Deep Quantile Regression for QoT Inference and Confident Decision Making

Tania Panayiotou, Hafsa Maryam, Georgios Ellinas
KIOS Research and Innovation Center of Excellence
Department of Electrical and Computer Engineering, University of Cyprus
{panayiotou.tania, maryam.hafsa, gellinas}@ucy.ac.cy

2021 IEEE Symposium on Computers and Communications (ISCC) | 978-1-6654-2744-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISCC53001.2021.9631468

***Abstract*—This work examines deep quantile regression for quality-of-transmission (QoT) estimation and accurate decision making in optical networks. Quantile regression is applied to approximate QoT models capable of inferring QoT bounds for any future lightpath, according to a predefined level of certainty, for confident decision making, without the need to consider traditional margins at decision time. It is shown, that quantile regression automatically accounts for such margins, in a discriminative fashion, leading to a significant margin reduction and subsequently to more accurate inference of the QoT of unestablished lightpaths, when compared to the traditional margin-based decision approaches. Specifically, deep quantile regression for QoT estimation ensures that lightpaths with insufficient QoT will be accurately identified and rejected, while also identifying correctly lightpaths with sufficient QoT, making it a confident decision making tool for the planning of optical networks.**

***Index Terms*—QoT estimation; margin reduction; deep quantile regression; machine learning.**

I. Introduction

Traditional estimation of QoT in optical networks involves the utilization of static Q-factor models that consider the physical layer impairments (PLIs) [1] (e.g., crosstalk, polarization mode dispersion, amplifier noise), that however may lead to a waste of spectrum resources, mainly due to the overestimation of the non-linear PLIs [2]. To address this issue, machine learning (ML) techniques have been proposed for the inference of QoT models [3]–[5].

Regression is often applied to estimate a QoT model in order to conduct inference about the QoT value of any unestablished lightpath. Commonly, with regression, a least squares loss function is minimized to find a QoT model, which is, in essence, an approximation of the relationship between the input variables (i.e., lightpaths) and the output variables (i.e., QoT values). By definition, the output of the QoT model is an estimate or an approximation, hence containing some uncertainty. In fact, this estimate is the mean response value, with the variability around that mean representing the uncertainty resulting from the errors in the model itself and the noise in the input data. Obviously, in order to make confident decisions regarding the QoT of unestablished lightpaths in optical networks, a point estimate is not enough, as it may lead to the establishment of lightpaths with true QoT values that violate some acceptable threshold. Thus, it is necessary to measure the certainty level of an estimate, essentially building prediction intervals [6].

A prediction interval is an interval (i.e., range of values) in which a single future observation will fall, with a certain probability, based on the existing model (i.e., given what has already been observed). In ML, several approaches exist for building prediction intervals, including conducting Bayesian inference, Monte Carlo dropout inference [7], and through the estimation of conditional quantile functions (e.g., deep quantile regression) [8]. In this work, as a first step towards the investigation of prediction intervals for QoT-based decision making, a deep quantile regression framework is adopted. In this framework, deep neural networks (DNNs) are trained to estimate conditional quantile functions (i.e., models). Each quantile function is estimated by minimizing an asymmetrically weighted sum of absolute errors [9]. This information can be subsequently used for confident decision making. While regression analysis has been previously explored for QoT model estimation [10], a mathematical framework for appropriately quantifying the model and input data uncertainty towards confident decision making has not been investigated.

Related work has demonstrated the ability of ML to find QoT models of sufficient accuracy [4], [11], ultimately leading to margin reduction [12], [13] compared to the margins considered in traditional physical layer models (PLMs) [1], [14], [15]. In both PLMs and ML-aided QoT models, the usage of a margin aims to fix model inaccuracies (i.e., uncertainty) by shifting the model estimates accordingly; that is, the QoT values considered in decisions making are worst than the model estimates to reduce or even alleviate the possibility of establishing into the network lightpaths with insufficient QoT (i.e., the estimated Q-factor is decreased by this margin). In that case, however, the possibility of over-provisioning a lightpath is increased, effectively wasting spectrum resources. Clearly, margin estimation has a significant impact on both decision making (regarding the QoT of unestablished lightpaths) and network efficiency, hence it must be treated with care.

