Uncertainty-Aware QoT Forecasting in Optical Networks with Bayesian Recurrent Neural Networks

Nicola Di Cicco*, Jacopo Talpini†, Mëmëdhe Ibrahimi*, Marco Savi†, Massimo Tornatore*,
*Department of Electronics, Information and Bioengineering (DEIB), Politecnico di Milano, Italy
†Department of Informatics, Systems and Communication (DISCo), University of Milano-Bicocca, Italy

***Abstract*—We consider the problem of forecasting the Quality-of-Transmission (QoT) of deployed lightpaths in a Wavelength Division Multiplexing (WDM) optical network. QoT forecasting plays a determinant role in network management and planning, as it allows network operators to proactively plan maintenance or detect anomalies in a lightpath. To this end, we leverage Bayesian Recurrent Neural Networks for learning uncertainty-aware probabilistic QoT forecasts, i.e., for modelling a probability distribution of the QoT over a time horizon. We evaluate our proposed approach on the open-source Microsoft Wide Area Network (WAN) optical backbone dataset. Our illustrative numerical results show that our approach not only outperforms state-of-the-art models from literature, but also predicts intervals providing near-optimal empirical coverage. As such, we demonstrate that uncertainty-aware probabilistic modelling enables the application of QoT forecasting in risk-sensitive application scenarios.**

***Index Terms*—Quality-of-Transmission, Machine Learning, Uncertainty, Regression, Forecasting**

I. Introduction

In the era of *5G-and-beyond* communications, optical networks are expected to support applications requiring unprecedented capacity, low latency, and high reliability. As such, innovative network design and optimization techniques are required to cope with such tight requirements.

Traditionally, optical network design was performed by employing canonical optimization techniques, e.g., by leveraging closed-form formulas for physical layer modeling or heuristic approaches for resource allocation. More recently, Machine Learning (ML)-based approaches have been demonstrated to be accurate, reliable, and fast in performing the same tasks while meeting network designers' needs [1]. In the last few years, several ML-based optimization approaches have been proposed in the area of optical networking, e.g., to estimate QoT of unestablished lightpaths [2], [3], for failure management [4], for traffic prediction [5] and for QoT forecast of already established lightpaths [6]. More specifically, while a handful of works have employed ML-based probabilistic approaches for *QoT estimation of unestablished lightpaths*, e.g., [2], [3], [7], no previous work on *QoT forecasting* employs an uncertainty-aware probabilistic approach, but rather only provide point estimates, which do not provide any information on the uncertainty of the prediction itself.

In this paper, we investigate the problem of forecasting QoT of deployed lightpaths in a Wavelength Division Multiplexed (WDM) network. While predicting the QoT of unestablished lightpaths has been significantly investigated in recent years [8], the problem of forecasting the QoT of already established lightpaths over a time horizon is a less investigated topic. In particular, in contrast to previous works which provide only point-estimate forecasts, we leverage Bayesian Recurrent Neural Networks to forecast a probability distribution of established lightpaths' QoT over a time horizon.

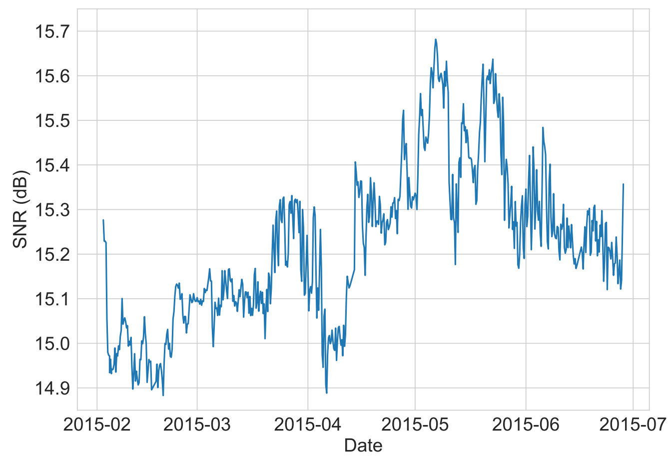


Fig. 1. SNR values over time of a single illustrative lightpath from the publicly-available Microsoft WDM optical backbone dataset. SNR values are averaged over windows of six hours.

