



## A comparison between black-, gray- and white-box modeling for the bidirectional Raman amplifier optimization

Yankov, Metodi P.; Soltani, Mehran; Carena, Andrea; Zibar, Darko; Da Ros, Francesco

*Published in:*  
Optical Fiber Technology

*Link to article, DOI:*  
[10.1016/j.yofte.2024.104060](https://doi.org/10.1016/j.yofte.2024.104060)

*Publication date:*  
2025

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
Yankov, M. P., Soltani, M., Carena, A., Zibar, D., & Da Ros, F. (2025). A comparison between black-, gray- and white-box modeling for the bidirectional Raman amplifier optimization. *Optical Fiber Technology*, 89, Article 104060. <https://doi.org/10.1016/j.yofte.2024.104060>

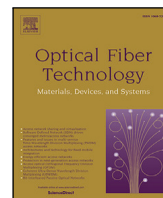
---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



## Regular Articles

# A comparison between black-, gray- and white-box modeling for the bidirectional Raman amplifier optimization

Metodi P. Yankov<sup>a</sup>, Mehran Soltani<sup>a</sup>, Andrea Carena<sup>b</sup>, Darko Zibar<sup>a</sup>, Francesco Da Ros<sup>a,\*</sup>

<sup>a</sup> Department of Electrical and Photonic Engineering, Technical University of Denmark, Ørstedsgade 343, Kgs. Lyngby, 2800, Denmark

<sup>b</sup> Dipartimento di Elettronica e Telecomunicazioni (DET), Politecnico di Torino, Corso Duca degli Abruzzi, 24, Torino, 10129, Italy

## ARTICLE INFO

## Keywords:

Raman amplifiers  
Machine learning  
Modeling

## ABSTRACT

Designing and optimizing optical amplifiers to maximize system performance is becoming increasingly important as optical communication systems strive to increase throughput. Offline optimization of optical amplifiers relies on models ranging from white-box models deeply rooted in physics to black-box data-driven and physics-agnostic models. Here, we compare the capabilities of white-, gray- and black-box models on the challenging test case of optimizing a bidirectional distributed Raman amplifier to achieve a target frequency-distance signal power profile. We show that any of the studied methods can achieve similar frequency and distance flatness of between 1 and 3.6 dB (depending on the definition of flatness) over the C-band in an 80-km span. Then, we discuss the models' applicability, advantages, and drawbacks based on the target application scenario, in particular in terms of flexibility, optimization speed, and access to training data.

## 1. Introduction

The behavior of optical amplifiers directly affects the signal performance in optical communication systems [1] and their impact grows as communication systems get extended to multi-band [1] and/or space-division multiplexing [2] operation. Therefore, modeling and more importantly optimizing the amplifier response is highly desirable, and considerable research effort has been dedicated to the topic [3–6]. Substantial earlier work on modeling optical amplifiers has been mainly focused on describing the physical processes leading to optical amplification [7–10], referred to as white-box models. With such models, offline optimization of the gain profile of the amplifiers can be carried out. This process is generally cumbersome due to the often complicated differential equations involved in predicting the amplifier behavior. Furthermore, the physical parameters involved, e.g. gain-medium properties such as emission spectra [8], gain coefficient spectra (Raman) [9, 10], etc., are not always easy to characterize. Model fitting is therefore applied [4, 11–13] but inaccuracies between the fitted model and the physical amplifier directly impair the optimization. Additionally, iterative optimization may suffer from a significant computational complexity if such white-box models are non-differentiable.

Alternatively, online optimization relying on gradient-free methods can be applied directly to the physical device. These mainly rely on the use of evolutionary algorithms, e.g. genetic algorithms [14–16]. While they do not require a model, the optimization of live systems is challenging and time-consuming due to the requirement of feedback from

the physical system, system downtime during the optimization process, and the related measurement uncertainties, which might mislead the optimization.

More recently, black-box models, based e.g. on neural networks (NNs), have been proposed [6, 17–24]. These models can be trained on either synthetic or experimental data and can directly implement inverse models – given the target amplifier response they provide the amplifier parameters – enabling nearly real-time reconfigurability once the models are trained. However, a relatively large training dataset is normally required, especially if generalizable models are desired, e.g. in terms of multiple Erbium-doped fiber amplifier (EDFA) units [25] or Raman amplifier relying on different fiber types [26].

Finally, the use of physics-informed NN models has gained increasing interest following their first introduction in [27]. These models combine the data-driven approach of black-box NN with the knowledge from the underlying physics, generally included through an ad-hoc loss function which includes terms from both the data-driven and the physics-informed sides. Such an approach has also seen preliminary applications to nonlinear propagation in optical fibers [28] more specifically to Raman amplifiers [29–31].

All of the above approaches, from the white-box physical modeling to the black-box NN-based models, have advantages and drawbacks depending on the target application [32]. Additionally, a systematic comparison of their performance on the same test case has yet to be reported.

