Physical Communication 58 (2023) 102042

Contents lists available at ScienceDirect

Physical Communication

journal homepage: www.elsevier.com/locate/phycom

Full length article

Active learning for OPM in FMF systems

M.A. Amirabadi, M.H. Kahaei, S.A. Nezamalhosseini *

School of Electrical Engineering, Iran University of Science and Technology (IUST), Tehran, 1684613114, Iran

ARTICLE INFO

Article history:
Received 31 October 2022
Received in revised form 24 January 2023
Accepted 28 February 2023
Available online 9 March 2023

Keywords:
Active learning
Optical performance monitoring
Few-mode fiber
Nonlinearity

ABSTRACT

Optical performance monitoring (OPM) is essential to guarantee the robust and reliable operation of few-mode fiber (FMF)-based transmission. The available OPM methods including the analytical models such as the enhanced Gaussian noise model provide high accuracy along with high computational complexity which makes them improper for real-time implementations. As an alternative approach, machine learning (ML)-based OPM removes this barrier at the cost of leveraging a large training dataset. However, generating a field or synthetic dataset for FMF-based transmission is very hard and time-consuming. As a specific ML deployment, active learning (AL) is designed to work with a small training dataset, therefore, in this paper, we employ AL for OPM in FMF-based transmission. Results indicate that the proposed AL-based OPM can properly estimate the generalized signal-to-noise ratio by using a very small training dataset and achieve the root mean squared error similar to that obtained by working on large training datasets.

1. Introduction

Mode division multiplexing (MDM) through few-mode fibers (FMFs) is a new multiplexing technique applicable for data rate enhancement in optical fiber communications [1–3]. FMFs always have a substantial linear coupling among modes, especially in long-range transmission, therefore, the receiver needs to be equipped with multi-input multi-output (MIMO) digital signal processing (DSP) like those widely used in radio frequency communication [4]. The available FMFs are step-index or graded-index where the second case is optimized to have the minimum differential modal group delay and a simple receiver [5]. There are special optical amplifiers used, mostly erbium-doped fiber amplifiers (EDFAs), for maintaining launched power in FMFs, and are optimized to have the minimum mode-dependent gain variation [6]. The inter and intra-modal Kerr-based nonlinear (coupling) effects severely degrade the performance of the FMF system.

Optical performance monitoring (OPM) is essential for improving control of transmission and physical layer fault management in high-capacity FMF-based transmission and switching [7,8]. OPM evaluates the signal quality by measuring optical phenomena and physical characteristics such as generalized signal-to-noise ratio (GSNR) without directly assessing the transmitted data sequence. GSNR estimation can be formulated in terms of amplified spontaneous emission (ASE) noise and nonlinear interference (NLI) noise [7]. The linear noise calculation is simple, while the NLI noise computation is challenging. To predict the NLI, the Gaussian noise (GN) model [9,10] and enhanced GN (EGN) model [11–13] are the most popular methods which consider the FMF nonlinearity as an additive Gaussian noise source. The GN model assumes Gaussian distribution for the propagating signal while the EGN model takes into account different modulation formats. These methods have computational complexity which makes them inappropriate for real-time applications.

1.1. Literature review

Recently, machine learning (ML) has been widely used as a fast and accurate GSNR estimation method [14–17]. ML models can be trained to learn the relationship between GSNR and the FMF system/link parameters such as transmitted power, span length, number of spans, ASE noise power, polarization/mode-dependent attenuation, mean attenuation, chromatic dispersion, and modal dispersion. ML-based OPM requires a huge dataset which slows down the training process, in addition, a lot of new training samples should be added when FMF system/link conditions change which increases training epochs [18–20]. However, generating a large field/synthetic dataset for FMF is very hard and consumes a lot of time [14]. In this context, active learning (AL), a special ML implementation, can be leveraged which explicitly requests label-specified features to improve training performance. AL can be adopted for cleverly choosing the training points to be acquired for improving OPM accuracy by minimizing an acquisition function tailored to the ML-based OPM model.

* Corresponding author.
E-mail addresses: m_amirabadi@elec.iust.ac.ir (M.A. Amirabadi), kahaei@iust.ac.ir (M.H. Kahaei), nezam@iust.ac.ir (S.A. Nezamalhosseini).

https://doi.org/10.1016/j.phycom.2023.102042
1874-4907/

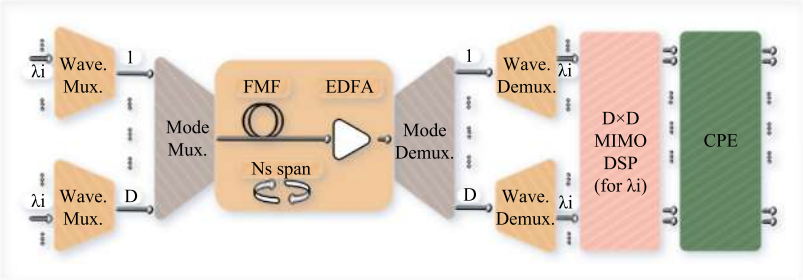


Fig. 1. Schematic of the considered FMF system model.

1.2. Motivations, novelties, and contributions

FMFs encounter additional impairments to those that exist in the single-mode fiber (SMF) including mode-dependent attenuation, mode-dependent gain, modal dispersion, linear coupling, and nonlinear coupling. Although FMF linear effects are reversed by the MIMO DSP, they have a consistent impact on FMF nonlinearity. This impact and the nonlinear coupling are the specificities of FMF nonlinearity that should be considered while designing an AL model for FMF systems. Therefore, the trained AL model for SMF transmission cannot be applied to FMF transmission and need to be re-tuned and re-trained.

The potential deployment of FMF in future optical networks necessitates investigating the applicability of AL-based OPM for FMF-based transmission. Despite this importance, the only existing works consider the quality of transmission estimation in SMF networks [18–20] and there is no investigation over AL-based OPM for FMF-based transmission. Therefore, in this paper, we present and develop AL-based OPM for GSNR estimation in FMF-based transmission by utilizing FMF system/link configurations as features. The novelties and contributions of this work include

- Proposing AL-based OPM in FMF transmission, having superior capabilities of fast data resource reservation enabling the possibility for real-time applications,
- Presenting a comprehensive investigation over AL-based OPM considering different performance criteria, training dataset length, acquiring dataset length, and ML hyperparameters.