The challenge of forecasting lightpaths' QoT comes from the fact that the optical transmission medium is very sensitive to environmental conditions, e.g., temperature changes, wind, mechanical movements of the fiber, and vibrations, to the point that it has been proven to be a reliable sensing instrument [9]. In this context, QoT forecasting allows network operators to estimate fast time-varying effects on a QoT metric (e.g., on Q-factor, Bit-Error Rate (BER) or Signal-to-Noise Ratio (SNR)). Since QoT may vary significantly over time due to fast time-varying effects, performing a probabilistic estimation of the channel's behavior allows network operators to assess more accurately the uncertainty associated with the estimation, thus allowing to perform more informed risk evaluations and take the corresponding measures accordingly. For example, a network operator may leverage probabilistic forecasts to estimate the probability of the QoT remaining above a desired performance level during its lifetime, thus setting safe but tight expectations when signing Service Level Agreements. Another application of interest is anomaly detection: when

N. Di Cicco and J. Talpini are co-first authors. Our source code is publicly available at https://github.com/bonsai-lab-polimi/icc2023-qot.

one or multiple QoT measurements fall outside the predicted intervals of high-probability, a network operator can infer that the lightpath has deviated from its "normal" behavior, and thus plan preventive maintenance prior to the occurrence of severe faults. While some works have utilized probabilistic QoT estimation of unestablished lightpaths, e.g., [2], [3], [7], to the best of our knowledge, this is the first work that addresses lightpath QoT forecasting through an uncertainty-aware probabilistic estimation lens.

As a motivating example, Fig. 1 illustrates the SNR values of a lightpath from a real-world publicly-available dataset [10]. We can observe that the lightpath's SNR values significantly vary over time, with a dynamic range of approximately 1dB. Most importantly, we can deduce that, in general, the temporal dynamics of the SNR cannot only be imputed to random zero-mean noise. While a physical justification on *why* these oscillations happen over time is at present an open research question, our goal is to leverage Machine Learning for modeling in a probabilistic way *how* these oscillations happen over time.

Our illustrative experimental results on the open-source Microsoft dataset [10] show that our probabilistic forecasts yield more accurate mean QoT values than state-of-the-art point-estimate approaches from literature. Moreover, we show that our estimated confidence intervals provide good empirical coverage of the ground-truth QoT values. We therefore pave the road for the application of QoT forecasting in risk-sensitive applications, such as estimating compliance to Service Level Agreements and anomaly detection.

The remainder of the paper is organized as follows. In Section II, we revise some relevant related works, emphasizing the differences to our proposed approach. In Section III, we describe the proposed *uncertainty-aware QoT forecasting* solution. In Section IV, we describe the dataset utilized. In Section V we provide and discuss the numerical results, and conclude the paper by highlighting the main takeaways and lessons learned in Section VI.

II. Related Works

The application of Machine Learning to optical networks has been receiving considerable attention as it enables automated network (re)configuration and fast decision-making by leveraging historical data.

We make use of a publicly-available Microsoft WDM backbone dataset [10]. Several works analyzed this dataset, e.g., [10]–[12], for investigating optical layer failures and capacity gains through elastic modulation.

In [6] authors provide a tutorial on lightpath QoT forecasting with the objective of identifying the most suitable ML techniques to address the challenge of QoT forecasting. Authors make use of historical data from three separate datasets and compare the performance of four variants of neural networks: Long Short-Term Memory (LSTM), Encoder-Decoder LSTM, Gated Recurrent Unit (GRU), and Multilayer Perceptron (MLP). Numerical results show that QoT forecasters based on linear regression and MLP can outperform more complex RNN models. Additionally, aggregate performance metrics illustrate that outliers in the training data can significantly impact the final performance.

Several other works have investigated the application of ML-based approaches to perform QoT forecasting [13]–[16]. In [14] authors propose an encoder-decoder LSTM to forecast the SNR of an established lightpath. The authors show how an operating lightpath might experience SNR variations up to 3 dB during its operation, possibly leading to outages. Results compare GRU to LSTM in terms of Root Mean Square Error (RMSE), and show that the performance of the employed model depends on the forecast horizon. In [15] authors employ a Convolutional Neural Network for QoT forecasting over forecast horizons up to 24 hours. The authors show that it is possible to capture and correctly predict the temporal lightpath SNR changes 24 hours before. Similarly, in [13] authors propose MLP and LSTM deep neural network models to forecast the minimum QoT of deployed lightpaths over time horizons up to 72 hours. Finally, in [16] authors propose two multivariate neural networks based on LSTM and GRU showing that they outperform their counterpart univariate models in terms of Absolute Maximum Error (AME).

