

Using machine learning in an open optical line system controller

Original

Using machine learning in an open optical line system controller / D'Amico, A., Curri, V., Straullu, S., Nespola, A., Khan, I., London, E., Virgillito, E., Piciaccia, S., Tanzi, A., Galimberti, G.. - In: JOURNAL OF OPTICAL COMMUNICATIONS AND NETWORKING. - ISSN 1943-0620. - 12:6(2020), pp. C1-C11. [10.1364/JOCN.382557]

Availability:

This version is available at: 11583/2807512 since: 2020-03-30T19:53:54Z

Publisher:

Institute of Electrical and Electronics Engineers Inc.

Published

DOI:10.1364/JOCN.382557

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

Optica Publishing Group (formely OSA) postprint/Author's Accepted Manuscript

“© 2020 Optica Publishing Group. One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modifications of the content of this paper are prohibited.”

(Article begins on next page)

Using machine learning in an open optical line system controller

ANDREA D'AMICO,^{1,*} STEFANO STRAULLU,² ANTONINO NESPOLA,² IHTESHAM KHAN,¹ ELLIOT LONDON,¹ EMANUELE VIRGILLITO,¹ STEFANO PICIACCIA,³ ABERTO TANZI,³ GABRIELE GALIMBERTI,³ AND VITTORIO CURRI¹

¹Department of Electronics and Telecommunications, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

²LINKS Foundation, Via Pier Carlo Boggio 61, 10138 Torino, Italy

³Cisco Photonics, Via S. M. Molgora 48/C, 20871 Vimercate, Italy

*Corresponding author: andrea.damico@polito.it

Received 5 November 2019; revised 10 January 2020; accepted 10 January 2020; published 14 February 2020 (Doc. ID 382557)

The reduction of system margin in open optical line systems (OLSs) requires the capability to predict the quality of transmission (QoT) within them. This quantity is given by the generalized signal-to-noise ratio (GSNR), including both the effects of amplified spontaneous emission (ASE) noise and nonlinear interference accumulation. Among these, estimating the ASE noise is the most challenging task due to the spectrally resolved working point of the erbium-doped fiber amplifiers (EDFAs), which depend on the spectral load, given the overall gain profile. An accurate GSNR estimation enables control of the power optimization and the possibility to automatically deploy lightpaths with a minimum margin in a reliable manner. We suppose an agnostic operation of the OLS, meaning that the EDFAs are operated as black boxes and rely only on telemetry data from the optical channel monitor at the end of the OLS. We acquire an experimental data set from an OLS made of 11 EDFAs and show that, without any knowledge of the system characteristics, an average extra margin of 2.28 dB is necessary to maintain a conservative threshold of QoT. Following this, we applied deep neural network machine-learning techniques, demonstrating a reduction in the needed margin average down to 0.15 dB. © 2020 Optical Society of America

<https://doi.org/10.1364/JOCN.382557>

Provided under the terms of the [OSA Open Access Publishing Agreement](#)

1. INTRODUCTION

Data traffic demand will experience a dramatic increase over the next few years, driven by the implementation of 5G access and the expansion of bandwidth-hungry applications, such as high definition video and virtual- and augmented-reality content [1]. These applications will boost cloud computing and cloud-storage-related data exchange, causing traffic expansion both within and between data centers. Optical networks will sustain this growth trend, particularly within their backbone portion. These backbone networks already carry massive amounts of data, and a further push will be required to match the required transmission capacity over the next five years. A key operator request is the ability to fully exploit existing infrastructure in order to maximize returns from investments. This need is directly related to the capability of orchestrating all network layers, allowing the data transport to reach the maximum available capacity [2–6]. In optical networks, the enabler for optimal exploitation of data transport—the dense wavelength division multiplexed (DWDM) transmission—is the control layer. In particular, software-defined network

controllers rely on a network abstraction. Nowadays, optical networks are fast moving toward partial disaggregation, with a final goal of full disaggregation; a disaggregated network has subsystems that are managed independently from one another by relying on common data structures and API (application program interface). Contrary to aggregated networks, disaggregated networks can be open and multivendor but are not able to have closed management. These features pave the road for a software-defined controller that is able to manage separately the working points of the various network elements, enabling the management to be user-customizable.

The first step in disaggregating the network is to consider the optical line systems (OLSs) that connect the network nodes. In this framework, the quality of transmission (QoT) degradation depends on the capability of OLS controllers to operate at the optimal working point [7,8]. The more accurately this demand is reached, the lower the margin for traffic deployment and, thus, the larger the deployed traffic rate. Moreover, there is the potential for the recovery of network failures to be automated, reducing downtime. Therefore, to reduce the margin, it is

mandatory to rely on a QoT estimator (QoT-E) that is able to reliably predict lightpath (LP) performance before its actual deployment, i.e., the generalized signal-to-noise ratio (GSNR), that includes the effects of amplified spontaneous emission (ASE) noise and provides both the optical SNR (OSNR) and nonlinear interference (NLI) accumulation [9]. The interaction between ASE noise and NLI [10,11] occurs in the case of very low operational GSNR, namely for extremely long OLSs, which require several amplification points. These conditions are verified in submarine point-to-point networks but have negligible effects within terrestrial networks. In this work we focus on terrestrial regional and notional backbone networks for which transparent propagation is over much smaller distances, meaning that considerable ASE-NLI interactions are not produced. Among the ASE noise and NLI contributions, the former is the most dominant, because it is twice the NLI when the system operates at optimal power [7,12]. Remarkably, it is also the most challenging to estimate. In fact, the ASE noise magnitude depends on the working point of erbium-doped fiber amplifiers (EDFAs) [13]; this in turn depends on the spectral load [14]. On the contrary, the NLI can be accurately predicted when the ASE noise accumulation is well characterized [15].

The purpose of this work is to investigate the reduction of uncertainty in the OSNR prediction and, consequently, to enable the network controller to reliably deploy the LP at the minimum margin. In this work we suppose the worst case of a completely agnostic scenario, by relying only on data coming from the optical channel monitor (OCM) available at the end of the line system. The uncertainty on the working point of the EDFAs is typically induced by a mixed effect of physical phenomena [14] and implementation issues, meaning that an analytic approach is almost impossible to achieve in an open environment. To counteract this, we opted to use machine-learning (ML) techniques, a tactic that has already been effectively tested when managing optical networks; see [16–19] for performance monitoring applications, [20,21] for prediction estimation of the ML approach, and [22] for both. An overall survey of ML applied in optical networks can be found in [23].

