Accurate Closed-Form GN/EGN-Model Formula Leveraging a Large QAM-System Test-Set

Accurate Closed-Form GN/EGN-model Formula Leveraging a Large QAM-System Test-Set

M. Ranjbar Zefreh, A. Carena, F. Forghieri, S. Piciaccia, P. Poggiolini, *Fellow,* OSA

***Abstract*—We tested the accuracy of a fully-closed-form approximate GN-model formula over 3,000 different C-band fully-loaded WDM-QAM system scenarios and 1,200 partially-loaded ones. By leveraging the large system test-set, we modified the formula to obtain a closed-form formula approximating the Enhanced GN-model (or EGN-model). The combined high accuracy and very fast computation time (ms) of such formula potentially make it an effective tool for real-time physical-layer-aware optical network management and control.**

***Index Terms*—NLI, GN-model, EGN-model, WDM networks, coherent transmission, physical layer aware, control plane, big-data**

I. Introduction

PHYSICAL-layer-aware control and optimization is of primary importance for the efficient operation of flexible ultra-high-capacity optical networks. One of the key enablers is the availability of an accurate analytical model of the Non-Linear-Interference (NLI) noise generated by signal propagation in the fiber. At the same time, the computational effort of the model must be light enough to allow real-time use. One possible way to comply with these requirements is by resorting to *approximate closed-form formulas* derived from existing NLI models.

Various NLI models have been proposed over the years, such as 'time-domain' [1], [2], GN [3], EGN [2], [4], [5], as well as others such as [6], [7] and several more (see refs. in [8]). In this paper, we concentrate on the GN/EGN model class. Many approximate closed-form (ACF) versions of the GN-model have been derived in the past, especially in its *incoherent* version (see iGN-model [3]). Most of them addressed idealized systems consisting of WDM combs with equally spaced identical channels and identical spans, such as [9]. A few tackled more general scenarios. In particular, Eqs. (41)-(43) from [3] could handle arbitrarily different channels in the WDM comb and arbitrarily different spans in the link. They were upgraded in [11] to make them capable of dealing with fully-loaded C-band and (C+L)-band systems, by including frequency-dependent loss, gain and dispersion, as well as ISRS (Interchannel Stimulated Raman Scattering). A similar approach towards upgrading [3] was recently undertaken in [10], as well. The two efforts were performed independently, with [10] appearing in Aug. 2018 and [11] in Sept. 2018. In this paper, we focus on the ACF GN-model formula [11].

M. Ranjbar Zefreh, A. Carena and P. Poggiolini are with Politecnico di Torino, Torino, Italy, e-mail: see www.optcom.polito.it . F. Forghieri and S. Piciaccia are with CISCO Photonics, Vimercate (MB), Italy.

This study was supported by the CISCO Sponsored-Research Agreement (SRA) SMART-LINKS, and by the PhotoNext Center of Politecnico di Torino.

Manuscript received March 15th, 2019.

First, an extensive test of the accuracy of [11] was performed. A large number (4,200) of highly-randomized C-band WDM systems was considered, of which 3,000 were fully-loaded and 1,200 partially (50%) loaded. We used as *benchmark* the system Optical Signal-to-Noise Ratio (OSNR), *inclusive of NLI*, estimated using the full-fledged numerically-integrated EGN-model. The latter has been shown to be very accurate in a wide variety of system scenarios [5], [8].

The results of the comparison, despite the several approximations involved in the derivation of [11] and the challenging features of the system test-set, showed a rather good match overall. However, [11] showed an average tendency to *underestimate* the OSNR. This could be expected since [11] was derived from the GN-model, whose known behavior is to somewhat overestimate NLI [3]. In addition, we also observed a non-negligible *variance* of the error vs. the benchmark.

