Politecnico di Torino

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Electrical, Electronics and Communications Engineering
(34^{th}cycle)

Machine Learning Applications to Optical Communication Systems

By

Ann Margareth Rosa Brusin

******

Supervisor(s):
Prof. Andrea Carena, Supervisor

Doctoral Examination Committee:
Prof. Chigo Okonkwo, Referee, Eindhoven University of Technology
Prof. Qunbi Zhuge, Referee, Shanghai Jiao Tong University
Prof. Roberto Gaudino, Politecnico di Torino
Prof. Cristina Emma Margherita Rottondi, Politecnico di Torino
Prof. Paolo Serena, Università degli Studi di Parma

Politecnico di Torino
2022

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Ann Margareth Rosa Brusin
2022

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

Acknowledgements

The three years and a half of my PhD journey has come to an end and none of all I achieved would have been possible without the support and contribution of many people, both from a scientific and personal point of view.

My first big thank goes to my supervisor Andrea Carena, who has advised me since the days of my Master's thesis. I was not able to thank him at that time, so I take here the opportunity to express my deep gratitude for all the knowledge he passed to me and the constant help he provided me over these years, especially during more critical periods.

Then, I would like to thank all people of the OptCom group: the professors, in particular Vittorio Curri and Pierluigi Poggiolini, and all researchers, post-docs and PhD students with whom I have shared these years. A special thank is for the current and past members of the Optica Student Chapter Torino (previously OSA Student Chapter Torino), especially Andon Bano, Giacomo Borraccini, Bruno Correia, Elliot London, Rasoul Sadeghi and Myriam Tipan Riofrio, for all the efforts made in our dissemination activities. I also would like to thank Antonino Nespola from Links Foundation, for the support from the laboratory, and the IT staff of our department, above all Marco Bertino, for the technical support which has been particularly fundamental over the last two years due to the COVID-19 pandemic.

It is my sincere desire to acknowledge the collaborations I had the opportunity to be part of. From the academia, I would like to thank Darko Zibar, Uiara C. de Moura and Francesco da Ros from DTU Fotonik for the fruitful discussions that gave me a lot of inspirations. I also want to thank the Aston Institute of Photonic Technologies, especially Wladek Forysiak and Sergei Turitsyn, for having hosted me for a two-weeks abroad research activity. During my PhD I also had the chance of collaborating with the industry, and for this I would like to thank Fabrizio Forghieri

from Cisco Photonics for the precious hints and comments he provided me during the weekly conference calls.

Another special thank is for Fernando P. Guiomar from Instituto de Telecomunicações of Aveiro, Portugal, who guided me during my first research activity in my Master's degree thesis and convinced me in pursuing a PhD program.

Last but not least, I would like to thank my parents, my sister, Tiziana and my friends that encouraged me to always believe in myself and on my capabilities.

Abstract

In this thesis the application of machine learning (ML) and artificial neural networks (ANNs) to optical communication systems is investigated in the context of multi–band systems (O–, E–, S–, C– and L–bands). Indeed, systems operating beyond the commercial C–band ($\sim$ 5 THz) are envisioned in future optical networks to sustain the ever increasing capacity demand due to the exponential growth of Internet traffic. Consequently, upgrades and new technological advances need to be introduced in the already deployed telecommunication infrastructures to handle multi–band transmissions.

At the same time, the already existing analytical models for system performance assessment should take into account those fiber impairments that were negligible in the limited C–band which, however, become important when extending the transmission bandwidth. Among these, inter–channel stimulated Raman scattering (ISRS), responsible of a power transfer from higher to lower frequency carriers that also depends on the spectral load at the input of the fiber, may significantly affect the performance of real–time flexible multi–band systems subjected to dynamic change of data traffic.

For this reason, it is fundamental to accurately determine the impact of these effects in a fast way for a quick and suitable system response. Nevertheless, standard approaches based on analytical closed–form models and numerical solvers face an increase of computational time and complexity that follows the extension of bandwidth. Here, we propose the use of ANNs in substitution of numerical solvers and in support of analytical models in the evaluation of power profiles evolution along optical fibers, especially including ISRS. The analysis is carried out for thousands of different input spectral loads and different fiber span lengths.

Together with the extension of transmission bandwidth, new amplification technologies are required, as the currently most deployed erbium doped fiber amplifiers

(EDFAs) only operate in the C–band with an extension to the L–band. Among possible solutions, Raman amplifiers (RAs) represent an interesting cost-effective option because, by exploiting the stimulated Raman scattering (SRS) effect present in any fiber, they can provide amplification gain potentially over all optical bands without the need of installing new fibers. Furthermore, with RAs it is possible to obtain gain profiles of any shape by simply tuning the Raman pump lasers in number, wavelength (frequency) and power, solving an inverse design problem.

Based on its great potential in solving complex problems reducing their computational complexity, in this thesis ML is also applied to RAs. Precisely, we present two ANN types: the *direct* ANN used to evaluate gain and ASE noise profiles of RAs, and the *inverse* ANN to design RAs for desired gain profiles. The two ANN models are proposed as possible alternatives to standard approaches based on the numerical solution of sets of nonlinear differential equations describing the Raman effect, which indeed become highly computational demanding in case of multi–band transmission as the complexity scales with the number of equations to solve. The high prediction ability of the two ANN models is investigated for a broad range of scenarios over both synthetic and experimental data, spanning from *full* to *partial* spectral loads, analyzing different fiber types and considering different transmission bandwidths (C–, C+L– and S+C+L–band).

Contents

List of Figures **x**

List of Tables **xxi**

Nomenclature **xxiv**

1 Introduction **1**

2 Optical Communication Systems and Machine Learning Overview **4**

2.1 Optical Communication Systems 4

2.1.1 Principles of Light Propagation in Optical Fibers 5

2.1.2 Evolution of Optical Communication Systems 10

2.1.3 Raman Amplifiers 13

2.2 Machine Learning and Artificial Neural Networks 18

2.2.1 Artificial Neural Networks Concepts and Methods 20

2.2.2 Back-propagation 25

2.2.3 Random Projection 26

2.2.4 Machine Learning Libraries 28

3 Machine Learning for Power Profile Prediction in Presence of Inter-channel Stimulated Raman Scattering **29**

3.1 Introduction 29

3.2 Simulation Setup and Data–sets Generation 33
3.3 Artificial Neural Network Models 36
3.4 Prediction Results and ANNs Performance Analysis 38
3.5 Conclusions 43

4 Machine Learning for Raman Amplification Analysis 45
4.1 Introduction 45
4.2 Raman Amplification Analysis in Full Spectral Load Condition . . . 46
4.2.1 Machine Learning Framework and Data–sets Generation . . 47
4.2.2 Testing Results and Accuracy Analysis 50
4.3 Raman Amplification Analysis in Partial Spectral Loads Condition . 56
4.3.1 Data–sets Generation and Machine Learning Framework . . 57
4.3.2 Artificial Neural Network Testing and Prediction Performance 67
4.4 Generalization of ML-based Raman Amplification Analysis to Multiple Fibers 85
4.4.1 Data–sets Generation 85
4.4.2 Machine Learning Framework 86
4.4.3 Testing results 89
4.5 Conclusions 91

5 Machine Learning for Raman Amplification Design 92
5.1 Introduction 92
5.2 Machine Learning Framework 94
5.3 Machine Learning for Raman Amplifier Design over Synthetic Data 96
5.3.1 Raman Amplifier Design over C–band 98
5.3.2 Raman Amplifier Design over C+L–band 107
5.3.3 Raman Amplifier Design over C+L–band Considering Partial Spectral Loads 111

5.4 Machine Learning for Raman Amplifier Design over Experimental Data 118
5.4.1 Raman Amplifier Design over C–band 118
5.4.2 Raman Amplifier Design over C–band: Extension to Different Input Signal Power Profiles and Fiber Types 123
5.4.3 Raman Amplifier Design over C–band with Joint Noise Figure Prediction 139
5.4.4 Raman Amplifier Design over C+L–band 147
5.4.5 Raman Amplifier Design over S+C+L–band 157
5.5 Conclusions 164

6 Conclusions **166**

References **169**

List of Figures

2.1 Normalized Raman gain profile of fused silica fiber for co-polarized pump and signal. 8

2.2 Attenuation coefficient in [dB/km] of standard single-mode fiber (SSMF)/nonzero-dispersion-shifted fiber (NZ-DSF) and of low-water-peak fiber (LWPF) with respect to optical bands. (Source for attenuation coefficient profiles: [100]) 13

2.3 Raman amplification pumping schemes: (a) co-propagating and (b) counter-propagating. 14

2.4 Example of numerically simulated composite flat Raman on-off gain profile (solid line) of a 100-km-long distributed Raman amplifier with five counter-propagating pump lasers at different frequencies and powers (vertical bars). Individual gain contribution of each pump laser is plotted in dashed line. 17

2.5 Generic architecture of a fully connected feed-forward artificial neural network. 21

2.6 Scheme of an artificial neuron (node). 21

2.7 (a) Popular activation functions and (b) their derivatives. 23

3.1 Simulation setup for data–sets generation in path 1, ANN models used for power profiles prediction in path 2 and for prediction of parameters α_0, α_1 and σ used in the CFM in path 3. 34

3.2 Testing process considering channels for (a) ANN Model 1 and (b) ANN Model 2. Loss **L**, input spectral load **S** and output power **P** profiles are in logarithmic units. 37

3.3 Probability density functions of RMSE and Err_{MAX} when the ANN Model 1 trained with the *2-power-levels* data–set is tested over (a) sub–bands considering the *2-power-levels* testing data–set and (b) over channels. 39

3.4 Candle plots of prediction error E with respect to distance when the ANN Model 1 trained with the *2-power-levels* data–set is tested over (a) sub–bands considering the *2-power-levels* testing data–set and (b) over channels. 40

3.5 Probability density functions (pdfs) of Err_{MAX} of power profiles predicted using ANN Model 1 trained with the three power levels granularity data–sets for different sub–span sections: (a) 1–10 km, (b) 11–20 km, (c) 21–50 km and (d) 51–100 km. 41

3.6 Probability density functions of RMSE and Err_{MAX} of power profiles determined using α_0, α_1 and σ coefficients predicted by ANN Model 2 considering two distance ranges: (a) 1 km to 30 km and (b) 1 km to $L_{fiber} = [50, 100]$ km. 42

4.1 Schematic of simulation setup. 47

4.2 General model of a multi-layer feed-forward artificial neural network for which model averaging is applied. 48

4.3 Probability density functions of Raman on-off gain profiles prediction errors over testing data–set evaluated in terms of (a) RMSE and (b) Err_{MAX}. 51

4.4 Probability density functions of Raman on-off gain profiles prediction errors over testing data–set evaluated in terms of (a) RMSE and (b) Err_{MAX}. 52

4.5 Target and predicted Raman on-off gain profiles in case of largest maximum absolute error Err_{MAX} for training algorithms and activation functions in Table 4.3. 53

4.6 Probability density functions of ASE noise profiles prediction errors over testing data–set evaluated in terms of (a) RMSE and (b) Err_{MAX}. 54

4.7 Cumulative distribution functions of ASE noise profiles prediction errors over testing data–set evaluated in terms of (a) RMSE and (b) Err_{MAX}. 55

4.8 Target and predicted ASE noise profiles in case of largest maximum absolute error Err_{MAX} for training algorithms and activation functions of Table 4.3. 56

4.9 Simulation setup of the optical communication systems under analysis. 58

4.10 Impact of partial spectral loads with respect to full spectral load on Raman on-off gain profiles assuming fixed pump powers and frequencies when a strong tilt is present. 60

4.11 Impact of partial spectral loads with respect to full spectral load on ASE noise profiles assuming fixed pump powers and frequencies when a strong tilt is present. 61

4.12 Impact of partial spectral loads with respect to full spectral load on Raman on-off gain profiles assuming fixed pump powers and frequencies in case of almost flat profile. 62

4.13 Impact of partial spectral loads with respect to full spectral load on ASE noise on-off gain profiles assuming fixed pump powers and frequencies in case of almost flat profile. 63

4.14 Artificial neural network models for RA gain or ASE noise profiles prediction (a) not considering and (b) considering spectral loads at the input. 64

4.15 Probability density functions of RMSE and Err_{MAX} of predictions of *on-off gain* profiles for the four different scenarios: (a) LU-NN trained and tested with full–load data–sets, (b) LU-NN trained with full–load data–set and tested with partial–load data–set, (c) LU-NN trained and tested with partial–load data–sets and (d) LA-NN trained and tested with partial–load data–sets. 68

4.16 Probability density functions of RMSE and Err_{MAX} of predictions of *ASE noise* profiles for the four different scenarios: (a) LU-NN trained and tested with full–load data–sets, (b) LU-NN trained with full–load data–set and tested with partial–load data–set, (c) LU-NN trained and tested with partial–load data–sets and (d) LA-NN trained and tested with partial–load data–sets. 69

4.17 Cumulative distribution functions of RMSE and Err_{MAX} of predictions of *on-off gain* profiles for the four different scenarios of Table 4.5. *Scenario 1*: LU-NN trained and tested with full–load data–sets; *scenario 2*: LU-NN trained with full–load data–set and tested with partial–load data–set; *scenario 3*: LU-NN trained and tested with partial–load data–sets; *scenario 4*: LA-NN trained and tested with partial–load data–sets. 70

4.18 Cumulative distribution functions of RMSE and Err_{MAX} of predictions of *ASE noise* profiles for the four different scenarios of Table 4.6. *Scenario 1*: LU-NN trained and tested with full–load data–sets; *scenario 2*: LU-NN trained with full–load data–set and tested with partial–load data–set; *scenario 3*: LU-NN trained and tested with partial–load data–sets; *scenario 4*: LA-NN trained and tested with partial–load data–sets. 70

4.19 (a) Probability density functions of the Err_{MAX} evaluated for the two groups of profiles in which the partial–load data–set can be split: only fully loaded cases (class C+L sub–class 22) in blue and non–fully loaded cases in red; (b) occurrences of Err_{MAX} not exceeding 0.1 dB with respect to the number of sub–bands ON in the non–fully loaded group of profiles. 71

4.20 Average Err_{MAX} with respect to the number of sub–bands ON for the prediction of Raman on-off gain (on the left) and ASE noise (on the right) in partial load condition (partial–load testing data–set) considering different ANN models: the LU-NN trained with the full–load (FL) data–set, the LU-NN trained with the partial–load (PL) data–set and the LA-NN. The different classes (class C, class L and class C+L) are organized by rows. 75

4.21 Percentage of $\mathrm{Err}_{\mathrm{MAX}}$ under 0.5 dB with respect to the number of sub–bands ON for the prediction of Raman on-off gain (on the left) and ASE noise (on the right) in partial load condition (partial–load testing data–set) considering different ANN models: the LU-NN trained with the full–load (FL) data–set, the LU-NN trained with the partial–load (PL) data–set and the LA-NN. The different classes (class C, class L and class C+L) are organized by rows. 79

4.22 Testing prediction errors of the *Raman on-off gain* profiles in partial load condition (partial–load testing data–set) with respect to sub–bands identifiers (ordered in frequency) for different scenarios: (a) LU-NN trained with the full–load data–set, (b) LU-NN trained with the partial–load data–set and (c) LA-NN. Median values are plotted with the black cross marker, boxes and whiskers include the 50% and the 90% of cases, respectively, and the remaining cases are plotted as individual markers. 81

4.23 Testing prediction errors of the *ASE noise* profiles in partial load condition (partial–load testing data–set) with respect to sub–bands identifiers (ordered in frequency) for different scenarios: (a) LU-NN trained with the full–load data–set, (b) LU-NN trained with the partial–load data–set and (c) LA-NN. Median values are plotted with the black cross marker, boxes and whiskers include the 50% and the 90% of cases, respectively, and the remaining cases are plotted as individual markers. 84

4.24 Graphical visualization of Raman on-off gain profiles of training data–set for different fiber types and span lengths. 87

4.25 Graphical visualization of ASE noise profiles of training data–set for different fiber types and span lengths. 88

4.26 Schematic of the considered artificial neural network model. 89

4.27 (a) Probability density function (pdf) and (b) cumulative distribution function (cdf) of RMSE and $\mathrm{Err}_{\mathrm{MAX}}$ for predicted *Raman on-off gain* profiles of testing data–set. 90

4.28 (a) Probability density functions (pdf) and (b) cumulative distribution functions (cdf) of RMSE and Err_{MAX} of *ASE noise* profiles of testing data–set. 90

5.1 Machine learning framework for inverse system design: (a) inverse ANN model, (b) direct ANN model and (c) fine–optimization routine based on gradient descent (GD) optimization. 95

5.2 Simulation set–up considered for the data–sets generation. 96

5.3 Probability density functions (pdfs) of (a) RMSE and (b) Err_{MAX} computed between the target *arbitrary* gain profiles $\mathbf{G}^{targ}$ and the validated gain profiles $\mathbf{G}^{valid}$ over the C–band for two, three and five pumps. 100

5.4 Probability density functions (pdfs) of RMSE (left column) and Err_{MAX} (right column) of the validated *arbitrary* gain profiles over the C–band given by the predicted pump powers and wavelengths configurations before and after fine–optimization (FO) for: (a)-(b) two, (c)-(d) three and (e)-(f) five pumps. 101

5.5 Target and validated *flat* gain profiles over the C–band before and after fine–optimization (FO) in case of: (a)-(b) two, (c)-(d) three and (e)-(f) five pumps. 102

5.6 RMSE and Err_{MAX} with respect to gain levels of *flat* gains over the C–band before and after fine–optimization (FO) for: (a)-(b) two, (c)-(d) three and (e)-(f) five pumps. 104

5.7 Target and validated *tilted* gain profiles over the C–band before and after fine–optimization (FO) in case of: (a)-(b) two, (c)-(d) three and (e)-(f) five pumps. 105

5.8 RMSE and Err_{MAX} with respect to gain levels of *tilted* gains over the C–band before and after fine–optimization (FO) for: (a)-(b) two, (c)-(d) three and (e)-(f) five pumps. 106

5.9 Probability density functions (pdfs) of (a) RMSE and (b) Err_{MAX} computed between the target *arbitrary* gain profiles $\mathbf{G}^{targ}$ and the validated gain profiles $\mathbf{G}^{valid}$ over the C+L–band for four pumps. . . 108

5.10 Target and validated *flat* gain profiles over the C+L–band before and after fine–optimization (FO) in case of four pumps. 109

5.11 RMSE and Err_{MAX} with respect to gain levels of *flat* gains over the C+L–band before and after fine–optimization (FO) for four pumps. . 110

5.12 Target and validated *tilted* gain profiles over the C–band before and after fine–optimization (FO) in case of four pumps. 110

5.13 RMSE and Err_{MAX} with respect to gain levels of *tilted* gains over the C–band before and after fine–optimization (FO) for four pumps. 110

5.14 Schematic of the inverse load-aware neural network ANN^{LA}_{inv}. 112

5.15 (a) Probability density function (pdf) and (b) cumulative distribution function (cdf) of pump power prediction error for all five pump lasers. 113

5.16 Predicted pumps powers plotted with respect to the target true values. 114

5.17 (a) Probability density function (pdf) and (b) cumulative distribution function (cdf) of validated gain profiles given by the predicted pumps powers configurations. 115

5.18 Validated flat on–off gain profiles for predicted pumps power configurations considering different input spectral loads. 116

5.19 Probability density functions (pdfs) of RMSE and Err_{MAX} of validated flat on–off gain profiles for different partial input spectral loads with respect to target flat profiles with gain levels: (a) 4 dB, (b) 6 dB, (c) 8 dB and (d) 10 dB. 117

5.20 (a) Mean value and (b) standard deviation of RMSE and Err_{MAX} of different validated gain profiles with respect to gain levels. 117

5.21 Experimental setup used to collect data–sets and to validate the predicted pumps configuration. 119

5.22 Probability density functions (pdfs) of (a) RMSE and (b) Err_{MAX} computed between the target *arbitrary* gain profiles $\mathbf{G}^{targ}$ and the validated gain profiles $\mathbf{G}^{valid}$ over the C–band for four pumps. . . . 120

5.23 Target and validated *flat* gain profiles over the C–band before and after fine–optimization (FO) in case of four pumps. 121

5.24 RMSE and $\mathrm{Err}_{\mathrm{MAX}}$ with respect to gain levels of *flat* gains over the C–band before and after fine–optimization (FO) for four pumps. . . 122

5.25 Target and validated *tilted* gain profiles over the C–band before and after fine–optimization (FO) in case of four pumps. 122

5.26 RMSE and $\mathrm{Err}_{\mathrm{MAX}}$ with respect to gain levels of *tilted* gains over the C–band before and after fine–optimization (FO) for four pumps. 123

5.27 Experimental setup used to collect data–sets for artificial neural network training, validation and testing. It is also used to further validate the predicted pumps configurations. 124

5.28 Different *inverse* artificial neural network models ANN_{inv} depending on the information considered at the input of the network: (a) $\mathbf{P_{in}}$-unaware, (b) $\mathbf{P_{in}}$-aware and (c) $\Delta\mathrm{P_{in}}$-aware. 127

5.29 Different *direct* artificial neural network models ANN_{dir} depending on the information considered at the input of the network: (a) $\mathbf{P_{in}}$-unaware, (b) $\mathbf{P_{in}}$-aware and (c) $\Delta\mathrm{P_{in}}$-aware. 127

5.30 True power versus predicted power for *pump 2* over $\mathscr{D}_{\Delta\mathrm{P_{in}}}$-rand and $\mathscr{D}_{\Delta\mathrm{P_{in}}}$–fixed using (a) $\mathbf{P_{in}}$–unaware, (b) $\mathbf{P_{in}}$–aware and (c) $\Delta\mathrm{P_{in}}$–aware inverse ANN models. 130

5.31 Machine learning framework including fine–optimization procedure. According to the type of ANN models, input signal PSD information can be considered or not. 131

5.32 Probability density functions of $\mathrm{Err}_{\mathrm{MAX}}$ computed between target and predicted gain profiles for different ANN direct models ANN_{dir} over different testing data–sets: (a) $\mathbf{P_{in}}$–unaware model over $\mathscr{D}_{\Delta\mathrm{P_{in}}}$–fixed for various optical fibers, (b) $\mathbf{P_{in}}$–unaware model, (c) $\mathbf{P_{in}}$–aware model, and (d) $\Delta\mathrm{P_{in}}$–aware model for 100 km SSMF over both $\mathscr{D}_{\Delta\mathrm{P_{in}}}$–fixed and $\mathscr{D}_{\Delta\mathrm{P_{in}}}$–rand. 132

5.33 Probability density functions of $\mathrm{Err}_{\mathrm{MAX}}$ computed between target and validated *arbitrary* gain profiles for different types of fiber before and after fine–optimization (FO). $\mathbf{P_{in}}$–unaware ANN models are considered over $\mathscr{D}_{\Delta\mathrm{P_{in}}}$–fixed data–sets. 133

5.34 Cumulative distribution functions of $\mathrm{Err_{MAX}}$ computed between target and validated *arbitrary* gain profiles for different types of fiber before and after fine–optimization (FO). $\mathbf{P_{in}}$–unaware ANN models are considered over $\mathscr{D}_{\Delta \mathrm{P_{in}}}$–fixed data–sets. 134

5.35 Probability density functions (pdfs) of $\mathrm{Err_{MAX}}$ computed between target and validated *arbitrary* gain profiles for 100 km SSMF considering $\mathbf{P_{in}}$–unaware, $\mathbf{P_{in}}$–aware and $\Delta \mathrm{P_{in}}$–aware over both $\mathscr{D}_{\Delta \mathrm{P_{in}}}$–fixed and $\mathscr{D}_{\Delta \mathrm{P_{in}}}$–rand before and after fine–optimization (FO). 135

5.36 Cumulative distribution functions (cdfs) of $\mathrm{Err_{MAX}}$ computed between target and validated *arbitrary* gain profiles for 100 km SSMF considering $\mathbf{P_{in}}$–unaware, $\mathbf{P_{in}}$–aware and $\Delta \mathrm{P_{in}}$–aware over both $\mathscr{D}_{\Delta \mathrm{P_{in}}}$–fixed and $\mathscr{D}_{\Delta \mathrm{P_{in}}}$–rand before and after fine–optimization (FO). 136

5.37 Maximum absolute errors $\mathrm{Err_{MAX}}$ with respect to gain levels for (a) flat, (b) positively tilted and (c) negatively tilted gain profiles for different types of fibers after fine–optimization. $\mathbf{P_{in}}$–unaware models are considered. 138

5.38 Experimental setup. 140

5.39 (a) Inverse ANN model ANN_{inv} and (b) direct ANN model ANN_{dir}. . 141

5.40 Probability density functions (pdfs) of $\mathrm{Err_{MAX}}$ before and after fine–optimization (FO) for (a) Raman on–off gain profiles and (b) effective noise figure $\mathrm{NF_{eff}}$ profiles. 143

5.41 Cumulative distribution functions (cdfs) of $\mathrm{Err_{MAX}}$ before and after fine–optimization (FO) for (a) Raman on–off gain profiles and (b) effective noise figure $\mathrm{NF_{eff}}$ profiles. 143

5.42 Worst case inverse design (using ANN_{inv}): (a) on–off gain profiles before and after fine–optimization (FO), effective noise figure $\mathrm{NF_{eff}}$ (b) before FO and (c) after FO. 144

5.43 (a) Target and corresponding measured gain profiles before fine–optimization (FO), (b) target and corresponding measured gain profiles after FO, (c) target (predicted by ANN_{dir}) and corresponding NF_{eff} profiles, and (d) Err_{MAX} with respect to flat gain levels for gain design (before and after FO) and NF_{eff} profiles predictions both by ANN_{inv} and ANN_{dir}. 145

5.44 (a) Experimental setup for C+L–band Raman amplifier, (b) input optical signal spectrum and (c) qualitative representation of the normalized pump lasers spectrum and their expected contribution to the overall Raman gain. 147

5.45 Measured on–off gain profiles for different pump lasers configurations for (a) distributed and (b) discrete Raman amplifiers over the C+L–band. 149

5.46 (a) Probability density functions (pdfs) and (c) cumulative distribution functions (cdfs) of the relative absolute errors of pumps currents predicted with the ANN_{inv} for both distributed and discrete RAs. . . 151

5.47 Probability density functions (pdfs) of (a) RMSE and (b) Err_{MAX} for the arbitrary gain profiles predicted by the ANN_{dir} for both distributed and discrete RAs. 151

5.48 Probability density functions and cumulative density functions of RMSE for *arbitrary* gain profiles before and after fine–optimization for (a),(b) distributed RA and (c),(d) discrete RA. 153

5.49 Probability density functions and cumulative density functions of Err_{MAX} for *arbitrary* gain profiles before and after fine–optimization for (a),(b) distributed RA and (c),(d) discrete RA. 154

5.50 Validated and target profiles for flat and tilted gains in case of (a),(b) distributed RA and (c),(d) discrete RA. 155

5.51 (a),(b) RMSE/BW and (c),(d) Err_{MAX}/BW with respect to gain levels for flat and tilted gain profiles for both distributed and discrete RAs. 156

5.52 (a) Experimental set–up for the dual stage discrete Raman amplifier over the S+C+L–band, (b) input optical signal spectrum and (c) qualitative representation of the normalized pump lasers spectrum and their expected contribution to the overall Raman gain. 157

5.53 Measured on–off gain profiles for different pump lasers configurations for dual–stage discrete Raman amplifier over the S+C+L–band. 159

5.54 (a) Probability density function (pdf) and cumulative distribution function (cdf) of relative absolute errors of pumps currents and voltages predicted with the ANN_{inv}, and (b) pdfs of RMSE and $\mathrm{Err}_{\mathrm{MAX}}$ for the arbitrary gain profiles predicted by the ANN_{dir} for S+C+L–band dual–stage discrete RA. 161

5.55 Probability density functions and cumulative density functions of (a)-(b) RMSE and (c)-(d) $\mathrm{Err}_{\mathrm{MAX}}$ for *arbitrary* gain profiles before and after fine–optimization for S+C+L–band dual–stage discrete RA. 162

5.56 Validated and target profiles for flat and tilted gains in case of S+C+L–band dual–stage discrete RA. 163

5.57 (a) RMSE/BW and (b) $\mathrm{Err}_{\mathrm{MAX}}$/BW with respect to gain levels for flat and tilted gain profiles for for S+C+L–band dual–stage discrete RA. 164

List of Tables

2.1 Mathematical expressions of activation functions and their derivatives. 23

3.1 Hyper–parameters settings for different ranges of distances in case of ANN Model 1. 39

4.1 Parameters of SSMF where α_s and α_p are the attenuation coefficients respectively for the signal and the pumps, A_{eff} is the effective area, D is the chromatic dispersion, γ is the nonlinear coefficient and C_r is the Raman gain efficiency coefficient. 47

4.2 Pump lasers wavelength and power ranges. 49

4.3 ANNs hyper-parameters considered for different training method for both gain and ASE noise profiles prediction. 50

4.4 Pump lasers frequencies and powers. 57

4.5 Scenarios considered for LU-NN and LA-NN training, validation and testing for RA gain predictions. The optimal hyper-parameters settings are reported. 66

4.6 Scenarios considered for LU-NN and LA-NN training, validation and testing for ASE noise predictions. The optimal hyper-parameters settings are reported. 66

4.7 Parameters for different types of fibers. 86

5.1 Raman pumps power and wavelength ranges. 97

5.2 ANN topologies and hyper–parameters. 99

5.3 Pump lasers frequencies and powers. 119

5.4 Hyper–parameters setting for ANN_{inv} and ANN_{dir} operating over the C–band. 120

5.5 Pump lasers frequencies and powers. 124

5.6 Optical fibers parameters, where α: attenuation, A_{eff}: effective area, C_r: Raman gain coefficient, L_{span}: span length. 125

5.7 Neural network parameters for the different models. (*) Except SSMF 100-km; (**) Nguyen-Widrow initialization algorithm [106]; (***) Dynamically modified during training according to [49]. . . . 128

5.8 Maximum prediction error between true and predicted pumps powers using the ANN_{inv} $\mathbf{P_{in}}$–unaware model for different types of fibers considering their relative $\mathscr{D}_{\Delta P_{in}}$-fixed data–sets. 129

5.9 Maximum prediction error between true and predicted pumps powers for SSMF 100 km over $\mathscr{D}_{\Delta P_{in}}$–fixed and $\mathscr{D}_{\Delta P_{in}}$–rand data–sets for different input signal PSD awareness. 129

5.10 Hyper-parameters for inverse and direct ANN models. 142

5.11 Pump lasers wavelengths, frequencies and current ranges for both distributed and discrete RAs. 148

5.12 Artificial neural network parameters for both distributed and discrete RAs. f_{act}: activation function, N_{HL}: number of hidden layers, N_{HN}: number of hidden nodes, N_{ANN}: model combination parameter, σ_{ANN}: standard deviation of normal distribution considered for ANN weights initialization, λ_{ANN}: regularization parameter. (*) Nguyen-Widrow initialization algorithm [106]. (**) Dynamically modified during training according to [49]. 150

5.13 Pumps currents configurations for flat and tilted gain profiles after fine–optimization for distributed RA. 154

5.14 Pumps currents configurations for flat and tilted gain profiles after fine–optimization for discrete RA. 155

5.15 Pumps lasers parameters for the S+C+L–band Raman amplifier. . . 158

5.16 Artificial neural network parameters for S+C+L–band dual–stage discrete RA. f_{act}: activation function, N_{HL}: number of hidden layers, N_{HN}: number of hidden nodes, N_{ANN}: model combination parameter, σ_{ANN}: standard deviation of normal distribution considered for ANN weights initialization, λ_{ANN}: regularization parameter. 159

5.17 Pumps currents and voltage configurations for flat gain profiles after fine–optimization. 163

5.18 Pumps currents and voltage configurations for tilted gain profiles after fine–optimization. 163

Nomenclature

Acronyms

AI	Artificial Intelligence
ANN	Artificial Neural Network
ASE	Amplified Spontaneous Emission
BDFA	Bismuth Doped Fiber Amplifier
BP	Back-propagation
cdf	Cumulative Distribution Function
CFM	Closed-Form Model
CW	Continuous Wave
DCF	Dispersion Compensating Fiber
DL	Deep Learning
EDFA	Erbium Doped Fiber Amplifier
EGN	Enhanced Gaussian Noise
ELM	Extreme Learning Machine
FL	Full Load
FWM	Four-Wave Mixing
GD	Gradient Descent

GN Gaussian Noise

GSNR Generalized Signal-to-Noise Ratio

GVD Group Velocity Dispersion

HL Hidden Layer

HN Hidden Nodes

HNLF Highly Nonlinear Fiber

IDF Inverse Dispersion Fiber

IM-DD Intensity-Modulated Direct-Detection

IoT Internet-of-Things

ISRS Inter-channel Stimulated Raman Scattering

KNN K-Nearest Neighbors

LM Levenberg-Marquardt

LWPF Low Water Peak Fiber

M2M Machine-to-Machine

MAE Maximum Absolute Error

MCF Multi-Core Fiber

ML Machine Learning

MMF Multi-Mode Fiber

MPF Multiple Parallel Fibers

MSE Mean Square Error

NF Noise Figure

NLI Nonlinear Interference

NLSE Nonlinear Schrödinger Equation

NZ-DSF Non-zero Dispersion Shifted Fiber

O-E-O Optical-Electrical-Optical

ODE Ordinary Differential Equation

OPA Optic Parametric Amplifier

pdf Probability Density Function

PDFA Praseodymium Doped Fiber Amplifier

PDM Polarization Division Multiplexing

PL Partial Load

PSD Power Spectral Density

QAM Quadrature Amplitude Modulation

QoT Quality of Transmission

QPSK Quadrature Phase Shift Keying

RA Raman Amplifier

RMSE Root Mean Square Error

ROADMs Reconfigurable Optical Add/Drop Multiplexer

RP Random Projection

RS Raman Solver

SBS Stimulated Brillouin Scattering

SDM Space Division Multiplexing

SLFN Single Hidden Layer Feed-forward Neural Network

SOA Semiconductor Optical Amplifier

SPM Self-Phase Modulation

SRS Stimulated Raman Scattering

SSMF Standard Single Mode Fiber

SVM support vector machine

TDFA Thulium Doped Fiber Amplifier

ULLF Ultra-low Loss Fiber

VOA Variable Optical Attenuator

VPN Virtual Private Network

WDM Wavelength Division Multiplexing

WSS Wavelength Selective Switch

XPM Cross-Phase Modulation

List of Publications

International Peer-Reviewed Journals

1. A. M. Rosa Brusin, F. P. Guiomar, A. Lorences-Riesgo, P. P. Monteiro, and A. Carena, "Enhanced resilience towards ROADM-induced optical filtering using subcarrier multiplexing and optimized bit and power loading," Opt. Express 27, 30710-30725 (2019)

2. D. Zibar, A. M. Rosa Brusin, U. C. de Moura, F. Da Ros, V. Curri, and A. Carena, "Inverse System Design Using Machine Learning: The Raman Amplifier Case," in Journal of Lightwave Technology, vol. 38, no. 4, pp. 736-753, 15 Feb.15, 2020, doi: 10.1109/JLT.2019.2952179

3. A. M. Rosa Brusin, U. C. de Moura, V. Curri, D. Zibar, and A. Carena, "Introducing Load Aware Neural Networks for Accurate Predictions of Raman Amplifiers," in Journal of Lightwave Technology, vol. 38, no. 23, pp. 6481-6491, 1 Dec.1, 2020, doi: 10.1109/JLT.2020.3014810

4. U. C. de Moura, A. M. Rosa Brusin, A. Carena, D. Zibar, and F. Da Ros, "Simultaneous gain profile design and noise figure prediction for Raman amplifiers using machine learning," Opt. Lett. 46, 1157-1160 (2021)

5. U. C. de Moura, M. A. Iqbal, M. Kamalian, L. Krzczanowicz, F. Da Ros, A. M. Rosa Brusin, A. Carena, W. Forysiak, S. Turitsyn, and D. Zibar, "Multi–Band Programmable Gain Raman Amplifier," in Journal of Lightwave Technology, vol. 39, no. 2, pp. 429-438, 15 Jan.15, 2021, doi: 10.1109/JLT.2020.3033768

6. U. C. de Moura, F. D. Ros, A. M. Rosa Brusin, A. Carena, and D. Zibar, "Experimental Characterization of Raman Amplifier Optimization Through

Inverse System Design," in Journal of Lightwave Technology, vol. 39, no. 4, pp. 1162-1170, 15 Feb.15, 2021, doi: 10.1109/JLT.2020.3036603

International Conferences

1. A. M. Rosa Brusin, V. Curri, D. Zibar, and A. Carena, "An ultra-fast method for gain and noise prediction of Raman amplifiers," 45th European Conference on Optical Communication (ECOC 2019), 2019, pp. 1-4, doi: 10.1049/cp.2019.0976

2. U. C. de Moura, F. Da Ros, A. M. Rosa Brusin, A. Carena, and D. Zibar, "Experimental Demonstration of Arbitrary Raman Gain-Profile Designs using Machine Learning," 2020 Optical Fiber Communications Conference and Exhibition (OFC), 2020, pp. 1-3

3. A. M. Rosa Brusin, U. C. de Moura, A. D'Amico, V. Curri, D. Zibar and A. Carena, "Load Aware Raman Gain Profile Prediction in Dynamic Multi-Band Optical Networks," 2020 Optical Fiber Communications Conference and Exhibition (OFC), 2020, pp. 1-3

4. D. Zibar, U. C. de Moura, H. M. Chin, A. M. Rosa Brusin, N. Jain, F. Da Ros, S. Kleis, C. Schaeffer, T. Gehring, U. L. Andersen and A. Carena, "Advancing Classical and Quantum Communication Systems with Machine Learning," 2020 Optical Fiber Communications Conference and Exhibition (OFC), 2020, pp. 1-3

5. F. Da Ros, U. C. de Moura, R. S. Luis, G. Rademacher, B. J. Puttnam, A. M. Rosa Brusin, A. Carena, Y. Awaji, H. Furukawa, and D. Zibar, "Optimization of a Hybrid EDFA-Raman C+L Band Amplifier through Neural-Network Models," 2021 Optical Fiber Communications Conference and Exhibition (OFC), 2021, pp. 1-3

6. U. C. de Moura, D. Zibar, A. M. Rosa Brusin, A. Carena, and F. Da Ros, "Generalization Properties of Machine Learning-based Raman Models," 2021 Optical Fiber Communications Conference and Exhibition (OFC), 2021, pp. 1-3

7. A. Nespola, S. R. Sandoghchi, L. Hooper, M. Alonso, T. D. Bradley, H. Sakr, G. T. Jasion, E. Numkam Fokoua, S. Straullu, F. Garrisi, G. Bosco, A. Carena, A. M. Rosa Brusin, Y. Chen, J. R. Hayes, F. Forghieri, D. J. Richardson, F. Poletti, and P. Poggiolini, "Ultra-Long-Haul WDM Transmission in a Reduced Inter-Modal Interference NANF Hollow-Core Fiber," 2021 Optical Fiber Communications Conference and Exhibition (OFC), 2021, pp. 1-3

8. D. Zibar, U. C. de Moura, A. M. Rosa Brusin, A. Carena and F. Da Ros, "Machine learning enabled Raman amplifiers," 2021 Conference on Lasers and Electro-Optics Europe & European Quantum Electronics Conference (CLEO/Europe-EQEC), 2021, pp. 1-1, doi: 10.1109/CLEO/Europe-EQEC52157.2021.9541957

9. U. C. de Moura, F. Da Ros, D. Zibar, A. M. Rosa Brusin and A. Carena, "Optimization of Raman amplifiers using machine learning," 2021 IEEE Photonics Society Summer Topicals Meeting Series (SUM), 2021, pp. 1-2, doi: 10.1109/SUM48717.2021.9505708

10. A. M. Rosa Brusin, M. Ranjbar Zefreh, P. Poggiolini, S. Piciaccia, F. Forghieri, and A. Carena, "Machine Learning for Power Profiles Prediction in Presence of Inter-channel Stimulated Raman Scattering, " 47th European Conference on Optical Communication (ECOC 2021), 2021, Tu1D.4

11. M. Ranjbar Zefreh, A. Carena, A. M. Rosa Brusin, F. Forghieri, S. Piciaccia and P. Poggiolini, "Real Time Closed-Form Model for Nonlinearity Assessment of Fibre Optic Links with Lumped Loss," 47th European Conference on Optical Communication (ECOC 2021), 2021, Tu1D.5

Chapter 1

Introduction

Currently, the majority of Internet data traffic relies on optical fiber systems providing an infrastructure of terrestrial and submarine networks. Initially employed in long–haul and metro networks, over the last few years optical fibers widely spread in access networks and data centers as well [5]. Starting from their deployment, optical communication systems have been able to keep up with the continuously growing demand of capacity and throughput thanks to the introduction of new enabling technologies, like wavelength division multiplexing (WDM), optical amplifiers, coherent detection with advanced modulation formats and polarization division multiplexing (PDM) [3]. New Internet applications have emerged during the years, such as video streaming, Internet-of-Things (IoT), cloud computing, machine-to-machine (M2M) communications and so on, and many others will emerge in the future, with a consequently exponential increase of Internet data traffic [24, 110]. Therefore, a critical disparity in growth rate has been observed in the past years, and is still forecasted, between fiber capacity and Internet traffic, which indeed scale at $\sim$20% and $\sim$60% per year, respectively, leading towards the so–called "capacity crunch" [111]. Moreover, due to the COVID–19 pandemic, a further increase of Internet demand occurred, above all during strict lockdown periods [39].