While previous literature considers point-estimate forecasts, our proposed forecaster models the probability distribution of lightpaths' QoT. By characterizing lightpaths' QoT as a probability distribution, a network operator is able to safely estimate the fast time-varying behavior of established lightpaths over arbitrary time horizons. Moreover, previous literature focuses on forecasting QoT values for single lightpaths only. In this work, we propose a multi-input multi-output model that jointly forecasts the QoT of multiple lightpaths over time. As such, our goal is to learn not only the temporal behavior of single lightpaths, but also the temporal correlations between channels and optical links in the WDM network.

III. Uncertainty-Aware QoT Forecasting

Our problem statement is as follows: given C input QoT time-series $\mathbf{x}_{1:T}$ of T time-steps, each one from a different lightpath, our goal is to forecast QoT values $\hat{\mathbf{y}}_{T+1:T+H}$ up to H time-steps in the future for each lightpath.

While ML-based QoT forecasters have been proposed in literature, they consider only point predictions, which do not provide information on their predictive uncertainty (i.e., in the form of confidence intervals) as pointed out in Section II. In the context of QoT forecasting, truthful confidence intervals allow an operator to (probabilistically) infer credible ranges for the QoT values. In addition, in the context of anomaly detection, truthful confidence intervals allow for early identification of potentially anomalous trends, with fine-grained control on the ratio between true anomalies and false alarms. Overall, we argue that probabilistic outputs are non-negotiable for real-world, risk-sensitive applications of ML in optical networks.

Moreover, while previous approaches focus on forecasting QoT for a single lightpath, we consider a model that takes as an input multiple QoT time-series, and jointly forecasts QoT values for all of the considered lightpaths. As such, the model can learn and exploit correlations between different lightpaths.

Therefore, our objective is to develop a probabilistic QoT multi-input multi-output forecaster providing truthful probabilistic predictions. To this end, we leverage Bayesian Recurrent Neural Networks (RNNs) for implementing our model. Bayesian Neural Network models allow for estimating the model's predictive uncertainty, while RNN models allow for efficient learning on QoT time-series data.

A. Background on Bayesian Neural Networks

Traditional Neural Networks (NNs) are not able to capture predictive uncertainty [17], a crucial aspect for many risk-sensitive applications. However, it is possible to tackle this issue in a principled way by coupling NNs with Bayesian probability theory, leading to the formulation of Bayesian Neural Networks (BNNs). The most distinguishing property of a BNN is marginalization, i.e., rather than using a single set of weights $\hat{\mathbf{w}}$ determined at the end of the training phase, BNNs rely on the computation of the predictive distribution for a given input $\mathbf{x}$, as follows [18]:

$$p(y|\mathbf{x}, \mathcal{D}) = \int p(y|\mathbf{x}, \mathbf{w}) p(\mathbf{w}|\mathcal{D}) d\mathbf{w} \tag{1}$$

where: $p(\mathbf{w}|\mathcal{D})$ is the posterior distribution of the weights of the considered model (i.e. a NN), given a training dataset $\mathcal{D} = \{(\mathbf{x}_i; y_i)\}_{i=1}^N$ of input-output pairs. The posterior distribution over the weights in the previous equation allows for capturing the model uncertainty, arising from the uncertainty associated with the parameters of the model, given the training dataset.

Unfortunately, the exact evaluation of the predictive distribution is computationally intractable for neural networks of practical size. To get around this problem, several approximate inference methods were proposed. Most approaches rely on Variational Inference for finding a tractable approximation to the Bayesian posterior distribution of the weights [19] or on Deep Ensembles, where the same NN architecture is trained multiple times and then the resulting models are averaged [18].

A different approach for approximating the Bayesian predictive distribution, named Monte-Carlo (MC) Dropout, was first proposed in [17] for NNs with Dropout layers.

We decided to employ MC-Dropout since it is efficient and easy to implement, as it does not require any change to an existing NN architecture. MC-Dropout is based on keeping Dropout layers activated also during the inference phase, randomly deactivating (dropping out) hidden units by sampling from a Bernoulli distribution N times, thus generating an ensemble of N different models. To summarize the ensemble predictions from MC-Dropout for a given input $\mathbf{x}$, it is possible to rely on the mean of the predictive distribution, obtained by averaging multiple stochastic forward passes [17]:

$$\hat{y} \equiv \mathbb{E}_{p(y|\mathbf{x},\mathcal{D})}[y] \approx \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{x}; \mathbf{w}_i) \tag{2}$$

where $f(\mathbf{x}; \mathbf{w}_i)$ denotes the i-th forward pass for a given NN.

