JLT-34070-2024

Physics-based Modeling for Hybrid Data-Driven Models to Estimate SNR in WDM Systems

Mariane Mansour, Md Saifuddin Faruk, *Senior Member, IEEE*, Charles Laperle, *Member, IEEE,* Michael Reimer, *Member, IEEE,* Maurice O'Sullivan, and Seb J. Savory, *Fellow, IEEE, Fellow, Optica*

***Abstract*—Recently several machine learning methods have been proposed to estimate the SNR, based on launch data and other system factors. These data-driven methods typically require a large number of datasets for training and generally are not interpretable. In this paper, we propose an alternative approach that requires less data and is interpretable, specifically a hybrid algorithm combining a physical model with Gaussian process regression. We develop a measurement-informed physical model, systematically reducing the number of independent parameters based on the underpinning physics and improve the overall performance of the physical model marginally. The model is validated using measurements performed on a 15-channel wavelength-division multiplexed system propagating over 1,000 km of standard single-mode fiber. The proposed hybrid model is not only interpretable but also obtains better agreement with measurements than a Gaussian process regression model and a simple neural network model for a given number of training datapoints.**

Index Terms—Gaussian noise model, long-haul transmission, wavelength division multiplexing, signal-to-noise ratio estimation, hybrid model, physical model, machine learning.

I. INTRODUCTION

THE available signal-to-noise ratio (SNR) is core to the performance of optical fiber communication systems, be this as a metric for network routing [1] or the digital signal processing (DSP) and forward error correction (FEC) within a modem [2].

In this paper, we tackle the task of accurate measurement-informed SNR estimation at the receiver for a wavelength division multiplexed (WDM) system. The SNR estimation can be done analytically using the Gaussian noise (GN) model, which is a relatively simple and sufficiently reliable tool for performance estimation [3-4]. As with any analytical model, the accuracy is only as good as the input model parameters, so for the GN model to give accurate predictions, the system parameters need to be measured. Alternatively, SNR can be estimated through a data-driven machine learning (ML) approach [5-8]. For example, in [5], the training for SNR estimation was performed on the channel launch powers, the amplifier gains and their noise figures, and the launch power at each fiber span. Hybrid models are also being used in the field of optical communication [9, 10, 11], as they combine the advantages of the physical model in terms of interpretability and that of the data-driven model in terms of adaptability.

Unlike an ML approach, which provides little insight into the system, an analytical approach allows a better understanding of the sources of the noise present, and their impact on the SNR. In turn, this helps localize faulty components, allowing for a fast rerouting of the corresponding signal. Hence, it is important for optical performance monitoring (OPM) which is the continuous monitoring of the signal and the characterization of its parameters [12,13].

The work in this paper is an extension of our previous work [14] which compares the use of ML versus the use of the physical model to estimate the SNR of a long-haul optical communication system. This paper considers the different ways of implementing the physical model and diving into the details of such implementation. We also investigate integrating it in a hybrid model with machine learning techniques.

The paper is organized as follows. In section 2, the most relevant noise sources in long-haul optical communications are identified and discussed. Section 3 covers multiple methods of implementing the physical model. Section 4 focuses on detailing the experimental model that we used. Section 5 is dedicated to comparing the results from using the physics-based methods, data-driven methods which are based on machine learning and the hybrid method.

II. NOISE SOURCES IN WDM SYSTEMS

For the analysis included in this paper, we will partition the noise into three parts, such that the total noise-to-signal ratio (NSR) for the ith channel is given by

$$NSR_i = NSR_{M,i} + NSR_{L,i} + NSR_{NL,i} \tag{1}$$

where $NSR_{M,i}$ is the modem noise, $NSR_{L,i}$ is the linear noise and $NSR_{NL,i}$ is the nonlinear interference noise.

In a WDM system where N_{ch} channels are multiplexed, the modem noise-to-signal ratio term, $NSR_{M,i}$ is independent of the

Manuscript received February 02, 2024.

For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising. Data underlying the results presented here are at https://doi.org/10.17863/CAM.105797

M. Mansour and S.J. Savory are with the Fibre Optic Communication Systems Laboratory, Department of Engineering, Cambridge, CB3 0FA (e-mail: mm2537@cam.ac.uk; sjs1001@cam.ac.uk)

M. S. Faruk is with the School of Computer Science and Engineering, Dean Street, Bangor, LL57 1UT.

C. Laperle, M. Reimer, and M. O'Sullivan are with Ciena Corporation, Ottawa, Ontario K2K 0L1, Canada.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org

signal launch power. In contrast, linear noise related term is given as, $NSR_{L,i} = P_{ASE,i}/P_i$, where $P_{ASE,i}$ is the total noise power from the amplified spontaneous emission (ASE within the channel) of the optical amplifiers and P_i is the launch power of ith channel. The nonlinear noise contribution is given by $NSR_{NL,i} = P_{NLI,i}/P_i$ where the nonlinear interference on the ith channel is given by $P_{NLI,i} = P_i \sum_{j=1}^{N_{ch}} \eta_{ij} P_j^2$ where η_{ii} is the self-phase modulation (SPM) factor, and η_{ij} for $j \neq i$ is the cross-phase modulation (XPM) factor capturing the effect of the jth channel launch power P_j on signal i [15]. The effects of four-wave mixing are ignored given that the conditions for phase-matching are not met. We can then rewrite Eq. (1) in terms of corresponding NSRs as:

$$NSR_i = NSR_{\mathrm{M},i} + \frac{P_{ASE,i}}{P_i} + \sum_{j=1}^{N_{ch}} \eta_{ij} P_j^2 \tag{2}$$

We note that, given this equation, at higher launch powers it is the nonlinear noise that dominates.