This scenario poses new challenges to the already deployed systems, which will require to be suitably upgraded or even substituted in the next future. Transmission over multiple bands is one of the possible and promising approaches currently investigated by researchers to extend the capacity with a linear increase up to a factor of 10 with respect to the current SSMF [73]. Another widely investigated

technology is space division multiplexing (SDM) that can be implemented through multiple parallel fibers (MPF), multi-core fibers (MCF) and/or multi-mode fibers (MMF). In this case the capacity scales with the number of fibers, cores and/or modes, respectively, up to 2–3 orders of magnitude the capacity of the current SSMF [73, 96].

Unfortunately, the high capacity provided by SDM requires the deployment of new fiber plant with consequently high costs. On the other hand, multi-band transmission, since allows to exploit the broad spectrum available in already deployed single mode fibers, represents an interesting cost–effective solution most likely to be deployed in the next few years [73].

When extending the transmission bandwidth beyond the standard C–band ($\sim$ 5 THz), in addition to nonlinear Kerr effects already occurring in C–band systems, other nonlinear impairments become non–negligible, such as the stimulated Raman scattering (SRS) that also depends on the spectral load at the input of the fiber, further degrading the transmission quality [4]. Consequently, analytical models estimating system performance in real–time systems should account for SRS as well [15, 98, 99]. Unfortunately, with wider transmission bandwidth, also the computational complexity of system performance evaluation increases, with the risk of not meeting anymore speed requirements of real–time systems.

Another challenge to face in multi–band systems lies in the selection of the optical amplifier, as the commercial and mostly deployed erbium doped fiber amplifiers (EDFAs) operate in the limited range of C+L–band. Among all possible amplification technologies, Raman amplification exploiting SRS is a suitable solution to provide gains over a wide range of frequencies and potentially over all optical bands (O–, E–, S–, C– and L–bands) [51]. Most importantly, unlike other amplification technologies, Raman amplifiers (RAs) can provide any desired gain profile by properly adjusting wavelengths and powers of laser pumps. This ability is particularly relevant to achieve flat gains over the entire transmission bandwidth, which is highly desirable for WDM channels to have same transmission quality.

In the context of flexible optical networks, the system needs to react in real–time to any change in power or configuration. For this reason it is fundamental to quickly determine the impact of linear and nonlinear impairments affecting the quality of transmission and adapt accordingly. Standard approaches based on the solution of

a set of nonlinear differential equations can be time and computational demanding, thus alternative solutions should be found.

Lately, machine learning and artificial neural networks have been applied in optics and optical communication systems to solve several problems, such as inverse systems design, fiber nonlinear impairments mitigation and quality of transmission evaluation [64, 86].

In this thesis, through simulations and experimental data, we investigate the use of ML and ANNs to assist analytical models in evaluating the power evolution of the transmitted signal along the fiber in presence of ISRS. Moreover, ANNs are here applied to RAs modeling and to the design of RAs to obtain desired gain profiles (flat, tilted and arbitrary).

The thesis is organized as follows. Chapter 2 presents a brief overview of optical communication systems (principles of light propagation and systems evolution) and the machine learning methods and algorithms adopted in our investigations. In Chapter 3, two ANN models are proposed to determine power profiles evolution in presence of ISRS, which are required by analytical closed–form models (CFMs) to assess linear and nonlinear system impairments. Chapter 4 proposes the use of ANNs to predict RAs gain and ASE noise profiles over the C+L–band both considering full and partial spectral loads, as in real–time systems the network load changes dynamically. A generalization to multiple fibers is also provided. Chapter 5 presents the use of ANNs to design RAs providing arbitrary, flat and tilted gain profiles over different bandwidths: C–, C+L– and S+C+L–band. Moreover, the validity of the approach with respect to different fiber types and input spectra is shown. Finally, in Chapter 6 we draw the conclusions.

Chapter 2

Optical Communication Systems and Machine Learning Overview

In this Chapter a quick overview is given on optical communication systems to provide the context of our research. After that, machine learning methods and artificial neural networks are presented, with a more detailed discussion on the algorithms and network architectures considered in our investigations.

2.1 Optical Communication Systems

Currently, optical fibers are the most deployed technology in telecommunications, employed both in terrestrial and submarine communication networks. Indeed, a few years after their proposal in 1966 by Kao and Hockham [63], thanks to advances in the fabrication procedure, optical fibers quickly started to replace the communication systems infrastructure utilized until that time. Particularly, the migration towards optical communication systems was motivated by the significant advantages provided by optical fibers, such as low transmission loss, ultra–wide potential bandwidth, robustness and flexibility, small size and weight, and potential low cost [100].

2.1.1 Principles of Light Propagation in Optical Fibers

The light propagating along the optical fiber is subjected to linear and non–linear impairments, which, for single mode fibers assuming single polarization and neglecting polarization effects, can be modelled by the scalar nonlinear Schrödinger equation (NLSE) [2]:

$$\frac{\partial E(z,t)}{\partial z} = -\alpha E(z,t) - j\beta_0 E(z,t) - \beta_1 \frac{\partial E(z,t)}{\partial t} + j\frac{\beta_2}{2}\frac{\partial^2 E(z,t)}{\partial t^2} + \frac{\beta_3}{6}\frac{\partial^3 E(z,t)}{\partial t^3} + \\ - j\gamma\left[(1-f_{\mathrm{R}})\,|E(z,t)|^2 + f_{\mathrm{R}}\int_0^t h_{\mathrm{R}}(t-t')|E(z,t')|^2 \mathrm{d}t'\right] E(z,t) \tag{2.1}$$

where the terms in the first row refer to linear effects, while the terms in the second row refer to nonlinear effects.

Regarding linear effects, the first term describes the signal attenuation along distance z due to fiber loss, where α is the fiber loss coefficient in units [km^{-1}], such that signal *power* is attenuated over a span as [83]:

$$P(z) = P(0)\, e^{-2\alpha z}. \tag{2.2}$$

β_i with $i = [0,\ldots,3]$ are the coefficients of the Taylor series expansion of the propagation constant β truncated to the third order:

$$\beta(\omega) = \beta_0 + \beta_1\Delta\omega + \frac{1}{2}\beta_2\Delta\omega^2 + \frac{1}{6}\beta_3\Delta\omega^3, \tag{2.3}$$

with $\Delta\omega = 2\pi(f - f_0)$, where f_0 being approximately the center of the transmission band, and each β coefficient is responsible for a specific effect. β_0 determines a frequency-independent phase velocity:

$$v_p = \omega_0/\beta_0, \tag{2.4}$$

β_1 instead determines the group velocity:

$$v_g = \frac{1}{\beta_1} = \left(\frac{d\beta}{d\omega}\right)^{-1}, \tag{2.5}$$

which is characteristic of a specific spectral component at frequency ω. Hence, *fiber chromatic dispersion*, also called *group velocity dispersion* (GVD), originates. As a matter of fact, slightly different group velocities are experienced by different pulse spectral components, which consequently do not arrive concurrently at the output of the fiber, but with a delay referred to as *group delay*. Given any two frequencies, this delay is defined as:

$$\Delta T = L\beta_2\Delta\omega, \tag{2.6}$$

where L is the fiber length and β_2 is the second order coefficient of the Taylor expansion in Eq. (2.3), also called GVD parameter. Consequently, this frequency–dependent effect causes a pulse broadening that is responsible of undesired inter-symbol interference during propagation.

Another common and widely used way of expressing the group delay is with respect to a range of wavelengths $\Delta\lambda$ by means of the *dispersion coefficient D*. Thus, the group delay is expressed as:

$$\Delta T = DL\Delta\lambda, \tag{2.7}$$

where D is expressed in units of [ps/(km·nm)] and relates to β_2 through the following:

$$D = \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{1}{v_g}\right) = -\frac{2\pi c}{\lambda^2}\beta_2, \tag{2.8}$$

with c being the speed of light. From Eq. (2.8), D clearly depends on λ, therefore by computing its derivative with respect to λ, we can determine the *dispersion slope S* as:

$$S = \frac{\mathrm{d}D}{\mathrm{d}\lambda} = \left(\frac{2\pi c}{\lambda^2}\right)^2\beta_3 + \left(\frac{4\pi c}{\lambda^3}\right)^2\beta_2, \tag{2.9}$$

where β_3 is the third order coefficient in Eq. (2.3).

As long as the power of the electric field is relatively low (below 1 mW), only linear effects occur, otherwise when power is higher, nonlinear effects become significant. Among them, the Kerr effect, originating from the interaction of the electric filed with atoms and molecules of the dielectric, is the most prominent one. It corresponds to a variation of the fiber refractive index as a function of the optical power intensity P:

$$n(z,t) = n_L + n_2\frac{P(z,t)}{A_{\text{eff}}}, \tag{2.10}$$

where n_L is the conventional refractive index of the material, n_2 is the nonlinear-index coefficient, that assumes values around 2.6×10^{-20} m^2/W for silica fibers, and A_{eff} is the optical mode effective area. It is worth noticing that the variation of the refractive index is instantaneous with optical power variations. From Eq. (2.10), the nonlinear part of the refractive index is given by:

$$\Delta n = n - n_L \equiv n_2 \frac{P(z)}{A_{\text{eff}}}. \tag{2.11}$$

Although Δn results to be rather small (below 10^{-12} at 1 mW of power), such that it can be studied considering perturbation techniques, its impact over propagation is non-negligible. In particular, it follows that the mode profile does not change significantly, as it is second-order with respect to Δn, while the propagation constant β, being first-order with respect to Δn, strictly depends on power. Therefore, recalling the definition:

$$\beta \triangleq \frac{2\pi}{\lambda_{mode}} = \frac{2\pi}{\lambda} n_{\text{eff}} \approx \frac{2\pi}{\lambda} n, \tag{2.12}$$

with λ and λ_{mode} being the wavelengths in vacuum and of the mode, respectively, a variation Δn in the refractive index induces a variation $\Delta\beta$ in the propagation constant:

$$\Delta\beta = \frac{2\pi}{\lambda}\Delta n = \underbrace{\frac{2\pi}{\lambda}\frac{n_2}{A_{\text{eff}}}}_{\gamma} P(z) = \gamma P(z) \tag{2.13}$$

where

$$\gamma \triangleq \frac{2\pi n_2}{\lambda A_{\text{eff}}} \tag{2.14}$$

is called *nonlinear coefficient* and is expressed in units of [W^{-1}km^{-1}]. This γ coefficient generates a self-induced nonlinear phase shift which corresponds to the *self-phase modulation* (SPM) phenomenon.

When two or more channels are transmitted at the same time, like in WDM systems, another nonlinear effect arises called *cross-phase modulation* (XPM). In this case, the resulting phase shift for each channel accounts for the power of other channels as well. In addition, the nonlinear interaction between photons of three co-propagating frequency carriers f_1, f_2 and f_3 along the fiber generates a new photon at a fourth frequency $f_4 = f_1 + f_2 - f_3$, such that the energy is conserved. This phenomenon is know as *four-wave mixing* (FWM) and is responsible of inter-

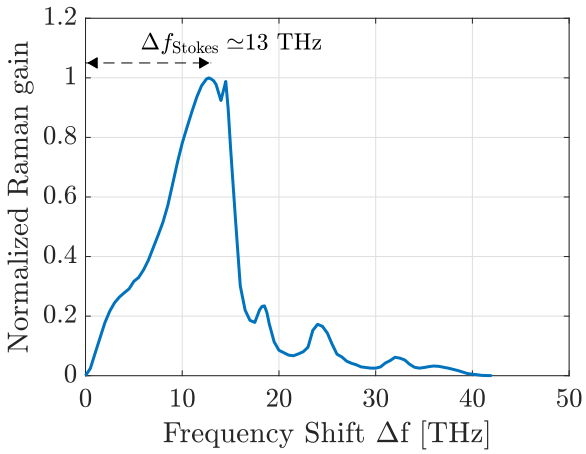


Fig. 2.1 Normalized Raman gain profile of fused silica fiber for co-polarized pump and signal.

channel cross-talk, which becomes more relevant when the fiber chromatic dispersion is very close to zero [100].

Other nonlinear effects manifest as inelastic nonlinear light scattering that causes an optical power transfer from high frequencies (pumping) to lower frequencies (scattered) providing optical gain with a shift in frequency. For this reason, these nonlinear phenomena can be exploited for optical amplification [100]. Stimulated Brillouin scattering (SBS) and stimulated Raman scattering (SRS) are the most important phenomena of this type and consist in the scattering of a photon to a second photon of lower energy, followed by the generation of a phonon. For SBS the resulting phonon is an acoustic phonon, while for SRS is an optical phonon [3].

The two phenomena are similar but some differences are present. Indeed, whether SRS can occur both in forward and backward directions, SBS occurs only in backward direction. Moreover, the SBS provides a very narrow gain spectrum with bandwidth lower than 100 MHz for which the peak of gain occurs at around $\sim$10 GHz of frequency shift. Instead, the SRS is able to provide particularly large gain spectrum, extending over 20–30 THz, with the peak at around 13 THz of frequency shift. Generally, the frequency shift between pumping frequency and scattered frequency is called Stokes shift [3]. The Raman gain profile can change for different types of fibers and depends on fiber effective area A_{eff} and fiber Raman gain coefficient g_R. Particularly, g_R is measured in [m/W] and scales linearly with pump frequency. Fig. 2.1 illustrates the normalized Raman gain profile of fused silica fiber for co-polarized pump and signal. As a first approximation, Raman gain profile of different

fibers can be obtained applying a proper scaling factor C_r to the normalized gain spectrum in Fig. 2.1 [81, 94]:

$$C_r = \frac{g_R}{A_{eff}} \tag{2.15}$$

C_r is named Raman gain efficiency coefficient and is measured in units of $[W^{-1}km^{-1}]$. From Eq. (2.15) we can infer that, assuming same g_R value, higher gains are given by fibers with smaller effective area, such as non-zero dispersion shifted fiber (NZ-DSF), for instance the Truewave-RS®, while low gains are given by large effective area fibers, like ultra-low loss fibers (ULLF) for submarine networks (TeraWave® SCUBA 125 and TeraWave® SCUBA 150).

In Eq. (2.1), the nonlinear contributions we account for are Kerr effects and SRS, whereas SBS is neglected. In particular, h_R is a complex function that represents the Raman impulse response to the vibrations of silica molecules induced by the optical field and f_R is the fractional Raman contribution [4]. When $h_R = 0$, only Kerr effects are present and the propagation equation simplifies as:

$$\frac{\partial E(z,t)}{\partial z} = -\alpha E(z,t) - j\beta_0 E(z,t) - \beta_1 \frac{\partial E(z,t)}{\partial t} + j\frac{\beta_2}{2}\frac{\partial^2 E(z,t)}{\partial t^2} - j\gamma|E(z,t)|^2 E(z,t) \tag{2.16}$$

Notice that the term proportional to the third-order dispersion coefficient β_3 disappears, as its contribution becomes significant only for the specific case of ultra-short pulses [4].

Instead, the Raman effect is modelled by the term:

$$-j\gamma f_R \int_0^t h_R(t-t')|E(z,t')|^2 \mathrm{d}t' E(z,t) \tag{2.17}$$

where f_R is about 0.18 [4]. Due to the amorphous nature of silica fibers, it is difficult to analytically determine the impulse response function $h_R(t)$. Some analytical approximations have been proposed, but they are not general and apply only in specific conditions [4]. A viable but not so practical solution is to compute it indirectly through an experimental approach based on the fact that the Raman gain spectrum is related to the imaginary part of the Fourier transform of $h_R(t)$ [4]. Then, once the imaginary part is determined, it is possible to compute the real part by means of the Kramers-Kronig relation [13].

In general, it does not exist an analytical solution for Eq. (2.1), therefore to evaluate nonlinear effects in optical fibers, simplified analytical models, such as

Gaussian Noise (GN)/Enhanced Gaussian Noise (EGN) models and closed form formulas, and numerical methods are exploited.

2.1.2 Evolution of Optical Communication Systems

Starting from the mid–1990s with the introduction of WDM, that allows to carry multiple optical channels over the same fiber, optical communication systems have undergone a rapid increase of capacity and achievable bit rates. This was possible also thanks to the continuous evolution of optical transmitters that have been able to reach increasingly higher bit rates. Indeed over the years, for intensity-modulated direct-detection (IM-DD) systems, the number of channels handled by WDM systems significantly grows from less than ten, each operating at 2.5 Gb/s, up to hundreds working at 10-40 Gb/s [3, 109].

At the same time, the introduction and the deployment of EDFAs to recover from fiber losses along the optical link, allow to fully take advantage of WDM technology [32], as their typical operational region corresponds to the frequency range where the fiber loss profile is at its minimum, i.e. the commercial band (C–band) between 191 THz and 196 THz. Operating directly in the optical domain, EDFAs completely avoid the optical-electrical-optical (O-E-O) conversion that characterized the previous amplification technology. This is beneficial also from a cost-effective point of view, as EDFAs amplify all WDM carriers simultaneously without the need of a dedicated device per wavelength. Moreover, the signal amplification can be provided at arbitrary high bit rates for each wavelength, which means that optical systems can be properly upgraded to higher per wavelength bit rates without replacing the already deployed amplifiers [5].

WDM technology also brings advantages from a networking viewpoint since, together with reconfigurable optical add/drop multiplexers (ROADMs), it is the key enabler of fully reconfigurable flexible optical networks [5], which benefit from automated connection provisioning, flexible adjustments of bandwidth and network self-healing/restoration [5].

With the introduction of coherent detection technology that allows to exploit both amplitude and phase components of the light, enabling higher order modulation formats (quadrature phase-shifted keying (QPSK) and quadrature amplitude modulation (QAM)), a further substantial boost was given to the achievable per-wavelength

bit rates. But at the same time, with higher order modulation formats, a reduction of transmission reach is observed. In addition, polarization division multiplexing (PDM), taking advantage of the dual–polarization nature of light, started to be implemented to double the per-wavelength rates, reaching values beyond 100 Gb/s [5].

Another degree of freedom to exploit for capacity increase is space division multiplexing (SDM), for which the simplest approach consists in just adding several C-band WDM systems on parallel fibers. Although this solution does not require the development of new components and sub-systems, it can be highly expensive as new fiber infrastructure should be deployed [80]. On the contrary, more sophisticated and potential approaches based on multi–core and multi–mode fibers would required a complete replacement of the infrastructure and the development of dedicated components [5, 80].

As forecasted by the Cisco annual internet report, capacity and bandwidth demand is continuously growing, following the increase of Internet applications, such as video streaming, cloud computing and services, machine-to-machine (M2M) communication, and the increase in number of devices [24]. The increase of Internet traffic further exacerbated with the COVID-19 outbreak, especially at the beginning, when people had to stay home and fully rely on the Internet for work, education, social interaction, and entertainment. In particular, more than 200% of growth has been experienced in the Internet traffic generated by applications used in remote working, like virtual private network (VPN) connectivity and video-conferencing [39]. Therefore, at some point in the next future the capacity available on the already deployed optical communication systems will saturate [110].

To cope with this increasing demand of capacity, different directions are currently being explored by researchers [80, 110], some of them already mentioned above. For instance the improvement of transmission technologies to reach higher and higher per-wavelength bit rates and of digital signal processing (DPS) techniques, the use of advanced modulation formats, such as probabilistic constellation shaping [19], or the already mentioned SDM.

Another viable and currently largely investigated solution is the increase of the optical transmission bandwidth beyond C–band towards multi–bands (O, E, S, C and L bands) [40, 80, 110]. In fact, by exploiting the broad optical spectrum characterizing optical fibers, it is possible to reuse legacy fibers with great cost-effective advantages. However, despite the costs reduction due to the fiber infrastructure reuse,

the development of new dedicated components is still required to efficiently handle transmission over multiple bands [73]. While the technology becomes mature, for short-term deployment, some studies on using C–band commercial components over S+C+L–band have been carried out [33, 37]. Regarding signal amplification, unfortunately, since commercial EDFAs operate only in the C–band with an extension to L–band, in a multiple bands scenario, other optical amplification technologies should be considered. Possible solutions rely on rare–earth-doped fiber amplifiers (xDFAs), like in [40] where xDFAs with different doping glass materials, each characterized by different noise figure (NF) values, have been used over different optical bands: Praseodymium DFA (PDFA) for O–band (NF=7 dB), Bismuth DFA (BDFA) for E–band, Thulium DFA (TDFA) for S-band and EDFA for C– and L–bands. Examples of single BDFAs operating over multiple bands can be found in [108], where the BDFA operates over the O+E–band covering 17.56 THz with gains >20 dB up to 31 dB and noise figure of 4.6$\sim$7.1 dB, and in [34], where instead the BDFA covers 10.56 THZ over the E+S–band with 27.9 dB of maximum gain and $\sim$5 dB of minimum noise figure. Also semiconductor optical amplifiers (SOAs) and optic parametric amplifiers (OPAs) are interesting solutions to provide gain in multi–band systems, like the SOA reported in [89] covering 12.7 THz over the S+C+L–band achieving 17 dB average gain and 6.5 dB average noise figure, and the OPA in [66] operating over a 10 THz bandwidth in the S+C+L–band with gains above 15 dB and noise figure below 5.1 dB. Raman amplifiers can be used in pure configuration [50, 58] or in hybrid configurations with xDFAs [43, 44, 50, 84], SOAs [7, 113] and OPAs [47].

Among all these solutions, pure RAs are particularly interesting because of their ability of providing gains potentially over all optical bands [51], as illustrated in Fig. 2.2 where, for reference, the profiles of attenuation coefficient in [dB/km] of standard single mode fiber (SSMF)/nonzero-dispersion shifted fiber (NZ-DSF) and of low water peak fiber (LWPF) are plotted with respect to optical bands. Moreover, unlike the other amplification technologies, RA gain spectrum can be tailored by simply adjusting the pump wavelengths and powers [60], which allows to obtain any desired gain profile with potentialities in different applications. However, for bandwidths beyond C+L–band, some portions of the WDM spectrum cannot be used as they will be occupied by the Raman pumps [80]. Nevertheless, the resulting band gaps are limited compared to the total optical bandwidth, with more pros than cons.

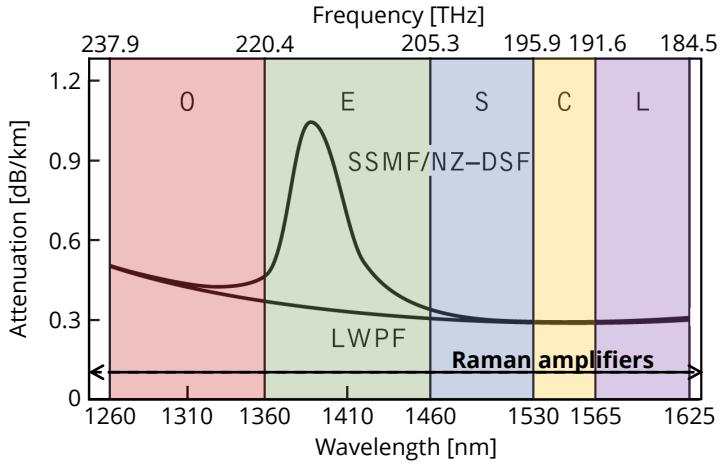
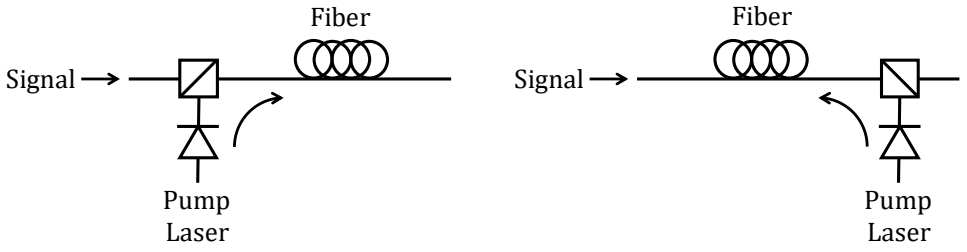


Fig. 2.2 Attenuation coefficient in [dB/km] of standard single-mode fiber (SSMF)/nonzero-dispersion-shifted fiber (NZ-DSF) and of low-water-peak fiber (LWPF) with respect to optical bands. (Source for attenuation coefficient profiles: [100])

Based on these considerations, together with the fact that distributed RAs have lower noise figure than other amplification schemes [51], in this thesis we only consider pure Raman amplifiers. Next, we provide a brief overview on Raman amplification and Raman amplifiers theory.

2.1.3 Raman Amplifiers

Raman amplifiers are based on stimulated Raman scattering consisting in the transfer of energy from a pumping photon at higher frequency to a lower energy photon at the signal frequency (lower frequency), with the remaining energy absorbed by the material as optical phonons (molecular vibrations). Depending on whether the pumping is before or after the fiber, two pumping schemes can be distinguish: co-propagating (forward) or counter-propagating (backward) pumping schemes. In Fig. 2.3, the respective schematics are shown. Also bidirectional configurations considering simultaneously co- and counter-propagating pumping are possible, but in general counter-propagating pumping schemes are preferred because of lower noise figure and reduced optical power at the beginning of the fiber [3], as the pump power does not add to the signal power. The amplification schemes discussed so far are referred to as first-order amplification, as pump and signal are separated by one Stokes shift. Also higher order Raman amplifiers are possible, where higher order pumps are used to amplify lower order pumps, instead of directly amplifying



(a) *Co-propagating* (b) *Counter-propagating*

Fig. 2.3 Raman amplification pumping schemes: (a) co-propagating and (b) counter-propagating.

the signal. These amplification schemes are generally used to further increase the reach [85].

Another classification of RAs is based on the amplifier design. If the fiber used to amplify the signal is the actual transmission fiber, then the amplification is distributed along the span and the RA is referred to as *distributed* Raman amplifier. In this case, the amplifier length corresponds to the fiber span length, with typical values above 40 km. Instead, if the RA is confined within a box placed before or after the fiber span, resulting in a lumped amplification, the RA is called *discrete* or *lumped* Raman amplifier. Compared to distributed RAs, in discrete RAs a different fiber is used to amplify the signal, with typical span lengths around 5 km [51].

In section 2.1.1, we have seen that the broad Raman gain spectrum corresponds to an instantaneous response in the time domain, with values of the order of femtoseconds, which are even shorter than typical pulse widths [4]. Consequently, it is easier to analyze the SRS effect in the frequency domain. Thus, in presence of Raman amplification, the spectral evolution of pump power P_p and power signal P_s along the fiber at distance z can be analytically expressed by the following coupled nonlinear ordinary differential equations [3, 14, 51]:

$$\frac{\mathrm{d}P_s}{\mathrm{d}z} = -2\alpha_s P_s + \mathrm{C_r}(f_s, f_p) P_p^{\pm} P_s, \tag{2.18}$$

$$\pm \frac{\mathrm{d}P_p^{\pm}}{\mathrm{d}z} = -2\alpha_p P_p^{\pm} - \left(\frac{f_p}{f_s}\right) \mathrm{C_r}(f_s, f_p) P_s P_p^{\pm}, \tag{2.19}$$

where $\mathrm{C_r}$ is the Raman gain efficiency, f_s and f_p are the frequencies respectively for signal and pump, and α_s and α_p are the respective attenuation coefficients.

In Eq. (2.19), the signs $\pm$ correspond to co-propagating pump P_p^+ and counter-propagating pump P_p^-, respectively. Analyzing the terms in the right-hand side of Eqs. (2.18) and (2.19), the first contribution corresponds to the power loss of signal and pump, respectively. Instead, the second term in Eq. (2.18) determines the contribution of signal gain, while in Eq. (2.19) the second term represents the pump depletion due to SRS.

In the *small-signal* condition, i.e. $P_s \ll \alpha_p/\mathrm{C_r}$, pump depletion term in Eq. 2.19 can be neglected, thus it is possible to analytically quantify the Raman amplification gain. Two metrics are used: the net gain and the on–off gain. The net gain is the amplifier gain at the fiber output with respect to input power and is computed as [14]:

$$\mathrm{G_{net}} = \frac{P_s(L)}{P_s(0)} = e^{-2\alpha_s L}\exp(\mathrm{C_r} L_{\mathrm{eff}} P_p^{\pm}), \tag{2.20}$$

where L is the fiber length and L_{eff} is the effective length defined as:

$$L_{\mathrm{eff}} = \frac{[1-\exp(-2\alpha_p L)]}{2\alpha_p}. \tag{2.21}$$

Instead, the Raman on–off gain represents the increase in signal power at the output of the amplifier when the pump lasers are turned on, which is given by [14]:

$$\mathrm{G_{ON-OFF}} = \frac{P_s(L) \text{ pumps ON}}{P_s(L) \text{ pumps OFF}} = \exp(\mathrm{C_r} L_{\mathrm{eff}} P_p^{\pm}). \tag{2.22}$$

In both Eqs. (2.20) and (2.22), in case of co-propagating pump pump power is taken at the fiber start, i.e. $P_p^+(0)$, whereas in case of counter-propagating pump the pump power is taken at the fiber end, that is $P_p^-(L)$.

Moreover, based on $\mathrm{G_{net}}$ and $\mathrm{G_{ON-OFF}}$ expressions, further considerations on the difference between distributed and discrete RAs can be drawn. If we neglect the dependency on pump power, the other two terms on which gain depends are $\mathrm{C_r}$ and L_{eff}, which are fiber-related. Considering that discrete RAs are constrained by size, the fiber length is quite reduced compared to distributed RAs. Thus, to achieve the same gain levels as distributed RAs, the only possibility is to use fibers with high Raman gain efficiency $\mathrm{C_r}$, i.e. with smaller effective area. This explains why highly nonlinear fibers, such as dispersion compensating fiber (DCF), highly nonlinear fiber (HNLF) and inverse dispersion fiber (IDF), are used to fabricate discrete RAs.

As for all amplifiers, together with amplification also noise is generated. In case of RAs, the major source of noise is amplified spontaneous emission (ASE) generated by spontaneous Raman scattering. Like for signal and pumps, also the power evolution of ASE noise over frequency and distance can be analytically determined and modelled as [14]:

$$\pm \frac{\mathrm{d}P_A^{\pm}}{\mathrm{d}z} = -2\alpha_A P_A^{\pm} + \mathrm{C_r}(f_A, f_p) P_p P_A^{\pm} + \mathrm{C_r}(f_A, f_p)[1+\eta(T)] h f_A B_{\mathrm{ref}} P_p \quad (2.23)$$

where $P_A^{\pm}$ is the ASE noise power in one polarization component, f_A is the frequency of the ASE, B_{ref} is the reference bandwidth over which noise is evaluated, and P_p is the total depolarized pump power at distance z. As for pumps power evolution, here as well the ASE noise propagation can occur in both directions. $\eta(T)$ is the phonon occupancy factor for the additional inhomogeneous source term, which is given by:

$$\eta(T) = \frac{1}{\exp\left(\dfrac{h\Delta f}{k_B T}\right) - 1} \quad (2.24)$$

with h being the Planck's constant, Δf is the frequency separation between pump and signal, k_B the Boltzmann's constant, and T the fiber temperature in units of Kelvin [8]. Typical boundary conditions for ASE noise are $P_A^{+}(0) = 0$ and $P_A^{-}(L) = 0$.

In WDM systems, it is fundamental to provide uniform gain to all WDM channels, especially when they are large in number, like in future ultra–wideband optical systems, requiring broadband optical amplifiers. As we have seen in the previous section, this can be achieved by employing pure Raman amplifiers operating in multiple pumps configuration, i.e. the Raman gain profile is the result of the super–position of multiple Raman gains provided by the corresponding pumps. Indeed, considering that SRS is an intrinsic phenomenon occurring in fused silica, for a given fiber, the Raman gain location and magnitude depend on the pump frequency position and power, respectively. Therefore, as the Raman gain spectrum extends over a bandwidth of 20–30 THz, with a quite flat spectrum over a pretty wide range of frequencies, by properly selecting the number of pump lasers and the corresponding pumps frequencies (wavelengths) and powers, potentially it is possible to obtain flat gain profiles over all optical bands. Fig. 2.4 shows an example of flat on–off gain profile around 8 dB with a gain flatness of ± 0.6 dB obtained by the super–position

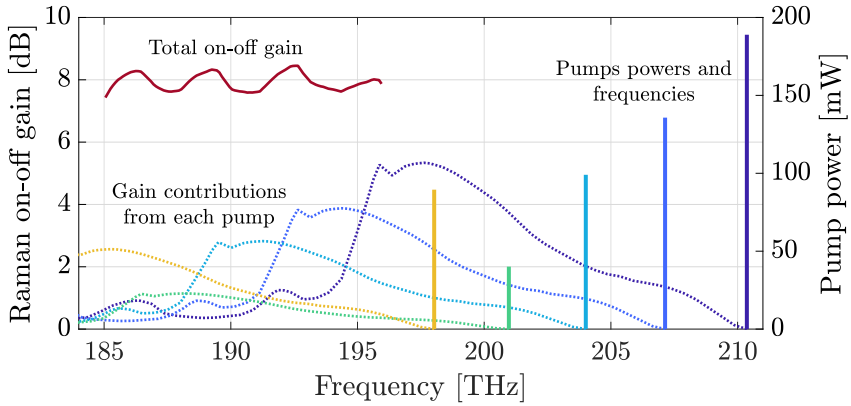


Fig. 2.4 Example of numerically simulated composite flat Raman on-off gain profile (solid line) of a 100-km-long distributed Raman amplifier with five counter-propagating pump lasers at different frequencies and powers (vertical bars). Individual gain contribution of each pump laser is plotted in dashed line.

of the Raman gain profiles given by five counter-propagating pump lasers for a 100-km-long distributed RA operating over the C+L–band.

Besides flat gains, except for physical hardware limitations, the proper tuning of pumps configuration allows to potentially obtain any desired gain shape, such as tilted and arbitrary gain profiles, that can be used in other applications, like to counteract undesired phenomena. For instance, Raman amplifiers can be used as gain flattening tools in hybrid RA+EDFA or in case of frequency combs.

It is worth to notice that the coupled equations in (2.18) and (2.19) need to be solved for each signal and pump frequency. Hence, in case of WDM systems it results in a set of equations, in numbers proportional to the number of channels and pumps. Similarly for the ASE noise.

When designing broadband Raman amplifiers (C+L–band and beyond) for WDM systems, in addition to the pump–signal interactions analyzed so far, also pump–pump and signal–signal interactions have to be taken into account, which instead are negligible for systems operating in C–band. In particular, signal–signal interactions generate ISRS, also known as nonlinear Raman–induced crosstalk, resulting in a power transfer from higher to lower frequencies and producing an SRS gain tilt [42]. Therefore, as shown in Fig. 2.4, to obtain flat gains, most of the contribution to the gain is given by the highest frequency pump laser, for which indeed higher power is required.

Considering a general scenario with N_s signals and N_p pumps, Eqs. (2.18) and (2.19) modify as follows [42]:

$$\frac{dP_{si}}{dz} = \left[-2\alpha_{si} + \sum_{j=1}^{N_p} C_{ps,ji}P_{pj}^{\pm} + \sum_{j=1}^{i-1} C_{ss,ji}P_{sj} - \sum_{j=i+1}^{N_s}\left(\frac{f_{si}}{f_{sj}}\right)C_{ss,ij}P_{sj}\right]P_{si} \quad (2.25)$$

$$\pm\frac{dP_{pi}}{dz} = \left[-2\alpha_{pi} + \sum_{j=1}^{N_s}\left(\frac{f_{pi}}{f_{sj}}\right)C_{ps,ij}P_{sj} + \sum_{j=1}^{i-1} C_{pp,ij}P_{pj}^{\pm} - \sum_{j=i+1}^{N_p}\left(\frac{f_{pi}}{f_{pj}}\right)C_{pp,ji}P_{pj}^{\pm}\right]P_{pi}^{\pm} \quad (2.26)$$

where P_{si} and P_{pi} are the instantaneous power at distance z of the i^{th} signal at frequency f_s and i^{th} pump at frequency f_p, respectively, while C_{ps}, C_{ss} and C_{pp} are the Raman gain efficiency for pump-signal, signal-signal and pump-pump interactions, respectively. Subscripts i and j are ordered in frequency from lower to higher. Compared to Eq. (2.18), in Eq. (2.25) a third and fourth term are introduced, where the third term represents the power gained by lower frequency signals and the fourth term corresponds to the power lost by higher frequency signal. Similarly for Eq. (2.26), where the third term accounts for the power gained by lower frequency pumps and the fourth term accounts for the power lost by higher frequency pumps.

Also Eq. (2.23) describing propagation and accumulation of ASE noise is modified accordingly for multiple co- and counter-propagating frequency pumps [42]:

$$\pm\frac{dP_A^{\pm}}{dz} = -2\alpha_A P_A^{\pm} + \sum_{i=1}^{N_p} C_{ri}P_{pi}\left[P_A^{\pm} + (1+\eta_i(T))\,hf_A B_{\text{ref}}\right] \quad (2.27)$$

where the subscript i refers to the i^{th} Raman pump. All the other variables have same meaning as in Eq. (2.23).

2.2 Machine Learning and Artificial Neural Networks

With the introduction of programmable computers, starting from the 1950s the concept of artificial intelligence (AI) arose, based on the idea that an algorithm is able to imitate, replicate and overcome intelligent human behavior [46]. In the 1980s, it followed the introduction of machine learning, a type of AI that provides computers with the ability of learning from data without being explicitly programmed to provide a desired output. Specifically, feature engineering, consisting in extracting

relevant information from raw data, it is fundamental to obtain accurate predictions. Afterwards, from the 2010s, an important class of ML, identified as deep learning (DL), started to be widely used. The main difference compared to ML resides on its ability of extracting useful information (patterns) directly from raw data without the need of data pre-processing.

Although at the beginning AI was limited to academia, nowadays it is largely used in the industry as well in many applications such as speech recognition, email spam detection, web search, automatic decision-making, images classification and so on. This wide spread was motivated by several progresses achieved in increased computational power, reduced cost and availability of data–sets and pre-trained models. According to the type of supervision they get, machine learning systems can be classified in: supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning [48].

In case of *supervised learning*, the algorithm is trained with examples of inputs (*features*) and corresponding outputs (*labels*). More precisely, in the training phase, supervised learning algorithms learn the proper mapping function describing the input-output relation. Afterwards, the trained model is used to make predictions on unseen data. Typical supervised learning tasks are *regression* and *classification* problems, which differ for the type of output. Whether for regression the output is a numerical real number, in case of classification the output is categorical. Among the most important supervised learning algorithm we can list: k-nearest neighbors (KNN), support vector machine (SVM), decision trees and random forests and artificial neural networks.