The rest of this paper is organized as follows; Section 2 describes the system model, and Section 3 explains the proposed AL-based OPM approach. Simulation results and discussions are provided in Section 4, and Section 5 is the conclusion of this work.

2. System model

We consider the FMF system depicted by Fig. 1 composing of N_s spans with EDFA at the end of each span for maintaining signal power. The propagating signal is a multiplexing of 2 polarizations, D spatial modes, and N_{ch} frequency channels. The FMF propagation suffers from linear and nonlinear effects such as modal dispersion, chromatic dispersion, linear coupling, Kerr-based nonlinearity, and nonlinear coupling. At the receiver, after the demultiplexer, the MIMO DSP compensates for linear effects, and the carrier phase estimator (CPE) recovers the nonlinear phase rotation. Based on the well-known EGN model, the received signal after CPE can be modeled as a summation of the transmitted signal and ASE and NLI noise [11,21]. Therefore, the GSNR of nth channel and pth mode after CPE can be formulated as [11]

$$GSNR_{n,p} = \frac{P_{n,p}}{\sigma^2_{ASE,n,p} + \sigma^2_{NLI,n,p}}, \tag{1}$$

where $P_{n,p}$, $\sigma^2_{ASE_{n,p}}$, and $\sigma^2_{NLI,n,p}$ respectively are launched power, ASE and NLI noise variances of nth channel and pth mode [11].

Demonstrating AL's ability to learn a well-behaved summation of triplets (EGN) is not beyond the scope of original AL demonstrations. Since EGN does not capture the stochastic performance variability associated with the general case. Moreover, using the EGN model in this form to propagate the optical signal makes the problem to be a trivial solution for AL. Therefore, here we consider the same approach as [22] and account for the uncertainties by modeling the gain and noise figure ripples as functions of mode-frequency and introducing randomization in the amplifier gain and noise figure, to simulate uncertainties in these parameters. In particular, we introduce a random offset and tilt into the noise figure ripple samples while preserving their correlation with the mode-frequency.

3. Proposed AL-based OPM approach

The proposed AL-based OPM relies on Gaussian process regression (GPR) [23], here, we first bring preliminaries about GPR and then explain the procedure of the proposed GPR-based AL approach.

3.1. GPR preliminaries

GPR is a probabilistic non-parametric regression algorithm that provides both a prediction and a quantification of prediction uncertainty [23]. GPR reliability and fast-to-compute uncertainty quantification are the two characteristics that make it proper for AL. In particular, we assume that only a small training dataset is available. The GPR observes a training set of l points $\mathbf{X}_l = \{\mathbf{x}_1, \ldots, \mathbf{x}_l\}$, coupled with response values $\mathbf{y} = \{y_1, \ldots, y_l\}$ where $y_i = f(\mathbf{x}_i) + \epsilon$; $i = 1, \ldots, l$ and $\epsilon \sim N(0, \sigma^2_{noise})$. This observation model expresses that the GSNRs calculated by the latent function $\mathbf{f}$ are corrupted by measurement noise ϵ due to the randomization in EDFA gain and noise figure. We denote the latent function values by $\mathbf{f} = (f(x_1), \ldots, f(x_l))$. The observation model can be summarized by $p(\mathbf{y}|\mathbf{f}) = N(\mathbf{f}, \sigma^2_{noise})$ which permits to handle dataset points wherein two observations with the same input features generate slightly different GSNR values, to increase dataset variability and produce uncertainty.

GPR considers $\mathbf{f}$ as a realization of a GP, therefore, in a Bayesian sense, we can assume that the $\mathbf{f}$ has a prior distribution given by $p(\mathbf{f}) = N(m(\mathbf{X}_l), K)$ with $m(\mathbf{X}_l) = (m(\mathbf{x}_1), \ldots, m(\mathbf{x}_l))$ and K where $K_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$ is a positive definite kernel. The mean and kernel functions should be selected before observing any data and encoding prior knowledge. The prior mean encodes trends of the latent function known before observing the data, and the prior kernel defines the smoothness of the GPR fit. Note that, GSNR does not have a priori specific properties, we select zero mean with a kernel from a stationary family, to have a dependency on a few hyperparameters to be tuned based on the input dataset. Based on the Bayes theorem, we combine the prior and

Algorithm 1: Proposed AL-based OPM approach.

Initialization: iteration counter $l = 0$, and training dataset $T_0 = T$;
while *The number of acquirable samples is exhausted* **do**
  Train regressor model with T_l ;
  Build acquisition function based on the observation model and prior variance;
  Find next input $\mathbf{x}_{l+1}$ maximizing accusation function;
  Update training set with $T_{l+1} = T_l \bigcup \{(\mathbf{x}_{l+1}, \mathbf{y}_{l+1})\}$;
  Update $l = l + 1$;
end

observation model to obtain the posterior distribution $p(\mathbf{f}|\mathbf{y}) = p(\mathbf{f})p(\mathbf{y}|\mathbf{f})/p(\mathbf{y})$ which is normally distributed with analytical mean and covariance expressions (see [23], chapter 2).

3.2. GPR-based AL

The AL-based OPM adaptively (not randomly) increases both training dataset size and regressor confidence by adding new dataset points from the pooling dataset by minimizing an acquisition function. The training and pooling datasets respectively include the labeled points and the potential points to query for a label. The acquisition function provides an uncertainty measure of a pooling dataset point, and is the variance in GPR. GPR associates with any new untried input the posterior variance of the GP, assuming this new point is added to the training set. For more details about GPR-based AL, we refer to [24,25].