ML-aided QoT models have been shown to provide margin reductions, as compared to traditional PLMs, positively impacting lightpath provisioning decisions and resulting in network capacity savings. However, QoT model margins are rather roughly approximated, with the potential of ML to achieve further improvements in decision making remaining greatly unexplored. Specifically, margin estimation is currently based on worst-case performance criteria of the obtained QoT model, with decision making inheriting a tendency to largely underestimate the true QoT (e.g., Q-factor) of a lightpath. Briefly, a common practice is to approximate a QoT model by minimizing the mean squared error (MSE) loss function over a training dataset, with a test dataset used to evaluate the performance accuracy and error spread of the model [12], [13]. Error spread information is subsequently used for margin estimation, with the estimated margin being applied in decision making for all unestablished lightpaths; that is, a constant margin is considered for all lightpaths during decision making (i.e., non-discriminative decision making) .

In this work, it is shown how prediction intervals, and specifically deep quantile regression can be used to achieve discriminative decision making (i.e., having a different margin treatment for different lightpaths), resulting in an overall margin that is significantly reduced, yet leading to accurate decisions for both possible classes of lightpaths; that is, the class of lightpaths with a true QoT that is sufficient, and the class of lightpaths with a true QoT that is insufficient. In contrast, it is shown that existing (traditional) margin estimation methods [12], [13] lead to accurate decisions only in the latter class, significantly reducing the accuracy of the former; an indicator that existing methods highly underestimate the true QoT of the lightpaths.

II. Deep Quantile Regression for QoT Inference

The proposed framework is based on learning conditional quantile functions with DNNs. These functions can then be used for building prediction intervals, estimating appropriate margins for each unestablished lightpath, and ultimately utilizing this information for confident decision making. This section provides the preliminaries of this work by first formally defining the conditional quantile functions and then describing how such functions are learned by means of deep regression.

A. Conditional Quantile Function

Let X and Y be random variables with a conditional cumulative distribution function $F(y|X) = Pr(Y \leq y|X)$. For $0 < \tau < 1$, the conditional quantile function of Y given X is defined as [9]

$$Q_Y(\tau|X) = \inf\{y : F(y|X) \geq \tau\}, \tag{1}$$

that is, conditional quantile function $Q_Y(\tau|X)$ returns the minimum value of y from amongst all those values whose conditional cumulative distribution function value exceeds τ. Equivalently, $Q_Y(\tau|X)$ returns the y value such that

$$F(y|X) := Pr(Y \leq y|X) = \tau. \tag{2}$$

Like the conditional distribution function ($F(y|X)$), $Q_Y(\tau|X)$ provides a complete description of the statistical properties of the random variable Y given X and it is referred to as the τ quantile which in essence determines the percentage of a population that is above or below a certain threshold.

In the QoT estimation problem, let Y be the random variable of the QoT values (i.e., measured in dB) for a set of lightpaths (i.e., a population), and X be the random variable of the lightpath observations. Then, the τ quantile returns, for any given lightpath $X = x$, a QoT value y such that the probability that the true QoT value of x will be less than or equal to y is equal to τ (Eq. (2)). In general, it is desirable that this probability is low (i.e., $\tau \to 0$) to provide a lower bound for the estimations. In contrast, as $\tau \to 1$, upper bounds are derived; that is, the probability that the true QoT value of x will be above y is close to zero.

B. Calculating Prediction Intervals

A prediction interval is therefore calculated between τ_u and τ_l quantiles, where $\tau_u > \tau_l$, with each quantile returning y_u and y_l values, respectively, given a lightpath x. According to these quantiles the probability of the true QoT value, y_t, of lightpath x to fall between y_u and y_l is given by

$$Pr(y_l \leq y_t \leq y_u) = \tau_u - \tau_l. \tag{3}$$

Depending on the use case, these calculated predictions intervals must be able to support confident decision making. For example, for the QoT decision making problem, it is important that the lower τ_l quantile is estimated according to a small value. Nevertheless, the optimal τ_l value, cannot be known a-priori, as it depends on how each selection affects decision making. Specifically, it depends on the impact that each τ_l selection has on the classification accuracy of lightpaths within each class of interest and the targeted performance accuracies.