Moreover, other than providing a single-point estimate, BNNs can also quantify the uncertainty associated with each prediction. For regression problems, we estimate the predictive

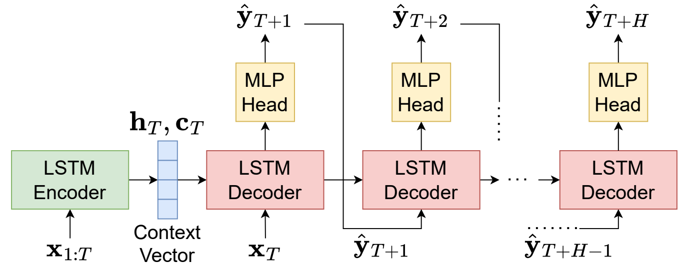


Fig. 2. Proposed neural network architecture. The LSTM encoder produces a fixed-size context vector from the input QoT time-series $\mathbf{x}_{1:T}$. The LSTM decoder recursively produces the forecasted QoT time-series $\hat{\mathbf{y}}_{T+1:T+H}$.

uncertainty as the variance of the predictive distribution. In particular, the variance of the predictive distribution can be decomposed via the law of total variance [17], [20], as follows:

$$\underbrace{\mathrm{Var}_{p(y|\mathbf{x},\mathcal{D})}[y]}_{\text{Total Uncertainty}} = \underbrace{\mathrm{Var}_{p(\boldsymbol{w}|\mathcal{D})}\left[\mathbb{E}_{p(y|\mathbf{x},\mathbf{w})}[y]\right]}_{\text{Model Uncertainty}} + \underbrace{\sigma^2}_{\text{Inherent Noise}} \tag{3}$$

This equation shows that the total variance (i.e., the prediction uncertainty) can be decomposed into two terms: the first one captures the model (or epistemic) uncertainty, while the second one the inherent noise on the target values.

In particular, for estimating the model uncertainty via MC-Dropout, one can compute the sample variance of the considered N stochastic forward passes [17], as follows:

$$\mathrm{Var}_{p(\boldsymbol{w}|\mathcal{D})}\left[\mathbb{E}_{p(y|\mathbf{x},\mathbf{w})}[y]\right] \approx \frac{1}{N} \sum_{i=1}^{N} (\hat{y} - f(\mathbf{x}; \mathbf{w}_i))^2 \tag{4}$$

For estimating the inherent noise level, we adopt the approach proposed in [20]. Specifically, for a given channel, denoting the mean i-th prediction as $\hat{y}$ and the corresponding ground-truth target as y_i, the inherent noise is estimated through the variance of the residuals as follows:

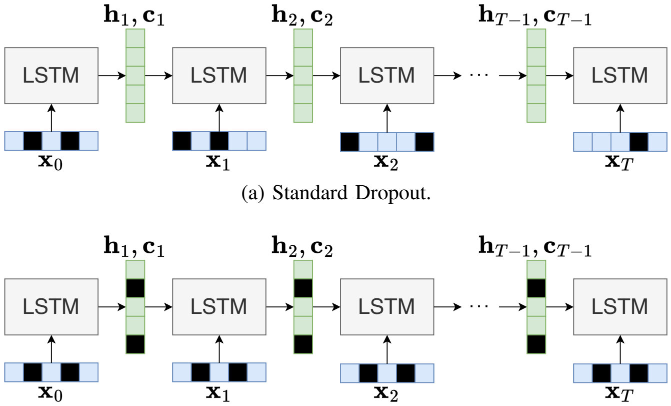
$$\sigma^2 \approx \frac{1}{V} \sum_{i=1}^{V} (y_i - \hat{y}_i)^2 \tag{5}$$

where V is the number of points used to train the model.

B. Proposed model architecture

Our neural network architecture is illustrated in Fig. 2. We implemented a classical encoder-decoder architecture with multiple LSTM layers. Briefly, an LSTM layer extends a standard RNN layer, such that each ordinary recurrent node is instead replaced by a memory cell. Each memory cell contains an internal state (i.e., a node with a self-connected recurrent edge of fixed weight) which ensures that gradients can flow across many time-steps without vanishing [21]. At each time-step t, an LSTM layer outputs a hidden state $\mathbf{h}_t$, encoding the working memory of the RNN, and a cell state $\mathbf{c}_t$, encoding the long-term memory of the LSTM.

