Neural Network based Linear and Nonlinear Noise Estimation

F.J. Vaquero Caballero[(1)], D. J. Ives [(1)], C. Laperle[(2)], D. Charlton[(2)], Q. Zhuge[(2)], M. O'Sullivan[(2)], Seb J. Savory[(1)]

Abstract—**Operators are pressured to maximize the achieved capacity over deployed links. This can be obtained by operating in the weakly nonlinear regime, requiring a precise understanding of the transmission conditions.**

Ideally, optical transponders should be capable of estimating the regime of operation from the received signal and feeding that information to the upper management layers to optimizate the transmission characteristics, however this estimation is challenging.

This paper addresses this problem by estimating the linear and nonlinear SNR from the received signal. This estimation is performed by obtaining features of two distinctive effects: nonlinear phase noise and second-order statistical moments. A small neural network is trained to estimate the SNRs from the extracted features.

Over extensive simulations covering 19,800 sets of realistic fibre transmissions, we verified the accuracy of the proposed techniques. Employing both approaches simultaneously gave a measured performances of 0.04 and 0.20 dB of std error for the linear and nonlinear SNR, respectively.

Index Terms—**Coherent communications, Metrology, Optical performance monitoring, Machine Learning**

I. INTRODUCTION

Growing traffic demands increase the pressure on operators to maximise the capacity over their deployed networks. Flexible Optical Networking and Impairment Aware Networking will heavily rely on signal quality information for routing decisions, requiring information extracted from the received signal.

Optical Performance Monitoring (OPM) is the field of optical communications that aims to characterise the impairments suffered through transmission from features extracted from the received optical signal [1]. The extracted information is especially relevant for the upper management layers since it provides

F.J. Vaquero Caballero, D. J. Ives, and Seb J. Savory, are with the Electrical Engineering Division, Department of Engineering, University of Cambridge, Cambridge CB3 0FA, U.K. (e-mail: fjv24@cam.ac.uk; e-mail: di231@cam.ac.uk; e-mail: sjs1001@cam.ac.uk)

C. Laperle, D. Charlton, Q. Zhuge, and M. O'Sullivan, are with Ciena Corporation, Ottawa, ON K2K 0L1, Canada. (e-mail: claperle@ciena.com; e-mail: dcharlto@ciena.com; e-mail: qzhuge@ciena.com; e-mail: mosulliv@ciena.com)

accurate information of the current state of transmission, enabling the identification of sources of under-performance and possible countermeasures to improve performance.

Most of the linear impairments suffered during transmission, such as inter-symbol interference (ISI) from chromatic dispersion (CD), and state-of-polarization rotation (SOPR), can be compensated by DSP with insignificant penalties [2].

Therefore, the source of impairments that penalize signal performance in an optical link is noise. Two sources of transmission noise can be considered: amplified spontaneous emission (ASE) and nonlinear interference noise (NLI). ASE noise is a result of the amplification from the Erbium doped fibre amplifiers (EDFA), which compensates for the signal attenuation. Nonlinearities are caused by the power dependence of the fibre refractive index, where the resultant nonlinear interference is proportional to the cube of the signal power, p.

The transmitter and receiver of the optical signal are also subjected to penalties due to non-ideal components, internal amplifications, shot noise, and quantisation, undermining the maximum achievable performance in the optical link [3]. These effects are commonly modelled as transceiver noise. Although significant, it is known as its characterisation is usually provided by the equipment manufacturer.

The total signal-to-noise ratio (SNR) in an optical link can be defined as:

$$SNR^{-1} = \underbrace{\frac{\sigma_{ASE}^2}{p} + \sigma_{TRx}^2}_{SNR_{LIN}^{-1}} + \underbrace{\eta_{NL}p^2}_{SNR_{NLI}^{-1}}, \qquad (1)$$

where σ_{ASE}^2, $\sigma_{TRx}^2 p$, and $\eta_{NL}p^3$, model the ASE, transceiver, and nonlinear noise, respectively. Instead of defining 3 $SNRs$ accounting for the individual contribution, we define two SNRs accounting for the linear (ASE and transceiver), and nonlinear contributions: SNR_{LIN} and SNR_{NLI}.

Analytical models such as the GN [4] and the EGN [5] models provide acurate predictions of nonlinear noise, they require an accurate characterisation of the fibre light-path, such as its fibre parameters, and information of the neighbouring channels such as their modulation format, symbol-rate, and transmitted

power. The preceding requirements are impractical as they are not readily available in heterogeneous dynamic networks.

Already several approaches have been considered for the estimation of nonlinearities, both [6], [7] based their study on the correlations in the received signal due to nonlinearities. Characterization of a auto-covariance function over nonlinear noise by principal component analysis and its prediction through neural networks was proposed in [8]. But very little progress has been made on the estimation of the linear noise from the received signal. It is possible to perform a 3-parameter fit of Equation 1, [9], but it requires to modify the transmitted power which would compromise performance and be impossible on real networks. We recently proposed a new method capable of jointly estimating the linear and nonlinear SNR, based on the measurement of the nonlinear phase noise [10].

This paper expands our last contribution over an extensive set of realistic simulation data, discussing the limitations of the proposed technique in terms of modulation format and reach. Additionally, we propose a new approach based on the estimation of the temporal properties of nonlinearities through time-varying ISI matrices, whose estimation is not limited to those scenarios where nonlinear phase noise is a major nonlinear contribution.

II. THEORETICAL BACKGROUND

In this section we introduce the time-domain first order regular perturbation solution of the nonlinear Schroedinger equation, suitable for modelling self-phase modulation (SPM) and cross-phase modulation (XPM) under the framework of pulse collisions [11]. The additional assumptions are that four-wave mixing (FWM) contributions are insignificant and there is polarisation aligment between channels. More complex models considering additional nonlinearities and different polarisation states can be found in [12] and [13]. This model is not only capable of accurately modelling the pulse propagation in the fibre and the characterisation of nonlinearities, but also provides an intuitive approach to understand the nature of the nonlinearities.

During this paper we employ the Bra–ket notation [14]: $|b\rangle$ denotes 2-D complex Jones column vectors $[b_x, b_y]^T$, $\langle b|$ denotes its Hermitian transpose $[b_x^*, b_y^*]$, $\langle b|c\rangle$ denotes the scalar product $b_x^* c_x + b_y^* c_y$, and $|b\rangle\langle c|$ is their dyadic operator:

$$\begin{bmatrix} b_x c_x^* & b_x c_y^* \\ b_y c_x^* & b_y c_y^* \end{bmatrix}, \tag{2}$$

We also denote: $\langle b \rangle$ as the average of $b^{(k)}$. In a WDM set-up, the nonlinear interference noise caused by an interfering channel B into channel A in the symbol

TABLE I: Characteristics of the different pulse collisions as noted by [13].

