

Introducing Load Aware Neural Networks for Accurate Predictions of Raman Amplifiers

A. Margareth Rosa Brusin, Uiara C. de Moura, Vittorio Curri, Darko Zibar, and Andrea Carena

Abstract—An ultra-fast machine learning based method for accurate predictions of gain and amplified spontaneous emission (ASE) noise profiles of Raman amplifiers is introduced. It is an alternative to high-complexity and time-consuming standard approaches, which are based on the numerical solution of sets of nonlinear differential equations. Main relevance resides on its possible application in real-time network controllers for future multi-band optical line systems where Raman amplification will be required to cope with capacities beyond the standard C-band. Here we consider as an example the C+L-band scenario with different input load conditions: full load and partial loads. For the case of full load it has been recently shown a neural network (NN) capable of highly accurate predictions. Real optical networks are not usually operated only in full load conditions: the load can dynamically vary over time and the behavior of the Raman amplifier depends on it. In this paper we introduce a new NN model and we show its higher accuracy when the line system is not fully loaded: we define it as the load aware neural network. Applying this new approach we can predict both gain and ASE noise profiles in Raman amplifiers with high accuracy under any load conditions: we demonstrate almost 100% of maximum prediction errors to be lower than 0.5 dB.

Index Terms—optical communication, optical amplifiers, machine learning, neural networks.

I. INTRODUCTION

RECENTLY, machine learning (ML) has gained large attention in the optical communication community. In general, it allows to solve highly complex problems through low complexity approaches, therefore several applications have been already proposed. Mostly known for its use at network operation level, such as in software-defined networking (SDN) [1], ML has also been successfully applied in different fields of optical communications. It has been proposed for nonlinear compensation [2], optical performance monitoring (OPM) and modeling [3], fault detection/prevention [4]

A. M. Rosa Brusin, V. Curri and A. Carena are with Dipartimento di Elettronica e Telecomunicazioni, Politecnico di Torino, Torino, Italy (e-mail: ann.rosabrusin@polito.it).

U. C. de Moura and D. Zibar are with DTU Fotonik, Department of Photonics Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark.

Manuscript received xx xx, 20xx; revised xx xx, 20xx.

and control of Erbium Doped Fiber Amplifiers (EDFAs) [5].

Due to the fast increase in capacity demand, the need of moving towards multi-band systems, with occupation up to O+E+S+C+L [6], led to the resorting of stimulated Raman scattering (SRS) based amplification schemes [7]. In particular, by increasing the number of pumps and properly tuning the pump powers and wavelengths, distributed Raman amplifiers (RAs) are able to provide gains over such large bands with low noise, an advantage with respect to other possible amplification solutions, as Semi-conductor Optical Amplifiers (SOAs) [8] and EDFAs [9].

The physical behavior of Raman amplifiers is modeled with good accuracy by a set of nonlinear ordinary differential equations (ODEs) [10]. The only drawback of this approach is the complexity of the model, for which the determination of the numerical solution is quite time consuming. Moreover, the time needed for the convergence to an accurate solution increases when considering broadband systems requiring a large number of pumps. In this scenario, the proper setting of RA pump powers and wavelengths for a specific gain profile, also known as the RA design, is commonly done using standard multi-variable optimization process or by genetic algorithms [11]. Both these approaches rely on the iterated numerical solution of the ODEs by means of a solver: every new design requires to start the process from the beginning.

Lately, new approaches based on machine learning has been presented to determine the optimal pumps allocation for a target Raman gain profile [12]–[14]. In [12], a first proof-of-principle of the *inverse model* for a pure RA is shown on synthetic data, whilst in [13] the complete and extensive ML framework is reported and experimentally demonstrated. In addition, a fine tuning of pump powers and wavelengths is implemented exploiting the so called ML-based *forward model* over both synthetic and experimental data. In this application the *forward model* is used instead of the numerical solver to instantaneously predict gain profiles for the gradient descent algorithm that adjusts the pumps. A similar goal

was targeted in [15] where they also developed a *forward model* for the RA in order to avoid time consuming numerical solutions of model equations.

In the first part of this paper we present an extension of [16], where we introduced the *forward model* for the prediction of both gain and noise profiles for Raman amplifier covering the C+L-band, using a fully ML-based approach with hyper-parameters optimization. Recently another experimental validation both for the *inverse* and *forward model* have been published [17] confirming the effectiveness of the ML approach and the interest of the community for a practical tool to design RA. All studies published up to now [12]–[17] have considered only the full load input condition: this limits the applicability of this approach to real networks where traffic patterns can vary over time and rarely reach the full-load condition.

In the second part of this paper we extend the results first introduced in [18] for the gain prediction only: we consider Load-Aware neural networks capable of accurate gain and ASE noise profiles predictions of Raman amplifiers under any load conditions. In particular, the main principle adopted to reduce the data-set dimensionality is to split the 11 THz overall C+L-band between 185 THz and 196 THz into $N_{\text{sb}} = 22$ sub-bands of 500 GHz, each of them carrying ten 50 GHz slots. The input load of the RA is considered on a sub-band basis to reduce the number of conditions to be analyzed.

The major relevance of this work that defines a highly accurate *forward model* for the prediction of gain and noise behavior in a Raman amplifier with variable loads resides on its possible application in real-time network controllers. With the introduction of multi-band optical line systems where Raman amplification is required to extend the gain window beyond the C-band, having a fast tool for describing the behavior of a RA will be a requirement.

In the following, we present the paper structure. In section II, the machine learning framework for the forward model, the generation of the data-sets and the neural network (NN) models are described. In section III, the application of the proposed Load-Aware approach is shown and a statistical analysis of its performance is presented. Finally, conclusions are discussed in section IV.

II. MACHINE LEARNING FRAMEWORK, DATA-SETS GENERATION & NEURAL NETWORK MODELS

A. Machine Learning Framework

Our goal is to design a NN able to substitute the *forward model* that relates output \mathbf{Y} (gain or noise profiles) to input \mathbf{X} (pumps powers and in some cases

input load). Generalizing the problem, we target the forward mapping function $\mathbf{Y} = f(\mathbf{X})$ that in this case is a complex relationship described by a set of ODE, with $\mathbf{X} = [x_1, \dots, x_M]^T$ and $\mathbf{Y} = [y_1, \dots, y_N]^T$ being the transpose input and output vectors, respectively. M and N are their respective length. Given a training data-set $\mathcal{D}^{K \times (N+M)} = \{\mathbf{Y}_k^T, \mathbf{X}_k^T | k = 1, \dots, K\}$, where K is the data-set size, we train a multi-layer neural network to learn an approximation of the mapping function $f(\cdot)$. Particularly, the training determines the set of neural network weights $\mathbf{W} = [\mathbf{W}^{(1)}, \dots, \mathbf{W}^{(N_{HL})}]$, where N_{HL} is the number of hidden layers of the NN. Then, once the mapping has been learned, the NN can be used to predict \mathbf{Y} given a new value of \mathbf{X} . We consider the NN introduced in [16] assuming the full-load input condition as Load-Unaware (LU) and we introduce a new model that we define as Load-Aware (LA). Depending on the parameters at the input \mathbf{X} of the neural network, we analyze the performance of two models, as shown in Fig. 1:

- LU-NN, a Load-Unaware Neural Network taking as input \mathbf{X} the pump powers $\mathbf{P} = [P_1, \dots, P_{N_p}]$, see Fig. 1(a);
- LA-NN, a Load-Aware Neural Network taking as input \mathbf{X} the pump powers $\mathbf{P} = [P_1, \dots, P_{N_p}]$ and the load information $\mathbf{S} = [S_1, \dots, S_{N_{\text{sb}}}]$, see Fig. 1(b).

N_p is the number of pumps used in the RA: in our study it is set to 5. N_{sb} is the number of sub-bands in which we split the input: in our study it is set to 22.