We will refer as characterizing parameters the ones that, given the launch powers of the channels, allow us to calculate the corresponding NSRs. In this case, they are $\boldsymbol{NSR_{\mathrm{M}}}$ which is the modem noise to signal vector, $\boldsymbol{P_{ASE}}$ the ASE noise power vector, and the $\boldsymbol{\eta}$ matrix. We define N_p as the number of characterizing parameters involved in an equation.

From the GN model [3], the XPM efficiency parameter is dependent on the walk-off parameter between two channels, which is a function of their frequency separation $\Delta f = |f_i - f_j|$, with closed-form solutions for the XPM and SPM efficiency factors over one span given by equations (3) and (4) respectively; these equations were transcribed to maintain notational consistency:

$$\eta_{ij} = \frac{32}{27} \frac{\gamma^2}{4\pi\alpha\beta_2 B^2} \ln\left[\frac{\Delta f + \frac{B}{2}}{\Delta f - \frac{B}{2}}\right] \quad for\ i \neq j \tag{3}$$

$$\eta_{ij} = \frac{32}{27} \frac{\gamma^2}{4\pi\alpha\beta_2 B^2} \ln\left|\frac{\pi^2 \beta_2}{\alpha} B^2\right| \quad for\ i = j \tag{4}$$

where α is the attenuation coefficient, γ is the nonlinear coefficient, L_s is the span length, β_2 is the chromatic dispersion coefficient which within the GN model is assumed to be constant over all channels, and B is the symbol rate.

It has been shown [3, 16, 17] that this nonlinear efficiency factor η_{ij} decreases as the channel separation increases and is approximately inversely proportional to the channel frequency separation.

The total XPM-induced phase shift on a channel is computed by summing terms from all other channels. Accordingly, and assuming that the XPM-induced phase shift is only dependent on channel separation, the $\boldsymbol{\eta}$ matrix comprising these factors can be a symmetric Toeplitz matrix.

III. Methods for $\boldsymbol{\eta}$ Matrix Characterization

Each of the following methods represents a different implementation of Eq. (2) based on different restrictions enforced on the $\boldsymbol{\eta}$ matrix and the other characterizing parameters. We define a datapoint as a set of N_{ch} values where each value respectively corresponds to a channel, during one instance of the experiment. Each instance of the experiment is characterized by a datapoint of per-channel launch power values, which generates a datapoint of channel NSRs; in our case, one datapoint of 15 per-channel launch powers, generates one datapoint of 15 NSR values.

For a certain number of characterizing parameters, we can compute the theoretical minimum number of datapoints to use to compute them using $N_d = \lceil N_p/N_{ch} \rceil$ where $\lceil . \rceil$ represents the ceiling function. However, we note that in order to maintain the stability of the algorithm and ensure positive NSR values, additional datapoints can be added. The number of variables for each method and the minimum number of datapoints required for each method are shown in Table 1.

A. Method 1

In this method, we assume that all characterizing parameters are unknown and that they are channel-dependent. We do not enforce any specific requirements on $\boldsymbol{\eta}$.

B. Method 2

This method relies on the GN model which assumes that channel separation is the only determining factor for the $\boldsymbol{\eta}$ matrix values. That is, in addition to being symmetric, $\boldsymbol{\eta}$ has the same value along each of its diagonals. In mathematical terms, we are assuming $\boldsymbol{\eta}$ to be a symmetric Toeplitz matrix.

C. Method 3

This method is similar to method 2, however, additionally and to further decrease the number of variables involved, we assume that the modem noise and the linear noise are channel independent.

TABLE 1. Number of Characterizing Parameters and Minimum Number of Datapoints for each Method

m	$N_{p,m}$	$N_{d,m}$
1	$N_{ch}^2 + 2N_{ch}$	$N_{ch} + 2$
2	$3N_{ch}$	3
3	$N_{ch} + 2$	$\lceil 1 + 2/\mathrm{N_{ch}} \rceil$

This table lists the values of $N_{p,m}$ and $N_{d,m}$ representing the number of parameters and the minimum number of datapoints, respectively, for each method m.

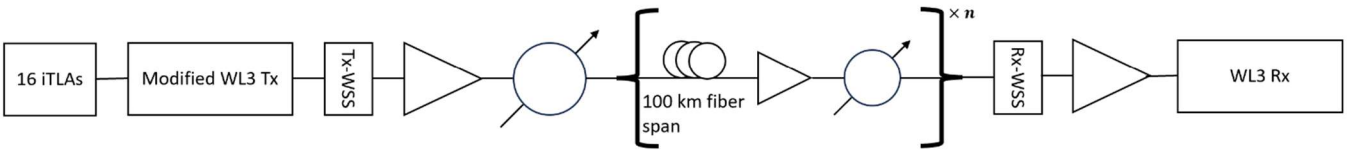

Fig. 1. Experimental setup used to evaluate the theoretical fitting model, where iTLA refers to integrated Tunable Lasers Assemblies, Tx-WSS and Rx-WSS refer to the transmitter's and the receiver's wavelength selective switch respectively, while WL3 refers to the Ciena WaveLogic 3 linecard. The number of spans is given by n.

