Machine Learning Techniques for Quality of Transmission Estimation in Optical Networks

YVAN POINTURIER[1]

[1] *Huawei Technologies France, 20 quai du Point du Jour, 92100 Boulogne-Billancourt, France*
[*] *Corresponding author: yvan@ieee.org*

Compiled January 7, 2021

The estimation of the Quality of Transmission (QoT) in optical systems with machine learning (ML) has recently been the focus of a large body of research. We discuss the sources of inaccuracy in QoT estimation in general, we propose a taxonomy for ML-aided QoT estimation, we briefly review ML-aided optical performance monitoring, a tightly related topic, and we review and compare all recently published ML-aided QoT articles.

http://dx.doi.org/10.1364/ao.XX.XXXXXX

1. INTRODUCTION

Machine learning has regained popularity in the past few years as a tool capable of solving typically highly nonlinear classification or regression problems for which there is no known analytic solution or when known solving methods, analytic or heuristic, fail to give an answer in a reasonable time. Machine learning has already been extensively leveraged in the optical communication and networking community, with:

- 6 surveys within 3 years [1–6];
- 3 workshops or dedicated sessions: at NIST [7, 8] in 2019, OFC conference in 2020 [9] and APC conference in 2020 [10];
- 2 tutorials [11, 12] at the OFC conference in 2020.

ML-based Quality of Transmission estimation (ML-QoT) has been the topic of 2 recent conference talks [13, 14] and the ambition of this paper is to make a complete survey of the literature body for ML-QoT.

Network designers have long been interested in accurate, fast QoT estimation for services to be established in a future or existing network. Accuracy is important as modeling errors translate into design margins [15, 16], which in turn translate into wasted capacity [17] or unwanted regeneration [18, 19]. In the case of network planning (either for a new network or when planning the establishment of a new service in an operating network), second-timescale computations are appropriate. In the case of online provisioning, computations must be much faster, for instance well below 1 second per service. One example of online provisioning is restoration.

The main QoT metric of interest to the network designer is service or light path Bit Error Rate (BER), whose value determines whether a service is acceptable and "error free" (BER below a predefined threshold) or not. Today's transmission systems use Forward Error Correction (FEC) and the threshold is usually expressed in terms of pre-FEC BER. The BER of a service is in turn tightly related to its signal-to-noise ratio (SNR). Given a modulation format, there exists a one-to-one mapping between BER and SNR. In addition, a metric inherited from non-coherent transmission systems is the Q-factor, which again is directly mapped to BER ($Q = \sqrt{2}\mathrm{erfc}^{-1}(2BER)$); in that respect, (pre-FEC) BER, SNR, and Q-factor are equivalent and refer to a measure of the end-to-end service performance.

QoT estimation is only one sub-field within optical communication that has been addressed through machine learning techniques; however, even for a sub-field, the literature is already very rich. This can be explained by the complexity of the problem itself; QoT as a whole (BER, SNR, Q) can be estimated through machine learning, but QoT can also be decomposed into several terms, each of which can be the topic of machine learning estimation.

In this paper, we survey and discuss previous research that attacks the QoT estimation problem for transport networks by means of any machine learning-based technique. We restrict the review to coherent transmission networks, as non-coherent transmission is typically no longer deployed in transport networks. We acknowledge that QoT estimation in non-coherent transmission line has also been addressed with machine learning, but we leave this topic outside of the scope of this review, which targets practical, current transport networks. We have also left out of the scope of this survey papers that leverage machine learning to make a choice on a component parameter (e.g., Erbium Doped Fiber Amplifier (EDFA) setting [20]) or a network design parameter (e.g., wavelength allocation [20–22]) based on a QoT related metric (e.g., SNR maximization, minimization of the penalty of adding a channel on existing signals) as those papers do not estimate quality of transmission.

Conversely, the problems of optical performance monitoring (OPM) and QoT estimation are tightly related, as monitoring is needed to feed most QoT estimation frameworks. For this reason, we provide a brief review of machine learning-aided OPM.

This survey is organized as follows. Acronyms are expanded in Table 1. In Section 2, we discuss the two key sources of inaccuracy for QoT estimation: the (physical) model itself, and its inputs. In Section 3, we review the taxonomy further used in the article. In Section 4, we review machine learning-aided techniques for optical performance monitoring. The key contribution of this survey is Section 5, where all machine learning-based QoT estimation papers are summarized and compared according to the taxonomy introduced in Section 3; in particular, the crux of this survey consists of the summary Tables 2-5.

Table 1. Acronyms

ADC	Analog to Digital Converter
ANN	Artificial Neural Network
ASE	Amplified Spontaneous Emission
AUR	Area Under the ROC
BER	Bit Error Rate
DNN	Deep Neural Network
DSP	Digital Signal Processing
EDFA	Erbium Doped Fiber Amplifier
FEC	Forward Error Correction
GPR	Gaussian Process Regression
KNN	K Nearest Neighbors
ML	Machine Learning
NN	Neural Network
OPM	Optical Performance Monitoring
OSA	Optical Spectrum Analyzer
OSNR	Optical Signal-to-Noise Ratio
PDL	Polarization Dependent Loss
PDM	Polarization Division Multiplexing
QAM	Quadrature Amplitude Modulation
QoT	Quality of Transmission
QPSK	Quadrature Phase-Shift Keying
RMSE	Root Mean Square Error
ROADM	Reconfigurable Optical Add-Drop Multiplexer
ROC	Receiver Operating Characteristic
SNR	Signal-to-Noise Ratio
SSF	Split Step Fourier
SVM	Support Vector Machine
WDM	Wavelength Division Multiplexing
WSS	Wavelength Selective Switch

2. SOURCES OF INACCURACY IN QOT ESTIMATION

Before we discuss why QoT estimation generally remains inaccurate, we need to delve into the main sources of impairments that are accounted for in a QoT metric such as SNR.

SNR can be split into several terms, corresponding to different noise sources. In general, we can write $SNR^{-1} = \sum_k SNR_k{}^{-1}$ where SNR is the SNR of a service and SNR_k is the contribution of the physical effect indexed by k, for instance:

- linear Amplified Spontaneous Emission (ASE) noise injected by optical amplifiers (in which case the contribution is called OSNR, optical signal-to-noise ratio),
- nonlinear noise caused by the fiber Kerr effects,
- filtering penalties due to filters being narrower than the signal's bandwidth (and, conversely, node crosstalk, originating from imperfectly filtered signals from adjacent or other ports),
- transponder back-to-back penalty,
- polarization dependent losses (PDL).

The first three aforementioned effects have been modeled with machine learning and are discussed further in this article. The second to last effect is difficult to estimate through machine learning as it depends on each pair of transceivers and is usually measured. The last effect (PDL) depends on sources that are external, independent from the telecommunication infrastructure; indeed, fiber birefringence can be changed for instance by mechanical stress on a fiber, resulting in varying polarization state at the input of further components (amplifiers, filters, etc.). If those components exhibit PDL, then the random input polarization state in turn induces a random loss. Since PDL sources are inherently impossible to monitor, and thus cannot be used as features in a machine learning framework, PDL is difficult to model through machine learning, and static (e.g., worst-case) margins are often used.

A. Trade-off accuracy/speed

Many physical, non-machine learning based models for QoT estimation exist; accurate models exist for each effect, but the speed bottleneck is the modeling of nonlinear effects, for which methods range from the Split Step Fourier method to analytic models; the SSF method is very accurate and versatile as it can address complex scenarios including the mix of non-coherent and coherent signals in networks with dispersion management; however the trade-off is with speed, as SSF is very slow — minutes or more per service.

The (coherent) Gaussian Noise model [23] is much faster and is accurate only within its application domain, which does not include for instance lines with dispersion management; an approximation of the coherent Gaussian Noise model is the incoherent Gaussian Noise model [23], which is faster but less accurate than the coherent version; other variations exist [24, 25]. Experimental validations of the Gaussian Noise model can be found in [26–28].

QoT estimation through physical modeling is still a very active field, with a goal to always capture more effects applicable to more diverse scenarios, such as modeling Stimulated Raman Scattering, which can no longer be neglected in C+L systems [29].