Given a training set $\mathbf{X}_l$ and a new point $\mathbf{x}_{l+1}$, we can compute the posterior variance $s^2_{l+1}(\mathbf{x})$ at any input location $\mathbf{x}$ and define the acquisition function with the formula

$$s^2_{l+1}(\mathbf{x}) = s^2_l(\mathbf{x}) - \frac{k_l(\mathbf{x}, \mathbf{x}_{l+1})^2}{k_l(\mathbf{x}_{l+1}, \mathbf{x}_{l+1}) + \sigma^2_{noise}}, \tag{2}$$

where $s^2_l(.)$ and $k_l(.,.)$ are respectively the posterior variance and covariance kernel given $(\mathbf{X}_l, y_l)$. Here, we do not need the value of the latent function $\mathbf{f}$ at $\mathbf{x}_{l+1}$, thus, $s^2_{l+1}(\mathbf{x})$ can be computed before measuring GSNR. Then we select the next input location $\mathbf{x}_{l+1}$ by minimizing the acquisition function. To do so, we envision a strategy that improves the training set by adding instances in such a way that the future uncertainty (GPR variance over the pooling dataset) is minimized. Thus, at each step, we find (and add) only one point with the highest uncertainty (variance) from the pooling dataset to the training dataset. In other words, we pre-select a pool of possible points and calculate s^2_{l+1} at these points and select the maximizer point as $\mathbf{x}_{l+1}$. The proposed AL-based OPM approach is summarized in Algorithm 1.

4. Simulation results and discussions

In this section, we first explain dataset generation and GPR hyperparameter tuning. Then, we provide the performance and complexity analysis of the proposed AL-based OPM approach for FMF-based transmission. Simulations are done in Python environment [26,27], Scikit-learn library [28].

4.1. Dataset generation

FMF system and link configuration: Dataset generation is the main issue while training an ML model for OPM in FMF. Since generating a large field dataset is quite hard while considering different system and link configurations, also generating a large synthetic dataset by split-step Fourier method (SSFM) simulation is impractical while dealing with a few modes and the whole C-band [29]. For instance, in our case, we had access to an Intel Xeon CPU with 32 cores and 64 GB RAM by which we were able to apply SSFM simulation considering only 3 spatial modes and 9 channels (0.45 THz bandwidth).

Table 1
Simulation parameters.

Parameter	Value
Number of modes	3
Number of channels	66
Center frequency	193.5 THz
Symbol rate	64 GBaud
Channel bandwidth	75 GHz
Launched power	−5 dBm to 5 dBm
Number of spans	1–8
Span length	80–120 km
Coupling length	80 km

Table 2
Nonlinear coupling coefficient between pth and qth mode for FMF type 1 [30] and 2 [31].

Type 1				Type 2			
pq	LP01	LP11a	LP11b	pq	LP01	LP11a	LP11b
LP01	1	0.661	0.661	LP01	0.73	0.36	0.36
LP11a	0.660	1.053	1.053	LP11a	0.36	0.55	0.18
LP11b	0.660	1.053	1.053	LP11b	0.36	0.18	0.55

In this paper, we synthetically generated a dataset based on the EGN model [11] considering system and link parameters defined in Tables 1, 2, and 3. At each realization, we randomly select one of the two FMF types reported in the Tables 2 and 3. and then consider a random 0.1 variation for each of the reported FMF parameters (i.e. mode-dependent attenuation, modal dispersion, chromatic dispersion, and nonlinear (coupling) coefficients), to produce more variability in the dataset and prevent the proposed AL-based model to be highly specific to these FMF types. We consider 3 spatial modes and 66 channels (5 THz bandwidth, C-band). Transceivers operate at 193.5 THz center frequency, 64 Gbaud symbol rate, and 75 GHz optical bandwidth. We use polarization multiplexed-quadrature phase shift keying modulation format. Transmission adopts launched powers per channel and mode uniformly selected between −5 dBm to 5 dBm. Fiber spans are assumed to be uniformly selected between 1–8 in the range 80–120 km (with 1 m granularity). Both weak and strong linear coupling regimes are considered with a coupling length of 80 km (i.e., after 80 km propagation the spatial modes are strongly coupled together). It should be noted that linear coupling accommodates continuously along spans while propagation, in the sense that after 80 km propagation, all spatial modes encounter strong coupling even if all spans have 80 km length. At the end of each span, an EDFA compensates for mode-dependent attenuation. We calculate EDFA gain by considering 0.8 dB EDFA gain tilt between LP01 and LP11a/b modes [32–35]. We consider EDFA gain ripples randomly chosen between $[-0.1, 0.1]$ dB [36]. Then, we compute the noise figure by $F = (2n_{sp}(G-1)+1)/G$ [37] with $n_{sp} = 1.58$ [37,38]. The motivation behind the randomization in gain and noise figure is to simulate uncertainties in these parameters.

Feature vector: The feature vector contains launched power, modal dispersion, chromatic dispersion, nonlinearity coefficient, nonlinear coupling coefficient, span length, number of spans, mode-dependent attenuation, noise figure, EDFA gain, as well as indices of channel and mode under test. We selected these features as they affect the GSNR value (label) through the EGN model. The GSNR calculation is replicated 50 times per FMF system/link configuration, and the GSNR value at each realization

Table 3
Nonlinearity coefficient, attenuation, modal dispersion, and chromatic dispersion for FMF type 1 [30] and 2 [31].

Type 1				Type 2			
Parameter	LP01	LP11a	LP11b	Parameter	LP01	LP11a	LP11b
Nonlinearity coefficient [1/(Watt km)]	1.3	1.3	1.3	Nonlinearity coefficient [1/(Watt km)]	1	1	1
Mode dependent attenuation [dB/km]	0.221	0.226	0.226	Mode dependent attenuation [dB/km]	0.201	0.206	0.206
Modal dispersion [ps/km]	0	6.5	6.5	Modal dispersion [ps/km]	−0.29	−0.66	−0.66
Chromatic dispersion [ps^2/km]	31.9	34.8	34.8	Chromatic dispersion [ps^2/km]	28.3	28.2	28.2

Table 4
Considered kernel formulations and their selected hyperparameter values.