D) Least squares solution:

To illustrate the technique used to calculate the coefficients, we start with Eq. (2)

$$NSR_i = NSR_{\mathrm{M},i} + \frac{P_{ASE,i}}{P_i} + \sum_{j=1}^{N_{ch}} \eta_{ij} P_j^2$$

Without loss of generality, we begin by expressing a two-channel case in matrix form given by

$$\begin{bmatrix} NSR_1 \\ NSR_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & P_1^{-1} & 0 & P_1^2 & P_2^2 & 0 & 0 \\ 0 & 1 & 0 & P_2^{-1} & 0 & 0 & P_1^2 & P_2^2 \end{bmatrix} \begin{bmatrix} NSR_{M,1} \\ NSR_{M,2} \\ P_{ASE,1} \\ P_{ASE,2} \\ \eta_{11} \\ \eta_{12} \\ \eta_{21} \\ \eta_{22} \end{bmatrix}$$

There are 8 unknowns in the vector hence we need a minimum of 4 separate datapoints that we denote A, B, C, D of measuring the NSR on the two channels and hence

$$\begin{bmatrix} NSR_{1A} \\ NSR_{2A} \\ NSR_{1B} \\ NSR_{2B} \\ NSR_{1C} \\ NSR_{2C} \\ NSR_{1D} \\ NSR_{2D} \end{bmatrix} = \begin{bmatrix} 1 & 0 & P_{1A}^{-1} & 0 & P_{1A}^2 & P_{2A}^2 & 0 & 0 \\ 0 & 1 & 0 & P_{2A}^{-1} & 0 & 0 & P_{1A}^2 & P_{2A}^2 \\ 1 & 0 & P_{1B}^{-1} & 0 & P_{1B}^2 & P_{2B}^2 & 0 & 0 \\ 0 & 1 & 0 & P_{2A}^{-1} & 0 & 0 & P_{1B}^2 & P_{2B}^2 \\ 1 & 0 & P_{1C}^{-1} & 0 & P_{1C}^2 & P_{2C}^2 & 0 & 0 \\ 0 & 1 & 0 & P_{2C}^{-1} & 0 & 0 & P_{1C}^2 & P_{2C}^2 \\ 1 & 0 & P_{1D}^{-1} & 0 & P_{1D}^2 & P_{2D}^2 & 0 & 0 \\ 0 & 1 & 0 & P_{2D}^{-1} & 0 & 0 & P_{1D}^2 & P_{2D}^2 \end{bmatrix} \begin{bmatrix} NSR_{M,1} \\ NSR_{M,2} \\ P_{ASE,1} \\ P_{ASE,2} \\ \eta_{11} \\ \eta_{12} \\ \eta_{21} \\ \eta_{22} \end{bmatrix}$$

By defining suitable vectors and matrices this can be written as

$$\boldsymbol{NSR} = \boldsymbol{Ax}$$

which in general has a solution

$$\boldsymbol{x} = \boldsymbol{A}^{+}\boldsymbol{NSR}$$

where $\boldsymbol{A}^{+} = (\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\boldsymbol{T}}$ denotes the pseudo-inverse of $\boldsymbol{A}$ where $\boldsymbol{A}^{\boldsymbol{T}}$ denotes the transpose of $\boldsymbol{A}$ (which for this square matrix just becomes $\boldsymbol{A}^{-1}$). This method can hence be extended to an arbitrary number of channels with a suitably defined vector of $\boldsymbol{NSR}$ and corresponding matrix $\boldsymbol{A}$ to give the vector of model parameters $\boldsymbol{x}$.

IV. Experimental Setup

The purpose of these experiments is to evaluate the characterizing parameters of Eq. (2) by measuring multiple datapoints of launch powers of each channel and corresponding NSR values. Conversely, once the parameters are found, we can compute the expected NSR at the receiver end given a set of per channel launch powers.

The experimental setup is similar to that in [14] and is shown in Fig. 1. The transmitter comprises 16 integrated tunable laser assemblies (iTLAs) bulk modulated by a modified Ciena WaveLogic 3 line card. The modulated output is then passed through a transmitter wavelength selective switch (Tx-WSS), allowing the selection of certain channels and dropping the rest. This is then followed by a booster amplifier. The multiplexed channels are then sent through n different spans, each span consisting of 100 km standard fiber followed by a fixed gain inline-EDFA. We recall that the gain is not uniform over the channels.

All of the connections are made using a Polatis 32×32 fiber switch, which also allows the control of the equal total launch power at the beginning of each span, that is it permits an exact compensation of the total power loss during propagation with its attenuation control function. At the receiver end, a Rx-WSS allows channel selection for demodulation with a Ciena WaveLogic 3 line card receiver.

Given the measurements are performed after 10 spans and temporal decorrelation of the channels occurs within the first span due to chromatic dispersion no additional decorrelation is included. We measure the BER at the receiver from which we can calculate the SNR (and hence the NSR). For a 16-QAM modulated signal that is Gray coded the relevant equation is given by [17]:

$$BER \approx \frac{3}{8}\mathrm{erfc}\left(\sqrt{\frac{SNR}{10}}\right) \tag{5}$$

where $\mathrm{erfc}(\cdot)$ is the complementary error function. The inverse complementary error function $\mathrm{erfc}^{-1}(\cdot)$ allows the SNR to be calculated from BER using $SNR \cong 10 \times \{\mathrm{erfc}^{-1}(8 \times \mathrm{BER}/3)\}^2$.