Further, regarding the upper τ_u quantile, clearly, this information does not directly add any value in QoT decision making. However, in this work an upper quantile is considered for completeness and to illustrate the extend of variability between the prediction intervals derived for each individual lightpath. This information can be used in future research efforts to investigate the reasons why the possible QoT values of some lightpaths are bounded according to small prediction intervals, while others are bounded according to larger intervals (i.e., the reasons why some lightpaths are subject to higher uncertainty). Identification of lightpaths that are subject to higher uncertainty may lead to further improvement in QoT estimation and decision making (e.g., such lightpaths can be characterized by certain qualities that the model was not trained to learn/does not know).

C. Learning Quantiles with DNNs

In general, quantile regression is an extension of classical least squares estimation of conditional mean models to the estimation of an ensemble of models for several conditional quantile functions. Specifically, conditional quantile functions are estimated by minimizing the asymmetrically weighted sum of absolute errors [9]:

$$L_\tau = \frac{1}{n}\sum_{i=1}^{n} \rho_\tau\big(y_t^i - \hat{Q}_Y(\tau|x^i)\big), \tag{4}$$

where

$$\rho_\tau(z) = \begin{cases} \tau z, & \text{if } z \geq 0, \\ (\tau - 1)z, & \text{if } z < 0, \end{cases} \tag{5}$$

n is the number of observed lightpaths, $x^i \in \mathbb{R}^d$ is a vector describing the i-th lightpath of X, $y_t^i \in \mathbb{R}$ is the true QoT value of this lightpath, and $\hat{Q}_Y(\tau|x^i)$ is an approximation of the τ quantile returning QoT estimate (i.e., prediction) $\hat{y}_\tau^i$ for lightpath i, $\forall\ i = 1, .., n$. Note that a special case of Eq. (4) is the median regression estimator which minimizes the mean absolute error (MAE) loss function (i.e., $\tau = 0.5$).

In this work, the τ quantile is parameterized by a DNN model that is trained to estimate $\hat{Q}_Y(\tau|X, \Theta)$, where Θ are the unknown parameters of the quantile DNN model. Specifically, the model is optimized to minimize the loss function of Eq. (4), given a training dataset $D = (X, Y) = \{x^i, y_t^i\}_{i=1}^n$. For training, the Adam optimization algorithm [16] was used.

III. Dataset Generation

The dataset was generated for the Deutsche Telekom (DT) network topology (Fig. 1), with the link distances scaled down three times to match the specifications of the Q-factor tool [14], [17] used for generating the ground truth QoT values of the lightpaths. The network capacity was set to 32 C-band wavelengths for each link in the network. Two thousand unicast and multicast lightpath requests were randomly generated, with a varying multicast group size of up to 3 destination nodes. Requests were generated according to a Poisson process with exponentially distributed holding times with a unit mean, for a network load of 100 Erlangs.

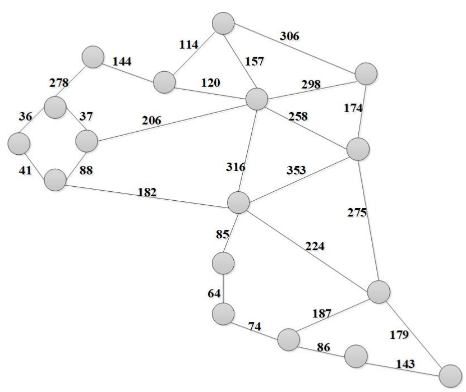

Fig. 1. DT network topology.

Multicast requests were routed according to the Steiner tree heuristic, unicast requests were routed according to the Dijkstra's algorithm, and the first-fit algorithm was applied for wavelength allocation. All provisioned light-trees were decomposed to a set of lightpaths, creating, together with the unicast connections, a dataset D with $n = 5497$ lightpaths in total. For each lightpath i, vector $x^i = [x_1^i, x_2^i, .., x_m^i] \in X$ was created according to the set of features considered in [3], [18] (i.e., x_1^i represents path length, x_2^i represents maximum link length, x_3^i represents allocated wavelength, etc.), shown to sufficiently describe a lightpath for ML-aided QoT modeling. The ground truths $y_t^i \in Y$ were computed by the Q-factor tool (i.e., having values in dB).