Our encoder-decoder architecture, known in literature as Seq2Seq [22], is well suited for extracting meaningful representations from multiple input time-series. The encoder and the decoder networks consist both of multiple LSTM layers,



(a) Standard Dropout.

(b) Variational Dropout.

Fig. 3. Standard Dropout and Variational Dropout in an LSTM layer. Masked inputs are depicted in black. In Variational Dropout, recurrent inputs (i.e., hidden state and cell state) are also masked, and the same mask is kept for the whole duration of the input sequence.

with the decoder having an additional Multi-Layer Perceptron (MLP) head that outputs the final forecast. The encoder network maps the input SNR time-series $\mathbf{x}_{1:T} \in \mathbb{R}^{C\times T}$ to a context vector of fixed size (i.e., the last hidden state and cell state of the encoder LSTM network), while the decoder network recursively predicts the future SNR time-series $\hat{\mathbf{y}}_{T+1:T+H} \in \mathbb{R}^{C\times H}$ from the context vector.

Additionally, to make our Seq2Seq architecture output probabilistic predictions, we exploited a particular implementation of the general MC-Dropout approach, described in the previous subsection, called Variational Dropout [23] and specifically designed for RNNs. Variational Dropout randomly masks (drops) some network units with their inputs, output, and recurrent connections at each time-step [23]. As shown in Fig. 3, in contrast to standard Dropout, Variational Dropout applies a mask also on the recurrent connections, and it keeps the same mask for each time-step of the input sequence. As shown in [23], a recurrent network with Variational Dropout is equivalent to an approximate Bayesian model. As for MC-Dropout, mean predictions and predictive uncertainty are estimated by aggregating multiple predictions over different Dropout masks.

IV. Dataset Description

We consider an open-source dataset of traces from a Microsoft optical WDM backbone [10]. The dataset aggregates measurements from 4000 frequency channels over 115 optical paths. The exact physical layer topology is not disclosed. Each frequency channel includes four main features, namely i) signal Q-factor, ii) transmit power, iii) chromatic dispersion, and iv) polarization mode dispersion. Measurements were collected at 15-minute sampling intervals over a 14-month observation period. As in [13], we considered only the time-series regarding the Q-factor, as this quantity is highly correlated with the SNR. Specifically, since all lightpaths in the dataset employ PDM-QPSK, the reported Q-values are approximately equivalent to the corresponding SNR values [24].

A. Data preprocessing

An exploratory analysis of the time-series revealed the presence of several outliers. We identified outliers by looking at samples falling beyond $\sim 4\sigma$-equivalent from the median value and we replaced them with linear interpolation, as proposed in [13]. Moreover, we smoothed each time-series by taking the mean over non-overlapping windows of 6 hours, to reduce the noise without compromising the informational content. The underlying assumption is that, during the window time scale, we considered the physical properties of the network as being nearly constant and that the observed variability is due to only statistical fluctuations. Then, the dataset is partitioned into 60% training, 20%, validation, and 20% testing set, and standardized using the training data, so that each feature distribution presents a zero mean and unit variance. For training, each time-series has been divided into batches of data having an observation window of 25 days and a target series of 12.5 days, corresponding to T=100 and H=50 input and output time-steps, respectively.

V. Illustrative Numerical Results

In this Section, we first illustrate the overall performance of our QoT forecaster by comparing aggregate performance metrics with a state-of-the-art baseline from the literature. We then discuss the uncertainty estimation capabilities of our forecaster by analyzing the quality (i.e., the empirical ground-truth coverage) of the estimated confidence bands.

To analyze the capabilities of our approach to exploit correlations between different channels and optical links, we explored two different use cases.

Scenario A: we considered all the channels present in a single optical link, for a total of C=50 channels. A trained model can therefore be leveraged for monitoring all lightpaths traversing the same location in the backbone network.

Scenario B: we randomly selected C=50 channels, each one from a distinct optical link. A trained model can therefore be leveraged for monitoring selected (potentially risk-sensitive) lightpaths in different locations in the backbone network.