We start our experiment by selecting a number of channels to multiplex, and the number of fiber spans. In this paper, we work with 15 WDM channels with central frequency 193.4

THz and 50 GHz channel spacing, and each channel is modulated at 34.5 GBd. We drop one of the 16 C-band channels available to conform with the convention of having a central channel. Next, we regulate the channels' baseline launch powers to ensure similar optical SNR (OSNR) at the transmitter end. The latter is measured using a Finisar WaveAnalyzer (model WA 1500S). We set the received optical power at -4 dBm, as per the equipment's requirements.

We then proceed to select a power range for all the channels. We do so by studying the performance of the central channel over $[-4\text{ dBm}, 4\text{ dBm}]$ Tx total launch power value range as per the equipment limitations. We measure the optimum launch power around 0 dBm. The test power range is then chosen to take into account the rate at which the BER is affected by the launch power. We recall that at lower launch powers linear noise dominates, whereas nonlinear noise dominates in the higher power regime. We denote this test power range as P_{range} which will denote the range of test powers that allows us to study Eq. (2).

Each instance of the experiment is based on a datapoint of launch powers of each channel and results in an NSR datapoint. Each experiment includes a pre-determined number of instances, and each of these instances results in a measured NSR datapoint for a launch power datapoint. These instances are then divided into training datapoints which allow finding the value of the characterizing parameters, and test datapoints allowing for the evaluation of each of the methods described previously.

The most dominant term in the measurement uncertainties is set by the launch power with an uncertainty of ± 0.05 dB.

V. Results and Discussion

To compare the performance of the methods, we first consider a certain number of training datapoints and compute the parameters of Eq. (2) based on each method. We then consider the test datapoints and use these characterizing parameters to calculate $\widehat{SNR}_{dB}$, the model-based SNR value, for each method. These are compared with the experimentally measured SNR value SNR_{dB} using the root mean squared error (RMSE).

We consider overall RMSE which is computed as follows [18]:

$$RMSE_{dB,overall} = \left\{\frac{1}{N_{ch}\times N_{test}}\sum_{c=1}^{N_{ch}}\sum_{t=1}^{N_{test}}\left\|SNR_{dB}(c,t) - \widehat{SNR}_{dB}(c,t)\right\|^2\right\}^{1/2} \quad (6)$$

where N_{test} is the number of test datapoints. On the other hand, the per channel RMSE is computed for a certain channel which is computed as:

$$RMSE_{dB,c} = \left\{\frac{1}{N_{test}}\sum_{t=1}^{N_{test}}\left\|SNR_{dB}(c,t) - \widehat{SNR}_{dB}(c,t)\right\|^2\right\}^{\frac{1}{2}} \quad (7)$$

A minimum of $N_{d,m}$ power and NSR datapoints are enough to compute the characterizing parameters of each method m. However, to prevent negative NSR values in the testing phase, we increased the number of training datapoints. We denote this new number of datapoints as $N'_{d,m}$; in our case, we used a maximum of 6 additional datapoints.

This section is divided as follows: we first consider the overall performance of these methods, then their per-channel performance. We then study the accuracy of fit of the XPM parameters compared to the theoretical model. We also consider the comparison of the physics-based model with the data-driven approach. The last sub-section is dedicated to investigating the hybrid model approach.

A. Physics-Based Method

1) Overall Methods Evaluation: We ran the first experiment with 15 channels over 10 spans, (a total of 1,000 km). We gather 700 datapoints, and we randomly select 500 datapoints to extract the characterizing parameters, and the rest of the datapoints are used for testing. The RMSE is calculated from the test datapoints, based on the number of datapoints used for training.

As shown in Fig. 2, the curve corresponding to method 2 shows the fastest convergence to an RMSE value of around 0.03 dB. Method 3 is the worst performing in terms of root mean square of 0.06 dB, while method 1 shows a slight relative improvement of 0.05 dB.

The lower performance of method 1 highlights the problems of having a model with 255 free parameters, compared to the simpler method 2 with 45. In contrast the poorer performance of method 3 highlights that could be due to the oversimplification of the system by imposing channel independence on the modem and linear noises. We note that the plots in Fig. 2 have different starting points as the minimum number of free parameters differs between methods.

The distribution of the estimation error for each method is shown in Fig.3. Method 3 shows the highest variance, as it fails to capture the channel dependence of the linear and modem terms in Eq. (2). On the other hand Method 1 is more prone to noise integrating the model given the high number of parameters involved. Additionally, Fig. 3 shows that method 2 provides the lowest absolute maximum error at 0.14 dB, while methods 1 and 3 result in a maximum absolute error of 0.20 dB, and 0.21 dB respectively.

2) SPM and XPM nonlinear terms: As expected from the GN model, for all three methods the SPM and XPM factors follow a distribution inversely proportional to the distance between the channels, with

$$y = \frac{a}{x+b} + c \quad (8)$$

where y stands for the XPM values and x stands for the distance between channels.[2]

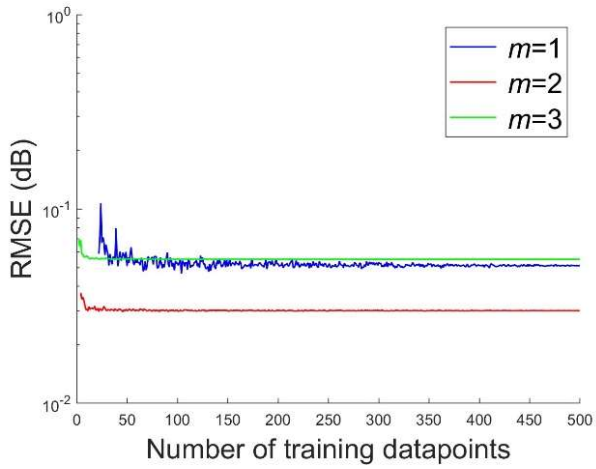

Fig. 2. Overall RMSE of the 200 test datapoints, based on the number of data points used to compute the characterizing parameter values for each method.