\* Corresponding author.

E-mail address: [fdro@dtu.dk](mailto:fdro@dtu.dk) (F.D. Ros).

In this work, we compare four optimization methods spanning the categories just described. Their performance is studied on the particularly challenging test case of a bi-directional distributed Raman amplifier with first and second-order pumping where the wavelength division multiplexed (WDM) signal power evolution over frequency and fiber distance is jointly optimized, aiming for maximizing flatness over both dimensions. Then, the applicability of each method to different optimization scenarios is discussed. In Section 2, the considered optimizers are briefly described. The test scenario is defined in Section 3. The methods' performance and applicability are reported in Sections 4 and 5, respectively. Section 6 summarizes the main findings.

## 2. Raman optimizers

The signal and pump power evolution through a Raman amplifier can be numerically calculated by solving a system of nonlinear partial differential equations [33]:

$$\frac{\partial P_m(f, z)}{\partial z} = \pm 2\alpha_m(f)P_m(f, z) + \sum_{f_n \in \mathcal{F}} \frac{g_R(f_n - f_m)}{A_{eff}} P(f_m, z) \cdot P(f_n, z), \quad (1)$$

where the indexes  $m$  refers to the current carrier (pump or signal), the index  $n$  indexes the co- and counter-propagating carriers (positive and negative sign on the first right-hand side term) and  $\mathcal{F} = \mathcal{F}_p \cup \mathcal{F}_s$  is the set of active carriers including the set of pumps  $\mathcal{F}_p$  and the set of signals  $\mathcal{F}_s$ .  $P(f, z)$  is the carrier power at frequency  $f$  and distance  $z$ ,  $g_R(f_n - f_m)$  is the Raman gain coefficient for a given offset between the frequencies  $f_n$  and  $f_m$  (here, assumed independent of the reference wavelength and constant over the C-band),  $\alpha(f_n)$  is the fiber loss at frequency  $f_n$  and  $A_{eff}$  is the fiber effective area. This model does not allow for analytical optimization, as it lacks a closed-form analytical solution. Therefore different approaches have been proposed and are briefly described below.

**Optimizers based on NNs.** Neural networks can approximate complex nonlinear relations which may be challenging to express analytically with tractable models. That is indeed the case for the relation between a Raman amplifier response (signal 2D evolution in frequency and fiber distance) and its input parameters (pump powers and wavelengths). For the 2D evolution, we rely on the convolutional NN (CNN) proposed in [34]. The network takes in input the desired 2D signal evolution in frequency and distance and outputs the predicted set of pump powers and wavelengths expected to lead to such a target evolution, i.e. an inverse model. The CNN extracts the key features in the input space through a series of convolutional and pooling layers and makes the final prediction through a fully connected regression layer. Such an offline model can be trained using synthetic [34] or experimental [35] data. Once the model is trained, the inference is nearly instantaneous as it simply relies on a sequence of convolution operations, matrix multiplications, and nonlinear activation functions (ReLU in our specific implementation). As the model is trained on a specific dataset, the CNN is constrained to the configurations (number of pumps, number of signal channels, fiber properties, etc.) that have been used to generate such a dataset, and extrapolation beyond that typically leads to poor accuracy. In this work, a training dataset of 3500 samples is generated using the Raman solver provided by the GNPpy library [36]. The topology of the CNN is optimized to 3 convolutional layers with 32 filters each with a kernel size of  $3 \times 3$  in each layer followed by max-pooling, and finally two fully-connected regression layers with 40 hidden nodes. The spatial resolution for the network input is fixed to 500 m. More details can be found in [34].

**Gradient-free optimizers.** Gradient-free optimization algorithms can be applied to numerical and analytical models which are non-differentiable and more importantly to experimental setups where a gradient cannot be explicitly calculated. Among gradient-free optimizers, evolutionary algorithms, such as differential evolution (DE), rely on the evolution of a population of solutions that approach the desired target by iteratively penalizing the individuals of the population that

achieve high values of the cost function. Here, we apply a DE algorithm to optimize the pump parameters providing target 2D power profile which are calculated by the numerical Raman solver. The algorithm uses the optimized parameters of [33] for our specific test scenario: a crossover probability, mutation factor, and population size of 0.5, 0.8, and 30, respectively.

**Neural network + gradient-free optimization.** As gradient-free optimizers operate iteratively, their convergence (if any) can be highly dependent on the initial conditions provided to the optimizer. In order to speed up the convergence and the quality of the final solution, an approximate initial condition can be provided. In the proposal of [33], the initial population provided to the DE is constructed by choosing pump powers and wavelengths within a small range of the prediction provided by the CNN-based model. This choice focuses the DE to a promising subset of the overall search space, thus significantly reducing the number of iterations needed for convergence but at the cost of requiring a pre-trained CNN model and assuming that the CNN prediction is of a sufficiently high accuracy. However, DE can then extrapolate beyond the CNN training data set without the accuracy penalty affecting the CNN. A similar approach has been applied to the more recent work of [37], where the gradient-free optimizer can be used to correct for modeling inaccuracies or variations over time. This combination of a white-box model (DE) with a black-box (NN)-based initialization can be considered a gray-box model, similar to physics-inspired NN models [27–30].