Instead, when *unsupervised learning* is considered, labels are not provided to the learning algorithm, but patterns recognition is left to the algorithm itself. Thus, the training data–set is said to be unlabeled. Also for this type of ML systems, different algorithms are available. An example of unsupervised learning task is *clustering*, also called *data mining*, for which the algorithm has to detect the similarities in the data.

Semi-supervised learning is placed in between of supervised and of unsupervised learning, as the algorithm receives partially labeled training data–set.

On the contrary, *reinforcement learning* operates on a completely different manner. In fact, the algorithm learns from the environment on a reward-penalty basis in

order to make the right decision/action. These types of algorithms are mostly used in robots and games like Go [48].

ML systems can also be classified based on their ability of learning incrementally or not. If the system does not have this ability, all the training data need to be used in the training phase, resulting in a highly computational time and resource demanding process which is performed offline. In this case we talk about *offline learning* or *batch learning*. When new data are available, the system has to be retrained from scratch on the entire data–set, including old data as well. Instead, with *online learning* the system is trained incrementally considering individual or mini-batches of training data–set samples. This approach is faster and cheaper than offline training and allows to easily include new data in the training. It is particularly suitable in case of a continuous flow of data [48].

Furthermore, ML systems can be classified as *instance-based learning* and *model-based learning* depending on how the generalization is performed [48]. In case of model-based learning, the system learns the example features with their patterns. Then, the generalization on new data is based on the learned patterns. Clustering is an example of instance-based learning. Whereas, when *model-based learning* is considered, the generalization is given by a model, suitably selected and trained to describe the training data–set examples and then used to make prediction on new data.

The problems studied in this thesis are all regression problems, since for a given input we aim to predict the corresponding output. This means that the ML systems considered in the next chapters are ANNs relying on supervised, offline and model-based learning. In the following we provide a brief overview of the ANN architectures and ML framework limited to our case of use. More detailed information on other ML techniques and architectures can be found in [10, 46, 48].

2.2.1 Artificial Neural Networks Concepts and Methods

Artificial neural networks, initially inspired by biological neural networks of neurons in the brain, are a ML model based on a set of connected artificial neurons, also called units or nodes [48]. Fig. 2.5 illustrates the generic schematic of a fully connected feed-forward neural network, the simplest ANN architecture where the signal flows only in one direction from the inputs to the outputs. We can identify an input layer,

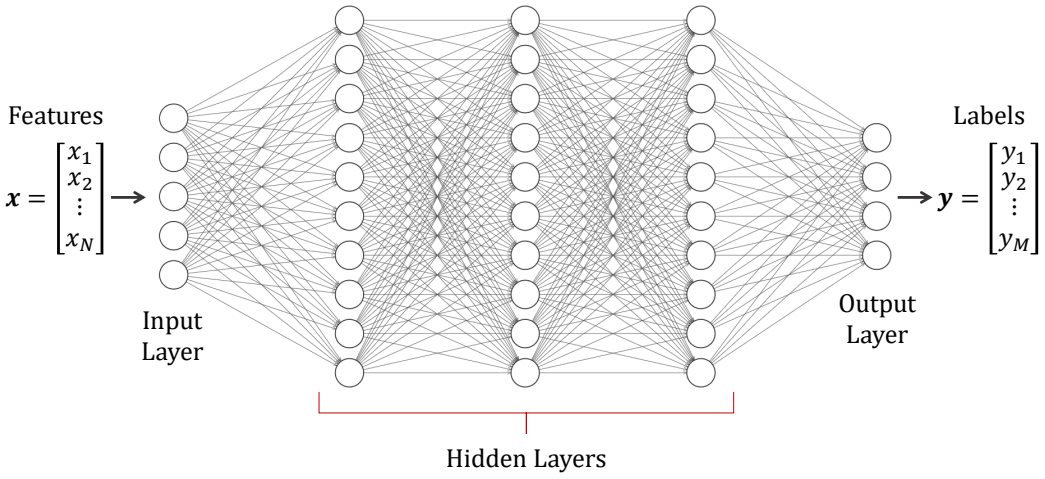


Fig. 2.5 Generic architecture of a fully connected feed-forward artificial neural network.

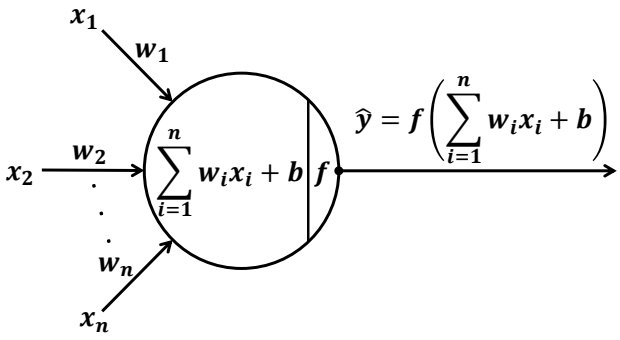


Fig. 2.6 Scheme of an artificial neuron (node).

an output layer and multiple hidden layers, where nodes are connected to each other.

In Fig. 2.6, a graphical representation of a single node is shown, where we can see that the connection between each input x_i and the node is associated to a weight w_i. To determine the output $\hat{y}$, the node computes the weighted sum of its vector of inputs $\mathbf{x} = [x_1, \ldots, x_i, \ldots, x_N]^{\mathrm{T}}$ (the T stands for transpose), adding a bias b (generally set to 1, but not strictly required), and applies a differentiable activation function $f(\cdot)$. Generally, in ML nomenclature, column vectors are considered [10]. Weights and bias are the parameters of the ANN. To better learn complex representations of

data, nonlinear functions are used to activate the node, but also linear functions can be considered, like in the output layer in case of regression problems.

The learning process consists in finding the most suitable values of ANN weights such that a cost function $f_{cost}(\cdot)$, also called objective function, is minimized. In other words, the distance between predictions and desired outputs should be minimized. The update of ANN weights is commonly implemented through gradient descent (GD) optimization algorithm. Starting from the randomly initialized weights $\mathbf{w}$, it computes the local derivative, more precisely the partial derivative, of the loss function with respect to the weights and modifies their values accordingly to decrease the cost function. The updated values of weights are computed as [48]:

$$\mathbf{w}^{(next)} = \mathbf{w} - \eta \nabla_{\mathbf{w}} f_{cost}(\mathbf{w}), \tag{2.28}$$

where η is the learning rate that determines the size of the step. This process is performed iteratively until the algorithms converges to a minimum corresponding to the optimized $\hat{\mathbf{w}}$. It is worth to notice that, for the convergence of the algorithm, the learning rate η should be properly selected. Indeed, if it is too small the algorithm will require many iterations before converging to the optimal solution, whereas if η is too large the algorithm may diverge. Depending on the loss function and on its derivative, the GD optimization algorithm might be stuck in a *local minimum* instead of reaching the *global minimum*, resulting in a sub-optimal solution.

For regression problems, the most commonly used cost functions are the mean square error (MSE) (or the root mean square error RMSE) and the mean absolute error (MAE) defined as [48]:

$$\text{MSE} = \sum_{i=1}^{M} \left(\hat{y}^{(i)} - y^{(i)} \right)^2 = \sum_{i=1}^{M} \left(f(\mathbf{w}^{\mathrm{T}}, \mathbf{x}^{(i)}) - y^{(i)} \right)^2 \tag{2.29}$$

and

$$\text{MAE} = \sum_{i=1}^{M} \left| \hat{y}^{(i)} - y^{(i)} \right| = \sum_{i=1}^{M} \left| f(\mathbf{w}^{\mathrm{T}}, \mathbf{x}^{(i)}) - y^{(i)} \right| \tag{2.30}$$

where $\hat{y}$ and y are the vectors of predicted and true values, respectively, with M being the number of elements in the vectors. The motivation behind the derivability requirement on the activation function is clear if we look at Eqs. (2.29) and (2.30), where we can notice that the prediction $\hat{y}$, present in the far right side of the loss function definition, is actually computed from the activation function. Consequently,

Activation function	$f(x)$	$f'(x)$
Logistic sigmoid (*logsig*)	$\dfrac{1}{1+e^{-x}}$	$\dfrac{e^{-x}}{(e^{-x}+1)^2}$
Hyperbolic tangent (*tanh*)	$\dfrac{e^{2x}-1}{e^{2x}+1}$	$\dfrac{4e^{-2x}}{(1+e^{-2})^2}$
Rectified linear unit (*ReLU*)	$\begin{cases} x & \text{if } x>0 \\ 0 & \text{if } x\leq 0 \end{cases}$	$\begin{cases} 1 & \text{if } x>0 \\ 0 & \text{if } x\leq 0 \end{cases}$

Table 2.1 Mathematical expressions of activation functions and their derivatives.

the activation function must be derivable as well. Popular activation functions are the logistic sigmoid (*logsig*), the hyperbolic tangent (*tanh*) and the rectified linear unit (*ReLU*). Their expression with the corresponding derivative are reported in Table 2.1 and graphically plotted in Fig. 2.7.

So far, for simplicity we have used vectors in the dissertation, but ML considers a large amount of data organized in matrices or tensors, therefore all the formulas written above can be rewritten accordingly. The collection of data used to train an ANN is called *training data–set* and each instance of the set is named *sample*. For the ANN to have a good generalization ability, it is important that the training data–set contains good and representative samples.

When implementing the GD algorithm over a training data–set, three different implementations can be identified based on how gradients are computed. We talk about *batch gradient descent* if at each GD step the derivatives are computed over the entire training data–set. Although it allows to reach the minimum with a smooth

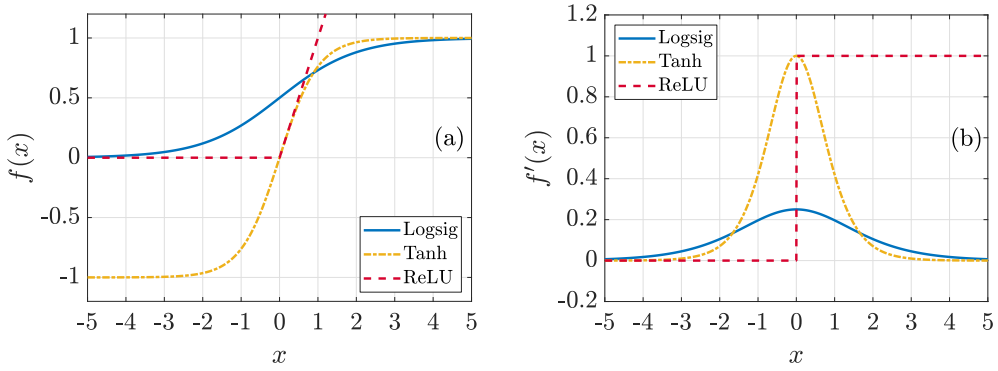


Fig. 2.7 (a) Popular activation functions and (b) their derivatives.

decreasing trend, unfortunately it becomes particularly slow when the training data–set is very large [48].

To speed up the optimization process, *stochastic gradient descent* can be implemented, for which at each step the gradient is computed with respect to a single sample randomly extracted from the training data–set, making this algorithm more suitable for large data–sets. Due to the randomness introduced in the gradient computation, the decrease of the cost function is not so regular as for batch GD. This may lead to a sub-optimal solution because the algorithm will keep bouncing around the minimum without stabilizing. Even though it might seem a drawback, this randomness proves to be efficient in escaping from local minima in case of particularly irregular cost functions [48]. A trade-off between the two approaches in represented by the *mini-batch gradient descent*, for which the gradient is computed over a small set of randomly selected samples called *mini-batches*.

To help and to speed up the algorithm convergence, it is good practice to scale features at the input of the ANN, especially when data have very different scale. Common scaling transformations are the *min-max scaling*, also known as *normalization*, for which all features are limited between 0 and 1, and *standardization*, for which data are scaled with respect to mean and standard deviation. After training, the generalization ability of the ANN is assessed over the *testing data–set*, composed of new samples that have not been included in the training.

When training an ANN, two main issues might occur that should be strongly avoided: *overfitting* and *underfitting*. Overfitting occurs when the ANN is overtrained, which means that very accurate performance is obtained on the training data–set, but the network is not able to provide a good generalization on new data. Typically, to prevent the ANN from overfitting, during the training phase part of the training data–set, called *validation data–set*, is kept aside to evaluate the generalization performance of the model. If the ANN performs well on the training data–set and not on the validation data–set, it means that the ANN is overfitted. If the validation data–set is too small, then the generalization performance might be biased by the data. A possible solution is represented by *K-fold cross–validation*, which consists in splitting the training data–set in K equal parts where at each iteration the k^{th} part is used for validation and the remaining parts are used for training. The average gives the ANN performance. On the contrary, underfitting occurs when the ANN model is not able to learn the input–output mapping, showing poor prediction performance on

the training data–set as well. Possible reasons might be that the training algorithm has not reached a minimum yet or that the model is too simple or even that features are not enough representative [48].

In general, good prediction performance is obtained considering a suitable ANN model, which is the result of a proper optimization of the ANN hyper–parameters: number of hidden layers (HLs), number of hidden nodes (HNs), activation function, learning rate η and so on. These hyper–parameters can be optimized following a grid search approach or through Bayesian optimization.

Next, we are going to describe the two training algorithm considered in this thesis: *back-propagation* (BP) and *random projection* (RP).

2.2.2 Back-propagation

Initially proposed in [95] dating back to 1986 and currently being one of the most widely used training algorithm in feed-forward neural networks, back-propagation iteratively optimizes ANN parameters based on an automatic computation of gradients. In general, the algorithm operates on mini-batches instead of considering the entire full training data–set, identifying an *iteration*. The pass over the entire training data–set identifies an *epoch*. The algorithm is composed of two main steps: the *forward pass* and the *backward pass* [48]. In the forward pass, starting from the input layer and going through the hidden layers until the output layer is reached, the algorithm computes the output of all neurons for each sample of each mini-batch, which are stored for the backward pass.

At the output layer, the algorithm measures the error between the actual network output and the desired output by means of a loss function. After that, by applying the chain rule of calculus, it determines the contribution of each output connection to this error. This process is repeated at each layer with respect to the previous one, back-propagating the error gradients from the output layer to the input layer.

Then, a gradient descent step is performed to modify the ANN weights such that the error is reduced. A more detailed description of the algorithm, with all mathematical details, can be found in [46].

Being an iterative process, one of the drawbacks of back-propagation is that it can be very time-consuming in reaching convergence, in particular when training

data–set and network architecture are particularly large, as thousands of epochs might be needed.

2.2.3 Random Projection

To overcome the training time issue of BP, in [54] random projection (RP) algorithm, also known as extreme learning machine (ELM), has been proposed for **s**ingle hidden **l**ayer **f**eed-forward neural networks (SLFNs). Indeed, the authors demonstrated that, if the activation function of the hidden layer is infinitely differentiable, by keeping fixed the randomly assigned input weights and hidden layer biases of the SLFN, the output weights linking the hidden layer to the output layer can be simply determined analytically by just inverting the hidden layer output matrices [54]. The method was proved to be thousands of times faster than BP and it was further extended in [52, 53]. Since this method is not so known as the BP algorithm, the main equations of the derivation are provided below for a better understanding of the reader.

Given a training data–set with M different samples $(\mathbf{x}_i, \mathbf{t}_i)$, where $\mathbf{x}_i = [x_{i,1}, x_{i,2}, \ldots, x_{i,n}]^{\mathrm{T}}$ is the vector of inputs and $\mathbf{t}_i = [\mathrm{t}_{i,1}, \mathrm{t}_{i,2}, \ldots, \mathrm{t}_{i,m}]^{\mathrm{T}}$ is the vector of desired output, the SLFN with L hidden nodes and activation function $f(x)$ can be modelled as [54]:

$$\sum_{i=1}^{L} \beta_i f_i(x_j) = \sum_{i=1}^{L} \beta_i f(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{o}_j, \quad j = 1, \ldots, M, \tag{2.31}$$

where $\beta_i = [\beta_{i,1}, \beta_{i,2}, \ldots, \beta_{i,m}]^{\mathrm{T}}$ is the vector of weights connecting the i^{th} hidden node and the output nodes, $\mathbf{w}_i = [w_{i,1}, w_{i,2}, \ldots, w_{i,m}]^{\mathrm{T}}$ is the vector of weights connecting the input nodes to the i^{th} node, b_i is the bias of the i^{th} hidden node, and $\mathbf{o}_i = [o_{i,1}, o_{i,2}, \ldots, o_{i,m}]^{\mathrm{T}}$ is the vector of predicted outputs. The weights are randomly initialized based on a normal distribution with zero mean and standard deviation σ, while the biases are set to 1. The output nodes are assumed to be linear (as we are considering regression problems). Given these assumptions, there exist β_i, $\mathbf{w}_i$ and b_i such that [54]:

$$\sum_{i=1}^{L} \beta_i f(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{t}_j, \quad j = 1, \ldots, M, \tag{2.32}$$

which means that $\sum_{j=1}^{L} \| \mathbf{o}_j - \mathbf{t}_j \| = 0$.

Considering that Eq. (2.32) can be written in a more compact way in matrix notation as $\mathbf{H}\beta = \mathbf{T}$, where [54]:

$$\begin{aligned}\mathbf{H}(\mathbf{w}_1,\ldots,\mathbf{w}_L,b_1,\ldots,b_L,\mathbf{x}_1,\ldots,\mathbf{x}_M) = \\ = \begin{bmatrix} f(\mathbf{w}_1\cdot\mathbf{x}_1+b_1) & \ldots & f(\mathbf{w}_L\cdot\mathbf{x}_1+b_L) \\ \vdots & \ldots & \vdots \\ f(\mathbf{w}_1\cdot\mathbf{x}_M+b_1) & \ldots & f(\mathbf{w}_L\cdot\mathbf{x}_M+b_L) \end{bmatrix}_{M\times L}\end{aligned} \tag{2.33}$$

$$\beta = \begin{bmatrix} \beta_1^{\mathrm{T}} \\ \vdots \\ \beta_L^{\mathrm{T}} \end{bmatrix}_{L\times m} \quad \text{and} \quad \mathbf{T} = \begin{bmatrix} \mathrm{t}_1^{\mathrm{T}} \\ \vdots \\ \mathrm{t}_M^{\mathrm{T}} \end{bmatrix}_{M\times m} \tag{2.34}$$

with $\mathbf{H}$ being the hidden layer output matrix of the ANN, the training of the neural network consists in finding the least-squares solution $\hat{\beta}$ such that:

$$\begin{aligned}\| \mathbf{H}(\mathbf{w}_1,\ldots,\mathbf{w}_L,b_1,\ldots,b_L)\hat{\beta} - \mathbf{T} \| = \\ = \min_{\beta} \| \mathbf{H}(\mathbf{w}_1,\ldots,\mathbf{w}_L,b_1,\ldots,b_L)\beta - \mathbf{T} \|\end{aligned} \tag{2.35}$$

Thus, the smallest norm least-squares solution to Eq. (2.32) is:

$$\hat{\beta} = \mathbf{H}^{\dagger}\mathbf{T}, \tag{2.36}$$

where $\mathbf{H}^{\dagger}$ is the Moore–Penrose generalized inverse of matrix $\mathbf{H}$ [54]. The computation of this matrix can be performed considering different methods, such as orthogonal projection method, orthogonalization method, iterative method and singular value decomposition [52]. In this thesis, we consider the orthogonal projection method, for which the Moore-Penrose matrix $\mathbf{H}^{\dagger}$ is defined as [52]:

$$\mathbf{H}^{\dagger} = \mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{H}^{\mathrm{T}}\right)^{-1} \tag{2.37}$$

if $\mathbf{H}\mathbf{H}^{\mathrm{T}}$ is non-singular or

$$\mathbf{H}^{\dagger} = \left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}} \tag{2.38}$$

if $\mathbf{H}^{\mathrm{T}}\mathbf{H}$ is non-singular. Then, to have more stable and generalized performance in the calculation of the output weights β, a positive value $1/\lambda$ is added to the diagonal of $\mathbf{H}\mathbf{H}^{\mathrm{T}}$ (or $\mathbf{H}^{\mathrm{T}}\mathbf{H}$), where λ is the regularization parameter [52]. Therefore,

substituting in Eq. (2.36), we obtain:

$$\beta = \mathbf{H}^{\mathrm{T}} \left(\frac{\mathbf{I}}{\lambda} + \mathbf{H}\mathbf{H}^{\mathrm{T}} \right)^{-1} \mathbf{T} \tag{2.39}$$

or

$$\beta = \left(\frac{\mathbf{I}}{\lambda} + \mathbf{H}^{\mathrm{T}}\mathbf{H} \right)^{-1} \mathbf{H}^{\mathrm{T}}\mathbf{T} \tag{2.40}$$

depending on the definition of $\mathbf{H}^{\dagger}$, in Eq. (2.37) or Eq. (2.38), respectively. As we can see from the equations above, standard deviation σ for weights initialization and regularization parameter λ hyper–parameters play a fundamental role in the determination of the optimal solution of β, therefore they should be properly selected.

Like in case of training with BP, to improve prediction performance optimization of hyper–parameters can be performed, both through grid search or Bayesian optimization.

2.2.4 Machine Learning Libraries

Currently, thanks to the spread of machine learning use and application, several machine learning libraries are available. Among them, in this thesis to train our ANNs with the BP algorithm, we use the Deep Learning Toolbox™ [104] from MathWorks® in MATLAB® environment, and TensorFlow™ [1] and Keras [20] libraries, which instead are written in Python language. Instead, the RP algorithm has been developed in MATLAB® environment based on the equations reported in section 2.2.3 [52–54].

Chapter 3

Machine Learning for Power Profile Prediction in Presence of Inter-channel Stimulated Raman Scattering

In this Chapter we propose the use of machine learning and artificial neural networks for the prediction of power profiles to support the quality of transmission (QoT) estimation in multi-band system scenarios, where the ISRS effect becomes relevant. Our analysis is carried out considering systems operating in the C+L–band, but it can be easily extended to larger bandwidths. Specifically, to determine the impact of both linear and nonlinear interference (NLI) effects, an accurate and fast evaluation of power profiles over frequency and distance is required, therefore two ANN models are introduced as support. One model is used to directly predict the evolution of power profiles to assess the impact of linear impairments, while the other is adopted to determine specific parameters used in closed-form model formulas for the NLI estimation. The results discussed in this Chapter have been presented in [93].

3.1 Introduction

In the last decade, for the analysis and the optimization of optical networks performance, several analytical models can be found in the literature to describe the

propagation over optical fibers and the behavior of the physical layer. Mainly introduced when transmission was limited to the C–band only, these models account for ASE noise and NLI impairments [16, 27, 62, 83, 97, 101]. In Chapter 2 we have seen that, with the optical communication systems evolving towards multi-band scenarios, in addition to the Kerr effect causing the NLI generation between channels, another source of nonlinearity becomes relevant, the ISRS, which must be accounted in analytical models [15, 99]. Causing a power transfer from higher frequencies to lower frequencies, most importantly ISRS largely depends on the spectral load at the input of the fiber.

Given a WDM comb with N channels, its propagation along the fiber can be modeled by a set of coupled nonlinear ordinary differential equations (ODEs), one for each channel j, as [88, 102]:

$$\begin{cases} \dfrac{\mathrm{d}P_1(z)}{\mathrm{d}z} = P_1(z)\left[-2\alpha_1 + \displaystyle\sum_{i=2}^{N} \mathrm{C_r}(f_i - f_1)P_i(z)\right] \\ \vdots \\ \dfrac{\mathrm{d}P_j(z)}{\mathrm{d}z} = P_j(z)\left[-2\alpha_j + \displaystyle\sum_{i=j+1}^{N} \mathrm{C_r}(f_i - f_j)P_i(z) - \sum_{i=1}^{j-1} \frac{f_j}{f_i}\mathrm{C_r}(f_j - f_i)P_i(z)\right] \\ \vdots \\ \dfrac{\mathrm{d}P_N(z)}{\mathrm{d}z} = P_N(z)\left[-2\alpha_N - \displaystyle\sum_{i=1}^{N-1} \frac{f_N}{f_i}\mathrm{C_r}(f_N - f_i)P_i(z)\right] \end{cases} \tag{3.1}$$

where j is in the range $2 \leq j \leq (N-1)$, z is the distance with respect to the origin of the fiber where the WDM comb is launched, α_j is the fiber loss coefficient, that can be different for every channel, and $\mathrm{C_r}$ is the Raman gain efficiency coefficient, which is specific to the fiber and depends on its effective area. Unfortunately, there is no analytical solution to Eq. (3.1), therefore it can be solved only numerically by means of numerical Raman solvers. Alternatively, to guarantee real–time predictions, introducing some approximations, simplified closed–form models (CFMs) have been derived to compute the NLI contribution, including the ISRS [23, 87, 88, 98].

In the context of multi–band systems, considering that larger bandwidths correspond to a larger number of channels, we assist to an increase of computational effort required to numerically determine the power profile evolution, as also the number of equations to solve increases. This drawback, together with the increase of the convergence time to the solution, makes numerical solvers unsuitable to be

used in real–time applications. Especially in quality of transmission (QoT) evaluation, where to determine the impact of both linear and nonlinear effects, a fast and accurate evaluation of power profiles over distance and frequency is required. The computational effort increases also in case of CFMs, like in [88] where the CFM relies on specific parameters which are determined channel by channel by fitting the actual power profile (evaluated numerically) with the CFM simplified analytical expression.

During the last few years, machine learning and artificial neural networks have become quite popular in the optical communications community thanks to the promising results obtained in different applications, such as in the analysis and design of Raman amplifiers [55, 56, 70, 91, 113, 117]. As an alternative to numerical solvers and CFMs, but also to meet speed requirements and reduce the complexity of the problem, following the considerations made so far on ML and ANNs, we propose two ANN models to support the evaluation of spectrally and spacially resolved power profiles. One model is used to directly predict the power profiles for the linear impairments and the other one to predict parameters α_0, α_1 and σ used to determine power profiles evolution inserted in CFM formulae to compute the nonlinear interactions [88].

Determination of coefficients α_0, α_1 and σ

In the following we report the main steps of the derivation presented in [88] to determine the coefficients α_0, α_1 and σ used to compute the power profile evolution inserted in CFMs. The authors start considering the following equation:

$$P_j(z) \cong P_j(0) \times \exp\left(-2\alpha_{0,j}z + \frac{2\alpha_{1,j}}{\sigma_j}\left(e^{-\sigma_j z} - 1\right)\right), \tag{3.2}$$

which is an approximated solution for the power evolution over channel j originally demonstrated in [23] assuming specific conditions which are: triangular approximation for the Raman efficiency profile, WDM bandwidth smaller than the ISRS bandwidth f_{ISRS} ($(f_N - f_1) \leq \Delta f_{\text{ISRS}}$, with N being the number of WDM channels), identical loss for all frequencies in absence of ISRS ($\alpha_1 = \cdots = \alpha_N = \alpha_0$), equally spaced WDM channels (Δf_{ch}), negligible photon conversion loss and same power for all WDM channels ($P_1(0) = P_2(0) = \cdots = P_N(0) = P_0$). In addition, a further

assumption is made [88]:

$$N^2 P_0 \times \frac{C_{r,max}}{\Delta f_{\text{ISRS}}} \times \Delta f_{ch} \times \frac{1-\exp(-2\alpha_0 z)}{2\alpha_0} \ll 1, \forall z \tag{3.3}$$

where $\text{C}_{\text{r},max}$ is the *ISRS gain profile maximum*. α_0, α_1 and σ in Eq. (3.2) are defined as [88]:

$$\alpha_{0,j} \triangleq \alpha_0 \tag{3.4}$$

$$\alpha_{1,j} \triangleq -\frac{NP_j(0)}{4} \frac{\text{C}_{\text{r},max}}{\Delta f_{ISRS}} (f_N + f_1 - 2f_j) \tag{3.5}$$

$$\sigma_j \triangleq 2\alpha_0 \tag{3.6}$$

From their definition, we can see that $\alpha_{0,j}$ and σ_j are frequency invariant, while $\alpha_{1,j}$ can be different channel by channel. Assuming a more realistic scenario, in [82], for a more detailed and accurate description of the system behavior, the possibility of having channel-dependent α_0, α_1 and σ is introduced. This leads to the following approximated model equation [88]:

$$\begin{aligned} P(f,z) \cong P(f,0) \times \exp\left(-2\alpha_0(f)z + \frac{2\alpha_1(f)}{\sigma(f)}\left(e^{-\sigma(f)z}-1\right)\right) = \\ = P(f,0) \times \exp\left(-2\int_0^z \left[\alpha_0(f) + \alpha_1(f)e^{-\sigma(f)z}\right] dz\right) \end{aligned} \tag{3.7}$$

which is still a simple and compact formula to be inserted into the GN-model closed-form formula to evaluate the NLI contribution [82]. Thus, introducing an approximation error $\varepsilon(f,z)$, the formula in Eq. (3.7) can be used to determine the *true* power evolution $P_{num}(f,z)$ [88]:

$$P_{num}(f,z) = P(f,0) \times \exp\left(-2\alpha_0(f)z + \frac{2\alpha_1(f)}{\sigma(f)}\left(e^{-\sigma(f)z}-1\right)\right) + \varepsilon(f,z), \tag{3.8}$$

with $|\varepsilon(f,z)| \ll |P_{num}(f,z)|$ and where in general $\varepsilon(f,z) \neq 0$.

The optimal values of $\alpha_0(f)$, $\alpha_1(f)$ and $\sigma(f)$ are determined by fitting the actual power profile $P_{num}(f,z)$, evaluated numerically with high accuracy, such that the error $\varepsilon(f,z)$, defined in [88] as:

$$\varepsilon(f,z) \cong P_{num}(f,z) \times \left[ln\left(\frac{P_{num}(f,z)}{P(f,0)} + 2\alpha_0(f)z + \frac{2\alpha_1(f)}{\sigma(f)} - \frac{2\alpha_1(f)}{\sigma(f)}e^{-\sigma(f)z}\right)\right] \tag{3.9}$$

is minimized. To obtained more precise estimations of $\alpha_0(f)$, $\alpha_1(f)$ and $\sigma(f)$ that takes into account the dependency of $\varepsilon(f,z)$ on distance z and the higher NLI generation when the power is higher, a new cost function is introduced [88]:

$$f_{cost} \triangleq \int_0^{\mathrm{L_{span}}} (P_{num}(f,z))^{m_c} \times \left(\frac{\varepsilon(f,z)}{P_{num}(f,z)}\right)^2 dz \cong \int_0^{\mathrm{L_{span}}} (P_{num}(f,z))^{m_c} \times \left[ln\left(\frac{P_{num}(f,z)}{P(f,0)}\right) + 2\alpha_0(f)z + 2\alpha_1(f)\cdot\frac{1-e^{-\sigma(f)z}}{\sigma(f)}\right]^2 dz \quad (3.10)$$

where $m_c \geq 0$ is a constant optionally selected for better estimations. In our analysis its value is set to $m_c = 2$. In particular, this cost function looks at the errors over the span length $\mathrm{L_{span}}$ assigning higher weights when the powers are higher. This means that the estimation of $\alpha_0(f)$, $\alpha_1(f)$ and $\sigma(f)$ is more accurate in the first kilometers of the span. For all the details on the derivation of $\alpha_0(f)$, $\alpha_1(f)$ and $\sigma(f)$ please refer to [88].

Eqs. (3.8)–(3.10) are continuous in frequency f and distance z, but in our analysis we will use their discretized version. In particular, frequencies are discretize according to the frequency resolution, while distance is discretized with respect to the spacial resolution. Therefore, the following notation holds: $\alpha_0(f) = \alpha_0[n] = \alpha_{0,n}$, $\alpha_1(f) = \alpha_1[n] = \alpha_{1,n}$, $\sigma(f) = \sigma[n] = \sigma_n$, where n is an integer number assuming values in range $n = [1,\ldots,\mathrm{N_f}]$, with $\mathrm{N_f}$ being the number of frequency points.

Once $\alpha_{0,n}$, $\alpha_{1,n}$ and σ_n have been optimized for each frequency point, they can be inserted in Eq. (3.11) [88]:

$$P_n[z] = P_n[0] \times \exp\left(-2\alpha_{0,n}z + \frac{2\alpha_{1,n}}{\sigma_n}\left(e^{-\sigma_n z} - 1\right)\right) \quad (3.11)$$

to evaluate the power profile evolution over frequencies and distances of interest.

3.2 Simulation Setup and Data–sets Generation

Fig. 3.1 shows the simulation setup considered to generated the synthetic data–sets used to train, validate and test the ANN (path 1). Over the 11 THz C+L–band defined between 185 THz and 196 THz, we consider a WDM comb of $\mathrm{N_{ch}} = 220$ channels in the 50 GHz frequency grid and each channel can be turned ON or OFF

with same probability. When turned ON, it carries 1 mW of power. Then, the comb is propagated over a single span of SSMF with $L_{span} = 100$ km length and $\alpha = 0.21$ dB/km. To study the impact of ISRS, the evolution of power profiles is determined over frequency and distance by means of the numerical Raman solver available within the open source library GNPy [45] with a spatial resolution of 1 km.

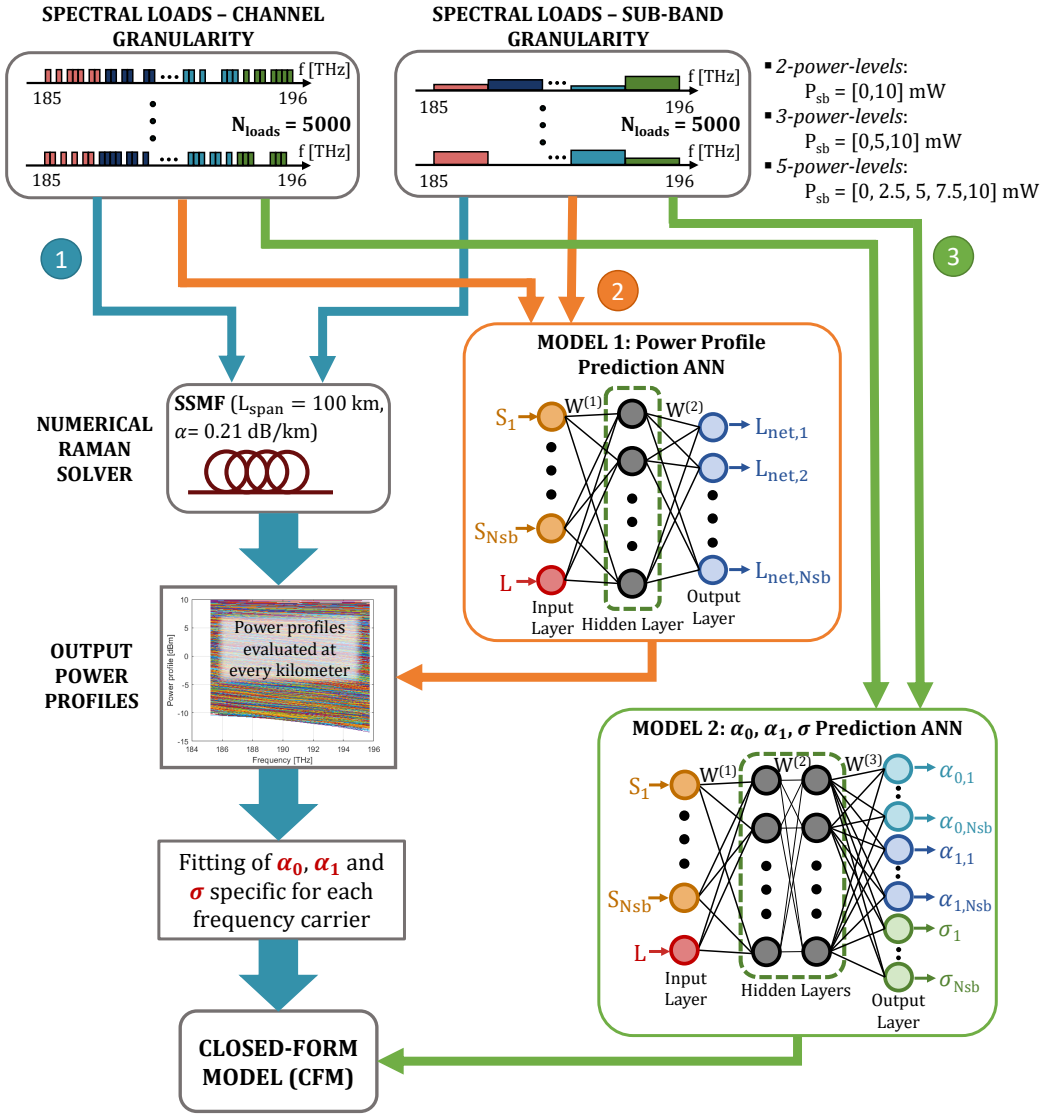


Fig. 3.1 Simulation setup for data–sets generation in path 1, ANN models used for power profiles prediction in path 2 and for prediction of parameters α_0, α_1 and σ used in the CFM in path 3.

Using this numerical solver, we generate two independent and uncorrelated data–sets: a *training data–set* for ANN training and validation, and a *testing data–set* to test the ANN over unseen data. Validation is performed to monitor and avoid any possible *underfitting* or *overfitting* of the ANN. To handle different load conditions, the training data–set should include a comprehensive and representative set of spectral load cases. Unfortunately, given the number of channels and the on/off states, the total number of combinations for the input spectral load would be 2^{220}, which is an unmanageable dimension to generate. Based on the fact that power loss profiles do not change significantly with respect to frequency of adjacent channels, to reduce the dimension of the space to be visited for the training data–set, we can implement an approach similar to the concept of super-channel, where a set of very tightly spaced channels goes through routing devices as a single entity [12, 114]. In our case, ten adjacent channels are grouped together to form a sub–band of 500 GHz, for a total of $N_{sb} = 22$ sub-bands [91]. Setting the frequency granularity to 10 channels per sub–band is a good compromise between accuracy and complexity, which indeed is reduced by a factor of 10. Furthermore, the generation of the training data–set on the sub–band basis gives a double advantage: the reduced number of spectral conditions to be analyzed and the reduced ANN complexity, since the spectral resolution of the output profiles decreases, with the number of elements reducing from 220 to 22.

In the generation of partial–loads, since the frequency position of channels ON is randomly selected, they are not uniformly distributed over sub–bands. For this reason, in the emulation of partial–loads, sub–bands can assume multiple power levels P_{sb} with different granularity. In particular, depending on the power levels granularity, we identify and generate three types of data–sets for ANN training: with *2-power-levels*, i.e. $P_{sb} = [0, 10]$ mW, which corresponds to have all channels OFF or ON, respectively, with *3-power-levels*, i.e. $P_{sb} = [0, 5, 10]$ mW, and with *5-power-levels*, i.e. $P_{sb} = [0, 2.5, 5, 7.5, 10]$ mW. For all the three types of multi–power levels data–sets, $N_{load} = 5000$ different spectral load conditions are extracted with random frequency position and power level for sub–bands.

For ANN testing, a second and independent data–set with *2-power-levels* is generated on sub–band basis for a preliminary assessment of the ANN performance. Then, another testing data–set is generated on a channel–granularity basis, as in general in realistic scenarios we are interested in evaluating the system performance channel by channel. Each channel can be OFF or ON with same probability, with

$P_{ch} = [0, 1]$ mW. Also in this case we generate $N_{load} = 5000$ different partial–load cases.

Following path 1 in Fig. 3.1, if we apply the procedure proposed in [88] to the generated data–sets, each power profile can be fitted to determine the triplet of parameters $\{\alpha_0, \alpha_1, \sigma\}$ necessary to enable the CFM with ISRS defined in Eq. (3.11). Considering that the NLI contribution is not the same along the fiber span and that the NLI is mostly generated in the first kilometers of the span where the power is higher (approximately from the span origin to its effective length), the cost function used in the fitting procedure assigns different weights to the errors along the span. Especially, higher weights are assigned to regions experiencing higher power in order to have a more accurate fitting, whereas lower weights are assigned to the remaining portions of the span.