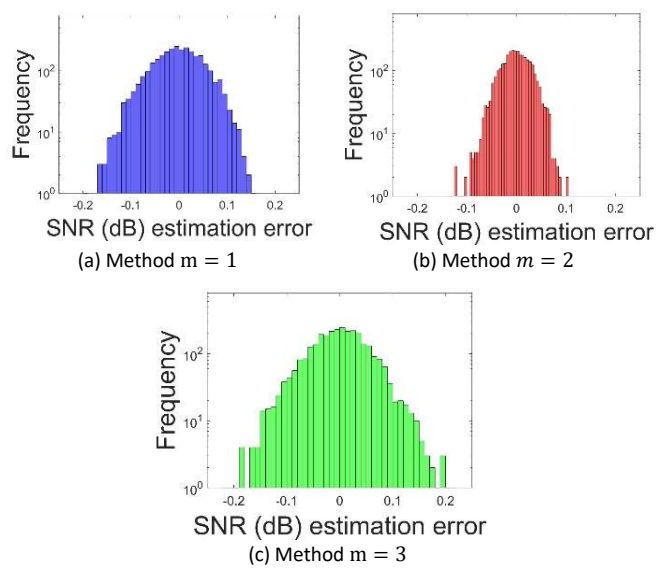

Fig. 3. The distribution of the estimation error over all channels and for each method. The estimation error is defined as the difference between the experimentally measured SNR and the SNR prediction based on each method's model, respectively. The means of the distributions are at -2×10^{-3}, -1×10^{-3}, and -2×10^{-4} dB, respectively. The second method shows the lowest standard deviation at 0.03 dB. In comparison, methods 1 and 3 result in a higher standard deviation of around 0.05 dB and 0.06 dB, respectively.

Fig. 4 shows the distribution of these factors providing confidence in the three methods. We notice close η_{ij} values for the three methods, despite the substantial difference in the reported RMSE values shown in Fig. 2. In choosing the number of parameters, the goal is to find a balance between allowing for enough parameters to accommodate channel dependency and restricting the number of parameters to prevent noise from playing a role in their estimation. Method 3 does not accommodate enough of all the channels, as it assumes the same P_{ASE} and NSR_M for all of them. In contrast, method 1 allows a leniency in the η_{ij} values, and while these when average over the diagonals $+k$ and $-k$ result in values similar to those in Method 2, this does not guarantee similar RMSE performances. Method 2 manages to capture the channel dependence, all while restricting the number of variables just enough to prevent the noise from playing a role in the estimations.

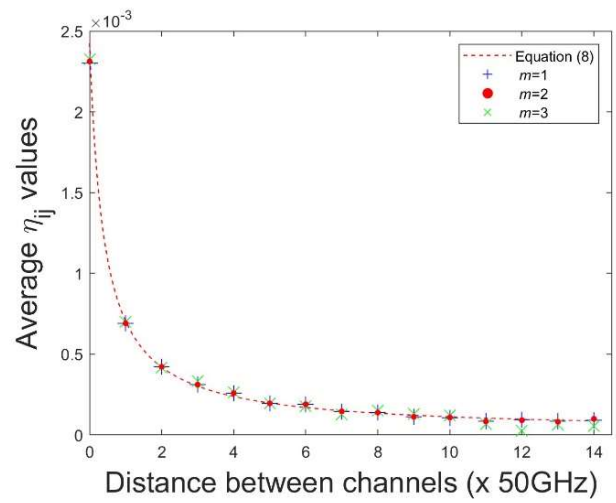

Fig. 4. Distribution of the XPM parameters based on the distance between channels, and the fitting of the XPM parameter values from method 2. The values on the parameters for the fit based on Eq. (8) are: $a = 9.73\times10^{-4}Hz$, $b = 0.40Hz$, $c = 2.00\times10^{-5}$. Their respective 95% confidence intervals are: $[7.94\times10^{-4}, 11.23\times10^{-4}]Hz, [0.33, 0.46]Hz,\ [-1.48\times10^{-5}, 5.47\times10^{-5}]$.

3) Per-channel evaluation: The RMSE performance of all the 15 channels is shown in Fig.5. The best performing channels are the non-central inner channels, especially channels 4, 5, 6, and channels 10, 11, 12 using method 2. The minimum maximum SNR error is for channel 10 at 0.06 dB when using method 2. For all the channels, method 2 is the best-performing one.

4) Modem SNR estimation: We next consider investigating the estimation of the modem SNR. As mentioned previously, the modem SNR is independent of the propagation in the fiber, or of the number of spans this propagation involves. The per-channel distribution of the estimation of the modem SNR for 100 training datapoints is shown in Fig. 6. We notice that this distribution is approximately flat. The maximum difference in estimation between channels is for method 2, with around 0.50 dB difference between channels 7 and 1. Channel 1 is the most sensitive to the method used with around 0.30 dB difference in modem SNR values. Channel 4 is the most resilient for change in methods when estimating this value. Method 3 shows an average of 15.41dB, with a standard deviation of 0.02 dB.