Kernel	Formula	Parameter
RBF	$k(x_i, x_j) = exp(-\frac{\lvert x_i - x_j\rvert^2}{2l^2})$	$l = 1$
Matern	$k(x_i, x_j) = \frac{1}{\Gamma(\nu)2^{\nu-1}}(\frac{\sqrt{2\nu}}{l}\lvert x_i - x_j\rvert^2)^{\nu} K_{\nu}(\frac{\sqrt{2\nu}}{l}\lvert x_i - x_j\rvert^2)$	$l = 1, \nu = 1.5$
White kernel	$k(x_i, x_j) = c_1\delta(x_i, x_j)$	$c_1 = 1$
Constant kernel	$k(x_i, x_j) = c_2$	$c_2 = 1$

is different due to the introduced randomization. We generate 2000, 1000, and 10 000 training, pooling, and test dataset points. We used different random seeds while dataset generation and shuffled them at the end to have statistically independent dataset points.

Feature preprocessing: The ML models tend to weight higher the features with bigger variances, i.e. the features with variances orders of magnitude higher than the other features prohibit learning from others. Thus, these features should be presented in a fixed range, e.g. by normalization scales each feature to have a unit norm. Here, we employ Min–Max normalization as a widely used method for scaling the features between 0 and 1. Min–Max normalization scales the ith feature, $x_i \in [x_{i,min}, x_{i,max}]$ into $\bar{x}_i \in [0, 1]$ as $\bar{x}_i = (x_i - x_{i,min})/(x_{i,max} - x_{i,min})$ with $x_{i,min}$, $x_{i,max}$ as minimum and maximum of ith feature.

Feature space: The feature space sweeps an infinite area, in the following we explain more. The channel and the mode indices, and the number of spans take 66, 3, and 8 integer values, respectively. The span length and launched power of the channel and mode under test are uniformly distributed and impose infinite values. Mode-dependent attenuation takes $198(3 \times 66)$ different values. Each of the modal dispersion, chromatic dispersion, nonlinearity coefficient, nonlinear coupling coefficients, noise figure, and EDFA gain are fully randomized and take infinite values. Calculating GSNR for each FMF system/link configuration takes almost half an hour using the EGN model, besides, a small change in FMF system/link configuration changes the GSNR value (see [1,6]). In conclusion, substituting the EGN model with even a huge look-up table based on pre-calculated EGN values is not practical.

4.2. Hyperparameter tuning

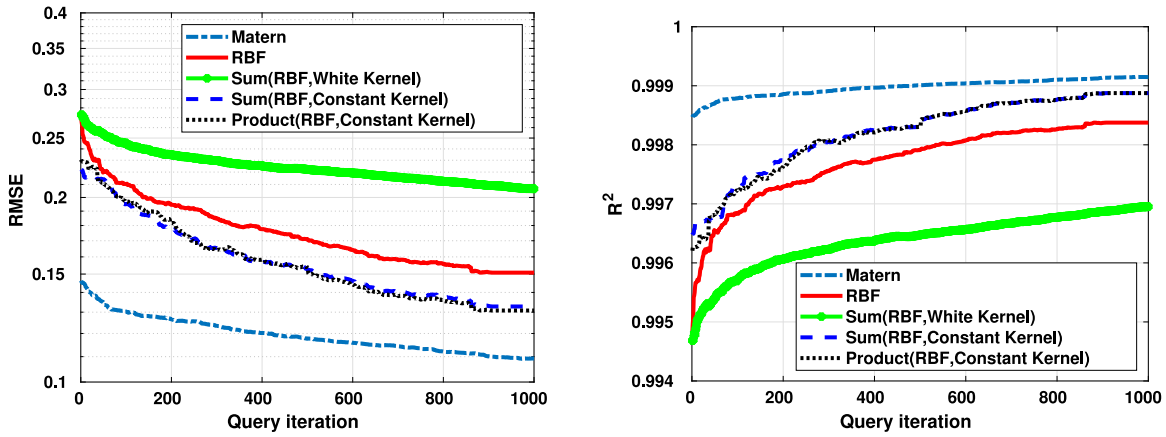
Optimizing GPR hyperparameters: Although GPR is a non-parametric ML method, it is specified by mean and kernel functions which should be defined before GPR deployment. As explained in Section 3.1, GSNR does not have a priori specific properties, therefore, we select zero mean with a kernel from a stationary family which has a dependency on a few hyperparameters. The kernel hyperparameters are optimized while training by maximizing the log marginal likelihood by using an optimizer (gradient-based optimizers are typically used for efficiency). The log marginal likelihood may have different local optima, therefore, the optimization should be done repeatedly from different starting points. Here, we optimize kernel hyperparameters during training by Limited memory Broyden Fletcher Goldfarb Shanno (L-BFGS or LM-BFGS) optimizer [39,40].

Considered kernel: In this paper, we consider standard kernels including the radial basis function (RBF), the product of RBF and constant kernel, the sum of RBF and constant kernel, the sum of RBF and White kernel, and Matern kernel [23,41]. The considered kernel formulations and hyperparameter values are presented in Table 4. The RBF is the most popular kernel which enforces infinite smoothness on the function and thus is suitable if the data is very smooth. The kernels are closed under sum and product operations, i.e., two base kernels can be combined into a new kernel. Therefore, other convenient selections are the product of RBF and constant kernel to scale the RBF magnitude, and the sum of RBF and constant kernel to modify the mean value. Moreover, the sum of RBF and White kernel can be used to explain the noise component of the signal. The Matern kernel is a generalization of RBF which represents a very flexible and general family of distributions and does not assume quite as much smoothness as the RBF kernel. For more details, we refer to chapter 4 of [23], and for guidance on how to best combine different kernels, we refer to [41].

Noise level: The noise level in the labels can be specified by passing it through parameter α. A moderate noise level can also help deal with numeric issues while training since it is effectively implemented as Tikhonov regularization, i.e., by adding it to the diagonal of the kernel matrix. As an alternative for specifying the noise level, a White kernel component can be included in the kernel for predicting the global noise level from data. Here, we add $\alpha = 10^{-10}$ to the kernel matrix diagonal during training. It can prevent a potential numerical issue during training, by ensuring that the calculated values form a positive definite matrix, and also can be interpreted as the variance of measurement noise.

