Calibrated Probabilistic QoT Regression for Unestablished Lightpaths in Optical Networks

Nicola Di Cicco
Politecnico di Milano
nicola.dicicco@polimi.it

Mëmëdhe Ibrahimi
Politecnico di Milano
memedhe.ibrahimi@polimi.it

Cristina Rottondi
Politecnico di Torino
cristina.rottondi@polito.it

Massimo Tornatore
Politecnico di Milano
massimo.tornatore@polimi.it

***Abstract*—Quality-of-Transmission (QoT) regression of unestablished lightpaths is a fundamental problem in Machine Learning applied to optical networks. Even though this problem is well-investigated in current literature, many state-of-the-art approaches either predict point-estimates of the QoT or make simplifying assumptions about the QoT distribution. Because of this, during lightpath deployment, an operator might take either overly-aggressive or overly-conservative decisions due to biased predictions. In this paper, we leverage state-of-the-art Gradient Boosting Decision Tree (GBDT) models and recent advances in uncertainty calibration to perform QoT probabilistic regression for unestablished lightpaths. Calibration of a regression model allows for an accurate modeling of the QoT Cumulative Distribution Function (CDF) without any prior assumption on the QoT distribution. In our illustrative experimental results, we show that our calibrated GBDT model's predictions provide accurate confidence interval estimates, even when only few samples per lightpath configuration are available at training time.**

***Index Terms*—Quality-of-Transmission, Machine Learning, Regression, Lightpath, Estimation**

I. Introduction

The introduction of coherent transmission and high-rate transponders in optical networks have enabled a whole new range of configurable transmission parameters to optimize network performance, at the expense of increased network-design complexity. In this context, a network design challenge that has lately received noticeable attention is the estimation of Quality-of-Transmission (QoT) of unestablished lightpaths. The main challenge to accurately estimate lightpaths' QoT comes from the non-linear nature of signal propagation in optical fibers and lack of complete knowledge of equipment parameters. Traditionally, QoT estimation in optical networks has been addressed using exact numerical solvers or margined formulas, with both approaches having non-negligible shortcomings. Exact numerical solvers, e.g., the Split-Step Fourier method [1], are computationally heavy, while margined formulas, e.g., the GN-model [2], are computationally fast but lead to under-utilization of network resources.

An alternative to traditional approaches for QoT estimation is the application of Machine Learning (ML). ML has proven effective for solving network design problems [3], among which QoT estimation [4], [5]. ML-based approaches overcome the shortcomings of both exact approaches and margined formulas by estimating lightpaths' QoT in reasonable time and by modelling the uncertainties not captured by the physical layer models. Seminal works that adopt ML for QoT estimation make use of ML-classification approaches [6]–[8] and aim to answer the following question: given a lightpath configuration, is the Signal-to-Noise Ratio (SNR) above a given threshold? However, a classification-based QoT estimation can only provide limited information. In particular, a classification-based QoT estimation does not allow to predict the SNR distribution and as a result *i)* it does not provide information on how far the predicted SNR is from the threshold and *ii)* it does not tell if it is possible to upgrade the lightpath configuration, e.g., from 100 Gbps DP-QPSK to 200 Gbps DP-16-QAM. An alternative is to use regression-based approaches [9]–[12]. In the context of QoT estimation, regression-based approaches can be used with three different purposes: *i)* estimation of the SNR distribution, i.e., quantifying if and how far the SNR is from the threshold (which is pivotal in presence of uncertainties introduced by fast time-varying penalties or by crosstalk effects due to the dynamic allocation of alien wavelengths), *ii)* probabilistic QoT modelling, i.e., estimating SNR to capture unknown effects such as amplifier gain ripple or connector losses with the objective of reducing design margins (time-varying penalties margins are still assumed to be fixed), and *iii)* estimation of the values of input parameters, i.e., gathering SNR information to refine the knowledge of input parameters (e.g., amplifier noise figure, nonlinear and dispersion coefficient and fiber attenuation coefficient). Most regression-based approaches, e.g., Refs. [9]–[12], estimate a scalar value of SNR, i.e., they do not capture the uncertainty associated to the SNR value due to fast time-varying penalties. Additionally, works that estimate SNR distribution such as [13] assume some prior knowledge about the SNR distribution, estimating the moments of a skewed-Gaussian. Note that, the QoT research problem has been investigated also in other areas of research, e.g., ad-hoc networks [14], however, applied in a different context to the one studied in our work.