The output \mathbf{Y} can be associated either to the gain profile or the amplified spontaneous emission (ASE) noise profile of the Raman amplifier under analysis.

Based on the results presented in [16], showing a comparison between two training methods, the *Back-propagation* (BP) [19] and the *Random Projection* (RP), also known as extreme learning machine (ELM) [20], in this work we consider the RP only. In fact, unlike the most known BP method, in RP the learning of the input-output mapping does not rely on gradient descent algorithms. Indeed, in the RP method, after a random initialization, the NN weights are not iteratively optimized, but they are fixed and the learning relies on matrix inversion and multiplication only at the last layer of the NN. This results in a reduced computational complexity required to train the network, much lower than in the BP case, making the RP a method with an ultra-fast training phase. Moreover, its high prediction accuracy is guaranteed by a proper initialization of two key parameters, the standard deviation of the weights and the regularization parameter [20].

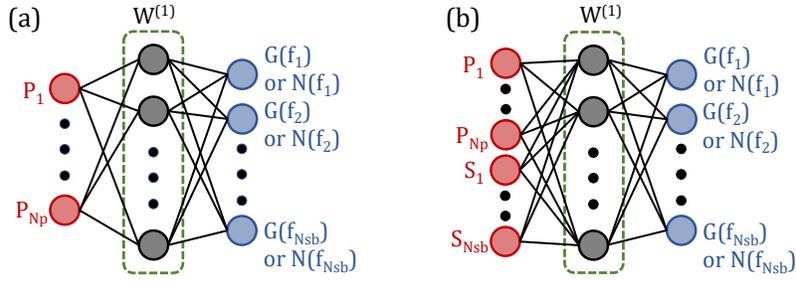


Fig. 1. Neural network models considered in the paper: (a) LU-NN and (b) LA-NN.

To improve the accuracy of predictions obtained with neural networks, an extensive hyper-parameters optimization has been performed over the number of hidden layers, the number of hidden nodes and the activation function. To reduce the impact of the randomly initialized weights, N_{NN} parallel and independent NNs are trained. Thus, the predictions are the result of a model averaging, i.e. they are the average of the outputs of all the N_{NN} networks.

To determine the goodness of the predictions, each predicted profile \mathbf{Y}^{pred} is compared to its corresponding target profile \mathbf{Y}^{targ} determined by the numerical solver. The selected metrics are the root-mean-square-error ($RMSE$) and the maximum error ($Error_{MAX}$), defined as

$$RMSE = \sqrt{\frac{1}{N_{sb}} \sum_{n=1}^{N_{sb}} (Y_n^{pred} - Y_n^{targ})^2} \quad (1)$$

$$Error_{MAX} = \max_{n \in \{1, \dots, N_{sb}\}} |Y_n^{pred} - Y_n^{targ}| \quad (2)$$

respectively.

B. Data-sets generation

In this section we describe the generation of the synthetic data-sets used to train and validate the NNs under analysis. A numerical solver, available within the open source library GNPY [21], solving the ODEs describing the Raman effect [10], is used to generate the data-sets.

As shown in Fig. 2, we consider a distributed Raman amplifier with $N_p = 5$ counter-propagating pumps over a span of single mode fiber (SMF) of length $L_{span} = 100$ km. Here we analyze a single span, but it is possible to consider more cascaded spans and also a combination of RA with Erbium Doped Fiber Amplifiers (EDFAs) to form a hybrid amplification scheme (hybrid EDFA+RA). Fiber parameters are

the following: the attenuation coefficients for the input signal and for the pumps are $\alpha_s = 0.21$ dB/km and $\alpha_p = 0.25$ dB/km, respectively. Pumps power values are drawn from a uniform distribution $P_i \sim U[0, 250]$ mW with $i \in \{1, \dots, 5\}$. On the opposite, pump frequencies are fixed, as usual in standard commercial modules. We set them at the following values: $f_1 = 210.37$ THz, $f_2 = 207.14$ THz, $f_3 = 204.01$ THz, $f_4 = 200.97$ THz and $f_5 = 198.03$ THz.

These values have been selected to enable gains over the 11 THz C+L-band, defined between 185 THz and 196 THz. At the input of the optical system we consider a Wavelength Division Multiplexed (WDM) comb composed of up to $N_{ch} = 220$ Nyquist-shaped channels, sitting in a 50 GHz grid.

To generate the data-set for the partial input load case, we face a problem of complexity because of the huge dimension of the space to be analyzed. Considering input channels all at the same power level (1 mW) but assuming they can be ON or OFF, we have a total of 2^{220} conditions. Moreover such cases must be combined together with random pumps. Approaching the problem in this form gives poor results because only a limited number of cases can be generated for the data-set, due to the complexity of the numerical evaluation of the model describing the RA. To reduce the dimension of the input load space we approach the data-set generation on a sub-band basis instead of on a channel basis. Thus, 10 adjacent slots identify a 500 GHz sub-band, for a total of $N_{sb} = 22$ sub-bands over the whole C+L-band. We then associate 10 slots to the C-band and 12 slots to the L-band.

Consider that our approach based on channels grouped in sub-bands can also be interpreted as a direct mapping of the concept of Super-Channels, that has been proposed and extensively studied in literature [22], [23]. In the following we do not refer specifically to a Super-Channel based systems because our approach can have a more

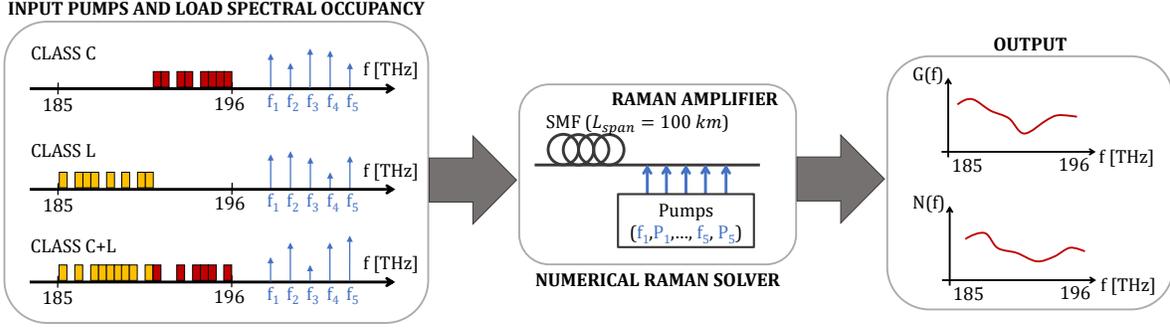


Fig. 2. Schematic representation of the optical system considered in our analysis.

TABLE I
OPTIMAL HYPER-PARAMETERS VALUES FOR THE LU-NN
TRAINED USING THE FULL-LOAD DATA-SET

Profile type	Gain	ASE noise
Activation function ($actFun$)	tanh	tanh
# hidden nodes (N_{HN})	1980	1900

TABLE II
OPTIMAL HYPER-PARAMETERS VALUES FOR THE LU-NN
TRAINED USING THE PARTIAL-LOAD DATA-SET

Profile type	Gain	ASE noise
Activation function ($actFun$)	logsig	logsig
# hidden nodes (N_{HN})	640	1920

general application.

Each sub-band can assume two states with same probability: ON, which means that the sub-band is carrying 10 mW power (1 mW per slot), and OFF, which means that the sub-band is not carrying any power. Depending on the load, three classes of data-set can be identified: i) class C, for which only sub-bands in C-band are considered, ii) class L, for which only sub-bands in L-band are considered, and iii) class C+L, where the entire C+L-band can be used. The left-hand side of Fig. 2 offers a pictorial representation of the classes showing a single partial load condition for every class.