[2] This form comes from noticing that the logarithmic term in Eq. (3) and Eq. (4) can be re-written as $\ln\left(1+\frac{1}{\Delta f/B-1/2}\right)$ and then approximated as a Taylor series.

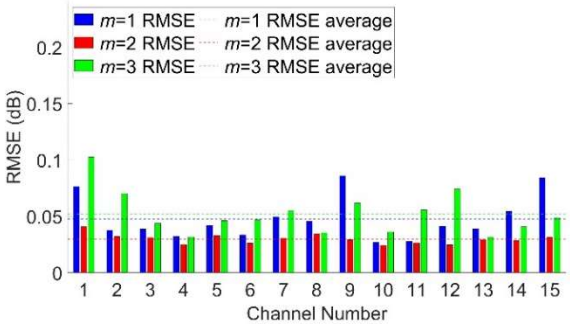

Fig. 5. The performance of each channel based on RMSE error.

B. Data-Driven Approaches

For the data-driven approaches, we use the neural network (NN) and Gaussian process regression (GPR). For NN, we use a feed-forward network with two hidden layers, each including 128 nodes to capture the dimensionality of the system for a fair comparison with the physical model. The input layer consists of 15 nodes, each corresponding to a channel launch power value, and the 15 output nodes represent the resulting channel NSR values. We use the Exponential Linear Unit as our activation function and run the experiment for training datapoints ranging between $N_{p,m}$ and 500 as our input. We use a mini-batch size of 100 and a constant learning rate of 10^{-3}. Final hyperparameter values were chosen based on systematic experimentation and optimization and selected based on the best performance.

Gaussian processes are a probabilistic machine learning approach, which produces an output with a clear probabilistic interpretation [19] and hence confidence margins. In addition, their use of a kernel-based approach allows modeling of the relationship between the data and also permits working in a feature space. We run the GPR for the same range of training datapoints and use the squared exponential as our kernel function.

Fig. 7 shows that the performance of the NN is not as good as that of the physical model (method 2). The average RMSE value for the NN is 0.04 dB compared to method 2 which plateaus at 0.03 dB. The GPR's performance is marginally better with a higher number of training datapoints; it requires at least 250 training datapoints before we start observing this improvement. For 500 training datapoints, the GPR's RMSE reaches an RMSE level lower than that of the other two methods. Whilst this improvement at a relatively large number of datapoints is not very significant, however, it does highlight the fact that the GPR is capturing dependencies that are not accounted for in the physical model which eventually plateaus.

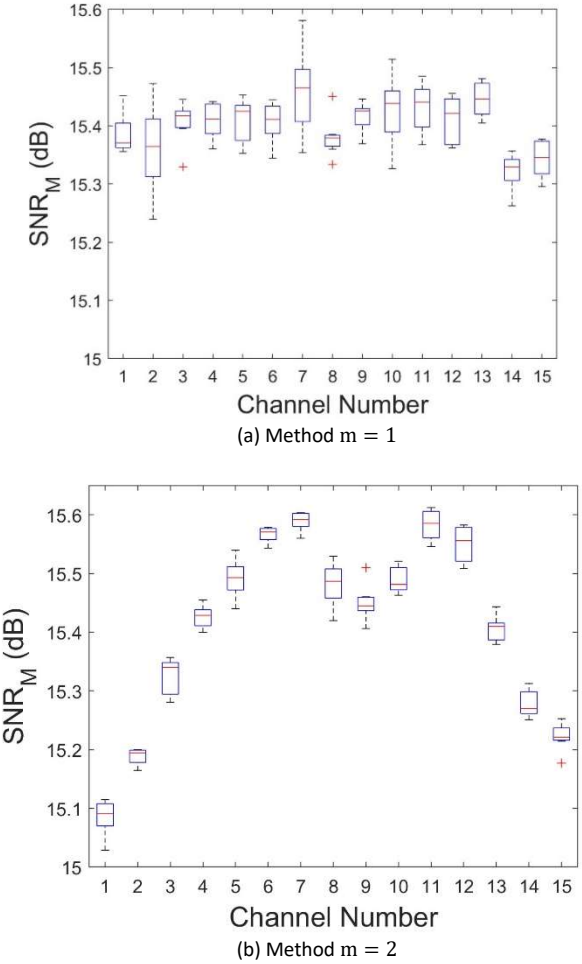

Fig. 6: Per-channel modem noise distribution for (a) method 1 and (b) method 2. These values were estimated by dividing our 700 datapoints into groups of 100, each of which was used to estimate a value for model noise leading to 7 estimations per channel.

We also would like to note that in [14] the NN was trained for 7 channels and that the physical model used was slightly different than method 1 used here with the modem SNR being experimentally measured and channel independent. This could explain why in [14], the NN performs better than the other proposed methods. In the case of the experiments run in this paper, there might be a potential to find a different NN that captures the interactions between the 15 channels and outperforms the one proposed in this paper. We highlight the fact that NN setups are generally found by systematic experimentation.