To improve the convergence of [11] vs. the EGN benchmark, we leveraged the large available 4,200 system test-set to find a simple correction law which contains both physical system parameters and best-fitted coefficients, with the goal of turning [11] into an accurate ACF *EGN*-model formula. This approach proved effective: the OSNR estimation error average and standard deviation vs. the EGN benchmark dropped to almost negligible, with the only exception of very-low dispersion set-ups ($D < 1$ ps/(nm km)). In the following, we first introduce the ACF GN-model formula [11]. Then, the features of the randomized system large-test-set are described. Next, we show the accuracy results of the ACF GN-model formula. Following, we introduce the best-fit correction aimed at achieving an ACF *EGN*-model formula, and discuss its accuracy and computational effort. Conclusion follow.

In this paper, ISRS is not considered. The test and possible best-fit of the version of [11] which supports ISRS is left for future investigation. A preliminary version of this paper was reported on in [12]. Here we use a much larger and more diversified test-set, as well as new best-fit corrections formulas.

II. The ACF GN/EGN formula

The ACF GN/EGN overall formula is shown as Eqs. (1)-(5). A coherent set of units is provided for the readers' convenience. Also, notice that *all quantities* bearing a superscript '(n)' or '(k)' are referred to the n-th or k-th span in the link.

$G_{\mathrm{NLI}}^{\mathrm{Rx}}\left(f_{\mathrm{CUT}}\right)$ in Eq. (1) is the total power-spectral-density (PSD) of NLI at the receiver (Rx) and at the frequency f_{CUT} of the Channel-Under-Test (CUT). Frequencies and bandwidths are in THz and PSDs are in W/THz. Eq. (1) shows $G_{\mathrm{NLI}}^{\mathrm{Rx}}\left(f_{\mathrm{CUT}}\right)$ to be the sum of $G_{\mathrm{NLI}}^{(n)}\left(f_{\mathrm{CUT}}\right)$, i.e., of the PSD

$$G_{\mathrm{NLI}}^{\mathrm{Rx}}\left(f_{\mathrm{CUT}}\right)=\sum_{n=1}^{N_{\mathrm{span}}}\left(G_{\mathrm{NLI}}^{(n)}\left(f_{\mathrm{CUT}}\right)\prod_{k=n+1}^{N_{\mathrm{span}}}\Gamma^{(k)}(f_{\mathrm{CUT}})\cdot e^{-2\cdot\alpha^{(k)}(f_{\mathrm{CUT}})\cdot L_{\mathrm{span}}^{(k)}}\right) \tag{1}$$

$$G_{\mathrm{NLI}}^{(n)}\left(f_{\mathrm{CUT}}\right)=\frac{16}{27}(\gamma^{(n)})^{2}\,\Gamma^{(n)}(f_{\mathrm{CUT}})\cdot e^{-2\alpha^{(n)}(f_{\mathrm{CUT}})\cdot L_{\mathrm{span}}^{(n)}}\cdot\bar{G}_{\mathrm{CUT}}^{(n)}\cdot\left(\rho_{\mathrm{CUT}}^{(n)}\cdot\left[\bar{G}_{\mathrm{CUT}}^{(n)}\right]^{2}I_{\mathrm{CUT}}^{(n)}+\sum_{n_{\mathrm{ch}}=1\ ,\ n_{\mathrm{ch}}\neq n_{\mathrm{CUT}}^{(n)}}^{N_{\mathrm{ch}}^{(n)}}2\rho_{n_{\mathrm{ch}}}^{(n)}\cdot\left[\bar{G}_{n_{\mathrm{ch}}}^{(n)}\right]^{2}I_{n_{\mathrm{ch}}}^{(n)}\right) \tag{2}$$

$$I_{\mathrm{CUT}}^{(n)}=\frac{1}{2\pi\mid\bar{\beta}_{2,\mathrm{CUT}}^{(n)}\mid\cdot 2\alpha^{(n)}(f_{\mathrm{CUT}})}\cdot\operatorname{asinh}\left(\frac{\pi^{2}}{2}\left|\frac{\bar{\beta}_{2,\mathrm{CUT}}^{(n)}}{2\alpha^{(n)}(f_{\mathrm{CUT}})}\right|R_{\mathrm{CUT}}^{2}\right) \tag{3}$$