**Gradient-based optimizers.** The signal and pump power evolution through a Raman amplifier can be numerically calculated by using finite difference methods to solve (1), i.e. in a white-box approach. Such a system of equations can be mostly expressed in a frequency-differentiable form with the key exception of the Raman gain efficiency which is normally either calculated by numerically computing an integral or defined by a look-up table based on experimental measurements. Neither option is differentiable. In order to overcome this limitation, in [3,38] it was proposed to fit a differentiable function to the Raman gain efficiency. If a sufficiently accurate fit of the Raman gain is achieved (e.g. with an overfitted neural network [3]), the accuracy of the solver solution is not compromised and the full system of equations becomes fully differentiable. Gradient descent (GD)-based optimization can therefore be applied allowing differentiability over pump, signal, and fiber parameters. Remark that the use of a GD-based optimizer for backward or bidirectional Raman amplifiers does not require the use of time-consuming iterative algorithms (e.g. shooting algorithms) that are normally required to tackle the boundary-value problem. GD-based optimizers simply fine-tune the powers at the input of the fiber for all pumps. The input pump powers, then, uniquely determine the powers at the output of the fiber for the counter-propagating pumps [38]. Therefore, the ability to use GD-based optimization over the system of (1) can provide a powerful alternative to physics-inspired NN models for specific applications.

## 3. Setup

The considered setup for the distributed Raman amplifier is shown in Fig. 1 and is based on  $L = 80$  km of a standard single-mode fiber (SSMF) with 0.2 dB/km, 0.25 dB/km and 0.32 dB/km of attenuation for 15xx-nm signal and 14xx-nm and 13xx-nm pumps, respectively. Effective area, nonlinear coefficient and peak Raman gain coefficient are set to  $80 \mu\text{m}^2$ ,  $1.26 \text{ 1/W/km}$ ,  $0.3841 \text{ 1/W/km}$ . Raman gain efficiency spectrum of an SSMF is assumed [36].

The optimization is carried out for 40 100-GHz spaced signal channels at a launch power of 0 dBm/channel spanning the C-band (192 THz to 196 THz). Bidirectional pumping is assumed with three first-order pumps and one second-order pump per propagation direction (eight pumps in total). The pump wavelengths and pump power constraints are defined in Table 1. The impact of noise is not considered in this work but can be included without significantly altering the methods

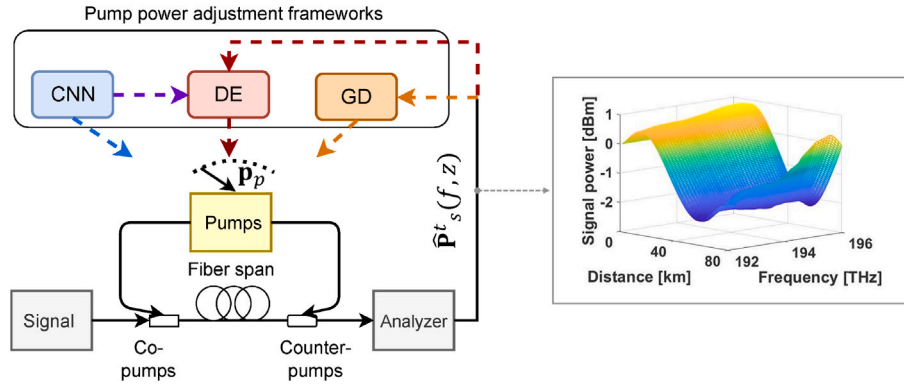


Fig. 1. The proposed distributed Raman amplifier setup with different pump power adjustment frameworks. The frameworks are discriminated with different colors (CNN: Blue, DE: Red, CNN+DE: Purple+Red, GD: Orange). The inset shows an example of a simulated 2D power profile for a given pump configuration and flat input power channel load.

Table 1  
Raman pump wavelengths and power ranges.

Co-pumps	$p_1$	$p_2$	$p_3$	$p_4$
$\lambda$ , nm	1366	1425	1455	1475
$[P_{LB}, P_{UB}]$ , mW	[200,1200]	[5,150]	[5,150]	[5,150]
$[P_{LB}, P_{UB}]$ , dBm	[23.0,30.8]	[7.0,21.8]	[7.0,21.8]	[7.0,21.8]
Counter-pumps	$p_5$	$p_6$	$p_7$	$p_8$
$\lambda$ [nm]	1366	1425	1455	1475
$[P_{LB}, P_{UB}]$ , mW	[200,1200]	[5,150]	[5,150]	[5,150]
$[P_{LB}, P_{UB}]$ , dBm	[23.0,30.8]	[7.0,21.8]	[7.0,21.8]	[7.0,21.8]

by simply increasing the dimensionality of the inputs and outputs of the models to account for the noise at each channel.

