Nonlinearity Mitigation Using a Machine Learning Detector Based on *k*-Nearest Neighbors

Danshi Wang, Min Zhang, Meixia Fu, Zhongle Cai, Ze Li, Huanhuan Han, Yue Cui, and Bin Luo

***Abstract*— A powerful machine learning detector based on the *k*-nearest neighbors (KNN) algorithm is proposed to overcome system impairments. The zero-dispersion link (ZDL), dispersion managed link (DML), and dispersion unmanaged link (DUL) are considered. Meanwhile, an improved algorithm, the distance-weight KNN, is introduced, which outperforms the conventional maximum likelihood-post compensation approach. The numerical results show that KNN is feasible for overcoming various impairments, especially for non-Gaussian symmetric noise, such as laser phase noise and nonlinear phase noise in the ZDL or DML.**

***Index Terms*— Machine learning, *k*-nearest neighbors (KNN), digital signal process (DSP), nonlinear phase noise (NLPN).**

I. Introduction

COHERENT optical fiber transmission systems are subject to performance degradation induced by noises, among which the amplified spontaneous emission (ASE) noise from the inline amplifier is a major source of linear noise [1], and meanwhile, the laser phase noise from the transmitter and local oscillator (LO) is another factor that strongly affects the system performance [2]. Particularly, nonlinear phase noise (NLPN), induced by the interaction between the signal and ASE noise via the fiber Kerr nonlinearity, is one of the major nonlinear distortion factors [3]. With the development of digital coherent detection, a lot of algorithms based on digital signal processing (DSP) have been proposed to mitigate the noise-induced impairments by using a nonlinear digital filter at the transmitter [4], a Volterra series-based compensator at the receiver [5], a maximum likelihood-based detector [6], and a digital back-propagation equalizer [7]. However, some of the above mentioned methods suffer from complexity and, additionally, the achievable gain in the nonlinear tolerance is dependent on particular transmission scenarios. Hence, the DSP algorithms for nonlinearity compensation require further research.

Manuscript received February 5, 2016; revised March 21, 2016; accepted April 12, 2016. Date of publication April 21, 2016; date of current version August 18, 2016. This work was supported in part by the National Natural Science Foundation of China Project under Grant 61372119, in part by the Doctoral Scientific Fund Project within the Ministry of Education of China, under Grant 20120005110010, and in part by the Beijing University of Posts and Telecommunications Excellent Ph.D. Students Foundation under Grant CX2015306. *(Corresponding author: Min Zhang.)*

D. Wang, M. Fu, Z. Cai, Z. Li, H. Han, Y. Cui, and B. Luo are with the Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: danshi_wang@bupt.edu.cn; meixiafu125@163.com; caizhongle@bupt.edu.cn; vlize@bupt.edu.cn; hanhh1314@126.com; cuiyue0114@bupt.edu.cn; luobin@bupt.edu.cn).

M. Zhang is with the State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: mzhang@bupt.edu.cn).

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

Digital Object Identifier 10.1109/LPT.2016.2555857

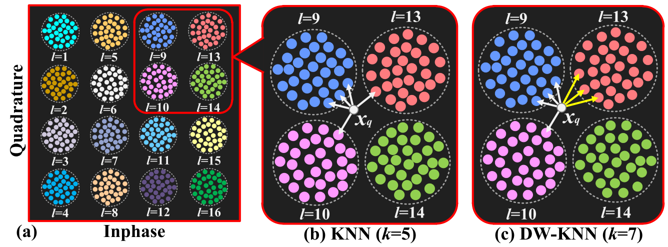

Fig. 1. (a) 16QAM constellations with the corresponding labels; (b) an example of one testing data point $\boldsymbol{x}_q$ detected by KNN (k =5); (c) the testing data point $\boldsymbol{x}_q$ detected by DW-KNN (k = 7).