$$I_{n_{\mathrm{ch}}}^{(n)}=\frac{\operatorname{asinh}\left(\pi^{2}\left|\frac{\bar{\beta}_{2,n_{\mathrm{ch}}}^{(n)}}{2\alpha^{(n)}\left(f_{n_{\mathrm{ch}}}^{(n)}\right)}\right|\left[f_{n_{\mathrm{ch}}}^{(n)}-f_{\mathrm{CUT}}+\frac{R_{n_{\mathrm{ch}}}^{(n)}}{2}\right]R_{\mathrm{CUT}}\right)-\operatorname{asinh}\left(\pi^{2}\left|\frac{\bar{\beta}_{2,n_{\mathrm{ch}}}^{(n)}}{2\alpha^{(n)}\left(f_{n_{\mathrm{ch}}}^{(n)}\right)}\right|\left[f_{n_{\mathrm{ch}}}^{(n)}-f_{\mathrm{CUT}}-\frac{R_{n_{\mathrm{ch}}}^{(n)}}{2}\right]R_{\mathrm{CUT}}\right)}{4\pi\left|\bar{\beta}_{2,n_{\mathrm{ch}}}^{(n)}\right|\cdot 2\alpha^{(n)}\left(f_{n_{\mathrm{ch}}}^{(n)}\right)} \tag{4}$$

$$\bar{\beta}_{2,\mathrm{CUT}}^{(n)}=\beta_{2}^{(n)}+\pi\beta_{3}^{(n)}\left[2f_{\mathrm{CUT}}-2f_{c}^{(n)}\right]\qquad,\qquad\bar{\beta}_{2,n_{\mathrm{ch}}}^{(n)}=\beta_{2}^{(n)}+\pi\beta_{3}^{(n)}\left[f_{n_{\mathrm{ch}}}^{(n)}+f_{\mathrm{CUT}}-2f_{c}^{(n)}\right] \tag{5}$$

of NLI produced in the n-th span at f_{CUT}, over all spans. The product operator 'Π' in Eq. (1) accounts for the linear propagation of each $G_{\mathrm{NLI}}^{(n)}\left(f_{\mathrm{CUT}}\right)$ from the n-th span to the Rx. The other symbols in Eq. (1), related to the n-th span, are (see also Fig.1): $\Gamma_n(f)$, the power-gain/loss at frequency f due to lumped elements, such as amplifiers and gain-flattening filters (GFFs), placed at the end of the span fiber; $2\alpha_n(f)$, the fiber power-loss coefficient (1/km) at frequency f; $L_{\mathrm{span}}^{(n)}$, the span length (km).

The $G_{\mathrm{NLI}}^{(n)}\left(f_{\mathrm{CUT}}\right)$ terms that feed Eq. (1) are found through Eq. (2), where all quantities are related to the n-th span. In it, γ_n is the fiber non-linearity coefficient 1/(W·km). $\bar{G}_{\mathrm{CUT}}^{(n)}$ and $\bar{G}_{n_{\mathrm{ch}}}^{(n)}$ are the *average* PSDs of the CUT and of the n_{ch}-th WDM channel (see Fig.1). They are defined as: $\bar{G}_{\mathrm{CUT}}^{(n)}=P_{\mathrm{CUT}}^{(n)}/R_{\mathrm{CUT}}$ and $\bar{G}_{n_{\mathrm{ch}}}^{(n)}=P_{n_{\mathrm{ch}}}^{(n)}/R_{n_{\mathrm{ch}}}^{(n)}$, where $P_{\mathrm{CUT}}^{(n)}$, $P_{n_{\mathrm{ch}}}^{(n)}$ are the launched power (W) and R_{CUT}, $R_{n_{\mathrm{ch}}}^{(n)}$ are the symbol rates (TBaud), for the CUT and the n_{ch}-th channels, respectively. The round bracket on the right of Eq. (2) contains two terms. One includes the factor $I_{\mathrm{CUT}}^{(n)}$ and accounts for NLI due to the self-channel interference (SCI) of the CUT onto itself. The other term, which includes the factors $I_{n_{\mathrm{ch}}}^{(n)}$, accounts for the cross-channel interference (XCI) of each WDM channel with the CUT. The summation runs over all the WDM channels indices $n_{\mathrm{ch}}=1,\ldots,N_{\mathrm{ch}}^{(n)}$, except the CUT index $n_{\mathrm{CUT}}^{(n)}$. Note that the ACF GN/EGN-model formulas Eqs. (1)-(5) allow for the WDM comb to be different at each span. This is why all channel-related parameters, including the total number of channels $N_{\mathrm{ch}}^{(n)}$, depend on the span index n. The only channel that is assumed to propagate across the whole link is the CUT.