C. Hybrid Method

In this sub-section, we investigate the use of machine learning to decrease the residual error from the physics-based

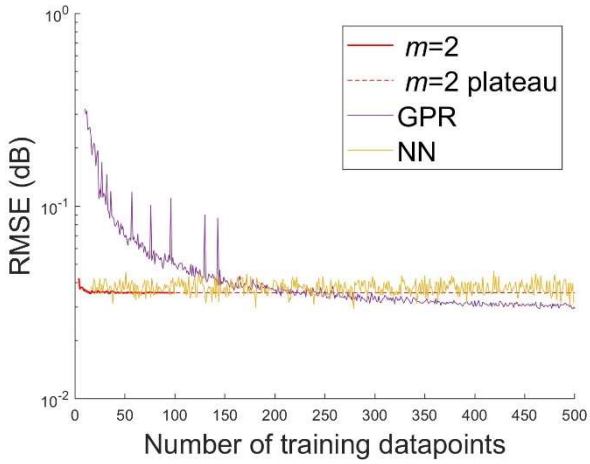

Fig. 7. Comparison between the physical model and machine learning for estimating the SNR.

method. This allows for the advantages of the machine learning approach in terms of accuracy to be used, all while maintaining the physical understanding of the physical model which also provides the advantage of training speed.

In our algorithm, we use GPR rather than NN for two main reasons. According to Fig. 7, GPR performs better than NN. In addition, GPR does not require an estimation of the dimensions of the system which is key in working with NN. For the data-driven approach, the estimation of the dimensionality was straightforward given the knowledge of the number of characterizing parameters. However, this is not simple when targeting residual errors. In addition, in the hybrid approach, we consider method 2 since it shows a better overall performance.

Starting with 700 datapoints, we randomly select 100 datapoints to extract the physical model based on method 2, we then randomly select another 400 datapoints to be used as training for the GPR, and the rest (200 datapoints) are used for testing. We start by training the physical model and then use the extracted characterizing parameters to compute the NSR residual error for the GPR training datapoints. The GPR model, with a squared exponential kernel function, is hence trained to predict residual errors based on launch power. The final predicted NSR from the hybrid model becomes:

$$NSR_{hybrid,i} = NSR_{physical,i} + (residual\ error)_{GPR,i} \quad (9)$$

Therefore, given a datapoint of launch powers, the hybrid NSR prediction is performed as shown in Fig. 8.

Compared to using GPR on its own, the hybrid method provides a better estimation with an overall RMSE error of 0.02 dB with the lowest RMSE error per channel being for channel 11 at 0.02 dB as shown in Fig. 9. In this figure, we also show the results from training the physical model on a maximum of

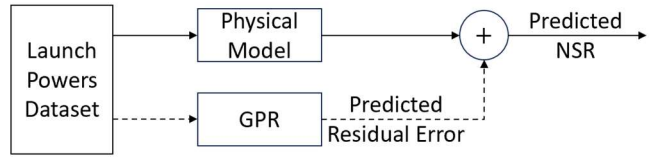

Fig. 8. Diagram representing the hybrid model. The dashed lines represent the optional GPR residual training.

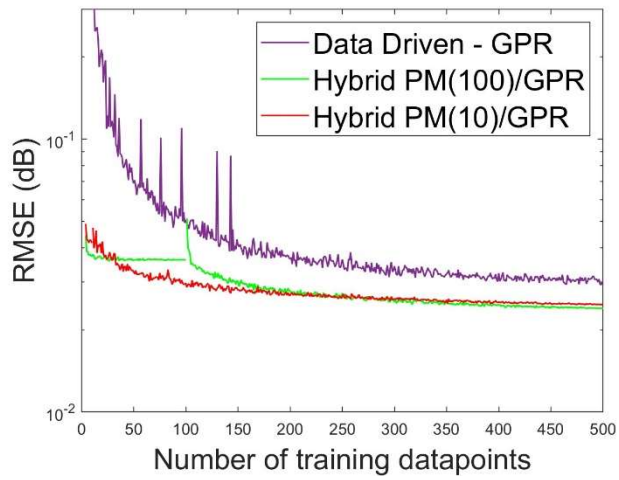

Fig. 9. Comparison of the performance of the hybrid method with the data-driven approach. The hybrid model is based on $m = 2$ and is trained based on a physical model from 10 datapoints (red) and 100 datapoints (green).

10 datapoints while the rest of the 490 datapoints dedicated to training are used in the GPR.

This improvement in performance in the hybrid model could be due to the GPR capturing other impairments that were assumed negligible in the formulation of the physical model. In our case, a simple kernel function is used for both the data-driven GPR and the hybrid model. It could be the case that the kernel function in the latter has the potential to target impairments that are not captured in the physical model. This is unlike using GPR on its own, where this simple kernel function has to capture the main impairments, that are the most dominating trend in the data. We believe that both of the performances can be further improved by constructing a more sophisticated kernel function. We also believe that hybrid models, in general, have the potential to tackle other impairments, especially in the case of tight filtering penalty, and polarization-dependent loss as ML will help to bridge the gap between the values estimates from the physical model and the experimentally measured ones.

VI. Conclusion

We have experimentally demonstrated SNR estimation techniques for WDM systems using physics-based, data-driven and hybrid methods. We conducted a 15-channel WDM transmission experiment over 1,000 km. We have shown that the method in which the physical model is implemented affects

its performance. In particular, we have shown that limiting the XPM efficiency matrix to a symmetric Toeplitz matrix decreases the number of characterizing parameters and hence the minimum number of datapoints needed. In our case, this decrease is from 17 datapoints to 3 datapoints. The RMSE for this implementation is at 0.03 dB compared to an RMSE of 0.05 dB when no restrictions on the XPM efficiency matrix are involved. We note that the number of datapoints can be further decreased by employing the nonlinear parameter fit and estimating its unknowns, though this construction of the equations requires a nonlinear solver. With a GPR-based machine learning method, we can achieve a slightly better performance, however, it requires a larger number of training datapoints (250 datapoints to reach the same performance). Finally, we implemented the hybrid model combining the physical model based on the symmetric Toeplitz matrix and the GPR, which for training over 500 datapoints performs better than the physical model or the GPR used on their own, with an RMSE of 0.02 dB.

