Data-Driven Erbium-doped Fiber Amplifier Gain Modeling Using Gaussian Process Regression

Harvey, Calum; Faruk, Md Saifuddin; Savory, Seb J.

IEEE Photonics Technology Letters

DOI:
10.1109/LPT.2024.3441110

E-pub ahead of print: 15/09/2024

Peer reviewed version

Data-Driven Erbium-doped Fiber Amplifier Gain Modeling Using Gaussian Process Regression

Calum Harvey, *Graduate Student Member, IEEE,* Md Saifuddin Faruk, *Senior Member, IEEE,* and Seb J. Savory, *Fellow, IEEE*

***Abstract*—We propose a data-driven erbium-doped fiber amplifier (EDFA) gain model utilizing Gaussian process regression (GPR). An additive Laplacian and radial-basis function kernel is proposed for the GPR and was found to outperform deep neural network (DNN) methods while additionally providing prediction uncertainty. Performance is measured using mean absolute error (MAE) averaged across five different EDFAs with three manufacturers. The GPR achieves an MAE of 0.1 dB using 30 training samples in contrast to the DNN that achieves an MAE of 0.25 dB using 3000 training samples. Additionally, we demonstrate that active learning can be used to improve robustness and repeatability of convergence.**

***Index Terms*—EDFA modeling, Gaussian process regression, active learning.**

I. Introduction

WITH the advancements in coherent transceivers and the subsequent improvements in metrology available from optical links, there has been an increase in work discussing how this new data can be used for measurement-informed modeling and, potentially, digital twins [1]. Another key use of this monitoring data is the introduction of software-defined networks (SDN) which is an approach to network management that enables dynamic reprogrammability for optimization of metrics such as margin allocation. Optimal margin allocation needs accurate quality of transmission (QoT) estimation tools which in turn need highly accurate amplifier noise figure and gain models.

Erbium-doped fiber amplifier (EDFA) gain modeling has been explored using numerous different approaches range from fully analytical to black box neural network methods. The original equations proposed in [2] provide an accurate estimate of gain, however, they require the use of internal parameters that in practice are not easily accessible. To address this, simplified analytical approaches such as the centre of mass example shown in [3] have been proposed, however, at the cost of decreased accuracy which significantly reduces their usability.

Data-driven approaches can fill this gap by learning the mapping between input and output powers without the need for explicit knowledge of the internal parameters but with more accuracy than the simplified analytical models. Typical approaches involving the use of artificial neural networks have been widely explored for gain modeling [4]–[8] through varying approaches such as training a model for each channel [7] or incorporating the simplified analytical model as input parameters [8]. These approaches, although provide good results, typically require a very large quantity of training data which can be costly to obtain for EDFA modeling. Additionally, they provide a point estimate, meaning the confidence of the prediction is unknown.

A solution to both these issues is to utilize probabilistic methods, more specifically Gaussian process regression (GPR) [9] which uses the properties of the Gaussian distribution to provide uncertainty estimations. GPR has already been shown to have significant benefits within optical communication [10], however, has yet to be applied to EDFA modeling, an area where measurement collection is particularly costly.

In this paper, we propose a data-driven EDFA model utilizing GPR to infer gain from any spectrum of channel dependent input powers. Our GPR uses an additive radial-basis function and Laplacian kernel to learn the relationship between input and output. The resulting approach is shown to only require 30 training spectra to converge the optimal mean absolute error (MAE) when tested on five distinct EDFA. Additionally, active learning is shown to increase monotonicity of convergence, therefore improving robustness, without reducing accuracy.

II. Proposed Data-Driven Model of an EDFA

Our data-driven model contains numerous elements functioning together with the goal of providing an setup with the ability to accommodate retraining. Our data collection setup is shown in Fig 1 which consists of a laser source, wavelength selective switch (WSS), and EDFA being monitored.

Measurement collection is controlled by a central controller that interfaces with the individual components. Reconfigurations are primarily made in the WSS with channel dependent input and output powers being measured using power monitors and written to a database.

A. Gaussian Processes

We utilize the GPR approach to infer the gain spectrum of an EDFA given the channel dependent input power spectrum while providing associated uncertainties relating to predictions. GPR is a non-parametric probabilistic modeling method that outputs Gaussian probability distributions instead of scalar values, allowing for quantifiable uncertainty [9]. Typically,

Calum Harvey and Seb J. Savory are with the Electrical Engineering Division, Department of Engineering, University of Cambridge, CB3 0FA Cambridge, U.K. (e-mail: ch982@cam.ac.uk; sjs1001@cam.ac.uk) and Md Saifuddin Faruk is with the School of Computer Science and Engineering, Bangor University, LL57 1UT Bangor, U.K. (e-mail: m.faruk@bangor.ac.uk).