The factors $I_{\mathrm{CUT}}^{(n)}$ and $I_{n_{\mathrm{ch}}}^{(n)}$, Eqs. (3),(4), derive from closed-form approximate solutions of GN-model integrals [11]. They contain: the center frequency f_{CUT} and $f_{n_{\mathrm{ch}}}^{(n)}$ (THz), and the *effective* dispersion $\bar{\beta}_{2,\mathrm{CUT}}^{(n)}$ and $\bar{\beta}_{2,n_{\mathrm{ch}}}^{(n)}$ (ps^2/km), of the CUT and of the n_{ch}-th channel, respectively. The effective dispersions are defined in Eq. (5), where $\beta_2^{(n)}$ and $\beta_3^{(n)}$ are the dispersion (ps^2/km) and dispersion slope (ps^3/km), respectively, of the n-th span fiber. The frequency $f_c^{(n)}$ is where $\beta_2^{(n)}$ and $\beta_3^{(n)}$ are calculated in the n-th span. Note that the 'effective dispersions' originate from an approximation needed to obtain a closed-form formula, which amounts to considering dispersion different from channel to channel, but constant over each individual channel bandwidth [11]. Finally, in Eq. (2) the two factors $\rho_{\mathrm{CUT}}^{(n)}$ and $\rho_{n_{\mathrm{ch}}}^{(n)}$ are best-fit functions meant to turn Eqs. (1)-(5) from an ACF of the *GN*-model into one of the *EGN*-model. They will be discussed in Sect. III-1.

III. Testing the ACF GN/EGN model formula

We focused on the OSNR of the CUT as the system performance parameter used to assess the accuracy of Eqs. (1)-(5):

$$\mathrm{OSNR}_{\mathrm{NL}}=\frac{P_{\mathrm{CUT}}}{P_{\mathrm{ASE}}+P_{\mathrm{NLI}}} \tag{6}$$

where P_{ASE} and P_{NLI} are the ASE and NLI noise powers, calculated over a bandwidth equal to the symbol rate. The subscript 'NL' indicates that NLI noise is included. Note that, assuming a homodyne receiver with a matched filter, $\mathrm{OSNR}_{\mathrm{NL}}$ coincides with the SNR measured on the received constellation (see [3], Sect. IV). For each tested system, $\mathrm{OSNR}_{\mathrm{NL}}$ was estimated for the CUT using both the ACF formulas Eqs. (1)-(5) and the full-fledged numerically-integrated EGN-model. The two results were then compared. Note that P_{NLI} was approximated as $G_{\mathrm{NLI}}^{\mathrm{Rx}}(f_{\mathrm{CUT}})\cdot R_{\mathrm{CUT}}$ when using Eqs. (1)-(5). Instead, when using the numerically-integrated EGN-model, the actual spectrum of NLI across the CUT bandwidth was calculated and then passed through the Rx matched filter.