Machine learning, considered as a powerful interdisciplinary tool, has been applied in various areas, such as data mining, pattern recognition, image processing, and artificial intelligence [8]. Recently, techniques of machine learning have also been used in optical communication systems [9]–[11]. In [9], we demonstrated the feasibility of support vector machine (SVM) for nonlinearity mitigation in a phase-shift keying system. However, the SVM is only a binary classifier; thus, numerous SVMs are necessary for high-order quadrature amplitude modulation (QAM). In [10], the expectation maximum (EM) depends on the parameters of the transmission link, and thus needs improvement for dynamic optical network links. In [11], the artificial neural network (ANN) requires much longer training time. Therefore, it is valuable to investigate a multi-class algorithm that is independent of the link information and free of a training process.

The k-nearest neighbors (KNN) algorithm is one of the top ten machine learning algorithms and has a powerful ability to solve classification and clustering problems [12], [13]. It is a straightforward and effective method, especially in the case of large data sets and low dimensions. However, to the best of our knowledge, the application of KNN to nonlinear impairments mitigation in the context of coherent optical communication has never been reported before.

In this letter, we propose a KNN-based detector to overcome system impairments in a 16QAM coherent optical system. Without any prior information and training process, KNN learns and captures the link properties from just a small set of training data. The zero-dispersion link (ZDL), dispersion managed link (DML), and dispersion unmanaged link (DUL) are considered. Compared with the conventional methods, KNN achieves higher linewidth tolerance, larger launch power range, and longer transmission distance, especially for non-Gaussian noise. Moreover, a modified KNN algorithm, namely the distance-weighted KNN (DW-KNN), is also introduced, which further improves the system performances.

II. Theory of *k*-Nearest Neighbors

A. KNN Algorithm for 16QAM Detection

As a classification algorithm, KNN can be used to identify different classes of data [14]. As shown in Fig. 1(a), the

16QAM signal consists of 16 constellation points, and each constellation point, like a cluster, is composed of many data points. Here, we regard the different constellation points of 16QAM as different classes of data in a two-dimensional (2D) space, where each data point $\boldsymbol{x}$ can be represented by its inphase and quadrature components, corresponding to a 2D vector $\langle I(\boldsymbol{x}), Q(\boldsymbol{x})\rangle$. The whole set of data points is divided into two groups: training data and testing data. The set of training data, $\{(\boldsymbol{x}_1, l_1), \ldots, (\boldsymbol{x}_N, l_N)\}$, consists of N data points $\boldsymbol{x}_i$, with the assigned labels $l_i \in \boldsymbol{L}$. For the 16QAM signal, the label set $\boldsymbol{L} = \{1, 2, \ldots, 16\}$ stands for the specific constellation points, as shown in Fig. 1(a).

Apart from the training data, the rest of the received 16QAM data comprise the set of testing data, where each data point with no label. Here, we take one testing data point $\boldsymbol{x}_q$ as an example to illustrate the principle of KNN, as shown in Fig. 1(b). From the enlarged figure, we can see that $\boldsymbol{x}_q$ is surrounded by four constellation points with labels "9," "10," "13," and "14."

To begin with, all the distances from the testing data point $\boldsymbol{x}_q$ to each training data point need to be calculated. The distance between $\boldsymbol{x}_q$ and an arbitrary training data point $\boldsymbol{x}_i$ is defined as:

$$d(\mathbf{x}_q, \mathbf{x}_i) \equiv \sqrt{(I(\mathbf{x}_q) - I(\mathbf{x}_i))^2 + (Q(\mathbf{x}_q) - Q(\mathbf{x}_i))^2}, \quad (1)$$