	2PC	3PC(I)	3PC(II)	4PC
Nature:	PN & PS	PN & PS	CN	CN
Mod. dependence.:	Yes	No	Yes	No
l condition:	$l = 0$	$l = 0$	$l \neq 0$	$l \neq 0$
k and m condition:	$k = m$	$k \neq m$	$k = m$	$k \neq m$

$n = 0$, commonly known as XPM, is given by:

$$|\Delta a^{(0)}\rangle = j\gamma \sum_{l,k,m} X_{l,k,m} \Big(\langle b^{(k)}|b^{(m)}\rangle \mathbf{I} + |b^{(m)}\rangle\langle b^{(k)}| \Big) |a^{(l)}\rangle, \tag{3}$$

where $\mathbf{I}$ is the 2x2 identity matrix. The transmitted symbols of the channel of interest and interfering channel are $|a_n\rangle$ and $|b_n\rangle$, respectively. The coefficients $X_{l,k,m}$ define the nature and efficiency of the NLI, and are dependent on the normalized waveform $g(z,t)$, symbol duration T, chromatic dispersion β_2, and channel spacing Ω [rad/s]. The analytical expression of $X_{l,k,m}$ is given by:

$$X_{l,k,m} = \int_0^L \int_{-\infty}^{-\infty} f(z)g^*(z,t)g(z,t-lT)$$
$$g^*(z,t-kT-\beta_2\Omega z)g(z,t-mT-\beta_2\Omega z)\,dt\,dz, \tag{4}$$

Equations 3 and 4 are directly obtained from the first-order regular perturbation approximation of the nonlinear Schroedinger equation [15], which is a widely used approximation of the Kerr effect. The previous equations are implicitly defined for inter-channel nonlinearities where the channel of interest is different from the interfering channels, although this model is also suitable for SPM.

$X_{l,k,m}$ analytical expression is very illustrative of the nature of the nonlinearities: nonlinearities are created by four pulse collisions, consisting of the match filter at the receiver, $g^*(z,t)$, including chromatic dispersion compensation (CDC) and three waveforms corresponding to the symbols: $|a_l\rangle$, $|b_k\rangle$, and $|b_m\rangle$.

Depending on the indexes, the contributions can be classified into two-pulse collisions (2PC, $l = 0, k = m$), three-pulse collisions (3PC, type I: $l = 0, k \neq m$, or type II: $l \neq 0, k = m$), and four-pulse collisions (4PC, $l \neq 0, k \neq m$). The different types of pulse collisions result in different noise contributions in the form of phase noise (PN), polarization-scattering (PS), and circular noise (CN). Table I summarizes the nature of the different pulse collisions.

An alternative form of Equation 3, consist of merging the contributions: $X_{l,k,m} \Big(\langle b^{(k)}|b^{(m)}\rangle \mathbf{I} + |b^{(m)}\rangle\langle b^{(k)}| \Big)$, into a set of 2x2 matrices $\mathbf{H}_l^{(n)}$. Since $\mathbf{H}_l^{(n)}$ includes the transmitted symbols, the matrices are time-varying. Due to its similarity to a linear filtering process where the filtering effect results in ISI, the set of $\mathbf{H}_l^{(n)}$ are commonly denoted as the 2x2 time-varying ISI matri-
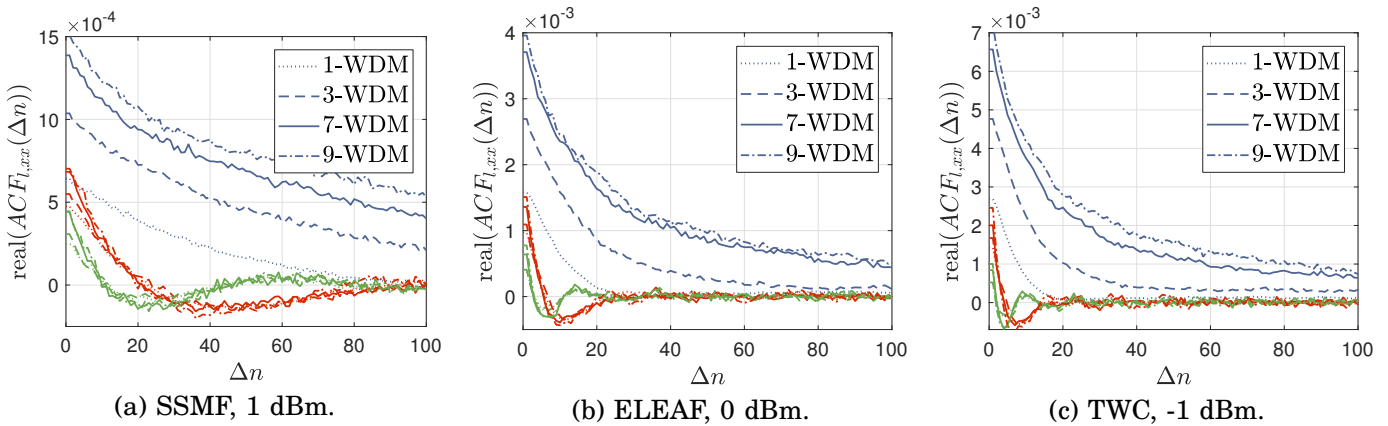
Fig. 1: ACF($H_l^{(n)}$) for 1,3,7, and 9 WDM channels. $l = 0$ (blue), $l = 1$ (red) and $l = 2$ (green). For different types of fibres simulated noise-less over 10 spans as defined in Table II.

ces, given by:

$$|\Delta a^{(n)}\rangle = j\gamma \sum_l \mathbf{H}_l^{(n)}|a^{(n-l)}\rangle, \quad \mathbf{H}_l^{(n)} = \begin{bmatrix} H_{l,xx}^{(n)} & H_{l,xy}^{(n)} \\ H_{l,yx}^{(n)} & H_{l,yy}^{(n)} \end{bmatrix} \tag{5}$$

where l indexes the different orders of the ISI matrices. The zero order of the ISI matrices, $\mathbf{H}_0^{(n)}$, gathers all the 2PC and 3PC(I) contributions, grouping all the elements that induce PS and nonlinear PN together. Although the ISI matrices are time-varying, their statistical properties can be studied through the auto-covariance function (ACF). In [16], a metric capable of extracting their temporal properties from the post-DSP received symbols, $\hat{a}_r^{(n)}$, was derived:

$$\hat{H}_{l,rs}^{(n)} = (\hat{a}_r^{(n)} - a_r^{(n)})/a_s^{(n-l)}, \quad r, s \in \{x, y\},$$

$$ACF_{l,rs}^{(\Delta n)} = \lim_{K \to \infty} \frac{1}{K} \sum_{n=1}^{K} \hat{H}_{l,rs}^{(n)} \hat{H}_{l,rs}^{*(n+\Delta n)} \tag{6}$$

Figure 1 shows the evolution of the real part of the most relevant orders of ACF for SSMF, ELEAF and TWC fibres over 10 spans, accounting for a total transmission of 1000km. The real part is the most significant contribution for the considered cases. Moreover, it is observed that the ACF of $l = 0$ is the strongest contribution for the three cases: $l = 0, 1, 2$.