Specifically, we also cite [24–27]. In [24], the authors utilize ML to predict the gain of a single EDFA and show that this method can provide improvements over an analytical model. In [25] ML is used to predict the output of an EDFA cascade; in particular, wavelength assignment over a specific network considered in its entirety is able to be automated. Reference [26] investigates how ML can mitigate the effect of the EDFA gain ripple on QoT-E within a simulated network and [27] demonstrates how ML may be used to automatically configure the gain required by amplifiers after deployment. The main difference between this previous research and the present work is that we focus on the OSNR response to specific configurations in a particular OLS that is considered as an element of a completely disaggregated network. Through this, we obtain an evaluation that can be combined with a nonlinear SNR prediction, in order to obtain a reliable QoT-E that can be used both in network planning and for the wavelength assignment in the online case.

In Section 2, we first address the issues related to the abstraction of the physical layer in order to effectively perform a multilayer optimization. In particular, we argue that an accurate QoT-E has a key role in minimizing the margin.

In Section 3, we describe the experiments performed to emulate an open OLS composed of 11 cascaded amplifiers and one booster amplifier. With this setup we have obtained a data set of measurements mimicking the power readings from an OCM, where different spectral loads have been generated by shaped ASE noise. Additionally, the EDFAs are used as black boxes, setting the average gain to the nominal level.

In Section 4, we statistically analyze the experimentally measured data set over all investigated bandwidths. Then, we present the variation in OSNR with respect to the spectral load configuration and discuss these fluctuations in light of physical considerations. Consequently, we derive the required margins, supposing a total absence of knowledge on the EDFA gain and of the noise figure per wavelength. These results show that the uncertainty induced by an agnostic use of the OLS may require the deployment of 2.28 dB of system margin, on average. Note that a closed OLS based on single-vendor equipment may largely reduce this uncertainty by characterizing the parameters of these devices. Nevertheless, aging and environmental effects may introduce some uncertainty even in this case.

In Section 5 we tested ML techniques. Here, we suppose that a training data set acquired before the deployment of real traffic has been collected in order to reduce the uncertainty of the estimated OSNR. We did not aim to develop a specific ML algorithm from scratch and instead aimed to show the effectiveness of ML in this scenario. For this reason, we relied upon the TensorFlow open source library [28]. We show that by utilizing and optimizing deep neural network (DNN) algorithms, we are able to reduce the required average margin on the OSNR prediction from the initial value of 2.28 dB down to 0.15 dB.

In Section 6, we give some overall comments that address possible further investigations.

2. PHYSICAL LAYER ABSTRACTION AND OPTIMIZATION IN TRANSPARENT OPTICAL NETWORKS

From a data transport point of view, an optical network is an infrastructure connecting site—in general with a meshed topology—where traffic is added/dropped or routed (see Fig. 1). Site-to-site links are bidirectional fiber connections implemented as one or more fiber pairs, with one fiber for each direction, that are periodically amplified by lumped and/or distributed amplification techniques: EDFAs optionally assisted by some distributed Raman amplification. These links are commonly defined as an OLS and are managed by a controller that has properly set the working point of the amplifiers and, consequently, the power spectral density (spectral load) at the input of each fiber span. State-of-the-art optical networks rely on coherent technology for optical transmission; routing operations are done at the optical transport layer thanks to reconfigurable optical add/drop multiplexers (ROADMs) that implement the transparency paradigm. The spectral usage of

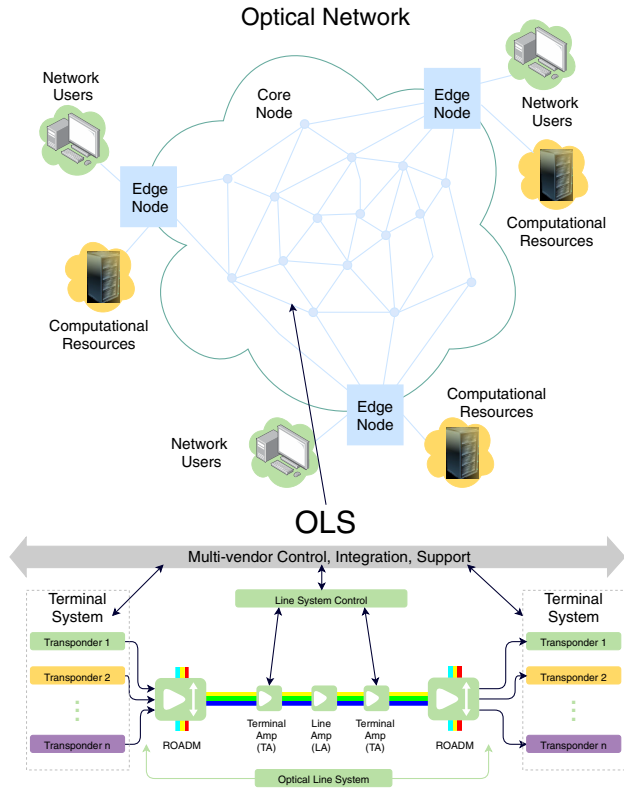


Fig. 1. Schematic description of an optical network as a topology of ROADM nodes connected by OLSs. The inset shows a general setup for an OLS that in this case is supposed to be open.

fiber propagation exploits DWDM to enable multichannel transmission over the C-band and, in the future, over multi-band systems, starting from the L-band. The DWDM spectral grid can be either fixed or flexible, according to the ITU-T recommendations [29] that define the spectral slots enabling transparent source-to-destination optical transport. Within this grid, the LPs are defined as the circuits describing the routing space, i.e., the set of possible connections that the routing wavelength assignment may rely on to set traffic transport (the LP deployment). Over a deployed LP, a polarization-division-multiplexed multilevel modulation format propagates transparently from source to destination, suffering from propagation impairments; this is summarized as ASE noise added by the amplifiers, fiber propagation effects, and ROADM filtering effects. It has been extensively demonstrated that the fiber propagation on an uncompensated OLS impairs the QoT of LPs operated with coherent technologies by introducing some amount of phase and amplitude noise [7,30–32]. This phase noise is typically well compensated by the carrier phase estimator module of the digital signal processing at the receiver. This disturbance must be considered only for high symbol rate transmission that is designed for short-reach, high-capacity transparent optical transmission or in the case of probabilistic shaping [32]. The amplitude noise that derives from fiber propagation, commonly defined as the NLI, always impairs performance as it is a Gaussian disturbance that sums with the ASE noise at the receiver. Additionally, the filtering

effects of ROADMs impact QoT degradation as an extra loss contribution.