Manuscript received April 19, 2021; revised August 16, 2021. This work was supported by the EPSRC Programme Grant TRANSNET [EP/R035342/1] and an EPSRC / Microsoft ICASE Studentship.

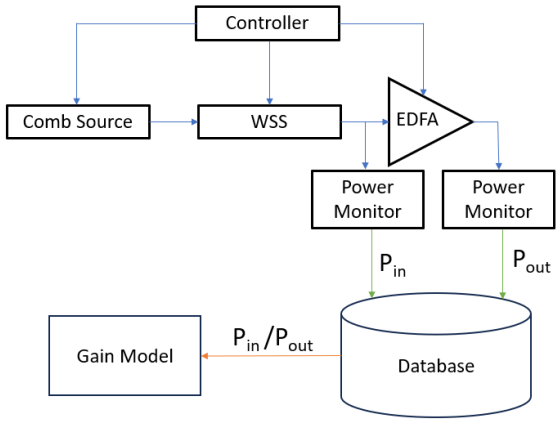

Fig. 1. Data retrieval and processing setup

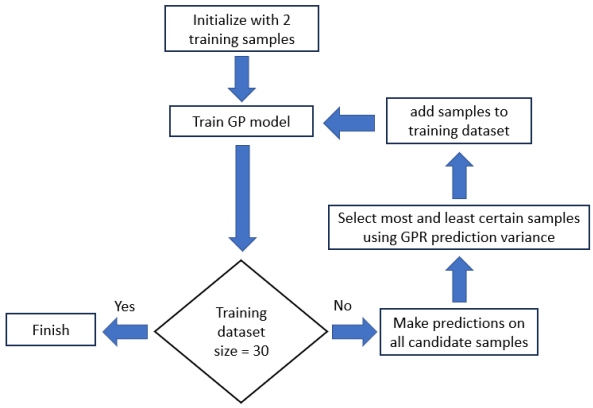

Fig. 2. Pool-based active learning approach using Gaussian processes to determine confidence of model predictions

GPR models are single output therefore an individual model is trained for each channel.

Additionally, GPR is a kernel method and therefore allows for the definition of a kernel function which allows us to embed our knowledge on how we think the individual data points should relate to each other.

There are two high level problems that need to be addressed by the kernel. Firstly, since a channel can either be 'on' or 'off', there are two distinct states that the kernel needs to be able to capture. Secondly, we know that input powers will not have uniform influence over the gain of a single channel. Therefore, we need to ensure that each inputs influence can be adjusted.

Our kernel, shown in Eq. 1, addresses the first problem by being an additive combination of two functions. The first component is a radial-basis function (RBF) which performs well for our main cluster of 'on' channels and a Laplacian function described by [11] which is robust to outliers [9] and therefore identifies 'off' channels more effectively. By making them additive, the value of the overall kernel will be large if either of the constituent parts is large. Allowing us to ensure similarity to correctly quantified in either of the two states.

The second can be resolved by ensuring the length scale is dimension dependent instead of a single scalar value. By making the length scale a d dimensional vector, the kernel can train the values to give each input dimension a different impact over the output of the model.

$$k(\Delta x) = s^2 \exp\left(\left(\frac{-\|\Delta x\|}{2l_1}\right)^2\right) + \exp\left(\frac{-\|\Delta x\|}{l_2}\right) \quad (1)$$

where $\Delta x = x_i - x_j$ with x_i and x_j denoting input vectors being compared. There are three optimizable hyperparameters in the kernel l_1 and l_2 being length scale of the respective functions and s^2 being the scaling factor of the RBF kernel. Additionally, the length scale of the RBF kernel l_1 is made dimension dependent has discussed above whereas l_2 is left as a scalar as 'off' channels cannot benefit from the additional complexity.

These hyperparameters are then optimized using the marigi-nal log likelihood loss function using the Adam optimizer with a learning rate of 0.1 and 2000 epochs. All data is normalized using min-max scaling to ensure all values are between 0 and 1 before training is undertaken.

B. Active Learning

Pool-based active learning [12] is an approach used in scenarios where labeling training data is expensive but unlabeled data is readily available. An example usage would be in image recognition, where images are easily obtainable but labeling the image content requires human input and therefore has a cost associated. However, we utilize this approach for an alternative purpose of distilling a large dataset into a small subset without losing accuracy. The primary motivation for this is the $O(n^3)$ time complexity of training the GPR, making large training datasets incompatible with GPR. Contrastingly, it has significantly lower inference complexity, meaning there are few drawbacks to performing large numbers of predictions.