According to the calculated distances ordered from the smallest to the largest, we can determine the k nearest training data points, and then obtain the set $N_k(\boldsymbol{x}_q)$ that is composed of their indices. The goal of KNN is to learn a function model f_{KNN} that can provide a reasonable prediction of the class label for the unknown testing data $\boldsymbol{x}_q$. Thus according to the predicted label, the testing data can be classified in terms of the sixteen constellation points. According to $N_k(\boldsymbol{x}_q)$, we can find k labels corresponding to the k nearest training data, and return the majority of these k labels as the predicted label for $\boldsymbol{x}_q$:

$$f_{KNN}(\mathbf{x}_q) = \arg\max_{l \in \mathbf{L}} \sum_{i \in N_k(\mathbf{x}_q)} \delta(l,\ l_i), \quad (2)$$

where $\delta(l, l_i) = 1$ if $l = l_i$, and $\delta(l, l_i) = 0$ if $l \neq l_i$. Note that if k is too large, the massive unrelated training data will be included in the set of $N_k(\boldsymbol{x}_q)$; thus, these unrelated data will interfere with the final classification results significantly.

In our example, k is assumed to be 5, and the 5 nearest data points with the labels {9, 9, 9, 10, 13} are displayed in Fig. 1(b). According to (2), the output of f_{KNN} $(\boldsymbol{x}_q)$ is "9." This is because "9" is the majority of these 5 labels, and more precisely, three-fifths of the nearest data points have the label "9." Finally, the testing data $\boldsymbol{x}_q$ is classified as data with the label "9."

As described above, the KNN model is based on the training data only, but independent of any parameters learned from the training process, so that the testing data can be detected directly. However, in other machine learning algorithms [9]–[11], the testing data cannot be detected directly. These algorithms not only need training data, but also need to finish the training process so as to obtain the parameters of their models, such as the intercept and slope terms of the hyperplane in SVM [9], the parameters that maximize the posteriori probability in EM [10], and the weight and threshold values of each neuron in ANN [11]. Only after that can the testing data be detected by the trained model.

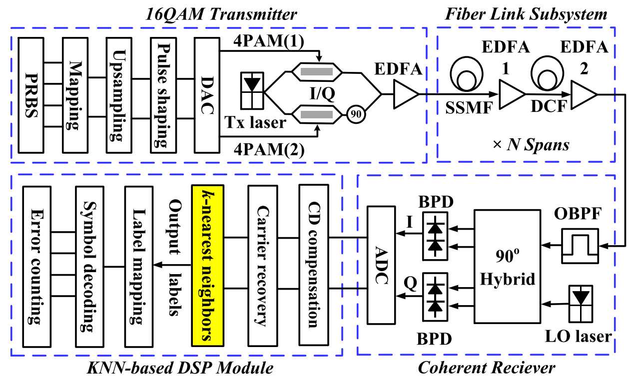

Fig. 2. Numerical model of a 16QAM coherent optical transmission system.

B. Distance-Weighted KNN Algorithm

In some cases, the KNN algorithm still has limitations. For example in Fig. 1(c), when k is set to be 7, among the 7 nearest training data points, the count of the labels "9" and "13" is both three. In this situation, the conventional KNN cannot identify the constellation point to which the testing data point $\boldsymbol{x}_q$ belongs. To address this problem, an improved algorithm, DW-KNN, is proposed to weight the contribution of the k nearest training data points according to their distance to the testing data point $\boldsymbol{x}_q$, giving greater weight to closer training data points. Accordingly, the inverse square of its distance from $\boldsymbol{x}_q$ is set as the weight coefficient. Then Eq. (2) is modified as

$$f_{\omega KNN}(\mathbf{x}_q) = \operatorname*{argmax}_{l \in \mathbf{L}} \sum_{i \in N_k(\mathbf{x}_q)} \omega_i \delta(l,\ l_i), \quad (3)$$

where $\omega_i \equiv 1/d(\boldsymbol{x}_q, \boldsymbol{x}_i)^2$. According to (3), the label "9" will be assigned to the $\boldsymbol{x}_q$ in Fig. 1(c), because the distances to the three training data with label "9" are closer than those with "13." To accommodate the case where the $\boldsymbol{x}_q$ exactly matches one of the training data points $\boldsymbol{x}_i$ for which the denominator $d(\boldsymbol{x}_q, \boldsymbol{x}_i)^2$ is zero, we assign l_i to $f_{\omega KNN}(\boldsymbol{x}_q)$.

