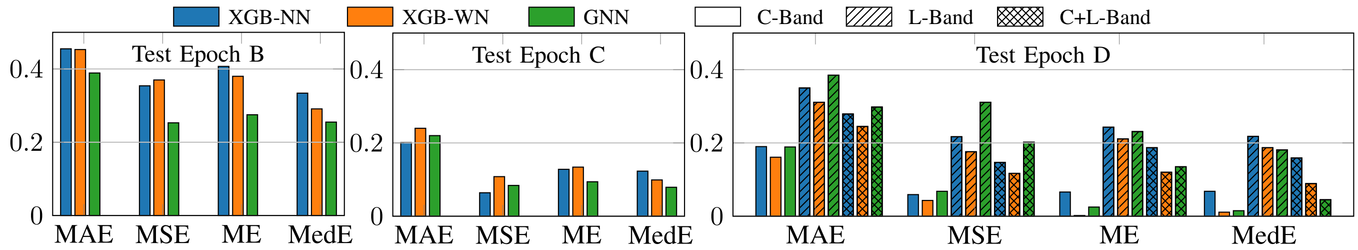

Fig. 2. Performance comparison across epochs and metrics (MAE, MSE, ME, MedE).

for bias reduction and for improving prediction reliability for the majority of lightpaths.

3) Epoch D: Once both C- and L-band traffic coexist and models are trained and evaluated across the two bands, XGB-WN achieves the lowest MAE, outperforming both XGB-NN and GNN. This suggests that, in a mature multiband network, carefully engineered features capturing local left–right spectral interactions together with coarse global occupancy statistics are sufficient to model QoT dependencies effectively. Nevertheless, the GNN achieves the lowest MedE, indicating fewer systematic over- or under-estimations.

Takeaway: In mature multiband scenarios, feature-engineered models are well suited for minimizing average error, while topology-aware learning primarily contributes to improved calibration by reducing large prediction errors affecting the majority of lightpaths.

4) Single Multiband vs. Ensemble Single-Band Models: When models are specialized on the C band, where the density of historical samples is highest and spectral loading is relatively stable, all approaches exhibit comparable MAE performance, with XGB-WN retaining a slight advantage. Here, the benefit of spectral awareness is more pronounced in bias-related metrics, as reflected by lower mean and median errors, while gains in MAE and MSE remain limited. When models are specialized for the L band, where training data are sparser and spectrum occupation is more heterogeneous, overall errors increase for all approaches. While XGB-WN continues to achieve the lowest MAE, the GNN attains the lowest MedE. Fig. 3 further compares the performance obtained at epoch D by combining in an ensemble the two above-assessed single-band models, each specialized for either the C or the L band and potentially relying on different learning approaches with that of the single multiband models trained jointly on C+L data. Results show that the four evaluated ensemble models never outperform the XGB-NN multiband one.

Takeaway: Models jointly trained on the C+L band outperform ensemble approaches combining two single-band models.

B. Operational Implications

From an operator's perspective, the direction of prediction errors is as important as their magnitude. Underestimation of SNR leads to conservative decisions, potentially resulting in over-provisioned margins or in unnecessarily unserved traffic. In contrast, overestimation of SNR is risky, as it may result in selecting transmission parameters that cannot be supported by the actual channel conditions, potentially causing service

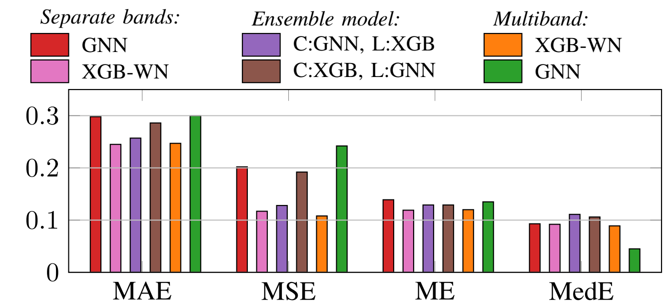

Fig. 3. Comparison of training strategies for Epoch D. **Separate bands:** Models trained independently on C-band and L-band, predictions combined. **Ensemble:** Different model types for each band. **Multiband:** Single model jointly trained on C+L bands.