III. METRICS AND TOOLS

In this section we introduce different strategies and metrics to estimate linear and nonlinear noise from the received signal. We review the most relevant metrics from the literature and propose a novel metric based on the ACF, covered in the aforementioned section. We also briefly introduce the foundations of dimensional reduction and neural networks relevant for our specific applications.

A. Amplitude Noise Covariance

Equation 3 illustrates the nature of nonlinear noise, consisting of the mixing of triplets of symbols from the channel of interest and the interfering channel, weighted by $X_{l,k,m}$. Consequently, the resultant non-linear noise exhibit correlation over the transmitted symbols.

In [6] by the definition of the amplitude noise covariance (ANC), the correlation characteristics of the non-linear noise were exploited to estimate the strength of the nonlinearities. For received symbols post-DSP, the noise is calculated in the normal direction of the constellation, as illustrated in Figure 2:

$$\Delta a_r^{(n)} = |\hat{a}_r^{(n)}| - |a_r^{(n)}|, \quad r \in \{x, y\} \tag{7}$$

Finally, ANC can be defined as:

$$ANC_{rs}^{(m)} = cov(\Delta a_r^{(k)}, \Delta a_s^{(k+m)}), \quad r, s \in \{x, y\}, \tag{8}$$

resulting in 3 ANCs: $\{xx, xy, yy\}$. The existence of 3 ANC components comes from the nature of nonlin-earities and the polarization effects: due the nature of nonlinearities, birefringence, and the random state-of polarization of the fibre, the nonlinear correlation is not limited to either polarization and is therefore present in both. The limitation of the study of the covariance to the normal direction is motivated by the effect of phase noise from the transmitted and received laser over the tangential components, which can alter the measured covariance.

A transformation to the ANC was proposed by [17]:

$$ALANC_{rs} = 10 \log_{10}(1/ \sum_{k=1}^{6} |ANC_{rs}^{(k)}|), \tag{9}$$

this metric adds the most significant elements of the covariance metric into a single term through a set of nonlinear transformations. In the following sections, we will refer to $ALANC_{xx}$ as simply ALANC, since very little improvement is obtained by considering the remaining $ALANC_{yy}$ and $ALANC_{xy}$ metrics in our estimations. The resultant metric has an almost linear mapping with SNR_{NLI} for high values of SNR_{LIN}.

Figure 3 shows the evolution of ALANC for noise-less SSMF transmission (red) and noise-loaded case (blue). We observe that the metric has very similar
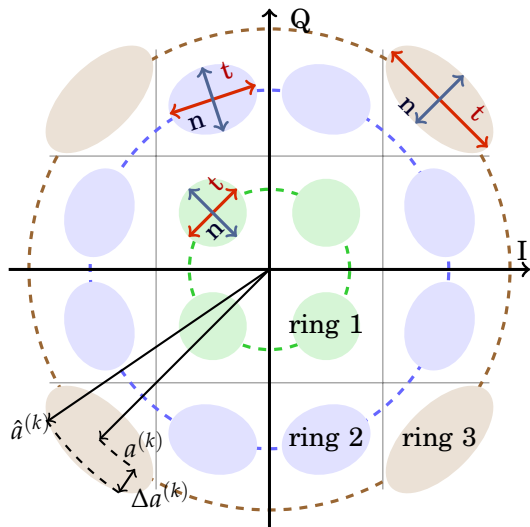
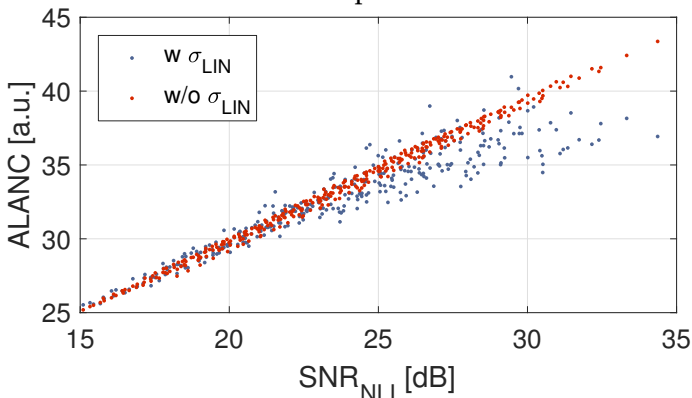Fig. 2: 16-QAM signal including Δa_k noise, and $\mathbf{t}$ and $\mathbf{n}$ components.

Fig. 3: ALANC for 1,3,5,7,9 WDM channels of SSMF, for 5 to 15 spans. NF=4.5dB and $SNR_{TRx} =$16dB.

behaviour for both noise-loaded and noise-less cases with $SNR_{NLI} < 25dB$.

B. Normal and Tangential Components

An alternative set of metrics can be obtained by decomposing the constellation into its normal, $\mathbf{n}$, and tangential, $\mathbf{t}$, components of the noise variance for each individual symbol. This approach enables us to quantify the different contributions of nonlinearities: the tangential components are affected by nonlinear PN, PS, and CN, whilst the normal components are only affected by nonlinear PS and CN. Alternatively, linear noise contributions from amplification and transceiver noise affects equally both $\mathbf{n}$ and $\mathbf{t}$ components. Figure 2 illustrates this decomposition for 3 constellation symbols in each of the power-constant rings of a 16-QAM signal.

Since phase noise induces broadening as a function of the magnitude of each ring, it is possible to average the $\mathbf{n}$ and $\mathbf{t}$ components for each constellation ring

resulting in N_v and T_v, where v indexes the constellation ring ($v = \{1, 2, 3\}$). These metrics correspond to a classification of the noise components of the error vector magnitude (EVM) metric, which can be directly related to SNR as:

$$SNR = \frac{1}{EVM^2} = \frac{4}{N_1 + 2N_2 + N_3 + T_1 + 2T_2 + T_3} \quad (10)$$

The most significant contributions of the phase noise were studied for a single polarization in [11], where the phase noise is proportional to the variance of the symbols' power. For the two polarizations case, the induced nonlinear phase noise increases monotonically with the fourth order modulation factor (FOMF) [13]:

$$FOMF = \frac{\langle |b|^4 \rangle}{\langle |b|^2 \rangle^2} \quad (11)$$

Therefore, this method may not be suitable for modulation formats such as QPSK due to its constant power and consequently low nonlinear phase noise. This metric is still relevant due to the growing interests in higher-order modulation formats such as 16-QAM, 32-QAM, and probabilistic shaping aiming for Gaussian-like power distributions. Those advanced modulation formats present a higher FOMF, producing more nonlinear phase noise.