A. Quality of Transmission Estimation Based on the GSNR

It is well accepted that the merit of QoT for deployed LPs is given by the GSNR, including both the effects of the accumulated ASE noise and NLI disturbance, defined as

$$\text{GSNR} = \frac{P_{\text{Rx}}}{P_{\text{ASE}} + P_{\text{NLI}}} = (\text{OSNR}^{-1} + \text{SNR}_{\text{NLI}}^{-1})^{-1}, \quad (1)$$

where $\text{OSNR} = P_{\text{Rx}}/P_{\text{ASE}}$, $\text{SNR}_{\text{NLI}} = P_{\text{Rx}}/P_{\text{NLI}}$, P_{Rx} is the power of the channel under test (CUT) at the receiver, P_{ASE} is the power of the ASE noise, and P_{NLI} is the power of the NLI. In particular, given the bit-error ratio (BER) versus the OSNR back-to-back characterization of the transceiver, the GSNR accurately predicts the BER, as has been extensively shown in multivendor experiments using commercial products [9]. P_{NLI} is generated by nonlinear effects and depends on the power of the CUT and on the spectral load with a cubic law [7]. This means that for each OLS there exists an optimal spectral load that maximizes the GSNR [8].

Given the cascade of N optical domains, each characterized by a generalized GSNR_i , where $i = 1, \dots, N$, it is straightforward to demonstrate that the overall QoT is given by the following expression:

$$\text{GSNR} = \left(\sum_{i=1}^N \frac{1}{\text{GSNR}_i} \right)^{-1}. \quad (2)$$

If we analyze the propagation effects on a given LP over a network route, we can abstract it as a cascade of the effects of each optical domain that introduces QoT impairments. Therefore, besides the effects of ROADMs, each LP experiences the cumulative impairments of all previously passed OLSs, where each introduces some amount of ASE noise and NLI. For QoT purposes, the OLS can be abstracted by a unique parameter commonly defined as SNR degradation that, in general, is frequency resolved ($\text{GSNR}_i(f)$), if the OLS controllers are able to keep the OLS operating at the optimal working point. Hence, with this condition, if the OLS controllers are able to expose the corresponding GSNR_i for QoT operations, a network can be abstracted as a weighted graph corresponding to its topology. The graph nodes are ROADM network nodes, while the edges are the OLSs and the weights on these edges are the $\text{GSNR}_i(f)$ degradations of the corresponding OLSs, as shown in Fig. 2. In particular, for a LP routed from A to F that passes through C and E, the QoT is

$$\begin{aligned} \text{GSNR}_{\text{AF}}^{-1}(f) &= \text{GSNR}_{\text{AC}}^{-1}(f) + \text{GSNR}_{\text{CE}}^{-1}(f) \\ &+ \text{GSNR}_{\text{EF}}^{-1}(f). \end{aligned}$$

Note that the network abstraction of the physical layer may be enriched with additional information, such as the latency or the accumulated chromatic dispersion. Both of these additional quantities sum on routes as the SNR degradation and are not exploited in this work. Once the network abstraction

availability of a device that is able to supply the OLS with various spectral load configurations and measure the OLS response in terms of OSNR. As OLSs are typically bidirectional, it is conceivable that a two-port portable device operating as an ASE-shaped generator at the output port and an accurate OCM at the input port can be used to retrieve these data. Moreover, a future implementation considers the possibility of these devices being built into the ROADMs nodes, allowing the data to be collected with periodical updates via streaming. Utilizing this data set enables a QoT-E based on the OSNR response to specific spectral load configurations, increasing the accuracy of OSNR predictions with respect to option (2), where only telemetry data is considered. Additionally, this approach does not require knowledge of the physical parameters of the OLS. This case provides an ideal scenario to apply ML, where the OLS is treated as a black box. In fact, a ML method using a training data set composed of past spectral load realizations can yield an accurate prediction for every newly generated spectral load realization.

In this work, we focus on option (3) and consider a realistic use case, namely, a scenario where the OLS controller wishes to allocate a new LP over the CUT, given an existing spectral load. In particular, we investigate the level of OSNR associated with this new LP.

3. EXPERIMENTAL SETUP

To obtain an experimental data set, we design and implement the experimental setup depicted in Fig. 4, based on commercial EDFAs [37] used as black boxes. Span losses are obtained by attenuators in order to focus only the OSNR and to avoid any NLI generation. The channel combs that provide the OLS spectral load have been obtained by shaping ASE noise. This approach does not limit the generality of the results because of the large time constant that characterizes the physical effects within EDFAs. The output of the ASE noise source is shaped by means of a programmable optical wave-shaper filter (Finisar 1000 S) to generate a 100 GHz-spaced, 35-channel WDM comb centered at 193.5 THz, amplified by a booster amplifier (EDFA₀ in Fig. 4). The choice of the 100 GHz spacing was forced by the hardware availability, as well as the overall frequency domain under investigation, which was limited to 3.5 THz (35 channels, each with 100 GHz spacing). These restrictions do not limit the generality of the results, as the OSNR values do not change appreciably within each channel bandwidth and all criticalities concerning the EDFA amplification process are properly captured. The optical line is composed of 11 spans, each made of a VOA, with the optical span attenuation set to 10 dB, each followed by an EDFA that operates at a constant output power of -10 dBm per channel. For the EDFAs, MATLAB control software has been developed to enable black box control. The OCM at the end of the OLS is mimicked by an optical spectrum analyzer (OSA). OCMs that are currently present in ROADMs nodes are not able to capture the noise floor due to their lack of sensitivity. As mentioned in option (3) within the previous section, for a real application scenario we suppose the presence of a specific device that is able to measure both the channel powers and the noise floor, or, to update the current OCM presence on the ROADMs nodes.

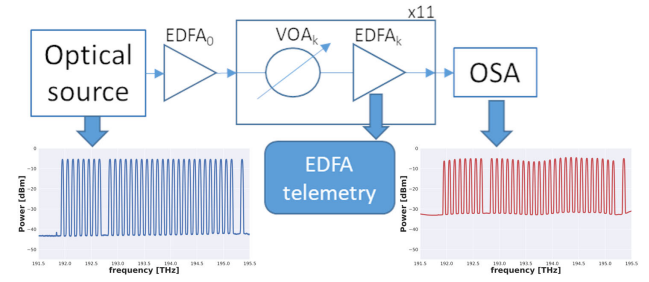


Fig. 4. Experimental setup: Here, the OLS under investigation is composed of an initial booster amplifier and a cascade of 11 spans, each containing a VOA and an EDFA. We show the input and output spectral power measurements obtained using an optical spectrum analyzer in blue and red, respectively.

Regarding the technical aspects of the data collection within this project, the experimental campaign lasted several days due to the OSA usage, taking significantly longer than an OCM. We expect that within a real application scenario the data collection process would last the duration of a single night before the in-service operation of the OLS, producing the required amount of data needed for training the ML.