The advantage of using machine learning and artificial neural networks is analyzed on standard machines over the testing data–set generated on channels–basis. To evaluate the power profiles evolution at every kilometer for the 5000 partial input spectral loads, the numerical Raman solver takes $9.5 \cdot 10^5$ s, i.e. on average around 190 s ($\sim$3 minutes) per profile, while fitting requires $2.5 \cdot 10^4$ s, i.e. 5 s on average per profile. Unfortunately, with these numbers, the numerical solvers do not satisfy the computational time required by real–time applications. On the contrary, the ANN complexity is all at the training stage performed offline. Thus, when used online after training, it provides an instantaneous evaluation of each power profile, requiring just only few tens of seconds for all the 5000 cases. As a consequence, ANNs turn out to be a potentially suitable solution to guarantee fast and highly accurate predictions.

3.3 Artificial Neural Network Models

We have seen that to have a real time evaluation of power profiles along the fiber span in frequency and distance and to overcome the intrinsic complexity resulting from the analytical solution of ODEs, machine learning and artificial neural networks can be used instead of numerical Raman solvers. In path 2 of Fig. 3.1, we report the ANN model, which we refer to as Model 1, used to predict power profiles. At its input, it receives the spectral load condition $\mathbf{S} = [S_1, \ldots, S_{Nsb}]^T$ per sub–band and the distance L at which we want to estimate the power profile. At the output of the ANN there is the loss profile $\mathbf{L_{net}} = [L_{net,1}, \ldots, L_{net,Nsb}]^T$ defined as the difference

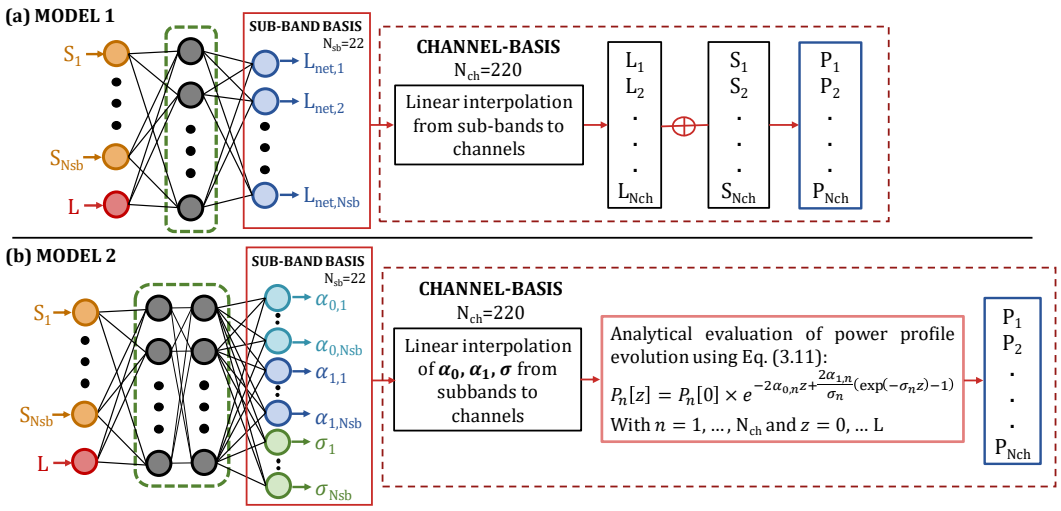


Fig. 3.2 Testing process considering channels for (a) ANN Model 1 and (b) ANN Model 2. Loss **L**, input spectral load **S** and output power **P** profiles are in logarithmic units.

in logarithmic units between the power profile at distance L and the corresponding input spectral load. Thus, based on this definition, we can easily obtain the output power profile by just summing up the input spectral load to the loss profile. As mentioned above, since sub–band granularity is considered for the ANN training, the dimensions of ANN input and output are quite limited, being 23 and 22, respectively.

To speed up the process, also the determination of the triplet of coefficients $\{\alpha_0, \alpha_1, \sigma\}$ used in the CFM derived in [88] for non-linearity assessment, in particular to compute the power spectral density of NLI, G_{NLI} (Eq. (92) in [88]), can be done by an ANN, that we refer to as ANN Model 2, shown in path 3 of Fig. 3.1. There we can see that the ANN receives the same input as model 1, i.e. 23 points, while the output is now represented by the the coefficients α_0, α_1, σ of each sub–band, i.e. 66 points. Details on ANNs hyper-parameters settings and training methods for both ANN models will be extensively described in the following section. For both models, input and output are properly normalized with respect to their mean and standard deviation.

The testing process for the two ANN models considering channels is illustrated in Figs. 3.2(a) and 3.2(b) for Model 1 and Model 2, respectively. At the input of both ANNs we provide the power per sub–band given by the contribution of the channels ON contained in each sub–band. At the output of Model 1, linear interpolation is performed for the 22 loss points over the 220 channels, which are then added to the

per–channel input power to get the actual output power profile. Linear interpolation is performed also in case of Model 2 to get coefficients α_0, α_1, σ over channels, which are then use in Eq. (3.11) to determine power profiles in order to evaluate the accuracy of their predictions.

The prediction accuracy is evaluated by computing the error E between the predicted power profile and the target power profile given by the numerical Raman solver, which is defined as $\mathrm{E} = P^{pred} - P^{targ}$. Furthermore, for each profile we also compute the root mean squared error (RMSE) and the maximum absolute error ($\mathrm{Err_{MAX}}$), defined in Eqs. (3.12) and (3.13):

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N_f}} \sum_{n=1}^{\mathrm{N_f}} (P_n^{pred} - P_n^{targ})^2} \tag{3.12}$$

$$\mathrm{Err_{MAX}} = \max_{n \in \{1,\ldots,\mathrm{N_f}\}} |P_n^{pred} - P_n^{targ}| \tag{3.13}$$

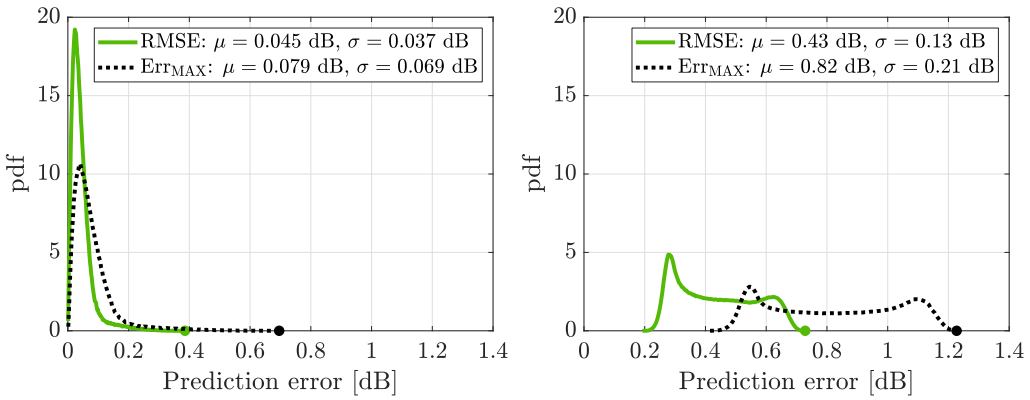
where $\mathrm{N_f}$ can be the number of sub–bands $\mathrm{N_{sb}}$ or the number of channels $\mathrm{N_{ch}}$ depending on the frequency granularity of the profile. To guarantee very high accurate predictions, we set a target maximum absolute prediction error $\mathrm{Err_{MAX}} \leq$ 0.1 dB.

3.4 Prediction Results and ANNs Performance Analysis

Starting with ANN Model 1, we first train this ANN model with the training data–set with *2-power-levels* considering all kilometers up to the fiber span length (1–100 km). We consider the *random projection* (RP) as training method with hyper–parameters settings indicated in Table 3.1: $N_{\mathrm{HL}} = 1$ hidden layer, $N_{\mathrm{HN}} = 1000$ hidden nodes, standard deviation for ANN weights initialization $\sigma_{\mathrm{ANN}} = 0.1$, regularization parameter $\lambda_{\mathrm{ANN}} = 10$ and hyperbolic tangent (tanh) as activation function. To improve generalization properties, as prediction accuracy may be affected by the random initialization of ANN weights, model combination is employed [46, 10]. This means that N_{ANN} different ANNs are trained and the predictions are given by averaging the predictions made by each single model.

Distance [km]	**1–100**	**1–10**	**11–20**	**21–50**	**51–100**
Activation function	tanh	tanh	tanh	tanh	tanh
Number of hidden layers N_{HL}	1	1	1	1	1
Number of hidden nodes N_{HN}	1000	1000	1000	1000	2000
Model combination parameter N_{ANN}	20	20	20	20	20
Standard deviation σ_{ANN}	0.1	0.1	0.1	0.1	0.1
Regularization parameter λ_{ANN}	10	10	10	10	10

Table 3.1 Hyper–parameters settings for different ranges of distances in case of ANN Model 1.

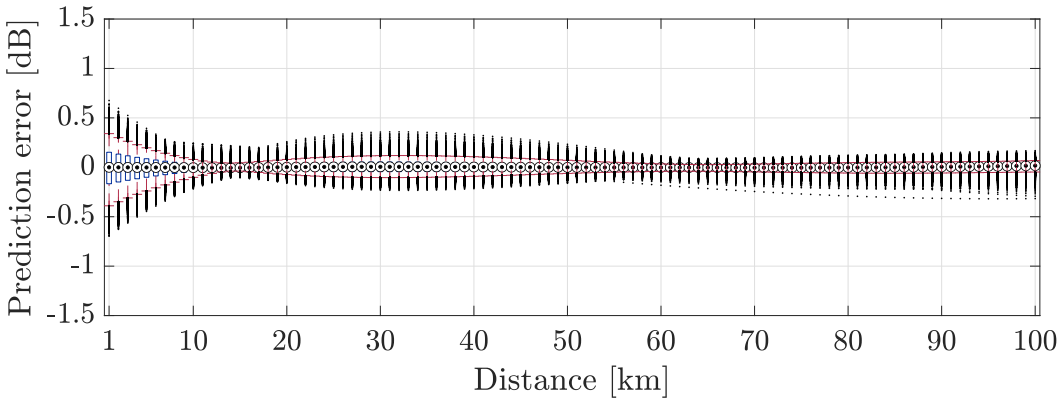


(a) *Sub–band basis* (b) *Channel basis*

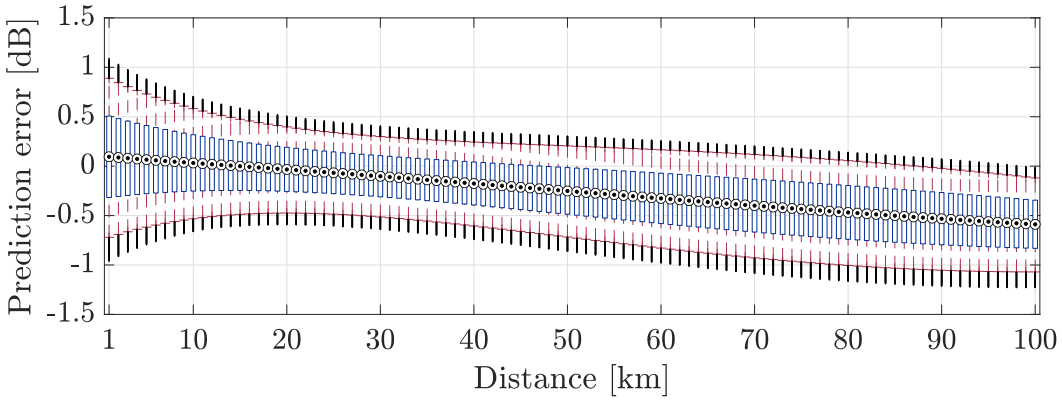
Fig. 3.3 Probability density functions of RMSE and Err_{MAX} when the ANN Model 1 trained with the *2-power-levels* data–set is tested over (a) sub–bands considering the *2-power-levels* testing data–set and (b) over channels.

Once trained, we make a preliminary assessment of the ANN prediction capabilities by testing it over the testing data–set with *2-power-levels* generated on a sub–band basis. The results are illustrated in Fig. 3.3(a), where we report the probability density functions (pdf) of the RMSE and the Err_{MAX}. Although the mean values of RMSE and Err_{MAX} are quite low, being 0.045 dB and 0.079 dB respectively, the goal of having all the Err_{MAX} below or equal to 0.1 dB is not achieved, with the tail of the pdf also reaching values up to 0.7 dB. In fact, $\text{Err}_{\text{MAX}} \leq 0.1$ dB for only 75.8% of cases.

At this point, it can be interesting to see how the prediction errors E are distributed with respect to distance. Since it is not practical to plot their probability density function for all distances, we introduce a more compact way of representing the pdf, which are depicted as candle plots, as shown in Fig. 3.4(a). Thus, for every kilometer,



(a) *Sub–band basis*

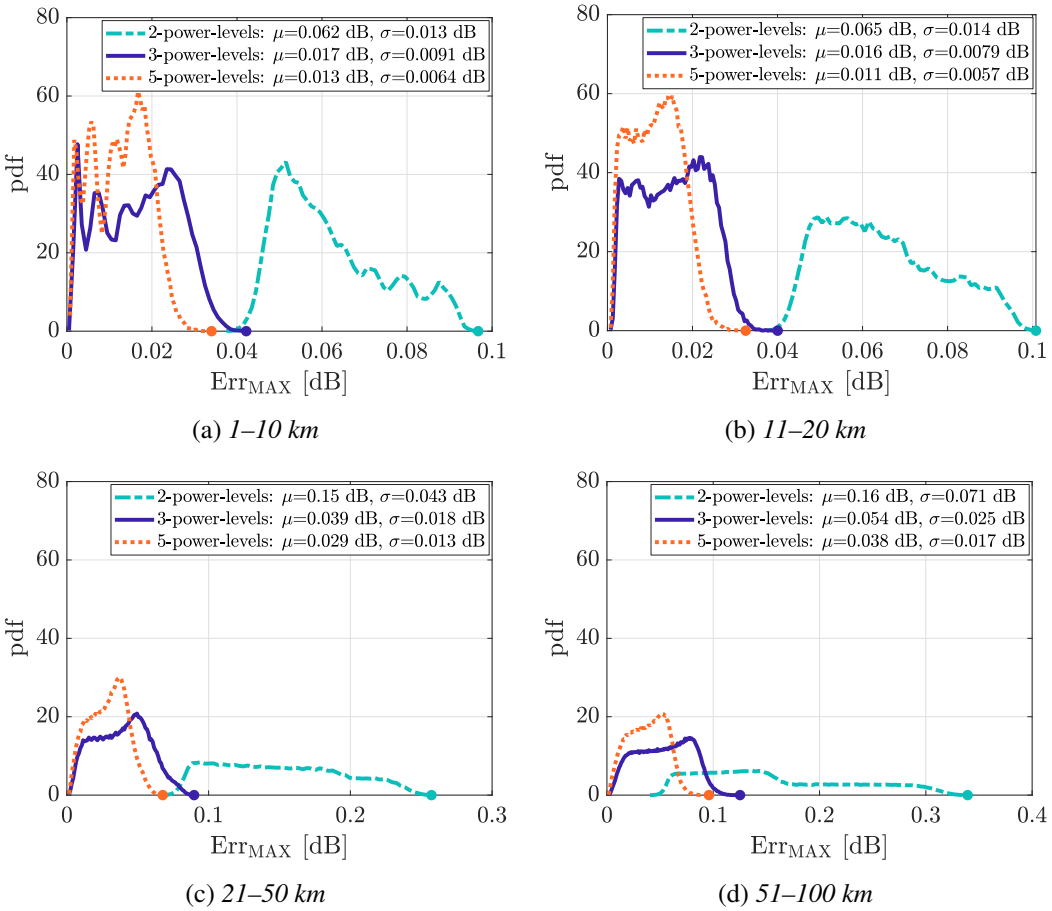


(b) *Channel basis*

Fig. 3.4 Candle plots of prediction error E with respect to distance when the ANN Model 1 trained with the *2-power-levels* data–set is tested over (a) sub–bands considering the *2-power-levels* testing data–set and (b) over channels.

the candle plot reports the median value of the pdf, indicated with a black circle, a box in blue containing the 50% of total cases between the 25% and 75% of the pdf, and whiskers in red including the 90% of cases between the 5% and 95% of the pdf. The remaining 5% of cases falling in the tails of each side of the pdf, are plotted as individual markers in black. From there, we can see that larger errors occur in the first kilometers, while starting from 50 km the errors are quite flat with distance.

Afterwards, as we want to use the ANN in a more realistic scenario, we also test it over channels considering the testing data–set generated on the channel basis. From the pdfs reported in Fig. 3.3(b), the performance of the ANN is even worse, since the average RMSE and $\mathrm{Err}_{\mathrm{MAX}}$ increase up to 0.43 dB and 0.82 dB, respectively, and the values of $\mathrm{Err}_{\mathrm{MAX}}$ are always above 0.4 dB. Moreover, compared to the sub–bands



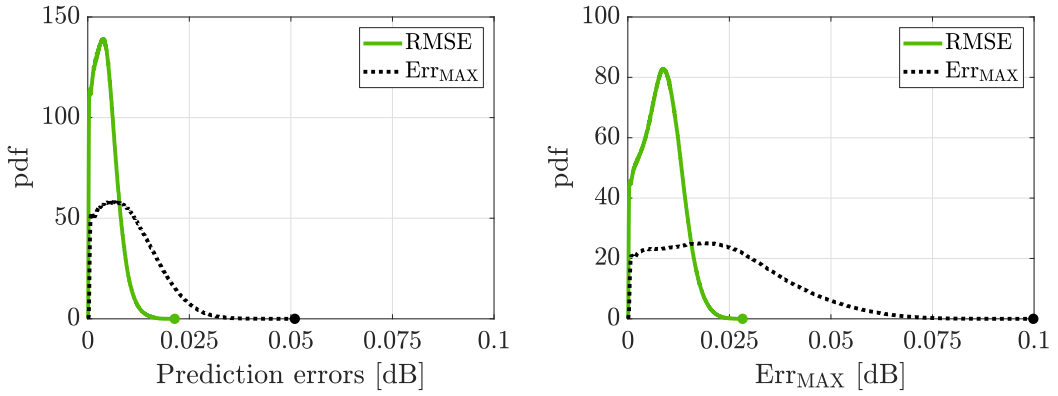
(a) *1–10 km* (b) *11–20 km*

(c) *21–50 km* (d) *51–100 km*

Fig. 3.5 Probability density functions (pdfs) of $\mathrm{Err}_{\mathrm{MAX}}$ of power profiles predicted using ANN Model 1 trained with the three power levels granularity data–sets for different sub–span sections: (a) 1–10 km, (b) 11–20 km, (c) 21–50 km and (d) 51–100 km.

case, the error values, especially the median, become more and more negative as the distance increases.

Since we are not able to reach our target accuracy $\mathrm{Err}_{\mathrm{MAX}} \leq 0.1$ dB over all channels and distances with a single ANN operating over the total span, two approaches are introduced: *spacial slicing*, consisting in splitting the span into sub–span sections, and consider *multi-power levels*, i.e. increase of the power levels granularity for sub–bands. This corresponds to train the ANN with the other two types of data–sets described in the previous section (*3-power-levels* data–set and *5-power-levels* data–set). Based on the errors distributions observed in Figs. 3.4(a) and 3.4(b), we identify four sub–span sections: 1–10 km, 11–20 km, 21–50 km and 51–100 km. In the first



(a) *1 km to 30 km* (b) *1 km to L_{fiber} =[50,100] km*

Fig. 3.6 Probability density functions of RMSE and Err_{MAX} of power profiles determined using α_0, α_1 and σ coefficients predicted by ANN Model 2 considering two distance ranges: (a) 1 km to 30 km and (b) 1 km to $L_{fiber} = [50, 100]$ km.

kilometers intervals are shorter because higher power levels produce a stronger ISRS effect.

The training is applied independently on each sub–span section and on all the three power levels granularity data–sets. ANN hyper–parameters settings are reported in Table 3.1.

The results applying both strategies are shown in Fig. 3.5, where we plot the pdfs of the Err_{MAX}. From there we can see that for sections 1–10 km (Fig. 3.5(a)) and 11–20 km (Fig. 3.5(b)) already training with the *2-power-levels* data–set we are able to reach the target accuracy of $\text{Err}_{\text{MAX}} \leq 0.1$ dB, with quite low average values, respectively 0.062 dB and 0.065 dB. Increasing the power levels granularity to 3 and 5, even more accurate predictions are achieved, as demonstrated by the reduced average values and maximum values reached by the pdfs tails. Instead, for sub–section 21–50 km (Fig. 3.3(c)), training with the *2-power-levels* data–set is not enough to have all Err_{MAX} below or equal 0.1 dB, which instead reaches values up to 0.26 dB. Only considering the *3-power-levels* and the *5-power-levels* data–sets for training, the target accuracy is guaranteed. Lastly, from Fig. 3.3(d) related to section 51–100 km, $\text{Err}_{\text{MAX}} \leq 0.1$ dB for all cases is obtained only when training the ANN with the *5-power-levels* data–set. In fact, with *2-power-levels* the Err_{MAX} are up to 0.34 dB, instead with *3-power-levels*, although for most of the cases $\text{Err}_{\text{MAX}} \leq 0.1$ dB, the tail of the pdf extends up to 1.3 dB.

The sub–span sections considered here represent just one of the possible solutions and certainly, other sub–span sections, different in numbers and ranges, can be selected. Then, from the results obtained up to this point, we can conclude that, with the proper trade–off between spacial slicing and multi–power levels granularity, and the proper ANN hyper–parameters optimization, we can potentially achieve any target accuracy for predictions.

Moving to ANN Model 2, a single ANN consisting of 2 HLs and 20 HNs is trained using only the *2-power-level* data–set, back–propagation (BP) as training method, which implements Levenberg-Marquardt (LM) algorithm, one of the fastest in terms of convergence time and most suitable for regression problems [103]. Based on the outcomes of previous investigations, we consider the logistic sigmoid (logsig) as activation function.

For ANN testing, with results shown in Fig. 3.6, coefficients α_0, α_1 and σ are predicted only for distances in the range $L_{\text{fiber}} = [50, 100]$ km, which are then used in Eq. (3.11) to determine the entire evolution of power profiles from the beginning of the fiber up to any distance L. The ANN prediction performance is assessed by computing the RMSE and the Err_{MAX} as in Eqs. (3.12) and (3.13), where P^{pred} and P^{targ} are now the power profiles given by inserting in Eq. (3.11) the triplet $\{\alpha_0, \alpha_1, \sigma\}$ predicted by the ANN model 2 or obtained from fitting, respectively. Since the NLI generation is not the same over the fiber span, but it is mostly generated when the power is higher, prediction errors are analyzed over two ranges of distances: from 1 km to 30 km, where actually most of the NLI is generated (Fig. 3.6(a)), and from 1 km to $L_{\text{fiber}} = [50, 100]$ km (Fig. 3.6(b)). For both cases, high prediction accuracy is reached as the RMSE and Err_{MAX} are always under 0.03 dB and 0.1 dB, respectively. Especially, errors in the first 30 km are even smaller, with Err_{MAX} never exceeding 0.05 dB.

3.5 Conclusions

This Chapter proposes the use of two ANN models, Model 1 and Model 2, to determine the evolution of spectrally and spacially resolved power profiles in presence of ISRS in an accurate and faster way than standard approaches requiring the solution of sets of ODEs. Morever, Model 2 aims at substituting, with an advantage in the computational time, the fitting procedure required to determine the coefficients used

in the CFMs [88]. Indeed, the accurate estimation of power profiles evolution along the fiber is fundamental in GN and EGN models to evaluate linear and nonlinear impairments needed in the network performance assessment. The complexity and computational time demand of the method is due to the data–sets collection and the ANNs training, except in case of RP, for which even the training stage is ultra-fast [54]. However, since data collection and ANN training are performed offline, they do not affect the computational speed when the ANN is used in the field, since the predictions of power profiles are obtained instantaneously.

Considering both models, in particular applying some slicing strategies in case of Model 1, we have shown the ability of our ANN models to achieve highly accurate power profiles evaluations, with $\mathrm{Err}_{\mathrm{MAX}}$ always below 0.1 dB.

Chapter 4

Machine Learning for Raman Amplification Analysis

In this Chapter we applied ML and ANNs for an ultra-fast and accurate prediction of RAs gain and noise profiles as alternative approach to the numerical solution of the set of ODEs modeling the Raman amplification phenomenon described in Chpater 2. Particularly, this method is proposed in the context of multi–band systems and fast and flexible reconfigurable optical networks and dynamic partial traffic allocation, typically required by future optical networks. Most of the results reported in this Chapter have been also presented in [90–92].

4.1 Introduction

As discussed in Chapters 1 and 2, the ever increasing capacity demand is leading research towards the extension of the transmission bandwidth from the standard 5 THz C–band (1530 nm–1565 nm), to multi-band systems (O, E, S, C and L), with occupation up to more than 53 THz (1260 nm–1625 nm) [73].

As already mentioned previously, among the available and recently proposed amplification technologies, such as SOAs, doped fiber amplifiers and hybrid amplifier configurations, such as hybrid RA+EDFA, in this thesis we will consider only pure RAs. Our choice is motivated by the great advantages supplied by RAs, like their ability of providing flexible amplification gains (flat, tilted and arbitrary) potentially

over all optical bands by simply tuning the pump lasers configuration in terms of power and wavelength (frequency). In addition, distributed RAs are characterized by a lower noise figure than EDFAs, which makes them even more appealing [3]. Indeed, typical noise figure values for distributed RAs are between -2 dB and 0 dB, corresponding to around 6 dB of improvement with respect to EDFA [25].

In general, the set of ODEs modeling the physical behavior of RAs [14] are solved numerically. Because of the complexity of the model, the numerical solution of these equations can be a quite time consuming process. Especially in a multi-band scenario, for which the number of equations to be solved grows with the number of channels and RA pump lasers, with a consequently increase of the convergence time needed for an accurate solution.

The performance of optical networks is continuously monitored to provide a good quality of service, therefore real-time network reconfiguration is performed to better exploit network resources. This requires the knowledge of all network components, including the behavior of optical amplifiers.

To cope with the speed requirements of real-time optical networks, alternative approaches based on ML and ANNs have been proposed to predict RAs gain and ASE noise profiles in an ultra-fast and accurate way [55, 90].

In particular in this thesis, after an initial study of a fully-loaded system with Raman pump lasers varying both in wavelengths and powers, a more extensive analysis is performed considering partial spectral loads. This is a more realistic scenario, as the load of the network is time-variant, and the SRS effect depends on the spectral load at the input of the fiber [3]. Moreover, for the study with partial spectral loads, Raman pump lasers are assumed to be fixed in wavelength (frequency), as in reality they do not change significantly once deployed.

4.2 Raman Amplification Analysis in Full Spectral Load Condition

The system under study is shown in Fig. 4.1 and consists of a single span of SSMF with span length L_{span}=100 km and fiber parameters reported in Table 4.1. At the input of the fiber, we launch WDM comb of 343 channels loaded over the 11 THz

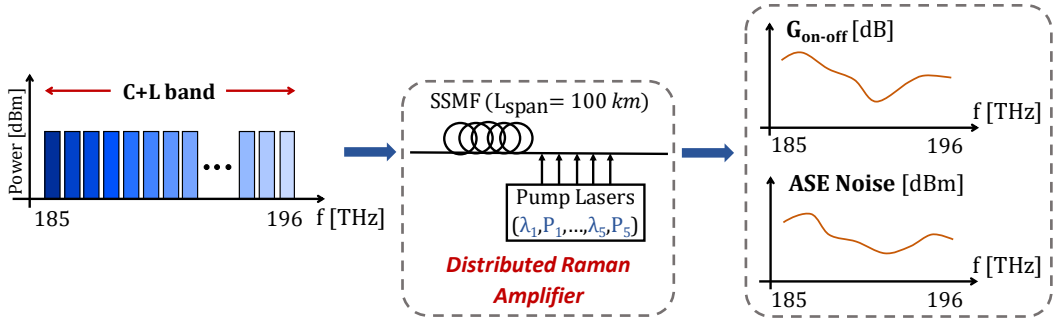


Fig. 4.1 Schematic of simulation setup.

α_s [dB/km]	α_p [dB/km]	A_{eff} [μm^2]	D [ps/nm/km]	γ [1/W/km]	C_r [1/W/km]
0.20	0.25	80	16.7	1.26	0.4125

Table 4.1 Parameters of SSMF where α_s and α_p are the attenuation coefficients respectively for the signal and the pumps, A_{eff} is the effective area, D is the chromatic dispersion, γ is the nonlinear coefficient and C_r is the Raman gain efficiency coefficient.

C+L–band, between 185 THz and 196 THz with a frequency resolution of 100 GHz, resulting in $N_f = 110$ points in frequency. Each channel is Nyquist shaped at 32 GBd symbol rate and carries 0 dBm of power. The distributed RA is made of five counter-propagating pump lasers with arbitrary wavelengths (frequencies) λ_i (f_i) and powers P_i, with $i = 1,\ldots,5$.

Although RA with co-propagating pumps has been demonstrated to have an even lower noise figure than RA with counter-propagating pumps [3], this latter case is preferable from a practical point of view. In fact, counter-propagating pumps prevent the system from damages due to high powers at the input of the fiber, since they are placed at the end where the power is already attenuated.

Then, at the output of the fiber we are interested in determining the Raman on-off gain and the ASE noise profiles. In simulations, these two profiles are evaluated numerically by means of a numerical Raman solver.

4.2.1 Machine Learning Framework and Data–sets Generation

As mentioned in the Introduction of this Chapter, given a set of pumps frequencies and powers, the goal is to determine the corresponding Raman on-off gain and ASE noise profiles in a fast way by means of ML and ANNs.

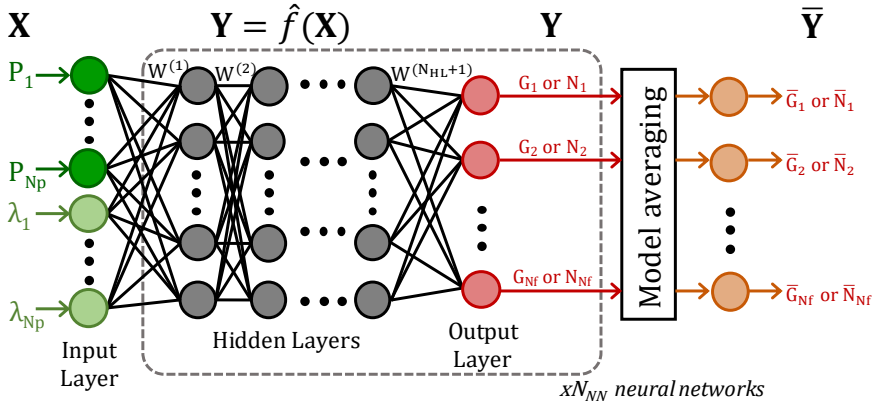


Fig. 4.2 General model of a multi-layer feed-forward artificial neural network for which model averaging is applied.

The problem under study is a simple regression problem that does not require complex ANN architectures. In Fig. 4.2, the general model of a multi-layer feed-forward neural network is shown. Gain and ASE noise profiles are predicted independently of each other. This *direct model* learns an approximation of the mapping function $\mathbf{Y} = f(\mathbf{X})$ describing the solution of the ODEs, where the input $\mathbf{X}$ corresponds to the RA pumps wavelengths and powers $\mathbf{X} = [\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, P_1, P_2, P_3, P_4, P_5]^{\mathrm{T}}$ and the output $\mathbf{Y}$ can be associated to the on-off gain profiles $\mathbf{G} = [\mathrm{G}_1, \ldots, \mathrm{G}_{\mathrm{N_f}}]^{\mathrm{T}}$ or the ASE noise profiles $\mathbf{N} = [\mathrm{N}_1, \ldots, \mathrm{N}_{\mathrm{N_f}}]^{\mathrm{T}}$. Each ANN is described by the matrices of weights $[\mathbf{W}^1, \ldots, \mathbf{W}^{\mathrm{HL}+1}]$, with HL being the number of hidden layers, connecting the input layer to the first hidden layer, hidden layers to hidden layers, and the last hidden layer to the output layer.

By means of an ODEs numerical solver, two different data–sets of $M = 5000$ elements are generated, one for ANN training and validation and the other for ANN testing. Pump wavelengths and powers are drawn from uniform distributions, whose range of values are shown in Table 4.2, which have been selected in order to obtain gain profiles in a range of practical interest. Maximum values of Raman on-off gain are between 13 dB (at higher frequencies) and 20 dB (at lower frequencies).

To train the ANNs, we consider two different algorithms: standard back-propagation (BP) [10], employing Matlab's Deep Learning ToolBox, and random projection (RP) [52, 54]. As described in Chapter 2, when BP is implemented, the weights of the ANN are iteratively optimized through a gradient-descent (GD) algorithm, such that the error between the target and the predicted values is minimized.

	Wavelength [nm]	Power [mW]
Pump 1	[1424,1436.2]	[0,160]
Pump 2	[1436.2,1458.4]	[0,160]
Pump 3	[1458.4,1480.6]	[0,160]
Pump 4	[1480.6,1502.8]	[0,160]
Pump 5	[1502.8,1525]	[0,160]

Table 4.2 Pump lasers wavelength and power ranges.

Since it is an iterative process, the training may require large computational time, in particular in case of more complex models. Instead, unlike BP, the RP algorithm does not keep updating the weights of the ANN on a GD based method. In fact, it consists on a random initialization of normally distributed weights for the hidden layers, which remain fixed during training. Just the weights of the last layer are updated in a single loop using least-square method, which guarantees the learning of the input-output mapping. In this case the computational effort is very limited and the training phase is completed almost in real-time.

To improve ANN prediction performance, hyper-parameters tuning can be performed. Instead, to improve generalization properties, as prediction accuracy may be affected by the random initialization of ANN weights, model combination is employed [10, 46]. This means that N_{ANN} different ANNs are trained and the predictions are given by *averaging* the predictions made by each single model.

In case of BP, the ANN model consists of 2 hidden layers, 10 hidden nodes and model averaging is performed over $N_{ANN} = 10$ independent ANNs. Also here, among the possible training algorithm available in the Matlab's Deep Learning ToolBox, we use the Levenberg-Marquardt (LM) algorithm. For model selection, we investigate logistic sigmoid and hyperbolic tangent activation functions.

When training using RP method, instead, we consider a single hidden layer feed-forward neural network (SLFN), meaning that the ANN model has only one hidden layer. We optimize the number of hidden nodes by sweeping them from 20 to 600 in steps of 20. The tested activation functions are hyperbolic tangent, logistic sigmoid and sine. The number of independent ANNs over which averaging is $N_{ANN} = 20$.

Table 4.3 sums up the hyper-parameters resulting from the optimization for both gain and ASE noise profiles predictions.

	BP		RP		
Activation function	tanh	logsig	sine	tanh	logsig
Number of hidden layers (N_{HL})	2	2	1	1	1
Number of hidden nodes (N_{HN})	20	20	400	120	240
Model combination parameter (N_{ANN})	10	10	20	20	20

Table 4.3 ANNs hyper-parameters considered for different training method for both gain and ASE noise profiles prediction.

For the ANN training, part of the training data–set is used to validate the network to prevent from *overfitting*. In case of BP, 70% of the data–set is used for training, 15% for validation and 15% for testing, while in case of RP K–fold cross-validation is performed considering $K = 10$, so the training–validation ratio is 90:10. Once the ANN has been trained, it can be used to make predictions over unseen data.

4.2.2 Testing Results and Accuracy Analysis

We test the trained ANNs over the testing data–set, which has been pruned selecting only the cases with gains in the range [4,12] dB. Gains below 4 dB are too low to implement an RA, while 12 dB corresponds to the threshold under which the RA operates in the moderate pumping regime (MPR) (60% of total span loss) to avoid saturation effects [26].

For each target profile Y^{targ} (gain or ASE noise), the accuracy of the predicted Y^{pred} is evaluated over the entire C+L–band in terms of root-mean-square-error (RMSE) and maximum error ($\mathrm{Err}_{\mathrm{MAX}}$), defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_{\mathrm{f}}} \sum_{n=1}^{N_{\mathrm{f}}} (Y_n^{pred} - Y_n^{targ})^2} \tag{4.1}$$

$$\mathrm{Err}_{\mathrm{MAX}} = \max_{n \in \{1,\ldots,N_{\mathrm{f}}\}} |Y_n^{pred} - Y_n^{targ}|. \tag{4.2}$$

Figs. 4.3 and 4.4 show the probability density functions (pdfs) and the cumulative distribution functions (cdfs), respectively, of RMSE and $\mathrm{Err}_{\mathrm{MAX}}$ of gain profiles prediction over the pruned testing data–set for the different combination of training algorithms and activation functions in Table 4.3.

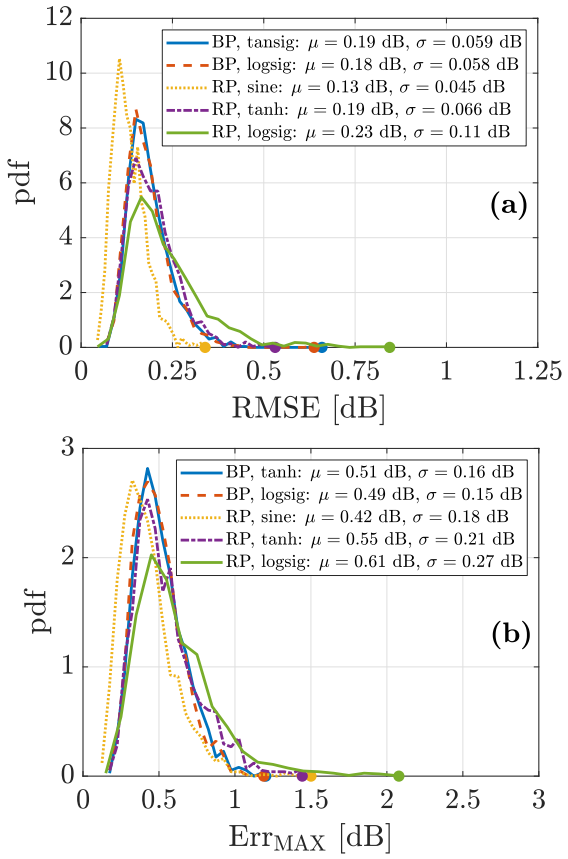


Fig. 4.3 Probability density functions of Raman on-off gain profiles prediction errors over testing data–set evaluated in terms of (a) RMSE and (b) Err_{MAX}.

Analyzing Fig. 4.3, we can observe that in case of BP, performance is similar regardless the activation function, for both RMSE (Fig. 4.3(a)) and Err_{MAX} (Fig. 4.3(b)). In case of hyperbolic tangent activation function, the RMSE has 0.19 dB average value and 0.059 dB standard deviation, with values up to 0.66 dB. Instead, in case of logistic sigmoid, the average value and the standard deviation of the RMSE are 0.18 dB and 0.058 dB, respectively, with values up to 0.64 dB.