It is worth mentioning that while the dataset was synthetically generated, this does not affect the scope of this work, which is to demonstrate the potential of quantile regression to achieve margin reduction and accurate decisions making, compared to the state-of-the-art least-square regression and error spread-based margin estimation approaches. In essence, to demonstrate the potential of quantile regression to capture model and input uncertainty, subsequently utilizing this information in decision making. All approaches are evaluated and compared on the same dataset. As uncertainty is expected to be higher with datasets obtained utilizing real networks (e.g., due to noisy inputs in optical performance monitoring and/or outliers), such frameworks are expected to be even more important in practical real-world scenarios.

IV. Model Training and Inference Accuracy

Several QoT models were trained, with each model optimized according to a different loss function. Specifically, for quantile regression the loss function in Eq. (4) was applied for $\tau_l = 0.1$ and $\tau_u = 0.95$, to create the lower ($\hat{y}_l^*$) and upper ($\hat{y}_u^*$) estimates, respectively, for any unseen lightpath pattern x^* (i.e., a lightpath pattern that was not used during training). Additionally, a QoT model was trained to minimize the MSE function to generate the estimate (i.e., prediction) $\hat{y}^*$. These predictions correspond to the mean response of the model, commonly considered in the QoT estimation literature. In this work these predictions are utilized to estimate the margins and perform decisions in the conventional way, hence allowing us to compare with the proposed approach (i.e., used as benchmarks).

Both the quantile and least squares QoT models were parameterized by a DNN model with 3 hidden layers of 64, 64, and 32 hidden units. The rectified linear unit (ReLU) activation function [19] was used for all hidden units. Training was performed according to 40 epochs, with a batch size equal to 100, and a learning rate equal to 0.001. Before training, validation, and testing, dataset D was scaled for the input features to be standard normally distributed. Seventy percent (70%) of the patterns in D were used for training from which 20% was used for model validation, with the remaining 30% of the patterns in D used for testing (i.e., 1650 test patterns in total).

Figure 2 illustrates the $L_{0.1}$ loss, $L_{0.95}$ loss, and MSE loss performance as training evolves. Clearly, the loss for both training and validation datasets reduces and eventually converges as the number of epochs increases for all loss functions considered. Specifically, performance accuracy (for the test dataset) is 0.0259 for the $L_{0.1}$ loss function, 0.0195 for the $L_{0.95}$ loss function, and 0.0258 for the MSE loss function. Note that all models achieved to similarly reduce their respective loss function.

The resulting prediction interval, formed by the 0.95 and 0.1 quantile models, was validated according to the test patterns resulting in 90% of the QoT ground truths falling within the inferred 0.85 prediction interval, 99% falling below the inferred upper quantile, and 91% falling above the inferred lower quantile. Note that even though upper and lower quantiles were trained to infer a 0.85 prediction interval, our validation

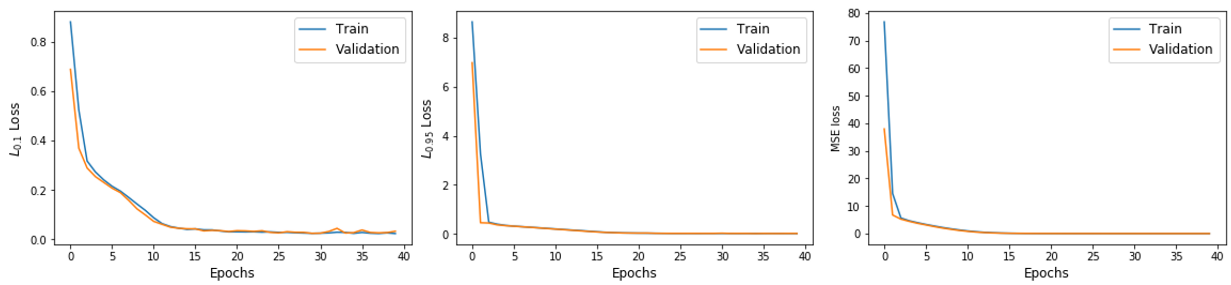

Fig. 2. Loss vs number of training epochs; $L_{0.1}$ loss function, $L_{0.95}$ loss function, and MSE loss function.

exceeded this probability. This is mainly due the fact that the quantiles are just approximations.