A potential physical-layer awareness enabling tool, such as the ACF GN/EGN-model addressed here, must be dependable over the widest range of possible situations. We therefore thoroughly randomized the generated test systems. For the *fully-loaded* system test-set, the whole C-band was occupied by channels that could *each* have a different: format, anyone among polarization multiplexed (PM) 16/32/64/128/256-QAM; symbol rate, anyone of 32, 64, 96 and 128 GBaud, with

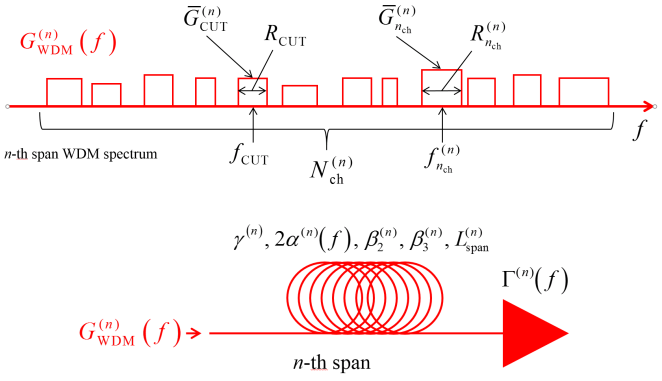

Fig. 1. Top: the WDM comb power spectrum $G^{(n)}_{\mathrm{WDM}}(f)$ at the input of the n-th span along the link. Bottom: the n-th span parameters, including fiber and the lumped elements transfer function $\Gamma^{(n)}(f)$.

associated channel spectral slot size of 43.5, 87.5, 131.25, 175 GHz, respectively; roll-off, uniformly distributed between 0.05 and 0.25. Each link was made up of any combination of three fiber types: SMF, NZDSF1 and NZDSF2. The fiber parameters were, respectively: α_{dB}, 0.21, 0.22, 0.22 dB/km; β_2, -21.3, -4.85, -2.59 ps^2/km; β_3, 0.1452, 0.1463, 0.1206 ps^3/km; γ, 1.3, 1.77, 1.35 1/(W·km). The parameters β_2 and β_3 were always referred to f_c =193.415 THz (1550 nm), which was also assumed as the overall WDM comb center frequency. The combs extended over 5 THz, in the range f_c±2.5 THz. Each span length was generated randomly according to a uniform distribution between 80 and 120 km. The EDFAs noise figure was fixed, at 6 dB.

For each system, $\mathrm{OSNR}_{\mathrm{NL}}$ was estimated at the span number corresponding to the *max-reach* for the CUT, assuming the following target SNRs for PM-16/32/64/128/256QAM, respectively: 11.48, 14.46, 17.00, 19.73, 22.32 dB. These SNRs correspond to a normalized generalized mutual information (NGMI) value of 0.87, for all formats. The WDM channels were launched, on average, at their optimal power into each span but, for further realism, we applied a random launch power deviation from optimum, uniformly distributed between ±30%, different for each channel in the comb. The CUT was instead launched at its optimal power. Finally, we considered three CUT locations: the lowest, mid and highest frequency channels in a comb (LF, MF and HF, for short). To increase system diversity, we generated a different system for each considered CUT. The resulting spread of system scenarios was quite extreme. To mention one indicator, maximum reach ranged from 1 span to 22, covering most of the practical range of terrestrial networks.

Even though the ACF GN/EGN-model formula Eqs. (1)-(5) support frequency-dependent loss and span-by-span changing WDM combs, we did not use these features in the randomization. We leave it for future investigation, where we also plan to address ISRS, which is supported by the ACF [11] too.

1) Accuracy results, full C-band loading: The comparison between the ACF GN/EGN-model formula and numerical integration of the full-fledged EGN-model is displayed in Fig. 2, as the ratio $\left(\mathrm{OSNR}^{\mathrm{CF}}_{\mathrm{NL}}/\mathrm{OSNR}^{\mathrm{EGN}}_{\mathrm{NL}}\right)_{\mathrm{dB}}$ for LF (top), MF (middle) and HF (bottom) CUTs. Full C-band loading is assumed. Each plot addresses 1,000 different systems, for a total of 3,000. In each graph, two histograms are plotted. Next to each histogram, the related mean, standard deviation σ and peak-to-peak spread are tabulated.