The novelty of our work and the research gap we address lies in the fact that i) we do not assume any prior knowledge on the SNR distribution, and, ii) we make use of recently-proposed probabilistic regression models to estimate the uncertainty associated with each prediction. We explore an uncertainty calibration procedure [15] that allows to produce accurate confidence intervals given a sufficient amount of i.i.d.

N. Di Cicco and M. Ibrahimi contributed equally to this work. Source code and datasets are publicly available at https://github.com/bonsai-lab-polimi/balkancom2022.

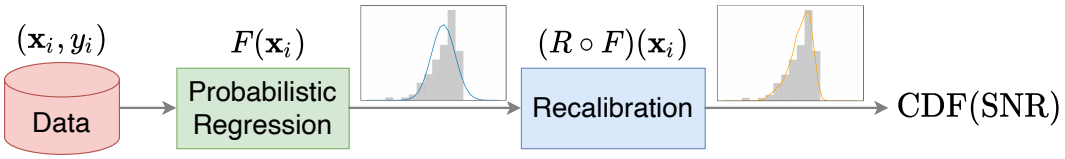

Fig. 1. Machine Learning pipeline: *Probabilistic Regression* followed by *Recalibration* results in accurate predictions of the SNR distribution.

data. With proper calibration, the regression model is able to accurately predict the SNR's CDFs even when a limited amount of samples per lightpath configuration are available (as it is expected to happen in realistic network scenarios). Indeed, one of the main challenges of applying ML approaches in optical networks is data scarcity, since in practice it is not very likely to have sufficient available SNR measurements for every possible lightpath configuration.

II. Related Work

ML-based QoT estimation has been investigated using both classification and regression approaches. Several ML-classification approaches for QoT estimation have been proposed [6]–[8], however, they provide no information about the actual numerical value that the SNR can take. Hence, in the following, we concentrate on those studies that have applied regression approaches to QoT estimation.

In [9], the authors make use of ML-regression to reduce design margins by estimating the penalties due to EDFA gain ripple and filtering uncertainties at ROADM nodes. Similarly, in [12], the authors investigate reducing the amplifier power profile and noise figure uncertainties by leveraging monitored values of lightpaths' SNR. In [10], the authors estimate the uncertainties of various input parameters, e.g., amplifier noise figure, nonlinear and dispersion coefficient, and fiber attenuation coefficient, utilizing SNR of established lightpaths to learn the network state and to train the parameters of the analytical physical layer model, e.g., the GN-model for estimating NLI impairments. In [11], the authors propose a ML method exploiting realistic network monitoring data and leveraging the physics behind light propagation to refine two main parameters highly affecting QoT estimation: span input/output lumped losses, and amplifier gain spectrum.

Compared to these works, we estimate the SNR distribution and determine not only whether a lightpath configuration is above/below threshold, but also enable making an informed decision on lightpath deployment by leveraging the knowledge on how close/far from the threshold is the SNR estimation. For example, if the SNR of a lightpath with a 32-QAM modulation format is predicted to be above threshold with less than 70% confidence, an operator may prefer a conservative approach and employ a 16-QAM modulation format to avoid a potentially faulty lightpath configuration. Similarly, if the same SNR estimation is predicted to be above threshold with 80% confidence, an operator may prefer a liberal estimation approach an decide to deploy the lightpath with a 32-QAM modulation format. Moreover, compared to previous works, the uncertainties in the SNR are due to fast time-varying penalties rather than due to unknown effects, e.g., amplifier gain ripple, or lack of complete knowledge of input parameters, e.g., span input/output lumped losses. Hence, we estimate an SNR distribution rather than a point estimate.