It should be mentioned that a large number of lightpaths was utilized for model training and testing. This was done to ensure that all models will converge by sufficiently minimizing their respective loss function, hence leading to a fair comparison between the models. As in practice, however, a large number of patterns may not readily available, especially at the deployment stage of the network, active and transfer learning techniques have emerged as promising techniques towards reducing the size of dataset required for QoT model training [20].

Furthermore, several other τ quantiles were trained, especially for estimating the lower τ_l quantile. In these preliminary results we opted to examine further only the 0.1 quantile that achieved significant margin reduction in conjunction with accurate QoT related decisions, as compared to the benchmark methods. Nevertheless, since further improvements may be possible by better fine tuning τ_l, examining the impact of various τ_l quantiles on margin reduction and decision accuracy is planned for future work.

V. Margin Estimation and Comparisons

Typically, a design margin is considered to compensate for the inaccuracies of ML-aided QoT models, approximated according to least squares techniques. Specifically, this margin is used to deal with the predictions overestimating the true QoT of lightpaths, hence penalizing accordingly all predictions. The two following approaches, reported in the literature for such margin estimation, are used as *benchmarks*:

- **Worst Error Margin** (M_w): is the maximum error resulting between the true and predicted QoT amongst all lightpaths with their true QoT overestimated by the predictions [13]:

$$M_w = \max_{y_t^i < \hat{y}^i} \{|y_t^i - \hat{y}^i|\}_{i=1}^{n'}, \tag{6}$$

where n' is the number of patterns in the test dataset, and $\hat{y}^i$ is the QoT prediction (i.e., mean response of the model).

- **Empirical Rule Error Margin** (M_e): is based on the empirical rule error stating that 99.7% of the data (i.e., errors) will be within three standard deviations of the mean error (i.e., $\mu \pm 3\sigma$), provided that the errors follow a Gaussian distribution. Hence, to account for the worst error [12],

$$M_e = \mu + 3\sigma, \tag{7}$$

where $\mu = \frac{1}{n'}\sum_{i=1}^{n'} |y_t^i - \hat{y}^i|$ is the mean absolute error resulting between the true and predicted QoT values in the test dataset, and $\sigma = \sqrt{\frac{1}{n'}\sum_{i=1}^{n'}(y_t^i - \hat{y}^i)^2}$ is the standard deviation of the error.

In the proposed framework, in essence, the τ_l quantile model automatically predicts a lower QoT estimate for each individual lightpath. Hence, in practice, margin estimation is not required, as the outputs of the lower quantile model can be directly used for decision making. However, for comparison purposes, to evaluate whether the proposed approach reduces the margins considered in the benchmark approaches, the quantile equivalent margin is utilized:

- **Equivalent Average Quantile Margin** ($\bar{M_q}$): is the mean difference between the predicted and the lower quantile QoT values of all lightpaths in the test dataset. Specifically,

$$\bar{M_q} = \frac{1}{n'}\sum_{i=1}^{n'}(\hat{y}^i - \hat{y}_l^i). \tag{8}$$

For each lightpath, the M_q margin is different, thus $\bar{M_q}$ is computed. Hence, just like M_w and M_e, $\bar{M_q}$ indicates the (average) equivalent penalty applied to a least squares prediction (i.e., indicates how much, on average, a prediction is reduced to reach the lower quantile value considered for decisions making).