For both experimental scenarios, we found the best hyperparameters leveraging on the Tree Parzen Estimator (TPE) Bayesian optimization algorithm implemented in the Optuna library [25]. Specifically, our encoder-decoder architecture consists of two LSTM layers with an embedding layer of size 18. We trained our model for 400 epochs with minibatch size 32 using the Huber loss [26]. The dropout rate was set to 0.2, as suggested in [23]. This architecture has been found to maximize the validation performance, with training times of less than ten minutes on a Macbook Pro M1 CPU. As a baseline, we consider the current state-of-the-art MLP-based forecasting approach proposed in [13]. Our models were implemented in PyTorch [27], and our source code is publicly available at https://github.com/bonsai-lab-polimi/icc2023-qot

A. Aggregate performance metrics

We evaluate the performance of our model by computing the average Mean Squared Error (MSE) and Mean Average

TABLE I
TEST SET PERFORMANCE METRICS FOR SCENARIOS A AND B.

Scenario A				
Horizon	**Our approach**		MLP [13]	
	MSE	MAPE%	MSE	MAPE%
150h	**0.0032**	**0.33**	0.0041	0.37
300h	**0.0029**	**0.32**	0.0036	0.37
600h	**0.0024**	**0.31**	0.0032	0.36

Scenario B				
Horizon	**Our approach**		MLP [13]	
	MSE	MAPE%	MSE	MAPE%
150h	**0.0050**	**0.45**	0.065	1.02
300h	**0.0053**	**0.46**	0.110	1.25
600h	**0.0054**	**0.47**	0.107	1.36

Percentage Error (MAPE) in the test set over different forecast horizons H, namely 150, 300, and 600 hours, corresponding to 25, 50, and 100 forecasting time-steps, respectively.

We report test set aggregate performance in Table I. For both considered scenarios, the averages from our probabilistic prediction yield values much closer to the ground-truth compared to the point-estimate forecasts that can be obtained using the MLP proposed in [13]. Empirically, we observed that this is because the MLP model, while providing reasonable forecasts when the SNR is almost stationary, completely fails to model more complex temporal patterns. Moreover, the MLP model takes as an input the time-series pertaining to a single lightpath, hence it cannot learn to exploit the temporal correlations between multiple lightpaths. Conversely, our proposed multi-input multi-output LSTM model is able to disclose complex temporal correlations in the input time-series, thus yielding more robust point-estimate forecasts. Finally, we underline that with our Seq2Seq approach we train *a single model* which we test for multiple forecast horizons, whereas MLPs require training *different models* for each forecast horizon, since their output dimensionality is fixed. Therefore, our approach generalizes well for forecast horizons different than training.

B. Quality of the predicted confidence intervals

Our probabilistic forecaster outputs not only a mean-value prediction, but also confidence intervals in which the ground-truth Q-value is presumed to lie. For estimating a confidence interval, we modeled each output time-step in our predicted forecast as a Normal distribution, with mean equal to the forecasted mean value, and variance equal to the total predictive uncertainty, computed as per Eq. (3). The choice of the Normal distribution makes the fewest assumptions and leads to the most conservative estimates, given only the knowledge of the mean and the variance [28] of the predictive distribution. We therefore compute the predicted centered 68.26%, 90%, 95% and 99% confidence interval, and we count the percentage of ground-truth points that fall inside. For a perfect estimation, we expect to observe on average, say, $x\%$ of ground-truth points falling inside the predicted $x\%$ confidence intervals.

In Table II we report the average empirical coverage of our confidence interval estimates averaged from both Scenario A and Scenario B. We observe that the predicted confidence intervals provide a reasonable empirical coverage, demonstrating that the proposed approach can indeed be used for predicting truthful confidence intervals.

TABLE II
PREDICTED CONFIDENCE INTERVALS FOR SCENARIOS A AND B

Interval Range	Ground-truth Coverage		
	Expected	Empirical (A)	Empirical (B)
$\mu \pm \sigma$	68.26%	71.34%	65.54%
$\mu \pm 1.62\sigma$	90%	90.41%	86.13%
$\mu \pm 1.96\sigma$	95%	94.04%	92.21%
$\mu \pm 2.58\sigma$	99%	98.39%	98,86%

C. Illustrative applications

We conclude our experimental evaluation with potential use cases for our forecast methodology. Namely, we consider i) performance monitoring and ii) anomaly detection for in-service lightpaths. Note that, apart from SNR drops of several dBs [10], which are "true" anomalies (and can be clearly isolated without the use of ML), the definition of anomaly is ultimately application-specific. For the sake of our study, we consider anomalous behaviors that qualitatively deviate from a consolidated temporal trend.