In our previous works [13], [17], we modelled the SNR as a random variable and used ML-regression approaches to estimate the parameters of the distribution. However, such approaches are not distribution-agnostic as they assume some prior knowledge about the SNR distribution. In this work, we utilize recent advances on the calibration of probabilistic regression models that enable us to accurately estimate uncertainties related to SNR, without requiring any prior knowledge on the underlying data distribution. Calibrating a regression model allows to avoid overconfident lightpaths' SNR estimations that would lead to deploying faulty lightpaths and allows to maximize network resource utilization, hence minimizing network cost.

III. Calibrated QoT Probabilistic Regression

Figure 1 shows a high-level view of the Machine Learning pipeline to solve the QoT estimation problem. Given testing data with its selected features $\mathbf{x}_i$, the *Probabilistic Regression* model estimates the SNR distribution and the *Recalibration* allows to accurately predict the SNR distribution (CDF(SNR)). In general, we would like a probabilistic regression model to be both calibrated and sharp. Intuitively, calibration measures how truthful the predicted confidence intervals are, whereas sharpness measures to how tight (i.e., close to the mean prediction) the aforementioned confidence intervals are. Figure 2 shows an illustrative example of *i)* uncalibrated but sharp estimation, *ii)* calibrated but unsharp estimation and *iii)* calibrated and sharp estimation. We observe that in case *i)* the confidence intervals are all very close to the mean prediction, but they miss most ground-truth datapoints (e.g., the 100% confidence intervals do *not* contain all the datapoints). On the other hand, when the estimation is calibrated but not sharp (case *ii)*), we have truthful but unnecessarily wide confidence intervals. In the context of QoT estimation, case *ii)* can be interpreted as the large system margins adopted in margined formulas. Finally, in *iii)* a sharp and calibrated estimation provides both truthful and tight confidence intervals.

A. Background on Gradient Boosting

QoT regression is a ML problem usually formulated on tabular data, i.e., data that can be represented as rows (one for each sample) and columns. Each column represents a lightpath feature (e.g., modulation format, path length, etc.). Among many different ML algorithms, at present the best-performing ones for tabular data are Gradient Boosted Decision Trees (GBDT) models [18], [19]. Let $\mathcal{D}_{\text{train}} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ be our

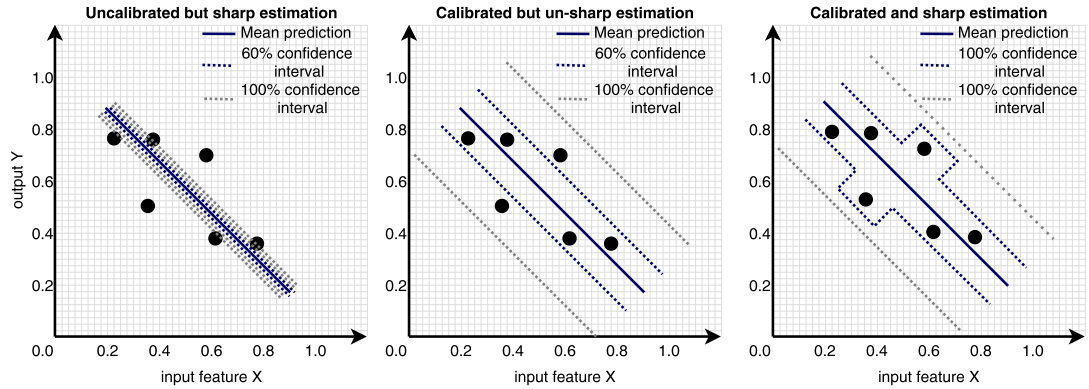

Fig. 2. Illustrative example of calibrated and sharp estimations. *Left*: uncalibrated but sharp estimation resulting in tight confidence intervals around the mean value but missing most datapoints. *Center*: calibrated but unsharp estimation with wide uniform width confidence intervals. The 100% confidence intervals cover all datapoints. *Right*: calibrated and sharp estimation. The 100% confidence intervals are tighter around the mean and cover all datapoints.