Although advanced modulation formats have limited reach, depending on the transmission distance different types of pulse collisions may dominate the nonlinear contributions. For long links, where 3PC and 4PC dominate, the amount of nonlinear phase noise relative to the total nonlinearities will decrease [13], [18], which can penalize the estimation of linear and nonlinear noise based on N_v and T_v metrics. These conditions of non-linear phase noise were not met for the simulation data considered in this paper.

C. PCA of $ACF_{l,rs}(\Delta n)$, and second-order statistics

We have shown in the previous subsections that by studying the covariance of the received signals it is possible to obtain estimates of the SNR_{NLI}, and by measuring the nonlinear phase noise, we can also obtain insightful metrics about the relationship between linear and nonlinear noise. In this subsection, we introduce a novel technique for the separation of SNR_{LIN} and SNR_{NLI}, based on second-order statistics of the auto-covariance function and a dimensional reduction technique. The proposed technique does not rely on the estimation of nonlinear phase noise.

Figure 1 shows the evolution of the $ACF_{l,rs}^{(\Delta n)}$ which provides a description of the time correlation properties of the nonlinearities in the fibre. We cannot directly rely on the ACF due to its high dimensionality, since $ACF_{l,rs}^{(\Delta n)}$ range of interest spread over a wide range of values for Δn and l. But several dimensionality reduction techniques can be applied to obtain a simpler representation of the ACF. Principal component analysis (PCA) presents a simple and intuitive
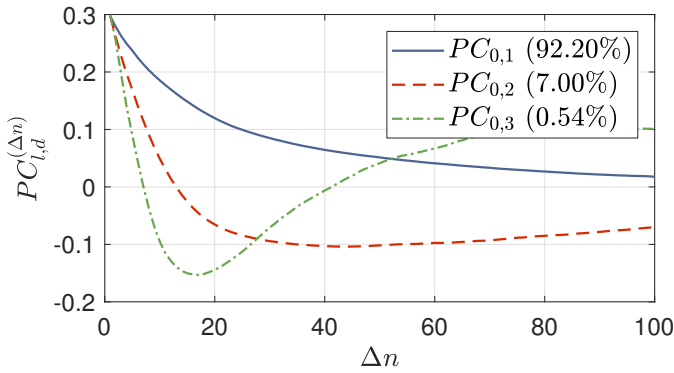
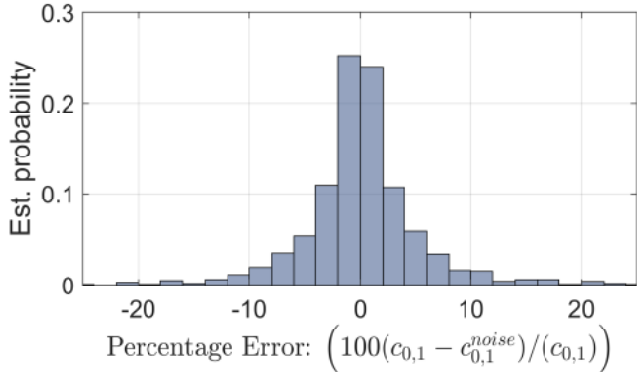Fig. 4: The three more relevant $PC_{0,d}$ coefficients for all noise-free scenarios from Table II.

Fig. 5: Percentage error of $c_{0,1}$ coefficients for data from Table II, NF=4.5dB and SNR_{TRx}=16 dB.

approach [19]. It was already used for nonlinearities identification in [8] and will be used in this paper. The process of the low dimensional reduction of $ACF_{l,xx}^{(\Delta n)}$ can be expressed as:

$$A\hat{C}F_{l,xx}^{(\Delta n)} = \sum_{d=1}^{D} c_{l,d} PC_{l,d}^{(\Delta n)}, \qquad (12)$$

where D is the number of dimensions considered. For simplicity, same number of dimensions are considered in all l, and the study is focused on $ACF_{l,xx}$ only. $PC_{l,d}^{(\Delta n)}$ are the principal components of the element d-dimension, $c_{l,d}$ are the coefficients of the PCA, and $A\hat{C}F_{l,rs}^{(\Delta n)}$ is the resultant approximation of the auto-covariance function.

For simplicity, we will focus on ACF for $l = 0$ since it is the major contribution. Figure 4 illustrates the three most significant $PC_{0,d}$ of $ACF_{0,xx}^{(\Delta n)}$, $\Delta n \neq 0$, with their percentage of explained variance. By making use the first two $c_{0,d}$, $d = \{1,2\}$, it is possible to explain 99.2% of the variance. Consequently, we can obtain a low-dimensional description of the ACF by just a few $c_{0,d}$ coefficients.

It is worth noting that the calculation of the ACF is based on second-order moment estimations where the linear noise also contributes. The linear components are mainly from ASE noise and quantization at the transceiver [3], both contributions can be modelled as additive white Gaussian noise (AWGN). The lin-

ear noise can be assumed to be uncorrelated. Consequently, in a noisy scenario the measured auto-covariance function, $ACF_{l,rs}^{noise,(\Delta n)}$, can be expressed as:

$$\begin{aligned} ACF_{l,rs}^{noise,(\Delta n)} &= ACF_{l,rs}^{(\Delta n)}, \quad \Delta n \neq 0, \\ ACF_{l,rs}^{noise,(0)} &= ACF_{l,rs}^{(0)} + K \cdot \sigma_{LIN}, \end{aligned} \qquad (13)$$

where K is a constant accounting for the signal normalization of the DSP and the effect of the weighted average by the transmitted symbols involved in the ACF calculation. Equation 13 illustrates the delta-behaviour of the ACF for AWGN, implying that all the difference between noise-less and noise-loaded scenarios is captured in $\Delta n = 0$.

A simple approach to exploit that effect can be to apply PCA analysis to $ACF_{0,rs}^{noise,(\Delta n)}$, $\forall \Delta n \neq 0$. By doing so, we obtain a description of the nonlinearities based on the tail of the ACF, where linear noise does not have an effect in ACF. From the estimate of $c_{0,d}$ the nonlinear contribution to $ACF(0)$ can be found; this allows the estimation of the linear noise contribution $K\sigma_{LIN}$. Algorithm 1 provides an step-by-step explanation of this process.

The $PC_{0,d}^{(\Delta n)}$ basis should be calculated under noise-less scenarios, to avoid PCA fitting the noise. Figure 5 shows the percentage error of the component $c_{0,1}$ obtained from a noise-less transmission compared to its value in a noisy transmission, $c_{0,1}^{noise}$. 75% of the components have an absolute percentage error smaller than 5%, and 92% smaller than 10%.

The approach described in this section may be suitable for different covariance metrics, and is not limited to the ACF described in this paper. It could be also applied to ANC, due to its similarity to ACF. We focus on ACF since its spread in Δn for nonlinearities is considerably longer than in the case of ANC. Since for $ANC_{rs}^{(m)}$ the elements of interest over m is limited to a few symbols. Other metrics for noise, with the previously mentioned properties, may be also possible.