III. Numerical Simulation and Results

A. Simulation System Description

Numerical simulations are conducted to demonstrate the feasibility of the KNN-based detector for a 16QAM coherent optical transmission system, as shown in Fig. 2. The laser phase noise from the transmitter and LO lasers (at 1550 nm) are modeled as a random walk Wiener process. The I/Q modulator is driven by two four-level pulse amplitude modulated (4PAM) electrical signals. A module of the 4PAM generator performs the generation of pseudo-random binary sequence with a length of 2^{16}-1, signal mapping, upsampling, multilevel pulse shaping with a raising time equal to 1/8 of the symbol duration, and digital-to-analog conversion. Through the I/Q modulator, a 16QAM signal is generated at 25 Gbaud (i.e., 100 Gbps).

Different fiber links may influence the signal propagation in different ways. Therefore, in the fiber link subsystem, the ZDL, DML, and DUL are considered. The DML consists of a different number of spans where each span consists of 80 km of standard single mode fiber (SSMF) and 17 km of dispersion compensating fiber (DCF). For SSMF and DCF, we have the following fiber parameters: loss coefficient $\alpha_{smf} = 0.2$ dB/km, $\alpha_{dcf} = 0.5$ dB/km; dispersion parameter

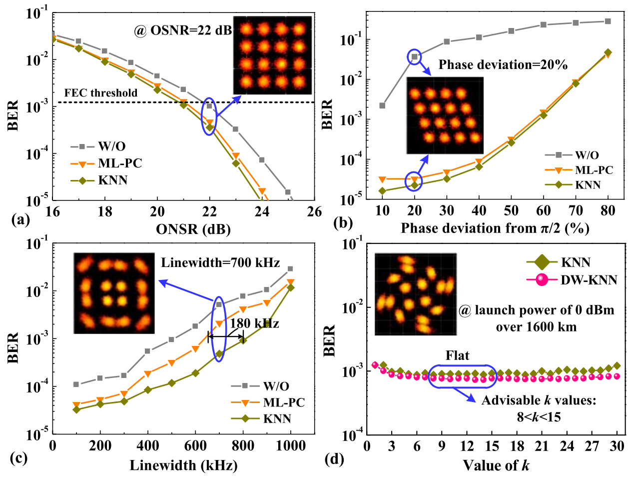

Fig. 3. BER as a function of (a) OSNR ($k = 10$); (b) phase deviation from $\pi/2$ ($k = 10$, OSNR $= 24$ dB); (c) linewidth ($k = 10$, modulation depth $= 2.2$, phase shift $= 5\%$); (d) k values (fiber length $= 1600$ km, launch power $= 0$ dBm).

$D_{smf} = 17$ ps/nm/km, $D_{dcf} = -80$ ps/nm/km; and nonlinear coefficient $\gamma_{smf} = 1.3$ W^{-1}km^{-1}, $\gamma_{dcf} = 5.3$ W^{-1}km^{-1}. For the DUL, each span only consists of an SSMF. An erbium-doped fiber amplifier (EDFA) with a noise figure of 6 dB is placed after the SSMF and DCF span to compensate for the fiber attenuation.

At the receiver, a Gaussian optical filter with 28 GHz bandwidth is used first to reduce the ASE noise. Then the signal is coherently detected in a 90° optical hybrid, photodetected, and sampled by 2 samples/symbol. Then, the sampled data are sent to the DSP module. In the DSP phase, the chromatic dispersion (CD) is compensated first by the method of [15], and then the carrier phase is recovered by the phase rotation method of [16]. In our simulations, we assume that the sampling signal has already been synchronized with the incoming signal so that the clock recovery is neglected. Then the sampled components are processed by the KNN-based detector to overcome the system impairments. The training data set is comprised of 1000 symbols, and the rest of the symbols are the testing data. Finally, the output labels from KNN are mapped to the corresponding symbols. After the decoding, bit-error rate (BER) is counted.