The four methods compared are: (i) the stand-alone black-box model based on a CNN, (ii) the gradient-free gray-box model based on the DE, (iii) the combined CNN+DE method, where the CNN provides the initial conditions to the DE-based optimizer, and (iv) the white-box model based on a differentiable implementation of Eq. (1) optimized through GD. As mentioned, the CNN does not require feedback from the channel and provides a solution for the pump configuration ‘instantaneously’. The DE and GD algorithms are iterative optimization processes. At every iteration, the frequency-distance power profile  $\hat{P}_s^t(f, z)$  at the channel frequencies is calculated using (1). The new profile is used to update the pump configuration for the next iteration. Remark that in the case of DE, the calculated power profile can be replaced by an experimental measurement, e.g. through a frequency-swept optical time-domain reflectometer as in [35].

The chosen test case focuses on optimizing just the pump powers. However, all four methods also allow for optimization of the pump wavelengths, as well as fiber parameters or input signal power, at the cost of a more complex optimization for DE and GD-based methods and a larger dataset required for the CNN-based methods. By its inherent design, the GD-method applied to the differentiable implementation of (1) can effectively calculate derivative with respect to any signal, pump, and fiber parameter.

#### 4. Performance comparison

A common scenario in multi-span transmission systems is to aim for flatness of the gain spectrum over distance and frequency. Depending on the application, flatness across the two dimensions is not of equal importance. It is therefore difficult to design a universal metric to benchmark the system and optimize the performance. Similarly, without such a metric, it is difficult to design a cost function for optimization. In the following, three heuristically chosen criteria and optimization strategies are proposed, aiming to cover a wide variety of applications. It is not the purpose of this paper to propose a universal

strategy but to compare different optimization methods for the same target.

Without loss of generality to the target frequency-distance evolution definition, performance is estimated based on three main criteria, following [33]. All criteria relate to the frequency-distance excursions of the signal carriers ( $f \in \mathcal{F}_s$ ) in dB scale. The first criterion is the maximum power excursion defined as

$$J_0(\mathbf{p}_p) = \max_{f \in \mathcal{F}_s, z} (P(f, z | \mathbf{p}_p)) - \min_{f \in \mathcal{F}_s, z} (P(f, z | \mathbf{p}_p)), \quad (2)$$

where  $P(f, z | \mathbf{p}_p)$  is the signal power at frequency  $f$  and distance  $z$  given the pump configuration  $\mathbf{p}_p$ . The criterion  $J_0$  is an ultimate goal and is the most difficult to achieve. Observe, that ultimate limits for  $J_0$  are not theoretically known, but are interesting areas for future research to explore. In particular, the possibility for  $J_0 = 0$  and its corresponding pump configuration would be of great theoretical and practical interest.

Second, the maximum power excursion over frequencies at any given distance is defined as

$$J_1(\mathbf{p}_p) = \max_z [\max_{f \in \mathcal{F}_s} (P(f, z | \mathbf{p}_p)) - \min_{f \in \mathcal{F}_s} (P(f, z | \mathbf{p}_p))]. \quad (3)$$

Minimizing  $J_1$  instead of  $J_0$  is typically easier under a given pump constraint, as it allows for the channel powers to vary along the distance dimension, as long as the WDM signal maintains a compact frequency spectrum.

Third, the maximum deviation from a loss-compensated transmission (0 dB net gain) over frequency

$$J_2(\mathbf{p}_p) = \max_{f \in \mathcal{F}_s} |P(f, L | \mathbf{p}_p) - P(f, 0 | \mathbf{p}_p)|. \quad (4)$$

Minimizing  $J_2$  is the easiest target and is the most practically relevant criterion, as it is directly related to the system requirement for additional frequency flattening filters and/or additional gain, e.g. through an EDFA [23].

A visual representation of the three loss criteria is shown in Fig. 2.

To provide a broader comparison of the four optimization methods, multiple application scenarios are considered by investigating different weighted cost functions  $\mathbf{m}^{(i)}$  [33]. This choice aims at covering applications that give different weight/priority to the different criteria:  $\mathbf{m}^{(0)} = J_0$ ,

$$\mathbf{m}^{(1)} = \frac{2}{3} J_0 + \frac{1}{3} J_1, \quad (5)$$

$$\mathbf{m}^{(2)} = \frac{2}{3} J_0 + \frac{1}{6} J_1 + \frac{1}{6} J_2.$$

The loss functions of Eq. (5) are used for all optimization schemes, except for the stand-alone CNN, for which the criteria  $J_0, J_1$ , and  $J_2$  cannot directly be incorporated in the training procedure as they are limited to the profiles available in the training set. Instead, the CNN is trained on the mean square error (MSE) over all distances