B. Inputs inaccuracy

Of course, a very accurate QoT model (based on physics or machine learning) is not useful when its inputs or parameters are not known accurately — "garbage in, garbage out:" how can we model QoT when the type of fiber may not even be known? [30] Hence, a source of inaccuracy in QoT estimation is that of the inputs themselves. As mentioned above, SNR can be decomposed into several terms (OSNR, nonlinear penalty, etc.), whose inputs may be subject to uncertainty. In Fig. 1, we review sources of parameter or input uncertainty with their respective impact on QoT estimation.

Linear noise is typically the main impairment in optical networks and is driven by per-channel power at the input of each span, which is not monitored due to a lack of optical spectrum analyzers in amplifiers in current networks. Input powers are in turn tightly related to per-channel EDFA gain, which is difficult to model and is not monitored, as will be discussed in Section 3-A. Span losses include both the attenuation of the fiber itself and lumped losses such as splices (which are more numerous in older fibers) or connector losses. Uncertainty for span losses lead to improper amplifier setting and increased linear noise. Span losses can be measured in an operating network but are difficult to know when designing a network, and they can change with time. Wavelength dependence of the span losses is incurred due to Stimulated Raman Scattering. Noise figures are usually known, although with limited accuracy or without any gain or wavelength dependence. Note that per-channel power also affects nonlinear-effects modeling.

Fiber type and presence or not of dispersion compensation/management strongly affects both nonlinear and linear impairments, and are usually (but not always) well known. Span losses (through the span input power) and fiber length both affect nonlinear effect modeling.

Filter responses within Wavelength Selective Switches (WSS) and transponder back-to-back penalties can be known from calibration data sheets, but may vary slightly from component to component and due to aging; services whose baudrate are close to the filter width will be more impacted by filter degradation.

Finally, the typical PDL of a component is usually given in data sheets.

C. Discussion

Please observe that Fig. 1 is as generic as possible, and strongly depends on the exact network/scenario under consideration: information about the infrastructure (e.g., fiber type) may be lost over time, or mis-typed in the network management tool; or not available at all by design in the case of disaggregation/alien wavelengths; similarly, calibration data for filters or transponders may or may not be available at all, depending for instance on when the network was built, the equipment vendor, and the operator.

Overall, the three bottlenecks for QoT estimation with traditional (physical) models are currently:

- the lack of accurate physical models for some of the components, in particular, amplifiers;
- the speed of the models for some of the impairments;
- the dependence on uncertain parameters or inputs.

Machine learning can mitigate each of these bottlenecks. Components that are difficult to model can be "learned;" physical effects that are complex/time-consuming to model can also be learned. Indeed many of the machine learning techniques have low computational complexity when performing the estimation itself as the computational complexity is actually moved to the training phase, which is usually done offline. Uncertain inputs that cannot be directly monitored could be inferred by machine learning.

In addition, machine learning can also be applied to learn trends for physical phenomena (prediction) [31], which can be leveraged as well to predict the evolution of the parameters or inputs of the models (traditional or machine learning-based), and hence improve the accuracy of the outputs of the models. However, when those variations are caused by external sources independent from the physics of the considered system, as is the case with PDL, machine learning will likely be less useful.

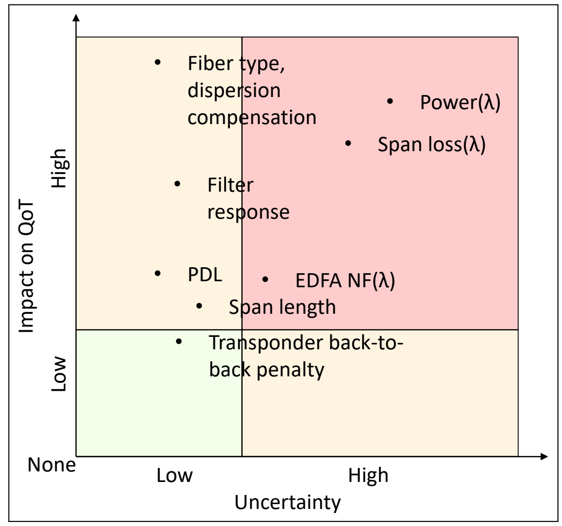

Fig. 1. Inaccuracy of typical QoT inputs.

3. TAXONOMY

This section introduces key notions used to classify and assess the methods proposed in the literature.

A. QoT models: classification, regression

Machine learning models for QoT can be split into two categories: classification and regression. Classification-based formulation typically solves the following problem: considering a service (typically to be established, or even already established if the establishment of a new service is expected to impact other services, in case of a low-margin scenario), will (is) its QoT be acceptable (BER below FEC limit) or not? In this case, QoT is modeled as a binary classifier. Classifiers are sufficient when the only information relevant to the network designer is the possibility to establish a service or not; however, more often, the network designer would like to know the exact value of the QoT, for instance to assess the robustness of the service to unplanned impairments (how far is the service from the threshold?), or to determine a capacity upgrade of the service (e.g., move it from 100 Gb/s PDM-QPSK (Quadrature Phase-Shift Keying) to 200 Gb/s PDM-16QAM (Quadrature Amplitude Modulation)). In this case, regression techniques are used. More rarely, classifiers with more than two classes have been proposed, each class corresponding to a QoT range, which can then be used to coarsely (depending on the number/granularity of the classes) assess the robustness or upgradability of a service.

B. Evaluation metrics

Depending on whether the QoT model is based on classification or regression, different metrics are used to evaluate the quality, or accuracy, of the model. Given a confusion matrix, where rows correspond to the predicted class, columns to the actual class, and cells to the number of instances from an actual class and predicted to be within a given class, the accuracy of a classifier is defined as the ratio between the number of correctly predicted instances with the total number of predicted instances, i.e., the proportion of true positives and negatives to the total estimated population size. Such a metric is known to fail when the number of samples in the two categories is greatly imbalanced; for instance, using this definition of accuracy, given a population with 99% of samples with class '1' and 1% of samples with class '0' then a trivial classifier that deterministically assigns class '1' to any sample will have an accuracy of 99%. For this reason, other metrics are also proposed, in particular, the area under the ROC curve (AUR), where the ROC (receiver operating characteristic) curve is the true positive vs. false positive rate curve parameterized by the decision threshold of the classifier. An AUR closer to 1 indicates a better classifier. A detailed discussion on this matter can be found in [32]. For simplicity, in the remainder of this paper, we will use wording "accuracy" for classifiers without further specifying the underlying metric.

Similarly, several metrics can be used to assess the accuracy of regression-based estimators. Typically, metrics are based on the distribution of the error, i.e., the difference between the estimated QoT $\hat{y}_i$ and the ground truth y_i for the service indexed by i. The average error $< |\hat{\mathbf{y}} - \mathbf{y}| >$ (where $< \cdot >$ denotes the mean of a vector) indicates on average how the estimator performs, but cannot be used to determine the margin to be taken to guarantee, with a certain probability, that the system will behave as expected, e.g., that the service will be error-free. The maximum error $\max_k |\hat{y}_k - y_k|$ would give such a margin with very high probability, while the cumulative distribution function of $(\hat{y} - y)$ can be used to obtain intermediate probabilities. Another popular metric, used more often in estimation frameworks due to its analytic properties, is the root mean square error (RMSE): $\sqrt{< (\hat{\mathbf{y}} - \mathbf{y})^2 >}$. Similarly, another popular metric is the R^2 parameter, which denotes the proportion of the variance of the data that can be explained by a model, i.e., $R^2 = 1 - \sum(y_i - \hat{y}_i)^2 / \sum(y_i - < y >)^2$; an R^2 parameter close to 1 indicates a better fit by a model.

C. Use cases: components, transmission lines, networks

To tackle the problem of estimating QoT for services network-wide, a divide-and-conquer approach can be used, by decomposing the network into transmission lines (the set of elements between two nodes or ROADMs), each of which could be modeled with machine learning techniques. Transmission lines can in turn be decomposed into a series of devices, e.g., amplifiers, fibers, WSS's/filters, each of which could be modeled with machine learning techniques. Assuming models exist for each device, it is nevertheless not guaranteed that their cascade is accurately modeled; indeed, small inaccuracies in the behavior of a component accumulate, leading to potentially large end-to-end inaccuracies. This is why network-wide or line-level methods are investigated in addition to the modeling of individual devices. Here we adopt the following criterion to distinguish between line and network studies. Whenever the proposed method uses network-wide data, i.e., the data used to train the model comes from anywhere in the network and not only from the line crossed by a considered service, we classify the paper in the "network" category. In cases where individual lines are considered and trained on synthetic data due to practical reasons, since a realistic scenario would require using training data from already established light paths to model the line, we also classify the paper as a "network" paper. If each transmission line is considered independently one from another with no input coming from any other line, the paper is classified as a "line" paper.