training dataset for QoT regression, where $\mathbf{x}_i$ are vectors of lightpath features and y_i are the corresponding ground-truth SNR values. Gradient Boosting is a function-approximation algorithm that iteratively updates its estimation f as follows:

$$f_t = f_{t-1} + \gamma h_t \tag{1}$$

where h_t is a base learner, and γ is the learning rate. In GBDT regression algorithms, base learners are regression trees. The intuition is that a base learner at iteration t is trained to correct the residual error of the approximation at iteration $t-1$. In particular, at each Gradient Boosting iteration, a base learner h_t is fitted on the training dataset with "pseudo-residuals" $r_{t,i}$ as targets. Pseudo-residuals are computed as follows:

$$r_{t,i} = -\left.\frac{\partial \mathcal{L}(y_i, f(\mathbf{x}_i))}{\partial f(\mathbf{x}_i)}\right|_{f=f_{t-1}} \tag{2}$$

where $\mathcal{L}$ can be any differentiable loss function. When $\mathcal{L}$ is the Mean Squared Error (MSE), the pseudo-residual is proportional to the true residual error. In general, for any differentiable loss function, each Gradient Boosting update (1) is equivalent to a gradient descent step in the function space.

B. Proposed QoT Regression Framework

To quantify the uncertainty in our predictions we need to perform probabilistic regression, i.e., our GBDT model must output a probability distribution. A common assumption in the literature is to model the output as a Normal distribution $p(y_i|\mathbf{x}_i) = N(\hat{\mu}_i, \hat{\sigma}_i^2)$, where means $\hat{\mu}_i$ and variances $\hat{\sigma}_i^2$ are predicted by our GBDT model from the input features $\mathbf{x}_i$. The model performs Maximum Likelihood Estimation (MLE) by minimizing the negative log-likelihood:

$$\mathcal{L}(y_i \mid \hat{\sigma}_i^2, \hat{\mu}_i) = \frac{1}{2}\left(\log(\hat{\sigma}_i^2) + \frac{(y_i - \hat{\mu}_i)^2}{\hat{\sigma}_i^2}\right) \tag{3}$$

However, it may be that the underlying distribution of the SNR is not Normal. In fact, we may not have any prior knowledge about the underlying SNR distribution at all. Because of that, this model alone may fail to accurately model the ground-truth SNR distribution.

To address this issue, we leverage recent advances on the calibration of probabilistic regression models. Calibration of a regression model allows to obtain accurate confidence interval estimates without specific assumptions on the underlying data distribution. In particular, we refer to the procedure developed in [15], which aims at improving the uncertainty estimates of Bayesian Neural Networks parameterizing a Normal distribution. In this paper, we apply this approach on GBDT models, which typically outperform neural networks on tabular data and are faster to train [18].

Let $F(\mathbf{x}_i)$ be the SNR's CDF predicted from lightpath features $\mathbf{x}_i$, and let its inverse $F^{-1}(\mathbf{x}_i)$ be the predicted quantile function. We say that a regression model is calibrated if the following holds:

$$\frac{\sum_{i=1}^{N} \mathbb{I}\{y_i \leq [F^{-1}(\mathbf{x}_i)](p)\}}{N} \xrightarrow[N\to\infty]{} p \quad \forall p \in [0,1] \tag{4}$$

For example, if our model is properly calibrated, we expect that approximately 95% of the ground-truth values y_i will be lower than their predicted 95-th quantiles $[F^{-1}(\mathbf{x}_i)](0.95)$.

Assume that our model outputs $[F(\mathbf{x}_i)](y_i) = 0.95$, i.e., it assigns the ground-truth value y_i to the 95-th quantile of the predicted SNR distribution given lightpath features $\mathbf{x}_i$. Moreover, assume that in our dataset only 80% of all ground-truth values y_i fall below their respective predicted 95-th quantiles, i.e., the inequality $[F(\mathbf{x}_i)](y_i) \leq 0.95$ holds for only 80% of the points in our dataset. This means that our model is uncalibrated (and overconfident) for 95-th quantile predictions. Because of this, to satisfy Eq. (4) we need to recalibrate 95-th quantile predictions to the 80-th.