A further classification within each class can be made based on the number of sub-bands which is ON. For each class, a fixed number of sub-bands turned ON is drawn with a randomly selected frequency position, forming a sub-class. In particular, to further reduce the number of cases, we consider only even numbers of sub-bands ON. For class C, the sub-classes correspond to a number of sub-bands which is ON from 2 to 10, for class L from 2 to 12, and for class C+L from 2 to 22. The last case, i.e. class C+L with 22 sub-bands ON, is the full load condition considered in [12]–[14], [16].

TABLE III
OPTIMAL HYPER-PARAMETERS VALUES FOR THE LA-NN

Profile type	Gain	ASE noise
Activation function ($actFun$)	tanh	tanh
# hidden nodes (N_{HN})	1980	1980

To summarize, depending on the load type (full or partial), we generate two families of data-sets:

- Full-load, consisting in a 5000-cases data-set generated with a *full* load input (class C+L and all 22 sub-bands turned ON), random pump powers $P_i \sim U[0, 250]$ mW, with $i \in \{1, \dots, 5\}$;
- Partial-load, for each class and sub-class 500-cases data-sets are generated considering random pump powers $P_i \sim U[0, 250]$ mW, with $i \in \{1, \dots, 5\}$, and fixed pump frequencies. Then, they are merged together to form an 11000-cases *partial* load data-set.

For each of the two data-sets, a completely independent and uncorrelated validation data-set is also generated in the same way, with an identical number of cases.

C. LU-NN and LA-NN models

In this sub-section, we study the application of Raman amplifiers in practical conditions: we do not consider only the full load condition but we assume that the load can be partial. Our goal is to design a NN capable of providing accurate gain and noise profile predictions also when the input load of the RA changes.

First, the LU-NN is trained using a full-load data-set. Then, to see if it is able to handle different input conditions, even if it is completely load-agnostic, it is re-trained with the partial-load data-set. Finally, we consider the newly proposed LA-NN that having as input the load conditions can better adapt to them.

In all cases we consider single hidden layer feedforward neural networks (SLFNs) [16], i.e. $N_{HL} = 1$, and we perform hyper-parameters optimization, testing three different activation functions: sine, hyperbolic tangent (*tanh*) and logistic sigmoid (*logsig*). The search of the optimum number of hidden nodes is done sweeping such value from 20 to 2000 in step of 20. We also optimize the values of the standard deviation for the random weights initialization and the regularization parameter. Then, to reduce the dependence on the randomly generated weights, we perform a model averaging over $N_{NN} = 20$ independent NNs. The optimal model is selected as the one providing the smallest prediction errors.

For the LU-NN trained with full-load data-set, optimized hyper-parameters values are reported in Table I for both RA gain and ASE noise profiles predictions. The activation function providing the smallest prediction errors is the hyperbolic tangent for both Raman gain and ASE noise predictions. In particular, the optimal NN requires 1980 and 1900 HNs for the gain and the ASE noise predictions, respectively.

Instead, Table II shows the optimized hyper-parameters obtained for the LU-NN trained with a partial-load data-set for both Raman gain and ASE noise predictions. The optimal activation function is the logistic sigmoid for both cases, with optimal number of HNs being 640 and 1920 for gain and noise predictions, respectively.

Finally, Table III contains the optimized hyper-parameters values for both Raman gain and ASE noise predictions in case of LA-NN. In this case, the optimized values are the same for the gain and noise predictions: both cases need 1980 HNs and the hyperbolic tangent as activation function.

III. SIMULATION RESULTS AND VALIDATIONS

In this section, the validation results are shown: prediction performances in terms of *RMSE* and *Error_{MAX}* are evaluated considering the optimal NN models resulting from Section II.C. First, the LU-NN is discussed and its limited performance with a partial input load is demonstrated. Then, the novel LA-NN approach is analyzed and its resilience in case of partial load conditions is shown. When performing the validation, the data-sets are pruned such that all gain profiles with points outside the interval [4;21] dB are discarded. Gains below the 4 dB threshold are too low and not useful in practical cases, while gains above 21 dB are not needed because they over-compensate the loss of the 100 km fiber span considered in this analysis.

The comprehensive statistical analysis of the prediction accuracy for the three evaluated models is reported

in Figs. 3 and 4 for gain and ASE noise profiles prediction of the Raman amplifier, respectively. Fig. 3(a) shows the validation results of the LU-NN carried out considering only full-load conditions: full-load data-set is used both in training and in validation. With an average *RMSE* of 0.0039 dB and an average *Error_{MAX}* of 0.015 dB, the LU-NN trained in full-load condition is able to provide highly accurate predictions, with very low errors. For the ASE noise prediction the accuracy of the LU-NN trained with the full-load data-set is similar, as the average errors are 0.0052 dB and 0.0027 dB for *Error_{MAX}* and *RMSE*, respectively (Fig. 4(a)).

Afterwards, we consider the more realistic condition given of partial-load and we use such data-set to validate the LU-NN trained using the full-load data-set. Since this neural network is Load Unaware, we can not feed the load information as input, but only the pumps powers. Validation results are reported in Fig. 3(b) and show a dramatic worsening of the accuracy of predictions: mean values of *Error_{MAX}* and *RMSE* increase to 0.47 dB and 0.41 dB, respectively, with *Error_{MAX}* pdf reaching up to more than 2 dB. This means that the LU-NN trained with the full-load data-set is not able to provide accurate predictions of gain profiles when applied in partial load conditions. In Fig. 4(b) it can be observed a similar result also for ASE noise profile predictions, even if the accuracy is slightly better than for the gain case. In fact, the mean value of the *Error_{MAX}* is 0.26 dB and the mean value of the *RMSE* is 0.17 dB, showing that the noise profile is less dependent on the input load condition compared to gain profiles.

Before analyzing the prediction performance of the LA-NN, we test a second version of the LU-NN, training it with a partial-load data-set, to see if it can adapt to variable load conditions even if it is intrinsically load-agnostic. The LU-NN can not receive the load condition as input, but since a change in the input load determines a change in the gain profile, the NN learns an average behavior of different loads: this behaviour could improve predictions. The validation results are shown in Fig. 3(c) and we can observe a slight improvement in performance with the average *RMSE* and the average *Error_{MAX}* reducing to 0.19 dB and 0.22 dB, respectively. And also the pdf of *Error_{MAX}* now extends only to about 1.5 dB.

Similarly, in Fig. 4(c) we have the validation results for the ASE noise predictions. Also in this case, as for the gain case, there is a reduction in the prediction errors with average *RMSE* and average *Error_{MAX}* decreasing to 0.095 dB and 0.15 dB, respectively. Compared to the results obtained for gain predictions, the performance for the noise are better, still confirming a reduced dependence of the noise profile on different

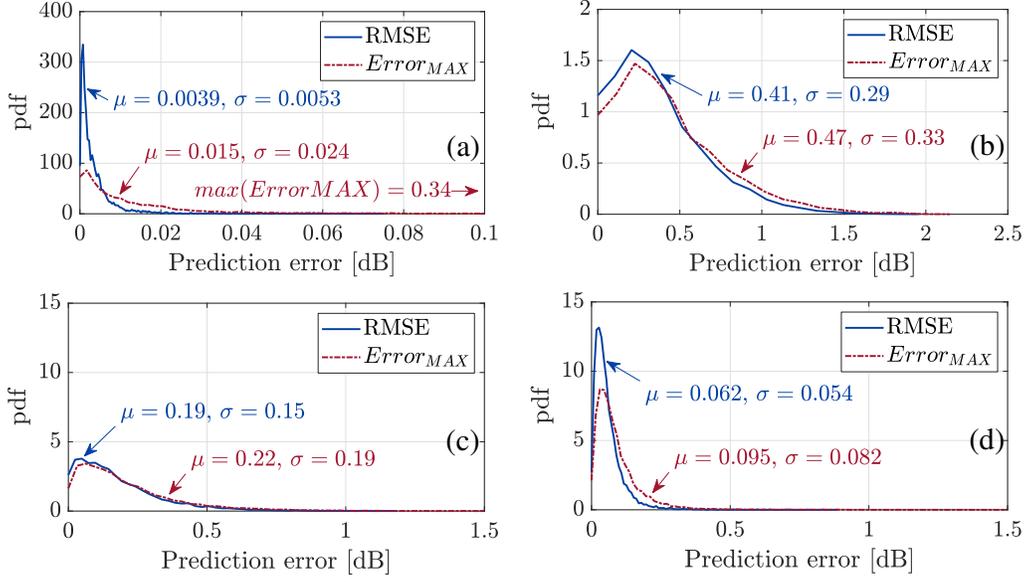


Fig. 3. Probability density functions of Raman gain prediction $RMSE$ and $Error_{MAX}$ for: (a) LU-NN trained and validated over full-load cases only, (b) LU-NN trained over full-load cases and validated over partial-load cases, (c) LU-NN trained and validated over partial-load cases and (d) LA-NN trained and validated over partial-load cases.