Although it was not observed in our simulations, in some systems the linear noise component maybe correlated over a few symbols over the ACF metric due to suboptimal equalization, match-filtering or big roll-off factors. Rather than discarding only $\Delta n = 0$ for the calculation of $c_{0,d}$, the set of indexes over which the correlation takes place must be discarded. By doing

---

Algorithm 1 separation SNR_{NLI} and SNR_{LIN}

1: pre-compute the PCA basis, $PC_{0,d}^{(\Delta n)}$, $\forall \Delta n \neq 0$, based on simulation noise-less data,
2: process $ACF_{0,rs}^{noise,(\Delta n)}$ of the received sequence
3: calculate $c_{0,d}$ of ACF, $\forall \Delta n \neq 0$
4: Use $c_{0,d}$ and $ACF_{0,rs}^{noise,(0)}$ as features to estimate both SNR_{NLI} and SNR_{LIN}
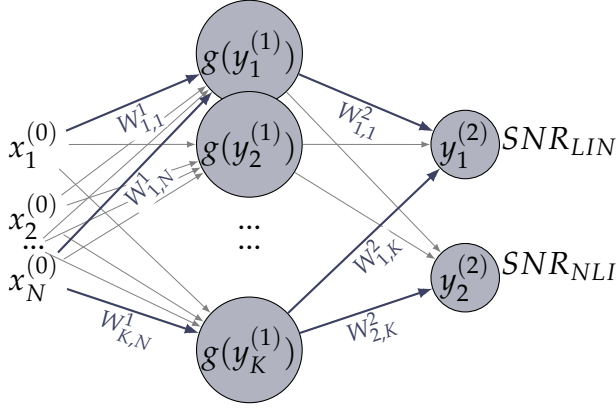
Fig. 6: Illustration of the structure of the single hidden layer neural network used in this paper.

so, we do not foresee significant penalties due to the length and smooth behaviour of ACF.

D. Neural networks based decomposition

In the previous subsections we have described several features that can be extracted from the transmitted data to estimate linear and nonlinear noise. A part from ALANC, which has an almost linear evolution with SNR_{NLI}, the rest of the introduced metrics present a nonlinear relationship between their value and the parameters to estimate. Neural networks are a powerful tool for learning those relationships and accurately provide estimates based on the input features.

Figure 6 illustrates a simple neural network consisting of N inputs $[x_1^0,..., x_N^0]$, 1-hidden layer with K nodes, and two outputs $[SNR_{LIN}, SNR_{NLI}]$. The equations that relates the inputs to the p-layer: $x_n^{(p)}$, to the outputs to the next layer: $x_n^{(p+1)}$ are given by [19]:

$$y_k^{(p)} = bias_k^{(p)} + \sum_{n=1}^{N} W_{k,n}^{(p)} x_n^{(p-1)}, \qquad (14)$$
$$x_n^{(p)} = g(y_n^{(p)}),$$

where $g(\cdot)$ is a nonlinear activation function such as the hyperbolic tangent function (tanh), sigmoid, or rectified linear unit (ReLU). In our case, we make use of tanh function.

The aim of the described neural network is to find the nonlinear mapping between the input features and the expected outputs, in our case SNR_{NLI} and SNR_{LIN}. We employ a 1-hidden layer neural network, since it proved to be sufficient for the transduction of the input features into accurate predictions for our application of interest. The network is trained by the standard 70/15/15 rule for the 3 sets: train/dev/test, with early stopping over the dev, converged by back-propagation. Intense monitoring was performed over the errors of the train/dev/test sets to avoid over-fitting.

The evolution of the convergence was studied over different scenarios to evaluate the likelihood of reaching a local minimum. The neural network weights
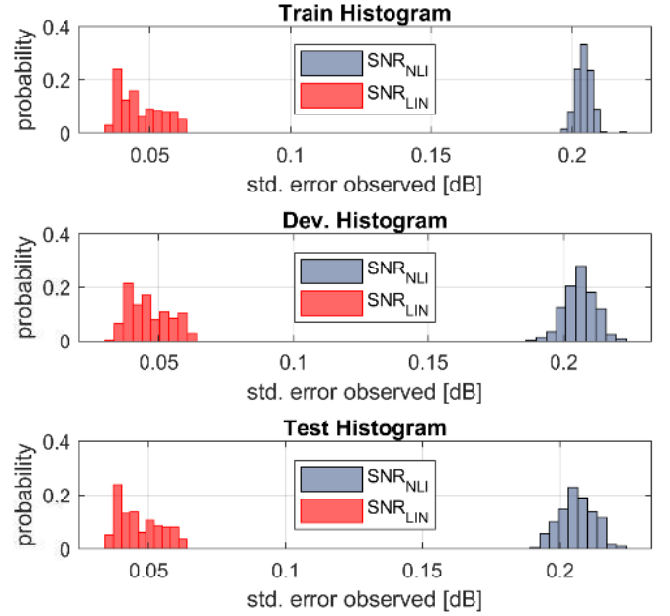

Fig. 7: Performance of 200 train runs of the neural network for the train/dev/test.
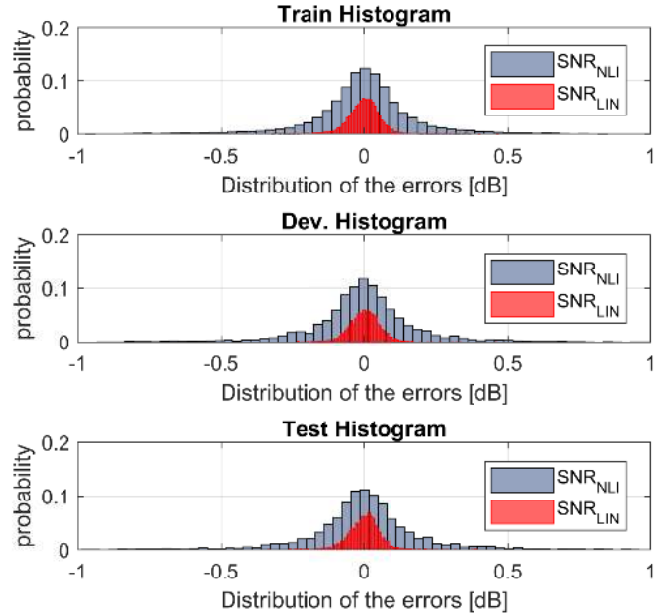

Fig. 8: Distribution of the errors for the case of 0.04 and 0.2 dB of std. of std. error for SNR_{LIN} and SNR_{NLI}

were randomly initialized and several trainings were performed to verify the achieved performance is consistent in all the runs. Figure 7 illustrates the spread of the performance of the neural network for the train/dev/test over 200 runs. It corresponds to the best estimation scenario considered in the results section of this paper, where 0.04 and 0.2 dB of std. error was reported for the SNR_{LIN} and SNR_{NLI}, respectively. Figure 8 shows the histogram of the performance for the same case, 0.04 and 0.2 dB of std. for the SNR_{LIN} and SNR_{NLI}. As expected, the error has a Gaussian shape.