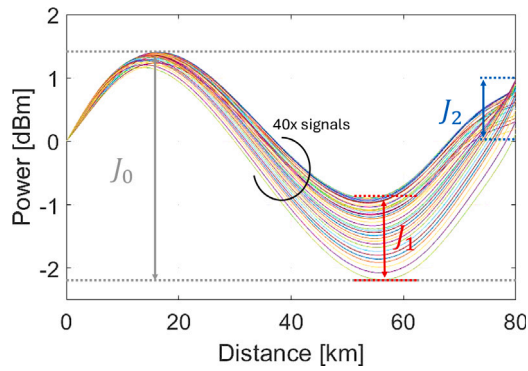


Fig. 2. Power evolution of the 40 signal channels over fiber distance showing the peak–peak estimation of the three loss criteria: maximum power excursion ( $J_0$ , gray), maximum power excursion over frequency ( $J_1$ , red), and deviation from loss-compensation ( $J_2$ , blue).

Table 2  
Comparison of accuracy performance achieved by the different methods.

Model	CNN		DE			CNN+DE			GD		
	Only	$m^{(0)}$	$m^{(1)}$	$m^{(2)}$	$m^{(0)}$	$m^{(1)}$	$m^{(2)}$	$m^{(0)}$	$m^{(1)}$	$m^{(2)}$	
$J_0$ [dB]	3.58	2.82	3.04	3.11	2.81	2.97	3.06	2.61	2.79	2.86	
$J_1$ [dB]	1.48	1.63	0.82	0.96	1.80	0.88	0.90	1.98	0.77	1.03	
$J_2$ [dB]	0.97	2.28	2.86	1.18	1.14	1.20	0.65	2.50	2.50	1.11	

and frequencies and is similarly producing a pump profile aiming at minimizing the MSE.

The cost function  $m^{(0)}$  aims at providing a baseline by minimizing the difficult target. The cost function  $m^{(1)}$  aims at relaxing the requirement for ultimate flatness in favor of minimizing a more practically relevant and potentially easier to achieve metric:  $J_1$ . Similarly,  $m^{(1)}$  further imposes the practical requirement that the gain spectrum is frequency flat. For a fair comparison, the performance in terms of  $J_0$ ,  $J_1$ , and  $J_2$  is evaluated based on the same solver applied to Eq. (1) for all modeling approaches, including the CNN-based. The performance w.r.t. each criterion for all optimization methods when each cost function is used is summarized in Table 2. For completeness,  $J_1$  is given also in the case when  $m^{(0)}$  is used, and  $J_2$  is given in the cases when  $m^{(0)}$  and  $m^{(1)}$  are used, i.e. cases in which the specific criterion evaluated was not included in the overall cost function during optimization. These values may also be of interest in the case when a multi-purpose configuration is required.

Generally, the different methods achieve similar performance in terms of  $J_0$  to within  $\pm 0.15$  dB, except the stand-alone CNN which is slightly penalized due to its dataset-restricted nature. The gradient-free stand-alone DE is slightly worse than its GD-based counterpart. The accuracy of DE is improved by initializing it using the CNN, in which case the overall best performance is achieved in terms of  $J_1$  and  $J_2$ . The GD was initialized with the maximum allowed values from Table 1.

It is noted that even though the performance is similar, it is achieved with different pump configurations in all cases as shown in Table 3, indicating the highly non-convex nature of the cost function. Initialization of all methods is thus important if more restrictive constraints on the pumps are to be imposed than the ones in Table 1.

Notice that the cost functions could be easily adapted to the desired optimization targets, e.g. in terms of specific signal gain spectrum, or output signal power profile. Similarly, if noise is taken into account, optimization in terms of noise figure or output signal-to-noise ratio could be targeted, e.g. following the approach shown in [39] for EDFA modeling. Finally, extension to multi-band operation is straightforward with only minor modification required to Eq. (1), e.g. to include the dependence of effective area and Raman gain on the pump frequency.

Table 3

Predicted pump power values in dBm. DE (stand-alone and CNN+DE) and GD are trained for  $m^{(2)}$ , while the CNN still uses MSE as cost function.

Pump \ Model	CNN	DE	CNN+DE	GD
$p_1$ [dBm]	25.2	29.1	26.5	26.9
$p_2$ [dBm]	15.2	14.2	16.7	16.8
$p_3$ [dBm]	21.6	16.0	18.8	17.4
$p_4$ [dBm]	10.8	16.5	11.5	16.1
$p_5$ [dBm]	30.1	28.2	29.9	30.5
$p_6$ [dBm]	10.8	15.9	7.8	10.4
$p_7$ [dBm]	12.8	17.1	13.2	11.5
$p_8$ [dBm]	16.3	17.1	18.0	16.9

Table 4

Applicability of the different methods.