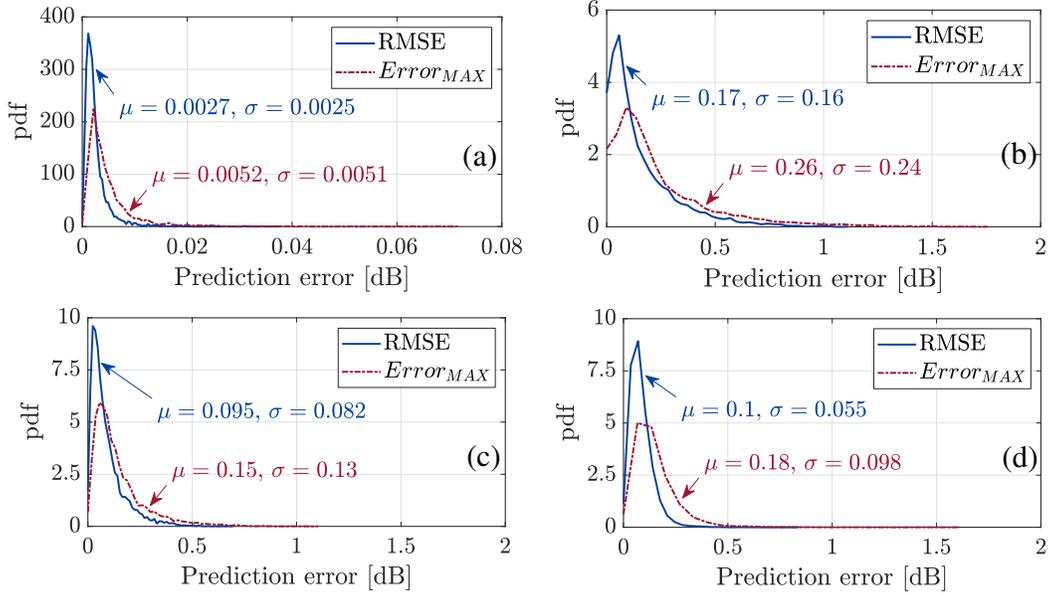


Fig. 4. Probability density functions of ASE noise prediction $RMSE$ and $Error_{MAX}$ for: (a) LU-NN trained and validated over full-load cases only, (b) LU-NN trained over full-load cases and validated over partial-load cases, (c) LU-NN trained and validated over partial-load cases and (d) LA-NN trained and validated over partial-load cases.

input loads.

From results discussed so far, it is clear that to fully recover the accuracy level in case of partial-load, in particular for the Raman gain prediction, a new NN model taking into account the information about the

input load condition is needed: we call this new model the Load Aware NN (LA-NN). After having trained the LA-NN using the partial-load data-set, we performed a comprehensive validation using the same type of data. In Fig. 3(d), we report the pdfs of both $Error_{MAX}$ and

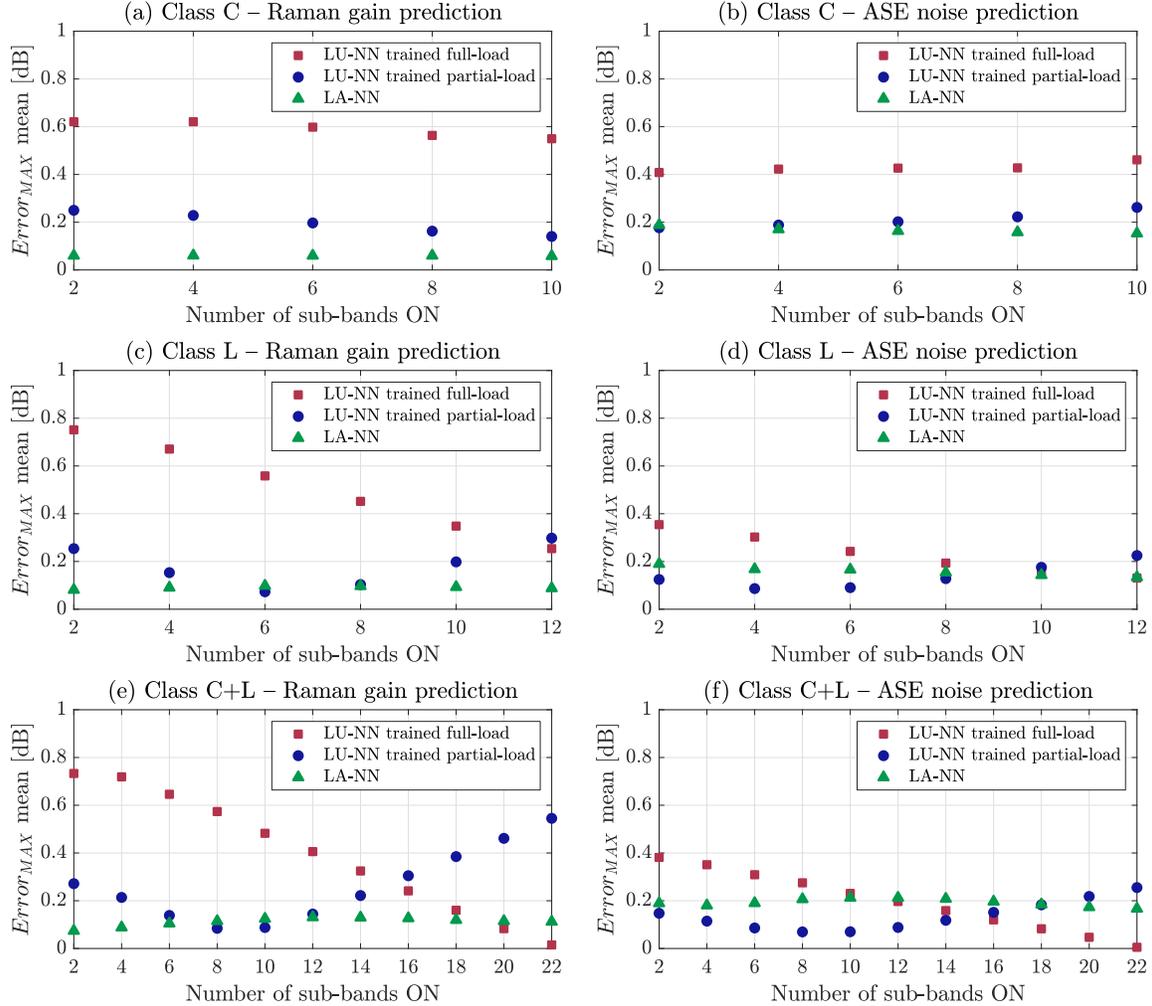


Fig. 5. Mean value of the $Error_{MAX}$ in dB with respect to the number of sub-bands turned ON for gain (on the left) and ASE noise (on the right) predictions. Each row corresponds to a load class: class C, class L and class C+L (from top to bottom). The comparison between the LU-NN trained with the full-load data-set, the LU-NN trained with the partial-load data-set and the LA-NN is shown.