B. Results and Discussions

1) Back-to-Back (BTB) Investigation: In the BTB case, we mainly investigate the feasibility of KNN for three kinds of system impairments: i.e., ASE noise, I/Q imbalance, and laser phase noise. Meanwhile, the popular algorithm of maximum likelihood-post compensation (ML-PC) [6] is selected for comparison. The ML-PC is performed based on the probability density function (PDF), and the parameters of PDF, such as expectation and variance, are estimated based on the maximum likelihood function. The Gaussian distribution is adopted in [6], and in our scheme, the number of training data for ML-PC is equal to that for KNN, namely 1000 symbols.

First, only ASE noise is added to the 16QAM signal in the back-to-back (BTB) case. The BER is measured by KNN and ML-PC as a function of the optical signal to noise ratio (OSNR). As a reference, the linear decision without any mitigation algorithm (W/O) is also tested, as shown in Fig. 3(a). We can see that the curves of KNN and ML-PC almost overlap, and for the case without mitigation, only a slight improvement is observed. From the constellation inset of Fig. 3(a), it is seen that each constellation point dominated by the ASE noise is a symmetric Gaussian distribution. In this case, the ML-PC based on PDF of Gaussian is effective, and the decision boundaries are linear; hence, there is little benefit from KNN.

Next, we investigate whether KNN can overcome the I/Q imbalance. When there is an I/Q imbalance, the phase difference between the I- and Q-branch might not be exactly 90°. The inset of Fig. 3(b) displays the 16QAM constellation whose phase difference deviates from 90° by 20%. We can see that the mismatch of the amplitude and phase between the I/Q components distorts the constellations seriously. Then we measure BER as a function of the phase deviation from 90°. The results show that without compensation, the BER performance sharply deteriorates from the phase deviation of 10%, and after 30%, the BER even raises up to 10^{-1}. On the other hand, KNN can effectively compensate the deterioration induced by the I/Q imbalance. Compared with ML-PC, although KNN achieves better performance, the improvement is insignificant. This is mainly because the I/Q imbalance distorts the overall shape of the constellation plane, but each constellation point still keeps its circular shape with a symmetric Gaussian distribution.

However, the I/Q imbalance is often accompanied by other impairments, such as laser phase noise induced by the laser linewidth. Thus, we investigate the combined effects of I/Q modulator imperfection and laser phase noise. The 16QAM constellation is clearly distorted by setting the modulation depth of the modulator to $m = V_{PP}/V_{\pi} = 2.2$. The phase deviation from 90° is 5%. In Fig. 3(c), the BER is plotted as a function of the combined laser linewidth and phase deviation of 5%. It is seen that, compared with that of ML-PC, the linewidth tolerance of KNN is increased by 180 kHz, which demonstrates that KNN is efficient for the laser phase noise.

2) Three Types of Fiber Links: First, to focus on the NLPN only, the dispersion of the fiber link is numerically set to zero, i.e., ZDL. After 1600 km transmission with 0-dBm launch power, the 16QAM signal is detected by the coherent receiver. Suffering from NLPN, the 16QAM signal (before carrier phase recovery) deteriorates significantly, as shown in Fig. 3(d). The entire constellation is distorted (phase offset introduced) and the outer points are compressed. First, the influence of k is studied. The NLPN-dominated signal is detected by the KNN and DW-KNN detectors. The values of k range from 1 to 30 and corresponding BERs are calculated. From Fig. 3(d), we can see that different k values may affect BER values, but the BER curves fluctuate gently over the entire k values, especially for the selected scope from 8 to 15. Therefore, the advisable k value should be in the flat range from 8 to 15, which can satisfy most of the practical optical communication systems.