IV. RESULTS

In this section, we evaluate the different methodologies introduced in this paper by computer simulations. For the validation of the algorithm, we consider 3 types of fibres: SSMF, TWC, and ELEAF. The span count is varied between 5 and 15 spans in steps of 1, the length of each span is set to 100 km for all the cases. The channel launch power is varied across 6 dB centred in the approximate optimum launched power (Figure 10). The gain of the amplifiers is set to compensate the loss of the span and their noise figure (NF) is varied uniformly between 4 and 6 dB, in steps of 0.5 dB. The SNR_{TRx} is similarly varied between 15 and 18 dB in steps of 1 dB. The number of WDM channels, C_W, is varied between 1, 3, 5, 7, and 9. All the split-step Fourier simulations are performed linear noise-free, where noise loading is performed at the receiver. By loading noise directly in the receiver, we are neglecting the nonlinear interactions of the noise and between the noise and the signal. This approach is commonly done is simulations [17] and should provide very similar results [20].

The channel spacing is set to 50 GHz, and the modulation format of choice is 16-QAM match-filtered with a 0.14 root-raised cosine (RRC) filter. A summary of the simulations is given in Table II. The total amount of transmission realisations simulated are 19800.

For all the results, the neural network responsible for the mapping between the input features, and the estimation is a 1-hidden-layer network with 7 nodes ($K = 7$). We chose a relatively small neural network due to the simplicity of the required nonlinear regression. By keeping the number of nodes small, we minimize the possibility of over-fitting. The performance was worse for a smaller number of nodes and it stabilized around 7 nodes where additional nodes resulted in similar performance.

Figure 9 shows the performance evolution for all the simulation data, showing the maximum spread of the error of the estimations.

TABLE II: Simulation parameters used.

Fibre types	SSMF	TWC	ELEAF
D [ps/nm/km]	16.7	2.8	4.3
α [dB]	0.2	0.21	0.21
γ [1/W/km]	1.3	2	1.47
Channel launch power [dBm]	[-2,3]	[-3,2]	[-4,1]
# spans	5:1:15		
WDM channels (C_W)	1, 3, 5, 7, 9		
NF [dB]	4,4.5,5,5.5,6		
SNR_{TRx} [dB]	15, 16, 17, 18		
Channel Spacing [GHz]	50		
Modulation format	16-QAM (0.14 RRC)		
Number of symbols	2^{18}		

A. N_v, T_v components, and ALANC

We first consider a neural network with input features: N_v and T_v, $j = \{1,2,3\}$, C_W, and the accumulated chromatic dispersion (ACD).

Figure 9a, and 9b illustrate the performance for SNR_{LIN}, and SNR_{NLI}. The top plot shows the evolution of the estimate as a function of the true SNR, while the bottom plot shows the evolution of the error. It is noticeable that the error for the SNR_{NLI} presents heteroscedastic behaviour: for high SNR_{NLI} the variance of the estimation is higher than for lower SNR_{NLI}. The std. error is 0.27, and 0.08 dB of for SNR_{NLI} and SNR_{LIN}, respectively.

By including ALANC of the x-polarization as an additional input, we observed a very similar performance.

B. PCA over ACF

Secondly, we consider the case of the PCA components where the inputs to the neural network are: $ACF_{0,xx}^{(0)}$, $c_{0,1}$, $c_{0,2}$, C_W, and ACD.

For simplicity, we focus on the xx component of the ACF for the case $l = 0$, which is the strongest contribution, and we only consider the first two components: $c_{0,1}$, and $c_{0,2}$, since they already account for 99.2% of the variance.

For SNR_{LIN}, the performance was very similar to the previous case: 0.08dB of std error: Figure 9c. For the case of SNR_{NLI}, the performance was slightly worse than in the case of N_v, and T_v components, resulting in 0.3dB of std error, Figure 9d.

C. Estimation based on joint approaches

Finally, we include all the explained metrics as input to the neural network to obtain a more accurate estimate. The inputs to the neural network are N_v, T_v, $ACF_{0,xx}^{(0)}$, $c_{0,1}$, $c_{0,2}$, C_W, and ACD.

The improvement is significant compared to previous cases: the SNR_{LIN} estimation std error is 0.04 dB, Figure 9e; while, the SNR_{NLI} is improved to 0.20 dB, Figure 9f.