$RMSE$ for the LA-NN in case of Raman gain prediction using the best NN model. The average $Error_{MAX}$ is 0.095 dB, and most of the values are below 0.5 dB.

Similar results are obtained validating the best model for the prediction of ASE noise profiles: see Fig. 4(d). The average values of $Error_{MAX}$ and $RMSE$ are 0.18 dB and 0.1 dB, respectively. Moreover, we can also observe that the performance of the LU-NN trained with the partial-load data-set and the LA-NN are similar. This means that the load information is not fundamental for the ASE noise prediction when operating in partial load condition.

In general, we can observe that when using the LA-NN it was not possible to achieve the same performance as the LU-NN trained and validated with full load only

(Figs. 3(a) and 3(d) for the gain and Figs. 4(a) and 4(d) for the noise) because when the load variations are considered, the dimension of the input space increases, thus the mapping is more complex.

At this point, we carried out a comprehensive statistical investigation on how the errors are distributed over each class (C, L and C+L) and sub-class (from 2 to 22 sub-bands ON) when we consider the partial load condition. Therefore, for all the three NN models studied (LU-NN trained with full-load data-set, LU-NN trained with partial-load data-set and LA-NN), the validation is carried out considering the partial-load data-set, properly pruned excluding all cases that at least for one sub-band present gains below 4 and above 21 dB, as already done in a previous section. Fig. 5 shows the mean value of

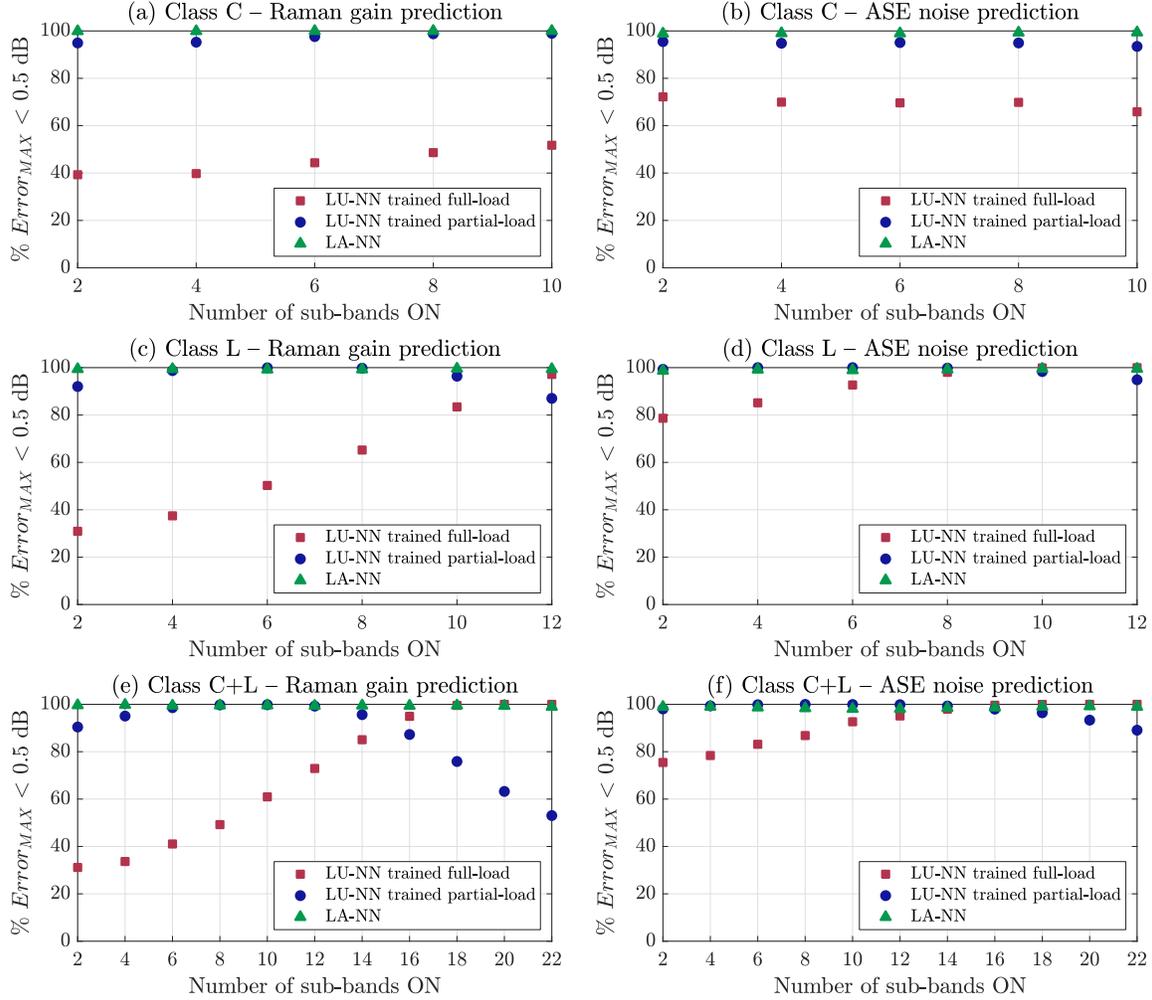


Fig. 6. Percentage of cases with $Error_{MAX}$ under 0.5 dB with respect to the number of sub-bands turned ON for gain prediction (on the left) and ASE noise prediction (on the right). Each row corresponds to a load class: class C, class L and class C+L (from top to bottom). The comparison between the LU-NN trained with the full-load data-set, the LU-NN trained with the partial-load data-set and the LA-NN is shown.

the $Error_{MAX}$ with respect to the number of sub-bands turned ON for the three load classes (class C, L and C+L, from top to bottom), for both gain (left column) and ASE noise profiles prediction (right column).

From Fig. 5(a) we can see that considering only sub-bands in C-band the average $Error_{MAX}$ is always larger than 0.5 dB when the LU-NN trained with the full-load data-set is used: it shows poor performance when working with input conditions different from full-load. Increasing the number of sub-bands turned ON from 2 to 10 sub-bands, the error reduces of just about 0.1 dB, because of the increased load. Instead, in case of LU-NN trained with partial-load data-set, as expected, the results improve with the mean of the $Error_{MAX}$ assuming values between 0.25 dB (when 2 sub-bands are ON)

and 0.14 dB (when 10 sub-bands are ON). Even better results are obtained with the LA-NN, for which the average $Error_{MAX}$ is always below 0.1 dB confirming the higher accuracy of this NN model.

For the ASE noise prediction in same condition (Fig. 5(b)), the accuracy of the LU-NN trained with full-load data-set is slightly better, as the average $Error_{MAX}$ is always lower than 0.5 dB, but it does not improve increasing the number of sub-bands ON. The LU-NN trained with the partial-load data-set and the LA-NN have similar performance, with the first one getting slightly worse with an increasing number of sub-bands turned ON.

Analyzing the gain prediction for the other two classes (Figs. 5(c) and 5(e)), we can see that in general, when

the LA-NN is used, the average $Error_{MAX}$ is always below 0.2 dB, clear sign of the capability of the LA-NN to adapt to partial loads. On the other hand, the LU-NN trained with full-load data-set shows poorer performance and it is characterized by the trend already discussed for class C: a reduction of errors when increasing the number of sub-bands ON. In fact, as we expect, the larger is the number of sub-bands turned ON, the condition is closer to the full-load case, and consequently we have a lower mean value of the $Error_{MAX}$. For full-load the LU-NN trained with full-load data-set is able to overcome the performance of the LA-NN showing an almost negligible error. Instead, the LU-NN trained with partial-load data-set shows a particular trend. In general, for low loads (2 and 4 sub-bands ON in case of Class L and 2, 4 and 6 in case of Class C+L) the average $Error_{MAX}$ assumes values pretty close to those provided by the LA-NN. Then, it increases dramatically when increasing the number of sub-bands ON. In particular for Class C+L, the average $Error_{MAX}$ reaches values of more than 0.5 dB. This means that the LU-NN trained with partial-load data-set is able to learn only the average effect of different loads on Raman gains and not all the variability. Only the LA-NN can effectively learn the actual dependence on input load and adapt to it.