Fig. 4a displays a forecast of a stable lightpath. While the past SNR values exhibited multiple temporal trends, our model was remarkably capable of outputting tight confidence intervals around the ground-truth and capturing the inception of a decreasing trend several time-steps in advance.

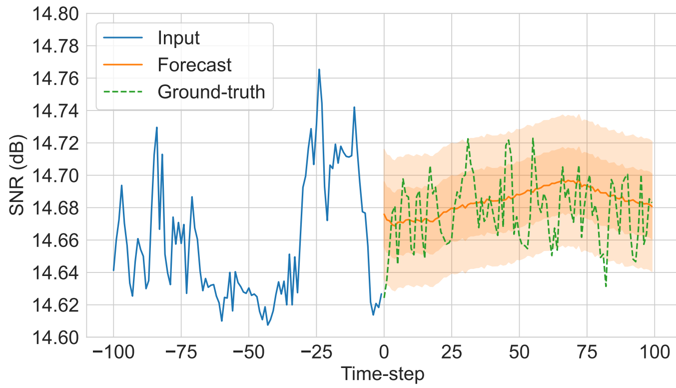
Fig. 4b displays a forecast on a lightpath that will likely be experiencing an anomalous event (namely, a change of stationary state) in the future. We can observe that, as expected, the anomalous ground-truth falls well outside the predicted confidence intervals. In this context, a network operator may define a threshold distance from the interval ranges, such that observations exceeding that distance raise an alarm. After that, the occurrence of n consecutive alarms may trigger a warning, with a warning severity growing together with n.

VI. CONCLUSION

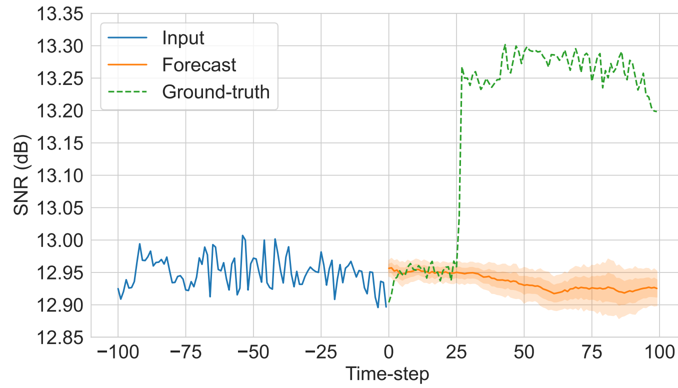
In this work, we proposed a novel uncertainty-aware approach for forecasting future QoT values in established lightpaths based on Bayesian RNNs. In particular, we illustrated a multi-input multi-output Seq2Seq architecture for exploiting the correlations between different frequency channels and fiber links in the physical network. Numerical results on a real-world dataset showed that not only our proposed approach outperforms previous literature, but also predicts confidence intervals providing reliable empirical coverage. Finally, we illustrated how our probabilistic forecasts could be applied for performance monitoring and anomaly detection. Future work will investigate attention-based models for interpreting correlations between different input QoT time-series.

VII. ACKNOWLEDGEMENT

This work was partially supported by the European Union under the Italian National Recovery and Resilience Plan



(a) Stable lightpath.



(b) Anomalous lightpath.

Fig. 4. Illustrative forecasts on a stable and an anomalous lightpath. The predicted mean forecast is represented by the solid orange line. Dark orange and light orange bands represent the predicted standard deviations and 95% confidence intervals, respectively.

(NRRP) of NextGenerationEU, partnership on "Telecommunications of the Future" (PE00000001 - program "RESTART").

References

[1] F. Musumeci *et al.*, "An overview on application of machine learning techniques in optical networks," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1383–1408, 2019.

[2] N. Di Cicco, M. Ibrahimi, C. Rottondi, and M. Tornatore, "Calibrated probabilistic qot regression for unestablished lightpaths in optical networks," in *2022 International Balkan Conference on Communications and Networking (BalkanCom)*, 2022, pp. 21–25.

[3] M. Ibrahimi *et al.*, "Machine learning regression for qot estimation of unestablished lightpaths," *Journal of Optical Communications and Networking*, vol. 13, no. 4, pp. B92–B101, 2021.

[4] F. Musumeci *et al.*, "Domain adaptation and transfer learning for failure detection and failure-cause identification in optical networks across different lightpaths [invited]," *Journal of Optical Communications and Networking*, vol. 14, no. 2, pp. A91–A100, 2022.