Following this line of reasoning, we can learn a recalibration function by training a simple univariate regression model R on a recalibration dataset constructed as follows:

$$\mathcal{D}_{\text{cal}} = \left\{[F(\mathbf{x}_i)](y_i), \hat{P}([F(\mathbf{x}_i)](y_i))\right\} \tag{5}$$
$$\hat{P}(p) = \frac{|y_i \mid [F(\mathbf{x}_i)](y_i) \leq p, i = 1, \ldots, N|}{N}$$

Where $\hat{P}(p)$ is the fraction of ground-truth values y_i that fall below their predicted p-th quantile. To avoid overfitting, the recalibration dataset can be constructed using data not seen during the training of the probabilistic regression model.

As suggested in [15], we use Isotonic Regression to learn the recalibration function from the recalibration dataset $\mathcal{D}_{\text{cal}}$.

Algorithm 1 Calibrated QoT Probabilistic Regression

Require: Training dataset $\mathcal{D}_{\text{train}} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$

1: Fit a GBDT probabilistic regression model F on the training dataset $\mathcal{D}_{\text{train}}$ minimizing the negative log-likelihood:

$$\mathcal{L}(y_i \mid \hat{\sigma}_i^2, \hat{\mu}_i) = \frac{1}{2}\left[\log(\hat{\sigma}_i^2) + \frac{(y_i - \hat{\mu}_i)^2}{\hat{\sigma}_i^2}\right]$$

2: Build a recalibration dataset $\mathcal{D}_{\text{cal}}$:

$$\mathcal{D}_{\text{cal}} = \left\{[F(\mathbf{x}_i)](y_i), \hat{P}([F(\mathbf{x}_i)](y_i))\right\}$$

$$\hat{P}(p) = \frac{|y_i \mid [F(\mathbf{x}_i)](y_i) \leq p, i = 1, \ldots, N|}{N}$$

(Optional): use $\{(\mathbf{x}_j, y_j)\}_{j=1}^{M}$ not in $\mathcal{D}_{\text{train}}$

3: Fit a recalibrator R on $\mathcal{D}_{\text{cal}}$ with Isotonic Regression

4: **return** GBDT model F, recalibrator R

Isotonic Regression fits on a dataset $\{x_i, y_i\}_{i=1}^{N}$ a univariate non-decreasing function $\hat{y}_i = f(x_i)$ by solving the following Quadratic Programming (QP) optimization problem:

$$\min \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \quad s.t. \;\; \hat{y}_i \leq \hat{y}_j \;\; \forall i, j : x_i \leq x_j \tag{6}$$

This is because the recalibration function to be learned is non-decreasing, and using Isotonic Regression ensures that this monotonicity constraint is satisfied.

After training a GBDT probabilistic regression model F and the corresponding recalibrator R, the predicted SNR's CDF given lightpath features $\mathbf{x}_i$ is given by $(R \circ F)(\mathbf{x}_i) = R(F(\mathbf{x}_i))$. The full training pipeline for calibrated QoT probabilistic regression is illustrated in Algorithm 1.

Time complexity. Let *n* be the number of training samples. Isotonic Regression is *O(n)* for already sorted data [16]. The worst-case complexity of GBDT is implementation-dependent, as it hinges on how trees are built at each GBDT iteration. In our results, we refer to the CatBoost GBDT implementation [21]. The per-iteration CatBoost complexity is *O(n)* times algorithm-specific constants. A detailed analysis of the CatBoost algorithm is beyond the scope of this paper, but it can be found in Appendix C of [21].