D. Comparison of results

Table III summarizes the results of this paper. Both approaches based on the estimation of phase noise (N_v and T_v), and the second moment statistics (ACF, $c_{0,1}$, and $c_{0,2}$) perform similarly. Both have a std error of 0.08 dB for the linear SNR. The performance of nonlinear SNR was 0.27 dB for the case of the method based on the estimation of the phase noise, while is was approximately 0.03 dB worse for the case of second moment statistics.
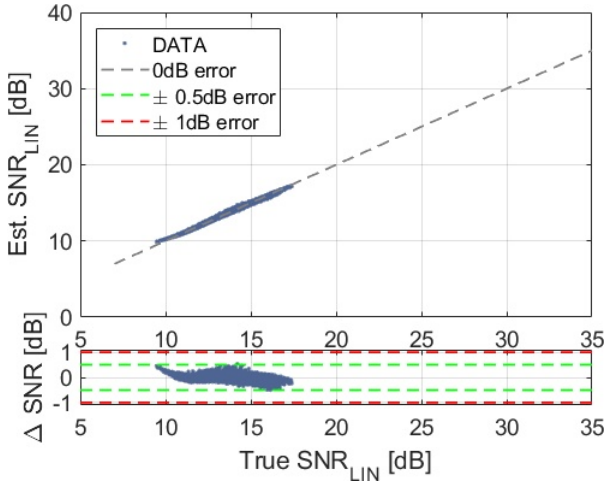
In the case where both approaches are jointly considered to estimate the linear and nonlinear SNR, the performance improves by 0.2 and 0.04 dB for the nonlinear and the linear SNR, respectively.
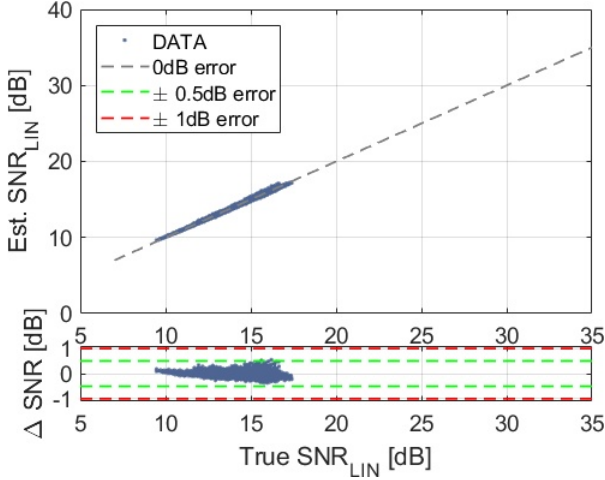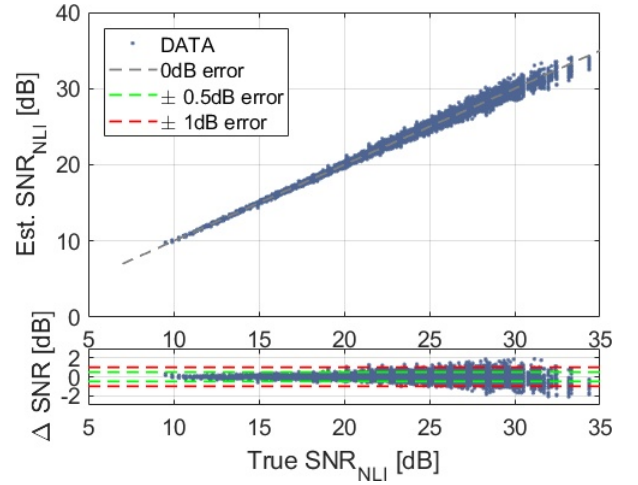
(a) True vs Estimated SNR_{LIN}, NN inputs: $[N_v, T_v, C_W, \text{ACD}]$. Std error: 0.08 dB.
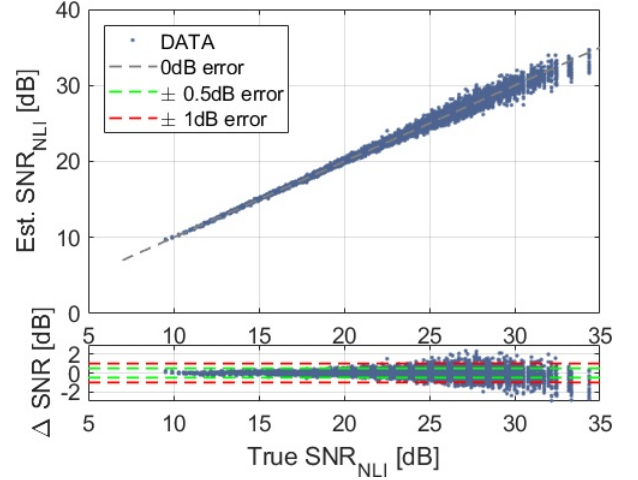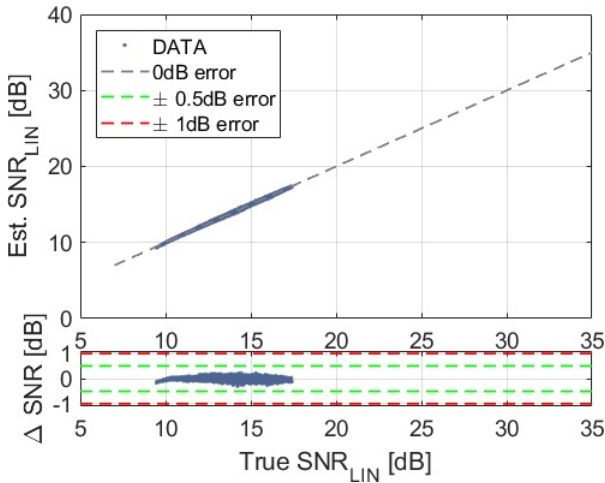
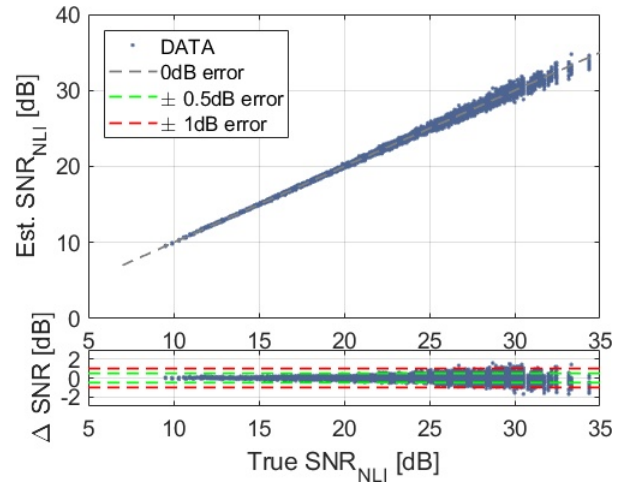(b) True vs Estimated SNR_{NLI}, NN inputs: $[N_v, T_v, C_W, \text{ACD}]$. Std error: 0.27 dB.

(c) True vs Estimated SNR_{LIN}, NN inputs: $[ACF_{0,xx}^{(0)}, c_{0,1}, c_{0,2}, C_W, \text{ACD}]$. Std error: 0.08 dB.

(d) True vs Estimated SNR_{NLI}, NN inputs: $[ACF_{0,xx}^{(0)}, c_{0,1}, c_{0,2}, C_W, \text{ACD}]$. Std error: 0.30 dB.

(e) True vs Estimated SNR_{LIN}, NN inputs: $[N_v, T_v, ACF_{0,xx}^{(0)}, c_{0,1}, c_{0,2}, C_W, \text{ACD}]$. Std error: 0.04 dB.

(f) True vs Estimated SNR_{NLI}, NN inputs: $[N_v, T_v, ACF_{0,xx}^{(0)}, c_{0,1}, c_{0,2}, C_W, \text{ACD}]$. Std error: 0.20 dB.
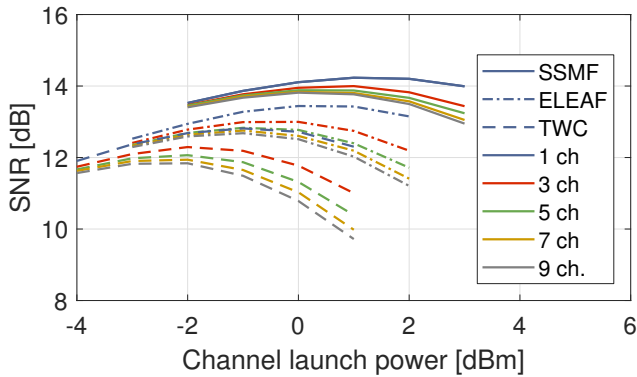
Fig. 9: Evolution of performance for the different cases

Fig. 10: Evolution of SNR for 10 spans of the considered set-ups, where $NF = 4.5$dB and $SNR_{TRx} = 16$dB.

TABLE III: Sumary of the results of this paper

Input Parameters	SNR_{NLI} std,%	SNR_{LIN} std,%
N_v, T_v, C_W, ACD	0.27 (6.4%)	0.08 (1.8%)
$c_{0,(1,2)}$, $ACF_{0,xx}^{(0)}$, C_W, ACD	0.30 (7.1%)	0.08 (1.8%)
All metrics	0.20 (4.7%)	0.04 (0.9%)

V. CONCLUSION

In this paper, we tackle the problem of estimating the linear and nonlinear SNR based on extracted features of the received signal. We explained a theoretical framework suitable for understanding nonlinearities and especially nonlinear phase noise. We also covered the extraction of the temporal description of the time-varying ISI matrices. Using this framework, we discuss the limitations of our previous metric introduced in [10], we note that the proposed technique is suitable for high order modulation formats where the induced nonlinear phase noise is significant. But it may not be suitable for dealing with power-constant modulation formats such as QPSK, or very long transmission distances because of the relative nonlinear contributions of the different types of pulse collisions, leading to more equal **n** and **t** components.