The red histograms were obtained using Eqs. (1)-(5) with the factors $\rho^{(n)}_{\mathrm{CUT}}$ and $\rho^{(n)}_{n_{\mathrm{ch}}}$ set to 1. This way, Eqs. (1)-(5) coincide with the ACF formula [11] which approximates the *GN*-model. The histograms show a mean error of about -0.6 dB on all three CUTs, mostly attributable to the known skew between GN and EGN model. The standard deviation is 0.22 dB on the LF-CUTs and MF-CUTs, and 0.28 dB on the HF-CUTs. We consider these overall results quite good, given the extent of the approximations used and the extreme diversity of systems addressed in the test-set. There are, though, rather flung-out outliers, causing a somewhat large peak-to-peak error spread in all red histograms, and in the HF-CUTs in particular, where the peak-to-peak error reaches 1.2 dB.

To improve accuracy, we decided to leverage the large system test-set to best-fit a correction that could turn Eqs. (1)-(5) into a closed-form approximation of the *EGN*-model. We defined the factors $\rho^{(n)}_{\mathrm{CUT}}$ and $\rho^{(n)}_{n_{\mathrm{ch}}}$ in Eq. (2) as:

$$\rho^{(n)}_{\mathrm{CUT}} = a_1 + a_2 \cdot (R_{\mathrm{CUT}})^{a_3} + a_4 \cdot \left(\left|\bar{\beta}_{2,\mathrm{acc}}(n, n_{\mathrm{ch}})\right| + a_5\right)^{a_6}$$
$$\rho^{(n)}_{n_{\mathrm{ch}}} = a_7 + a_8 \cdot \left(\left|\bar{\beta}_{2,\mathrm{acc}}(n, n_{\mathrm{ch}})\right| + a_9\right)^{a_{10}} \quad (7)$$
$$\bar{\beta}_{2,\mathrm{acc}}(n, n_{\mathrm{ch}}) = \sum_{k=1}^{n-1} \bar{\beta}^{(k)}_{2,n_{\mathrm{ch}}} \cdot L^{(k)}_{\mathrm{span}}$$

where $a_1 \ldots a_{10}$ are fitting parameters and $\bar{\beta}_{2,\mathrm{acc}}(n, n_{\mathrm{ch}})$ is the effective accumulated dispersion at the n_{ch}-th channel frequency, from link start till the input of the n-th span fiber. Eq. (7) was conceived based on clues from [9] and [13] and on an extensive numerical study of NLI estimation error sensitivity vs. various physical parameters, which turned out to favor R_{CUT} and $\bar{\beta}_{2,\mathrm{acc}}$ as the most effective ones. The best-fitting of the a_i was performed with the goal of minimizing the mean-square error (MSE) between the CUT NLI PSDs obtained using the ACF formula Eq. (2), $G^{(n)}_{\mathrm{NLI}}(f_{\mathrm{CUT}})$, and the corresponding PSDs obtained through the full-fledged numerically-integrated EGN-model $G^{(n)}_{\mathrm{NLI,EGN}}(f_{\mathrm{CUT}})$. The contributions to the MSE were evaluated *at each system span*, from n=1 to max reach, over a *training* set of 1,500 systems, resulting in the sum of about 20,000 error terms of the form: $|\ G^{(n)}_{\mathrm{NLI}}(f_{\mathrm{CUT}}) - G^{(n)}_{\mathrm{NLI,EGN}}(f_{\mathrm{CUT}})\ |^2$. Of the 1,500 training systems, 750 were fully-loaded and 750 were sparsely-loaded (see next section). The resulting values of $a_1 \ldots a_{10}$ were, in order: -3.1549, 5.5720, 8.5347e-3, -1.7293, 4.8072e-02, -2.0053e-2, -4.1167e-1, 6.1769e-1, 2.1726e1, 7.9148e-2.

