Fast and accurate nonlinear interference in-band spectrum prediction for sparse channel allocation

Isaia Andrenacci, Matteo Lonardi, Petros Ramantanis, Elie Awwad, Ekhine Irurozki, Stéphan Clémençon

Fast and accurate nonlinear interference in-band spectrum prediction for sparse channel allocation

Isaia Andrenacci
Physics of Optical Networks,
LTCI
Nokia Bell Labs, Télécom Paris
Nozay, France
isaia.andrenacci@nokia.com

Matteo Lonardi
Physics of Optical Networks
Nokia Bell Labs
Nozay, France
matteo.lonardi@nokia-bell-labs.com

Petros Ramantanis
Physics of Optical Networks
Nokia Bell Labs
Nozay, France
petros.ramantanis@nokia-bell-labs.com

Élie Awwad
LTCI
Télécom Paris
Palaiseau, France
elie.awwad@telecom-paris.fr

Ekhiñe Irurozki
LTCI
Télécom Paris
Palaiseau, France
irurozki@telecom-paris.fr

Stephan Clémençon
LTCI
Télécom Paris
Palaiseau, France
stephan.clemencon@telecom-paris.fr

***Abstract*— We propose and numerically evaluate a machine-learning-based nonlinear interference spectrum estimator for a coherent optical network. The solution shows a root-mean-squared error of about 0.13 dB compared with split-step Fourier simulation when estimating the nonlinear interference variance.**

Keywords—optical network, machine learning, fiber nonlinearity, nonlinear interference power spectral density

I. Introduction

Accurate prediction of the quality of transmission is critical for fine-tuning margins in monitoring-assisted optical networks. For this, several solutions are possible. One extreme is the split-step Fourier method (SSFM) [1] which allows accurate prediction of the waveform distortions and, consequently, the bit error ratio of a lightpath at the expense of high computational complexity. At the other extreme, the Gaussian noise model [2], in its simplest form, allows for a fast prediction of the nonlinear interference (NLI) variance at the price of a reduced accuracy when the main assumptions underlying the model are not met. Somewhere in between, the enhanced Gaussian noise (EGN) model [3] offers improved accuracy with an additional computational cost. At the same time, the technique discussed in [4] allows for a very fast assessment of the in-phase (IP) and quadrature (Q) NLI power spectral densities (PSDs), where the PSDs account for the fine impact of the receiver digital signal processing (DSP), most notably the carrier phase estimation (CPE).

Nevertheless, the computational complexity of EGN remains prohibitive for real-time applications, while the fixed-channel loading calibration step required in [4] might be impractical, particularly for heterogeneous systems. As an alternative, [5] proposes a machine learning (ML) model to reduce complexity and improve the performance prediction accuracy, however, without spectral characterization capabilities. Moreover, in [6], ML has been used to get the signal waveform after propagation, yet requiring different training for every link.

In this paper, to the best of our knowledge, we propose the first attempt using ML to achieve a low-complexity prediction of the NLI IP and Q PSDs with sparsely loaded spectra and a variety of transmission distances employing multi-polynomial regression [9]. Such an estimator is useful for typical scenarios of terrestrial optical networks. The paper is organized as follows: Section II reports the collection data phase and description of the adopted ML model. Section III will address the accuracy of this proposal for predicting the nonlinear spectrum shape and its variance. In Section IV, we investigate the ML proposal for short links with a few spans, and finally, in Section V, we draw our conclusion.

II. Methodology and Data Collection

In the first part of this section, we will address the data collection phase to construct two labeled datasets for PSD components. Then, we will describe the ML model and the learning procedure in the last three subsections. Finally, we will draw comments on the accuracy and complexity of this ML proposal.