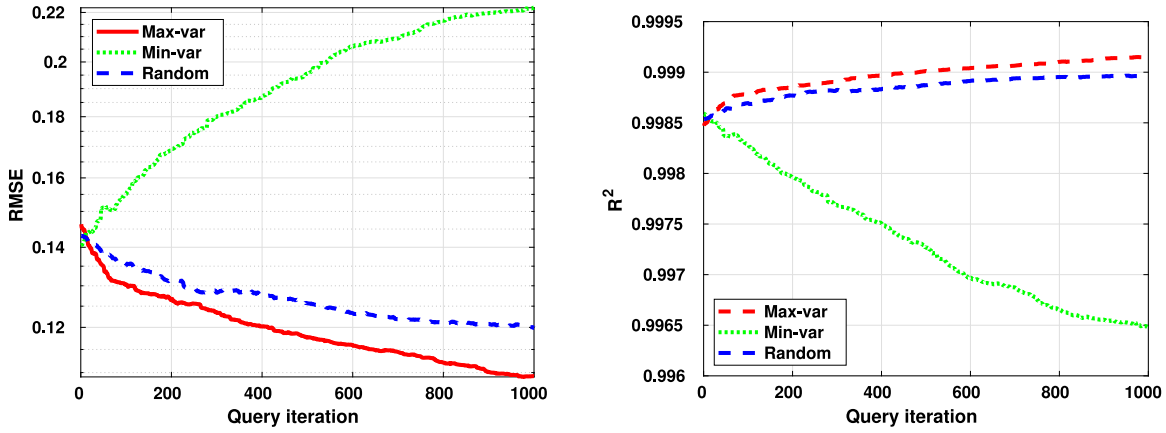
4.3. Performance analysis

We provide the performance analysis of the proposed AL-based OPM approach in terms of R^2 and root mean square error (RMSE) as defined in [28]. Figs. 2(a) and 2(b) respectively depict RMSE and R^2 values obtained by AL-based OPM approach versus query iteration, for Matern kernel, RBF kernel, the sum of RBF and White kernel, the sum of RBF and constant kernel, and product of RBF and constant kernel. Here, we start with 500 training samples and then increase the training dataset size by quarrying the next points from the pooling dataset by using the AL-based OPM approach. Matern kernel provides the best performance at all query iterations which shows the adaptability of this kernel to the distribution of our dataset. As explained in Section 4.2, RBF assumes a little too smoothness for data which might not be always true, however, the ν parameter in the Matern kernel controls the smoothness of the resulting function. The smaller ν the less smooth approximated function. Note that the Matern kernel becomes equivalent to RBF when $\nu \to \infty$. RBF performs



(a) (b)

Fig. 2. (a) RMSE and (b) R^2 values obtained by AL-based OPM approach versus query iteration, considering 500 training samples, for Matern kernel, RBF kernel, the sum of RBF and White kernel, the sum of RBF and constant kernel, and product of RBF and constant kernel.



(a) (b)

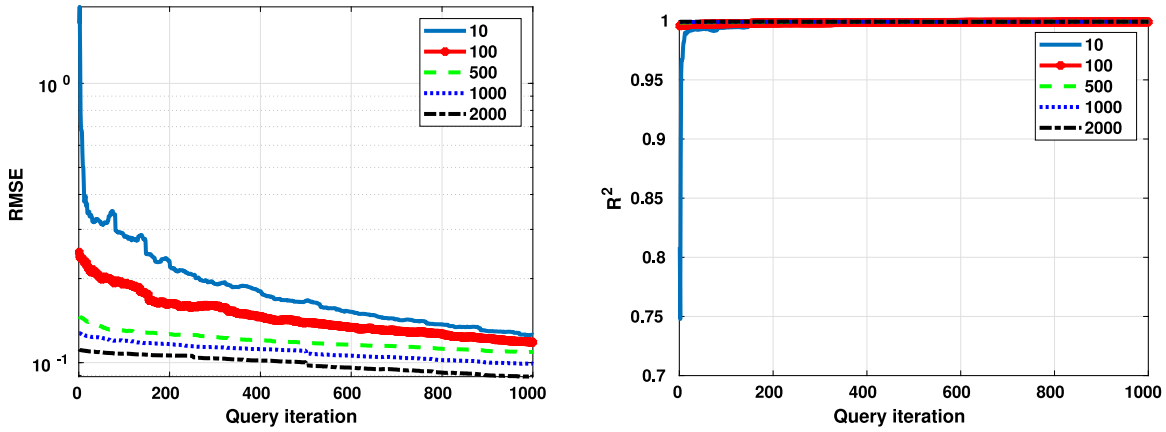
Fig. 3. (a) RMSE and (b) R^2 values versus query iteration, considering 500 training samples, for GPR with max-var, min-var, and random query strategies.

close to Matern with a little degradation. The RBF and sum of RBF with the White kernel have the same RMSE at low query iterations then a gap appears between them. There are two ways to specify the noise level for GPR, specifying α which just adds values to the diagonal as expected, or incorporating the noise level in the White kernel, in other words, specifying α is equivalent to adding a White kernel with $c_1 = \alpha$. Note that in RBF the $\alpha = 10^{-10}$, and we consider $c_1 = 1$ in the sum of RBF and white kernel. At low query iterations, the noise level (variance) is high, then by the proposed AL-based approach this level reduces where $c_1 = 1$ is no more a good choice. Even though the sum and product of RBF and constant kernel do not have good performance at the beginning they get close to the Matern kernel at high query iterations. Actually, in practice, the sum and product of RBF with the constant kernel can be used instead of shifting the mean and scaling the variance of the dataset. In conclusion, we use the Matern kernel in the following.

Defining overfitting as the situation where the training performance is better than testing performance: we do not observe any overfitting in any case as the train and test performance are almost the same. Moreover, the obtained train/test performance is quite well which means we neither encounter underfitting. This in turn indicates that the GPR hyperparameters are properly tuned.