Also if we compare them in terms of Err_{MAX}, the performance is almost the same. Indeed, the average and the standard deviation are respectively 0.51 dB and 0.16 dB for *tanh*, and 0.49 dB and 0.15 dB for *logsig*, while the maximum values of Err_{MAX} are 0.20 dB in case of *tanh* and 0.19 dB in case of *logsig*. On the contrary, when we consider RP, we can see that prediction performance slightly changes with the activation function. The steepest pdf corresponds to *sine*, for which we get the lowest average values of RMSE and Err_{MAX}, being 0.13 dB and 0.42 dB,

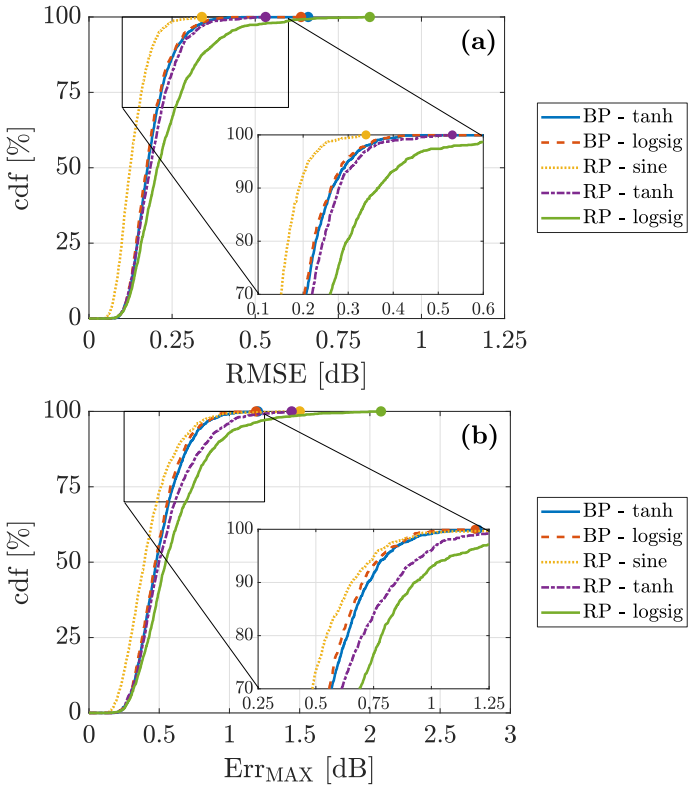


Fig. 4.4 Probability density functions of Raman on-off gain profiles prediction errors over testing data–set evaluated in terms of (a) RMSE and (b) Err_{MAX}.

respectively. The standard deviations are 0.045 dB and 0.18 dB, respectively. By comparing results with BP and with RP, the lowest RMSE is given when we consider RP with *sine* activation function, similarly for the Err_{MAX}, but with shorter pdf tails in case of BP (1.2 dB against 1.5 dB).

To better analyze the distribution of the prediction errors, we can look at the cdfs in Figs. 4.4(a) and 4.4(b) for RMSE and Err_{MAX}, respectively, and at their insets. From Fig. 4.4(a), we can see that the steepest cdf corresponds to RP with *sine* activation function, as can be also inferred from its pdf, while the less steep corresponds to RP with *logsig*. In particular, for all the training algorithm and activation functions, RMSE is always below 0.3 dB for more than 80% of cases, and for more than 97% of cases the RMSE is always below 0.5 dB.

If we analyze the cdfs of Err_{MAX} (Fig. 4.4(b)), only for RP with *logsig* activation function the percentage of cases that have Err_{MAX} below 0.5 dB is less than 50%, while for the others the cases are more than 50%, especially for the RP with *sine*

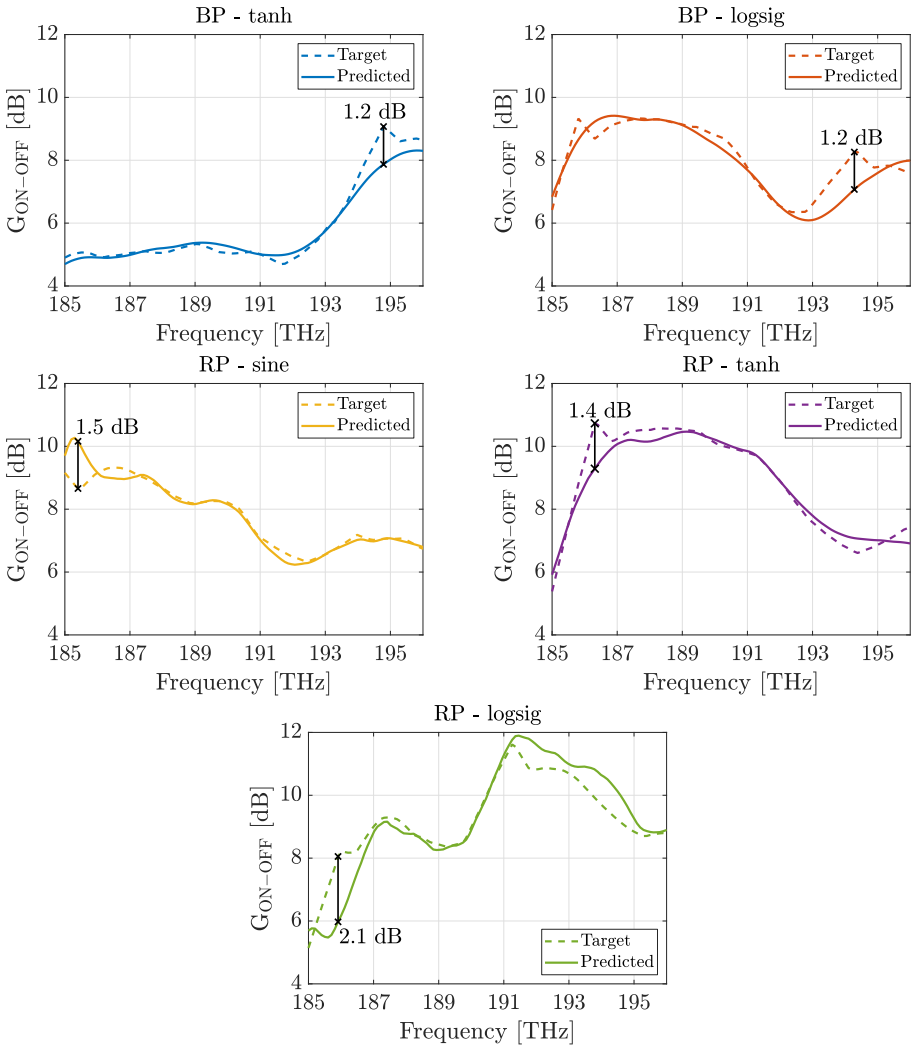


Fig. 4.5 Target and predicted Raman on-off gain profiles in case of largest maximum absolute error Err_{MAX} for training algorithms and activation functions in Table 4.3.

that reaches 75%. Moreover, all the cdfs show that for more than 90% of cases the Err_{MAX} is always smaller than 1 dB.

In Fig. 4.5 we plot the target and predicted gain profiles for the different training algorithms and activation functions of Table 4.3 for the worst case prediction, i.e. when the largest maximum absolute error Err_{MAX} is experienced. These values are not negligible, as they exceed 1 dB, but we can see that the impact of prediction errors is limited to small frequency regions, except for RP with *logsig*.

Similar considerations can be drawn in case of ASE noise profiles prediction, with pdfs and cdfs of RMSE and Err_{MAX} plotted in Figs. 4.6 and 4.7, respectively. Also

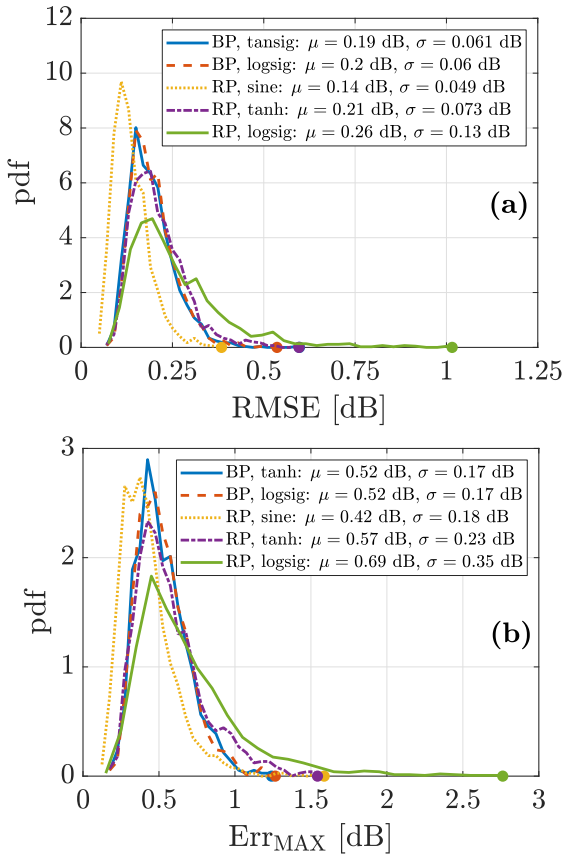


Fig. 4.6 Probability density functions of ASE noise profiles prediction errors over testing data–set evaluated in terms of (a) RMSE and (b) $\mathrm{Err_{MAX}}$.

for ASE noise prediction, on average, the lowest RMSE and $\mathrm{Err_{MAX}}$ are obtained when we consider RP with *sine* activation function, with values 0.14 ± 0.049 dB and 0.42 ± 0.18 dB, respectively. Accordingly, the associated cdfs are the steepest among the others.

The worst case predictions for the different training algorithms and activation functions are shown in Fig. 4.8, where, as already observed in case of gain profiles prediction, the large values of $\mathrm{Err_{MAX}}$ impact only a small portion of frequencies, with the exception of RP with *logsig*, for which the difference is not negligible for frequencies between 185 THz and 189 THz.

These results show that, with a suitable and properly optimized ANN model, it is possible to predict both RA gain and ASE noise with good accuracy. In fact, with a maximum absolute error $\mathrm{Err_{MAX}}$ lower than 0.5 dB in case of RP and *sine* activation function for both gain and noise profiles prediction, the proposed approach

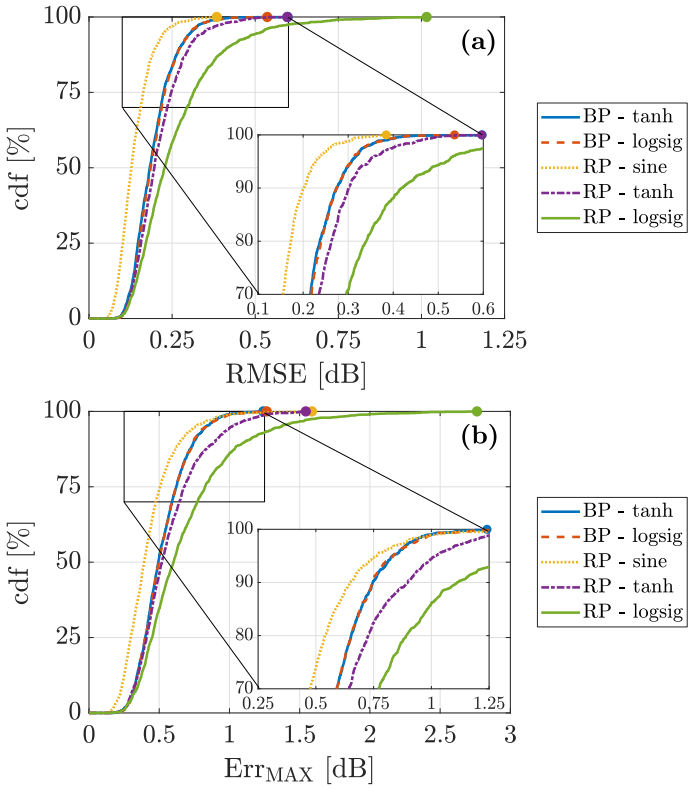


Fig. 4.7 Cumulative distribution functions of ASE noise profiles prediction errors over testing data–set evaluated in terms of (a) RMSE and (b) Err_{MAX}.

is a potential solution for integrated network controllers in next generation optical networks.

In this analysis, the two profile types are predicted using a dedicated ANN, but this does not exclude the possibility of building an ANN model able to predict gain and ASE noise simultaneously. In addition, the presented principles, here applied specifically over a C+L–band systems with five counter–propagating pumps and a 100 km single span of SSMF, can be potentially scaled up to a higher number of pumps to support larger bands and extended to other fiber types and span lengths. Although the ANNs complexity would scale linearly with the bandwidth, we expect the predictions to keep the same level of accuracy, also performing a further ANN hyper-parameters optimization.

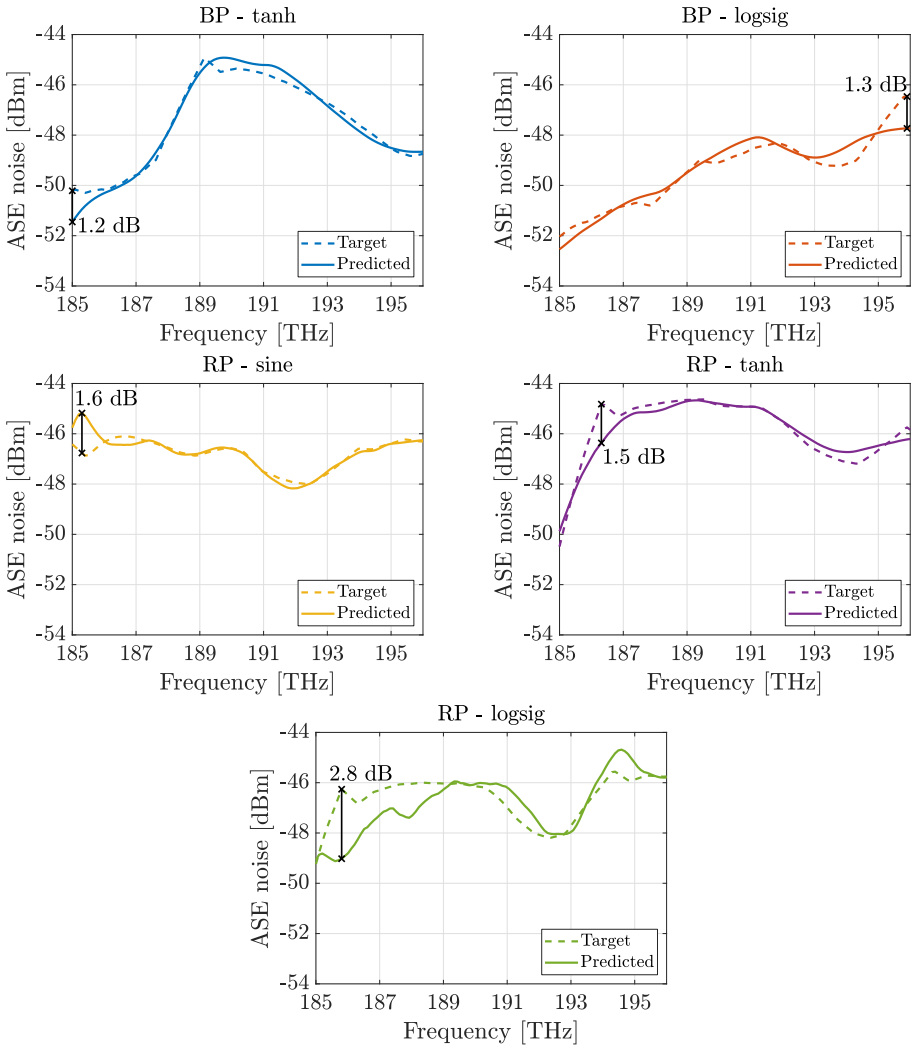


Fig. 4.8 Target and predicted ASE noise profiles in case of largest maximum absolute error $\mathrm{Err}_{\mathrm{MAX}}$ for training algorithms and activation functions of Table 4.3.

4.3 Raman Amplification Analysis in Partial Spectral Loads Condition

The results presented in the previous section, as well as other papers on RA design and analysis [17, 55, 90, 113, 116, 117], assume full spectral load condition, which is a limited scenario since in real networks traffic patterns can change over time and seldom reach the full spectral load. For this reason, in this section we extend the analysis to partial spectral loads, further showing the potential application of the

proposed method in real time network controllers requiring a fast tool to describe the RA behavior.

Like in Chapter 3, we consider an 11 THz C+L–band system between 185 THz and 196 THz with $N_{ch} = 220$ channel slots of 50 GHz and to emulate partial spectral loads, each channel can be turned ON or OFF. Also here, to reduce the dimension of the space to be visited for the data–set generation (2^{220}), we consider groups of ten adjacent channels, for a total of $N_{sb} = 22$ sub-bands of 500 GHz. Hence, the input load of the RA is generated on a sub-band basis.

4.3.1 Data–sets Generation and Machine Learning Framework

Data–sets generation

To generate the synthetic data–sets needed for ANNs training, validation and testing, we use the numerical Raman solver included in the GNPy open source library [45], which solves the ODEs equations describing the Raman effect [14]. The system under study is shown in Fig. 4.9 and it is composed of a distributed Raman amplifier over a single span of SSMF of length $L_{span} = 100$ km with $N_p = 5$ counter-propagating pumps. Attenuation coefficients of the fiber are $\alpha_s = 0.21$ dB/km, for the input signal, and $\alpha_p = 0.25$ dB/km, for the pumps. As a proof of principle and to keep the problem simple, the analysis is performed just over a single span of fiber, but it can be extended to more spans and to hybrid amplification schemes (hybrid EDFA+RA).

Unlike section 4.2, to mimic a more realistic scenario, the RA pump lasers are kept fixed in frequency, as in general they do not change significantly once the RA has been deployed. Instead, the pumps powers P_i, with $i \in \{1,\dots,5\}$, are drawn from a uniform distribution between 0 mW and 250 mW. Their values are shown in Table 4.4. At the input of the fiber, a WDM comb is considered over the C+L–band

	Frequency [THz]	Power [mW]
Pump 1	210.37	[0,250]
Pump 2	207.14	[0,250]
Pump 3	204.01	[0,250]
Pump 4	200.97	[0,250]
Pump 5	198.03	[0,250]

Table 4.4 Pump lasers frequencies and powers.

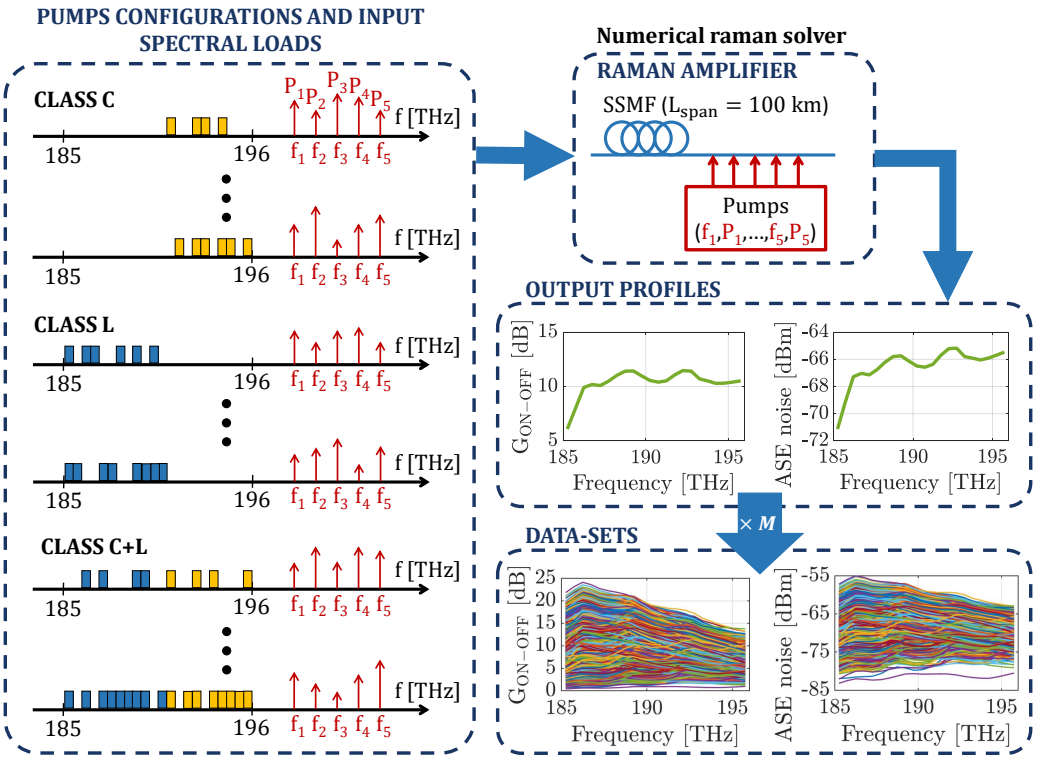


Fig. 4.9 Simulation setup of the optical communication systems under analysis.

from 185 THz to 196 THz, for a total of 11 THz, and it is made of $N_{ch} = 220$ Nyquist-shaped channels in the 50 GHz grid carrying 1 mW (0 dBm) of power.

Operating on a sub–band basis, the partial loads are generated playing on the sub-bands, with each of them having the same probability of being ON or OFF. When a sub-band is ON it carries 10 mW power, corresponding to having all the 10 channels ON, when it is OFF, no power is carried. To have a wide coverage of all the possible cases, we generate diversified types of partial spectral loads depending on the portion of the bandwidth occupied by the sub–bands. Depending on whether only C–band, only L–band or the entire C+L–band is considered, we identify three classes: *class C*, *class L* and *class C+L*. To reduce the number of cases, for each class, only even numbers of sub–bands ON are drawn with randomly selected frequency position. Within each class, cases with the same fixed number of sub–bands ON identify a sub–class. For *class C* the number of sub–bands ON, and thus the sub–classes, are from 2 to 10, for *class L* are from 2 to 12, and for *class C+L* are from 2 to 22, which corresponds to the full spectral load.

In our analysis, two types of data–sets are generated depending on whether *full–loads* or *partial–loads* are considered. Moreover, for each data–set type, we build two completely uncorrelated data–sets, one for ANN training and validation, and the other to test the ANN over unseen data. For the *full–load* (FL) data–sets, M=5000 different samples are extracted with full spectral load and random pump powers with values drawn from the uniform distribution $P_i \sim U[0,250]$ mW, with $i = 1,\ldots,5$ (Table 4.4). Instead, the *partial–load* (PL) data–sets consist of the M=11000 samples obtained by merging together the 500-samples data–sets generated for each class and sub–class. Also in this case, pumps frequencies are fixed and the randomness is on the pumps powers, which are drawn from uniform distribution $P_i \sim U[0,250]$ mW, with $i = 1,\ldots,5$.

Impact of Partial Spectral Loads on Raman on-off gain and ASE noise profiles

Since for a given configuration of pumps powers and frequencies, the RA gain, as well as the ASE noise, changes with varying input spectral loads, to further support the introduction of an ANN considering also the spectral load at its input, especially in a partial–load condition, here we report two practical examples comparing full–load and partial–load scenarios. We consider the same simulation setup of Fig.4.9, with five counter–propagating pumps with fixed frequencies and powers. The pumps frequencies are those of Table 4.4, i.e. $f_{i=1,\ldots,5}$=[210.37, 207.14, 204.01, 200.97, 198.03] THz. Also in this case we slice the C+L–band into 22 sub–bands, applying the classification into sub–classes based on the number of sub–bands ON. For each sub–class, 100 different partial spectral loads are extracted with randomly selected position of sub–bands ON and the corresponding RA gain and ASE noise profiles are numerically determined by means of the numerical Raman solver [45]. As reference, gain and noise profiles are evaluated in the full–load case as well.

In the first example, the pumps powers are set to $P_{i=1,\ldots,5}$=[246.7, 237.7, 194.2, 192.7, 168.8] mW. Fig. 4.10 shows the on-off gain profiles, classified by sub–classes, compared with the profile generated in full–load condition (black solid line). For simplicity, only a limited set of sub–classes is reported. In each subplot, circles of same color identify the profile obtained for a given input spectral load condition. In this case, due to pumps powers settings, gain profiles show a strong negative tilt, with several dBs of difference between higher and lower frequencies.

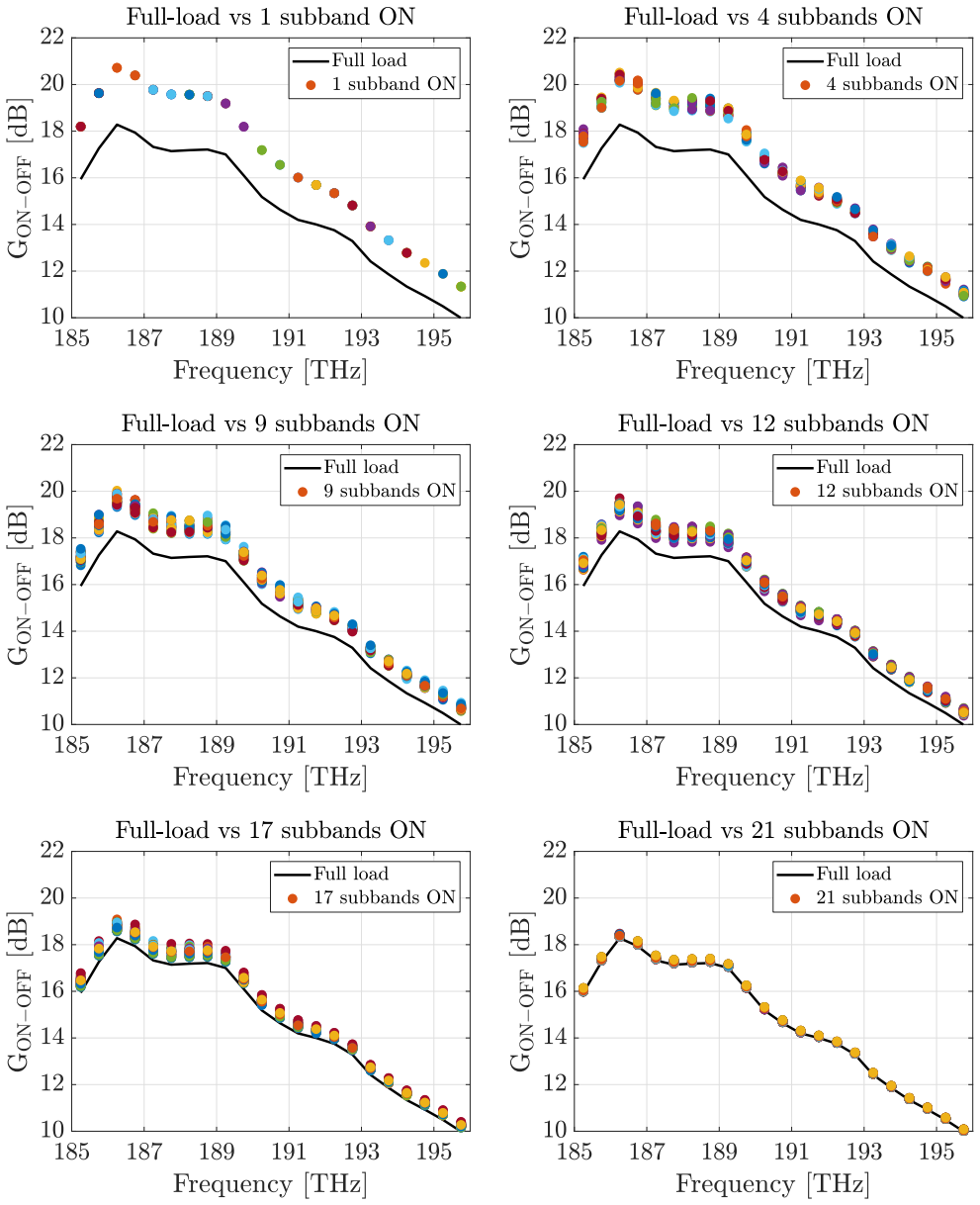


Fig. 4.10 Impact of partial spectral loads with respect to full spectral load on Raman on-off gain profiles assuming fixed pump powers and frequencies when a strong tilt is present.

Analyzing the gains for different sub–classes, we can observe that, when the number of sub–bands ON decreases, the difference with respect to the full–load increases, with variations even greater than 2 dB. Indeed, the few channels ON mostly benefit from the power transferred by the pump lasers, experiencing higher gains.

In addition, within a sub–class there are also variations between gain profiles with different input spectral load conditions, which is explained by the different pump-signal and signal-signal interactions. The greatest variability is observed for partial–loads close to the half–load, i.e. for sub–classes 10 to 16.

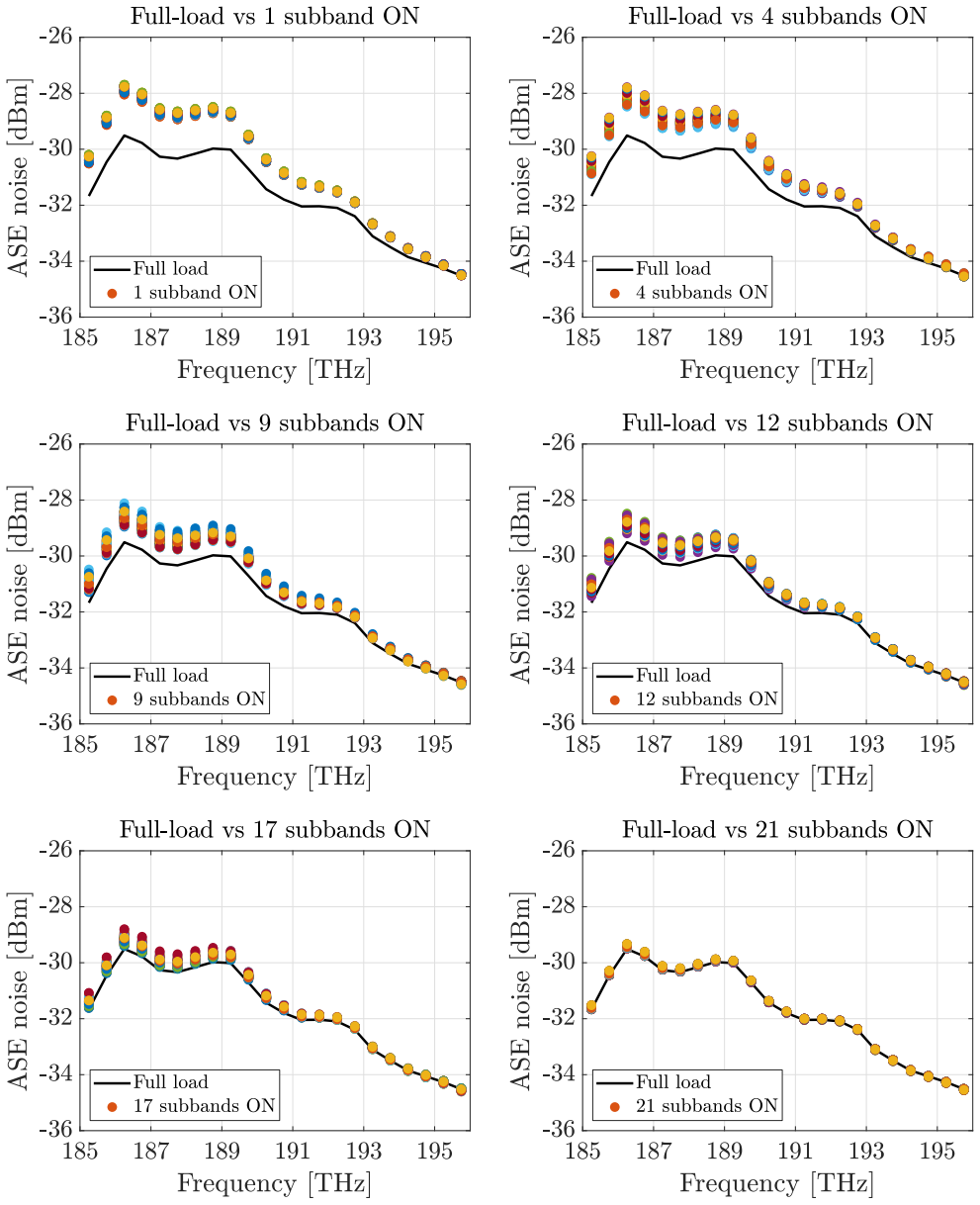


Fig. 4.11 Impact of partial spectral loads with respect to full spectral load on ASE noise profiles assuming fixed pump powers and frequencies when a strong tilt is present.

In Fig. 4.11, the ASE noise profiles corresponding to the gains of Fig.4.10 are illustrated. Also in this case, similar conclusions can be drawn, except that the variation of partial–loads with respect to the full–load is reduced, meaning that, compared to gain profiles, the ASE noise profiles have a reduced dependency on the input spectral load conditions.

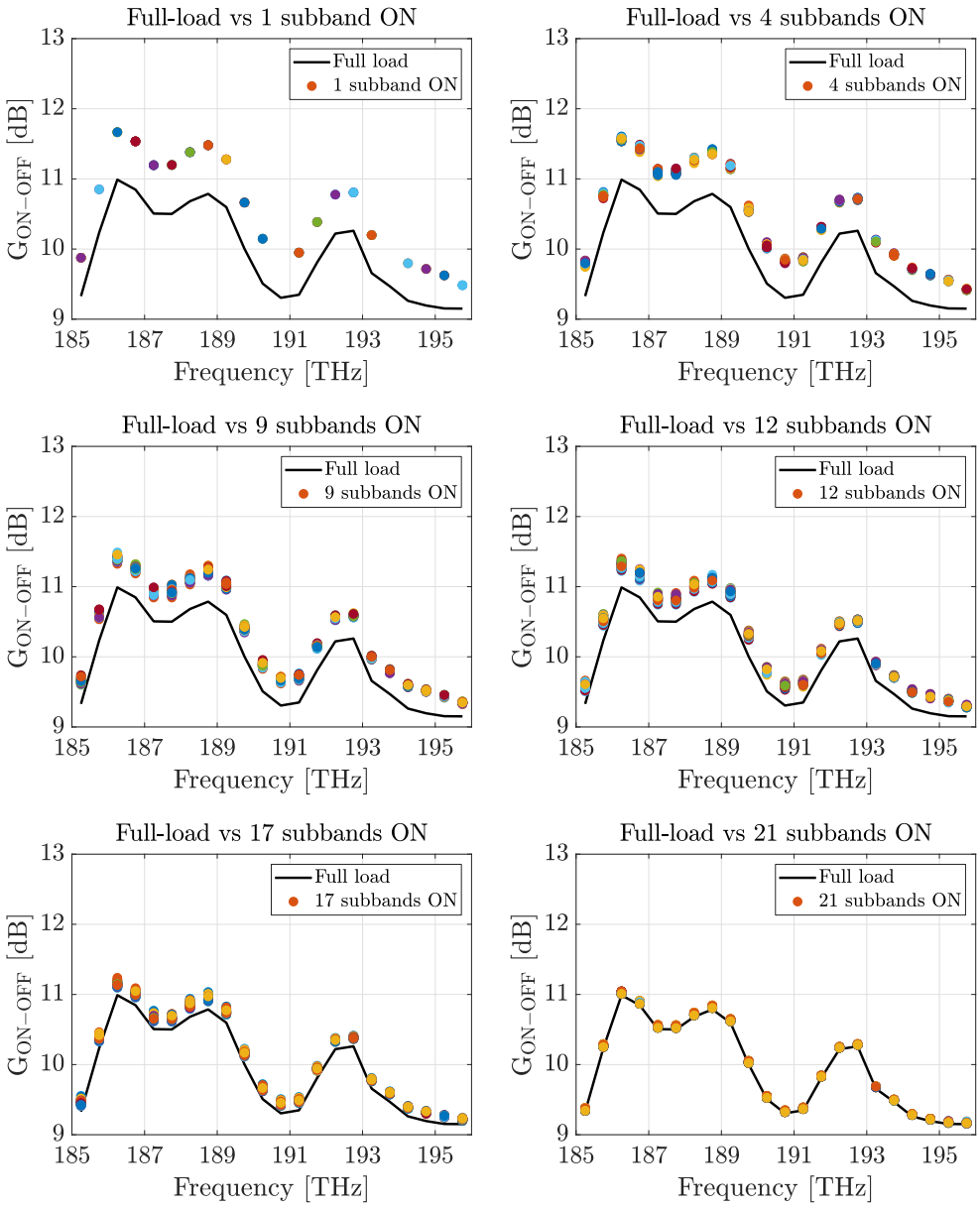


Fig. 4.12 Impact of partial spectral loads with respect to full spectral load on Raman on-off gain profiles assuming fixed pump powers and frequencies in case of almost flat profile.

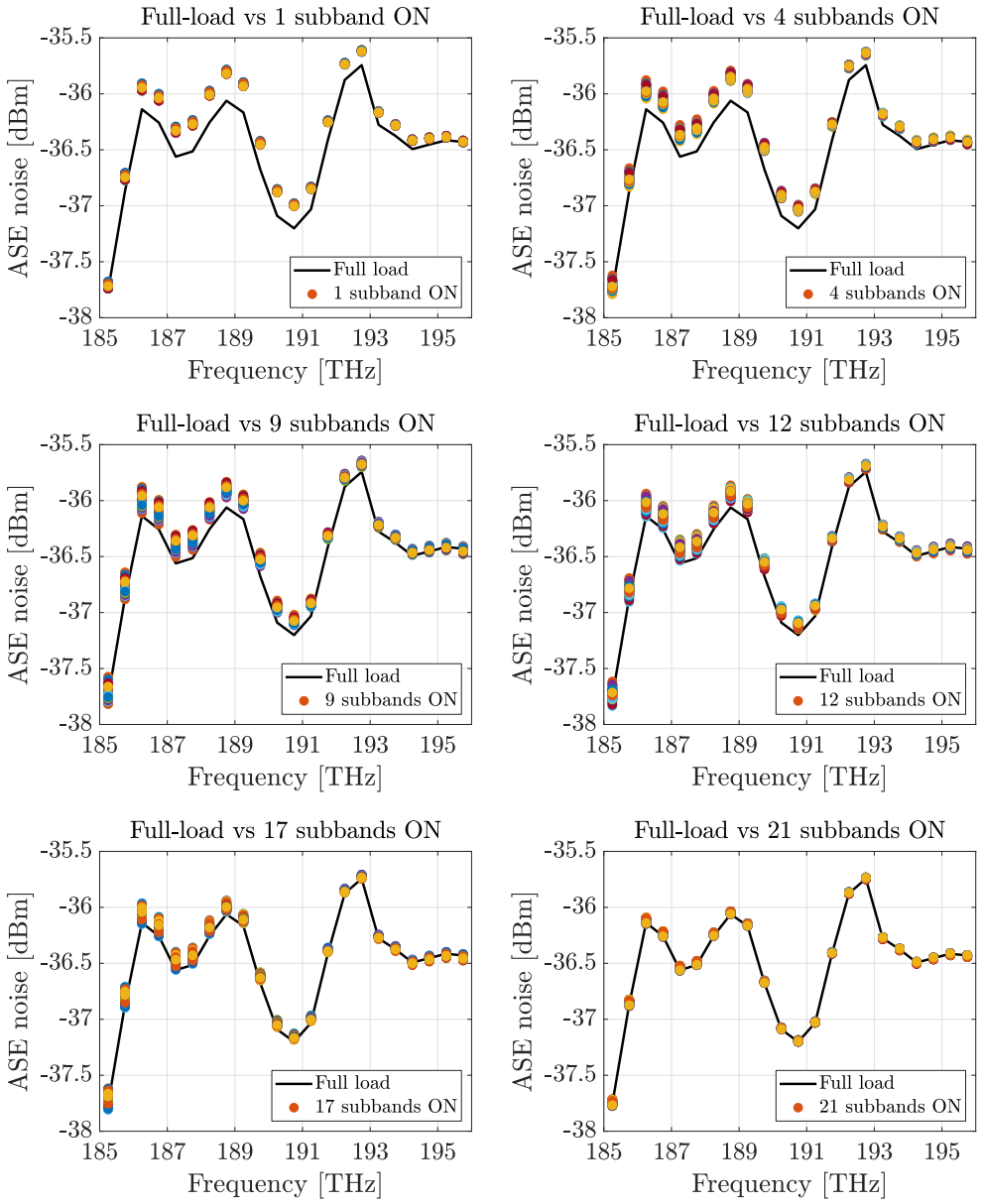
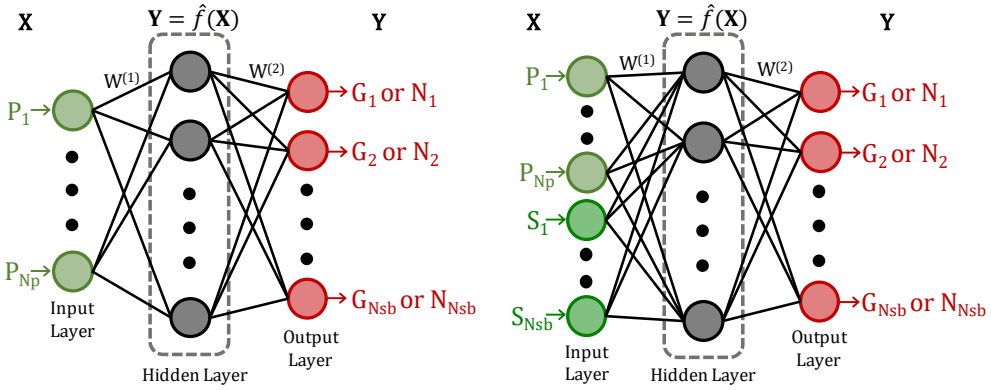


Fig. 4.13 Impact of partial spectral loads with respect to full spectral load on ASE noise on-off gain profiles assuming fixed pump powers and frequencies in case of almost flat profile.