D. Data source: greenfield vs. brownfield

Machine learning intrinsically relies on a training step, where data is gathered to train or calibrate a model. For instance, data can be:

- generated using a modeling tool known to be accurate but having some other drawback such as speed, for example the SSF method or the Gaussian Noise model [23];
- collected in the lab on the equipment prior to deployment or on equivalent equipment;
- collected directly in the field before/during commissioning, or even during network operation (online).

Data collection may be difficult, time-consuming or costly, and it is generally desirable to reduce the amount of training. In particular, when a model for some system (device, line, network) has been trained for a specific set of parameters (e.g., a modulation format), it is expected that adjusting the model for another yet similar parameter value (e.g., another modulation format) can be done without having to fully re-train the model. Some of the papers described here leverage transfer learning, which aims at minimizing the amount of new data needed to re-train a machine learning model while achieving the same accuracy when changing the value of a key parameter of the considered system.

As machine learning relies on a training step, which in turn relies on data collection, most of the machine learning techniques described in this survey apply to "brownfield" line/networks which have already been deployed and are running, such that it is possible to collect data directly from the field to train models. The typical scenario is capacity upgrade, where the operator needs to verify that QoT of a new service to be established is acceptable, or where the operator would like to determine the maximum capacity the service can carry given its estimated QoT. Additionally, the operator may want to verify that the introduction or restoration of the new service does not negatively impact the services already established, in case the network was designed with low margin [16]. It is possible to go beyond relying on data from existing services, and to actively probe the network to generate more data ("active learning"). However, some of the presented techniques also apply to lines/networks yet to be physically deployed, where field data is not available - the "greenfield" scenario. A typical scenario is the estimation of the QoT of all services to be established right after the line/network is deployed. An intermediate scenario is when models are trained using a limited amount of data collected right after a line/network is deployed, for calibration; networks cannot be designed using the model as training data is not yet available, however resource (power, capacity, routing, spectrum, etc.) allocation can still be computed before the network goes live. We consider this scenario as greenfield in this survey.

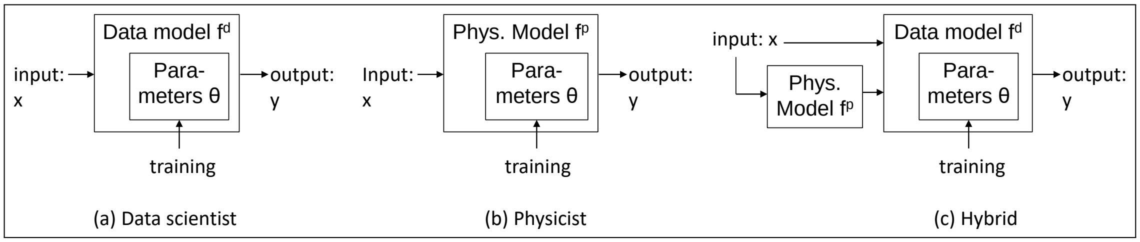

Fig. 2. Machine learning approaches for QoT estimation.

When a model has already been trained on a ("source") line or network, it is possible to adjust the model on a different ("target") line or network with only a small amount of data acquired on the target line/network, using a technique called transfer learning. In case transfer learning is used at the line level, the source and target lines may be in the same network — a brownfield scenario. In the other cases, i.e., the target is a line in a new network, or the source and targets are complete networks, then transfer learning applies to the greenfield case, with a caveat: as transfer learning relies on training on the target (line or network), a limited but nonzero amount of data collection is needed on the target, corresponding to the intermediate greenfield scenario mentioned above.

E. Learning Type: Data scientist, Physicist, Hybrid

We introduce in this survey a novel type of classification for machine-learning based algorithms applied to QoT estimation. Consider a physical model f (for instance, the computation of OSNR after a fiber span and an amplifier) parameterized by θ (for instance the noise figure and gain of the amplifier or the linear loss parameter of the fiber, known only from data sheets) to estimate y (for instance, said OSNR) based on inputs or features x (for instance, the launch power of a channel in the fiber): $y = f_\theta(x)$.

The traditional approach, which we call "data scientist" (Fig. 2a), models a system as a black box $f_\theta = f_\theta^d$ that has no a priori knowledge of the underlying physics of the system. This does not mean that a data scientist handling the QoT modeling problem should not understand the physics of the system to model; for instance, physical skills/insight will play a role when selecting/formatting/normalizing the features, etc. The black box (e.g., a neural network) is parameterized (e.g., θ implements the neural weights) and the training step consists of setting θ such that the resulting data model approximates well the underlying physical model. Most papers summarized in this survey use this approach.

In the second approach, which we call "physicist" (Fig. 2b), it is assumed that a physical model $f_\theta = f_\theta^p$ already exists for the model of interest, but that the parameters θ (for instance, the fiber length or linear attenuation) of the system are not known accurately. The "physicist" approach assumes that f_θ^p is correct but that QoT estimation inaccuracies stem from θ. The goal of the training step is then to refine or calibrate θ so as to make f_θ^p more accurate when presented with a new input (for instance, a new launch power).

In the third approach, which we call "hybrid" (Fig. 2c), the system of interest is again modeled as a black box $f_\theta = f_\theta^d$ as in the data scientist approach above (for instance, f_θ^d is a neural network), and the traditional inputs are supplemented with the output of the physical model corresponding to the system f^p (for instance, the analytic OSNR calculation for a fiber span); this biases the data scientist model towards the behavior of the physical model, while physical modeling deficiencies are addressed through training on the real system.

As a further illustrative example, [33] compares the first and third approaches with an analytic method (f^p is the GN model) in the same paper for SNR estimation; the **data scientist** method is exemplified by a neural network (f^d) trained on observed network states, while the **hybrid** method additionally feeds said artificial network with the output of the analytic model. All techniques are evaluated in the context of noisy model parameters (in this case, the noisy parameters are the fiber lengths): the analytic model is as accurate as its parameters (i.e., it is very accurate for non-noisy parameters but very inaccurate for noisy parameters), while the plain neural network, being trained on noisy parameters, outperforms the analytic model when parameters are noisy. As the hybrid method leverages a neural network trained on noisy parameters, it performs well in a noisy situation; being also trained for the non-noisy situation through the output of the GN model, the hybrid method also performs well in the non-noisy situation. In contrast, a **physicist** method would first learn (de-noise) the parameters and feed those into an analytic QoT model such as the GN model.

4. ML-OPM: MACHINE LEARNING-BASED OPTICAL PERFORMANCE MONITORING

ML-based OPM, or ML-OPM for short, consists in estimating a single component such as linear or nonlinear noise, or separating a compound QoT-related metric into separate components, such as the linear and nonlinear noise contributions from the received samples at a coherent receiver. ML-OPM has received a lot of attention and surveys can be found in [1, 3, 6]. ML-OPM essentially behaves as a software OPM and voids the requirement for dedicated hardware, although some of the proposed techniques rely on analysis of samples already available just after the analog to digital converters (ADC) or further down in the coherent receiver digital signal processing (DSP) chain: such processing clearly requires modifications on the coherent DSP chip, which may entail additional development cost and energy consumption.

ML-OPM permits one to gain more insight on the impairments sustained by signals already present in a communication system, and as such can be leveraged for ML-QoT; however, ML-OPM in itself is not a QoT estimation tool. For this reason, an exhaustive survey of ML-OPM is left out of the scope of this paper. However, because ML-OPM and ML-QoT are strongly related, we give a very brief overview of ML-OPM here, as related work.