A. Data Generation

Fig. 1 shows the SSFM simulation setup we carried out during the data generation stage. We considered a grid of 21 dual-pol QPSK channels at the transmitter (TX) with a fixed spacing Δf of 37 GHz. We used the central channel as a channel under test (CUT), which was always kept "on." At the same time, the interfering channels are randomly switched "on", and "off" in each of the $M = 691$ simulations. The launch power at

978-3-903176-54-6 © 2023 IFIP

the span ingress was 0 dBm. The transmitted channels had a root-raised cosine spectral shaping with a roll-off of 0.01

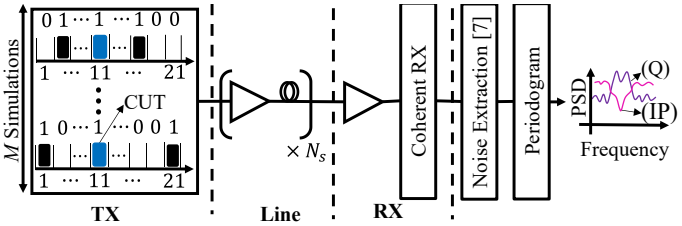

Fig. 1: SSFM simulations setup with different sparse channel allocation and link lengths used to collect the IP and Q PSDs.

and a symbol rate of 32 GBd. The number of simulated symbols in each run was 576000, with 54 samples per symbol. We considered dispersion unmanaged links with $N_{\text{spans}} = 50$ different lengths. All the simulations propagated through N_s spans for $N_s = 1, \ldots, N_{\text{spans}}$ of 100 km standard single-mode fiber and noiseless flat-gain amplifiers to focus on the NLI impairment. In the receiver (RX) section of Fig. 1 we first used a front end to down-convert the optical waveform of the CUT into the electric domain. Second, we applied a matched filter and an ideal chromatic dispersion compensation. We down-sampled to 1 sample per symbol and recovered the phase in a data-aided way, averaging over the transmitted sequence length. Third, we conducted a data-aided noise statistic evaluation to recover the imaginary and real parts of the variance, as explained in [7]. Finally, we applied the Welch modified periodogram method without overlap and a window length of 128 to evaluate the IP and Q components of the NLI PSD, thus obtaining two vectors $\mathbf{y}^{\mathbf{Q}}$ and $\mathbf{y}^{\text{IP}}$ with elements y_k for $k = 1, \ldots, N$, with $N = 256$ points. Finally, the total spectrum was recovered by summing up its two components.

B. Labeled Datasets Generation

We generate two different datasets, one for the IP and the other for the Q component. Let us introduce some notation first. The cross-channel NLI variance induced by an interfering channel to the CUT is approximately inversely proportional to the frequency spacing between the interferer and the CUT [8]. Following this idea, we define $\rho = (\mathbf{w} \cdot \mathbf{c})/\|\mathbf{w}\|$ where $\mathbf{w} = [1/10, \cdots 1/2, 1, 0, 1, 1/2, \cdots 1/10]/\Delta f$ is the weight vector, and $\mathbf{c} = [c_1, \cdots, c_m, \cdots, c_{21}]$ is the binary vector representing the transmitted channel allocation with $c_m = 1$ for "channel on" and $c_m = 0$ for "channel off", where $m = 1, \ldots, 21$. Therefore, the variable ρ describes the importance of the NLI due to the channel allocation, i.e., higher ρ corresponds to higher NLI. In addition, we define a normalized span number $n_s = N_s/N_{\text{spans}} - 1$ to have the input data in the same range for the ML algorithm. We define the input features vector $\mathbf{x} = [\rho, n_s, (n_s)^2, \cdots (n_s)^d]$, where d is the polynomial degree. Meanwhile, we normalized the two collected IP and Q PSDs by the cube of the CUT transmitted power referred to in the following as normalized PSD with the unit of measure of $\text{dB}(\text{mW}^{-2}/\text{Hz})$. This procedure permits us to have one PSD describing all the possible spectra with different CUT power levels. Finally, we construct two datasets, one for IP and one for Q labeling all the input vectors to the corresponding PSD in the natural logarithm scale. Therefore, the first dataset is formed by the pairs $(\mathbf{x}, \mathbf{y}^{\text{Q}})$, and the second one is formed by $(\mathbf{x}, \mathbf{y}^{\text{IP}})$.

C. Model

We used a different multi-polynomial regressor [9] for each of the N frequency bins representing the PSD. Hence, we assume that there is polynomial relation between $\mathbf{x}$ and y_k for each $k = 1, \ldots, N$, given by the equation $y_k = b_{k,0} + b_{k,1}\rho + b_{k,2}n_s + b_{k,3}n_s^2 + \cdots + b_{k,d+1}n_s^d$ where $b_{k,0}$ is the intercept and $b_{k,i}$ are the regression coefficients with $i = 1, \ldots, d+1$. We applied this model for both datasets since their inputs and outputs have similar relations.