Figs. 3(a) and 3(b) respectively describe the obtained RMSE and R^2 values versus query iteration, considering 500 training samples, for GPR with maximum variance (max-var), minimum variance (min-var), and random query strategies. The RMSE value decreases in max-var and random strategies and increases in the min-var case and vice-versa for R^2 value. Considering the max-var strategy, AL-based OPM, in an iterative process, minimizes the acquisition function and in turn minimizes the uncertainty (GPR variance) which leads to performance improvement. However, considering the min-var strategy, AL-based OPM, in an iterative process, minimizes the acquisition function which does not change the GPR variance since the data points with maximum variance are remaining in the pooling dataset. Considering the random selection, AL-based OPM, in an iterative process, updates the training dataset by randomly selecting points from the pooling dataset. Note that the variance (uncertainty) of the generated dataset is not so high, since we do not consider artifacts or additional noise sources. R^2 is the fraction of variance explained by the model with respect to the total variance of the test data, RMSE is also a measure of model accuracy which unlike R^2 is dependent on the scale of the label. In other words, $0 < R^2 < 1$ and $0 < RMSE < \infty$, thereby, the difference between different strategies is more in terms of RMSE rather than R^2. In conclusion, we use the max-var query approach in this paper.

Figs. 4(a) and 4(b) respectively demonstrate RMSE and R^2 values versus query iteration, for 10, 100, 500, 1000, and 2000 training dataset sizes. Here, we start with 10, 100, 500, 1000, and 2000 training dataset sizes, and then we increase the training



(a) (b)

Fig. 4. (a) RMSE and (b) R^2 values versus query iteration, for 10, 100, 500, 1000, and 2000 training dataset sizes.

dataset size by quarrying the next points from the pooling dataset by the AL-based OPM approach. The performance improvement by increasing training dataset size is more observable in small query iterations rather than high query iterations, in terms of both RMSE and R^2. Although we can see the effect of training dataset size even at high query iterations, this difference is negligible and the obtained results show the AL-based OPM approach works properly even at low training dataset size. The RMSE difference between 10 and 100, and between 100 and 500 is 0.01 dB and 0.02 dB, respectively, and the RMSE difference of 10 with respect to 1000 and 2000 is 0.03 dB and 0.04 dB, respectively. The same as Figs. 2 and 3, the difference between different cases is more obvious in RMSE than R^2 metric. Both RMSE and R^2 quantify how well a regression model fits a dataset, and it is useful to calculate both the RMSE and R^2 for a given model because each metric gives us different useful information. The first one tells us how well a regression model can predict the value of the response variable in absolute terms while the second one tells us how well a model can predict the value of the response variable in percentage terms. For instance, considering training datasets larger than 100, the RMSE tells us that the average deviation between the predicted GSNR made by the model and the true GSNR is 0.1 dB, and the R^2 value tells us that the predictor variables in the model can explain 99% of the variation in the GSNRs. Moreover, for training datasets larger than 100 the R^2 result looks the same which means that with 100 point the predictor variables in the model can explain 99% of the variation in the GSNRs. Considering obtained RMSE and R^2 values, we conclude that a training dataset with 100 points is enough. We report 1000 query iterations meaning that at the end, additional 1000 training data points have been added to the dataset. Regarding this, the performance of training data sets 500, 1000, and 2000 with 0 query iterations are respectively equal to the performance of training datasets 100 with 350 queries, 500 with 120 queries, and 1000 with 350 queries. In that sense, the total required training dataset is reduced, i.e. the total amount of generated data in AL is smaller than the dataset required for classical ML at the same prediction performance.

4.4. Complexity analysis

Table 5 shows the complexity analysis of the AL-based OPM method, SSFM, and EGN model. The proposed AL-based OPM method deploys the GPR algorithm at each iteration. As explained by [42], the complexity of GPR is mainly the calculation complexity of finding the kernel inverse defined by N_{train}^3 multiplications/summations where N_{train} is the number of training samples. As explained by [6,11], the SSFM simulation processing blocks include MDM and WDM multiplexer and de-multiplexer, EDFA at each span where each span has N_{step} steps, FMF linearity and nonlinearity implementation at each step, and dispersion compensation at receiver. The transmitted signal at each WDM comb has $2D$ dimensions with N_{sym} symbols at each dimension, therefore, the fast Fourier transform (FFT)/inverse FFT for dispersion implementation/compensation has $2DN_{sym}log(2DN_{sym})$ multiplications/summations. The $exp(x) = \sum_{i=0}^{n_1} x^i/i!$ can be calculated by $2n_1$ multiplications and n_1 summations, and the larger integer n_1 the better accuracy. The EGN model complexity analysis is based on equation (25) of [11]. The EGN model has four 3-dimensional (3D) integration, to calculate them numerically, n_2, n_3 and n_4 points with identical distances should be considered for the first, second, and third dimensions. Therefore, $n_2 \times n_3 \times n_4$ 3D areas appear, taking a 3D summation over these areas is equivalent to the principal 3D integration. The larger n_2, n_3, and n_4 the better accuracy. In sum, the complexity orders of the AL-based OPM method and EGN model are respectively $O(N_{train}^3)$ and $O(n_2n_3n_4)$. Considering the fact that $n_2n_3n_4 \gg N_{train}^3$, the AL-based OPM method provides much less complexity than the EGN model.

4.5. Discussion

To give an intuition on the practical relevance of the obtained results, we mention that the absolute error values ($|GSNR_{pred} - GSNR_{true}|$) are below 0.1 dB, 0.25 dB, and 0.75 dB in 58, 95, and 100% of cases, respectively. The $GSNR_{pred}$ values are obtained using small training and pooling datasets while $GSNR_{true}$ values are based on the EGN model which is an accurate analytical model. The proposed AL-based OPM approach and EGN model predict each GSNR value in 10^{-6} s and 10^3 s. Therefore, the proposed AL-based OPM speed-up the procedure 10^9 times and is proper for real-time GSNR estimation. Although we need to run the EGN model at each query to obtain the next query point, this does not change anything as here we refer to real-time GSNR estimation.

5. Conclusions

In this paper, we have presented a specific ML-based OPM approach relying on AL for GSNR estimation in FMF-based transmission. Results indicate that the proposed AL-based OPM method

Table 5
Computational complexity of AL-based OPM method, SSFM, and EGN model.