In the second example, we examine a gain profile almost flat around 10 dB, which is relevant from the network point of view, since same amplification gain is desired over the entire transmission bandwidth. In this case, the pumps powers are $P_{i=1,\ldots,5}$=[169.5, 232.3, 80.1, 107.1, 93.7] mW and, like in the previous example, 100 different partial loads are extracted with randomly selected position of sub–bands



(a) *LU-NN* (b) *LA-NN*

Fig. 4.14 Artificial neural network models for RA gain or ASE noise profiles prediction (a) not considering and (b) considering spectral loads at the input.

ON. The resulting RA gains and ASE noise profiles are depicted in Figs. 4.12 and 4.13, respectively, and are classified by sub–classes. Although the variations of the partial–load cases with respect to the full–load are reduced compared to the first example, these variations are still non-negligible, especially when few sub–bands are ON.

Despite the analyzed examples refer to specific configurations of pumps with a limited number of spectral loads, they allow an easy and intuitive understanding of the physical phenomenon that motivates our study on partial spectral loads and the introduction of a dedicated ANN to handle different spectral conditions.

Machine Learning Framework and Artificial Neural Network Models

Given the goal of determining the function $\hat{f}(\cdot)$ that approximates the *direct model* $\mathbf{Y} = f(\mathbf{X})$ describing the RA behavior and relating output $\mathbf{Y}$ (gain or noise profile) to input $\mathbf{X}$ (laser pumps powers or laser pumps powers+partial spectral loads), we consider two ANN models (Fig. 4.14), which differ for the parameters associated to the vector $\mathbf{X}$ at the input of the ANN. Fig. 4.14(a) shows the Load-Unaware neural network (LU-NN) whose input $\mathbf{X}$ is represented by the pump powers $\mathbf{P} = [P_1, \ldots, P_{\mathrm{Np}}]^{\mathrm{T}}$ only. Whereas, in Fig. 4.14(b) the Load-Aware neural network (LA-NN) is presented, which takes as input $\mathbf{X}$ both pumps powers $\mathbf{P} = [P_1, \ldots, P_{\mathrm{Np}}]^{\mathrm{T}}$ and spectral load $\mathbf{S} = [S_1, \ldots, S_{\mathrm{Nsb}}]^{\mathrm{T}}$. For both ANN models, the output $\mathbf{Y}$ can be the Raman on-off gain $\mathbf{G} = [G_1, \ldots, G_{\mathrm{Nsb}}]^{\mathrm{T}}$ or the ASE noise profile $\mathbf{N} = [N_1, \ldots, N_{\mathrm{Nsb}}]^{\mathrm{T}}$.

Based on the conclusions drawn in the previous section, since BP and RP provide similar results in terms of prediction accuracy, with RP being a faster algorithm than BP for ANN training, all the analysis is performed considering RP only. To guarantee high accuracy, two key parameters must be properly initialized: the standard deviation σ_{ANN} for the initialization of ANN weights and the regularization parameter λ_{ANN}. During the training phase, the 90% of the training data–set is used to train the ANN while the remaining 10%, called validation data–set, is used for ANN validation, which is fundamental to monitor any possible underfitting or overfitting of the ANN. The testing data–set is entirely used to test the ANN over unseen data.

The goodness of predictions is evaluated by computing the root-mean-square-error (RMSE) and the maximum absolute error ($\mathrm{Err}_{\mathrm{MAX}}$) between the predicted profile Y^{pred} and the target profile Y^{targ} defined in Eqs. (4.1) and (4.2).

With the goal of designing an ANN able to give fast and highly accurate predictions of RA gain and ASE noise profiles for any load condition, the LU-NN is trained and tested over both full and partial spectral loads. Being the LU-NN load–agnostic, to assess its ability to handle partial spectral loads as well, different combinations of training and testing data–sets are analyzed. Instead, the LA-NN is trained and tested over *partial–load* data–sets only. Thus, four different scenarios are identified, which are summarized in Tables 4.5 and 4.6.

As can be seen in Fig. 4.14, the ANN models present only one hidden layer ($N_{\mathrm{HL}} = 1$) [54], while an extensive hyper–parameters optimization is performed over the number of hidden nodes N_{HN} and the activation function. The number of hidden nodes is swept from 20 up to 2000 in step of 20, and the considered activation functions are *sine*, *tanh* and *logsig*. To avoid any possible impact of weights initialization, model averaging is carried out over $N_{\mathrm{ANN}} = 20$ parallel and independent ANNs. The optimal model is selected based on the lowest RMSE computed over the validation data–set.

When the LU-NN is trained with the full–load data–set (scenarios 1 and 2), the optimal number of hidden nodes is 1980 for the RA gain and 1900 for the ASE noise profiles predictions. Instead, among the tested activation functions, the smallest RMSE is given when the hyperbolic tangent is employed, for both gain and noise profiles. Afterwards, we optimize the LU-NN once again when it is trained with the partial–load data–set (scenario 3).

	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 4**
ANN model	LU-NN	LU-NN	LU-NN	LA-NN
Training data–set type	Full-load	Full-load	Partial–load	Partial–load
Testing data–set type	Full-load	Partial-load	Partial–load	Partial–load
Activation function	tanh	tanh	logsig	tanh
Number of hidden layers N_{HL}	1	1	1	1
Number of hidden nodes N_{HN}	1980	1980	640	1980
Model combination parameter N_{ANN}	20	20	20	20

Table 4.5 Scenarios considered for LU-NN and LA-NN training, validation and testing for RA gain predictions. The optimal hyper-parameters settings are reported.

	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 4**
ANN model	LU-NN	LU-NN	LU-NN	LA-NN
Training data–set type	Full-load	Full-load	Partial–load	Partial–load
Testing data–set type	Full-load	Partial-load	Partial–load	Partial–load
Activation function	tanh	tanh	logsig	tanh
Number of hidden layers N_{HL}	1	1	1	1
Number of hidden nodes N_{HN}	1900	1900	1920	1980
Model combination parameter N_{ANN}	20	20	20	20

Table 4.6 Scenarios considered for LU-NN and LA-NN training, validation and testing for ASE noise predictions. The optimal hyper-parameters settings are reported.

For Raman gain profiles, the optimization results in 640 hidden nodes, while for ASE noise profiles, it results in 1920 hidden nodes. For both cases, the logistic sigmoid results to be the optimal activation function.

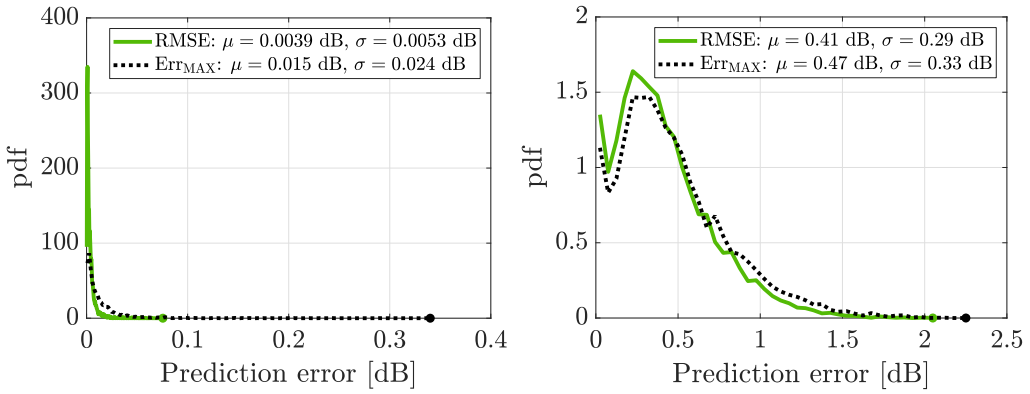
Eventually, for the LA-NN, introduced to handle partial-loads (scenario 4), the hyper–parameters optimization returns the same optimal number of HNs, 1980, and activation function, hyperbolic tangent, for both RA gain and ASE noise profiles predictions. A summary of the hyper-parameters chosen for each scenario for both gain and ASE noise is reported in Tables 4.5 and 4.6, respectively.

4.3.2 Artificial Neural Network Testing and Prediction Performance

For each scenario of Tables 4.5 and 4.6, once the approximating function $\hat{f}(\cdot)$ has been determined by training the ANN, it can be used to predict the RA gain or ASE noise profiles of the testing data–sets (full–load and partial–load data–sets). In particular, we prune the testing data–sets discarding all the cases with values of gain below 4 dB, which is too low to use in practice, and those exceeding 21 dB, since they would over-compensate the total loss of the fiber span.

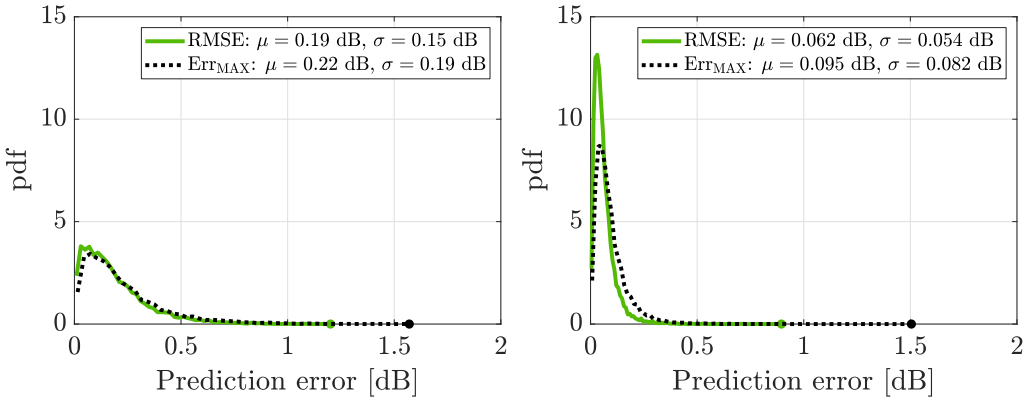
The accuracy of the predictions is evaluated in terms of RMSE and $\mathrm{Err}_{\mathrm{MAX}}$ computed between the predicted profile and the target one. By computing these errors for all the predicted profiles of the testing data–set, we obtain a statistical assessment of the performance of the ANN. For a more detailed analysis, in Figs. 4.15–4.18 we plot the probability density functions and the cumulative distribution functions of RMSE and $\mathrm{Err}_{\mathrm{MAX}}$ of RA gain and ASE noise profiles predictions for the four different scenarios of Tables 4.5 and 4.6.

Fig. 4.15(a), presents the pdfs of the RMSE and $\mathrm{Err}_{\mathrm{MAX}}$ computed for the predictions of the RA gain profiles of the testing data–set for scenario 1, i.e. when the LU–NN is trained and tested with full–load data–sets. In this case, with the training and testing data–sets generated with the same load condition, the ANN properly learns the mapping between the input pumps powers and the output gains, achieving very low errors and high accuracy over the predictions of the tested profiles. The average values of the RMSE and $\mathrm{Err}_{\mathrm{MAX}}$ are respectively 0.0039 dB and 0.015 dB, with very low standard deviations as well, corresponding to 0.0053 dB and 0.024 dB.



(a) *LU-NN trained FL and tested FL*

(b) *LU-NN trained FL and tested PL*



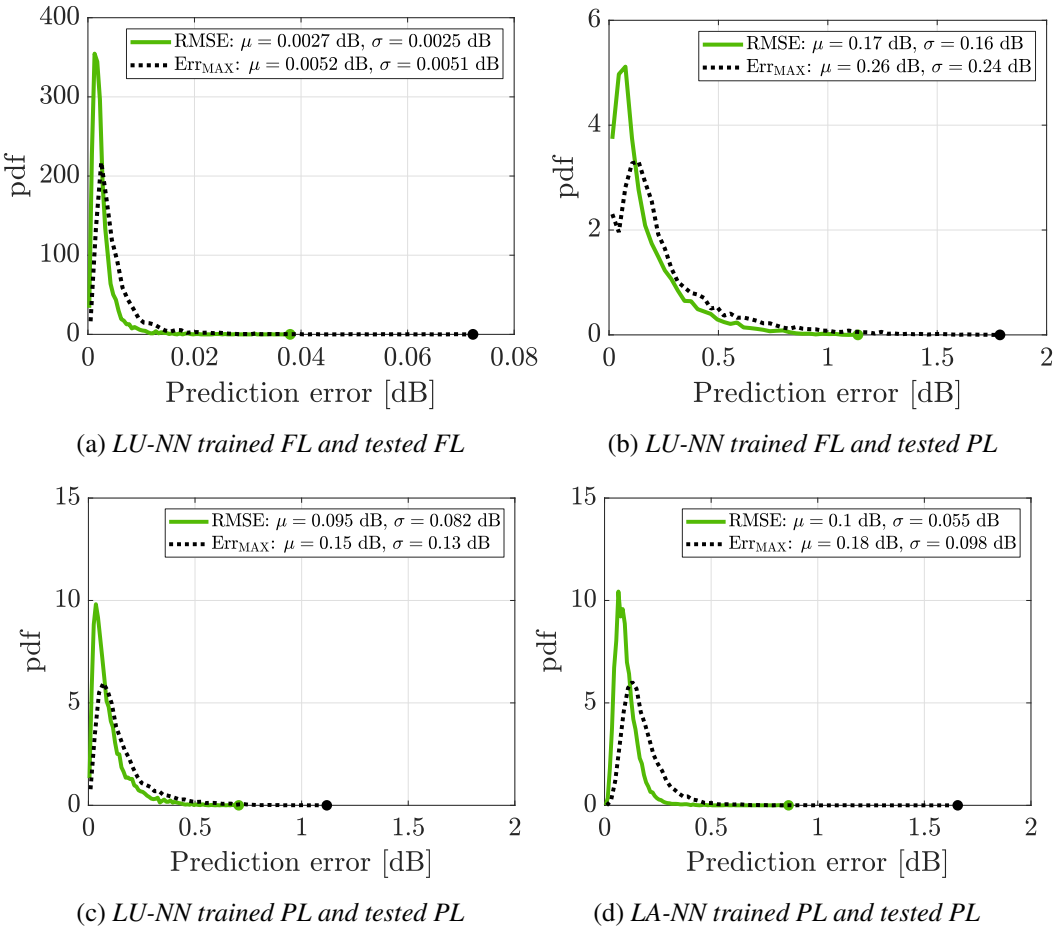
(c) *LU-NN trained PL and tested PL*

(d) *LA-NN trained PL and tested PL*

Fig. 4.15 Probability density functions of RMSE and Err_{MAX} of predictions of *on-off gain* profiles for the four different scenarios: (a) LU-NN trained and tested with full–load data–sets, (b) LU-NN trained with full–load data–set and tested with partial–load data–set, (c) LU-NN trained and tested with partial–load data–sets and (d) LA-NN trained and tested with partial–load data–sets.

As can be seen from the cdf curve in blue solid line on the right picture of Fig. 4.17, although the Err_{MAX} reaches values up to 0.34 dB, only few cases exceed 0.1 dB, with more than 98% of cases below 0.1 dB.

Similar high prediction performance is obtained for the ASE noise profiles, as illustrated in Fig. 4.16(a), where the pdfs of RMSE and Err_{MAX} are shown. The corresponding cdfs are plotted in blue solid line in Fig. 4.17. In this case, the RMSE is always below 0.04 dB, with mean equals to 0.0027 dB, while the Err_{MAX} has average 0.0052 dB and never exceeds 0.08 dB.



(a) *LU-NN trained FL and tested FL*

(b) *LU-NN trained FL and tested PL*

(c) *LU-NN trained PL and tested PL*

(d) *LA-NN trained PL and tested PL*

Fig. 4.16 Probability density functions of RMSE and Err_{MAX} of predictions of *ASE noise* profiles for the four different scenarios: (a) LU-NN trained and tested with full–load data–sets, (b) LU-NN trained with full–load data–set and tested with partial–load data–set, (c) LU-NN trained and tested with partial–load data–sets and (d) LA-NN trained and tested with partial–load data–sets.

Afterwards, in the perspective of supporting a more practical scenario where partial spectral loads are considered, we use this neural network to make predictions of RA gain profiles of the partial–load data–set (scenario 2). The testing results are shown in Fig. 4.15(b), where a dramatic worsening of the prediction performance can be observed. Now the mean value of the RMSE increases up to 0.41 dB and the one of the Err_{MAX} becomes 0.47 dB, with some cases greater than 2 dB.

Especially, for both RMSE and Err_{MAX}, two separate pdfs can be inferred, one covering very low errors and the other one including larger errors.

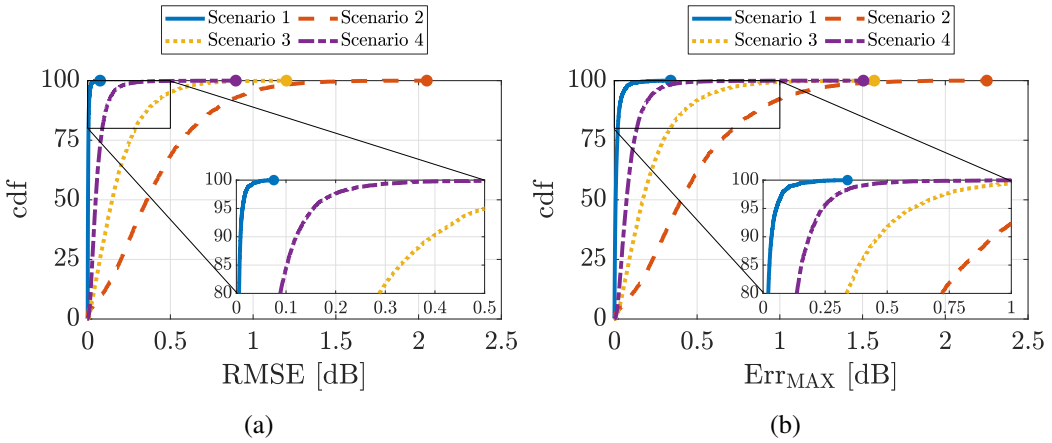


Fig. 4.17 Cumulative distribution functions of RMSE and Err_{MAX} of predictions of *on-off gain* profiles for the four different scenarios of Table 4.5. *Scenario 1*: LU-NN trained and tested with full–load data–sets; *scenario 2*: LU-NN trained with full–load data–set and tested with partial–load data–set; *scenario 3*: LU-NN trained and tested with partial–load data–sets; *scenario 4*: LA-NN trained and tested with partial–load data–sets.

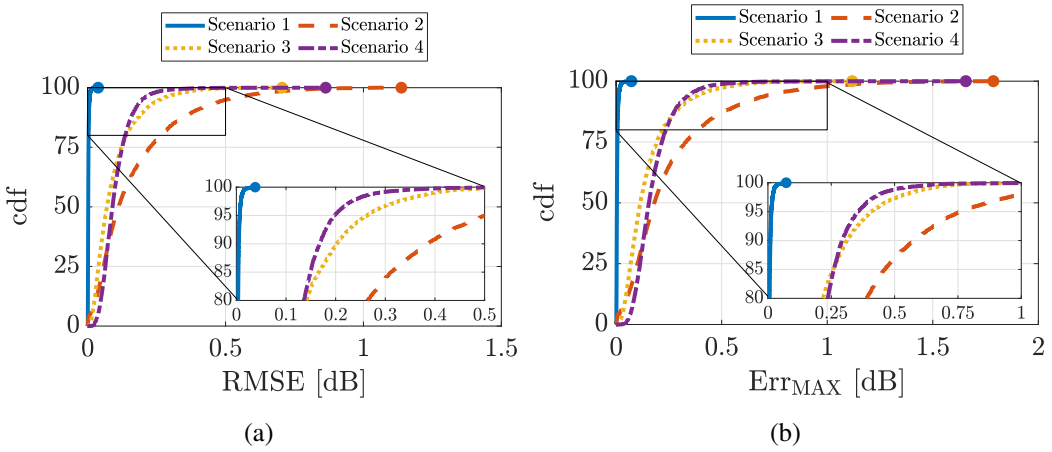
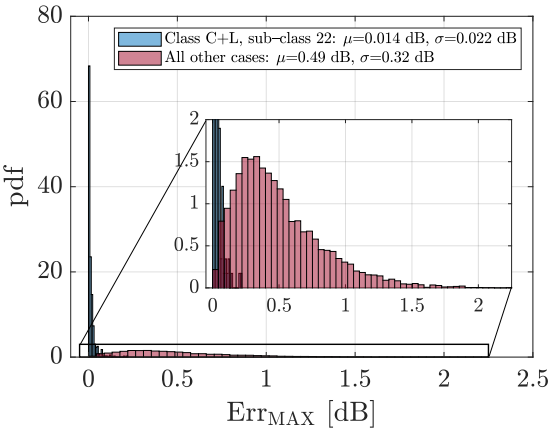
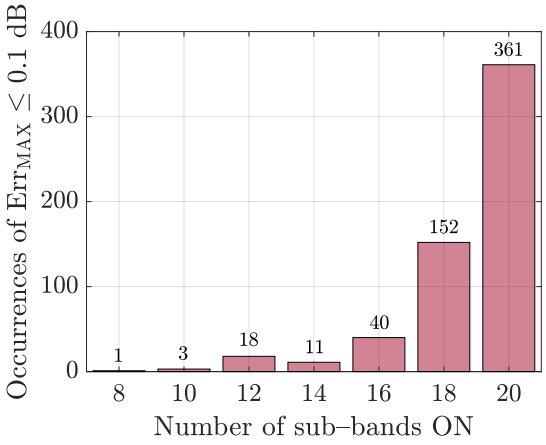


Fig. 4.18 Cumulative distribution functions of RMSE and Err_{MAX} of predictions of *ASE noise* profiles for the four different scenarios of Table 4.6. *Scenario 1*: LU-NN trained and tested with full–load data–sets; *scenario 2*: LU-NN trained with full–load data–set and tested with partial–load data–set; *scenario 3*: LU-NN trained and tested with partial–load data–sets; *scenario 4*: LA-NN trained and tested with partial–load data–sets.

The low errors most likely refer to the full–load cases present in the partial–load data–set, 500 profiles out of 11000 belonging to *class C+L* sub–class 22, while the other pdf contains all the other cases, the remaining 10500 samples. This supposition is confirmed by the results in Fig. 4.19.



(a)



(b)

Fig. 4.19 (a) Probability density functions of the $\mathrm{Err_{MAX}}$ evaluated for the two groups of profiles in which the partial–load data–set can be split: only fully loaded cases (class C+L sub–class 22) in blue and non–fully loaded cases in red; (b) occurrences of $\mathrm{Err_{MAX}}$ not exceeding 0.1 dB with respect to the number of sub–bands ON in the non–fully loaded group of profiles.

In Fig. 4.19(a), we show the pdfs of the $\mathrm{Err_{MAX}}$ obtained by splitting the full–load cases, the blue curve, from the other non–fully loaded cases (pure partial–load), the red curve. From there, we can see that for the two curves the average value of $\mathrm{Err_{MAX}}$ differs significantly, being 0.014 dB for the fully loaded cases and 0.49 dB for the others, with high prediction accuracy delivered only in case of full spectral loads. In fact, for the blue curve, the $\mathrm{Err_{MAX}}$ values are always below 0.22 dB, with the 99% of cases $\leq$0.1 dB.

Also the red curve presents some cases with $\text{Err}_{\text{MAX}} \leq 0.1$ dB. In Fig. 4.19(b), their distribution over the number of sub–bands ON is reported in terms of occurrences with respect to the total number of *pure* partial–load profiles. The results show that most of the cases fall in sub–classes with larger number of sub–bands ON, especially sub–classes 18 and 20. This is a clear sign that the LU-NN trained with full spectral loads is able to provide highly accurate predictions only when used to predict profiles generated in a fully loaded or almost fully loaded scenario, but not for any load condition.

In general, the overall performance of the LU-NN trained with the full–load data–set and tested over the partial–load data–set is poor. This can be easily seen in Fig. 4.17(b), where the relative cdf in orange dashed line (scenario 2) shows that only less than 62% of cases delivers $\text{Err}_{\text{MAX}} \leq 0.5$ dB.

The same analysis is performed for the prediction of ASE noise profiles. The resulting pdfs of RMSE and Err_{MAX} are shown in Fig. 4.16(b), where we can see that the mean value of the RMSE is 0.17 dB and the mean value of the Err_{MAX} is 0.26 dB with values up to 1.8 dB. These values are smaller compared to the gain case, with slightly improved prediction accuracy, because the ASE noise profiles are less dependent on the input spectral load, as discussed previously. This improvement is further demonstrated in Fig. 4.18, where $\text{Err}_{\text{MAX}} \leq 0.5$ dB for the 87% of cases of the partial–load testing data–set (cdf curve in orange dashed line).

Before introducing the LA-NN to handle any input spectral load condition, another intermediate step can be done to see if the LU-NN can adapt to different load conditions when trained with the partial–load data–set (scenario 3). Based on the fact that a variation in the input spectral load produces a variation in the RA gain profile, although load–agnostic because it does not receives the spectral load as input, the LU-NN can learn the average behaviour of different load conditions. With this approach, the ANN prediction performance further improves, as can be seen from the testing results illustrated in Fig. 4.15(c) for the gain profiles. Now the tail of the pdf of the Err_{MAX} goes down to about 1.5 dB and the average values of the RMSE and of the Err_{MAX} reduce to 0.19 dB and 0.22 dB, respectively.

The improvement of scenario 3 with respect to scenario 2 is even more evident in Figs. 4.17(a) and 4.17(b), where the cdfs related to scenario 3 (yellow dotted line) are steeper. In particular, in Fig. 4.17(b) more than 91% of cases have $\text{Err}_{\text{MAX}} \leq 0.5$ dB.

Similar testing results are obtained in case of ASE noise profiles prediction as illustrated in Fig. 4.16(c), where the average RMSE and the average Err_{MAX} decrease down to 0.095 dB and 0.15 dB, respectively. Like scenario 2, also in this case, the predictions of ASE noise profiles are more accurate than in case of gains, confirming that the noise is less impacted by different load conditions than gains.

Based on the obtained results, we can conclude that the LU-NN model analyzed so far is not able to provide highly accurate predictions at any load condition, especially for RA gain profiles. Thus, it is fundamental to introduce a load–aware neural network (LA-NN) that includes the spectral load information at its input. The LA-NN is trained and tested with partial–load data–sets (scenario 4). In Fig. 4.15(d), the pdfs of both RMSE and Err_{MAX} are reported in case of RA gain prediction. From there we can see that the prediction accuracy significantly increases, since the average RMSE is 0.062 dB and the average Err_{MAX} is 0.095 dB. Moreover, although some values of the Err_{MAX} reach 1.5 dB, for more than 99% of cases the Err_{MAX} is below 0.5 dB (see Fig. 4.17(b)).

As shown in Fig. 4.17, among all three scenarios with gain profiles prediction in partial–load condition, the scenario considering the LA-NN (scenario 4) is the most accurate one. Despite the highly accurate predictions given by the LA-NN, the introduction of the load condition at the input of the ANN adds a further degree of freedom, making the input-output mapping more complex. This is the reason why the LA-NN cannot reach the accuracy achieved by the LU-NN trained and tested with full–load data–sets (scenario 1), which indeed is a less complex scenario.

The LA-NN delivers high accurate predictions also in case of ASE noise, being the mean RSME and Err_{MAX} equal to 0.1 dB and 0.18 dB, respectively, as you can see in Fig. 4.16(d), but at the same time there is not a substantial improvement with respect to scenario 3 where the LU-NN is both trained and tested with the partial–load data–sets. On the contrary, their performance is rather similar, as can be seen in Fig. 4.18, meaning that the spectral load condition is not such a fundamental information in case of prediction of ASE noise profiles.

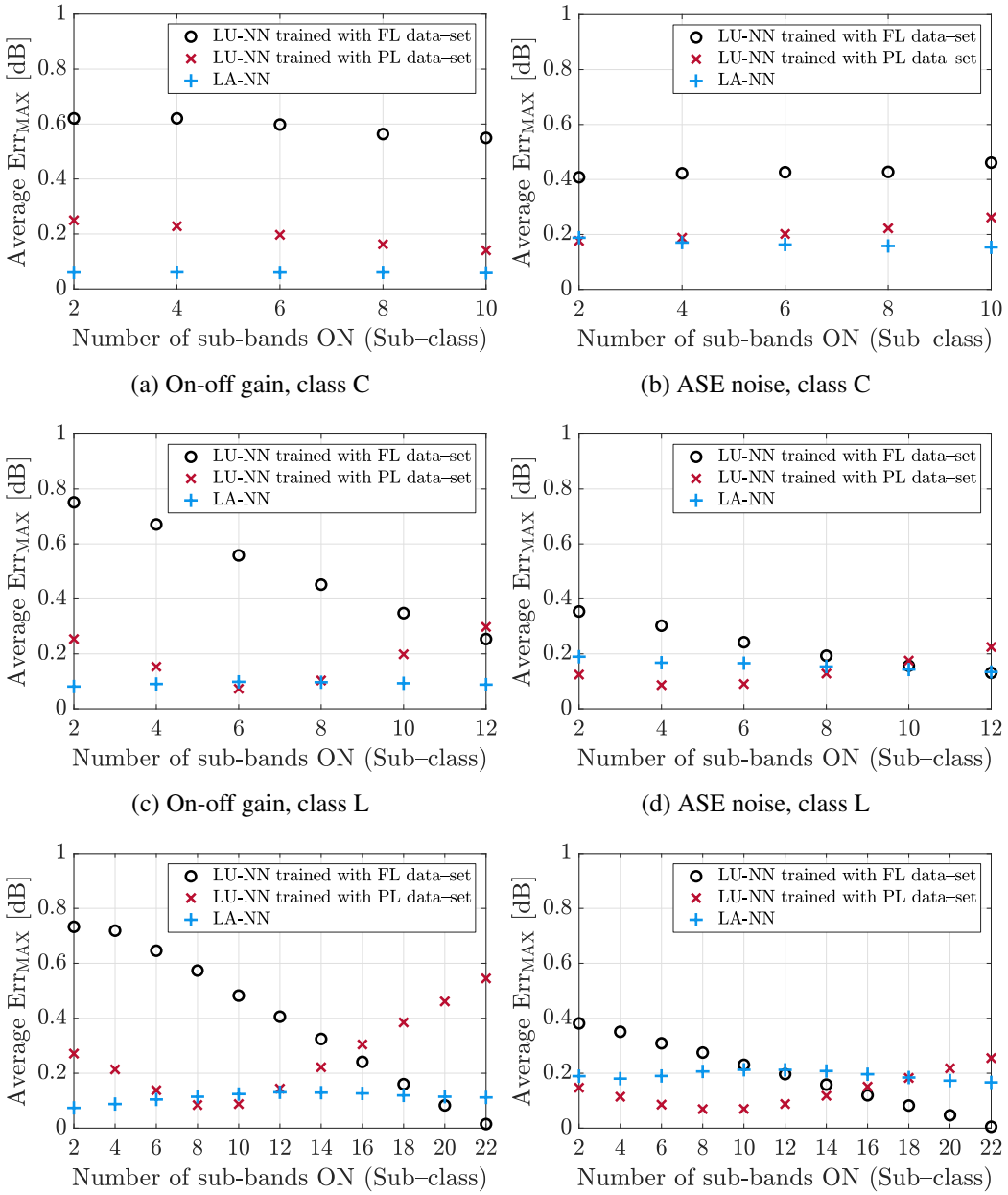
Statistical analysis of errors distribution of the partial–load testing data–set with respect to classes and sub–classes

The analysis carried out so far gives only a global overview of the ANNs performance in presence of partial spectral loads, without distinguishing the impact on the different classes (C, L and C+L) and sub–classes (from 2 to 22 sub–bands ON) of load. Thus, to fully understand how the prediction errors are distributed over the C+L–band and with respect to the number of sub–bands ON, we perform a comprehensive statistical analysis for both gain and noise profiles.

In Fig. 4.20, for each class of load and for each type of profile, the average $\mathrm{Err}_{\mathrm{MAX}}$ is shown with respect to the number of sub–bands ON. Classes are organized on rows with class C, class L and class C+L from top to bottom, while the profile types are organized on columns with the gains on the left and the noise on the right.

Fig. 4.20(a) shows the results in case of gain profiles and sub–bands in the C–band. From there we can see that, as expected, the worst performance is given when the LU-NN trained with the full–load data–set is used, as the average value of the $\mathrm{Err}_{\mathrm{MAX}}$ is always above 0.5 dB, with a slight improvement when the number of sub–bands ON increases. On the contrary, when the LU-NN trained with the partial–load data–set is considered, the mean $\mathrm{Err}_{\mathrm{MAX}}$ is always below or equal to 0.25 dB. Also in this case, the value of the average $\mathrm{Err}_{\mathrm{MAX}}$ decreases with the increasing number of sub–bands ON from 0.25 dB, for 2 sub–bands ON, down to 0.14 dB, for 10 sub–bands ON. As already known from the previous analysis, the highest prediction accuracy is obtained with the LA-NN. In fact, the average $\mathrm{Err}_{\mathrm{MAX}}$ is always around 0.06 dB regardless the number of sub–bands ON.

In case of prediction of ASE noise profiles (Fig. 4.20(b)), the mean $\mathrm{Err}_{\mathrm{MAX}}$ is always lower than 0.5 dB when the LU-NN trained with the full–load dat–set is used, showing better performance compared to the gain case, without showing an improvement with the number of sub–bands ON. Instead, with an average $\mathrm{Err}_{\mathrm{MAX}}$ always below 0.3 dB, similar performance is observed for the LU-NN trained with the partial–load data–set and the LA-NN. For the LU-NN trained with the partial–load data–set, the mean value of the $\mathrm{Err}_{\mathrm{MAX}}$ slightly increases in case of 8 and 10 sub–bands ON.



(a) On-off gain, class C

(b) ASE noise, class C

(c) On-off gain, class L

(d) ASE noise, class L

(e) On-off gain, class C+L

(f) ASE noise, class C+L

Fig. 4.20 Average Err$_{MAX}$ with respect to the number of sub–bands ON for the prediction of Raman on-off gain (on the left) and ASE noise (on the right) in partial load condition (partial–load testing data–set) considering different ANN models: the LU-NN trained with the full–load (FL) data–set, the LU-NN trained with the partial–load (PL) data–set and the LA-NN. The different classes (class C, class L and class C+L) are organized by rows.

The superiority of the LA-NN against the other ANN models in adapting to partial–loads and in providing high accurate predictions of RA gain profiles regardless the spectral load and the number of sub–bands ON is confirmed also for class L (Fig. 4.20(c)) and class C+L (Fig. 4.20(e)), with average Err_{MAX} assuming values between 0.07 dB and 0.13 dB. As already observed for class C, the prediction accuracy of the LU-NN trained with the full–load data–set gets better with the load condition getting closer to the full spectral load, but for class L and C+L, this improvement is even stronger. Indeed, for class L the mean Err_{MAX} improves of 0.5 dB, from the 0.75 dB of sub–class 2 to the 0.25 dB of sub–class 12. Similarly for class C+L, for which the LU-NN trained with the full–load data–set is able to overcome the LA-NN when the load condition belongs to sub–class 20 or sub–class 22. i.e. full–load. This is plausible because the LU-NN trained with the full–load data–set has been specifically optimized to provide highly accurate predictions in the full–load condition, while the LA-NN has not.

Taking for granted that the LA-NN is in general better than the LU-NN trained with the partial–load data–set, if we analyze the behavior of this latter ANN model, unlike class C that presents a clear decreasing trend with the number of sub–bands ON, we can see that the average Err_{MAX} shows a parabolic-like trend for class L and C+L. At their minimum, around sub–classes 6 and 8 for class L, and around sub–classes 8 and 10 for class C+L, the resulting values are comparable to those given by the LA-NN. Moreover, especially for class C+L, starting from sub–class 14 we observe a dramatic increase of the average Err_{MAX} reaching also values above 0.5 dB. This can be explained recalling that the LU-NN trained with the partial–load data–set, since it is load–agnostic, is able to learn only the average impact of different spectral loads on the RA gain and not all the possible variations.

Regarding the prediction of the ASE noise for class L and class C+L, the results are illustrated in Figs. 4.20(d) and 4.20(f), respectively. For both the two classes, when we consider the LU-NN trained with the full–load data–set, the average Err_{MAX} never exceeds 0.4 dB, showing higher accuracy compared to the gain prediction case. Moreover, like for gain profiles prediction, the performance improves with the increasing number of sub–bands ON, and in particular for more than half–load (i.e. from 12 to 22 sub–bands ON), the LU-NN trained with the full–load data–set overcomes the LA-NN performance, which instead has average Err_{MAX} always stable around 0.2 dB.

Finally, if we look at the performance of the LU-NN trained with the partial–load data–set, like for class C, also for classes L and C+L the performance is similar to those given by the LA-NN. Even lower average Err_{MAX} values are observed, especially in case of class C+L for sub–classes 4 to 16. At the basis of these results there are two considerations, which are related to each other: the reduced dependence of the ASE noise on the spectral load condition and the lower complexity of the LU-NN because it takes only the pumps powers at its input, simplifying the input-output learning. On the other hand, the load condition at the input of the LA-NN introduces more complexity, which justifies the slightly larger errors.

In general, we can observe that, in case of C–band, the performance of the LU-NN trained with FL data–set does not substantially improve, for gain prediction, or even gets worse, for ASE noise prediction, with increasing number of sub–bands ON compared to the L–band and C+L–band cases. To qualitatively explain why this happens, we can consider Figs. 4.10-4.13.

In case of on-off gain profiles (Fig. 4.20(a)), the average Err_{MAX} is around 0.6 dB, without showing a substantial improvement with increasing number of sub-bands ON. This can be explained by looking at Fig. 4.10, where, given a fixed configuration of pump lasers, the on-off gain profiles are plotted for different partial loads and different numbers of sub-bands ON. Focusing on the cases up to 9 sub-bands ON, there we can observe that in the C–band (191 THz to 196 THz), the deviation with respect to the full–load case, is reduced than in the L–band, with the gain profiles getting closer to the full–load case when the number of sub-bands ON is larger, but without showing particular difference among each other. This behavior can be observed also in Fig. 4.12, even though the gain profile is almost flat and not affected by a strong tilt as in Fig. 4.10. Thus, as the gain profiles are more similar regardless the input partial spectral load, also the corresponding errors with respect to the full–load case are similar.