For every spectral load, we measured the input and output spectrum in order to generate the final data set. Specifically, we measured the total power over each channel spectral bandwidth, i.e., the noise floor if the channel is *off*, or the channel power if the channel is *on*. In fact, since the channel bandwidth (32 GHz) is less than half of the channel spacing, we have been able to measure the noise floor even for the *on* channels, estimating their OSNR. An experimental data set has been generated with 4435 cases representing different spectral load configurations. For clarity, let us define N_{on} as the number of channels in the *on* state in a distinct configuration. Given this definition, the data set is composed of a scenario with all channels *on* ($N_{on} = 35$), the 35 cases where only one channel is *on* ($N_{on} = 1$), and 140 configurations for each $N_{on} = 2, \dots, 34$. This final set of configurations includes pairs of spectral loads that are identical, except for the CUT being either within the *on* or *off* state.

4. STATISTICAL ANALYSIS OF EXPERIMENTAL DATA

In this section, we statistically analyze the OSNR fluctuations produced by different spectral loads in order to obtain a quantitative estimation of the total OSNR uncertainty, given a static OLS (the OSNR values are calculated with a noise bandwidth of 12.5 GHz). Moreover, we use the experimental data set as outlined in option (3) in Section 2 to acquire a prediction of the OSNR responses. To summarize the data set characteristics, there are 4435 measurements of distinct spectral load configurations, which are a subset of the 2^{35} possibilities, given 35 channels. To populate the data set, we select a sample of spectral load configurations which is uniform over the number of channels in the *on* state. Moreover, for the set of configurations with the same N_{on} , the channels that are in the *on* state are chosen randomly, except for the CUT, which is equally divided between the *on* and *off* states. This specific data set selection

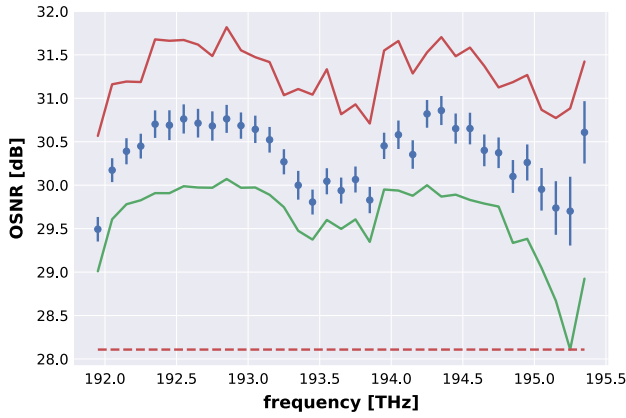


Fig. 5. Overall OSNR measurements in the frequency domain. The blue dots are the mean values over the entire sample for each channel; the error bars are equal to the standard deviations. In red and green the maximum and the minimum for each channel are outlined, respectively. The dashed red line indicates the overall OSNR minimum of 28.1 dB.

method is enacted in order to validate the prediction method on the CUT OSNR response. During the entire analysis, we have not taken into account any uncertainty in the measurements, as they are negligible with respect to the characteristic variances of the system. A few basic considerations arise by calculating the average of the OSNRs for each channel over the entire sample, presented in Fig. 5. These OSNR averages sketch a characteristic figure of the EDFA amplification process, which takes place between 29.5 and 30.9 dB, with standard deviations from 0.14 to 0.40 dB. In order to learn more about the EDFA cascade behavior, it is necessary to consider each configuration separately. In fact, the OSNR of each channel depends upon the state of every other channel within the spectral load. For example, as a primitive analysis in this direction, we investigate how the OSNR distributions change with regards to the number of *on* channels in the spectral load. Figures 6 and 7 present the distributions enclosed in Fig. 5 for a select subset of channels, plotted against the total number of *on* channels in the configurations: here these figures show the means and standard deviations, σ , of the channels, respectively. It must be noted that because the data set was further divided into chunks, the reliability of the averaged quantities is substantially decreased. This causes the standard deviation (presented in Fig. 7) to be far less uniform across all channels when only a small number of channels are in the *on* state. Regardless, Fig. 6 shows that for the CUT ($f = 195.25$ THz), there is an unquestionable increase in the OSNR as the line approaches a full load configuration. Moreover, for all channels σ decreases under the same conditions, meaning that the system tends toward a stable state. To further characterize the OSNR response with respect to a specific configuration, it is necessary to fully understand the intrinsic behavior of the amplification phenomenon.

A. Physical Considerations

Despite it being possible to obtain a precise physical description of the emission phenomenon involved in the amplification

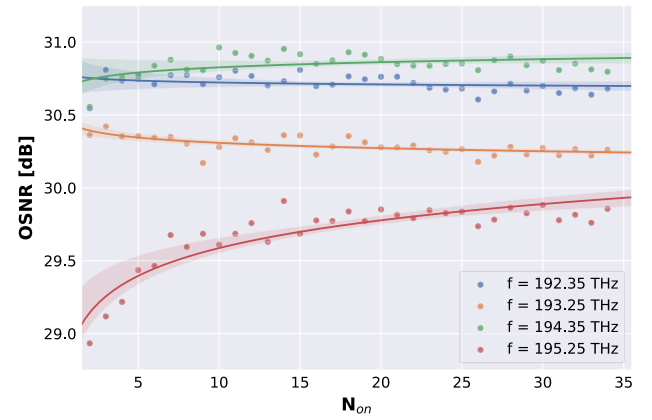


Fig. 6. Mean values of four channel OSNRs are plotted with respect to the configurations for an increasing N_{on} . In the legend, we report the central frequency of the channels considered. The colored lines and shaded areas are qualitative visual expressions of the trend of measured data.

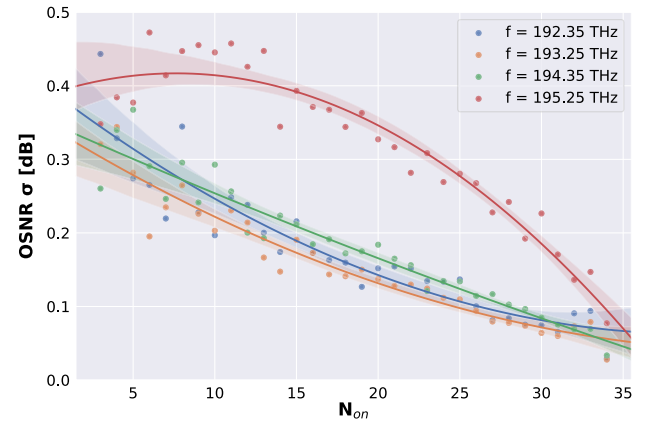


Fig. 7. Standard deviation values of the same configurations plotted in Fig. 6. As expected, the channel centered at 195.25 THz maintains the highest variance out of all of the configurations. The colored lines and shaded areas are qualitative visual expressions of the trend of measured data.