Table I, reports the margin values estimated according to the different approaches, given the trained models and the test dataset. Clearly, $\bar{M_q}$ outperforms both benchmarks, resulting in a 75% margin reduction (on average) when compared to M_w, and in a 66% margin reduction (on average) when compared to M_e. Insights on why $\bar{M_q}$ achieves such a reduc-

TABLE I
Estimated Margins (in dB)

M_w	M_e	$\bar{M_q} \pm std$
0.832	0.5938	$\mathbf{0.2 \pm 0.15}$

tion are given by observing Fig. 3, illustrating the prediction intervals between quantile estimates $\hat{y}_l^*$ and $\hat{y}_u^*$, mean response predictions $\hat{y}^*$, and ground truths y_t^*, for a small number of lightpaths randomly selected from the test dataset. Clearly, ground truths of some lightpaths are tightly bounded by the quantiles (i.e., by a small prediction interval), while for others the prediction intervals are larger. In general, the prediction intervals vary, depending on the uncertainty level of the model, for each input lightpath. The lower quantile estimate is capable of capturing exactly this fact, by distinguishing between the different inputs, hence returning a QoT estimate that is not lower than necessary, given the desired certainty level (i.e.., the τ value).

On the other hand, both M_w and M_e margins, have no way of distinguishing between the input patterns, unnecessarily penalizing all lightpaths. As an example, the M_w margin will unnecessary reduce the predictions of several lightpaths (e.g., lightpaths $1, 4, 8, 9, 10, 17$ in Fig. 3) by a large constant margin to mitigate the error of lightpath 11 (i.e., the maximum error occurs for lightpath 11). The same holds true for the M_e margin which is lower than M_w, but still unnecessarily reduces the predictions of these lightpaths. Specifically, the lower quantile for these lightpaths suggests that the probability that such (low) values will occur is below 10%.

Nevertheless, lower quantile estimates still have a probability of failing, depending on the τ quantile selected. This is clearly shown for lightpaths 11 and 18 in Fig. 3, where their ground truths fall below the 0.1 (lower) quantile estimates. As $\tau \rightarrow 0$, this probability will reduce to zero. However, in decision making, having this probability go to zero is not of the utmost importance. Instead, it is more important to sufficiently reduce this probability to the point where decision accuracy, for both classes of interest, is the best possible; that is, the probability to accurately identify the lightpaths with insufficient QoT is close to 1, without significantly reducing the probability to accurately identify the lightpaths with sufficient QoT.

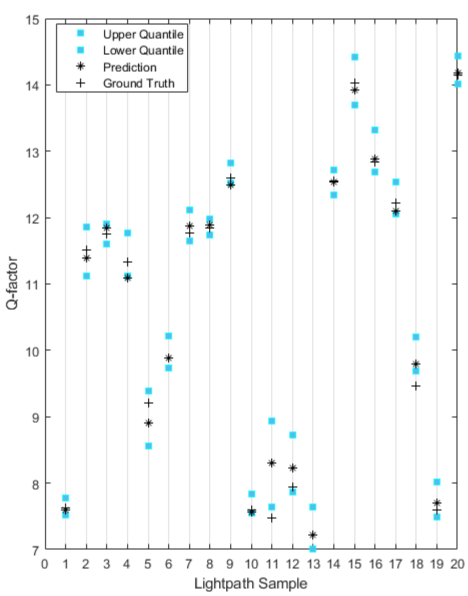

Fig. 3. QoT prediction intervals, ground truths, and predictions for a number of lightpaths.

VI. Decision Accuracy and Comparisons

To investigate how each margin estimation approach affects decision accuracy, the QoT models' outputs have been utilized to classify the estimated QoT of lightpaths in the two classes of interest, namely Class 1 and Class 2, that correspond to the lightpaths with insufficient QoT and with sufficient QoT, respectively.

Specifically, a lightpath belongs to Class 1 if its QoT is above a given QoT threshold (or equal to it) and to Class 2 otherwise. Several QoT thresholds are considered (Table II), with the accuracy evaluated on the lightpaths of the test dataset. Hence, decisions are taken according to the inferred QoT values reduced by the margin considered, and accuracy is tested against the ground truths in the test dataset. For the M_w and M_e margins, inference is performed according to the obtained least squares QoT model with the estimates reduced by M_w/M_e, while for the lower quantile approach, QoT estimates can be directly used for decision making, without any further reduction.