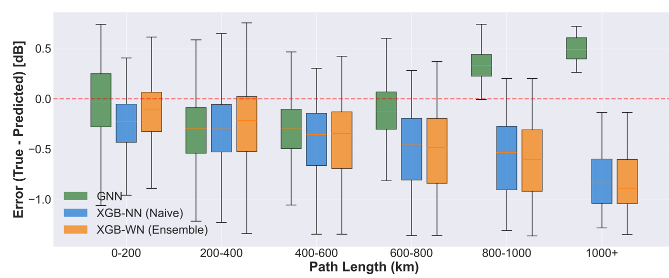

Fig. 4. Prediction error by path length ranges for Epoch B samples.

degradation. To investigate operational implications, in Fig. 4 we report the prediction error achieved by the three models (computed as $y_i - \hat{y}_i$) by path length ranges for epoch B samples. Positive values indicate SNR underestimations, whereas negative values indicate overestimations.

While for short lightpaths ($\leq$ 400 km) errors are roughly uniformly distributed across 0, a critical pattern starts emerging for lightpaths above 400km and becomes critical for long-haul lightpaths (800+ km): in the GNN case, errors progressively shift *positively* (above zero), whereas for the two XGBoost models errors shift *negatively* (below zero), indicating systematic overestimation. These opposite biases have significant consequences, since long-haul lightpaths usually operate with tighter margins and are more expensive to re-provision if they fail: the XGBoost models' tendency to overestimate SNR for such lightpaths could lead to service disruptions, whereas the GNN's conservative behavior provides an implicit "safety buffer" precisely where it matters most.

C. Practical Considerations

Beyond accuracy metrics, when compared to XGBoost-based models, GNN models offer two practical advantages

TABLE III
BEST-PERFORMING MODEL PER EPOCH AND SPECTRAL BAND FROM AN OPERATOR'S PERSPECTIVE

Epoch	Band	Preferred Model	Motivation
B	C	GNN	Significantly lower MAE ($\sim$21%) and reduced bias; full-spectrum awareness is particularly beneficial in sparse, early deployments.
C	C	XGB-NN (MAE/MSE)/ GNN (ME/MedE)	XGB-NN achieves the lowest MAE and MSE, while GNN provides lower mean and median errors, indicating reduced systematic bias.
D	C	XGB-WN	The large amount of available C-band data favors localized spectral information over full-network context.
D	L	XGB-WN (MAE/MSE/ME) / GNN (MedE)	XGB-WN minimizes average error metrics despite limited training data, while GNN yields lower median error, indicating improved robustness.
D	C+L	XGB-WN (MAE/MSE/ME) / GNN (MedE)	The XGB-WN multiband model minimizes aggregate error, whereas the GNN multiband model reduces large over- and under-estimations.

for network operators. First, they learn directly from network topology and spectrum occupation, eliminating the need for handcrafted features describing spectral neighbors. Second, by processing the complete network state, they can capture long-range spectral dependencies that pre-engineered features cannot easily represent. However, such advantages come at the price of a significantly higher computational complexity: GNN training required 625s versus 20s for both variants of XGBoost (31× faster) in the HPC cluster of Politecnico di Torino (NVIDIA GPU 40 with 26GB RAM). Based on the numerical assessment discussed above, Table III summarizes the key findings from an operator perspective, identifying the most suitable model for each epoch and spectral band. The results indicate that topology-aware learning delivers operational benefits that extend beyond conventional aggregate accuracy metrics: for operators prioritizing prediction reliability particularly during early deployment phases and for long-reach lightpaths, the GNN approach provides clear advantages over XGBoost models, despite sometimes exhibiting higher MAE and MSE.

VI. CONCLUSIONS

This work compared two machine learning approaches for SNR estimation of candidate lightpaths in multiband optical networks: a GNN exploiting information on the network topology and the detailed spectral occupation of each link, and two variants of an XGBoost regressor trained on a set of pre-defined lightpath features summarizing the spectral environment along the lightpath traversed links.

Simulation results obtained in an incremental traffic scenario show that while XGBoost achieves strong performance in terms of conventional accuracy metrics, it also exhibits the emergence of systematic prediction bias, particularly for long-haul lightpaths. In contrast, the GNN model, by leveraging full knowledge of the overall spectrum occupation, maintains more balanced and better-calibrated predictions, as reflected by lower median prediction error.

REFERENCES

[1] N. Deng, L. Zong, H. Jiang, Y. Duan, and K. Zhang, "Challenges and enabling technologies for multi-band WDM optical networks," *J. Lightw. Technol.*, vol. 40, no. 11, pp. 3385–3394, 2022.