process, without accurate knowledge of the OLS physical parameters it is not feasible to determine the evolution of the spectral load through the EDFA cascade. In a general scenario, this obstacle would be exacerbated by the embedded EDFA software controller, which, in order to maintain specific requirements, changes the spectral powers at the output of the amplifiers with an unknown algorithm. Properly addressing the cause of the OSNR fluctuations requires splitting the OSNR into its constituents: the received signal power and the ASE noise. An important point is that intensity of the signal amplification and the ASE noise are strictly related. Essentially, these quantities coincide with the stimulated and spontaneous emission of the amplifiers, respectively, and both depend on the population inversions of the erbium within the EDFAs [14]. As a rough summary, if no power is transmitted in a given frequency band, all the relative population inversion is utilized by the ASE noise, allowing it to reach a maximum value. In

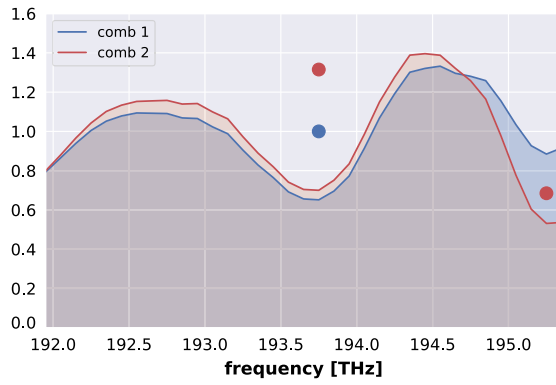


Fig. 8. Qualitative visualization of the OSNR fluctuations that arise from turning on a new channel, for both the ASE noise (shaded lines) and the power of the *on* channels (dots). Here, the $N_{on} = 1$ case is given in red, and the $N_{on} = 2$ case is given in blue. In this figure, all quantities are normalized in order to have a unitary mean value.

contrast, when the transmitted signal is amplified, a smaller amount of population inversion is present, resulting in a lower maximum noise value that may be attained. This effect is shown within Fig. 8, where two spectral load configurations are considered. Here, a clear reduction in ASE noise is observed by switching an extra channel *on*. This is the case for all channels, with the minimum amount of ASE noise being achieved when all channels are in the *on* state. Furthermore, it should be noted that among all possible configurations, the example shown in Fig. 8 experiences the wildest change in the noise figure. In fact, the channel switched to the *on* state has a frequency bandwidth centered at 195.25 THz, with a frequency close to the peak of the well-known spectral hole burning phenomenon [14]. Likewise, this behavior is also reflected by the large OSNR variance of this channel. Revisiting the data set, this feature is pictured in Fig. 9, where we plot the standard deviations of the overall OSNR measurements for each channel. Furthermore, in Fig. 8 it can be observed that even though channels have a frequency spacing of 32 GHz in this experiment, changing a channel to *on* can affect the power of the noise upon frequency bandwidths hundreds of gigahertz away. Since the EDFA energy level population inversion quantifies the intensity of both the amplification and of the noise, we can conclude that the state of a single channel impacts both the signal power and the ASE noise of channels within its frequency neighborhood. This cross-dependency between the power of the channel and the ASE noise, which depends on the state of the other channels, means that calculating the OSNR of every channel is challenging; this is not an intrinsic value of the channel but of the entire spectral load. Owing to the above considerations, it is not possible to further characterize the OSNR response for a particular configuration if the parameters of the OLS are not accurately known.

Apart from the statistical description of the entire data set and the heuristic analysis on OSNR fluctuations, we wish to use this data set as grounds for a realistic use case. In general, the required margin must be conservative and take into account the OSNR fluctuations, and depends upon the needs of the OLS operators; to be agnostic with respect to these needs and to compare different prediction methods in a fair manner,

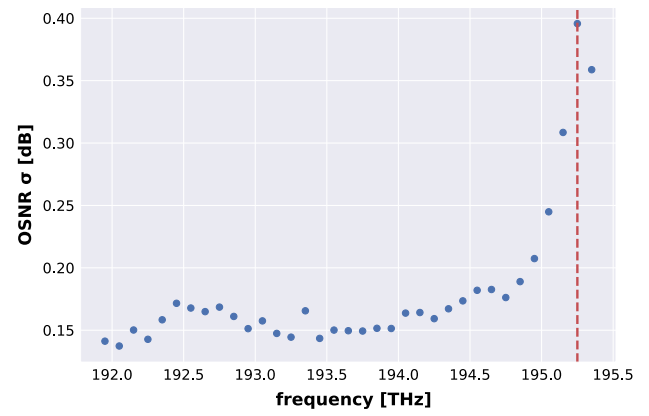


Fig. 9. Standard deviation trend over all of the channels, highlighting an increase as the OSNR approaches the frequency where the peak of the spectral hole burning occurs, given by the dashed red line in the figure.

we quantify an estimation of the average margin by calculating the root-mean-square (RMS) error, given by

$$\text{RMS} = \sqrt{\frac{\sum_{i=0}^D (\text{OSNR}_i^t - \text{OSNR}_i^p)^2}{D}}, \quad (4)$$

where OSNR_i^t and OSNR_i^p are the measured and predicted values of the CUT OSNR for the i th spectral load, respectively, and D is the dimension of the test data set. If nothing is known about the OSNR dependency upon frequency, the same OSNR threshold must be implemented for all channels with a magnitude lower than an overall expected minimum. In this case, the OSNR_i^p are set to the constant OSNR threshold of 28.1 dB, producing an average margin of up to 2.28 dB over a set of realizations equivalent to our sample.

Supposing the availability of stored data that describes the frequency-resolved OSNR response [option (3) in Section 2], one can reduce the margin by setting a minimum value for each channel that must lie beneath the respective minimum measurement (the continuous green line in Fig. 5). Although this solution is suboptimal, it is the best achievable result that is conservative and agnostic with regards to the specific spectral load configuration. This solution produces a limited improvement, compared to the initial value of 2.28 dB, as the average margin would lie between 1.72 and 0.46 dB, depending upon the channel. This result can be further improved by characterizing the OSNR fluctuation dependency upon the specific spectral load configuration; as the user knows the number of *on* channels for a given spectral load, they can set the threshold as the minimum value of the OSNR measurement for the given N_{on} . The result of this approach produces an RMS error which lies between 1.22 and 0.09 dB for the CUT (worst-case scenario), shown in Fig. 10. These improvements would reduce the margin in an effective manner; however, being highly dependent upon the sample features, their accuracy is limited by the statistical incidence of the sample over all possible realizations of the system. This means that having a reliable value for each channel may require considering a large number of instances. In light of this, a ML approach appears to