Model	CNN	DE	CNN+DE	GD
Inference speed	High	Medium	Low	Low
Training speed	Low	Low	n.a.	n.a.
Scalability to parameters	No <sup>a</sup>	Low	Medium	High
Experimental	Yes <sup>b</sup>	Yes	Yes <sup>b</sup>	Yes <sup>c</sup>

<sup>a</sup> Requires a new training dataset.

<sup>b</sup> Requires an experimental training dataset.

<sup>c</sup> Requires parameter fitting of the model.

## 5. Applicability and discussion

The applicability of the different methods is summarized in Table 4.

The CNN provides ‘instant’ inference at the cost of a long training time, approx. an hour for the specific scenario discussed in this work. The inference time is increased if higher accuracy is required by additional optimization using DE. The more general DE, and GD methods require significantly increased optimization time: in the order of minutes to hours, depending on the complexity of the scenario and computing power. In the specific scenario discussed in this work, DE required approx. a minute per training iteration while the GD optimization took a total of a few minutes on a graphical processing unit. On the other hand, the DE and GD do not require any pre-training and are fully flexible with dynamically changing scenarios, such as variable signal load and adding/removing a pump and/or other components in the setup. Whenever such change is required, the CNN needs a new dataset and re-training. This limitation can also be extended to physics-informed NN models, which are still partially constrained to the parameters varied during training [40]. However, the stand-alone CNN is not constrained by the underlying physical dynamics beyond what could be learned through the training dataset. This means that its predictions, especially ones that extrapolate beyond the dataset range, are not necessarily solutions to the true model in (1) and may thus be non-realizable in practice, or even nonphysical. Gray-box approaches such as CNN+DE, CNN+GD, and other physics-informed models [27] are among the methods to guide the optimization to extrapolate. Nevertheless, only applying GD or DE through a physically accurate white-box model ensures the prediction/solution follows physics laws. Physics-informed models may require additional validation of their predictions to ensure physical accuracy during inference.

The flexibility of the GD method, as well as the offline (i.e. based on a model) version of the DE, comes with the requirement that the optimization model matches to a high degree the real-life setup, which may in turn requires some parameter fitting (such as fiber parameters and connector losses). Assuming that, all methods can be applied to experimental and live network scenarios, as shown for DE and CNN+DE applied to the case under test in [35]. Considering the above requirements, advantages, and disadvantages of each method, the specific optimization scenario under consideration requires to evaluate what



method proves most useful. For example, in the case of a single static optimization of a fixed Raman amplifier, using DE might be the fastest as it does not require the high training overhead or model-fitting (for an experimental setting) of CNN and GD, respectively. However, if the target profile needs to be adapted numerous times on the same Raman amplifier (e.g. to changing quality of transmission requirements on different channels or dynamic channel add/drop), CNN provides the fastest inference time, compared e.g. with DE which would require nearly restarting the optimization from scratch at every new target. Only profiles similar to those within the original training dataset of the CNN can however be targeted. Such a limitation, as mentioned, does not limit DE and GD.

## 6. Conclusions

Popular strategies for the optimization of optical amplifiers were summarized and compared. The example of a bi-directional Raman amplifier was studied with the target of flattening the frequency and distance gain spectrum. All studied methods achieve similar performance with approx. 1 dB of frequency-flatness at loss-compensated transmission, and approx. 3 dB of maximum power excursion when both frequency and distance flatness is the target. Each method poses its advantages and disadvantages based on the application scenario and the required speed for training, inference, and parameter fitting, as well as support for dynamically changing environments.

## CRediT authorship contribution statement

**Metodi P. Yankov:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Conceptualization. **Mehran Soltani:** Writing – review & editing, Software, Methodology, Investigation, Data curation, Conceptualization. **Andrea Carena:** Writing – review & editing, Supervision, Methodology, Investigation, Data curation. **Darko Zibar:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **Francesco Da Ros:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Data curation, Conceptualization.

## 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.

## Acknowledgments

This work was supported by the Villum Foundation, Denmark through the Villum Young Investigator OPTIC-AI project (grant no. VIL29334), and the Villum Investigator POPCOM project (grant no. VIL54486), by the Danish National Research Foundation, Denmark through the Center of Excellence SPOC (grant no. DNRF123), and by Ministero dell'Università e della Ricerca, Italy through the PRIN 2017 project FIRST.

## Data availability

Data will be made available on request.