Similar results are shown in (Figs. 5(d) and 5(f)) for ASE noise profiles predictions: now the average $Error_{MAX}$ is always under 0.4 dB for LU-NN trained with full-load data-set as well, but LA-NN is able to maintain such error almost below 0.2 dB. The LU-NN trained with partial-load data-set has similar performance to the LA-NN, with a slightly increase of the average $Error_{MAX}$ for increasing number of sub-bands turned ON. For both class L and C+L (Fig. 5(d) and Fig. 5(f) respectively), as already observed in case of gain profiles prediction, the average $Error_{MAX}$ decreases when the number of sub-bands ON increase when the LU-NN trained with full-load data-set is considered. In fact for the class L, from almost 0.4 dB, when only 2 sub-bands are ON, the mean $Error_{MAX}$ reduces to a about 0.1 dB. Instead, when using the LA-NN model, this value is almost constant and always below 0.2 dB. When more than half the total number of sub-bands are turned ON (from 12 to 22), the LU-NN trained with full-load data-set shows better accuracy than the LA-NN. This might be explained by the use of an almost error-free LU-NN model obtained from the hyper-parameters optimization and the simpler mapping. In fact, this model is able to provide such small errors when predicting noise profiles generated in a full load condition. And for ASE noise profiles such errors do not significantly increase even if

we reduce the number of sub-bands ON. In addition, we can observe that there is not such a big difference in the values assumed by the average $Error_{MAX}$ when we consider the three different NN models. As already said before, the ASE noise profiles does not significantly change with different input loads.

Besides looking only at the average $Error_{MAX}$, it is important to consider also how errors are distributed and in particular if there are substantial portions of the population in the tails of the pdfs. For this purpose, in Fig. 6 we evaluate the percentage of the cases of $Error_{MAX}$ below 0.5 dB, a reasonable threshold to guarantee accurate predictions. This analysis is performed on a class and sub-class basis, as previously done.

As expected, for Raman gain predictions, shown in Figs. 6(a), 6(c) and 6(e), the percentage of $Error_{MAX}$ cases below 0.5 dB increases with an increasing input load for the LU-NN trained with full-load data-set. For class C, it starts at 39% and it reaches about 52% when all 10 C-band sub-bands are ON, showing once again the limited performance of the LU-NN trained with full-load data-set.

Instead for classes L and C+L, it starts from about 30% of the cases increasing up to 97% of the cases for L-class and even reaching 100% for C+L class. In particular, the LU-NN trained with full-load data-set is able to provide predictions with all $Error_{MAX}$ cases below 0.5 dB even when the input WDM comb is not fully loaded, but only in case of 18 and 20 sub-bands ON. In general, the LU-NN trained with partial-load data-set instead is able to always provide more than 90% of cases with $Error_{MAX}$ lower than 0.5 dB, except for class C+L when the number of sub-bands ON is 16 or more. In this case, the percentage dramatically decreases down to 53% of cases when the full-load condition is met. On the other hand, the LA-NN model always guarantees high accuracy in predictions: in all three classes and for any load conditions it reaches almost 100% of the cases having an $Error_{MAX}$ lower than 0.5 dB, with very few cases with $Error_{MAX}$ larger than 0.5 dB. This means that most of the population experiences very small errors.

Considering the results related to the ASE noise prediction (Figs. 6(b), 6(d) and 6(f)), we observe that, the percentage of cases of $Error_{MAX}$ under 0.5 dB is always above 60% also for the LU-NN trained with full-load data-set. Although this is a higher value compared to the one obtained in case of gain predictions, there is a non negligible section of the cases experiencing large (≥ 0.5 dB) errors. Similarly to the case of gain predictions, using the LA-NN model we have a higher accuracy that

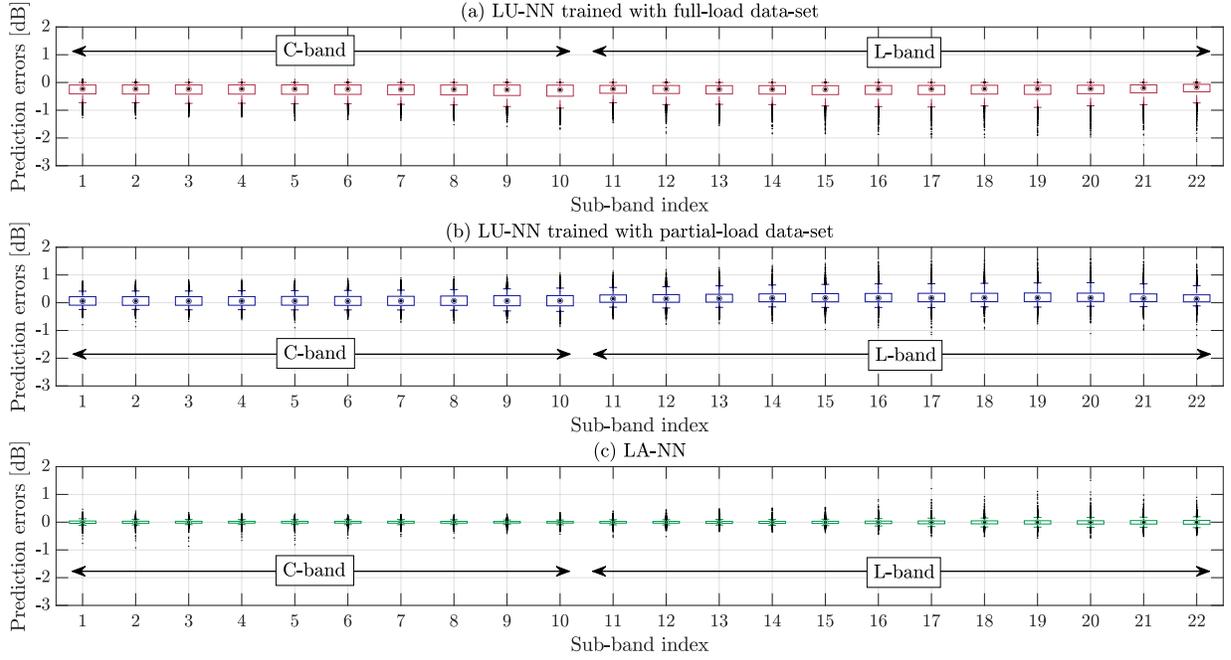


Fig. 7. Distribution of the *Raman gain* prediction errors of the 11000-cases partial-load data-set with respect to the sub-bands position when using (a) the LU-NN trained with full-load data-set, (b) the LU-NN trained with partial-load data-set and (c) the LA-NN. Sub-bands are ordered in wavelengths. Boxes and whiskers contain 50% and 90% of the cases, respectively. The remainder of the cases are plotted as individual markers.

allows to reach almost the 100% of cases having an $Error_{MAX}$ lower than 0.5 dB. Also the LU-NN trained with partial-load data-set is able to provide pretty good performance over the three classes, as the percentage of cases with $Error_{MAX} < 0.5$ dB is always greater than 89%.

After the analysis of the trend of the $Error_{MAX}$ with respect to the load, i.e. number of sub-bands turned ON, we study the spectral distribution of prediction errors. Defining such errors as the difference between the predicted profile and the target one, we perform a comprehensive statistical analysis considering all 11000-cases of the partial load data-set, i.e. the data-set obtained by merging the 500-cases data-set of each class and sub-class (transmission band and number of sub-bands turned ON for each class).