IV. Synthetic Data Generation

A frequent challenge for research on applying ML in optical networks is the lack of field data, due to confidentiality requirements imposed by operators and vendors. Even when such data are available, operators might not have sufficient data representing QoT-unfeasible lightpaths to train an ML model. As a result, the generation of synthetic data is often the *go-to* solution to demonstrate the effectiveness of newly proposed QoT estimation approaches. We utilize the *E-tool* transmission simulation tool [6], that assumes frequency slice units (FSU) of 12.5 GHz, a total 4 THz link capacity and transceivers operating at 28 GBaud with a 37.5 GHz channel bandwidth. We assume dispersion uncompensated transmission over standard single model fiber (SSMF) with a 0.2 dB/km attenuation coefficient and signal power restoration thanks to amplifiers spaced every 100km. We assume an additional per-link random penalty parameter according to an exponential distribution with a 1 dB average. This added penalty accounts for the unpredictability of fast time-varying impairments (e.g., polarization effects [20]), and it is chosen to have an exponential distribution according to the principle of maximum entropy. Moreover, adding a per-link stochastic penalty allows to account for the number of nodes along the path and differentiate between lightpaths with similar path lengths but that cross a different number of nodes.

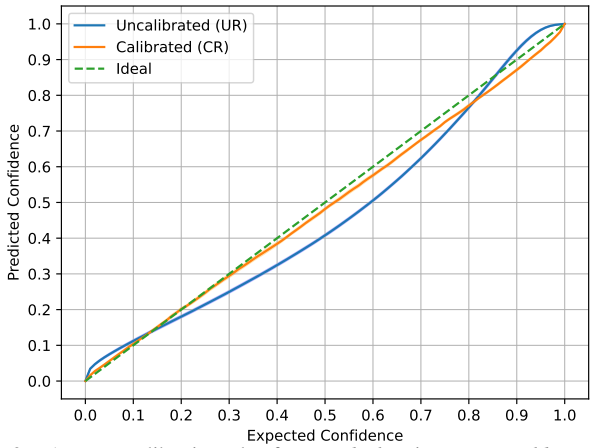

Fig. 3. Average calibration plot for an *Ideal* estimator, *Uncalibrated Regression* (UR) and *Calibrated Regression* (UR). Point below the *Ideal* line correspond to underconfident predictions, and viceversa.

For the training dataset, N lightpaths are randomly selected (e.g., N=300) by randomly choosing a bitrate in $[50, 500]$ Gbps range with 50 Gbps granularity and one of the $r \cdot M$ possible combinations (r=3 routes $\cdot$ $M = 6$ modulation formats, i.e., (DP)-BPSK, DP-QPSK and DP-n-QAM, with n = 8, 16, 32, 64). For each lightpath, the SNR calculation is repeated $k = 10$ times under different random penalty samples. To prevent overfitting, 10% of the training dataset is held-out as validation, and the model yielding the best loss on the validation set is retained. We did not use a separate dataset for recalibration, as we did not observe overfitting on the training dataset. For testing, we generated a larger dataset of 1800 lightpaths with $k = 100$ SNR samples each. While it is impractical to obtain such data in a real scenario, we use it at test time only to validate the performance of our proposed approach on data that is representative of the whole network and of the true SNR distribution. We chose 6-dimensional lightpath feature vectors $\mathbf{x}_i$ with the following information: modulation format, bitrate, shortest and longest link lengths in the path, total path length, number of hops in the path.

V. Illustrative Numerical Results

For all our experiments, we consider the 14-node Japan national topology [13] and report averaged results over 10 randomly generated training datasets. For the implementation of Algorithm 1, we used CatBoost [21] with default hyperparameters optimizing the "RMSEWithUncertainty" loss function, and Scikit-learn IsotonicRegression [22].