## References

- [1] L. Rapp, M. Eiselt, Optical amplifiers for multi-band optical transmission systems, *J. Lightwave Technol.* 40 (6) (2021) 1579–1589.
- [2] B.J. Puttnam, G. Rademacher, R.S. Luís, Space-division multiplexing for optical fiber communications, *Optica* 8 (9) (2021) 1186–1203.
- [3] M.P. Yankov, F. Da Ros, U.C. de Moura, A. Carena, D. Zibar, Flexible Raman amplifier optimization based on machine learning-aided physical stimulated Raman scattering model, *J. Lightwave Technol.* 41 (2) (2023) 508–514.
- [4] R.T. Jones, K.R. Bottrill, N. Taengnoi, P. Petropoulos, M.P. Yankov, Spectral power profile optimization of a field-deployed wavelength-division multiplexing network enabled by remote EDFA modeling, *J. Opt. Commun. Netw.* 15 (8) (2023) C192–C202.
- [5] X. Yang, A. Ferrari, D. Le Gac, G. Charlet, M. Tornatore, Y. Pointurier, Experimental impact of power re-optimization in a mesh network, *J. Opt. Commun. Netw.* 15 (7) (2023) C20–C28.
- [6] A. Minakhmetov, B. Prieur, M. Le Monnier, D. Rouvillain, B. Lavigne, Digital twin of unrepeatable line based on Raman and remote optically pumped amplifier machine learning models, in: 2023 Optical Fiber Communications Conference and Exhibition, OFC, IEEE, 2023, pp. 1–3.
- [7] C. Headley, G.P. Agrawal, Raman Amplification in Fiber Optical Communication Systems, Academic Press, 2005.
- [8] A. Saleh, R. Jopson, J. Evankow, J. Aspell, Modeling of gain in erbium-doped fiber amplifiers, *IEEE Photonics Technol. Lett.* 2 (10) (1990) 714–717.
- [9] W.S. Pelouch, Raman amplification: An enabling technology for long-haul coherent transmission systems, *J. Lightwave Technol.* 34 (1) (2016) 6–19.
- [10] M.N. Islam, Raman amplifiers for telecommunications, *IEEE J. Sel. Top. Quantum Electron.* 8 (3) (2002) 548–559.
- [11] G. Borraccini, V. Gatto, A. D'Amico, S. Straullu, F. Aquilino, A. Nespola, S. Piciaccia, A. Tanzi, G. Galimberti, V. Curri, Gain profile characterization and modelling for dual-stage EDFA abstraction and control, *IEEE Photonics Technol. Lett.* 36 (2) (2023) 107–110.
- [12] H. Hafermann, X. Zhao, S. Yu, Y. Frignac, Preemphasis-aware semiconductor optical amplifier model, in: Optical Fiber Communication Conference, Optica Publishing Group, 2023, pp. W1E–7.
- [13] T. Eldahrawy, A. Hraghi, A. Lorences-Riesgo, T.-H. Nguyen, I. Demirtzioglou, L. Godard, H. Hafermann, N. El Dahdah, Y. Zhao, Y. Frignac, et al., Parameter estimation of semi-conductor optical amplifier booster based on digital signal processing, in: 2024 Optical Fiber Communications Conference and Exhibition, OFC, IEEE, 2024, pp. 1–3.
- [14] B. Neto, A.L.J. Teixeira, N. Wada, P.S. André, Efficient use of hybrid genetic algorithms in the gain optimization of distributed Raman amplifiers, *Opt. Express* 15 (26) (2007) 17520–17528.
- [15] M. Li, L. Zhang, B. Liu, Q. Zhang, Y. Wang, Q. Tian, X. Yin, X. Xin, K. Wang, J. Zou, Design of multi-pumped Raman fiber amplifier by differential evolution optimization, in: 2014 13th International Conference on Optical Communications and Networks, ICOON, IEEE, 2014, pp. 1–4.
- [16] G. Borraccini, S. Straullu, S. Piciaccia, A. Tanzi, G. Galimberti, V. Curri, Cognitive Raman amplifier control using an evolutionary optimization strategy, *IEEE Photonics Technol. Lett.* 34 (4) (2022) 223–226.
- [17] J. Zhou, J. Chen, X. Li, G. Wu, Y. Wang, W. Jiang, Robust, compact, and flexible neural model for a fiber Raman amplifier, *J. Lightwave Technol.* 24 (6) (2006) 2362.
- [18] J. Chen, H. Jiang, Optimal design of gain-flattened Raman fiber amplifiers using a hybrid approach combining randomized neural networks and differential evolution algorithm, *IEEE Photonics J.* 10 (2) (2018) 1–15.
- [19] U.C. de Moura, et al., Multi-band programmable gain Raman amplifier, *J. Lightwave Technol.* 39 (2) (2020) 429–438.
- [20] M. Ionescu, A. Ghazisaeidi, J. Renaudier, Machine learning assisted hybrid EDFA-Raman amplifier design for C+L bands, in: 2020 European Conference on Optical Communications, ECOC, 2020, pp. 1–3.
- [21] X. Ye, A. Arnould, A. Ghazisaeidi, D. Le Gac, J. Renaudier, Experimental prediction and design of ultra-wideband Raman amplifiers using neural networks, in: Optical Fiber Communication Conference, Optica Publishing Group, 2020, pp. W1K–3.
- [22] C. Mineto, L.S. Schanner, L.G.M. Riveros, J.H. Da Cruz, F.D. Simões, T. Sutili, F.B.F. Lopes, R.C. Figueiredo, E. Conforti, Performance of artificial-intelligence-based modelling for distributed Raman amplification, in: 2021 SBMO/IEEE MTT-S International Microwave and Optoelectronics Conference, IMOC, IEEE, 2021, pp. 1–3.
- [23] F. Da Ros, U.C. de Moura, R.S. Luís, G. Rademacher, B.J. Puttnam, A.M.R. Brusin, A. Carena, Y. Awaji, H. Furukawa, D. Zibar, Optimization of a hybrid EDFA-Raman C+L band amplifier through neural-network models, in: 2021 Optical Fiber Communications Conference and Exhibition, IEEE, 2021, p. Tu1E.5.
- [24] A. Donodin, U.C. de Moura, A.M.R. Brusin, E. Manuylovich, V. Dvoyrin, F. Da Ros, A. Carena, W. Forsysiak, D. Zibar, S.K. Turitsyn, Neural network modeling of bismuth-doped fiber amplifier, *J. Eur. Opt. Soc.-Rapid Publ.* 19 (1) (2023) 4.
- [25] M.P. Yankov, U.C. de Moura, F. Da Ros, Power evolution modeling and optimization of fiber optic communication systems with EDFA repeaters, *J. Lightwave Technol.* 39 (10) (2021) 3154–3161.