As early as 2010, [34] proposed a neural network to estimate the OSNR, chromatic dispersion, and polarization-mode dispersion of an established service. The neural network was trained with parameters derived from asynchronous constellation diagrams. Since the monitoring information lies in the transitions rather than in the detection samples, the technique does not require sampling at the signal symbol rate and a photodiode with low-speed ADC and slow DSP are sufficient. The technique was applied for QPSK signals with no polarization diversity. A similar technique was presented in [35] for 16QAM signals with no polarization diversity.

OSNR monitoring was also the topic of [36–39], which all process receiver post-ADC samples with neural networks to estimate OSNR. Chromatic dispersion is additionally monitored in [40], also using the post-ADC samples, using a long short-term memory neural network. The requirement for processing at the coherent receiver was relaxed in [41], where a neural network is fed by samples from a single photodiode followed by a fast but asynchronous ADC, thereby enabling OSNR monitoring at any point on a line, at the expense of deploying extra hardware. This framework is extended in [42], where transfer learning is used to decrease the amount of retraining of a neural network trained for a modulation format (e.g., PDM-16QAM) but used for another modulation format (e.g., PDM-64QAM). The separation of linear and nonlinear noise was tackled in [43], which performs constellation analysis using a neural network — leading to an implementation in a commercial coherent modem in [44]; such separation was also addressed in [45], which estimates using a neural network the ratio between linear and nonlinear noises from the received spectrum obtained from the received samples or through a high resolution optical spectrum analyzer.

5. ML-QOT: MACHINE LEARNING-BASED QUALITY OF TRANSMISSION ESTIMATION

The summary of the key features according to the taxonomy presented in Section 3 and of the outcomes of the papers surveyed in the article can be found in Tables 2-5, each table corresponding to one of the subsections below: amplifiers modeling, line modeling with regression, network modeling with classification, network modeling with regression. Following the discussion in Section 3-D, for the transfer learning-related techniques we left empty the "data source" column. In the tables, we also indicate whether results were obtained experimentally or through simulation in the "Type of study" column. We also list the features needed to estimate the QoT metric in the "Inputs" column. All algorithms are supervised and rely on training data, which can be obtained through monitoring on a live network, emulation/calibration (e.g., from lab experiments) or simulation (e.g., using a physical model).

A. Amplifiers modeling

The first set of papers (Table 2) aim at modeling the behavior of amplifiers — a single EDFA ([46–49]) , the combination of Raman amplification with an EDFA [50], and a single Raman amplifier (along with a 100 km fiber span) over the C+L bands [51].

Modeling of amplifiers is key as most systems are OSNR-limited, i.e., the main impairment is the ASE noise originating from amplification. The amount (power variance) of noise generated by an amplifier is deceptively simple and given by $P_{ASE} = hf(GF-1)B$ where h is Planck's constant, f is the central frequency of the signal, B is the bandwidth in which the noise is considered, G is the amplifier gain and F is its noise figure. However, in reality EDFA gains are wavelength dependent; to make things more complicated, the wavelength dependence (the shape of the gain) nonlinearly depends on the input power spectrum of the EDFA, i.e., on which input channels are lit, and on the power of each of the input channels, due for instance to spectral hole burning, an effect that is difficult to model and affects amplifier gain at the short wavelengths in the C band [52]. This makes EDFAs good candidates for machine learning modeling. Similarly, Raman amplifiers are modeled using a set of ordinary differential equations [53] and are also good candidates for machine learning modeling.

The premises of all aforementioned papers in this section are essentially the same: given the input power spectrum of an amplifier, estimate the gain per channel or equivalently the output power spectrum using a neural network trained for instance in the lab, prior to deployment.

With the exception of [49], which uses the "hybrid" approach, all amplifier modeling papers use the "data scientist" approach. In [49] a neural network is fed with the outputs of an EDFA analytic model in addition to the standard modeling inputs; although the technique ultimately does not improve modeling accuracy, it improves the training time and reduces the amount of training samples.

Each study was done for a single amplifier, with a single setting, such as the amplifier gain or total input power, with the notable exception of [50], which additionally outputs the amplifier's settings to achieve a given response; as related work, the problem of finding amplifiers' settings to achieve a target gain shape is put in the broader context of inverse system design in [54], whereby a machine learning framework is used to learn what parameters yield a desired output for some system, such as an optical amplifier. Solving the problem for a single setting implies that training may have to be done for each amplifier's settings. The reported results, an error (RMSE or maximum) typically below 0.2 dB, should be considered in light of the application of the estimation framework. In a line with 10 or 20 such amplifiers, errors may accumulate to several dB in optical spectrum and then in SNR.

B. Line modeling: regression

We now consider the modeling of an individual optical point-to-point line (Table 3) [55–59]. All papers rely on neural networks except [55], which leverages Gaussian Process Regression (GPR). All follow a "data scientist" approach.

In [55, 57], the greenfield scenario is considered; [55] shows that training the GPR model on synthetic data enables accurate BER estimation in deployed systems, while [57] assumes that a line model is trained prior to deployment or at commissioning time based on measurements on the real line. Both [56, 57] model linear effects only ([56] in the very special case of a submarine link subject to strict power supply constraints) for two different reasons: in [56], operation far from the nonlinear regime is assumed, while in [57], it is assumed that nonlinear effects can be accurately modeled e.g., using the Gaussian Noise model — experiments are accordingly carried out on a cascade of amplifiers without any fibers, whose losses are emulated with variable optical amplifiers — and that most uncertainty on QoT modeling comes from the lack of an accurate model for the amplifiers, as discussed in Section 5-A. In all cases, estimation error of a fraction of a dB is reported.

Papers [58, 59] further propose transfer learning techniques to adjust a neural network trained on one line to another line with little amount of retraining, to estimate OSNR [58] or Q [59].

C. Network modeling: classification

In the following set of papers (Table 4), a classification framework is adopted, generally to estimate whether a service has an acceptable QoT or not [32, 60–65], and, in the case of [66], to give a coarse estimate of the value of the OSNR (as only ASE noise is modeled) through binning using a four-class classifier. For the sake of completeness, [61] is applied to multicast services and is generalized to multicore networks in [67], and to multi-slice networks with distributed controllers in [68], while [69] further leverages the work from [32] for improved resource allocation.

Typical classification accuracy is 90% and reaches 99% in several of the papers. A problem with classification is the lack of training data for the "unacceptable BER" class, as operators typically do not try to establish services they anticipate will fail, or at least target a very small blocking rate. This biases the classifier and is usually solved by generating data (services) with unacceptable BER by simulation. Such data generation is not practical in live networks, where forcing collection of samples in that class could instead be done through probing including establishing services with poor BER as in [32, 62]. In the case of probing, it should be ensured that sending new signals in a functioning network does not disturb the operation of the network; also additional equipment would be needed to probe the network as transponders are valuable resources, which are typically not left idle by operators except in the case of protection.

In [65], the amount of training data for the binary classifier proposed in [32] is reduced through transfer learning when the classifier is applied to a new line or network; similar accuracy can be obtained on the target without full retraining of the machine learning model. In [70] the transfer learning technique is compared with the probing-based active learning technique already mentioned above [62] in terms of samples to collect; compared with transfer learning, active learning requires fewer, but targeted samples that may be more difficult or costly to collect.

A large variety of machine learning techniques, all of the "data scientist" type, were proposed to tackle the classification problem: Support Vector Machines (SVM), K Nearest Neighbors (KNN), random forest, GPR, logistic regression, and neural networks, typically used for regression but here used for classification. All papers rely exclusively on simulations, based on synthetic data sets generated through one of the QoT analytic models introduced in Section 2-A, such as the GN model.

It should be noted here the remarkably small amount of required features needed to achieve high classification accuracy; in particular, none of the features needed to model nonlinear effects is present in some of the papers, for instance: per-channel or at least total input span power; length of each span: as nonlinear effects occur at the beginning of the spans, if all spans are sufficiently long, then nonlinear effects are similar for all spans and the exact lengths are not needed, but a high diversity in span lengths would yield a high diversity in nonlinear impairments; also span lengths drive amplification noise; this suggests that OSNR-limited scenarios are considered, or that even higher classification accuracy could be achieved with additional features. Note that this discussion can be generalized to the regression papers, which do not all use the features that are needed to physically model nonlinear effects.