Using the best-fitted correction formulas Eq. (7), we obtained the *green* histograms of Fig. 2. For the LF and MF CUTs, the mean error was reduced to negligible values. The standard deviation was brought down to extremely low values (0.05 dB). Peak-to-peak error was less than 0.35 dB. Regarding the HF-CUTs, the results were greatly improved too. However, all HF-CUT error parameters were more than double those of the LF and MF-CUTs. In particular, a few outliers caused the HF-CUTs peak-to-peak error to be 0.72 dB. We examined these outliers and found them to be those where NZDSF2 was prevalent, especially at the start of the link. NZDSF2 has dispersion of only 0.58 ps/(nm·km)) at the

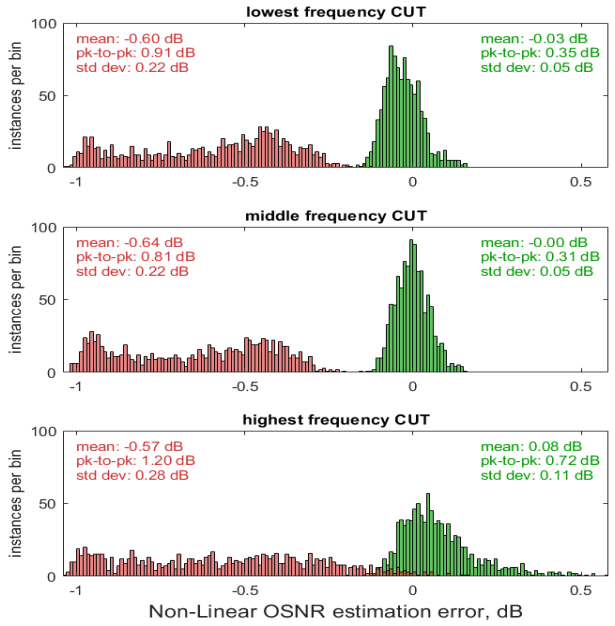

Fig. 2. Histograms of OSNR estimation error, including NLI, at system maximum reach, defined as: $(\mathrm{OSNR}_{\mathrm{NL}}^{\mathrm{ACF}}/\mathrm{OSNR}_{\mathrm{NL}}^{\mathrm{EGN}})_{\mathrm{dB}}$, where $\mathrm{OSNR}_{\mathrm{NL}}^{\mathrm{ACF}}$ is found using the approximate closed-form GN/EGN-model formula Eqs. (1)-(5), and $\mathrm{OSNR}_{\mathrm{NL}}^{\mathrm{EGN}}$ is found through numerical integration of the EGN-model. Each plot shows 1,000 different systems. CUT means 'channel-under-test'. Red histograms: no best-fit. Green histograms: best-fit using Eq. (7).

frequency of the HF-CUTs and, at such low dispersion value, some of the approximations used to derive the CF formulas may fail. We double-checked this explanation by generating a 500-system test-set using only SMF and NZDSF1. We indeed found HF-CUTs errors as low as those of the LF and MF CUTs. We also performed further tests taht showed that, with the current ACF EGN-model formulas, the value of about 1 ps/(nm km) should be considered a practical threshold below which $\mathrm{OSNR}_{\mathrm{NL}}$ estimation error may increase as shown.

2) Accuracy results, sparse C-band loading: To be considered reliable for practical use, an ACF NLI formula should provide accurate results not only at full-load, but also at partial load. We generated 1,200 systems using the same general randomization as for the full-load case. We then turned off anyone of the WDM channels with probability $1/2$, so that the test-set contained systems with an average load of 50% and a wide spread of actual loads. We used the same values of $a_1 \ldots a_{10}$ used in Sect. III-1: no re-fitting was performed since, as mentioned, we had already used sparsely-loaded systems to make up *half* of the 1,500 system training set. For the LF, MF and HF-CUTs, respectively, the mean errors were 0.01, 0.01 and 0.11 dB, the standard deviations were 0.07, 0.05 and 0.16 dB, and the peak-to-peak errors were 0.57, 0.29 and 0.90 dB. Again the HF-CUTs were the most critical, for the same reason as discussed above. Some degradation of accuracy was observed vs. full-load, but overall quite contained.

3) Computational effort: We used the ACF EGN-model formula to characterize all of the 3,000 fully-loaded test-set systems. We did it on a laptop, using interpreted Matlab(TM) code. It took on average about 5 ms to calculate the $\mathrm{OSNR}_{\mathrm{NL}}$ for *all* WDM channels of a single system. This is several orders of magnitude faster than using numerical integration of the EGN or even GN-model and certainly compatible with real-time use.