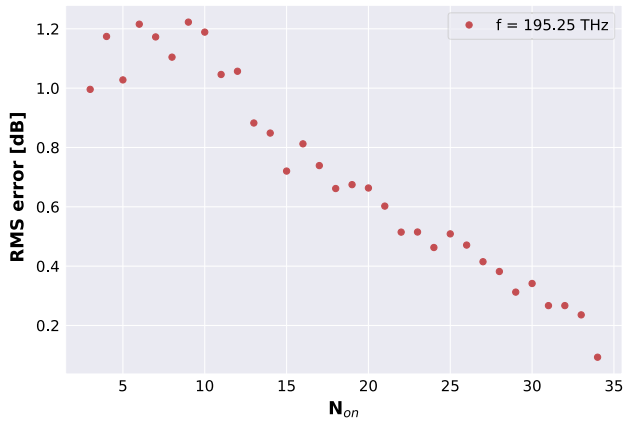


Fig. 10. RMS error for the worst-case scenario channel, with an increasing number of channels in the *on* state within the configuration, obtained considering the respective minimum measured OSNR value used as a margin threshold.

be an appropriate candidate to increase the accuracy of OSNR predictions, if the dimensions of the sample are fixed.

5. QOT-E BASED ON MACHINE-LEARNING

The prediction of the OSNR based upon a specific spectral load configuration is an ideal scenario for ML, especially within a case where the OLS is treated as a black box, as ML is able to compensate for the lack of knowledge of the OLS parameters. In order to measure the enhancement obtained using a ML approach, we focus on the real-case scenario outlined at the end of Section 2. Far from being an exhaustive description of ML applications, the goal of this work is to achieve a better prediction of OSNR using ML techniques in the scenario under investigation. First, it is necessary to divide the measurement data set into training and testing sets. The former represents the stored data set on which the OLS controller can base the OSNR predictions for a LP that will be allocated to the CUT. The latter represents a set of real outcomes that can be used to validate the accuracy of a particular prediction method. To estimate this accuracy we use the RMS error, considering $OSNR_i^m$ and $OSNR_i^p$ as the measured and predicted values of the CUT OSNR, restricted to the test subset of the data set. Setting a constant $OSNR_i^p$ for all i as the minimum measured value of the CUT OSNR yields a value of 1.63 dB RMS error over all the configurations in the test data set. Following this, we take advantage of the well-known TensorFlow platform [28] to perform ML, adapting various high-level features from this platform according to our requirements.

Before proceeding with implementing a ML technique to predict the OSNR of an OLS, we first undertook preliminary investigations in order to probe whether a neural network or a linear regression model provides superior performance—as a result we decided to utilize the DNN implemented in TensorFlow, which is a feed-forward multilayer (deep) neural network, because it outperforms a linear regression model in this scenario. We applied this DNN model to our data set, obtaining various levels of accuracy depending on the DNN

network parameters. We characterized this DNN model utilizing a proximal Adagrad optimizer (again, implemented in TensorFlow [28]) with a fixed learning rate of 0.1 and a regularization strength of 0.001. Most importantly, we have tuned the number of hidden layers and nodes in order to achieve the best trade-off between precision and computational time. These two parameters are linked to the complexity of the DNN, which in turn is tied to the complexity of the problem to be solved. Although increasing the number of layers and nodes improves the accuracy of the DNN, raising these values also has an adverse effect on computational time. In the end, we decided upon a DNN with three hidden layers, containing 32 nodes each, taking approximately 8 min to train (using a machine running with 32 GB of 2133 MHz RAM and an Intel Core i7 6700 3.4 GHz CPU), as increasing DNN complexity does not further improve the accuracy of the OSNR estimations. These quantities would be changed if we considered a system with a larger number of amplifiers, with the computation time increasing accordingly (a rough estimation obtained from our trials is that the computation time scales linearly with the number of nodes). Once the model has been trained it can be validated and utilized for any possible spectral load configuration, within the overall investigated bandwidth, for the OLS under consideration.

A. Data Set Preparation

Considering a single CUT (with $f = 195.25$ THz), we selected 30% of the data set to be designated as a testing subset. Because of the CUT being close to the spectral hole burning peak, this is a worst-case scenario for OSNR fluctuations; therefore, lower error predictions are expected for all other CUT selections. The testing subset was created by randomly choosing instances from the data set, with the only requirement being that the uniformity of the distribution with respect to the number of *on* channels in the configurations was preserved. This means that for each configuration subset with a given N_{on} we select 30% to be in the test data set. DNN training and prediction processes require the definition of features and labels, which indicate system inputs and outputs, respectively. As outlined in the previous section, the uncertainty of the system can be divided along the variances of the received signal power and the ASE noise. Therefore, we consider these two quantities as independent inputs of the system and set them as the DNN features. Correspondingly, the OSNR is the only system output under investigation and so is set as the DNN label. In order to properly address the aforementioned realistic scenario, the DNN features correspond to the quantities measured when the CUT is *off*, whereas the labels correspond to the CUT OSNR when the CUT is in the *on* state. As a consequence of this restriction, the final data set composed of the training and testing subsets is half the size of the original data set.

B. Results and Comments

In Fig. 11, we show the distributions of the measured OSNR for the CUT and the predictions of the DNN over the test data set. This figure highlights how the DNN predictions closely resemble the measured OSNR values, having a similar mean,

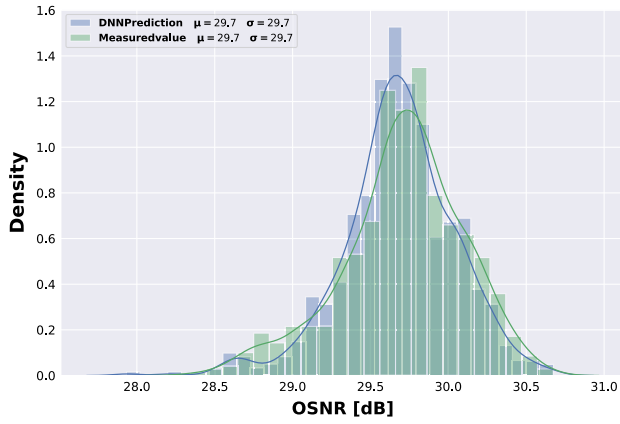


Fig. 11. Comparison of the OSNR distributions of the DNN guesses and the measured values, respectively.