Methods	Number of multiplications	Number of summations
AL-based OPM	N_{train}^3	N_{train}^3
SSFM simulation	$(N_{step}N_s)(8DN_{sym}log(DN_{sym})+ 6DN_{sym}+2n_1+1)+ (4DN_{sym}log(DN_{sym})+4DN_{sym}+2n_1)$	$(N_{step}N_s)(8DN_{sym}log(DN_{sym})+ 2DN_{sym}+n_1+2)+ (2DN_{sym}log(DN_{sym})+4DN_{sym}+n_1)$
Integral-form EGN model	$8Dn_2n_3n_4(2n_1+4)$	$8Dn_2n_3n_4(n_1+6)$

is well designed to work with small training dataset, i.e. starting from a limited training dataset achieves capability similar to that obtained by working on large training datasets. The proposed AL-based OPM approach can achieve an RMSE value of 0.1 dB for GSNR estimation utilizing a few pooling dataset points without any training dataset.

CRediT authorship contribution statement

M.A. Amirabadi: Conception and design of study, Writing – original draft, Writing – review & editing. **M.H. Kahaei:** Conception and design of study, Writing – original draft, Writing – review & editing. **S.A. Nezamalhosseini:** Conception and design of study, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgment

All authors approved the version of the manuscript to be published.

References

[1] M.A. Amirabadi, M.H. Kahaei, S.A. Nezamalhosseini, F. Arpanaei, A. Carena, Closed-form EGN model for FMF systems, in: Asia Communications and Photonics Conference, 2021, pp. T4A–33.
[2] G. Rademacher, S. Warm, K. Petermann, Nonlinear interaction in differential mode delay managed mode-division multiplexed transmission systems, Opt. Express 23 (1) (2015) 55–60.
[3] G. Rademacher, R.S. Luís, B.J. Puttnam, R. Maruyama, K. Aikawa, Y. Awaji, H. Furukawa, K. Petermann, N. Wada, Investigation of intermodal nonlinear signal distortions in few-mode fiber transmission, J. Lightwave Technol. 37 (4) (2019) 1273–1279.
[4] G. Rademacher, R.S. Luis, B.J. Puttnam, H. Furukawa, R. Maruyama, K. Aikawa, Y. Awaji, N. Wada, Investigation of intermodal four-wave mixing for nonlinear signal processing in few-mode fibers, IEEE Photonics Technol. Lett. 30 (17) (2018) 1527–1530.
[5] G. Rademacher, F. Schmidt, K. Petermann, Optimum capacity utilization in space-division multiplexed transmission systems with multimode fibers, in: ECOC 2016; 42nd European Conference on Optical Communication, 2016, pp. 1–3.
[6] M.A. Amirabadi, M.H. Kahaei, S.A. Nezamalhosseini, L.R. Chen, Optimal power allocation in nonlinear MDM-wdm systems using Gaussian noise model, IET Optoelectron. (2022).
[7] Z. Dong, A.P.T. Lau, C. Lu, OSNR monitoring for QPSK and 16-QAM systems in presence of fiber nonlinearities for digital coherent receivers, Opt. Express 20 (17) (2012) 19520–19534.
[8] F.N. Khan, T.S.R. Shen, Y. Zhou, A.P.T. Lau, C. Lu, Optical performance monitoring using artificial neural networks trained with empirical moments of asynchronously sampled signal amplitudes, IEEE Photonics Technol. Lett. 24 (12) (2012) 982–984.
[9] G. Rademacher, K. Petermann, Nonlinear Gaussian noise model for multimode fibers with space-division multiplexing, J. Lightwave Technol. 34 (9) (2016) 2280–2287.
[10] P. Poggiolini, The GN model of non-linear propagation in uncompensated coherent optical systems, J. Lightwave Technol. 30 (24) (2012) 3857–3879.
[11] M.A. Amirabadi, M.H. Kahaei, S.A. Nezamalhosseini, L.R. Chen, Joint power and gain allocation in MDM-WDM optical communication networks based on extended Gaussian noise model, IEEE Access (2022).
[12] P. Poggiolini, Y. Jiang, A. Carena, F. Forghieri, A simple and accurate closed-form EGN model formula, 2015, arXiv preprint arXiv:1503.04132.
[13] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, F. Forghieri, The EGN model of non-linear fiber propagation, Opt. Express 22 (13) (2014) 16335–16362.
[14] P.J. Freire, D. Abode, J.E. Prilepsky, N. Costa, B. Spinnler, A. Napoli, S.K. Turitsyn, Transfer learning for neural networks-based equalizers in coherent optical systems, J. Lightwave Technol. 39 (21) (2021) 6733–6745.
[15] Y. Xu, D. Wang, M. Zhang, X. Zhou, Z. Zhang, J. Li, Y. Zhu, P. Xie, N. Paerhati, Deep transfer learning based multi-impairment diagnosis for PAM-4 optical communication systems, in: 2019 18th International Conference on Optical Communications and Networks, ICOCN, 2019, pp. 1–3.
[16] Y. Cheng, W. Zhang, S. Fu, M. Tang, D. Liu, Transfer learning simplified multi-task deep neural network for PDM-64QAM optical performance monitoring, Opt. Express 28 (5) (2020) 7607–7617.
[17] X. Zhu, B. Liu, X. Zhu, J. Ren, R. Ullah, Y. Mao, X. Wu, M. Li, S. Chen, Y. Bai, Transfer learning assisted convolutional neural networks for modulation format recognition in few-mode fibers, Opt. Express 29 (22) (2021) 36953–36963.
[18] D. Azzimonti, C. Rottondi, M. Tornatore, Reducing probes for quality of transmission estimation in optical networks with active learning, J. Opt. Commun. Netw. 12 (1) (2020) A38–A48.
[19] D. Azzimonti, C. Rottondi, A. Giusti, M. Tornatore, A. Bianco, Active vs transfer learning approaches for QoT estimation with small training datasets, in: Optical Fiber Communication Conference, 2020, pp. M4E–1.
[20] C. Rottondi, R. di Marino, M. Nava, A. Giusti, A. Bianco, On the benefits of domain adaptation techniques for quality of transmission estimation in optical networks, J. Opt. Commun. Netw. 13 (1) (2021) A34–A43.
[21] P. Serena, C. Lasagni, A. Bononi, The enhanced Gaussian noise model extended to polarization-dependent loss, J. Lightwave Technol. 38 (20) (2020) 5685–5694.
[22] M. Ibrahimi, H. Abdollahi, C. Rottondi, A. Giusti, A. Ferrari, V. Curri, M. Tornatore, Machine learning regression for QoT estimation of unestablished lightpaths, J. Opt. Commun. Netw. 13 (4) (2021) B92–B101.
[23] M. Seeger, Gaussian processes for machine learning, Int. J. Neural Syst. 14 (02) (2004) 69–106.
[24] A. Kapoor, K. Grauman, R. Urtasun, T. Darrell, Active learning with gaussian processes for object categorization, in: 2007 IEEE 11th International Conference on Computer Vision, 2007, pp. 1–8.
[25] E. Pasolli, F. Melgani, Gaussian process regression within an active learning scheme, in: 2011 IEEE International Geoscience and Remote Sensing Symposium, 2011, pp. 3574–3577.
[26] M.A. Amirabadi, M.H. Kahaei, S.A. Nezamalhosseini, Novel suboptimal approaches for hyperparameter tuning of deep neural network [under the shelf of optical communication], Phys. Commun. 41 (2020) 101057.
[27] M.A. Amirabadi, M.H. Kahaei, S.A. Nezamalhosseini, V.T. Vakili, Deep learning for channel estimation in FSO communication system, Opt. Commun. 459 (2020) 124989.
[28] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, Scikit-learn: ML in python, J. ML Res. 12 (2011) 2825–2830.
[29] M.A. Amirabadi, M.H. Kahaei, S.A. Nezamalhosseini, A. Carena, Deep learning regression vs. Classification for QoT estimation in SMF and FMF links, in: ICOP, 2022.