In Figs. 7(a), 7(b) and 7(c), we report the spectral distribution for gain prediction errors when the LU-NN trained with full-load data-set, the LU-NN trained with partial-load and the LA-NN are used, respectively. The figures show the prediction error distribution for the 22 sub-bands: sub-bands are identified by an index from 1 to 22, ordered in the wavelength domain from C to L band. This means that the sub-band with index 1 is the one at lowest wavelength (center at 1532.6 nm)

and the sub-band with index 22 is the one at highest wavelength (center at 1619.4 nm). For each sub-band the box captures 50% of the cases in the data-set, between 25% and 75% of the cdf of errors. Whiskers capture 90% of the cases, going from the 5% to the 95% of the cdf. The marker inside the box indicates the median value. Outside black dots markers correspond to outliers, i.e. predictions in the tails of the pdf of errors exceeding the 90% of cases captured by whiskers.

Analyzing results for the LU-NN trained with full-load data-set (Fig. 7(a)), we can observe that most of the errors are negative. In fact, such NN has been trained in a full load condition, with a fixed input power level. In case of partial loads, the input power level is usually lower than the state considered in training, because not all the sub-bands are ON. A lower input only maintains or increases the gain. Consequently, most of the predicted gains are lower than the target gains, and the errors are negative. Moreover, we also observe that the tails of the errors are larger in the upper region of the L-band, at longer wavelengths. Median values range between -0.26 dB and -0.17 dB without a clear dependence on wavelength. Box sizes have an extension ranging from 0.27 dB in sub-band 22 to 0.40 dB in sub-band 10, and whiskers range is between 0.73 dB in sub-band 2 and

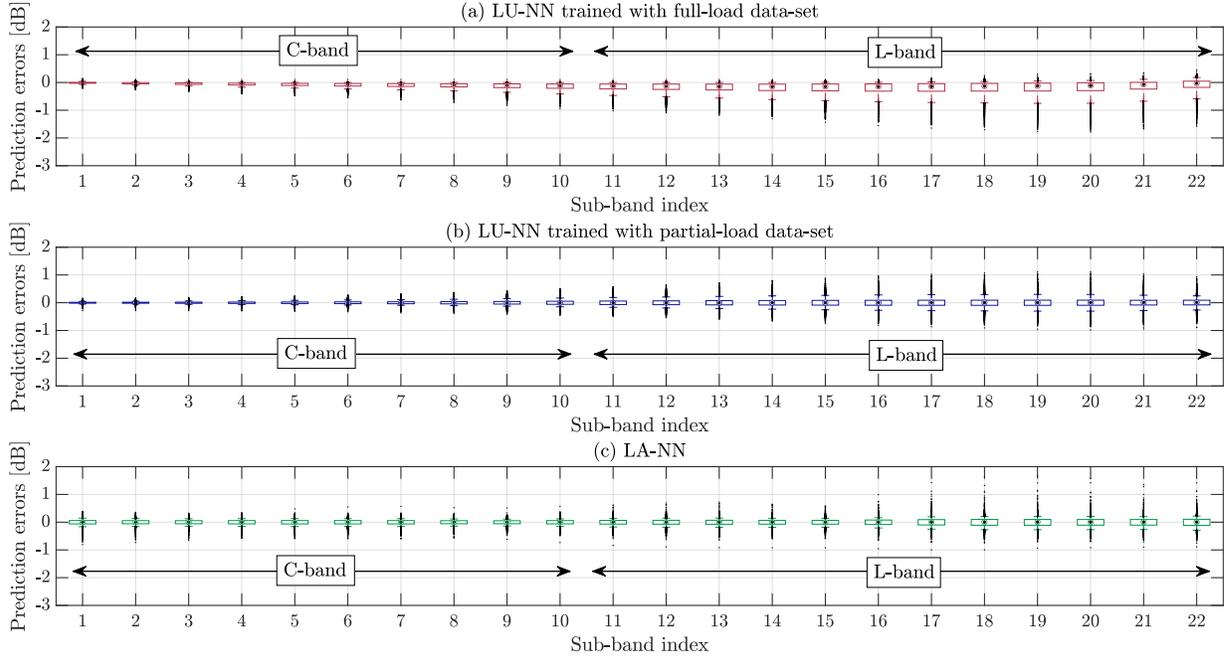


Fig. 8. Distribution of the ASE noise profile prediction errors of the 11000-cases partial-load data-set with respect to the sub-bands position when using (a) the LU-NN trained with full-load data-set, (b) the LU-NN trained with partial-load data-set and (c) the LA-NN. Sub-bands are ordered in wavelengths. Boxes and whiskers contain 50% and 90% of the cases, respectively. The remainder of the cases are plotted as individual markers.

0.92 dB in sub-band 10. Outliers extend almost only in the bottom part, reaching a maximum error of -2.25 dB in sub-band 21.

Considering the LU-NN trained with partial-load data-set (Fig. 7(b)), the prediction errors are no longer mainly negative, as the median assumes values between 0.05 dB and 0.18 dB. Similarly to the LU-NN trained with full-load condition, the box sizes range from 0.25 dB in sub-band 22 to 0.37 dB in sub-band 10, the whiskers range from 0.66 dB in sub-band 1 to 0.87 dB in sub-band 19. The outliers do not exceed -0.92 dB and 1.50 dB, in sub-bands 2 and 20 respectively.

Moving to Fig. 7(c) where we report the gain prediction errors for the LA-NN, at a glance we have a further proof of the higher accuracy of this new NN when dealing with partial loads compared to both the LU-NNs analyzed. In this case, the median values are all almost equal to 0 dB, and most significantly, the size of the boxes and whiskers is much reduced with respect to the LU-NN case, being the maximum 0.14 dB and 0.39 dB, respectively. In this case, errors for most of the cases does not show a spectral dependence and distributions are not biased. Only for the outliers we observe a slight spectral dependence: their top range has a broader extension in the upper portion of the

L-band. For the LA-NN almost all outliers fall in the range ± 1 dB, again indicating an improved adaption to partial load of this NN compared to the LU-NN where maximum errors extend beyond -2 dB.

In general, for all the three NN models, we can see that prediction errors tend to be larger in the L-band, especially for long wavelengths (high sub-band indices). This might be related to higher order interactions between signal and pumps due to SRS.

Similar considerations can be drawn observing plots for the case of the ASE noise prediction errors reported in Figs. 8(a), 8(b) and 8(c), for LU-NN trained with full-load data-set, LU-NN trained with partial-load data-set and LA-NN models, respectively. Nevertheless, as already observed in previous analysis, unlike the prediction of the Raman gain profiles, the errors in the prediction of the ASE noise profiles are usually smaller, specially for the C-band, even when using the LU-NN trained with full-load data-set. This is again the sign of a reduced dependence of the noise profiles with respect to the input load conditions. In fact, as it can be seen in Fig. 8(a) for the LU-NN trained with full-load data-set, the median values vary from -0.15 dB and -0.01 dB, and the maximum box size and whisker range are reduced to 0.28 dB and 0.82 dB. Similar results are observed for the

LU-NN trained with partial-load data-set, for which the errors extend in the positive region as well. In this case the median value is almost equal to 0 dB, the maximum box size and whisker range are 0.19 dB and 0.60 dB and the outliers assume values from -1 dB to 1.66 dB in sub-bands 17 and 19, respectively. Also for ASE noise predictions the errors are higher in the upper part of the L-band and show a bias toward negative values, for the same reason discussed above for gains.

On the contrary, in Fig. 8(c) when the LA-NN is used, the trend is very similar to the one observed for the case of gain prediction, with errors more uniformly distributed over each sub-band. Also in this case, just the outliers in the upper portion of the L-band are showing slightly larger errors with respect to those in the other spectral regions, but they again fall almost all in the range ± 1 dB.

As already discussed in previous section, for ASE noise prediction, the performance of the LU-NN trained with partial-load data-set and of the LA-NN are quite similar.