The process is illustrated in Fig. 2 where the aim of the active learning process is to intelligently create the most data efficient training dataset possible where every training sample is adding some empirical value to the accuracy of the model. Our heuristical approach involves utilizing the uncertainty produced by the trained GPR model when predictions are made, adding the most and least certain samples from our large pool of 3000 unlabeled data points during each iteration.

Firstly, an initialization process needs to be developed to allow an initial model to be training, with which to complete the active learning. To ensure maximum usage of the active learning process, we initialize with a single random sample within our training dataset that is used to train a model.

After initialization, the active learning process follows the following procedure. Each iteration the highest and lowest variance samples are selected, labeled, and added into the training dataset until the training dataset reaches a size of 100 samples. Each time new samples are added, the model is retrained to incorporate the newly added training data.

It was found that convergence to a low error occurred significantly before 100 samples, however, to ensure convergence under any circumstance 100 was chosen as the termination point.

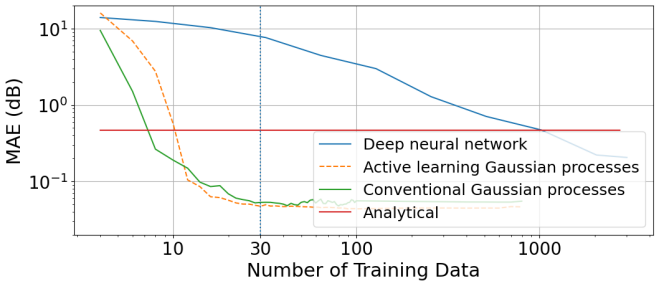

Fig. 3. Comparison of mean absolute error of EDFA gain modeling methods simplified analytical, deep neural network, naive Gaussian processes, and active learning Gaussian processes

III. Results and Discussion

Our experiment utilizes measurements from two Manlight EDFA, one FSCOM EDFA, and the published dataset found at [13] which provides ROADM measurements for 8 preamp and 8 booster Lumentum EDFA at 15/18/21 dB fixed gain values. By showing good performance over a range of manufacturers, the generalizability of the modeling approach can be easily shown. The experimental setup seen in Fig 1 was used for the two Manlight and FSCOM EDFA whereas the published dataset was written to the database manually to ensure the retreival of data for the modeling was identical for each EDFA independent of where the data came from.

In our measurements, each channel has a 33% chance of being off for any given partial-load configuration, additionally, channels that are 'on' are set to a random power of between -22 and -18 dBm. The EDFA being modeled is set to a fixed gain of 25 dB and a tilt of 0 dBm. A new configuration is periodically generated and applied to the WSS through the controller. Once a change is detected in the channel configuration, measurements are taken pre and post amplification and written to the database. The model then utilizes the database to access the data for training the model.

For the published data, we use the 15 dB gain measurements, again with 0 dB tilt, to validate performance at high and low gain values.

In Fig. 3, we compare our GPR approach with and without active learning to the simplified analytical and a deep neural network approach similar to the architecture proposed in [4]. By using a more generic architecture instead of a highly optimized approach as proposed in [13], we allow the flexibility to adapt based on the number of channels being used as it differs for our measurements. GPR without active learning consists of adding random samples instead of using the active learning approach.

Our GPR approaches significantly outperform the DNN and analytical methods, converging with only 30 training samples to an MAE of 0.05 dB and 0.04 dB for conventional GPR and active learning GPR respectively. Although this is to be expected due to the data efficiency of GPR, we found that even if the training data provided to the DNN is increased by two orders of magnitude to 3000 datapoints, the MAE is 0.3 dB which is still significantly higher than the GPR approaches.

Spectrum predictions are shown in Fig. 4 for the partial and fully loaded cases when trained on 30 samples. Each predicted