D. Training Stage

We trained with 80% of the database, i.e., 552 simulations drawn randomly from all the simulations. In the learning stage we estimate the $\hat{b}_{k,0}$ and the $\hat{b}_{k,i}$ regression coefficients for each PSD frequency bin at fixed d, that best fit the data in terms of mean squared error. For each PSD component, we repeat this stage for twenty datasets with different polynomial degrees for $d = 1, \ldots, 20$.

E. Test Stage

We utilized the remaining 20% of the data for the testing phase and measured their prediction accuracy. To this end, we used the L2 norm defined as $\text{L2} = [\sum_{k=1}^{N}(y_k - \hat{y}_k)^2]^{1/2}$ where y_k is the bin frequency PSD evaluated with SSFM and $\hat{y}_k = \hat{b}_{k,0} + \hat{b}_{k,1}\rho + \hat{b}_{k,2}n_s + \hat{b}_{k,3}n_s^2 + \cdots + \hat{b}_{k,d+1}n_s^d$ is the one predicted through ML. As reported in the next section, we repeat this accuracy investigation for all degrees to find the optimum d giving the lowest distance.

Fig. 2 illustrates how we use a test set sample to predict the NLI PSDs. The upper part of Fig. 2 shows the feeding process of the same input features vector into all regressors that give a vector of points in the natural logarithmic scale, and we pass it to the dB scale to obtain the $\hat{\mathbf{y}}^{\text{Q}}$ and $\hat{\mathbf{y}}^{\text{IP}}$ PSDs.

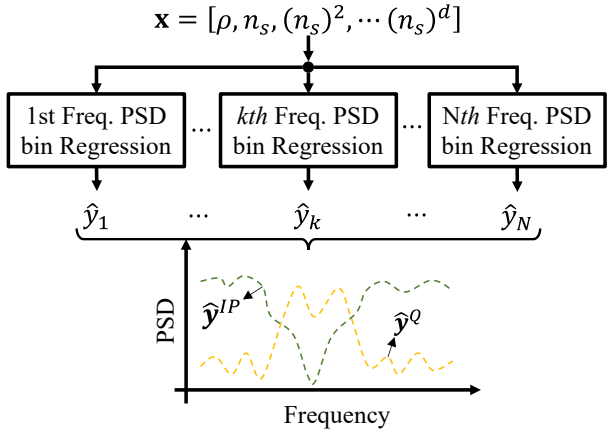

Fig. 2: Different N regressors to predict the NLI PSDs from an input feature vector $\mathbf{x}$ describing the channel allocation and the number of spans.

978-3-903176-54-6 © 2023 IFIP

III. Nonlinear Spectrum Prediction and Variance Estimation

In the first part of this section, we present an analysis of the shape of the PSD with two representative examples, and we assess the prediction accuracy as a function of the link length. In the second part, we report an investigation on the estimation of NLI variance.

A. Nonlinear Power Spectral Density Prediction

We provide two representative examples to show the prediction capabilities of this approach. Fig. 3(a) depicts a case in the test set of a link with one span, and $\rho = 0.43$. The x-axis represents the frequency, while the y-axis reports the normalized PSD values. On the left-hand side, we show the total PSD, where the solid line is the one evaluated with SSFM and the dashed line is the ML prediction, while on the right-hand side, we show the IP and Q PSDs. For the example of Fig. 3(a), the error is mainly found either at the edges or in the center of the spectrum. With the same structure we report in Fig. 3(b), an example of the PSD prediction for ρ = 0.43 and fifty spans. The error in the second example is less important than in the first example and lies mainly in the central frequency.