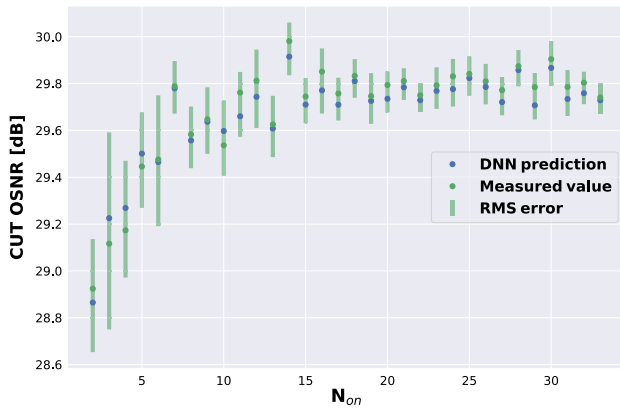


Fig. 12. Comparison of the OSNR averages of the DNN guesses and the measured values, respectively, presented in terms of the number of channels, N_{on} . With the error bars we indicate the RMS error.

μ , range, and standard deviation, σ . An average margin of 0.15 dB is obtained through this DNN estimation of the CUT OSNR, a significant improvement with respect to the previous solutions presented at the end of the previous section. To properly frame these results in the realistic use-case scenario, it must be underlined that despite the DNN providing a high level of accuracy, it may make predictions that are not conservative. For example, in this case 38% of the predictions are greater than the real values, even if the majority are greater by a marginal amount. This percentage of nonconservative predictions may be reduced by shifting the OSNR estimations of the DNN by a fixed amount. For example, to reach a scenario where less than 6% of the predictions are nonconservative, the DNN estimations must be shifted by a factor of 0.2 dB, giving an RMS error of 0.27 dB, which remains a significant improvement over the initial average margin estimations. Furthermore, it should be stressed that the data set used in this work contains fewer configurations where a small number of channels are *on*, visible in Fig. 12. The result is that these scenarios are underrepresented in the training data set, causing the accuracy of the DNN predictions to be lower when $N_{on} < 10$; ensuring that these cases are represented equally would reduce the overall RMS

error. Additionally, Fig. 12 reveals that all nonconservative cases in this investigation were given when $N_{10} = 10$ or less, further stressing that the criticalities of the DNN prediction depends upon the statistical incidence of the sample over all possible realizations. In light of these results a ML approach exhibits promising accuracy, and it seems that with further, more in-depth parameter selection and training that the DNN may eventually lead to an OSNR margin estimation that approaches zero, at least for similar use cases.

6. CONCLUSIONS

In this work we have addressed the system margin minimization enabled by a reliable prediction of the QoT given by the GSNR. The main idea of our approach is that in order to obtain the best estimation of the GSNR, this QoT-E must be separated into OSNR and nonlinear SNR components. In fact, because of the inaccuracy on the parameters and the software-defined EDFA behavior, the former cannot be analytically estimated in an accurate way, and so requires an adaptive approach. We focus on predicting the OSNR component of the GSNR, as opposed to the nonlinear SNR, as this term is both the most dominant and the most affected by uncertainties. We propose a ML approach to estimate the OSNR response over distinct spectral load configurations, leaving the estimation of the nonlinear SNR to an analytical model that may give a fast and accurate prediction, once the actual signal spectral powers are known.

We supposed an agnostic use of the OLS by operating the EDFAs as black boxes that set the nominal gain and by relying only on data from the OCM to predict the spectrally resolved GSNR. Experimentally, we obtained a data set from an OLS containing a cascade of 11 pairs of EDFAs and VOAs; we utilize the attenuators in place of the fiber in order to avoid any NLI generation and to focus our investigation only on the prediction of the OSNR.

We consider a realistic scenario where an OLS controller wishes to predict the OSNR of a LP over the CUT, given an existing spectral load. Supposing the availability of previously measured OSNR outputs, we give different predictions with different levels of accuracy by considering different OLS behavior awareness. First, we show that, without any specific knowledge of the OLS or the uncertainty fluctuations of the OSNR, deploying the minimum required conservative threshold produces an average margin of 2.28 dB. Next, by considering the minimum measurements for each channel as an OSNR threshold we evaluate a varying average margin that lies between 1.72 and 0.46 dB, depending upon the channel under consideration. This result can be further improved by assuming that the N_{on} is known, allowing the OSNR threshold to be set to the minimum values that have been measured within the respective set of configurations. An average margin between 1.22 and 0.09 dB is found in this case, which, nevertheless is not reliable as it depends strongly upon the statistical incidence of the analyzed sample over all possible realizations of the system. Finally, we demonstrate that DNN ML techniques from the TensorFlow platform enable an accurate OSNR estimation with an RMS error of 0.15 dB over the CUT, representing the worst-case scenario. By applying a

rigid shift to the DNN predictions, it is possible to guarantee a requested conservative percentage threshold, decreasing the DNN accuracy. For example, introducing a shift of 0.2 dB to the DNN estimations produces a result where 94% of the predictions are fully conservative and gives a reasonable RMS error of 0.27 dB.

To conclude, future analyses performed by also including telemetry data from the EDFAs may yield a further reduction in the residual uncertainty, consequently reducing the required system margin. Furthermore, a future investigation could exploit a ML algorithm that, during the training stage, penalizes prediction values that are higher than the measured values, obtaining a model that is predisposed to conservative predictions and ensuring that the model maintains reliability with high accuracy.

Funding. H2020 Marie Skłodowska-Curie Actions (814276).

Acknowledgment. The authors would like to thank Alessio Ferrari and Dr. Mattia Cantono for their fruitful suggestions.

Disclosures. The authors declare no conflicts of interest.