The BER as a function of the launch power with k assigned as 10 is shown in Fig. 4(a). The carrier phase recovery is added and the launch power of the 16QAM signal (over 1600 km ZDL) ranges from -8 dBm to 4 dBm. It is seen that as the launch power increases, the received OSNR grows higher, which indicates a gradually decreasing ASE noise and contributes to a better BER performance. As the launch power exceeds the optimal value, the received signal is impaired by a remarkable nonlinear effect and the BER performance is degraded again. The numerical results show that DW-KNN

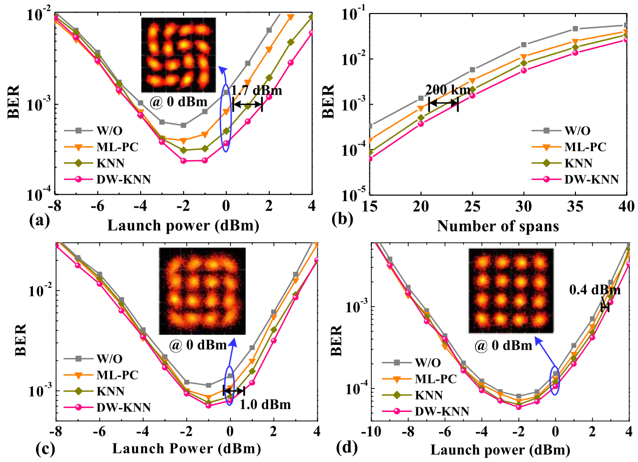

Fig. 4. BER as a function of (a) launch power (zero-dispersion link, fiber length = 1600 km), (b) transmission distance (zero-dispersion link, launch power = 0 dBm), (c) launch power (dispersion managed link, fiber length = 1200 km), and (d) launch power (dispersion unmanaged link, fiber length = 1200 km).

achieves the best performance, and the nonlinear tolerance is increased by ~1.7 dBm for a BER of 10^{-3}. Then we study the BER as a function of the transmission distance, as shown in Fig. 4(b), where the transmission distance at a BER of 10^{-3} is increased by ~200 km.

In addition to the ZDL, we also consider the DML where the CD is compensated by the DCF in the link, as well as the DUL where the CD is compensated by the DSP method [15] at the receiver. The BERs as functions of launch power are measured in DML and DUL, respectively, as shown in Fig. 4(c)–(d). The total transmission distance is about 1200 km and carrier phase recovery is performed. By employing DW-KNN, the improvement in the nonlinear tolerance is 1.0 dBm for DML and 0.4 dBm for DUL. This is because, for DUL, the interplay between dispersion and nonlinearities results in a more circularly symmetric Gaussian-like noise in contrast to the non-circular noise in the case of DML where the variance of phase noise is larger than that of amplitude noise [17]. To illustrate this visually, the 16QAM constellations at 0 dBm after phase recovery are shown in the insets of Fig. 3(a), (c), (d).

As demonstrated above, KNN has the potential to overcome various system impairments. However, compared with ML-PC, KNN contributes to different superior performances in different cases. The ML-PC based on the PDF of Gaussian is effective for the circularly symmetric Gaussian-like impairments, such as ASE noise, I/Q imbalance, and NLPN in the DUL. However, for non-Gaussian impairments, such as laser phase noise and the NLPN in the ZDL or DML, the ML-PC only yields suboptimal results. In contrast, the KNN based on the received data but independent of the PDF can not only be applied to various transmission links owing to its learning capacity, but also mitigate non-Gaussian impairments.

IV. Conclusions

In this letter, an effective detector based on the KNN method was proposed in the context of a coherent optical transmission system. Various system impairments were investigated. Compared with ML-PC, KNN increased the linewidth tolerance by 180 kHz, and improved the nonlinear tolerance for the dispersion neglected, managed, and unmanaged links by 1.7, 1.0, and 0.4 dBm, respectively. In the ever expanding optical networks, where all sorts of links coexist, the dynamic routing across several fiber spans plays an increasing number of roles in replacing the traditional semi-static single-type fiber connect. Therefore, we believe that KNN will be more useful to mitigate nonlinear impairments in the heterogeneous fiber spans. Note that we mainly investigated the single channel and single polarization case; other nonlinear effects among multiple channels and polarization effect are not taken into account. We think that associated with other DSP algorithms, KNN could also be feasible in these cases, which will be further demonstrated and investigated in our following work.