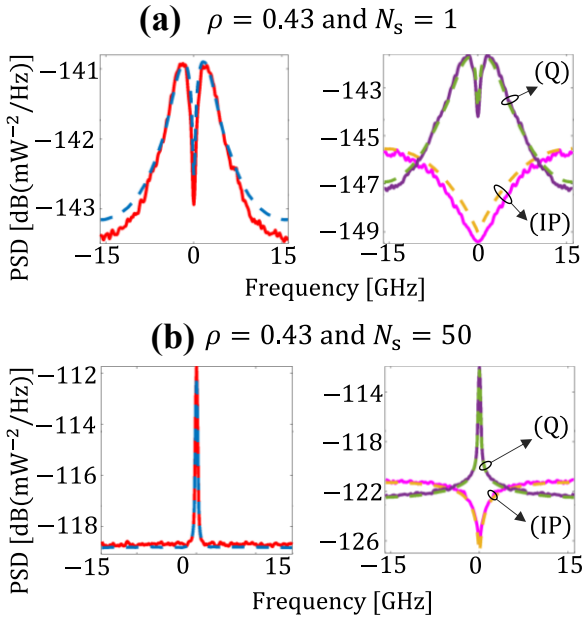

Fig. 3: (a) illustrates an example of PSD prediction for a link with ρ=0.43 for $\boldsymbol{N_s = 1}$ span and (b) with $\boldsymbol{N_s = 50}$ spans. Solid lines represent the PSDs evaluated through SSFM and dashed lines represent PSDs predicted by ML. Total PSDs (IP+Q) are reported on the left side, while the IP and Q PSDs are illustrated on the right.

We train and test the N regressors with different polynomial degrees d to find the one that leads to the lowest L2 norm, as mentioned in the previous section. Fig. 4(a) depicts the average distance for the IP (solid line) and Q (dashed line) components as a function of the polynomial degree. The different polynomial degrees are reported on the x-axis, while the y-axis represents the average L2 norm over the whole test set. We notice a decreasing behavior for both situations, with the lowest average L2 obtained for the Q component. In particular, the average distances associated with IP and Q reach a plateau at $d = 16$, respectively, around 2.30 dB and 2.19 dB.

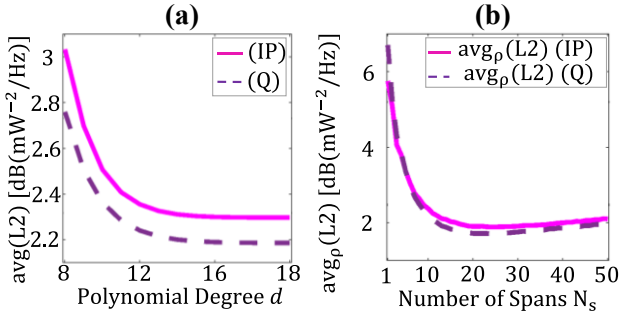

Fig. 4:(a) represents the average L2 norm over all the simulations test sets as a function of the polynomial degree. (b) depicts the average L2 norm over all different $\boldsymbol{\rho}$ in the test set as a function of the number of spans at $\boldsymbol{d = 16}$.

At this fixed degree $d = 16$, we investigate the average error distances as a function of link lengths to highlight the difference displayed in the previous examples of the prediction. To this end, we divided the test set into fifty subsets, i.e., one for each span number. Fig. 4 (b) depicts the average L2 norm over all the ρ in each subgroup representing the IP and Q PSD link length. The left y-axis represents the average distance over ρ, while the number of spans is reported on the x-axis. We obtain a larger distance for the first spans and a plateau after the tenth, which can be translated into two working regimes. The first one below or equal to ten spans has a decreasing average distance. The average L2 for IP PSD in this region has a maximum of 5.75 dB and a minimum of 2.38 dB, while 6.7 dB and 2.22 dB can be found for the Q component. A second regime above the tenth span, with a plateau around 2 dB and 1.84 dB, respectively, for the IP and Q PSD, with both minima located in the twenty-fourth span. Since it may not be evident how the differences in terms of L2 impact performance, in the following subsection, we study the impact of this ML proposal on the estimation of the nonlinear variance.

B. Nonlinear Variance Estimation

To understand the accuracy of PSD prediction, discussed in the previous Subsection III.A we show the error for estimating the a_{NL}, i.e., the integral of total normalized PSD.