D. Network modeling: regression

In the last set of papers (Table 5; see also Table 4 for [64]) [19, 64, 71–83], network-wide information is leveraged to train and estimate the QoT of a service in the brownfield/upgrade scenario with one exception: [81] also applies to greenfield. Those techniques are highly diverse: 8 papers use a data scientist approach, 5 a physicist approach, and 3 a hybrid approach, using such diverse techniques as neural networks, optimization through filtering, gradient descent, least squares or SVM regression; 10 papers report simulations and 5 experiments.

The physicist approach is used in [19, 72, 74, 79, 80] to learn the parameters of a physical model or to refine them in case those inputs are not accurately known; in particular, [80] focuses on learning EDFA gain ripple penalties from end-to-end measurements (as mentioned in Section 5-A, such technique can be cast in the more general framework of inverse system design), while [79] additionally learns filtering penalties. The other 3 papers [19, 72, 74] deal with end-to-end QoT as a whole. Note that [74] also uses the data scientist approach. The problem of handling parameter uncertainty is also tackled in [81] using the hybrid method, but without trying to refine the parameters.

The hybrid approach is used in [75, 76, 81], where artificial neural networks are fed both with network and signal-related features as in other works but also with the output of a modeling tool (e.g., the Gaussian Noise model) optionally complemented with monitoring data. Paper [75] additionally considers networks with mixed types of fibers, while [81] leverages information gained from a network to apply to another network, without relying on transfer learning.

The range of applicability of the proposed technique is strongly impacted by the type of study (simulation or experimental) and varies accordingly widely depending on the papers; as a representative example, in [64], a simulation paper, results are reported for four network topologies with high topological diversity (nodes, links, node degree, link lengths, fiber type), for SNR margin values spanning 40 dB; on the other extreme, [78] is based on a field trial over a single transmission line where most characteristics are fixed, such that the prediction range is limited to the interval [16, 16.6] dB.

6. CONCLUSION

QoT estimation is just one of many optical networking fields that have recently been tackled through machine learning. In only four years, dozens of articles were published on the specific topic of machine learning-aided QoT estimation. This field is actually very rich, as exemplified by the variety of the techniques/machine learning tools that are used and the targeted application, from individual components modeling to full network modeling, from the modeling of a single physical effect to the modeling of the end-to-end BER, from networks still at the design stage (greenfield) to networks already deployed and waiting for an upgrade (brownfield).

Machine learning proves to be useful when traditional modeling is difficult or slow, which is the case for some of the key optical networking components such as optical amplifiers, but also full lines (which cascade said components) and consequently whole networks. Machine learning can replace physical models through a black box approach, but also complement existing physical models through refinement of the model parameters. The results reported in the literature in terms of accuracy and even speed are very encouraging. Of course, as in any machine learning related research, the quality of the underlying data is essential.

In particular, machine learning relies on training, which is often specific to components or lines; solutions include training in the lab (at the expense of the real, field experience and possibly accuracy), online (which limits the use case to "brownfield" or capacity upgrade, but is likely more accurate through reliance on actual, monitored data) or through retraining a model already trained in the lab as with transfer learning techniques. Therefore, researchers have to find the right trade-off between synthetic data generation, where machine learning may simply end up learning the algorithm used to generate the data, and experimental studies, which are limited in scope due to the difficulty and cost of emulating a wide range of scenarios.

ACKNOWLEDGMENTS

The author thanks Gabriel Charlet for the insightful discussions.

REFERENCES

1. J. Mata, I. de Miguel, R. J. Durán, N. Merayo, S. K. Singh, A. Jukan, and M. Chamania, "Artificial intelligence (AI) methods in optical networks: A comprehensive survey," Elsevier Opt. Switch. Netw. **28**, 43–57 (2018).
2. D. Rafique and L. Velasco, "Machine learning for network automation: overview, architecture, and applications," IEEE/OSA J. Opt. Commun. Netw. **10**, D126–D143 (2018).
3. F. Musumeci, C. Rottondi, A. Nag, I. Macaluso, D. Zibar, M. Ruffini, and M. Tornatore, "An overview on application of machine learning techniques in optical networks," IEEE Commun. Surv. & Tutorials **21**, 1383–1408 (2019).
4. M. A. Amirabadi, "A survey on machine learning for optical communication [machine learning view]," arXiv:1909.05148 [eess.SP] (2019).
5. R. Gao, L. Liu, X. Liu, H. Lun, L. Yi, W. Hu, and Q. Zhuge, "An overview of ML-based applications for next generation optical networks," Sci. China Inf. Sci. **63** (2020).
6. X. Liu, H. Lun, M. Fu, Y. Fan, L. Yi, W. Hu, and Q. Zhuge, "AI-based modeling and monitoring techniques for future intelligent elastic optical networks," MDPI Appl. Sci. p. 363 (2020).
7. J. Gordon, A. Battou, M. Majurski, D. Kilper, U. Celine, D. Zibar, M. Tornatore, J. Pedro, J. Simsarian, and J. Westdorp, "Summary: Workshop on machine learning for optical communication systems," Https://doi.org/10.6028/NIST.SP.2100-04.
8. J. Gordon, A. Battou, and D. Kilper, "Workshop on machine learning for optical communication systems: a summary," in *Proc. IEEE/OSA Optical Fiber Communication Conference (OFC),* (San Diego, CA, USA, 2020), p. M3E.3.
9. "The role of machine learning for the next-generation of optical communication systems and networks," (2020). IEEE/OSA Optical Fiber Communication Conference (OFC), sessions M3E and M4E.
10. "AI-based optics," (2020). Asia Communications and Photonics Conference (ACP), workshop 1.
11. R. Morais, "Machine learning in multi-layer optical networks: Why and how," in *Proc. IEEE/OSA Optical Fiber Communication Conference (OFC),* (San Diego, CA, USA, 2020), p. M1B.1.
12. F. N. Khan, Q. Fan, J. Lu, G. Zhou, C. Lu, and A. Pak Tao Lau, "Applications of machine learning in optical communications and networks," in *Proc. IEEE/OSA Optical Fiber Communication Conference (OFC),* (San Diego, CA, USA, 2020), p. M1G.5.
13. R. Ayassi, A. Triki, M. Laye, N. Crespi, R. Minerva, and C. Catanese, "An overview on machine learning-based solutions to improve lightpath QoT estimation," in *Proc. International Conference on Transparent Optical Networks (ICTON),* (2020), p. We.A3.1 (Invited).
14. Y. Pointurier, "Machine learning-aided quality of transmission (QoT) estimation," in *Proc. International Conference on Transparent Optical Networks (ICTON),* (2020), p. We.D2.2 (Invited).
15. J.-L. Augé, "Can we use flexible transponders to reduce margins?" in *Proc. IEEE/OSA Optical Fiber Communication Conference (OFC),* (Anaheim, CA, USA, 2013), p. OTu2A.1.
16. Y. Pointurier, "Design of low-margin optical networks," IEEE/OSA J. Opt. Commun. Netw. **9**, A9–A17 (2017).
17. S. J. Savory, R. J. Vincent, and D. J. Ives, "Design considerations for low-margin elastic optical networks in the nonlinear regime," IEEE/OSA J. Opt. Commun. Netw. **11**, C76–C85 (2019).
18. I. Sartzetakis, K. Christodoulopoulos, C. P. Tsekrekos, D. Syvridis, and E. Varvarigos, "Quality of transmission estimation in WDM and elastic optical networks accounting for space–spectrum dependencies," IEEE/OSA J. Opt. Commun. Netw. **8**, 676–688 (2016).
19. M. Bouda, S. Oda, O. Vassilieva, M. Miyabe, S. Yoshida, T. Katagiri, Y. Aoki, T. Hoshida, and T. Ikeuchi, "Accurate prediction of quality of transmission based on a dynamically configurable optical impairment model," IEEE/OSA J. Opt. Commun. Netw. **10**, A102–A109 (2018).
20. U. Moura, M. Garrich, H. Carvalho, M. Svolenski, A. Andrade, A. C. Cesar, J. Oliveira, and E. Conforti, "Cognitive methodology for optical amplifier gain adjustment in dynamic DWDM networks," J. Light. Technol. **34**, 1971–1979 (2016).
21. Y. Huang, C. L. Gutterman, P. Samadi, P. B. Cho, W. Samoud, C. Ware, M. Lourdiane, G. Zussman, and K. Bergman, "Dynamic mitigation of EDFA power excursions with machine learning," Opt. Express **25**, 2245–2258 (2017).
22. W. Mo, C. L. Gutterman, Y. Li, S. Zhu, G. Zussman, and D. C. Kilper, "Deep-neural-network-based wavelength selection and switching in ROADM systems," IEEE/OSA J. Opt. Commun. Netw. **10**, D1–D11 (2018).
23. P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "The GN-model of fiber non-linear propagation and its applications," J. Light. Technol. **32**, 694–721 (2014).
24. R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Properties of nonlinear noise in long, dispersion-uncompensated fiber links," Opt. Express **21**, 25685–25699 (2013).
25. P. Serena and A. Bononi, "A time-domain extended Gaussian Noise model," J. Light. Technol. **33**, 1459–1472 (2015).
26. Q. Fan, J. Lu, G. Zhou, D. Zeng, C. Guo, L. Lu, J. Li, C. Xie, C. Lu, F. N. Khan, and A. Pak Tao Lau, "Experimental comparisons between machine learning and analytical models for QoT estimations in WDM systems," in *Proc. IEEE/OSA Optical Fiber Communication Conference (OFC),* (San Diego, CA, USA, 2020), p. M2J.2.
27. A. Ferrari, M. Filer, K. Balasubramanian, Y. Yin, E. Le Rouzic, J. Kundrát, G. Grammel, G. Galimberti, and V. Curri, "GNPy: an open source application for physical layer aware open optical networks," IEEE/OSA J. Opt. Commun. Netw. **12**, C31–C40 (2020).
28. A. Ferrari, K. Balasubramanian, M. Filer, J. Kundrát, Y. Yin, G. Grammel, G. Galimberti, E. Le Rouzic, and V. Curri, "Softwarized optical transport QoT in production optical network: a brownfield validation," in *Proc. European Conference on Optical Communication (ECOC),* (2020).
29. D. Semrau, R. I. Killey, and P. Bayvel, "The Gaussian Noise model in the presence of Inter-channel Stimulated Raman Scattering," J. Light. Technol. **36**, 3046–3055 (2018).
30. E. Seve, J. Pesic, C. Delezoide, A. Giorgetti, A. Sgambelluri, N. Sambo, S. Bigo, and Y. Pointurier, "Automated fiber type identification in SDN-enabled optical networks," J. Light. Technol. **37**, 1724–1731 (2019).
31. C. Wang, S. Fu, Z. Xiao, M. Tang, and D. Liu, "Long Short-Term Memory neural network (LSTM-NN) enabled accurate Optical Signal-to-Noise Ratio (OSNR) monitoring," J. Light. Technol. **37**, 4140–4146 (2019).
32. C. Rottondi, L. Barletta, A. Giusti, and M. Tornatore, "Machine-learning method for quality of transmission prediction of unestablished lightpaths," IEEE/OSA J. Opt. Commun. Netw. **10**, A286–A297 (2018).
33. M. Lonardi, J. Pesic, T. Zami, and N. Rossi, "The perks of using machine learning for QoT estimation with uncertain network parameters," in *Proc. OSA Advanced Photonics Congress,* (2020).
34. J. A. Jargon, X. Wu, H. Y. Choi, Y. C. Chung, and A. E. Willner, "Optical performance monitoring of QPSK data channels by use of neural networks trained with parameters derived from asynchronous constellation diagrams," Opt. Express **18**, 4931–4938 (2010).
35. T. Shun Rong Shen, K. Meng, A. Pak Tao Lau, and Z. Yang Dong, "Optical performance monitoring using artificial neural network trained with asynchronous amplitude histograms," Photonics Technol. Lett. **22**, 1665–1667 (2010).