- [26] U.C. de Moura, D. Zibar, A.M.R. Brusin, A. Carena, F. Da Ros, Fiber-agnostic machine learning-based Raman amplifier models, *J. Lightwave Technol.* 41 (1) (2022) 83–95.
- [27] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *J. Comput. Phys.* 378 (2019) 686–707.
- [28] X. Jiang, M. Zhang, Y. Song, H. Chen, D. Huang, D. Wang, Predicting ultrafast nonlinear dynamics in fiber optics by enhanced physics-informed neural network, *J. Lightwave Technol.* 42 (5) (2023) 1381–1394.
- [29] M. Mei, Y. Li, M. Niu, J. Zhu, W. Li, M. Luo, Z. Feng, X. Wu, L. Mei, Q. Hu, et al., Power evolution prediction of bidirectional Raman amplified WDM system based on PINN, *Opt. Express* 32 (4) (2024) 6587–6596.
- [30] L. Zhang, E. Pincemin, N. Genay, D. Zibar, Fast and efficient solver for Raman amplifier equations employing neural networks, in: *CLEO: Science and Innovations*, Optica Publishing Group, 2024, p. STh3J.5.
- [31] X. Liu, Y. Chen, Y. Zhang, Y. Liu, L. Yi, W. Hu, Q. Zhuge, Physics-informed EDFA gain model based on active learning, 2022, arXiv preprint arXiv:2206.06077.
- [32] F. Da Ros, M. Yankov, M. Soltani, A. Carena, D. Zibar, Modeling optical amplifiers: from inverse design to full system optimization, in: 2023 IEEE Photonics Society Summer Topicals Meeting Series, SUM, IEEE, 2023, pp. 1–2.
- [33] M. Soltani, F. Da Ros, A. Carena, D. Zibar, Spectral and spatial power evolution design with machine learning-enabled Raman amplification, *J. Lightwave Technol.* 42 (12) (2022) 3546–3556.
- [34] M. Soltani, F. Da Ros, A. Carena, D. Zibar, Inverse design of a Raman amplifier in frequency and distance domains using convolutional neural networks, *Opt. Lett.* 46 (11) (2021) 2650–2653.
- [35] M. Soltani, F. Da Ros, A. Carena, D. Zibar, Experimental validation of machine-learning based spectral-spatial power evolution shaping using Raman amplifiers, *Opt. Express* 30 (25) (2022) 45958–45969.
- [36] A. Ferrari, et al., GNPpy: an open source application for physical layer aware open optical networks, *IEEE/OSA J. Opt. Commun. Networking* 12 (2020) C31–C40.
- [37] X. Gao, R. Gu, Y. Liu, Y. Ji, et al., Fast online optimization of multi-pump Raman amplifiers for field deployment in multi-band optical networks, in: 2024 Optical Fiber Communications Conference and Exhibition, OFC, IEEE, 2024, pp. 1–3.
- [38] G. Marcon, A. Galtarossa, L. Palmieri, M. Santagiustina, Model-aware deep learning method for Raman amplification in few-mode fibers, *J. Lightwave Technol.* 39 (5) (2021) 1371–1380.
- [39] M.P. Yankov, P.M. Kaminski, H.E. Hansen, F. Da Ros, SNR optimization of multi-span fiber optic communication systems employing EDFAs with non-flat gain and noise figure, *J. Lightwave Technol.* 39 (21) (2021) 6824–6832.
- [40] Y. Song, M. Zhang, X. Jiang, F. Zhang, C. Ju, S. Huang, A.P.T. Lau, D. Wang, SRS-Net: a universal framework for solving stimulated Raman scattering in nonlinear fiber-optic systems by physics-informed deep learning, *Commun. Eng.* 3 (1) (2024) 109.