Fig. 5(a) illustrates with a dashed line the root-mean-square-error (RMSE) and, with a solid one, the maximum absolute error (MAE) as a function of the number of spans. As before, we evaluated the two error metrics over the different ρ of the test database to have an error as a function of the number of spans. Both RMSE and MAE capture the mismatch between a_{NL} evaluated from SSFM and $\hat{a}_{NL}$ calculated integrating the predicted PSD. Also, in this case, we found two working regimes. For the RMSE, we have a decreasing relation from 0.39 dB down to 0.11 dB in the first regime, i.e., for less or equal than ten spans. In the second one, there is slightly

978-3-903176-54-6 © 2023 IFIP

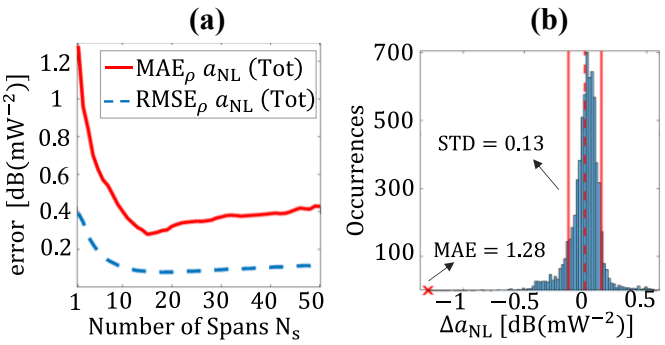

Fig. 5: (a) The y axis reports RMSE for $\boldsymbol{a_{\mathrm{NL}}}$ estimation versus different link lengths. (b) illustrates the histogram of the $\boldsymbol{a_{\mathrm{NL}}}$ estimation error for all the simulations in the test set.

increasing behavior towards the last span from a minimum of 0.08 dB for the twentieth span up to 0.12 dB.

Meanwhile, the MAE goes from 1.28 dB down to 0.36 dB in the first regime. The second regime shows a slightly increasing trend up to the 0.42 dB corresponding to the fiftieth span. These two curves help us clarify the link between the average distance for the PSDs components and the total nonlinear variance estimation error. For instance, in this specific case, a distance L2 of approximately 6 dB for the first span can be translated into an average a_{NL} estimation error of 0.4 dB, while we neglect here the impact of the carrier phase estimation on the NLI quadrature component [4].

To have a complete picture of the error statistics, we report in Fig. 5(b) the histogram of the a_{NL} estimation error for all the simulations in the test set. The x-axis reports the estimation error, i.e., $\Delta a_{NL} = a_{NL} - \hat{a}_{NL}$., while the y-axis represents the number of times this error occurs. The dashed line denotes the average error of -0.007 dB. Meanwhile, the standard deviation is 0.13 dB (solid line) over all spans and ρ of the test dataset. The estimation error can reach a maximum value larger than 1dB due to the shorter link lengths, as shown in Fig. 5(a).

C. *Comments on the Accuracy and Complexity*

These two previous investigations into the accuracy of this technique show two working regimes, the first relating to the first ten spans and the second to larger distances. In the first regime, we obtained a prediction worse than the second with an MAE less than 1.2 dB in the estimate of the variance of the NLI. On the contrary, we got an MAE below 0.42 dB in the second regime for the fiftieth span.

In terms of complexity, the main advantage of the polynomial regression strategy is that it only requires the computation of a simple polynomial for each frequency bin N of the PSD. This leads to a computation time of milliseconds for each link type, making it suitable for real-time applications. On the contrary, in the traditional EGN model, it is not possible to reach such low computational time since it has to calculate nested integrals to obtain the PSD of the NLI.

IV. Investigation of the Prediction for Short Links and the Impact of the Training Dataset

In Section III, we observe two working regimes, one below or equal to the tenth span and another above it. The results showed a larger average distance on the first regime with the higher value positioned in the first span. This can be translated into an RMSE of 0.4 dB for estimating the a_{NL} coefficient from the total PSD.

Hence, this behavior raises a question of how we can reduce the error for a small number of spans. In this section, we describe a simple strategy we adopt to see if it is possible to decrease this error. We compare three datasets with different maximum span lengths without changing the ML algorithm.

We construct two smaller datasets with N_{spans} equal to 10 and 20 from the entire dataset. Therefore, we select the samples with desired maximum span length to obtain two smaller ones, respectively, the dataset with $N_{spans} = 10$ and $N_{spans} = 20$.