Instead, regarding the ASE noise (Fig. 4.20(b)), the average Err_{MAX} assumes values between 0.4 dB and 0.5 dB, with no improvement when the number of sub-bands ON increases. Indeed, if we look at Fig. 4.11, regardless the number of sub-bands ON, the partially loaded cases overlap the full–load case in large part of the C–band (the high frequencies region). On the other hand, in the L–band the deviation with respect to the full–load case is more evident, showing that it is more subjected to the power transfer from higher to lower frequencies, a phenomenon

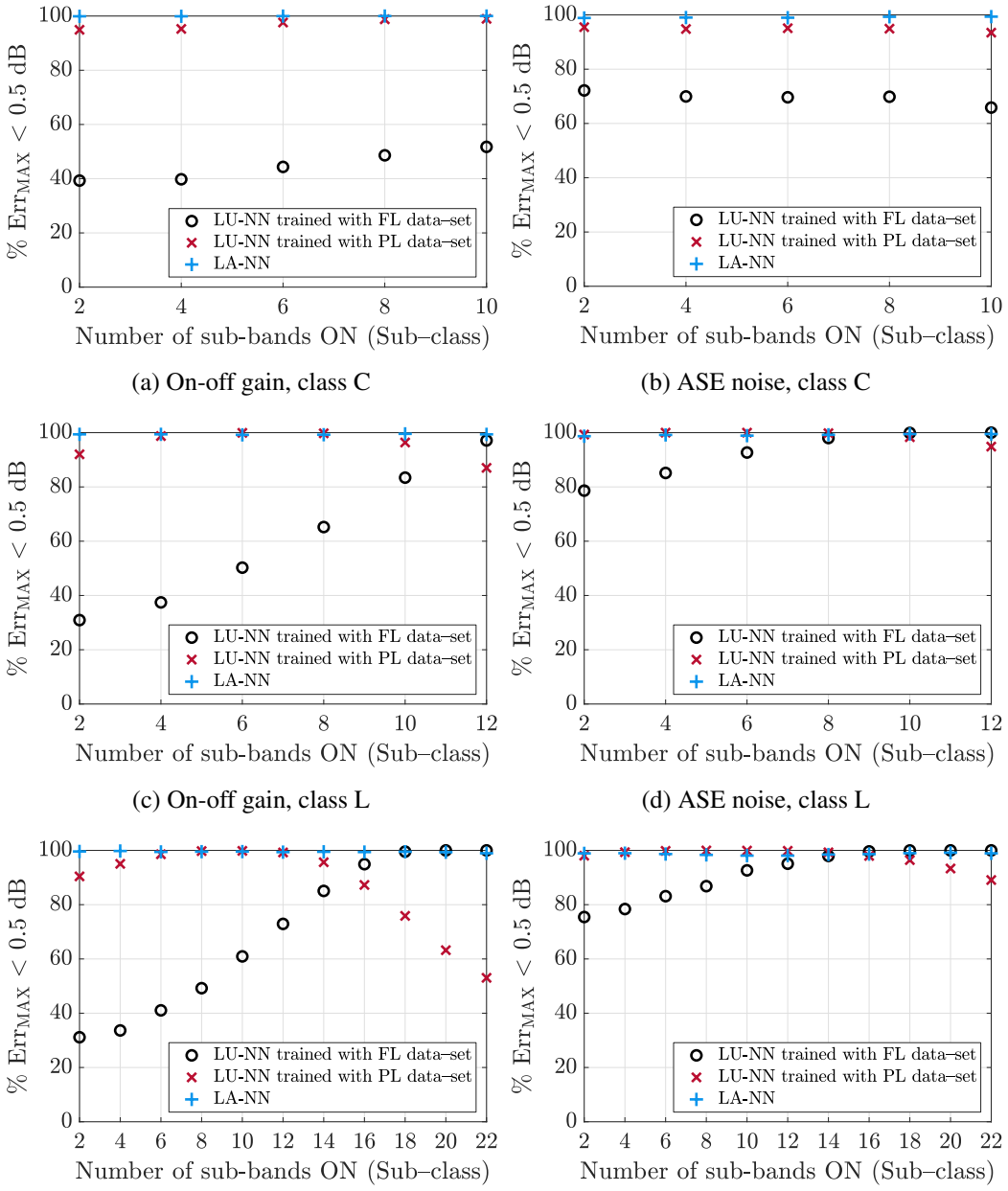
that is more mitigated when the number of sub-bands ON is larger. Similarly in Fig. 4.13, where the less strong tilt in the gain results in more similar ASE noise profiles, despite the different input partial loads and number of sub-bands ON, in particular in the C–band.

Therefore, since to quite different partial spectral loads correspond similar ASE noise profiles (in the C–band), this non-bijective relation between ANN input (partial-loads) and ANN output (ASE noise profiles) might prevent the ANN to correctly learn the input-output mapping. In particular, this might be more critical when the number of sub-bands ON is larger as there are more degrees of freedom, as more sub–bands are involved in pump-signal and signal-signal interactions.

Another interesting analysis to assess the reliability of the predictions of RA gain and ASE noise profiles, consists in looking at the distribution of the prediction errors over the tails of the pdfs, still with respect to classes and sub–classes. In Fig. 4.21 we report the percentage of cases of $\mathrm{Err_{MAX}}$ values below a reasonable threshold which guarantees accurate predictions with respect to the total number of profiles of the testing data–set. In our case this threshold is set to 0.5 dB. Results for the RA gain predictions are shown in Figs. 4.21(a), 4.21(c) and 4.21(e). For class C (Fig. 4.21(a)), since we have seen that for both the two LU-NN models the average $\mathrm{Err_{MAX}}$ reduces with the increasing number of sub–bands ON, consequently the percentage of $\mathrm{Err_{MAX}}$ cases under 0.5 dB increases with the sub–classes. Similarly for class L and class C+L (Figs. 4.21(c) and 4.21(e), respectively).

The limited ability of the LU-NN trained with the full–load data–set to provide highly accurate predictions at any load condition is confirmed by the low percentage of cases with $\mathrm{Err_{MAX}} < 0.5$ dB for most of the sub–classes of the three classes of load. In case of class C, the range of percentage is between 39% and 52%, with the lowest percentage referring to sub–class 2 and the highest to sub–class 10. Instead, for the other two classes, the increase with the number of sub–bands ON is stronger, from around 30% of cases up to 97%, for class L, and 100%, for class C+L.

A quite substantial improvement is observed for the LU-NN trained with the partial–load data–set. Indeed, the percentage of $\mathrm{Err_{MAX}}$ below 0.5 dB is more than 90% for the majority of cases with the exception of sub–classes from 16 to 22 of class C+L, for which the percentage of cases with $\mathrm{Err_{MAX}} < 0.5$ dB even reduces down to 53% in case of full–load condition.



(a) On-off gain, class C

(b) ASE noise, class C

(c) On-off gain, class L

(d) ASE noise, class L

(e) On-off gain, class C+L

(f) ASE noise, class C+L

Fig. 4.21 Percentage of $\mathrm{Err}_{\mathrm{MAX}}$ under 0.5 dB with respect to the number of sub–bands ON for the prediction of Raman on-off gain (on the left) and ASE noise (on the right) in partial load condition (partial–load testing data–set) considering different ANN models: the LU-NN trained with the full–load (FL) data–set, the LU-NN trained with the partial–load (PL) data–set and the LA-NN. The different classes (class C, class L and class C+L) are organized by rows.

Instead, with the LA-NN, highly accurate predictions of gain profiles are always guaranteed since almost 100% of cases shows Err_{MAX} below 0.5 dB for any load condition regardless the class and the sub–class, with a very limited number of cases exceeding this threshold.

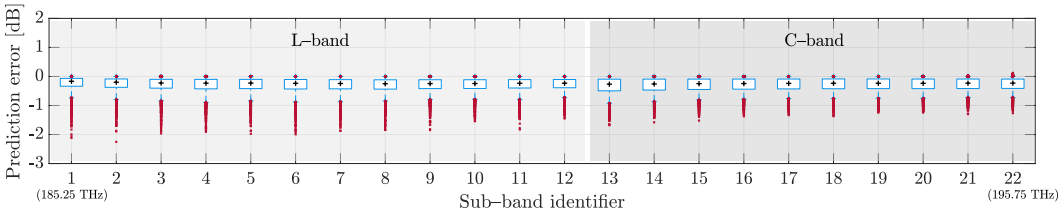
The same analysis is performed over ASE noise profiles predictions. The results are shown in Figs. 4.21(b), 4.21(d) and 4.21(f) for classes C, L and C+L, respectively. The higher accuracy compared to the gain case is further demonstrated by the higher percentage of cases with Err_{MAX} under 0.5 dB, with values always above 60% for all the three ANN models. More precisely, for the LU-NN trained with the partial–load data–set the percentage of Err_{MAX} below 0.5 dB is always above 89%, while for the LA-NN almost 100% of cases have Err_{MAX} lower than 0.5 dB.

From these analyses, it results that the LA-NN is the most suitable ANN model to provide accurate predictions at any load condition, especially for RA gain profiles, since in the majority of cases the maximum absolute prediction errors Err_{MAX} is lower than 0.5 dB.

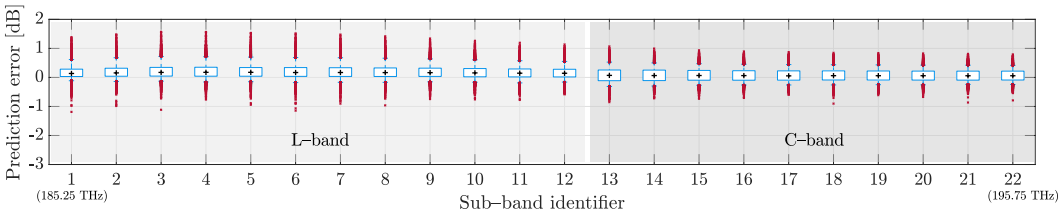
Statistical analysis of the spectral distribution of prediction errors under partial load conditions

Up to this point, for each predicted profile we evaluated the RMSE and the Err_{MAX} which provide an overall indication of the accuracy over the entire profile, but without giving any information on how the errors are spectrally distributed over the transmission bandwidth. For this purpose, for each gain or ASE noise profile of the partial–load testing data–set, we evaluate the errors between the predicted and the target profile at any frequency position. Unlike RMSE and Err_{MAX}, these errors can be both positive or negative.

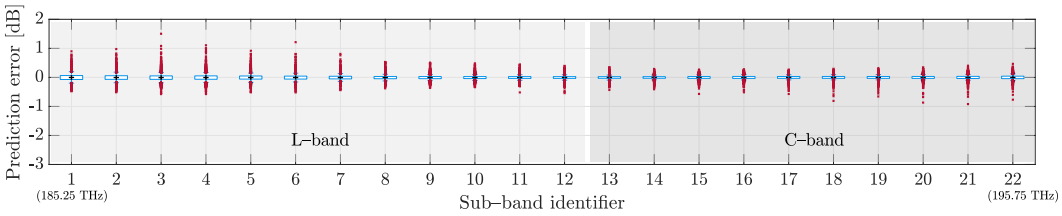
In Fig. 4.22, we show the prediction errors in case of RA gain profiles with respect to the sub–bands frequency position for the LU-NN trained with the full–load data–set (Fig. 4.22(a)), the LU-NN trained with the partial–load data–set (Fig. 4.22(b)) and the LA-NN (Fig. 4.22(c)). For the sake of simplicity, each sub–band is identified by a numerical identifier from 1 to 22 ordered in frequency, which means that 1 corresponds to the lowest frequency (185.25 THz) and 22 corresponds to the highest frequency (195.75 THz). Identifiers from 1 to 12 are in the L–band, while identifiers from 13 to 22 are in the C–band. Since it is not practical to plot the probability



(a) LU-NN trained with the full–load data–set



(b) LU-NN trained with the partial–load data–set



(c) LA-NN

Fig. 4.22 Testing prediction errors of the *Raman on-off gain* profiles in partial load condition (partial–load testing data–set) with respect to sub–bands identifiers (ordered in frequency) for different scenarios: (a) LU-NN trained with the full–load data–set, (b) LU-NN trained with the partial–load data–set and (c) LA-NN. Median values are plotted with the black cross marker, boxes and whiskers include the 50% and the 90% of cases, respectively, and the remaining cases are plotted as individual markers.

density function of the prediction errors for all sub–bands position and ANN models, we prefer to represent the pdfs in amore compact way as candle plots. We already introduced this type of representation in Chapter 3. Just to recall, for each sub–band, the candle plot reports the median value of the pdf indicated with a single black cross marker, a box containing the 50% of total cases between the 25% and 75% of the pdf, and whiskers including the 90% of cases between the 5% and 95% of the pdf. The remaining 5% of cases falling in the tail of the pdf, each per side, are plotted as individual markers in red.

Fig. 4.22(a) shows the results when the LU-NN trained with the full–load data–set is used to make predictions of the RA gain profiles in partial spectral load conditions. From there, we can see that the majority of the errors is negative. Indeed, as clearly

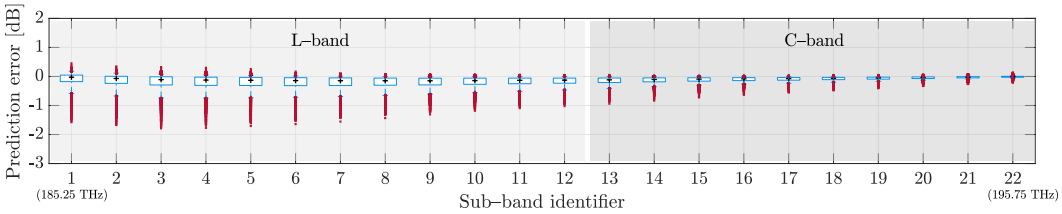
illustrated in Fig. 4.10, where we discussed the impact of full and partial spectral loads on gain profile, lower input power levels generate higher gains compared to the full–load case, because the reduced number of sub–bands ON corresponds to less interactions among sub–bands in terms of power transfer. Therefore, the gain profiles in partial–load condition predicted by the LU-NN trained with full–load data–set (scenario 2) are in general under-estimated and the prediction errors become negative. We can observe that the median does not seem to depend on the frequency position of sub–bands, assuming values between -0.26 dB and -0.17 dB, as well as for boxes and whiskers. The lower bound of boxes (25% of the pdf) assumes values between -0.49 dB and -0.34 dB, while the upper bound (75% of the pdf) is between -0.11 dB and -0.07 dB. Instead, the minimum box extension is experienced in sub–band 1 with value 0.27 dB and the maximum extension in sub–band 13 with value 0.40 dB. Regarding the whiskers, the lower bounds (5% of the pdf) are between -0.92 dB and 0.73 dB, while the upper bounds (95% of the pdf) are around 0 dB. With 0.73 dB of difference between the upper and the lower bounds, the minimum range of values occurs in sub–band 21, while the maximum is experienced in sub–band 13 with value 0.92 dB. The dependence on the sub–bands frequency position is observed in the errors plotted as individual markers, which mostly extend in the bottom part of the figure, with larger range in the lower region of the L–band (lower sub–bands identifier). The largest error, corresponding to -2.25 dB, occurs in sub–band 2.

If we analyze the LU-NN trained with the partial–load data–set, whose results are illustrated in Fig. 4.22(b), we immediately notice that the prediction errors are no longer mainly negative, but they can assume both negative and positive values with similar probability. As a matter of fact, the values of the median is between 0.05 dB and 0.18 dB, all positive, which means that there is a slight bias towards overestimation, in particular in the L–band, where we can see that the values are higher than in the C–band. The boxes have lower bounds in the range [-0.11, 0.046] dB and the upper bounds in the range [0.22, 0.35] dB, while the minimum extension of the box is 0.25 dB in sub–band 1 and the maximum is 0.37 dB in sub–band 13, showing a small reduction compared to the previous case. In addition, we can observe that all the boxes in the L–band contain positive errors, confirming the slight trend towards overestimation. Looking at the whiskers, their extension ranges from 0.66 dB in sub–band 22 to 0.87 dB in sub–band 4, with lower bounds between -0.31 dB and -0.11 dB and upper bounds between 0.42 dB and 0.72 dB. Then, the errors plotted as individual markers are always within -1.19 dB and 1.57 dB,

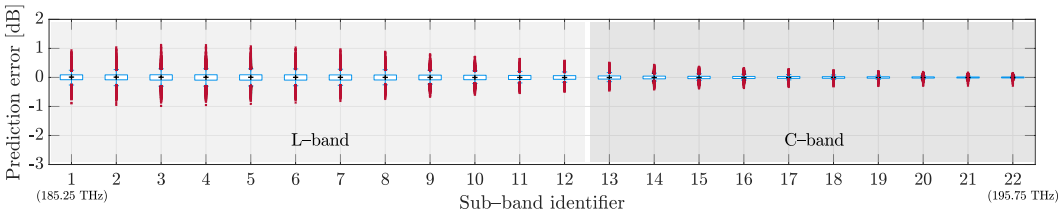
experienced in sub–band 1 and sub–band 3, respectively. Like for the LU-NN trained with full-load data–set, there is a slight spectral dependence, particular evident in the errors plotted as individual markers.

Finally, in Fig. 4.22(c), we report the prediction errors given by the LA-NN. There, its higher capability in providing accurate predictions of RA gain profiles at any load condition is further proved. Compared to both the LU-NN analyzed in Figs. 4.22(a) and 4.22(b), in this case, the median assumes values around 0 dB, therefore the ANN model is no longer biased to under-estimate or over-estimate the gain profiles. Moreover, a considerable reduction of the ranges of boxes and whiskers can be observed, with maximum values equals to 0.14 dB and 0.39 dB, respectively, and both experienced in sub–band 1. Also the range of extension of the errors plotted as individual markers is reduced, with most of the errors falling within the range [-1,1] dB. As for the other two ANN models, broader ranges are observed in the lower portion of the L–band. This might be caused by the higher order interactions induced by the SRS not only between pumps and sub–bands but also between sub–bands, with a power transfer from higher frequency sub–bands to lower frequency sub–bands.

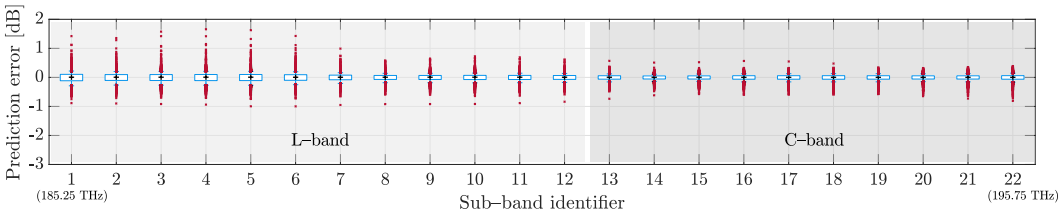
In Fig. 4.23 the candle plots of the prediction errors in case of ASE noise profiles are reported with respect to the sub–bands position for the LU-NN trained with full–load data–set (Fig. 4.23(a)), the LU-NN trained with the partial–load data–set (Fig. 4.23(b)), and the LA-NN (Fig. 4.23(c)). Although we have already seen that errors are smaller in case of noise predictions than in case of gain profiles, similar considerations can be drawn about the trends of candle plots, with the difference that in this case the extension of boxes and whiskers are reduced, confirming a less dependence of the noise on the input load conditions. Like for gains, also in this case, the noise profiles predicted with the LU-NN trained with full-load data–set are generally under-estimated, since most of the errors are negative. In fact, as we can see in Fig. 4.18(a), the median is always negative with values oscillating between -0.15 dB and -0.01 dB. Unlike the gain case, there is a stronger spectral dependence, with a significant difference between candles in the C–band and candles in the L–band. In the C–band the prediction errors are considerably smaller than in the L–band, with much reduced ranges for boxes and whiskers. With a difference of 0.033 dB between the upper and the lower bound, the boxes have their minimum extension in sub–band 22, as well as whiskers, with a difference of 0.10 dB.



(a) LU-NN trained with the full–load data–set



(b) LU-NN trained with the partial–load data–set



(c) LA-NN

Fig. 4.23 Testing prediction errors of the *ASE noise* profiles in partial load condition (partial–load testing data–set) with respect to sub–bands identifiers (ordered in frequency) for different scenarios: (a) LU-NN trained with the full–load data–set, (b) LU-NN trained with the partial–load data–set and (c) LA-NN. Median values are plotted with the black cross marker, boxes and whiskers include the 50% and the 90% of cases, respectively, and the remaining cases are plotted as individual markers.

The maximum extension between upper and lower bound is in sub–band 3 for both boxes and whiskers with values equal to 0.28 dB and 0.82 dB, respectively. Also the errors plotted as individual markers assume more limited values, never exceeding 1 dB in absolute value, in the C–band than in the L–band, where -1.8 dB is reached.

A similar spectral dependence of the prediction errors with respect to frequency position of sub–bands is observed in case of the LU-NN trained with the partial–load data–set (Fig. 4.23(b)), with larger errors in the lower portion of the L–band and very small errors in the upper portion of the C–band. The median is always around 0 dB, therefore the ANN model is not biased towards underestimation, as for the LU-NN trained with the FL data–set, or overestimation, like for the LU-NN trained with the

partial–load data–set used for gain prediction. The maximum sizes of boxes and whiskers occur in sub–band 4 and are equal to 0.19 dB and 0.60 dB, respectively.

Also for the LA-NN, shown in Fig. 4.23(c), there is a slight dependence of errors with respect to the frequency position of sub–bands, like in case of gain profiles predictions. Moreover, although performance of the LA-NN is similar to the LU-NN trained with partial–load data–set performance, with the LA-NN the errors are more evenly distributed over sub–bands. As observed in case of gain prediction, the majority of the errors plotted as individual markers are within -1 dB and 1 dB, with just few cases going beyond.

4.4 Generalization of ML-based Raman Amplification Analysis to Multiple Fibers

The system analyzed so far considers only one type of fiber, the SSMF, at a single span length, $L_{span} = 100$ km. Nevertheless, this is a limiting condition suitable only to provide a proof of principle, but not to be used in real field since different types of fiber at different span lengths are employed in optical communication networks.

Starting from the study carried out for the SSMF in partial–load conditions for the prediction of RA gain and noise profiles, we extend the analysis to three other types of commercially available fibers, the non-zero dispersion shifted fiber (NZ-DSF) Truewave™ and two kinds of ultra-low loss fiber (ULLF), TeraWave® SCUBA 125 and TeraWave® SCUBA 150. In addition to 100 km span, two other span lengths, 50 km and 75 km, are considered.

In the following we will present the details of the data–sets generation and the description of the ANN model upgraded to handle multiple fibers, with an extensive discussion on predictions accuracy.

4.4.1 Data–sets Generation

The system under analysis is the same as the one reported in Fig. 4.9 considering partial spectral loads, with the difference that now three other fiber types and two other fiber span lengths are added. In Table 4.7, we report the parameters for the

Fiber Type	α_s [dB/km]	α_p [dB/km]	A_{eff} [μm^2]	C_r [$W^{-1}km^{-1}$]	L_{span} [km]
SSMF	0.21	0.25	80	0.3841	[50,75,100]
Truewave-RS™ [81]	0.21	0.25	55	0.73	[50,75,100]
TeraWave® SCUBA 125 [75]	0.158	0.208	125	0.2458	[50,75,100]
TeraWave® SCUBA 150 [76]	0.155	0.205	153	0.2008	[50,75,100]

Table 4.7 Parameters for different types of fibers.

four different fibers: the attenuation coefficient for the signal α_s, the attenuation coefficient for the pumps α_p, the effective area A_{eff}, the Raman gain efficiency coefficient C_r and the span length L_{span}.

The generation of training and testing data–sets is done as previously described based on classes (spectral occupation) and sub–classes (number of sub–band ON). In particular, for both data–sets, for each combination of class, sub–class, fiber type and fiber length, 300 different partial spectral loads are generated with a corresponding pumps powers configuration drawn from a uniform distribution in the range [0,250] mW, for a total of M=79200 samples per data–set.

For each case, by means of the numerical Raman solver in [45], we determine the on–off gain and the ASE noise profiles. Depending on the fiber type and length, different ranges of gain are obtained. For this reason, gain profiles of the testing data–set are not pruned. Figs. 4.24 and 4.25 provide a graphical representation of Raman gain and ASE noise profiles, respectively, for the different fiber types and lengths. Please note that, because of pumps frequency allocation and pump-to-signal, signal-to-signal and pump-to-pump interactions following SRS, the upper limit of each gain range is not experienced at any frequency, but in general it is reached in the lower frequencies region.

4.4.2 Machine Learning Framework

Since partial spectral loads are considered, the ANN models used to predict gain and ASE noise profiles are LA-NN properly upgraded to support multiple fiber types and multiple span lengths. Fig. 4.26 shows the schematic of the ANN model, where at the input, apart from pumps powers $\mathbf{P} = [P_1, \ldots, P_{\mathrm{Np}}]^{\mathrm{T}}$ and spectral loads $\mathbf{S} = [S_1, \ldots, S_{\mathrm{Nsb}}]^{\mathrm{T}}$, other parameters are added to properly identify each case: fiber span length L, attenuation coefficient for signal α_s and pumps α_p, effective area

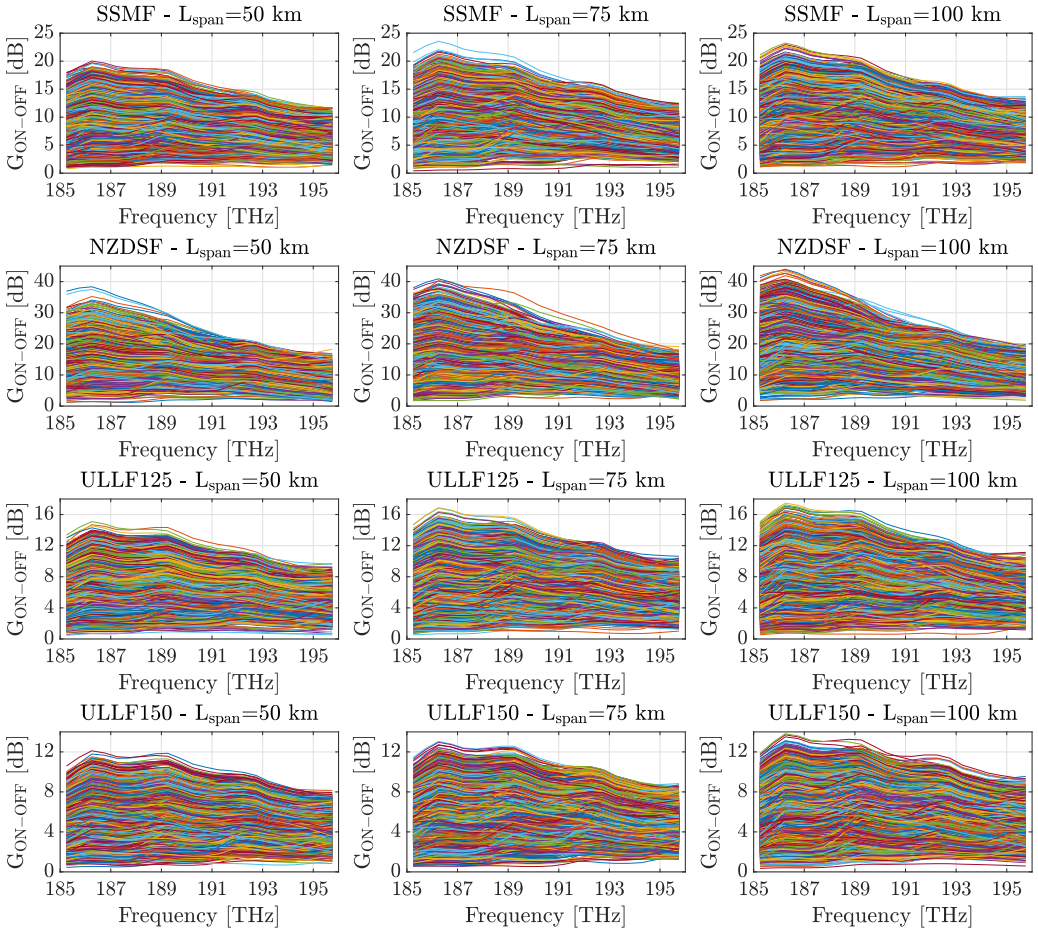


Fig. 4.24 Graphical visualization of Raman on-off gain profiles of training data–set for different fiber types and span lengths.

A_{eff} and Raman gain efficiency coefficient C_r. Depending on the type of profile to predict, at the output of the ANN we can have the Raman on–off gain profile $\mathbf{G} = [G_1, \ldots, G_{Nsb}]^T$ or the ASE noise profile $\mathbf{N} = [N_1, \ldots, N_{Nsb}]^T$.

Unlike the previous study, in this case we use Tensorflow and Keras libraries to train, validate and test the ANN models. For both gain and ASE noise profiles prediction we consider fully connected ANNs with the following hyper–parameters for ANN model: 2 hidden layers and 400 hidden nodes per layer. Input and output layers have 32 and 22 hidden nodes, respectively. The activation function used at input layer and hidden layers is the rectified linear unit (ReLU), while linear function is used at output layer. ANN features (data at the input of the ANN) are scaled to be in the [0,1] range [77].

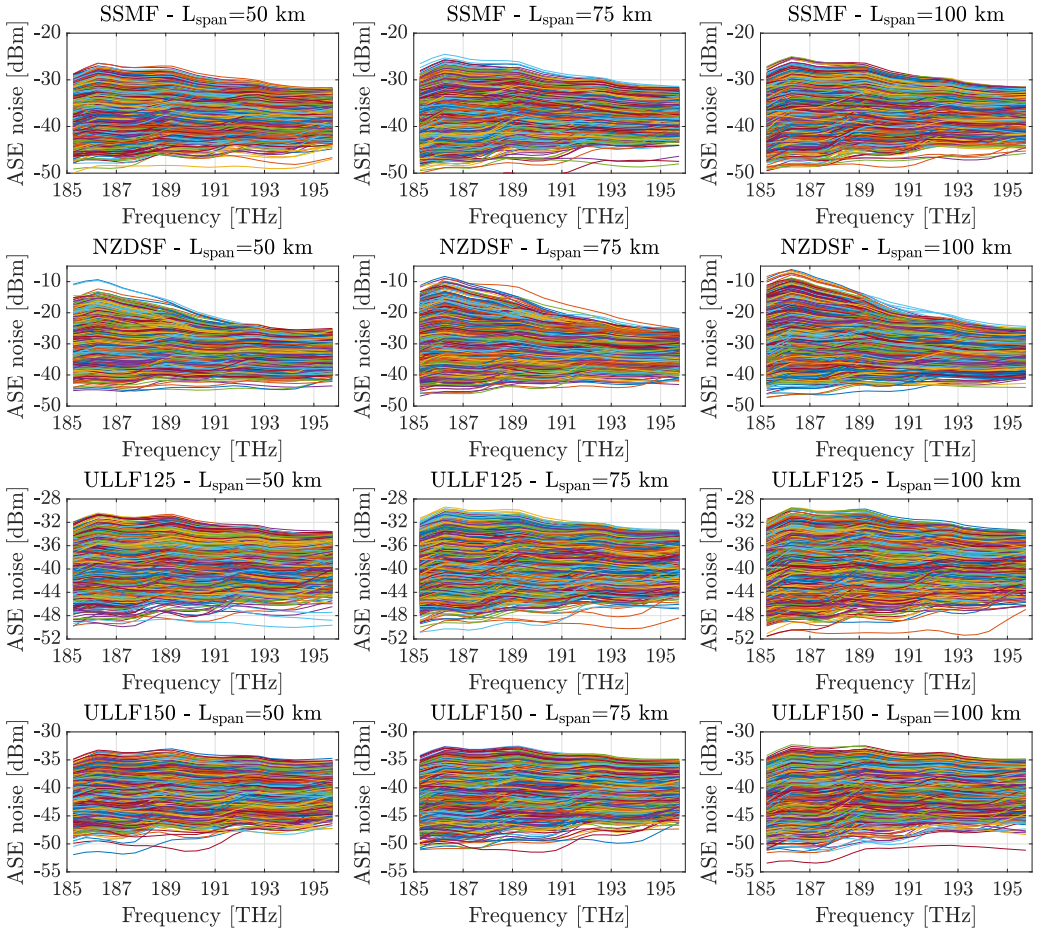


Fig. 4.25 Graphical visualization of ASE noise profiles of training data–set for different fiber types and span lengths.

ANN weights are initialized using default settings, i.e. Glorot uniform initialization, also known as Xavier uniform initialization [22]. For ANN training, we consider 0.1 validation split, which means that 90% of training data–set is actually used for training and 10% for validation, learning rate $\eta = 0.001$ and mini-batch size $N_{batch} = 100$. To optimize the ANN weights we use the Adam (adaptive moment estimation) optimizer [65], a stochastic gradient descent method based on adaptive estimation of first-order and second-order moments [21]. The loss function is the MSE and the ANN training is stopped after 10000 epochs.

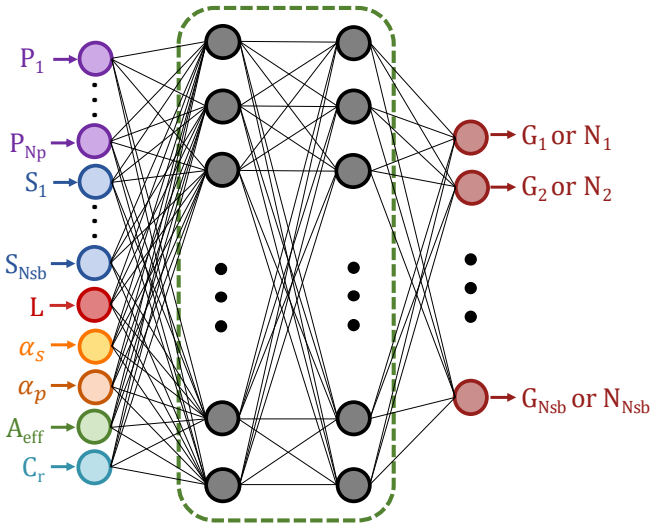


Fig. 4.26 Schematic of the considered artificial neural network model.

4.4.3 Testing results

After training, we test the ANNs over the unseen data of the testing data–sets. The prediction performance is evaluated in terms of RMSE and Err_{MAX} as previously done and computed between predicted and target profiles. Figs. 4.27(a) and 4.27(b) illustrate the pdf and the cdf, respectively, of RMSE and Err_{MAX} for all tested Raman on-off gain profiles. From there we can see that very highly accurate predictions are obtained, in particular we have that the average RMSE and Err_{MAX} are 0.023 dB and 0.036 dB, respectively. Although, values up to 1.86 dB are reached, they are very few with respect to the entire testing data–set population, since for 99% of cases the Err_{MAX} is below 0.17 dB and $\text{Err}_{\text{MAX}} \leq 0.5$ dB for almost 100% of cases. Analyzing the cases for which $\text{Err}_{\text{MAX}} > 0.5$ dB, it results that most of the cases correspond to NZ-DSF, which indeed has a broader range of gain levels.

Similar considerations can be drawn in case of prediction of ASE noise profiles (Fig. 4.28), for which the mean RMSE and Err_{MAX} are 0.03 dB and 0.046 dB, respectively. Also in this case, the tails of pdf and cdf extend to almost 1.5 dB, but the majority of cases have very low prediction errors, with 99% of cases having Err_{MAX} below 0.17 dB.

These results represent a further confirmation of the ability of direct ANN models to provide highly accurate predictions of Raman gains and ASE noise in presence

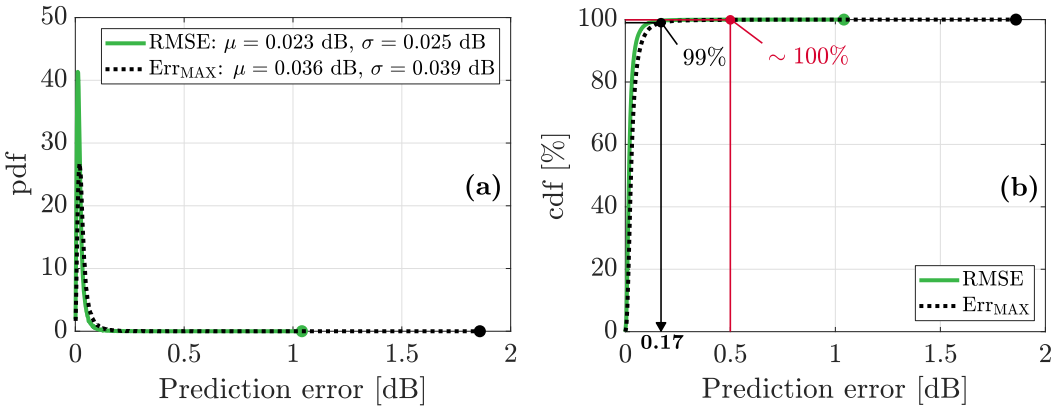


Fig. 4.27 (a) Probability density function (pdf) and (b) cumulative distribution function (cdf) of RMSE and Err$_{\text{MAX}}$ for predicted *Raman on-off gain* profiles of testing data–set.

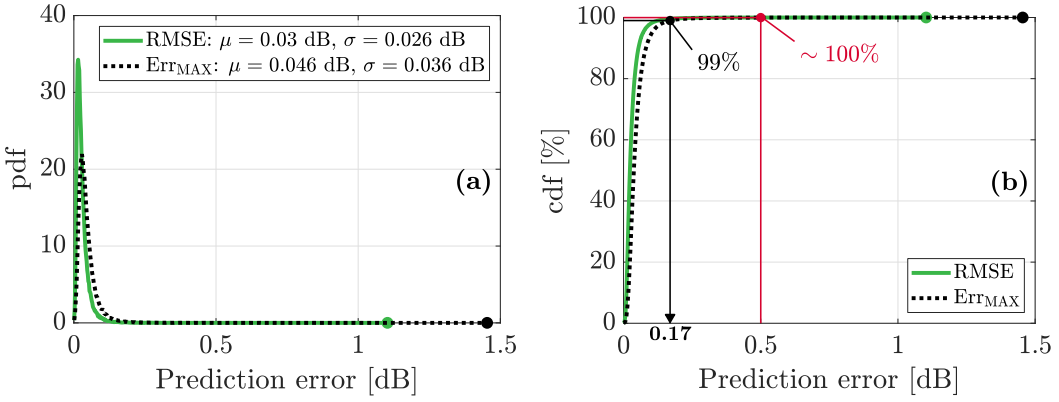


Fig. 4.28 (a) Probability density functions (pdf) and (b) cumulative distribution functions (cdf) of RMSE and Err$_{\text{MAX}}$ of *ASE noise* profiles of testing data–set.

of partial spectral loads also in a more general scenario considering different types and lengths of fibers. Remarkably, these results tell us that there is no need of a dedicated ANN for each specific fiber fiber, but a single ANN is able to manage all the scenarios, allowing to extend the range of possible applications of the proposed approach. A more extensive analysis of errors distribution with respect to fiber type, number of sub–bands ON and frequency position of sub–bands ON can be carried out, as done in section 4.3, but we are not going to provide it here as we expect trends similar to those obtained in section 4.3.2.

4.5 Conclusions

In this Chapter we have proposed the use of a *direct* ANN model to predict RA gain and ASE noise profiles in an ultra-fast and accurate way for multi–band systems (in our case over the C+L–band). Starting with the LU–NN, introduced for a fully spectral loaded scenario, then upgraded to the LA–NN to handle thousands of different partial spectral loads, we have shown the great ability of this direct ANN model to provide highly accurate predictions of both gain and noise profiles.

Furthermore, a single generalized model for multiple fiber types and span lengths has been presented, for which very high predictions are obtained on the testing data–sets profiles (both gain and ASE noise), as almost 100% of cases have Err_{MAX} below 0.5 dB.

Like for the prediction of power profiles discussed in Chapter 3, also here the complexity is only at the data collection and ANNs training stages. The prediction of gain and noise profiles is instantaneous, proving that the proposed approach can be suitable to be used in real time networks.

Chapter 5

Machine Learning for Raman Amplification Design

In this Chapter, a pure machine learning based approach is discussed for the inverse design of Raman amplifiers to obtain arbitrary, flat and tilted gain profiles in a multi–band scenario. Both synthetic and experimental data are considered. Moreover, a gradient descent based routine is introduced to fine–adjust the RA pumps configurations when the predictions given by the inverse ANN are not accurate enough. Most of the results here reported have been published in [28–31, 117].

5.1 Introduction

The determination of the optimal input parameters returning a desired output function, is a relevant problem in many fields of photonics, such as optical communications components and systems and network design and optimization [35, 69, 71, 78, 107, 115]. Among them, the design of Raman amplifiers has received considerable attention in the optical communications community, thanks to their potential in providing flexible gains over a broad range of frequencies, as required in future ultra–wideband optical networks.

In general, for Raman amplifiers the determination of pump lasers configuration in terms of powers and frequencies/wavelengths to obtain a targeted gain profile is a non–trivial, time–consuming and complex optimization procedure that relies on

the solution of the propagation equations describing the power evolution of signal and pumps along the fiber. Due to the relevance of the problem, some solutions have been proposed, such as the use of genetic algorithms jointly with propagation equations [41, 74], though resulting in time–consuming approach without guaranteed convergence to a solution, or in combination to direct model artificial neural networks to avoid solving the propagation equations [17].

Despite being powerful tools, genetic algorithms have two main limitations: they are based on a random optimization that may require non–negligible convergence time, in particular when initial conditions are not suitably set, and when a new gain profile is targeted, the optimization procedure requires to be started again from scratch.