36. F. N. Khan, K. Zhong, X. Zhou, W. H. Al-Arashi, C. Yu, C. Lu, and A. Pak Tao Lau, "Joint OSNR monitoring and modulation format identification in digital coherent receivers using deep neural networks," Opt. Express **25**, 17767–17776 (2017).
37. A. S. Kashi, Q. Zhuge, J. Cartledge, S. A. Etemad, A. Borowiec, D. W. Charlton, C. Laperle, and M. O'Sullivan, "Nonlinear signal-to-noise ratio estimation in coherent optical fiber transmission systems using artificial neural networks," J. Light. Technol. **36**, 5424–5431 (2018).
38. T. Tanimura, T. Hoshida, T. Kato, and S. Watanabe, "OSNR estimation providing self-confidence level as auxiliary output from neural networks," J. Light. Technol. **37**, 1717–1723 (2019).
39. H. J. Cho, S. Varughese, D. Lippiatt, R. Desalvo, S. Tibuleac, and S. E. Ralph, "Optical performance monitoring using digital coherent receivers and convolutional neural networks," Opt. Express **28**, 32087–32104 (2020).
40. C. Wang, S. Fu, H. Wu, M. Luo, X. Li, M. Tang, and D. Liu, "Joint OSNR and CD monitoring in digital coherent receiver using long short-term memory neural network," Opt. Express **27**, 6936–6945 (2019).
41. J. Thrane, J. Wass, M. Piels, J. C. M. Diniz, R. Jones, and D. Zibar, "Machine learning techniques for optical performance monitoring from directly detected PDM-QAM signals," J. Light. Technol. **35**, 868–875 (2017).
42. Y. Cheng, W. Zhang, S. Fu, M. Tang, and D. Liu, "Transfer learning simplified multi-task deep neural network for PDM-64QAM optical performance monitoring," Opt. Express **28**, 7607–7617 (2020).
43. F. J. V. Caballero, D. J. Ives, C. Laperle, D. Charlton, Q. Zhuge, M. O'Sullivan, and S. Savory, "Machine learning based linear and nonlinear noise estimation," IEEE/OSA J. Opt. Commun. Netw. **10**, D42–D51 (2018).
44. A. D. Shiner, M. E. Mousa-Pasandi, M. Qiu, M. A. Reimer, E. Young Park, M. Hubbard, Q. Zhuge, F. J. Vaquero Caballero, and M. O'Sullivan, "Neural network training for OSNR estimation from prototype to product," in *Proc. IEEE/OSA Optical Fiber Communication Conference (OFC),* (San Diego, CA, USA, 2020), p. M4E.2.
45. M. Lonardi, J. Pesic, P. Jennevé, P. Ramantanis, N. Rossi, A. Ghazisaeidi, and S. Bigo, "Optical nonlinearity monitoring and launch power optimization by artificial neural networks," J. Light. Technol. **38**, 2637–2645 (2020).
46. S. Zhu, C. L. Gutterman, W. Mo, Y. Li, G. Zussman, and D. C. Kilper, "Machine learning based prediction of erbium-doped fiber WDM line amplifier gain spectra," in *Proc. European Conference on Optical Communication (ECOC),* (Rome, Italy, 2018).
47. Y. You, Z. Jiang, and C. Janz, "Machine learning-based EDFA gain model," in *Proc. European Conference on Optical Communication (ECOC),* (Rome, Italy, 2018).
48. M. Freire, S. Mansfeld, D. Amar, F. Gillet, A. Lavignotte, and C. Lepers, "Predicting optical power excursions in erbium doped fiber amplifiers using neural networks," in *Proc. Asia Communications Conference (ACP),* (Hangzhou, China, 2018).
49. S. Zhu, C. Gutterman, A. D. Montiel, J. Yu, M. Ruffini, G. Zussman, and D. Kilper, "Hybrid machine learning EDFA model," in *Proc. IEEE/OSA Optical Fiber Communication Conference (OFC),* (San Diego, CA, USA, 2020), p. T4B.4.
50. M. Ionescu, "Machine learning for ultrawide bandwidth amplifier configuration," in *Proc. International Conference on Transparent Optical Networks (ICTON),* (Angers, France, 2019).
51. A. M. Rosa Brusin, U. C. De Moura, V. Curri, D. Zibar, and A. Carena, "Introducing load aware neural networks for accurate predictions of Raman amplifiers," J. Light. Technol. **38**, 6481–6491 (2020).
52. M. Bolshtyansky, "Spectral hole burning in erbium-doped fiber amplifiers," J. Light. Technol. **21**, 1032–1038 (2003).
53. J. Bromage, "Raman amplification for fiber communications systems," J. Light. Technol. **22**, 73–93 (2004).
54. D. Zibar, A. M. Rosa Brusin, U. C. De Moura, F. Da Ros, V. Curri, and A. Carena, "Inverse system design using machine learning: the Raman amplier case," J. Light. Technol. **38**, 736–753 (2020).
55. J. Wass, J. Thrane, M. Piels, R. Jones, and D. Zibar, "Gaussian process regression for WDM system performance prediction," in *Proc. IEEE/OSA Optical Fiber Communication Conference (OFC),* (Los Angeles, CA, USA, 2017), p. Tu3D.7.
56. J. Cho, S. Chandrasekhar, E. Sula, S. Olsson, E. Burrows, G. Raybon, R. Ryf, N. Fontaine, J.-C. Antona, S. Grubb, P. Winzer, and A. Chraplyvy, "Supply-power-constrained cable capacity maximization using multilayer neural networks," J. Light. Technol. **38**, 3652–3662 (2020).
57. A. D'Amico, S. Straullu, A. Nespola, I. Khan, E. London, E. Virgillito, S. Piciaccia, A. Tanzi, G. Galimberti, and V. Curri, "Using machine learning in an open optical line system controller," IEEE/OSA J. Opt. Commun. Netw. **12**, D84–D99 (2020).
58. L. Xia, J. Zhang, S. Hu, M. Zhu, Y. Song, and K. Qiu, "Transfer learning assisted deep neural network for OSNR estimation," Opt. Express **27**, 19398–19406 (2019).
59. J. Yu, W. Mo, Y.-K. Huang, E. Ip, and D. C. Kilper, "Model transfer of QoT prediction in optical networks based on artificial neural networks," IEEE/OSA J. Opt. Commun. Netw. **11**, C48–C57 (2019).
60. J. Mata, I. de Miguel, R. J. Durán, J. C. Aguado, N. Merayo, L. Ruiz, P. Fernández, R. M. Lorenzo, and E. J. Abril, "A SVM approach for lightpath QoT estimation in optical transport networks," in *Proc. IEEE BigData,* (Boston, MA, USA, 2017), pp. 4795–4797.
61. T. Panayiotou, S. P. Chatzis, and G. Ellinas, "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**, 98–108 (2017).
62. D. Azzimonti, C. Rottondi, and M. Tornatore, "Reducing probes for quality of transmission estimation in optical networks with active learning," IEEE/OSA J. Opt. Commun. Netw. **12**, A38–A48 (2020).
63. S. Aladin, A. V. S. Tran, S. Allogba, and C. Tremblay, "Quality of transmission estimation and short-term performance forecast of lightpaths," J. Light. Technol. **38**, 2807–2814 (2020).
64. R. M. Morais and J. Pedro, "Machine learning models for estimating quality of transmission in DWDM networks," IEEE/OSA J. Opt. Commun. Netw. **10**, D84–D99 (2018).
65. C. Rottondi, R. di Marino, M. Nava, A. Giusti, and A. Bianco, "On the benefits of domain adaptation techniques for quality of transmission estimation in optical networks," IEEE/OSA J. Opt. Commun. Netw. **13**, A34–A43 (2021).
66. A. Diaz-Montiel, S. Aladin, C. Tremblay, and M. Ruffini, "Active wavelength load as a feature for QoT estimation based on support vector machine," in *Proc. IEEE International Conference on Communications (ICC),* (San Diego, CA, USA, 2019).
67. T. Panayiotou, G. Savva, B. Shariati, I. Tomkos, and G. Ellinas, "Machine learning for QoT estimation of unseen optical network states," in *Proc. IEEE/OSA Optical Fiber Communication Conference (OFC),* (San Diego, CA, USA, 2019), p. Tu2E.2.
68. T. Panayiotou, G. Savva, I. Tomkos, and G. Ellinas, "Decentralizing machine-learning-based QoT estimation for sliceable optical networks," IEEE/OSA J. Opt. Commun. Netw. **12**, 146–162 (2020).
69. M. Salani, C. Rottondi, and M. Tornatore, "Routing and spectrum assignment integrating machine-learning-based QoT estimation in elastic optical networks," in *Proc. IEEE INFOCOM,* (Paris, France, 2019), pp. 1738–1746.
70. D. Azzimonti, C. Rottondi, A. Giusti, M. Tornatore, and A. Bianco, "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**, A56–A66 (2021).
71. P. Samadi, D. Amar, C. Lepers, M. Lourdiane, and K. Bergman, "Quality of transmission prediction with machine learning for dynamic operation of optical WDM networks," in *Proc. European Conference on Optical Communication (ECOC),* (Gothenburg, Sweden, 2017).
72. E. Seve, J. Pesic, C. Delezoide, S. Bigo, and Y. Pointurier, "Learning process for reducing uncertainties on network parameters and design margins," IEEE/OSA J. Opt. Commun. Netw. **10**, A298–A306 (2018).
73. R. Proietti, X. Chen, K. Zhang, G. Liu, M. Shamsabardeh, A. Castro, L. Velasco, Z. Zhu, and S. J. B. Yoo, "Experimental demonstration of machine-learning-aided QoT estimation in multi-domain elastic optical networks with alien wavelengths," IEEE/OSA J. Opt. Commun. Netw. **11**, A1–A10 (2019).