IV. CONCLUSION

We have introduced a machine learning based framework for ultra-fast and accurate predictions of gain and ASE noise profiles of a C+L-band Raman amplifier with any load conditions. The proposed technique is faster compared to standard approaches based on the numerical solution of ODEs describing the physical effect because NNs have a lower complexity.

Starting with the analysis of the Load Unaware model (LU-NN) trained with full-load data-set we have validated it with a data-set of the same type, obtaining both average $Error_{MAX}$ and standard deviation below 0.03 dB. Similar results have been demonstrated for the prediction of the ASE noise profiles.

At the same time, we have shown how LU-NNs are not able to deliver high accuracy when we assume a partial load data-set in validation: the average $Error_{MAX}$ is increased to 0.47 dB and 0.26 dB, for gain and ASE noise profiles, respectively. For this reason we have trained this LU-NN with the partial-load data-set, but also in this case it was not able to predict Raman gains with high accuracy when the input load condition changes, as the average $Error_{MAX}$ is 0.22 dB. Instead, better results have been found for the ASE noise predictions, showing a smaller dependency with respect to different loads.

Finally, we have introduced an advanced model, the Load Aware NN (LA-NN), a neural network model that takes into account the arbitrariness of the input load as well, to be able to deliver highly accurate predictions also under partial load. Specifically, for the prediction

of Raman amplification gains, we have obtained an average $Error_{MAX}$ always lower than 0.2 dB and in almost 100% of cases the $Error_{MAX}$ is under 0.5 dB. Comparable trends have been found also for ASE noise predictions, even if in this case the LU-NN trained with full-load data-set and partial-load data-set suffer less from partial load conditions showing reduced errors.

Analyzing the gain prediction error distributions as a function of sub-bands location, we have found that, when using the LA-NN, the errors are small independently on the spectral positioning. Only the distribution tails show slightly larger errors in the L-band. Similar trends are observed for the prediction of ASE noise profiles, for which we have found very low errors in the C-band using the LU-NN as well (trained both with full-load and partial-load data-sets).

Due to the reduced complexity of the RP approach, only a few seconds are needed for training a single NN and less than one second is required for testing.

This tool will be required in the control-plane of future multi-band networks to allow real-time analysis of Raman amplifiers. Our approach is a viable solution for practical implementation in network controllers. Other applications can be in the design and optimization of multi-band Raman amplifiers, as support for fine tuning design achieved through machine learning inverse design or as a fast solver for standard optimization methods.

ACKNOWLEDGMENT

This project has received funding from the European Research Council through the ERC-CoG FRECOM project (grant agreement no. 771878) and by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754462.

REFERENCES

- [1] D. Rafique and L. Velasco, "Machine learning for network automation: overview, architecture, and applications [Invited Tutorial]," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 10, no. 10, pp. D126–D143, October 2018.
- [2] S. Zhang, F. Yaman, K. Nakamura, T. Inoue, V. Kamalov, L. Jovanovski, V. Vusirikala, E. Mateo, Y. Inada, and T. Wang, "Field and lab experimental demonstration of nonlinear impairment compensation using neural networks," *Nature Communications*, vol. 10, no. 1, 2019.
- [3] 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," *Applied Sciences*, vol. 10, no. 1, 2020.
- [4] F. N. Khan, Q. Fan, C. Lu, and A. P. T. Lau, "An Optical Communication's Perspective on Machine Learning and Its Applications," *Journal of Lightwave Technology*, vol. 37, no. 2, pp. 493–516, January 2019.

- [5] 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," *Journal of Optical Communications and Networking*, vol. 12, no. 6, pp. C1–C11, June 2020.
- [6] A. Napoli, N. Costa, J. K. Fischer, J. Pedro, S. Abrate, N. Calabretta, W. Forsyia, E. Pincemin, J. P.-P. Gimenez, C. Matrakidis, G. Roelkens, and V. Curri, "Towards multiband optical systems," in *Advanced Photonics*, 2018, p. NeTu3E.1.
- [7] M. N. Islam, *Raman amplifiers for telecommunications 1: physical principles*. Springer-Verlag, 2004.
- [8] J. Renaudier, A. Arnould, A. Ghazisaeidi, D. L. Gac, P. Brindel, E. Awwad, M. Makhsiyani, K. Mekhazni, F. Blache, A. Boutin, L. Letteron, Y. Frignac, N. Fontaine, D. Neilson, and M. Achouche, "Recent Advances in 100+nm Ultra-Wideband Fiber-Optic Transmission Systems Using Semiconductor Optical Amplifiers," *Journal of Lightwave Technology*, vol. 38, no. 5, pp. 1071–1079, 2020.
- [9] E. Desurvire, J. R. Simpson, and P. C. Becker, "High-gain erbium-doped traveling-wave fiber amplifier," *Optics Letters*, vol. 12, no. 11, pp. 888–890, 1987.
- [10] J. Bromage, "Raman Amplification for Fiber Communications Systems," *Journal of Lightwave Technology*, vol. 22, no. 1, pp. 79–93, January 2004.
- [11] B. Neto, A. L. J. Teixeira, N. Wada, and P. S. Andre, "Efficient use of hybrid Genetic Algorithms in the gain optimization of distributed Raman amplifiers," *Optics express*, vol. 15, no. 26, pp. 17 520–17 528, 2007.
- [12] D. Zibar, A. Ferrari, V. Curri, and A. Carena, "Machine learning-based Raman amplifier design," in *Optical Fiber Communication Conference (OFC) 2019*. Optical Society of America, 2019, p. M1J.1.
- [13] 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 Amplifier Case," *Journal of Lightwave Technology*, vol. 38, no. 4, pp. 736–753, 2020.
- [14] M. Ionescu, "Machine learning for ultrawide bandwidth amplifier configuration," in *International Conference on Transparent Optical Networks, ICTON*, 2019, p. We.B7.3.
- [15] J. Chen and H. Jiang, "Optimal Design of Gain-Flattened Raman Fiber Amplifiers Using a Hybrid Approach Combining Randomized Neural Networks and Differential Evolution Algorithm," *IEEE Photonics Journal*, vol. 10, no. 2, pp. 1–15, 2018.
- [16] A. M. Rosa Brusin, V. Curri, D. Zibar, and A. Carena, "An ultra-fast method for gain and noise prediction of Raman amplifiers," in *European Conference on Optical Communication, ECOC*, 2019, p. Th.1.C.3.
- [17] X. Ye, A. Arnould, A. Ghazisaeidi, D. L. Gac, and J. R. and, "Experimental prediction and design of ultra-wideband Raman amplifiers using neural networks," in *Optical Fiber Communication Conference (OFC) 2020*. Optical Society of America, 2020, p. W1K.3.
- [18] A. M. Rosa Brusin, U. C. de Moura, A. D'Amico, V. Curri, D. Zibar, and A. Carena, "Load aware Raman gain profile prediction in dynamic multi-band optical networks," in *Optical Fiber Communication Conference (OFC) 2020*. Optical Society of America, 2020, p. T4B.3.
- [19] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.
- [20] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, December 2006.
- [21] GNPY. [Online]. Available: DOI: 10.5281/zenodo.3458320, <https://github.com/Telecominfraproject/oopt-gnpy>
- [22] G. Bosco, V. Curri, A. Carena, P. Poggiolini, and F. Forghieri, "On the performance of Nyquist-WDM Terabit Superchannels based on PM-BPSK, PM-QPSK, PM-8QAM or PM-16QAM subcarriers," *Journal of Lightwave Technology*, vol. 29, no. 2, pp. 53–61, January 2011.
- [23] T. Zami, "Co-Optimizing Allocation of Nyquist Superchannels and Physical Impairments Aware Placement of Regenerators in Elastic WDM Networks," *Journal of Lightwave Technology*, vol. 32, no. 16, pp. 2830–2840, January 2014.