IV. Conclusion

We tested the accuracy of a closed-form approximate GN-model formula over 3,000 fully-loaded and 1,200 partially-loaded, highly-randomized, C-band systems that used a variety of QAM formats, spacings, symbol rates and 3 different fibers. To the best of our knowledge, this is the first time such an extensive study has been performed. We then greatly improved the accuracy of the model by leveraging the large test-set available, mimicking 'big-data' approaches used in other contexts.

Away from pathological near-zero-dispersion situations, the formula showed very good accuracy in reproducing the results of the full-fledged, numerically-integrated EGN-model, at a comparatively negligible computational effort. We therefore believe the approximate closed-form formula proposed here could potentially provide an effective and accurate tool to support real-time physical-layer-aware management and control of optical networks. Further model upgrades to improve accuracy, encompass Gaussian-shaped and other formats, as well as testing with ISRS over C+L band, are in progress.

References

[1] A. Mecozzi and R.-J. Essiambre, 'Nonlinear Shannon limit in pseudolinear coherent systems,' *J. Lightwave Technol.*, vol. 30, no. 12, pp. 2011–2024, June 15$^{\mathrm{th}}$ 2012.

[2] 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, Nov. 2013.

[3] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, F. Forghieri, 'The GN model of fiber non-linear propagation and its applications,' *J. of Lightwave Technol.*, vol. 32, no. 4, pp. 694–721, Feb. 2014.

[4] 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, June 2014.

[5] P. Serena, A. Bononi, 'A Time-Domain Extended Gaussian Noise Model,' *J. Lightwave Technol.*, vol. 33, no. 7, pp. 1459–1472, Apr. 2015.

[6] M. Secondini and E. Forestieri, 'Analytical fiber-optic channel model in the presence of cross-phase modulations,' *IEEE Photon. Technol. Lett.*, vol. 24, no. 22, pp. 2016–2019, Nov. 15$^{\mathrm{th}}$ 2012.

[7] R. Dar, M. Feder, A. Mecozzi, M. Shtaif, 'Pulse collision picture of inter-channel nonlinear interference noise in fiber-optic communications,' *J. Lightwave Technol.*, vol. 34, no. 2, pp. 593–607, Jan. 2016.

[8] P. Poggiolini, Y. Jiang, 'Recent Advances in the Modeling of the Impact of Nonlinear Fiber Propagation Effects on Uncompensated Coherent Transmission Systems,' tutorial review, *J. of Lightwave Technol.*, vol. 35, no. 3, pp. 458-480, Feb. 2017.

[9] P. Poggiolini, 'The GN model of non-linear propagation in uncompensated coherent optical systems,' *J. of Lightwave Technol.*, vol. 30, no. 24, pp. 3857–3879, Dec. 15$^{\mathrm{th}}$ 2012.

[10] D. Semrau, R. I. Killey, P. Bayvel, 'A Closed-Form Approximation of the Gaussian Noise Model in the Presence of Inter-Channel Stimulated Raman Scattering,' *www.arXiv.org*, paper arXiv:1808.07940, Aug. 23rd 2018.

[11] P. Poggiolini 'A generalized GN-model closed-form formula,' *www.arXiv.org*, paper arXiv:1810.06545v2, Sept. 24th 2018.

[12] P. Poggiolini, M. Ranjbar Zefreh, G. Bosco, F. Forghieri, S. Piciaccia, 'Accurate Non-Linearity Fully-Closed-Form Formula based on the GN/EGN Model and Large-Data-Set Fitting,' in *Proc. of OFC 2019*, paper M1I.4, San Diego (CA), Mar. 2018.

[13] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, 'A simple and effective closed-form GN model correction formula accounting for signal non-Gaussian distribution,' *J. of Lightw. Technol.*, vol. 33, no. 2, pp. 459-473, Jan. 2015.