We proposed a novel nonlinear estimation technique which does not rely on nonlinear phase noise. The novel technique exploit the temporal properties of the time-varying ISI matrices and the lack of correlation of the Gaussian noise. The studied temporal properties are inherent to the nonlinearities of any modulation format and transmission distance, although the derivation of ACF comes from a XPM study, we also validate its applicability for single channel transmissions. By applying principal component analysis, we extracted 2 features capable of explaining over 99% of the variance.

The performance of the evaluated metric was investigated over 19800 realizations of optical fibre transmission, covering different fibre types, number of spans, numbers of WDM channels, launch powers, amplifier noise figures, and transceiver SNR.

The considered features have a nonlinear relationship with the target estimations, a small neural network of 1-hidden layer and 7 nodes was trained. We fed different subsets of the presented metrics into the neural network to evaluate its performance.

The performance of the metric based on nonlinear phase noise was 0.27 and 0.08 dB of std for the nonlinear and linear SNR, respectively. For the case of the second-order moment statistics, the performance was very similar: 0.30 and 0.08 dB, respectively. Finally, when applying both techniques jointly, std error of 0.20 and 0.04 dB was obtained for the nonlinear and the linear SNR, respectively. The performance and the input features to the neural networks are summarized in Table III.

We believe that the performance improvement observed compared to our previous results [10], specially for the SNR_{LIN}, is a result of the inclusion of SNR_{TRx} which reduces the range of SNR_{LIN} observed to approximately 9 dB to 17dB. Consequently, the linear noise is a mayor contribution of noise in all the simulated cases and its range of variation is smaller than SNR_{NLI}, which results in a smaller error on its estimation.

We would like to emphasize that in this paper we are measuring the inherent error of the estimator, for the study of the accuracy of different techniques to separate linear and nonlinear noise. We expect that the experimental verification of the proposed algorithms will have a higher uncertainty, mainly dominated for the uncertainty and noise from the experimental set-up, resulting in higher estimation errors for both SNR_{LIN} and SNR_{NLI}.

The considered future work includes the experimental verification of the proposed techniques, the exploration of alternative machine learning techniques, the study of the limitations of the proposed techniques, and inclusion of different auto-covariance components for the estimation of the linear and nonlinear SNR.

REFERENCES

[1] X. Zhou and C. Xie, *Enabling Technologies for High Spectral-efficiency Coherent Optical Communication Networks*. Wiley, 2016.
[2] S. J. Savory, "Digital filters for coherent optical receivers," *Optics Express*, vol. 16, no. 2, p. 804, 2008.

[3] K. Roberts, Q. Zhuge, I. Monga, S. Gareau, and C. Laperle, "Beyond 100 Gb / s : Capacity , Flexibility , and Network Optimization," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 9, no. 4, pp. 12–24, 2017.

[4] P. Poggiolini, "The GN model of non-linear propagation in uncompensated coherent optical systems," *Journal of Lightwave Technology*, vol. 30, no. 24, pp. 3857–3879, 2012.

[5] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri, "EGN model of non-linear fiber propagation," *Optics Express*, vol. 22, no. 13, pp. 16335–16362, 2014.

[6] Z. Dong, A. P. T. Lau, and C. Lu, "OSNR monitoring for QPSK and 16-QAM systems in presence of fiber nonlinearities for digital coherent receivers," *Optics Express*, vol. 20, no. 17, p. 19520, 2012.

[7] H. Choi, J. Chang, H. Kim, and Y. Chung, "Nonlinearity-tolerant OSNR estimation technique for coherent optical systems," in *Optical Fiber Communication Conference*, 2015.

[8] R. T. Jones, J. C. M. Diniz, M. Yankov, M. Piels, A. Doberstein, and D. Zibar, "Prediction of Second-Order Moments of Inter-Channel Interference with Principal Component Analysis and Neural Networks," in *European Conference on Optical Communication, ECOC*, no. 1, pp. 1–3, 2017.

[9] D. J. Ives, F. J. Vaquero Caballero, and S. J. Savory, "Remote Abstraction of an Installed Dark Fiber Network using Noise to Signal Ratio," in *Optical Fiber Communication Conference*, 2018.

[10] F. J. Vaquero Caballero, D. J. Ives, Q. Zhuge, M. O. Sullivan, and S. J. Savory, "Joint Estimation of Linear and Non-linear Signal-to-Noise Ratio based on Neural Networks," in *Optical Fiber Communication Conference*, 2018.

[11] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Properties of nonlinear noise in long, dispersion-uncompensated fiber links," *Optics Express*, vol. 21, no. 22, pp. 25685–25699, 2013.

[12] A. Ghazisaeidi, "A Theory of Nonlinear Interactions between Signal and Amplified Spontaneous Emission Noise in Coherent Wavelength Division Multiplexed Systems," *Journal of Lightwave Technology*, vol. 35, no. 23, 2017.

[13] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Pulse Collision Picture of Inter-Channel Nonlinear Interference in Fiber-Optic Communications," *Journal of Lightwave Technology*, vol. 34, no. 2, pp. 593–607, 2016.

[14] J. P. Gordon and H. Kogelnik, "PMD fundamentals: Polarization mode dispersion in optical fibers," *Proceedings of the National Academy of Sciences*, vol. 97, no. 9, pp. 4541–4550, 2000.

[15] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Properties of nonlinear noise in long, dispersion-uncompensated fiber links," *Optics Express*, vol. 21, no. 22, pp. 25685–25699, 2013.

[16] O. Golani, D. Elson, L. Domanc, L. Galdino, R. Killey, P. Bayvel, and M. Shtaif, "Experimental characterization of the time correlation properties of nonlinear interference noise," in *European Conference on Optical Communication*, 2017.

[17] A. S. Kashi, Q. Zhuge, J. C. Cartledge, A. Borowiec, D. Charlton, C. Laperle, and M. O'Sullivan, "Fiber Nonlinear Noise-to-Signal Ratio Monitoring Using Artificial Neural Networks," *European Conference on Optical Communication*, 2017.

[18] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Inter-Channel Nonlinear Interference Noise in WDM Systems : Modeling and Mitigation," vol. 33, no. 5, pp. 1044–1053, 2015.

[19] C. M. Bishop, *Pattern Recognition and Machine Learning*. 2013.

[20] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri, "On the Accuracy of the GN-Model and on Analytical Correction Terms to Improve It," *arXiv*, no. 1401.6946, 2014.