Before looking at the accuracy results, we first found the convergence degree for these new datasets. Fig. 6(a) illustrates the average distance for the IP (solid line) and Q (dashed line) components as a function of the polynomial degree. The y-axis represents the average L2 norm overall test set with $N_{spans} = 10$, while the x-axis describes the different polynomial degrees.

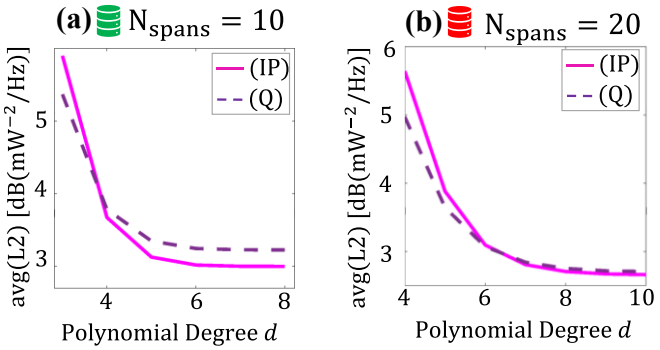

Fig. 6: (a) average L2 norm over all the simulations test datasets with $\boldsymbol{N_{spans}} = 10$ as a function of the polynomial degree. (b) larger test dataset with $\boldsymbol{N_{spans}} = 20$.

We notice the same descending behavior for both PSD components, with the lowest average L2 obtained for the IP component. In particular, the average distances associated with IP and Q reach a plateau at d =7, respectively, around 3.2 dB and 3 dB. Meanwhile, Fig. 6(b), with the same structure, reports the average L2 distance over the test set with $N_{spans} = 20$. In this case, the two components converge at d =9, around 2.7 dB for IP and 2.67 dB for Q.

We can draw two observations on the average distance with two different datasets from these two plots. First, we obtained a slightly larger average distance for both new datasets concerning the entire one. Second, the degree of the polynomial decreases with the reduction of the dataset, so we have a different multi-polynomial regression for each dataset. With these polynomial degrees, we evaluate the average distance for the PSD (Q) and the estimation error of the a_{NL}.

978-3-903176-54-6 © 2023 IFIP

Fig. 7(a) shows the average L2 distance over different ρ in the three different datasets as a function of the number of spans. The y-axis represents the average L2 distance, while the x axis reports the number of spans in the log scale. The solid line reports the average L2 evaluated over the entire dataset, as before. The novelties are the dotted and the dashed lines, representing the average distance for the dataset with N_{spans} equal to 10 and 20. The first thing that can be noticed is that decreasing the datasets produces a lower average distance for a low number of spans. This reduction is more prominent for the first span, in which we can gain a 1.3 dB between the dataset with $N_{spans} = 10$ and the one with $N_{spans} = 50$. However, this affirmation is no longer valid for the last spans of the smaller datasets, where we observe an average L2 distance comparable to (or larger than) the one for the entire dataset.

Moreover, we report the estimation error variance for the different datasets to have a complete vision of the ML proposal accuracy. Fig. 7(b) depicts the RMSE over all the ρ in estimating a_{NL} for three different datasets as a function of the number of spans. The y-axis describes RMSE over ρ, while the x-axis represents the number of spans in the log scale. This shows a decrease in RMSE with a smaller dataset for the first spans, as observed for the average distance L2 in Fig. 7(a). This illustration helps to connect a large gain obtained for distance to one slight improvement in the RMSE for estimating the a_{NL}. For instance, in the first span, the gain of 1.3 dB in the average distance can be translated into an improvement of 0.1 dB in the estimation of the a_{NL}.