REFERENCES

1. "Cisco Visual Networking Index: Forecast and Trends, 2017–2022," Cisco White Paper (2017), <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html>.
2. V. Curri, M. Cantono, and R. Gaudino, "Elastic all-optical networks: a new paradigm enabled by the physical layer. How to optimize network performances?" *J. Lightwave Technol.* **35**, 1211–1221 (2017).
3. D. J. Ives, P. Bayvel, and S. J. Savory, "Routing, modulation, spectrum and launch power assignment to maximize the traffic throughput of a nonlinear optical mesh network," *Photon. Netw. Commun.* **29**, 244–256 (2015).
4. Y. Pointurier, J.-L. Augé, M. Birk, and E. Varvarigos, "Introduction to the JOCN special issue on low-margin optical networks: publisher's note," *J. Opt. Commun. Netw.* **11**, 598 (2019).
5. Y. Pointurier, "Design of low-margin optical networks," *J. Opt. Commun. Netw.* **9**, A9–A17 (2017).
6. D. W. Boertjes, M. Reimer, and D. Côté, "Practical considerations for near-zero margin network design and deployment," *J. Opt. Commun. Netw.* **11**, C25–C34 (2019).
7. V. Curri, A. Carena, A. Arduino, G. Bosco, P. Poggiolini, A. Nespola, and F. Forghieri, "Design strategies and merit of system parameters for uniform uncompensated links supporting Nyquist-WDM transmission," *J. Lightwave Technol.* **33**, 3921–3932 (2015).
8. R. Pastorelli, "Network optimization strategies and control plane impacts," in *Optical Fiber Communication Conference* (OSA, 2015).
9. M. Filer, M. Cantono, A. Ferrari, G. Grammel, G. Galimberti, and V. Curri, "Multi-vendor experimental validation of an open source QoT estimator for optical networks," *J. Lightwave Technol.* **36**, 3073–3082 (2018).
10. A. Bononi, P. Serena, and N. Rossi, "Nonlinear signal-noise interactions in dispersion-managed links with various modulation formats," *Opt. Fiber Technol.* **16**, 73–85 (2010).
11. P. Poggiolini, A. Carena, Y. Jiang, G. Bosco, V. Curri, and F. Forghieri, "Impact of low-OSNR operation on the performance of advanced coherent optical transmission systems," in *The European Conference on Optical Communication (ECOC)* (IEEE, 2014), pp. 1–3.
12. A. Ferrari, G. Borracchini, and V. Curri, "Observing the generalized SNR statistics induced by gain/loss uncertainties," in *European Conference on Optical Communication (ECOC)* (IEEE, 2019).
13. B. Taylor, G. Goldfarb, S. Bandyopadhyay, V. Curri, and H.-J. Schmidtknecht, "Towards a route planning tool for open optical networks in the telecom infrastructure project," in *Optical Fiber Communication Conference and the National Fiber Optic Engineers Conference* (2018).
14. M. Bolshtyansky, "Spectral hole burning in erbium-doped fiber amplifiers," *J. Lightwave Technol.* **21**, 1032–1038 (2003).
15. G. Grammel, V. Curri, and J. L. Auge, "Physical simulation environment of the telecommunications infrastructure project (TIP)," in *Optical Fiber Communication Conference and the National Fiber Optic Engineers Conference* (2018).
16. 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 *Asia Communications and Photonics Conference (ACP)* (IEEE, 2018), pp. 1–3.
17. J. Thrane, J. Wass, M. Piels, J. C. Diniz, R. Jones, and D. Zibar, "Machine learning techniques for optical performance monitoring from directly detected PDM-QAM signals," *J. Lightwave Technol.* **35**, 868–875 (2017).
18. F. N. Khan, C. Lu, and A. P. T. Lau, "Optical performance monitoring in fiber-optic networks enabled by machine learning techniques," in *Optical Fiber Communication Conference and Exposition (OFC)* (IEEE, 2018), pp. 1–3.
19. L. Barletta, A. Giusti, C. Rottondi, and M. Tornatore, "QoT estimation for unestablished lightpaths using machine learning," in *Optical Fiber Communication Conference* (Optical Society of America, 2017), paper Th1J–1.
20. I. Sartzetakis, K. K. Christodoulopoulos, and E. M. Varvarigos, "Accurate quality of transmission estimation with machine learning," *J. Opt. Commun. Netw.* **11**, 140–150 (2019).
21. W. Mo, Y.-K. Huang, S. Zhang, E. Ip, D. C. Kilper, Y. Aono, and T. Tajima, "ANN-based transfer learning for QoT prediction in real-time mixed line-rate systems," in *Optical Fiber Communication Conference and Exposition (OFC)* (IEEE, 2018), pp. 1–3.
22. C. Rottondi, L. Barletta, A. Giusti, and M. Tornatore, "Machine-learning method for quality of transmission prediction of unestablished lightpaths," *J. Opt. Commun. Netw.* **10**, A286–A297 (2018).
23. J. Mata, I. De Miguel, R. J. Duran, N. Merayo, S. K. Singh, A. Jukan, and M. Chamania, "Artificial intelligence (AI) methods in optical networks: a comprehensive survey," *Opt. Switching Netw.* **28**, 43–57 (2018).
24. 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 *European Conference on Optical Communication (ECOC)* (IEEE, 2018), pp. 1–3.
25. C. L. Gutterman, W. Mo, S. Zhu, Y. Li, D. C. Kilper, and G. Zussman, "Neural network based wavelength assignment in optical switching," in *Proceedings of the Workshop on Big Data Analytics and Machine Learning for Data Communication Networks* (ACM, 2017), pp. 37–42.
26. A. Mahajan, K. Christodoulopoulos, R. Martinez, S. Spadaro, and R. Munoz, "Machine learning assisted EFDA gain ripple modelling for accurate QoT estimation," in *European Conference on Optical Communication (ECOC)* (IEEE, 2019).
27. M. Ionescu, "Machine learning for ultrawide bandwidth amplifier configuration," in *21st International Conference on Transparent Optical Networks (ICTON)* (IEEE, 2019).
28. <https://www.tensorflow.org/>.
29. <https://www.itu.int/rec/T-REC-G.694.1/en>.
30. D. J. Elson, G. Saavedra, K. Shi, D. Semrau, L. Galdino, R. Killely, B. C. Thomsen, and P. Bayvel, "Investigation of bandwidth loading in optical fibre transmission using amplified spontaneous emission noise," *Opt. Express* **25**, 19529–19537 (2017).
31. A. Nespola, S. Straullu, A. Carena, G. Bosco, R. Cigliutti, V. Curri, P. Poggiolini, M. Hirano, Y. Yamamoto, T. Sasaki, J. Bauwelinck, K. Verheyen, and F. Forghieri, "GN-model validation over seven fiber types in uncompensated PM-16QAM Nyquist-WDM links," *IEEE Photon. Technol. Lett.* **26**, 206–209 (2014).

32. D. Piliori, F. Forghieri, and G. Bosco, "Residual non-linear phase noise in probabilistically shaped 64-QAM optical links," in *Optical Fiber Communication Conference and the National Fiber Optic Engineers Conference* (2018).
33. R.-J. Essiambre and R. W. Tkach, "Capacity trends and limits of optical communication networks," *Proc. IEEE* **100**, 1035–1055 (2012).
34. A. Carena, V. Curri, G. Bosco, P. Poggiolini, and F. Forghieri, "Modeling of the impact of nonlinear propagation effects in uncompensated optical coherent transmission links," *J. Lightwave Technol.* **30**, 1524–1539 (2012).
35. M. Cantono, D. Piliori, A. Ferrari, C. Catanese, J. Thouras, J. L. Auge, and V. Curri, "On the interplay of nonlinear interference generation with stimulated Raman scattering for QoT estimation," *J. Lightwave Technol.* **36**, 3131–3141 (2018).
36. 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).
37. https://www.cisco.com/c/en/us/products/collateral/optical-networking/ons-15454-series-multiservice-transport-platforms/data_sheet_c78-658542.html.