74. I. Sartzetakis, K. Christodoulopoulos, and E. Varvarigos, "Accurate quality of transmission estimation with machine learning," IEEE/OSA J. Opt. Commun. Netw. **11**, 140–150 (2019).
75. Q. Zhuge, X. Zeng, H. Lun, M. Cai, X. Liu, L. Yi, and W. Hu, "Application of machine learning in fiber nonlinearity modeling and monitoring for elastic optical networks," J. Light. Technol. **37**, 3055–3063 (2019).
76. R. Morais, "Fast and high-precision optical performance evaluation for cognitive optical networks," in *Proc. IEEE/OSA Optical Fiber Communication Conference (OFC),* (San Diego, CA, USA, 2020), p. Th3D.3.
77. J. Pesic, M. Lonardi, N. Rossi, T. Zami, E. Seve, and Y. Pointurier, "How uncertainty on the fiber span lengths influences QoT estimation using machine learning in WDM networks," in *Proc. IEEE/OSA Optical Fiber Communication Conference (OFC),* (San Diego, CA, USA, 2020), p. Th3D.5.
78. Z. Gao, S. Yan, J. Zhang, M. Mascarenhas, R. Nejabati, Y. Ji, and D. Simeonidou, "ANN-based multi-channel QoT-prediction over a 563.4-km field-trial testbed," J. Light. Technol. **38**, 2646–2655 (2019).
79. A. Mahajan, K. Christodoulopoulos, R. Martínez, S. Spadaro, and R. Muñoz, "Modeling EDFA gain ripple and filter penalties with machine learning for accurate QoT estimation," J. Light. Technol. **38**, 2616–2629 (2020).
80. E. Seve, J. Pesic, and Y. Pointurier, "Associating machine-learning and analytical models for QoT estimation: Combining the best of both worlds," IEEE/OSA J. Opt. Commun. Netw. **13** (2021). To appear.
81. I. Khan, M. Bilal, M. Siddiqui, M. Khan, A. Ahmad, M. Shahzad, and V. Curri, "QoT estimation for light-path provisioning in un-seen optical networks using machine learning," in *Proc. International Conference on Transparent Optical Networks (ICTON),* (2020).
82. C.-Y. Liu, X. Chen, R. Proietti, and S. J. B. Yoo, "Evol-TL: Evolutionary transfer learning for QoT estimation in multi-domain networks," in *Proc. IEEE/OSA Optical Fiber Communication Conference (OFC),* (San Diego, CA, USA, 2020), p. Th3D.1.
83. J. Pesic, M. Lonardi, E. Seve, N. Rossi, and T. Zami, "Transfer learning from unbiased training data sets for QoT estimation in WDM networks," in *Proc. European Conference on Optical Communication (ECOC),* (2020).