At first glance, it seems rather counterintuitive that we have

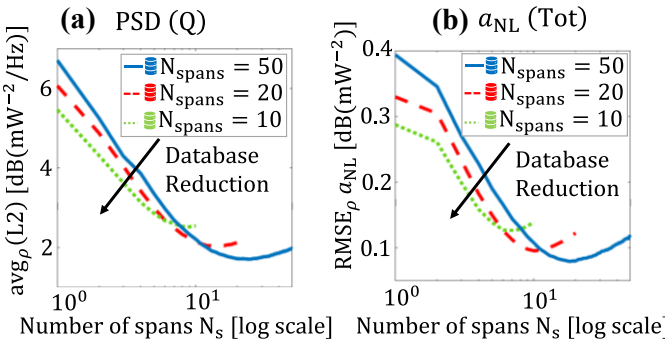

Fig. 7: (a) average distance over the different $\boldsymbol{\rho}$ of the three datasets, respectively $\boldsymbol{N_{spans}}$ equal to 10,20, and 50. (b) RMSE over all the $\boldsymbol{\rho}$ in the different datasets.

obtained an improvement in the error prediction with a decrease in the dataset size. However, we could have expected this behavior since we applied a polynomial regression. The reduction of the database allows a tailored fitting for the first spans. In fact, we have found different convergence polynomial degrees for two smaller datasets. To tackle the disparities in these two working regimes, for future work, we must consider alternative ML approaches and address an ample parameter space for the dataset to approximate a real network scenario.

V. Conclusion

In this paper, we propose and numerically assess an ML-based multi-polynomial regression estimation for Q and IP NLI PSD vectors to have a fast and accurate assessment of the performance that allows decoupling the line from the influence of receiver DSP. We demonstrated PSD estimation capability and assessed by reporting, against SSFM, the average L2 distance and, more conventionally, the NLI variance estimation characterization for sparsely allocated optical links of various lengths. For longer links above the tenth span, we report an average L2 distance over the different channel allocations of less than 2 dB for both the IP and Q PSD components while an RMSE on the NLI variance below 0.12 dB. Instead, the accuracy is worse for shorter links, with an average L2 below 6.7 dB and RMSE less than 0.39 dB. We also discuss the option of improving accuracy by considering a dataset with less spans a suitable option for networks with shorter links.

Acknowledgment

This work was partly funded by EU Horizon 2020 B5G-OPEN Project (grant agreement 101016663).

References

[1] G. Agrawal, *Nonlinear Fiber Optics*, 4th ed. Academic Press, 7 2006.
[2] P. Poggiolini, et al., "The GN-Model of Fiber Non-Linear Propagation and its Applications," *Lightwave Technology, Journal of*, vol.32, no.4, pp. 694–721, Feb 2014.
[3] R. Dar, et al., "Properties of nonlinear noise in long, dispersion-uncompensated fiber links," *Optics Express*, vol. 21, no. 22, pp. 25685–25699, 11 2013.
[4] M. Lonardi et al., "Power Spectral Density Estimation in Dispersion Unmanaged Coherent Metro Networks," in *Optical Fiber Communication Conference (OFC) 2019 (2019), paper M1I.3*. Optical Society of America, Mar. 2019, p. M1I.3.
[5] X. Ye and A. Ghazisaeidi, "KerrNet: Artificial neural networks to speed up perturbation analysis-based models by five orders of magnitude," in *2022 27th OptoElectronics and Communications Conference (OECC) and 2022 International Conference on Photonics in Switching and Computing (PSC)*, pp. 1–4.
[6] Q. Qiu, et al., "Fourier neural operator based fiber channel modeling for optical transmission," in *ECOC 2022*.
[7] N. Rossi, et al., "Nonlinear interference noise statistics in unmanaged coherent networks with channels propagating over different lightpaths," in *Optical Communication (ECOC), 2014 European Conference on*, Sept 2014, pp. 1–3.
[8] O. Rival and K. Mheidly, "Accumulation Rate of Inter and Intra-channel Nonlinear Distortions in Uncompensated 100G PDM-QPSK Systems," in *National Fiber Optic Engineers Conference*. Optical Society of America, 2012, p. JW2A.52.
[9] G. James, et al., *An introduction to statistical learning*. Springer, vol. 112.
[10] A. Bononi, M. Bertolini, P. Serena, and G. Bellotti, "Cross-Phase Modulation Induced by OOK Channels on Higher-Rate DQPSK and Coherent QPSK Channels," in Journal of Lightwave Technology, vol. 27, no. 18, pp. 3974-3983, Sept.15, 2009, doi: 10.1109/JLT.2009.2021537.