[30] S. Mumtaz, R.J. Essiambre, G.P. Agrawal, Nonlinear propagation in multimode and multicore fibers: Generalization of the Manakov equations, J. Lightwave Technol. 31 (3) (2012) 398–406.
[31] F.M. Ferreira, C.S. Costa, S. Sygletos, A.D. Ellis, Overcoming degradation in spatial multiplexing systems with stochastic nonlinear impairments, Sci. Rep. 8 (1) (2018) 1–10.
[32] Y. Jung, E.L. Lim, Q. Kang, T.C. May-Smith, N.H.L. Wong, R. Standish, F. Poletti, J.K. Sahu, S.U. Alam, D.J. Richardson, Cladding pumped few-mode EDFA for mode division multiplexed transmission, Opt. Express 22 (23) (2014) 29008–29013.
[33] V.A.J.M. Sleiffer, Y. Jung, V. Veljanovski, R.G.H. Van Uden, M. Kuschnerov, H. Chen, B. Inan, L.G. Nielsen, Y. Sun, D.J. Richardson, S.U. Alam, 73.7 Tb/s (96 x 3 x 256-Gb/s) mode-division-multiplexed DP-16QAM transmission with inline MM-EDFA, Opt. Express 20 (26) (2012) B428–B438.
[34] Y. Jung, V.A.J.M. Sleiffer, B. Inan, M. Kuschnerov, V. Veljanovski, B. Corbett, R. Winfield, Q. Kang, A. Dhar, J. Sahu, F. Poletti, Multimode EDFA performance in mode-division multiplexed transmission systems, in: Optical Fiber Communication Conference, 2013, pp. JW2A–24.
[35] Q. Kang, E. Lim, Y. Jun, X. Jin, F.P. Payne, S. Alam, D.J. Richardson, Gain equalization of a six-mode-group ring core multimode EDFA, in: 2014 the European Conference on Optical Communication, ECOC, 2014, pp. 1–3.
[36] R. Hashemi, H. Beyranvand, H. Rabbani, Joint channel power and amplifier gain optimization in coherent DWDM systems, Opt. Commun. 475 (2020) 126212.
[37] B. Mukherjee, I. Tomkos, M. Tornatore, P. Winzer, Y. Zhao, Springer HandBook of Optical Networks, Springer Nature, 2020.
[38] E. Desurvire, Erbium-Doped Fiber Amplifiers: Device and System Developments, Wiley-Interscience, 2002.
[39] R.H. Byrd, P. Lu, J. Nocedal, C. Zhu, A limited memory algorithm for bound constrained optimization, SIAM J. Sci. Comput. 16 (5) (1995) 1190–1208.
[40] C. Zhu, R.H. Byrd, P. Lu, J. Nocedal, Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization, ACM Trans. Math. Softw. 23 (4) (1997) 550–560.
[41] D. Duvenaud, The kernel cookbook: Advice on covariance functions, 2014.
[42] M. Belyaev, E. Burnaev, Y. Kapushev, Exact inference for Gaussian process regression in case of big data with the cartesian product structure, 2014, arXiv preprint arXiv:1403.6573.

Mohammad Ali Amirabadi was born in Zahedan, Iran, in 1993. He received the B.Sc. degree in Optics & Laser Engineering from Malek-e-Ashtar University of Technology, Isfahan, Iran, in 2015, and the M.Sc. degree in Communication Engineering from Iran University of Science and Technology, Tehran, Iran in 2017. Now he is studying Ph.D. in Communication Engineering in Iran University of Science and Technology, Tehran, Iran. His research interests include Multimode Fiber Optic Communication, Free Space Optical Communication, and Deep Learning.

Mohammad Hossein Kahaei received the B.Sc. degree in electrical engineering from Isfahan University of Technology, Isfahan, Iran, in 1986, the M.Sc. degree in adaptive signal processing from the University of the Ryukyus, Okinawa, Japan, in 1994, and the Ph.D. degree in signal processing from Queensland University of Technology, Brisbane, Australia, in 1998. Since 1999, he has been with the School of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran, where he is currently an Associate Professor and the Head of Signal and System Modeling laboratory. His research interests include array signal processing with primary emphasis on compressed sensing, sparse optimization problems, data science, localization, tracking, DOA estimation, blind source separation, and wireless sensor networks.

S. Alireza Nezamalhosseini received the B.Sc. degree in electrical engineering from Amirkabir University of Technology, Tehran, Iran, in 2006, and the M.Sc. and Ph.D. degrees in electrical engineering from Sharif University of Technology (SUT), Tehran, Iran, in 2008 and 2013, respectively. He is currently an assistant professor at Iran University of Science and Technology (IUST), Tehran. His research interests include underwater wireless optical communications, mode-division multiplexing in optical fibers, and visible light communications.