ACKNOWLEDGMENT

M. Mansour, Md. Faruk, and Seb J. Savory gratefully acknowledge the donation of equipment and support from Ciena. M. Mansour thanks Ciena for funding her Ph.D. studentship. Seb J. Savory acknowledge support from UK EPSRC TRANSNET project (EP/R035342/1). This research was performed under the auspices of a Ciena University collaborative research grant.

REFERENCES

1. S. Ganesh and R. Amutha, "Efficient and secure routing protocol for wireless sensor networks through SNR based dynamic clustering mechanisms," in *J. Opt. Commun. Netw.,* vol. 15, no. 4, pp. 422-429, Aug. 2013.
2. M. N. Sakib and O. Liboiron-Ladouceur, "A study of error correction codes for PAM signals in data center applications," in *IEEE Photon. Technol. Lett.*, vol. 25, no. 23, pp. 2274-2277, Dec.1, 2013, doi: 10.1109/LPT.2013.2286105.
3. P. Poggiolini, G. Bosco, A. Carena, *et al*, "The GN-model of fiber non-linear propagation and its applications," *J. Lightw. Technol.*, vol. 32, no. 4, pp. 694-721, 2014.
4. P. Poggiolini, "The GN model of non-linear propagation in uncompensated coherent optical systems," *J. Lightw. Technol.*, vol. 30, no. 24, pp. 3857-3879, 2012.
5. J. Lu, G. Zhou, Q. Fan, *et al*, "Performance comparisons between machine learning and analytical models for quality of transmission estimation in wavelength-division-multiplexed systems [Invited],"*J. Opt. Commun. Netw.,* vol. 13, no. 4, pp. B35-B44, 2021.
6. Y. Pointurier, "Machine learning techniques for quality of transmission estimation in optical networks," *J. Opt. Commun. Netw.,* vol. 13, no. 4, pp. B60-B71, 2021.
7. S. Allogba, S. Aladin and C. Tremblay, "Machine-learning-based lightpath QoT estimation and forecasting," *J. Lightw. Technol.*, vol. 40(10), pp. 3115-3127, 2022.
8. J. W. Nevin, S. Nallaperuma, N. A. Shevchenko, X. Li, M. S. Faruk and S. J. Savory, "Machine learning for optical fiber communication systems: An introduction and overview," *APL Photonics,* vol. 6, no. 12, pp. 121101, 1 Dec. 2021.
9. Meng Cai, Xiaomin Liu, Lei Liu, Huazhi Lun, Lilin Yi, Weisheng Hu, and Qunbi Zhuge, "Optical filtering impairment monitoring based on model fusion for optical networks," Opt. Express 30, 24639-24654 (2022)
10. S. Zhu *et al*., "Hybrid machine learning EDFA model," *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, San Diego, CA, USA, 2020, pp. 1-3.
11. J. W. Nevin, F. J. Vaquero-Caballero, D. J. Ives and S. J. Savory, "Physics-informed Gaussian process regression for optical fiber communication systems," in *J. Lightw. Technol.*, vol. 39, no. 21, pp. 6833-6844, 1 Nov.1, 2021
12. D. C. Kilper, R. Bach, D. J. Blumenthal, *et al*, "Optical performance monitoring," *J. Lightw. Technol.*, 22(1), 294-304, 2004.
13. Z. Dong, F. N. Khan, Q. Sui, K. Zhong, C. Lu and A. P. T. Lau, "Optical Performance Monitoring: A Review of Current and Future Technologies," *J. Lightw. Technol.*, vol. 34, no. 2, pp. 525-543, 2016.
14. M. S. Faruk, M. Mansour, C. Laperle, M. O'Sullivan, and Seb J. Savory, "Experimental investigation of physics- and ML-based QoT estimation for WDM systems," in *Proc. Eur. Conf. Opt. Comm.*, Paper Tu.A.2.1, 2023
15. H.-M. Chin, D. Charlton, A. Borowiec, *et al*, "Probabilistic design of optical transmission systems," *J. Lightw. Technol.*, vol. 35, no. 4, pp. 931-940, 2017.
16. S. J. Savory, "Approximations for the nonlinear self-channel interference of channels with rectangular spectra," in *IEEE Photon. Technol. Lett.*, vol. 25, no. 10, pp. 961-964, 2013.
17. A. Carena, V. Curry, G. Bosco, *et al*, " Modeling of the impact of nonlinear propagation effects in uncompensated optical coherent transmission links," *J. Lightw. Technol.*, vol. 30, no. 10, pp. 1524-1539, 2012.
18. Y. Pointurier, "Machine learning techniques for quality of transmission estimation in optical networks," *J. Opt. Commun. Netw.,* vol. 13, no. 4, pp. B60-B71, April 2021
19. C. E. Rasmussen, and C. K. I. Williams, "Relationships between GPs and Other Models" in *Gaussian Processes for Machine Learning*, the MIT press, 2005, ch. 6.