TABLE II
Classification Accuracy for Several QoT Thresholds

	Threshold 9.5 dB			Threshold 9 dB		
	M_w	M_e	M_q	M_w	M_e	M_q
Accuracy	0.92	0.93	0.96	0.88	0.91	0.95
Class 1 Accuracy	1	1	0.99	1	1	1
Class 2 Accuracy	0.82	0.87	**0.93**	0.79	0.84	**0.9**
	Threshold 8.5 dB			Threshold 8 dB		
	M_w	M_e	M_q	M_w	M_e	M_q
Accuracy	0.89	0.92	0.96	0.93	0.95	0.97
Class 1 Accuracy	1	1	1	1	1	0.99
Class 2 Accuracy	0.82	0.88	**0.93**	0.89	0.93	**0.96**
	Threshold 7 dB			Threshold 7.5 dB		
	M_w	M_e	M_q	M_w	M_e	M_q
Accuracy	0.74	0.76	0.86	0.67	0.82	0.96
Class 1 Accuracy	1	0.99	0.99	1	1	1
Class 2 Accuracy	0.7	0.73	**0.84**	0.66	0.82	**0.95**

Clearly, the classification accuracies of Table II indicate that while both M_w and M_e margin approaches succeed to accurately classify lightpaths in Class 1, this is done by highly underestimating the true QoT of all lightpaths, leading to low accuracy in Class 2. The quantile approach (i.e., M_q) similarly succeeds in Class 1 classification, yet it outperforms both M_w and M_e approaches in Class 2 classification; an indicator that M_q underestimates less the true QoT of lightpaths compared to the benchmarks. Specifically, the quantile approach outperforms benchmarks M_w/M_e up to $30\%/14\%$ in total accuracy, and up to $30\%/13\%$ in Class 2 accuracy, with up to 1% loss in Class 1 accuracy. Nevertheless, as previously mentioned, in the quantile approach, it is possible that this loss can be further reduced by fine tuning the τ_l value.

It should be noted that quantile regression has an additional advantage arising from the fact that the inferred quantiles (i.e., future observations combined with margins) do not depend on the statistical analysis of the error observed amongst the test patterns. Instead, it directly outputs QoT estimates combined with margins, for every future lightpath. In contrast, M_w and M_e, that depend on the statistical analysis of the error, may not yield appropriate margins, as future lightpaths may behave differently, radically changing the values of these margins (e.g., the worst error deviation possible may not be observed in the test patterns). Furthermore, M_e is based on the assumption that the error follows the Gaussian distribution, which may not be the case, especially for data obtained in the field which are subject to noisy observations.

Finally, it is important to be mentioned that quantile regression and decision making is not equivalent to the traditional classification approach previously applied in the literature for QoT related decisions [3], [10], [21]–[23]. The classification approach in essence combines prediction and decision making, which may lead to premature decisions. Specifically, while it has been shown in these works that high classification accuracies can be achieved, in total and in both classes of interest as well, traditional classification techniques tend to balance the classification error within the classes of interest, especially when enough observations (history) are present in both classes. On the other hand, in general, quantile regression has the advantage of controlling the classification error, depending on what is critical in decision making for the underlying use case.

VII. Conclusion

Deep quantile regression is a promising approach for accurate QoT decision making, without the need of significantly underestimating the QoT of all future lightpaths. Specifically, it is shown that quantile QoT models can achieve accurate decisions for the critical Class 1 (close or up to 100% accuracy), yet outperforming traditional margin-based approaches in total accuracy and in accuracy obtained within the class of lightpaths with sufficient QoT (Class 2). The latter improvement is based on that fact that quantile regression achieves to significantly reduce the traditional margins considered for QoT-related decisions. As margin reduction and accurate decision making ultimately save network capacity, an interesting future direction is the examination of capacity savings in conjunction with fine tuning the certainty level of the lower quantile (i.e., the probability τ) that is considered for decision making.

Acknowledgment

This work has been supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 739551 (KIOS CoE) and from the Government of the Republic of Cyprus through the Directorate General for European Programmes, Coordination and Development.