Figure 3 shows the calibration plots for Uncalibrated Regression (*UR*), which assumes a Normal distribution for the

TABLE I
AVERAGE RMSE VALUES FOR CALIBRATED REGRESSION (CR), UNCALIBRATED REGRESSION (UR) AND QUANTILE REGRESSION (QR)

Algorithm / Quantile	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90
CR	**1.43**	**1.18**	**1.05**	**0.98**	**0.91**	**0.86**	**0.81**	**0.78**	**0.78**
UR	**1.43**	1.26	1.24	1.24	1.19	1.08	0.93	0.79	0.93
QR	1.54	1.20	1.06	**0.98**	0.94	**0.86**	0.82	0.80	0.82

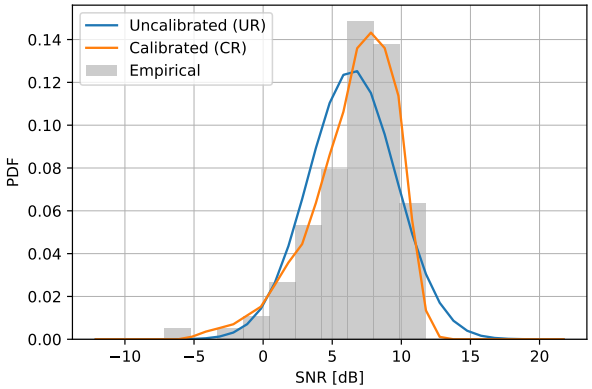

Fig. 4. Example of empirical SNR PDF (ground truth) and predicted SNR PDFs by *Uncalibrated Regression* (UR) and *Calibrated Regression* (CR).

SNR, and Calibrated Regression (*CR*), which results from the application of the learned recalibration function to *UR*. A perfectly calibrated model (*Ideal*) predicts exactly the expected empirical confidence and is represented as the diagonal in the calibration plot. Points below and above the *Ideal* line represent underconfident and overconfident predictions, respectively. We note that *UR* overestimates for the quantiles approximately above the 80-th and below the 12-th, and underestimates otherwise. After recalibration, *CR* almost-perfectly matches the *Ideal* calibration performance, being at worst slightly underconfident for the highest quantiles. In terms of the Mean Absolute Calibration Error, i.e., how much uncalibrated (either overconfident or underconfident) the model is with respect to the *Ideal* performance, *UR* shows on average a 5% error while *CR* shows on average a 1.7% error. To visually illustrate the results discussed above, we show the predicted SNR distributions of *CR* and *UR* together with the empirical ground-truth in Fig. 4 for an exemplificative lightpath. While *UR* models the output as a Normal and is therefore unable to capture the asymmetries in the SNR distribution, after recalibration *CR* provides a much more truthful match to the ground-truth.

Finally, we evaluate the performance in terms of RMSE with respect to nine equally-spaced empirical quantiles. As a baseline, we consider CatBoost Quantile Regression (*QR*), which estimates separately a fixed number of quantiles. Note that single-output quantile regression requires training one model per quantile, while our approach can estimate simultaneously all quantiles. Table IV shows the RMSE values for each quantile for *CR*, *UR* and *QR*. Overall, *CR* is the best performing approach, outperforming on average both *UR* and *QR*. In particular, RMSE for *CR* is up to 0.28 lower compared to *UR* and up to 0.11 lower compared to *QR*. This shows that *CR* is both calibrated and sharp, as quantile estimates are close to their corresponding ground-truth values.

VI. CONCLUSION

We proposed a ML-based probabilistic regression approach for estimating the SNR distribution and we explored a new calibration procedure to refine the model's predictions. The proposed *Calibrated Regression* allows to truthfully predict the SNR distribution without any prior knowledge on the underlying data distribution. We showed that a *Calibrated Regression* almost perfectly matches the performance of an *Ideal* estimator and provides better confidence interval estimates compared to an *Uncalibrated Regression*. Finally, we quantified the performance of *Calibrated Regression* in terms of RMSE with respect to ground-truth quantiles, and showed that it outperforms both an *Uncalibrated Regression* and single-output quantile regression.