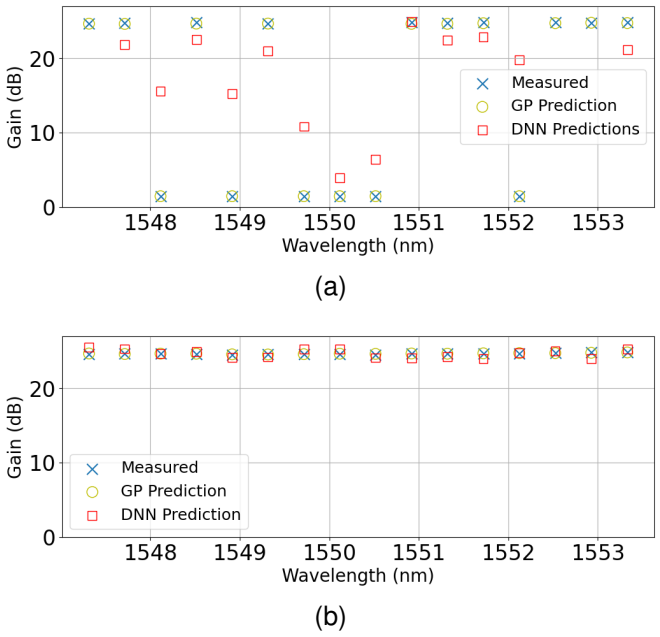

Fig. 4. Gain spectrum predictions for (a) partially-loaded and (b) fully-loaded configurations when trained on 30 samples

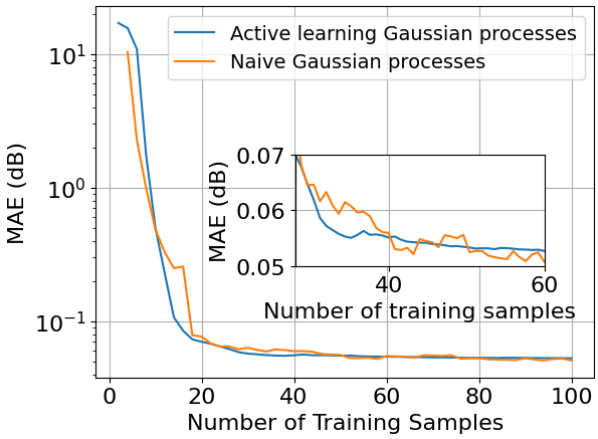

Fig. 5. Average performance of naive and active learning approaches at different training sample sizes over 5 runs. Inset shows increased variance of Naive compared to active learning.

point is the mean of prediction from the GPR, however, due to the variance being too small the associated uncertainty for each prediction is not visible.

The key benefit of the active learning approach over the typical GPR is the stability of its convergence that comes from strategically picking our training dataset. Fig.5 shows the average of the error over 5 unique runs of both naive and active learning based GPR. We see significantly more variance in error when samples are naively added compared to the active learning. By using the naive approach, it is not guaranteed that the model will decrease in error as the number of training samples increases, whereas, the active learning can be seen as monotonic and therefore provides greater guarantees of performance.

The results shown in Table I show that the GPR approaches have good performance on all EDFA that were tested, with

EDFA (Number of Channels)	DNN from [13] $M = 258 \times 10^9$	Simplified DNN $M = 66 \times 10^9$	GPR $M = 135 \times 10^6$	GPR w/AL $M = 2 \times 10^9$
Manlight 1 (16)	0.62 (-)	0.25 (-)	0.05 (0.26)	0.04 (0.26)
Manlight 2 (16)	0.77 (-)	0.42 (-)	0.08 (0.26)	0.08 (0.26)
FSCOM (16)	0.66 (-)	0.27 (-)	0.07 (0.26)	0.07 (0.26)
Lumentum 1 from [13] (95)	0.17 (-)	0.20 (-)	0.30 (0.58)	0.18 (0.55)
Lumentum 2 from [13] (95)	0.16 (-)	0.18 (-)	0.15 (1.09)	0.14 (0.98)
Average Error	0.34	0.25	0.13	0.10

TABLE I
MAE IN DECIBELS OF THE SIMPLIFIED DNN, DNN PROPOSED IN [13], GPR, AND GPR WITH ACTIVE LEARNING FOR FIVE DIFFERENT EDFA TRAINED WITH 30 SAMPLES FOR GPR AND 3000 FOR ACTIVE LEARNING AND DNNS. M IS THE TOTAL NUMBER OF MULTIPLICATIONS REQUIRED FOR TRAINING.

slightly higher error on the published Lumentum dataset. This could primarily be caused by the dataset being field fiber measurements and subsequently more unpredictable. The average error was found to be 0.1 dB and there was no degradation in performance by using active learning. Additionally, due to the usage of a probabilistic approach we can define a standard deviation associated with each prediction we make. This can therefore be used to quantify how certain we are of a prediction, however, for the DNN this is not available.

The DNN proposed in [13] is also evaluated and although it performs well with data that it was designed for. However, performance is significantly degraded on our measurements whereas the GPR manages to maintain a much lower average error.

Furthermore, computational complexity for each approach is calculated. For GPR, we use n^3 with active learning having an additional factor i denoting the number of iterations needed: $i(n^3)$. Our DNN is computed by $nt(ab + bc + cd + de + ef)$ where t is number of epochs and a,b,c,d,e,f are the number of nodes in each layer.