[5] D. Andreoletti, S. Troia, F. Musumeci, S. Giordano, G. Maier, and M. Tornatore, "Network traffic prediction based on diffusion convolutional recurrent neural networks," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2019, pp. 246–251.

[6] S. Allogba, S. Aladin, and C. Tremblay, "Machine-learning-based lightpath qot estimation and forecasting," *Journal of Lightwave Technology*, vol. 40, no. 10, pp. 3115–3127, 2022.

[7] M. Ibrahimi, H. Abdollahi, A. Giusti, C. Rottondi, and M. Tornatore, "Machine learning regression vs. classification for qot estimation of unestablished lightpaths," in *Photonic Networks and Devices*. Optica Publishing Group, 2020, pp. NeM3B–1.

[8] Y. Pointurier, "Machine learning techniques for quality of transmission estimation in optical networks," *J. Opt. Commun. Netw.*, vol. 13, no. 4, pp. B60–B71, Apr 2021. [Online]. Available: https://opg.optica.org/jocn/abstract.cfm?URI=jocn-13-4-B60

[9] G. A. Wellbrock, T. J. Xia, E. Ip, Y.-K. Huang, M.-F. Huang, T. Wang, and Y. Aono, "Field trial of vibration detection and localization using coherent telecom transponders over 380-km link," in *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, 2021, pp. 1–3.

[10] M. Ghobadi and R. Mahajan, "Optical layer failures in a large backbone," in *Proceedings of the 2016 Internet Measurement Conference*. Association for Computing Machinery, 2016, p. 461–467.

[11] M. Filer *et al.*, "Elastic optical networking in the microsoft cloud [invited]," *Journal of Optical Communications and Networking*, vol. 8, no. 7, pp. A45–A54, 2016.

[12] M. Ghobadi, J. Gaudette, R. Mahajan, A. Phanishayee, B. Klinkers, and D. Kilper, "Evaluation of elastic modulation gains in microsoft's optical backbone in north america," in *2016 Optical Fiber Communications Conference and Exhibition (OFC)*, 2016, pp. 1–3.

[13] H. Chouman, P. Djukic, C. Tremblay, and C. Desrosiers, "Forecasting lightpath qot with deep neural networks," in *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, 2021, pp. 1–3.

[14] S. Aladin, S. Allogba, A. V. S. Tran, and C. Tremblay, "Recurrent neural networks for short-term forecast of lightpath performance," in *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, 2020, pp. 1–3.

[15] A. Mezni, D. W. Charlton, C. Tremblay, and C. Desrosiers, "Deep learning for multi-step performance prediction in operational optical networks," in *2020 Conference on Lasers and Electro-Optics (CLEO)*, 2020, pp. 1–2.

[16] S. Allogba, S. Aladin, and C. Tremblay, "Multivariate machine learning models for short-term forecast of lightpath performance," *Journal of Lightwave Technology*, vol. 39, no. 22, pp. 7146–7158, 2021.

[17] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*. PMLR, 2016, pp. 1050–1059.

[18] A. G. Wilson and P. Izmailov, "Bayesian deep learning and a probabilistic perspective of generalization," *Advances in neural information processing systems*, vol. 33, pp. 4697–4708, 2020.

[19] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, "Weight uncertainty in neural network," in *International conference on machine learning*. PMLR, 2015, pp. 1613–1622.

[20] L. Zhu and N. Laptev, "Deep and confident prediction for time series at uber," in *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2017, pp. 103–110.

[21] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with lstm," *Neural Computation*, vol. 12, no. 10, pp. 2451–2471, 2000.

[22] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *Advances in neural information processing systems*, vol. 27, 2014.

[23] Y. Gal and Z. Ghahramani, "A theoretically grounded application of dropout in recurrent neural networks," *Advances in neural information processing systems*, vol. 29, 2016.

[24] C. Delezoide, P. Ramantanis, and P. Layec, "Leveraging field data for the joint optimization of capacity and availability in low-margin optical networks," *Journal of Lightwave Technology*, vol. 38, no. 24, pp. 6709–6718, 2020.

[25] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

[26] P. J. Huber, "Robust estimation of a location parameter," *Annals of Mathematical Statistics*, vol. 35, pp. 492–518, 1964.

[27] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 2019.

[28] P. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica® Support*. Cambridge University Press, 2005.