References

[1] B. Shariati et al., "Physical-layer-aware performance evaluation of SDM networks based on SMF bundles, MCFs, and FMFs," *IEEE/OSA J. Opt. Commun. Netw.*, 10(9):712—722, 2018.

[2] M. Klinkowski and K. Walkowiak, "Impact of crosstalk estimation methods on the performance of spectrally and spatially flexible optical networks," *Proc. ICTON*, 2018.

[3] T. Panayiotou et al., "Performance analysis of a data-driven quality-of-transmission decision approach on a dynamic multicast-capable metro optical network," *IEEE/OSA J. Opt. Commun. Netw.*, 9(1): 98-108, 2017.

[4] R. Proietti et al., "Experimental demonstration of machine-learning-aided QoT estimation in multi-domain elastic optical networks with alien wavelengths," *IEEE/OSA J. Opt. Commun. Netw.*, 11(1), 2019.

[5] R. Morais and J. Pedro, "Machine learning models for estimating quality of transmission in DWDM networks," *IEEE/OSA J. Opt. Commun. Netw.*, 10(10):D84—D99, 2018.

[6] W.Q. Meeker et al., "Statistical intervals: A guide for practitioners and researchers", Wiley, 2017.

[7] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning", arXiv:1506.02142, 2016.

[8] F. Rodrigues and F.C. Pereira, "Beyond expectation: Deep joint mean and quantile regression for spatiotemporal problems", *IEEE Trans. Neural Netw. Learn. Syst.*, 31(12):5377-5389, 2020.

[9] R. Koenker, "Fundamentals of quantile regression", In *Quantile Regression*, pp. 26-67, Cambridge University Press, 2005.

[10] Y. Pointurier, "Machine learning techniques for quality of transmission estimation in optical networks," *IEEE/OSA J. Opt. Commun. Netw.*, 13(4):B60–B71, 2021.

[11] Z. Gao et al., "ANN-based multi-channel QoT-prediction over a 563.4-km field-trial testbed," *IEEE/OSA J. Lightw. Technol.*, 38(9), 2020.

[12] E. Seve et al., "Learning process for reducing uncertainties on network parameters and design margins," *IEEE/OSA J. Opt. Commun. Netw.*, 10(2):A298–A306, 2018.

[13] I. Sartzetakis et al., "Accurate quality of transmission estimation with machine learning," *IEEE/OSA J. Opt. Commun. Netw.*, 11(3), 2019.

[14] G. Ellinas et al., "Multicast routing algorithms based on Q-factor physical-layer constraints in metro networks," *IEEE Photon. Technol. Lett.*, 21(6):365—367, 2009.

[15] A. Carena et al., "EGN model of non-linear fiber propagation," *OSA Opt. Express*, 22(13):16335-16362, 2014.

[16] D.P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv:1412.6980, 2014.

[17] N. Antoniades et al., "Performance engineering and topological design of metro WDM optical networks using computer simulation," *IEEE J. Sel. Areas Commun.*, 20(1):149—165, 2002.

[18] T. Panayiotou et al., "A data-driven QoT decision approach for multicast connections in metro optical networks," *Proc. ONDM*, 2016.

[19] P. Ramachandran et al., "Searching for activation functions," *Proc. ICLR*, 2018.

[20] D. Azzimonti et al., "Comparison of domain adaptation and active learning techniques for quality of transmission estimation with small-sized training datasets," *IEEE/OSA J. Opt. Commun. Netw.*, 13(1), 2021.

[21] T. Panayiotou et al., "Decentralizing machine-learning-based QoT estimation for sliceable optical networks," *IEEE/OSA J. Opt. Commun. Netw.*, 2(7):146-162, 2020.

[22] J. Mata et al., "A SVM approach for lightpath QoT estimation in optical transport networks," *Proc. IEEE BigData*, 2017.

[23] S. Aladin et al., "Quality of transmission estimation and short-term performance forecast of lightpaths," *IEEE/OSA J. Lightwave Technol.* 38(10):2807—2814, 2020.