Table I shows the results of these calculations. Both DNN approaches are significantly more computationally complex than the simple GPR. Although it is expected for the simple GPR to be less complex, it is worth noting that even with the added complexity of the active learning the complexity is still less than the DNN.

IV. CONCLUSION

We have proposed and experimentally validated a data-driven EDFA gain modeling approach utilizing Gaussian processes on five separate EDFAs. We have reduced the number of training data required to 30 samples while still having an average MAE of 0.1 dB across all EDFA tested while providing prediction uncertainty. This is a significant improvement on the DNN approach which averages an MAE of 0.25 dB for 3000 training samples. Additionally, we show that active learning can be incorporated to improve monotonicity of error reduction in relation to number of training samples, improving the robustness of the model training process. Our results show that for EDFA gain modeling a Gaussian processes approach is superior to neural network methods due to less training samples needed and the inherent prediction uncerainty.

V. ACKNOWLEDGEMENTS

For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising. Data underlying the results presented in this paper are available at https://doi.org/10.17863/CAM.107402

REFERENCES

[1] M. S. Faruk and S. J. Savory, "Measurement informed models and digital twins for optical fiber communication systems," *Journal of Lightwave Technology*, vol. 42, no. 3, pp. 1016–1030, 2024.

[2] A. A. M. Saleh, R. M. Jopson, J. D. Evankow, and J. Aspell, "Modeling of gain in erbium-doped fiber amplifiers," *IEEE Photonics Technology Letters*, vol. 2, no. 10, pp. 714–717, 1990, publisher: Institute of Electrical and Electronics Engineers (IEEE).

[3] K. Ishii, J. Kurumida, and S. Namiki, "Experimental investigation of gain offset behavior of feedforward-controlled WDM AGC EDFA under various dynamic wavelength allocations," *IEEE Photonics Journal*, vol. 8, no. 1, pp. 1–13, 2016.

[4] Y. You, Z. Jiang, and C. Janz, "Machine learning-based EDFA gain model," in *2018 European Conference on Optical Communication (ECOC)*. IEEE, 2018.

[5] F. D. Ros, U. C. d. Moura, and M. P. Yankov, "Machine learning-based EDFA gain model generalizable to multiple physical devices," *2020 European Conference on Optical Communications (ECOC)*, 2020, _eprint: 2009.05326.

[6] A. Raj, Z. Wang, F. Slyne, T. Chen, D. Kilper, and M. Ruffini, "Self-normalizing neural network, enabling one shot transfer learning for modeling EDFA wavelength dependent gain." [Online]. Available: http://arxiv.org/abs/2308.02233

[7] S. Zhu, C. L. Gutterman, W. Mo, Y. Li, G. Zussman, and D. C. Kilper, "Machine learning based prediction of erbium-doped fiber WDM line amplifier gain spectra," in *2018 European Conference on Optical Communication (ECOC)*. IEEE, 2018.

[8] S. Zhu, C. Gutterman, A. D. Montiel, J. Yu, M. Ruffini, G. Zussman, and D. Kilper, "Hybrid machine learning EDFA model," in *Optical Fiber Communication Conference (OFC) 2020*. Optica Publishing Group, 2020, p. T4B.4, journal Abbreviation: Optical Fiber Communication Conference (OFC) 2020. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=OFC-2020-T4B.4

[9] C. E. Rasmussen, *Gaussian processes for machine learning*. MIT Press, 2006.

[10] J. W. Nevin, F. J. Vaquero-Caballero, D. J. Ives, and S. J. Savory, "Physics-informed gaussian process regression for optical fiber communication systems," *Journal of Lightwave Technology*, vol. 39, no. 21, pp. 6833–6844, 2021.

[11] A. Feragen, F. Lauze, and S. Hauberg, "Geodesic exponential kernels: When curvature and linearity conflict."

[12] B. Settles, "From theories to queries: Active learning in practice," in *Active Learning and Experimental Design workshop In conjunction with AISTATS 2010*, ser. Proceedings of Machine Learning Research, I. Guyon, G. Cawley, G. Dror, V. Lemaire, and A. Statnikov, Eds., vol. 16. Sardinia, Italy: PMLR, 16 May 2011, pp. 1–18. [Online]. Available: https://proceedings.mlr.press/v16/settles11a.html

[13] Z. Wang, D. C. Kilper, and T. Chen, "Open EDFA gain spectrum dataset and its applications in data-driven EDFA gain modeling," *Journal of Optical Communications and Networking*, vol. 15, no. 9, pp. 588–599, 2023, conference Name: Journal of Optical Communications and Networking. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10210309