REFERENCES

[1] J. Shao *et al.*, "Comparison of Split-Step Fourier Schemes for Simulating Fiber Optic Communication Systems," in IEEE Photonics Journal, 6 (4), pp. 1-15, 2014.
[2] P. Poggiolini *et al.*, "The GN-Model of Fiber Non-Linear Propagation and its Applications," in Journal of Lightwave Technology, 32 (4), pp. 694-721, 2014.
[3] F. Musumeci *et al.*, "An Overview on Application of Machine Learning Techniques in Optical Networks," in IEEE Communications Surveys and Tutorials, 21 (2), pp. 1383-1408, 2019
[4] M. Ibrahimi, C. Rottondi and M. Tornatore, "Machine Learning methods for Quality-of-Transmission estimation", In Machine Learning for Future Fiber-Optic Communication Systems (pp. 189-224).
[5] Y. Pointurier, "Machine learning techniques for quality of transmission estimation in optical networks," in Journal of Optical Communications and Networking, 13 (4), pp. B60-B71, 2021.
[6] C. Rottondi *et al.*, "Machine-learning method for quality of transmission prediction of unestablished lightpaths," IEEE/OSA Journal of Optical Communications and Networking, 10 (2), pp. A286-A297, 2018
[7] T. Jimenez *et al.*, "A Cognitive Quality of Transmission Estimator for Core Optical Networks," IEEE Journal of Lightwave Technology, 31 (6), pp. 942-951, 2013
[8] R. M. Morais *et al.*, "Machine learning models for estimating quality of transmission in DWDM networks," IEEE/OSA Journal of Optical Communications and Networking, 10 (10), pp. D84-D99, 2018
[9] A.Mahajan *et al.*, "Modeling EDFA Gain Ripple and Filter Penalties With Machine Learning for Accurate QoT Estimation," IEEE Journal of Lightwave Technology, 38 (9), pp. 2616-2629, 2020
[10] I. Sartzetakis *et al.*, "Accurate quality of transmission estimation with machine learning," IEEE/OSA Journal of Optical Communications and Networking, 11 (3), pp. 140-150, 2019
[11] N. Morette *et al.*, "On the Robustness of a ML-based Method for QoT Tool Parameter Refinement in Partially Loaded Networks," in OFC, pp. 1-3, 2022
[12] E. Seve *et al.*, "Associating machine-learning and analytical models for quality of transmission estimation: combining the best of both worlds," IEEE/OSA Journal of Optical Communications and Networking, 13 (6), pp. C21-C30, 2021
[13] M. Ibrahimi *et al.*, "Machine learning regression for QoT estimation of unestablished lightpaths," IEEE/OSA Journal of Optical Communications and Networking, 13 (4), pp. B92-B101, 2021
[14] L.A. Grieco *et al.*, "Ad-Hoc, Mobile, and Wireless Networks: 19th International Conference on Ad-Hoc Networks and Wireless," ADHOC-NOW 2020, Bari, Italy, October 19–21, 2020, Proceedings. Vol. 12338. Springer Nature, 2020.
[15] V. Kuleshov *et al.*, "Accurate Uncertainties for Deep Learning Using Calibrated Regression," Proceedings of the International Conference on Machine Learning (ICML), 2018
[16] M.J. Best, and N. Chakravarti, "Active set algorithms for isotonic regression; A unifying framework," Mathematical Programming 47, 425–439, 1990
[17] M. Ibrahimi *et al.*, "Machine Learning Regression vs. Classification for QoT Estimation of Unestablished Lightpaths," in OSA Advanced Photonics Congress (AP), paper NeM3B.1, 2020
[18] R. Shwartz-Ziv *et al.*, "Tabular data: deep learning is not all you need," 8th ICML Workshop on Automated Machine Learning (AutoML), 2021
[19] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," The Annals of Statistics, 29 (5), pp. 1189-1232, 2001
[20] Y. Pointurier, "Design of low-margin optical networks," in Journal of Optical Communications and Networking, 9 (1), pp. A9-A17, 2017.
[21] L. Prokhorenkova *et al.*, "CatBoost: Unbiased boosting with categorical features," Proceedings of the 32nd International Conference on Neural Information Processing Systems (NeurIPS), 2018
[22] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011