References

[1] J. Lin, L. Xi, J. Li, X. Zhang, X. Zhang, and S. A. Niazi, "Low noise optical multi-carrier generation using optical-FIR filter for ASE noise suppression in re-circulating frequency shifter loop," *Opt. Exp.*, vol. 22, no. 7, pp. 7852–7864, 2014.

[2] T. Okamoto and F. Ito, "Laser phase noise characterization using parallel linear optical sampling," *J. Lightw. Technol.*, vol. 32, no. 18, pp. 3119–3125, Sep. 15, 2014.

[3] N. Ekanayake and H. M. V. R. Herath, "Effect of nonlinear phase noise on the performance of M-ary PSK signals in optical fiber links," *J. Lightw. Technol.*, vol. 31, no. 3, pp. 447–454, Feb. 1, 2013.

[4] K. Roberts, C. Li, L. Strawczynski, M. O'Sullivan, and I. Hardcastle, "Electronic precompensation of optical nonlinearity," *IEEE Photon. Technol. Lett.*, vol. 18, no. 2, pp. 403–405, Jan. 15, 2006.

[5] F. P. Guiomar, J. D. Reis, A. L. Teixeira, and A. N. Pinto, "Digital postcompensation using volterra series transfer function," *IEEE Photon. Technol. Lett.*, vol. 23, no. 19, pp. 1412–1414, Oct. 1, 2011.

[6] A. Tan, H. Wymeersch, P. Johannisson, E. Agrell, P. Andrekson, and M. Karlsson, "An ML-based detector for optical communication in the presence of nonlinear phase noise," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2011, pp. 1–5.

[7] A. Napoli *et al.*, "Reduced complexity digital back-propagation methods for optical communication systems," *J. Lightw. Technol.*, vol. 32, no. 7, pp. 1351–1362, Apr. 1, 2014.

[8] C. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[9] D. Wang *et al.*, "Nonlinear decision boundary created by a machine learning-based classifier to mitigate nonlinear phase noise," in *Proc. Eur. Conf. Exhibit. Opt. Commun. (ECOC)*, Sep./Oct. 2015, pp. 1–3, paper p.3.16.

[10] D. Zibar *et al.*, "Nonlinear impairment compensation using expectation maximization for dispersion managed and unmanaged PDM 16-QAM transmission," *Opt. Exp.*, vol. 20, no. 26, pp. B181–B196, 2012.

[11] M. A. Jarajreh *et al.*, "Artificial neural network nonlinear equalizer for coherent optical OFDM," *IEEE Photon. Technol. Lett.*, vol. 27, no. 4, pp. 387–390, Feb. 15, 2015.

[12] O. Kramer, *K-Nearest Neighbors: Dimensionality Reduction With Unsupervised Nearest Neighbors*. Berlin, Germany: Springer, 2013.

[13] P. Harrington, *Machine Learning in Action*. Littleton, CO, USA: Manning Company, 2012.

[14] T. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.

[15] S. J. Savory, "Digital filters for coherent optical receivers," *Opt. Exp.*, vol. 16, no. 2, pp. 804–817, Jan. 2008.

[16] A. P. T. Lau and J. M. Kahn, "Signal design and detection in presence of nonlinear phase noise," *J. Lightw. Technol.*, vol. 25, no. 10, pp. 3008–3016, Oct. 2007.

[17] S. Kumar, "Effect of dispersion on nonlinear phase noise in optical transmission systems," *Opt. Lett.*, vol. 30, no. 24, pp. 3278–3280, 2005.