[2] F. Musumeci, C. Rottondi, A. Nag, I. Macaluso, D. Zibar, M. Ruffini, and M. Tornatore, "An overview on application of machine learning techniques in optical networks," *IEEE Commun. Surv. Tut.*, vol. 21, no. 2, pp. 1383–1408, 2018.

[3] R. Ayassi, A. Triki, N. Crespi, R. Minerva, and M. Laye, "Survey on the use of machine learning for quality of transmission estimation in optical transport networks," *J. Lightw. Technol.*, vol. 40, no. 17, pp. 5803–5815, 2022.

[4] O. Ayoub, S. Troia, D. Andreoletti, A. Bianco, M. Tornatore, S. Giordano, and C. Rottondi, "Towards explainable artificial intelligence in optical networks: the use case of lightpath QoT estimation," *J. Opt. Commun. Netw.*, vol. 15, no. 1, pp. A26–A38, 2022.

[5] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, 2020.

[6] B. Shariati, A. Mastropaolo, N.-P. Diamantopoulos, J. M. Rivas-Moscoso, D. Klonidis, and I. Tomkos, "Physical-layer-aware performance evaluation of SDM networks based on SMF bundles, MCFs, and FMFs," *J. Opt. Commun. Netw.*, vol. 10, no. 9, pp. 712–722, 2018.

[7] T. Panayiotou, G. Savva, B. Shariati, I. Tomkos, and G. Ellinas, "Machine learning for QoT estimation of unseen optical network states," *Optical Fiber Communications Conf. and Exhibition*, 2019.

[8] R. Proietti, X. Chen, K. Zhang, G. Liu, M. Shamsabardeh, A. Castro, L. Velasco, Z. Zhu, and S. J. B. Yoo, "Experimental demonstration of machine-learning-aided QoT estimation in multi-domain elastic optical networks with alien wavelengths," *J. Opt. Commun. Netw.*, vol. 11, no. 1, pp. A1–A10, 2019.

[9] X. Chen, R. Proietti, H. Lu, A. Castro, and S. J. B. Yoo, "Knowledge-based autonomous service provisioning in multi-domain elastic optical networks," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 152–158, 2018.

[10] A. Prakash and S. Kar, "Supervised graph convolution networks for OSNR and power estimation in optical mesh networks," *J. Opt. Commun. Netw.*, vol. 14, no. 6, pp. 469–480, 2022.

[11] G. Savva, T. Panayiotou, I. Tomkos, and G. Ellinas, "Deep graph learning for QoT estimation of unseen optical sub-network states," *J. Lightw. Technol.*, vol. 40, no. 4, pp. 921–934, 2022.

[12] X. Chen, Z. Ouyang, H. Gao, Q. Lin, and Z. Zhu, "Meta-learning-aided QoT estimator provisioning for a dynamic VNT configuration in optical networks," *J. Opt. Commun. Netw.*, vol. 17, no. 1, pp. A103–A111, 2025.

[13] M. Ibrahimi, H. Abdollahi, C. Rottondi, A. Giusti, A. Ferrari, V. Curri, and M. Tornatore, "Machine learning regression for QoT estimation of unestablished lightpaths," *J. Opt. Commun. Netw.*, vol. 13, no. 4, pp. B92–B101, 2021.

[14] G. Sticca, M. Ibrahimi, F. Musumeci, N. Di Cicco, A. Castoldi, R. Pastorelli, and M. Tornatore, "Selective hybrid EDFA/Raman amplifier placement to mitigate lightpath degradation in (C+L) networks," *J. Opt. Commun. Netw.*, vol. 15, no. 8, pp. C232–C241, 2023.

[15] G. S. Sticca, M. Ibrahimi, N. Di Cicco, F. Musumeci, and M. Tornatore, "Incremental planning with dual-fiber distributed Raman amplification in (C+L+S) networks," *J. Opt. Commun. Netw.*, vol. 17, no. 9, pp. D156–D166, 2025.

[16] D. Semrau, R. I. Killey, and P. Bayvel, "A Closed-Form Approximation of the Gaussian Noise Model in the Presence of Inter-Channel Stimulated Raman Scattering," *J. Lightw. Technol.*, vol. 37, no. 9, pp. 1924–1936, 2019.

[17] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proc. 25th ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining*, 2019, pp. 2623–2631.

[18] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," in *Int. Conf. Learning Representations (ICLR)*, 2017.