AUTHOR BIOGRAPHY

Yvan Pointurier (S'02-M'06-SM'12) received a Ph.D. from the University of Virginia, USA in 2006. Between 2006 and 2009, he was a postdoctoral fellow at McGill University in Montreal and then a senior researcher at AIT, Greece. In 2009, Dr. Pointurier joined Alcatel-Lucent (now Nokia) Bell Labs as a research engineer and later became a research team leader. In 2020 he joined Huawei as a research team leader. Dr. Pointurier has authored or co-authored more than 15 European and US patents, and over 100 technical papers in leading journals, key conferences (OFC, ECOC, ACM Internet Measurement Conference, IEEE Infocom, ICC, Globecom), and book chapters. He received a best paper award at the IEEE ICC conference in 2006 and an IEEE Communication Letters Exemplary Reviewer award in 2014 and 2016 (top 3% of the reviewers).

Table 2. Comparison of machine learning-based Quality of Transmission estimation frameworks for amplifiers

Paper	Use case	Data source	Learning type	Algorithm	Type of study	Inputs (features)	Output	Key results
[46]	Single EDFA	Green, brown	Data scientist	DNN	Exp.	Per channel input power	Optical power spectrum	RMSE $<$ 0.2 dB
[47]	Single EDFA	Green, brown	Data scientist	DNN	Exp.	Per channel input power	Gain	0.18 dB max. gain estimation error
[48]	Single EDFA	Green, brown	Data scientist	ANN	Sim.	Per channel input status (on/off)	Optical power spectrum	Error $<$ 0.2 dB in 90% of the cases
[49]	Single EDFA	Green, brown	Hybrid	DNN	Exp.	Per channel input power; per channel output power from analytic model	Optical power spectrum	RMSE $<$ 0.2 dB; faster convergence with inputs from analytic model
[50]	Single Raman / EDFA link	Green, brown	Data scientist	DNN	Exp.	Amplifiers' pump powers (which adjust gain and tilt)	Optical power spectrum	Typical error in output power $<$ 0.2 dB (max.: 0.3 dB)
[51]	Single Raman link	Green, brown	Data scientist	ANN	Sim.	Per subband input power, pump powers	Gain and ASE noise prediction	Max error $<$ 0.5 dB on gain and noise power

Table 3. Comparison of machine learning-based Quality of Transmission estimation frameworks for lines: regression

Paper	Use case	Data source	Learning type	Algorithm	Type of study	Inputs (features)	Output	Key results
[55]	Line	Green, brown	Data scientist	GPR	Sim. / Exp.	TX power, number of spans, baudrate, interchannel spacing	BER / Q	Q error $<$ 0.3
[56]	Line	Brown	Data scientist	DNN	Exp.	Per-channel received signal and noise	OSNR	0.5 dB max. OSNR error
[57]	Line	Green, brown	Data scientist	ANN	Exp.	End of line spectrum	OSNR	Error on OSNR $<$ 0.2 dB
[58]	Line		Data scientist	DNN, transfer learning	Exp.	Amplitude histogram of received samples	OSNR	Retraining test size divided by 5, training time divided by 10
[59]	Line		Data scientist	DNN, transfer learning	Exp.	Q from other lines and new line	Q	Retraining test size divided by 50, training time divided by 4

Table 4. Comparison of machine learning-based Quality of Transmission estimation frameworks for networks: classification

Paper	Use case	Data source	Learning type	Algorithm	Type of study	Inputs (features)	Output	Key results
[60]	Network	Brown	Data scientist	SVM	Sim.	Light path length, link lengths, wavelength, statistics on co-propagating light paths	BER (2 classes)	$>$ 99.9% correct decisions
[61]	Network	Brown	Data scientist	ANN	Sim.	Path length, number of EDFAs, max. link length, destination node degree, wavelength	BER (2 classes)	Accuracy $>$ 90%
[32]	Network	Brown	Data scientist	KNN, Random Forest	Sim.	Total length, number of links, maximum link length, demand capacity, modulation format, guardband, modulation format, and traffic volume of nearest left and right neighbor	BER (2 classes)	Typical accuracy $>$ 95%
[62]	Network	Brown	Data scientist	GPR, active learning	Sim.	Total length, number of links, maximum link length, demand capacity, modulation format	BER (2 classes)	Accuracy $>$ 99% by complementing a small data set of existing services with only a few well-chosen light paths used as probes
[63]	Network	Brown	Data scientist	SVM, ANN	Sim.	Total link length, span length, channel launch power, modulation format, data rate	SNR (2 classes)	$>$ 99% correct decisions with ANN and SVM, ANN 10x faster (sub-ms)
[64]	Network	Brown	Data scientist	KNN, logistic regression, SVM (classification), ANN (classification, regression)	Sim.	Number of hops and spans, light path length, average and maximum link length, average span attenuation, average dispersion, modulation format	SNR (continuous or 2 classes)	Typical classification accuracy $>$ 95%, average SNR error $<$ 0.4 dB
[66]	Network	Brown	Data scientist	SVM	Sim.	Number of ROADMs, of links, of fiber spans, Length of fiber span, launch channel power, EDFA Pre- and EDFA Post-amplifier gain, wavelength	OSNR (4 classes)	$>$ 95% correct decisions
[65]	Network		Data scientist	SVM, transfer learning	Sim.	Total length, number of links, maximum link length, demand capacity, modulation format	BER (2 classes)	Retraining test size divided by 20

Table 5. Comparison of machine learning-based Quality of Transmission estimation frameworks for networks: regression

Paper	Use case	Data source	Learning type	Algorithm	Type of study	Inputs (features)	Output	Key results
[71]	Network	Brown	Data scientist	ANN	Exp.	Not explicit	OSNR	Microsecond estimation, RMSE $<$ 0.2dB in 90% of cases

[19]	Network	Brown	Physicist	Filtering	Exp.	Pre-FEC BER	Q	0.6 dB max. estimation error on Q, millisecond estimation
[72]	Network	Brown	Physicist	Gradient descent	Sim.	Pre-FEC BER, fiber input power, NF	SNR	Typical error $<$ 0.1 dB
[73]	Network	Brown	Data scientist	DNN	Exp.	Channel, noise power on each link	OSNR	$<$ 5 dB OSNR error over range 20-40 dB
[74]	Network	Brown	Physicist; data scientist	Least squares	Sim.	Per span: fiber attenuation, dispersion, nonlinear coefficient (physicist); Number of EDFAs and hops, total path length, baud rate, "load metric" (data scientist)	SNR	SNR RMSE down to $<$ 0.1 dB
[75]	Network	Brown	Hybrid	ANN	Sim.	Gaussian Noise model output, span number, maximum and average lengths, launch power, link length, residual chromatic dispersion, average fiber nonlinear coefficient and attenuation, number of channels	Nonlinear SNR	Error <2 dB in 95% of the cases
[76]	Network	Brown	Hybrid	ANN	Sim.	Fiber attenuation, dispersion coefficient, effective area, and non-linear refractive index; span length, number of active channels, launch power, channel bandwidth and frequency, output of analytic modeling tool	Nonlinear SNR	$>$ 99.9% estimates within 0.5 dB, $<$ 10 ms computation
[77]	Network	Brown	Data scientist	ANN	Sim.	List of span lengths	SNR	RMSE $<$ 0.2 dB
[78]	Network	Brown	Data scientist	DNN	Exp.	Launch power, laser bias, EDFAs input/output powers	Q	RMSE $<$ 0.02dB with 600 training points
[79]	Network	Brown	Physicist	SVM regression	Sim.	Power per channel at each node, fine (sub-GHz) optical spectrum every few spans, pre-FEC BER	SNR	Typical error <0.2 dB
[80]	Network	Brown	Physicist	Gradient descent	Sim.	Pre-FEC BER, spectrum at gain equalizer	SNR	Typical error <0.1 dB
[81]	Network	Both	Hybrid	ANN	Sim.	Per channel power and frequency; number of spans; from physical model: ASE and nonlinear noises	SNR	Max. error $<$ 0.5 dB
[82]	Network		Data scientist	DNN, evolutionary transfer learning	Exp.	OSA power profile at each WSS	Q	Retraining test size divided by 10
[83]	Network		Data scientist	ANN	Sim.	List of span lengths	SNR	RMSE in dB divided by 2