Many other solutions can be found in the literature [6, 11, 38, 58, 61, 72, 79, 112], but they all rely on optimization loops. Therefore, although being speeded up through approximations and simplifications, these methods still require several iterations before reaching the optimal pumps configurations.

Recently, to avoid repeating the optimization procedure every time a new profile is targeted, pure machine learning based approaches have been successfully proposed for the inverse design of Raman amplifiers [30, 55, 56, 113, 116, 117], also in case of few–mode fibers [18, 70], with substantial advantages from the computational time point of view.

Thanks to their high flexibility, Raman amplifiers operating in multi–pump configurations can be potentially used not only to amplify signals, but also to perform gain shaping, like in frequency combs wavelength–dependent loss compensation or in the achievable information rate optimization by adjusting channels powers. Furthermore, RAs can be used in hybrid schemes to complement the gain of other amplification technologies [56].

In the following, we will only focus on pure Raman amplification over different transmission bands (C, C+L and S+C+L–bands) and considering different input signal conditions. After presenting the general ML–framework for the inverse system design, including a fine–optimization routine to improve the prediction of the input parameters, we then discuss the ANN inverse model performance in designing arbitrary, flat and tilted gain profiles for both synthetic and experimental data.

5.2 Machine Learning Framework

In the inverse model $\mathbf{X} = f^{-1}(\mathbf{Y})$, given a target output $\mathbf{Y}^{targ}$ the goal is to find the corresponding input $\mathbf{X}^{targ}$. To do this we have to determine the inverse mapping $f^{-1}(\cdot)$, a complex and non–trivial task, as for many systems the direct mapping function $f(\cdot)$ requires the numerical solution of nonlinear differential or integral equations that do not have analytical solutions. For other cases, $f(\cdot)$ is even unknown.

As already seen in Chapter 4, for the Raman amplifier case the direct mapping function $f(\cdot)$ corresponds to the ODEs equations describing the power evolution of channels and pumps along the fiber, whose solution may be complex and quite time–consuming. Thus, similarly to what we have done for the direct model, also here we propose the use of a machine learning and artificial neural network based method to learn the inverse mapping. The ML–framework is illustrated in Fig. 5.1. The inverse ANN model, referred to as ANN_{inv}, reported in Fig. 5.1(a), can theoretically learn the mapping $f^{-1}(\cdot)$ between the output vector $\mathbf{Y} = [\mathrm{Y}_1, \ldots, \mathrm{Y}_N]^{\mathrm{T}}$ and the input vector $\mathbf{X} = [\mathrm{X}_1, \ldots, \mathrm{X}_M]^{\mathrm{T}}$ with high enough accuracy, as long as the direct mapping is unique [46]. M and N are the length of input and output vectors, respectively.

Given the training data–set $\mathscr{D}_{inv}^{K\times(N+M)} = \{\mathbf{Y}_k, \mathbf{X}_k | k = 1, \ldots, K\}$, with K being the data–set size, the inverse model ANN_{inv} provides an approximation of the inverse mapping $\hat{f}^{-1}(\cdot)$.

Once the ANN_{inv} has been trained, it can be used to predict input $\mathbf{X}^{pred}$ corresponding to the target output $\mathbf{Y}^{targ}$. In case the prediction of the inverse model is still not enough accurate, $\mathbf{X}^{pred}$ can be iteratively fine–adjusted by applying a gradient descent based optimization until a desired tolerance is achieved, thus:

$$\mathbf{X}_{j+1}^{pred} = \mathbf{X}_j^{pred} - \eta \nabla E(\mathbf{X}_j^{pred}) \tag{5.1}$$

with j being the iteration number and $\eta > 0$ the learning rate. The error $E(\mathbf{X}_j^{pred})$, defined as:

$$E(\mathbf{X}_j^{pred}) = MSE(\mathbf{Y}^{targ}, f(\mathbf{X}_j^{pred})) \tag{5.2}$$

is given by MSE computed between $\mathbf{Y}^{targ}$ and $f(\mathbf{X}_j^{pred})$, respectively the target output and the output given by the direct model.

Unfortunately, running the direct model $f(\cdot)$ at every iteration can be quite time–consuming, failing in being an ultra–fast method, or even impossible, especially if

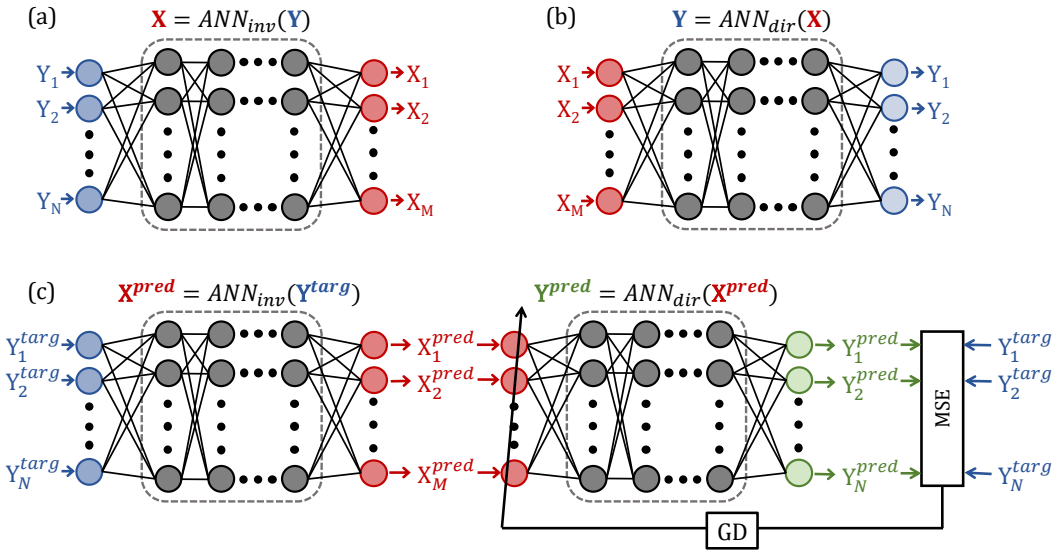


Fig. 5.1 Machine learning framework for inverse system design: (a) inverse ANN model, (b) direct ANN model and (c) fine–optimization routine based on gradient descent (GD) optimization.

the direct model is unknown. Therefore, instead of relying on the direct model $f(\cdot)$, to determine the output corresponding to $\mathbf{X}_j^{pred}$ we can use the direct ANN model ANN_{dir}, depicted in Fig. 5.1(b), whose high prediction accuracy has been extensively discussed in Chapter 4. The ANN_{dir} is completely independent from the ANN_{inv} and is trained with data–set $\mathscr{D}_{dir}^{K\times(M+N)} = \{\mathbf{X}_k, \mathbf{Y}_k | k = 1,\ldots,K\}$, which corresponds to the $\mathscr{D}_{inv}^{K\times(N+M)}$ data–set with swapped input and output. Accordingly, $E(\mathbf{X}_j^{pred})$ is redefined as:

$$E(\mathbf{X}_j^{pred}) = MSE(\mathbf{Y}^{targ}, ANN_{dir}(\mathbf{X}_j^{pred})). \tag{5.3}$$

The complete fine–optimization procedure is schematically summarized in Fig. 5.1(c). Furthermore, since $\mathbf{X}_j^{pred}$ should be already close to the optimal solution, not many iterations should be required for the gradient descent to converge, resulting in a low–complexity and computationally fast procedure.

Finally, to fully verify the accuracy of predicted inputs, we perform a further *validation* by feeding the original system with the predicted inputs (before and/or after fine–optimization). The system is the one used to generate the data–sets and can be simulative or experimental. Afterwards, the resulting outputs are compared with the target ones. Please note that this validation is different from the ANN validation adopted to keep under control ANN underfitting and overfitting.

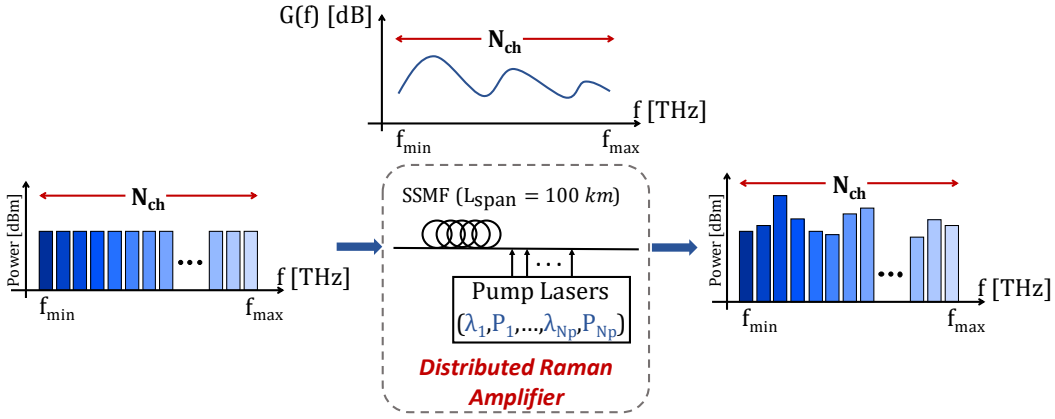


Fig. 5.2 Simulation set–up considered for the data–sets generation.

5.3 Machine Learning for Raman Amplifier Design over Synthetic Data

In this section, we investigate the ML–framework presented above in case of the design of RAs over C– and C+L–bands considering the simulation set–up depicted in Fig. 5.2.

We consider a single span distributed Raman amplifier operating in backward pumping scheme configuration and employing up to 5 pump lasers. As already discussed in Chapter 4, backward pumping is selected against forward and bidirectional pumping because of its greater practicality. The single span consist of $\mathrm{L_{span}} = 100$ km of SSMF with $\alpha_s = 0.2$ dB/km attenuation for the signal and $\alpha_p = 0.25$ dB/km attenuation for the pumps. The effective area is $\mathrm{A_{eff}} = 80\ \mu\mathrm{m}^2$, the chromatic dispersion $\mathrm{D} = 16.7$ ps/nm/km, the nonlinear coefficient $\gamma = 1.26\ \mathrm{W}^{-1}\mathrm{km}^{-1}$ and the Raman gain coefficient $\mathrm{C_r} = 0.4125\ \mathrm{W}^{-1}\mathrm{km}^{-1}$. Depending on the selected band, different pumping schemes are analyzed. They are reported in Table 5.1 together with ranges of wavelengths $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_M]^{\mathrm{T}}$ and powers $\mathbf{P} = [P_1, \ldots, P_M]^{\mathrm{T}}$ for Raman pumps, where M is the number of operating pumps that can assume values in $\mathrm{N_p} = \{2,3,4,5\}$.

At the input of the RA there is a comb of WDM channels carrying 0 dBm of power, operating at 32 GBd symbol rate and shaped at Nyquist limit, i.e. channel spacing equals to symbol rate, to fully load the transmission band with a flat optical spectrum.

Band	**C–band**			**C+L–band**
Number of Pumps	2 pumps	3 pumps	5 pumps	4pumps
P_1 [mW]	[0,400]	[0,300]	[0,160]	[0,200]
P_2 [mW]	[0,400]	[0,300]	[0,160]	[0,200]
P_3 [mW]	–	[0,300]	[0,160]	[0,200]
P_4 [mW]	–	–	[0,160]	[0,200]
P_5 [mW]	–	–	[0,160]	–
λ_1 [nm]	[1414,1449]	[1414,1437.3]	[1414,1428]	[1414,1441.8]
λ_2 [nm]	[1449,1484]	[1437.3,1460.6]	[1428,1442]	[1441.8,1469.5]
λ_3 [nm]	–	[1460.6,1484]	[1442,1456]	[1469.5,1497.3]
λ_4 [nm]	–	–	[1456,1470]	[1497.3,1525]
λ_5 [nm]	–	–	[1470,1484]	–

Table 5.1 Raman pumps power and wavelength ranges.

In case of C–band, defined between 192 THz and 196 THz (4 THz bandwidth), 125 channels are considered for a total WDM signal power of 21 dBm, while in case of C+L–band, defined between 185 THz and 196 THz (11 THz bandwidth), channels are 343 corresponding to 25 dBm of total power.

As done so far, the data–sets are generated by means of a numerical Raman solver (RS) solving the set of ODEs describing the evolution of power along the fiber in presence of Raman amplification with specified pumps configuration As done so far, by using a numerical Raman solver (RS), we solve the set of linear ordinary differential equations describing the evolution of power along the fiber in presence of Raman amplification with specified pumps configuration, to determine the gain profile $\mathbf{G} = [\mathrm{G}(f_1),\ldots,\mathrm{G}(f_N)]^{\mathrm{T}} = [\mathrm{G}_1,\ldots,\mathrm{G}_N]^{\mathrm{T}}$ at the end of the fiber, where N is the number of samples for the gain. Here we consider the on-off gain, defined between the power at the end of the fiber when the Raman pumps are ON and the power at the end of the fiber when the Raman pumps are OFF. Due to the resolution of the numerical Raman solver, corresponding to 100 GHz, $N = 40$ for the C–band and $N = 110$ for the C+L–band.

To generate the training and testing data–sets $\mathscr{D} = \{(\mathrm{G}_1^k,\ldots,\mathrm{G}_N^k,\lambda_1^k,\ldots,\lambda_M^k, P_1^k,\ldots,P_M^k)|k = 1,\ldots,K\}$, pumps wavelengths and powers are drawn from uniform distributions with ranges reported in Table 5.1. K is the data–set size, which for the training data–set is optimized between 2000 and 9000 to find a good trade–off

between accuracy and computational time, with the optimum corresponding to 4000 points. The testing data–set instead is composed of 5000 points containing gain profiles with values between 4 dB and 12 dB. Gains below 4 dB are too low, while gains above 12 dB do not satisfy the moderate pumping regime condition [26], corresponding to the 60% of 20 dB total loss of the span that prevents from gain saturation effects and loss of efficiency.

During the training stage, the inverse ANN model (ANN_{inv}) learns the mapping between the Raman gain $\mathbf{X} = [\mathrm{G}_1, \ldots, \mathrm{G}_N]^{\mathrm{T}}$ and the laser pumps configuration $\mathbf{Y} = [P_1, \ldots, P_M, \lambda_1, \ldots, \lambda_M]^{\mathrm{T}}$.

Once trained, the ANN_{inv} can be used to predict the pumps configuration to get a target gain profile $\mathbf{G}^{targ}$. To assess the ANN_{inv} prediction performance, instead of comparing the predicted and the true values of pumps powers and wavelengths, it would be interesting to compare the target gain profile and the one resulting from the predicted pumps configuration. For this purpose, we use again the numerical RS, which now at the input receives the predicted pumps powers and wavelengths. This is the process called *validation* and the gain profile corresponding to the predicted pumps configuration is referred to as validated gain $\mathbf{G}^{valid}$. The prediction accuracy is evaluated computing the RMSE and the $\mathrm{Err}_{\mathrm{MAX}}$ between the target gain profile $\mathbf{G}^{targ}$ and the gain profile resulting from the validation through the Raman solver $\mathbf{G}^{valid}$, as in Eqs. (5.4) and (5.5):

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(\mathrm{G}_n^{valid} - \mathrm{G}_n^{targ})^2} \tag{5.4}$$

$$\mathrm{Err}_{\mathrm{MAX}} = \max_{n\in\{1,\ldots,N\}} |\mathrm{G}_n^{valid} - \mathrm{G}_n^{targ}|. \tag{5.5}$$

To further improve the prediction accuracy of the inverse design, a further fine–optimization of the pumps configuration can be carried out. Then, the fine–optimized pumps powers and wavelengths are given at the input of the Raman solver for the gain profile validation.

5.3.1 Raman Amplifier Design over C–band

We first analyze the ability of the inverse neural network model ANN_{inv} in designing Raman amplifiers in the C–band, for which pumping schemes with 2, 3 and 5 pumps

ANN model	ANN_{inv}		ANN_{dir}	
Band	C	C+L	C	C+L
Training algorithm	RP	RP	LM	LM
Number of hidden layers N_{HL}	7	4	2	2
Number of hidden nodes N_{HN}	800	900	10	10
Activation function	tanh	logsig	tanh	logsig
Model combination parameter N_{ANN}	200	200	1	1

Table 5.2 ANN topologies and hyper–parameters.

are considered. To find the optimal network topology that minimizes the validation error, we perform hyper-parameters optimization. In this case, best performance are obtained when random projection (RP) method is used, with a great advantage in computational time. It results in an ultra–fast training, which also speeds up the selection of the optimal network topology following the optimization of hyper–parameters. The topology considered for all the three pumping schemes is reported in Table 5.2: $N_{\mathrm{HL}} = 7$ hidden layers, $N_{\mathrm{HN}} = 800$ hidden nodes and hyperbolic tangent as activation function of each node. Model averaging is performed over $N_{\mathrm{ANN}} = 200$ parallel and independent ANN.

Once the ANN is trained, it can be used to predict pumps configuration for flat, tilted or arbitrary gain profiles. If there is still not a sufficient agreement between the target gain profile and the one given by the predicted pumps configuration, to further improve the design of the Raman amplifier, pumps powers and wavelengths are fine–adjusted using the ANN_{dir} in combination with the gradient descent algorithm. As shown in Table 5.2, for all the three pumping schemes, the direct model ANN_{dir} is trained using the Levenberg–Marquardt learning algorithm considering $N_{\mathrm{HL}} = 2$ hidden layers, $N_{\mathrm{HN}} = 10$ hidden nodes and hyperbolic tangent as activation function. In this case, no model combination is performed as the weights of the ANN are iteratively optimized based on error back–propagation. In addition, the training of only one ANN model, helps in speeding up the entire process.

Design of Arbitrary Gain Profiles

First, we use the ANN_{inv} to predict the pumps powers and wavelengths to obtain the arbitrary gains of the testing data–set. In Fig. 5.3, we report the probability density functions of the RMSE and of the $\mathrm{Err}_{\mathrm{MAX}}$ between the target gain profiles $\mathbf{G}^{targ}$ and

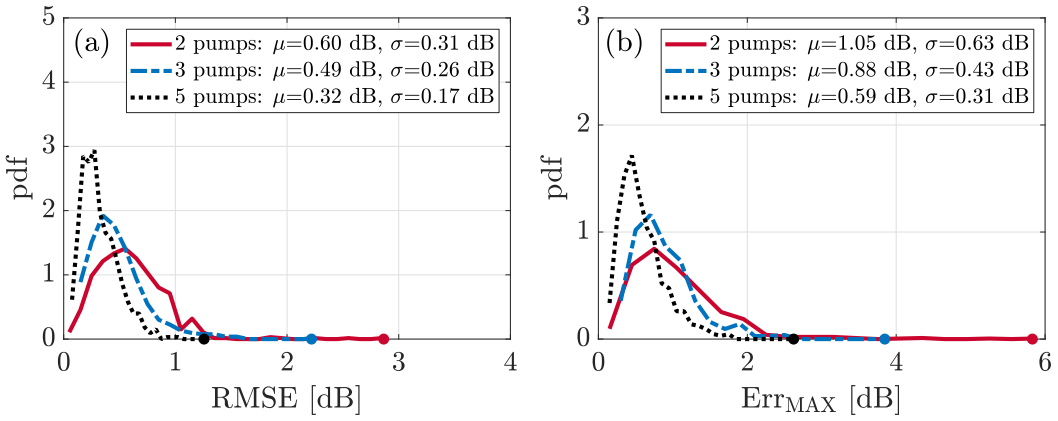


Fig. 5.3 Probability density functions (pdfs) of (a) RMSE and (b) Err_{MAX} computed between the target *arbitrary* gain profiles $\mathbf{G}^{targ}$ and the validated gain profiles $\mathbf{G}^{valid}$ over the C–band for two, three and five pumps.

the validated gain profiles $\mathbf{G}^{valid}$ given by the Raman solver for the predicted pumps powers and wavelength for all the three pumping schemes. Looking at the RMSE (Fig. 5.3(a)), we can see that its mean value is between 0.60 dB and 0.32 dB, while the standard deviation is between 0.31 dB and 0.17 dB, from which we can observe a slightly higher accuracy in case of 5 pumps. Also in case of Err_{MAX} (Fig. 5.3(b)), the trend is similar, with better performance in case of 5 pumps. Here the average Err_{MAX} assumes values between 1.05 dB and 0.59 dB and the standard deviation between 0.63 dB and 0.31 dB. Although for most of the cases the errors are pretty acceptable, there are some cases for which the Err_{MAX} is quite large, in particular for 2 and 3 pumps with values up to 5.8 dB and 3.8 dB, respectively.

For application not requiring very tight accuracy, the predictions provided by our ANN_{inv} are accurate enough, especially when 5 pumps are considered. Instead, if more accurate predictions are needed, we can perform a fine–optimization of pumps powers and wavelengths, as described in the section dedicated to the machine learning framework. In Fig. 5.4, we compare the pdf of the RMSE and the Err_{MAX} before and after fine–optimization for 2, 3 and 5 pumps. We can observe a substantial improvement of performance after the fine–optimization, with average values of RMSE and Err_{MAX} always below or equal to 0.12 dB and 0.29 dB, respectively. Moreover, also the tails of pdfs reduce and the Err_{MAX} never exceeds 1 dB.

The results obtained so far show that the ANN_{inv} is able to learn the inverse mapping and to predict pumps configurations for the target arbitrary gains with quite good accuracy, especially when fine–optimization is applied.

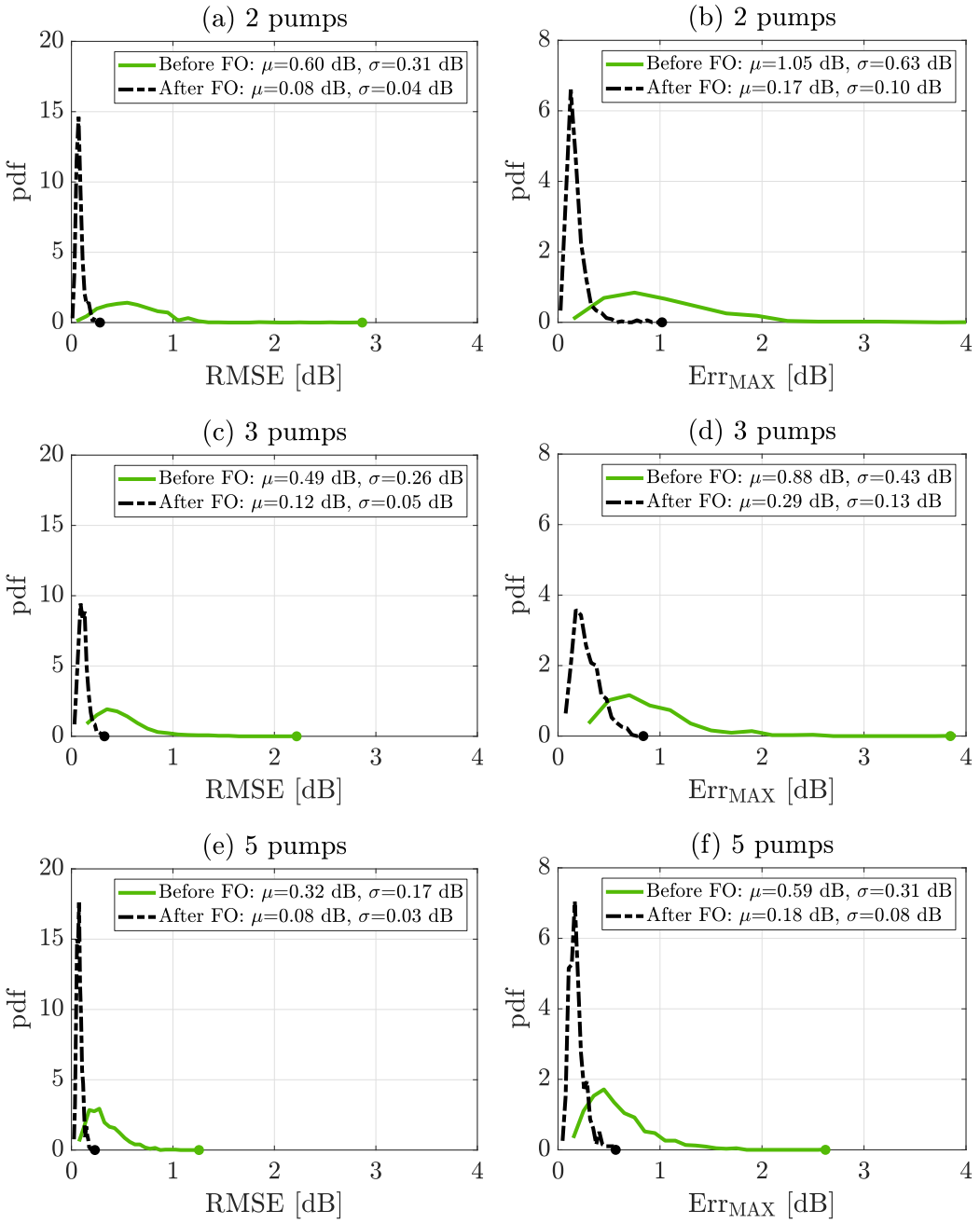


Fig. 5.4 Probability density functions (pdfs) of RMSE (left column) and Err_{MAX} (right column) of the validated *arbitrary* gain profiles over the C–band given by the predicted pump powers and wavelengths configurations before and after fine–optimization (FO) for: (a)-(b) two, (c)-(d) three and (e)-(f) five pumps.

Design of Flat and Tilted Gain Profiles

After the analysis over arbitrary gains, in this section, we explore the capability of the inverse model to design flat and tilted gain profiles. In our analysis, gain levels are in the range [5,12] dB and the slope of tilted gains is -0.25 dB/THz, corresponding to a 1 dB of negative tilt along the C–band.

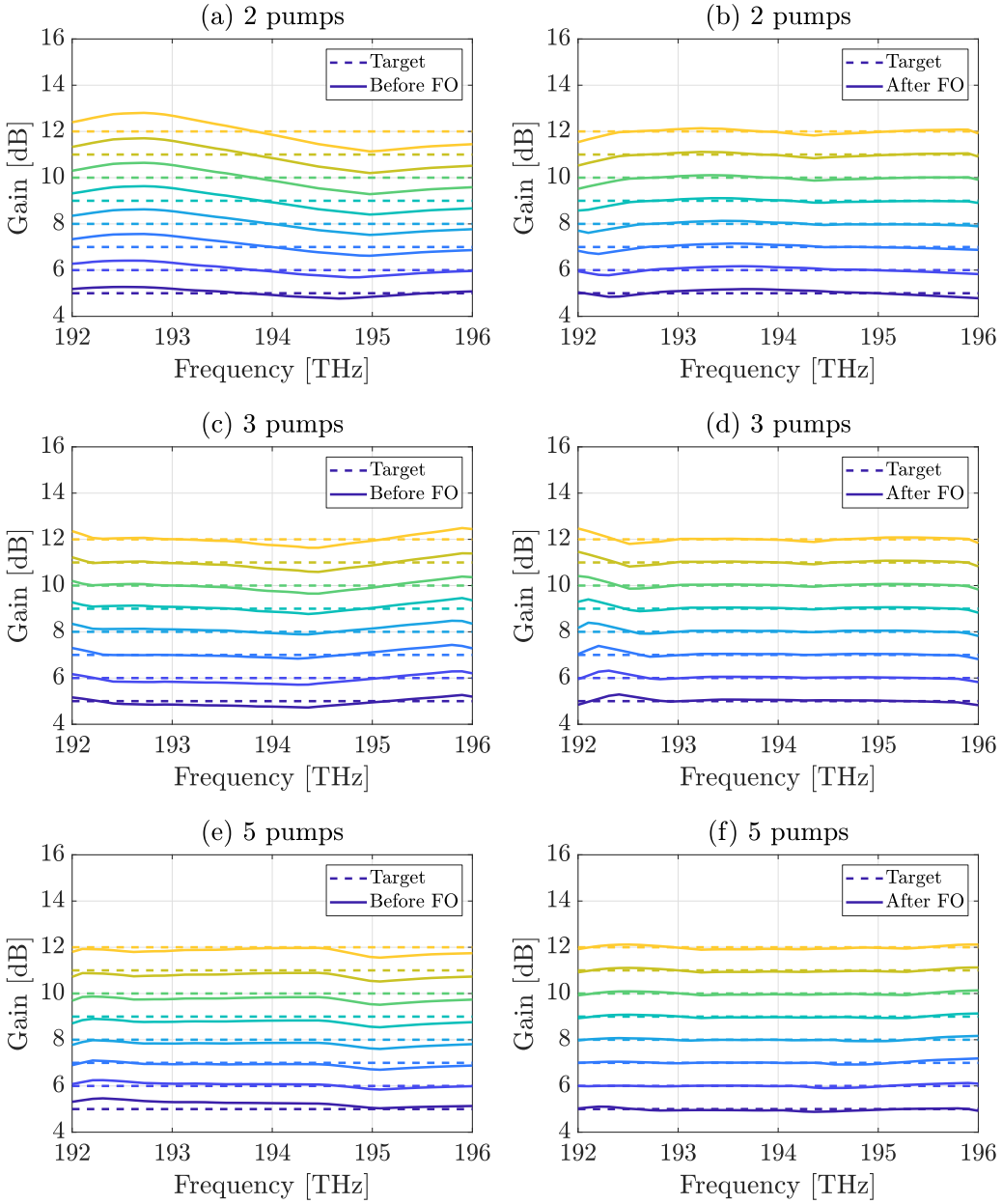


Fig. 5.5 Target and validated *flat* gain profiles over the C–band before and after fine–optimization (FO) in case of: (a)-(b) two, (c)-(d) three and (e)-(f) five pumps.

The qualitative representation of the target and validated *flat* gain profiles resulting from the pumps configurations predicted by the ANN_{inv} are illustrated in Figs. 5.5(a), 5.5(c) and 5.5(e) for two, three and five pumps, respectively. From there we can see that the inverse model is able to determine the pumps allocation with quite good precision, with increasing matching between target and validated gain profiles when the number of pumps increases, showing smoother profiles as well. In fact, although it is not trivial to infer the final gain profile, as it is given by the nonlinear interactions among pumps different in wavelengths and powers, on the other hand more pumps correspond to more degrees of freedom that can be tuned to obtain the target gain. Deviations are all below 1 dB, with the larger values observed for lower frequencies and in particular in case of 2 pumps. Applying fine–optimization, flat gains design further improves, as shown in Figs. 5.5(b), 5.5(d) and 5.5(f), where a higher flatness is observable.

A more precise analysis of the pumps prediction accuracy is reported in Fig. 5.6, where the RMSE and the Err_{MAX} are plotted with respect to the gain levels. If fine–optimization is not performed, in case of 2 pumps RMSE (Fig. 5.6(a)) and Err_{MAX} (Fig. 5.6(b)) increase with gain levels, reaching values up to 0.6 dB and 0.86 dB, respectively. Instead, in case of 3 and 5 pumps RMSE and Err_{MAX} values are much lower and do not show a specific dependency on gain levels. In particular, the RMSE is around $\sim$0.2 dB and the Err_{MAX} is always below 0.5 dB.

When pumps configurations are further fine–optimized, the RMSE is always much lower than 0.2 dB for all the three pumping schemes and furthermore it is flat with respect to gain levels. Although for 3 pumps there is not an appreciable improvement in terms of Err_{MAX} when performing the fine–optimization, in case of 2 and 5 pumps the reduction is significant. Indeed, for 2 pumps the Err_{MAX} never exceeds 0.5 dB, while for 5 pumps it always stays below 0.2 dB.

These results lead us to make some considerations about the number of pumps needed to design Raman amplifiers providing flat gains over the entire C–band. In fact, depending on the desired accuracy and on the agreed complexity, flat gains can be obtained with high accuracy considering all the analyzed pumping schemes, also with 2 pumps, if fine–optimization is performed.

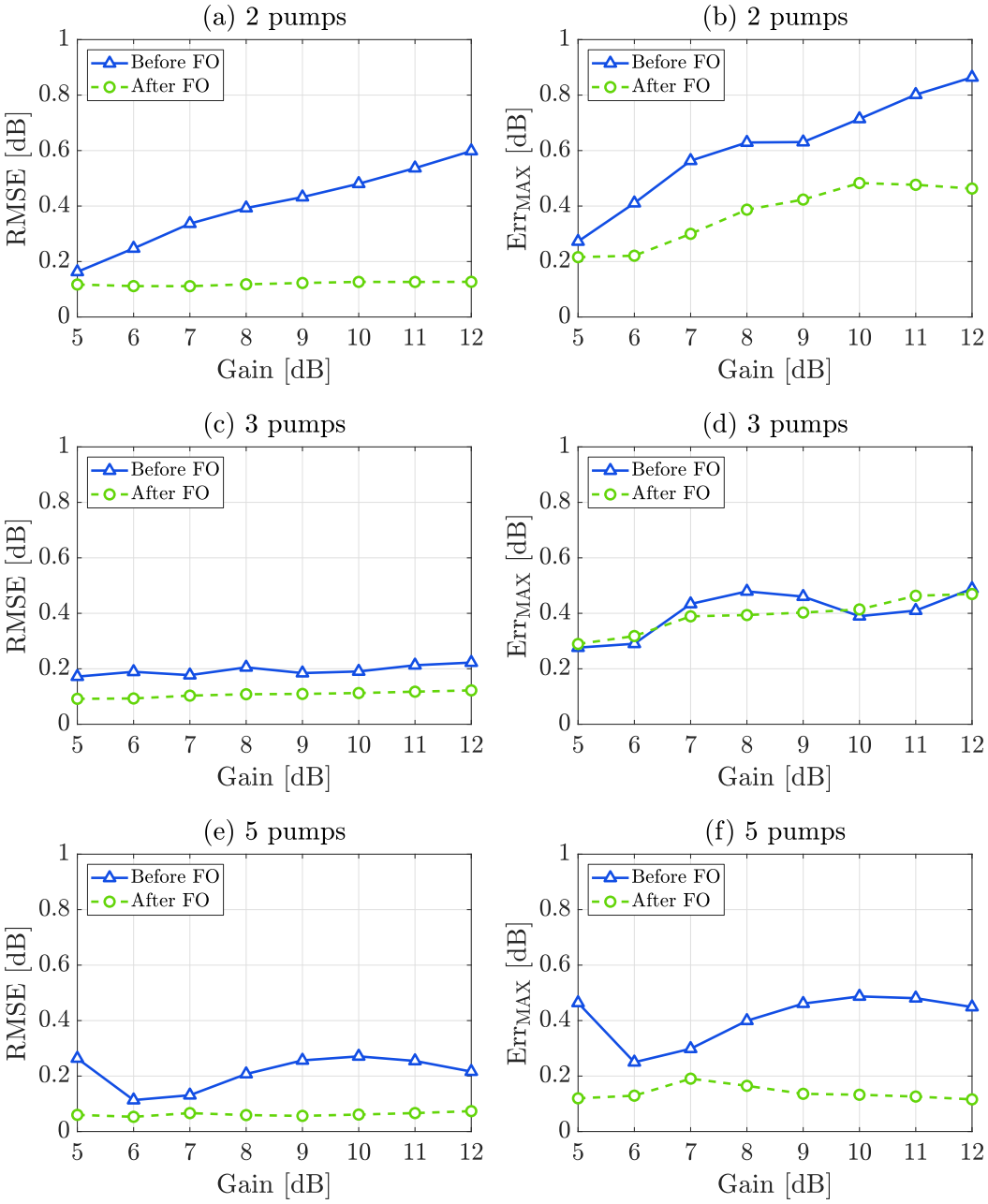


Fig. 5.6 RMSE and Err_{MAX} with respect to gain levels of *flat* gains over the C–band before and after fine–optimization (FO) for: (a)-(b) two, (c)-(d) three and (e)-(f) five pumps.

Moving to the tilted gain profiles, in Figs. 5.7(a), 5.7(c) and 5.7(e) we plot the target and validated profiles given by the pumps powers and wavelengths predicted by the ANN_{inv} for 2, 3 and 5 pumps, respectively.

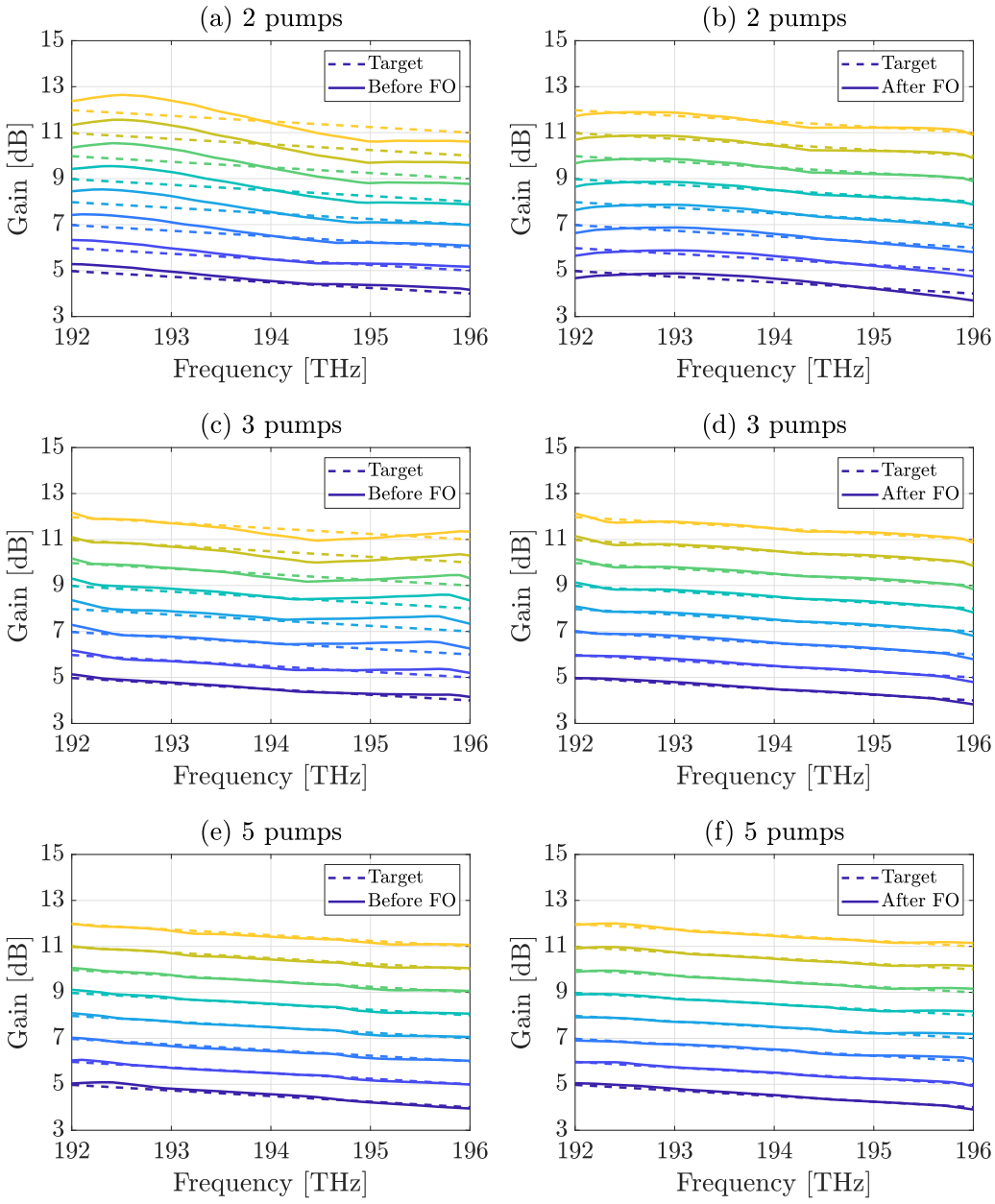


Fig. 5.7 Target and validated *tilted* gain profiles over the C–band before and after fine–optimization (FO) in case of: (a)-(b) two, (c)-(d) three and (e)-(f) five pumps.

Performance similar to the flat gain case is observed, since, also for tilted gains, higher deviations occur when 2 pumps are considered, especially for higher gain levels. Remarkably, in case of 5 pumps the validated gain profiles overlap the target ones almost perfectly without the need of fine–adjusting pumps configurations.

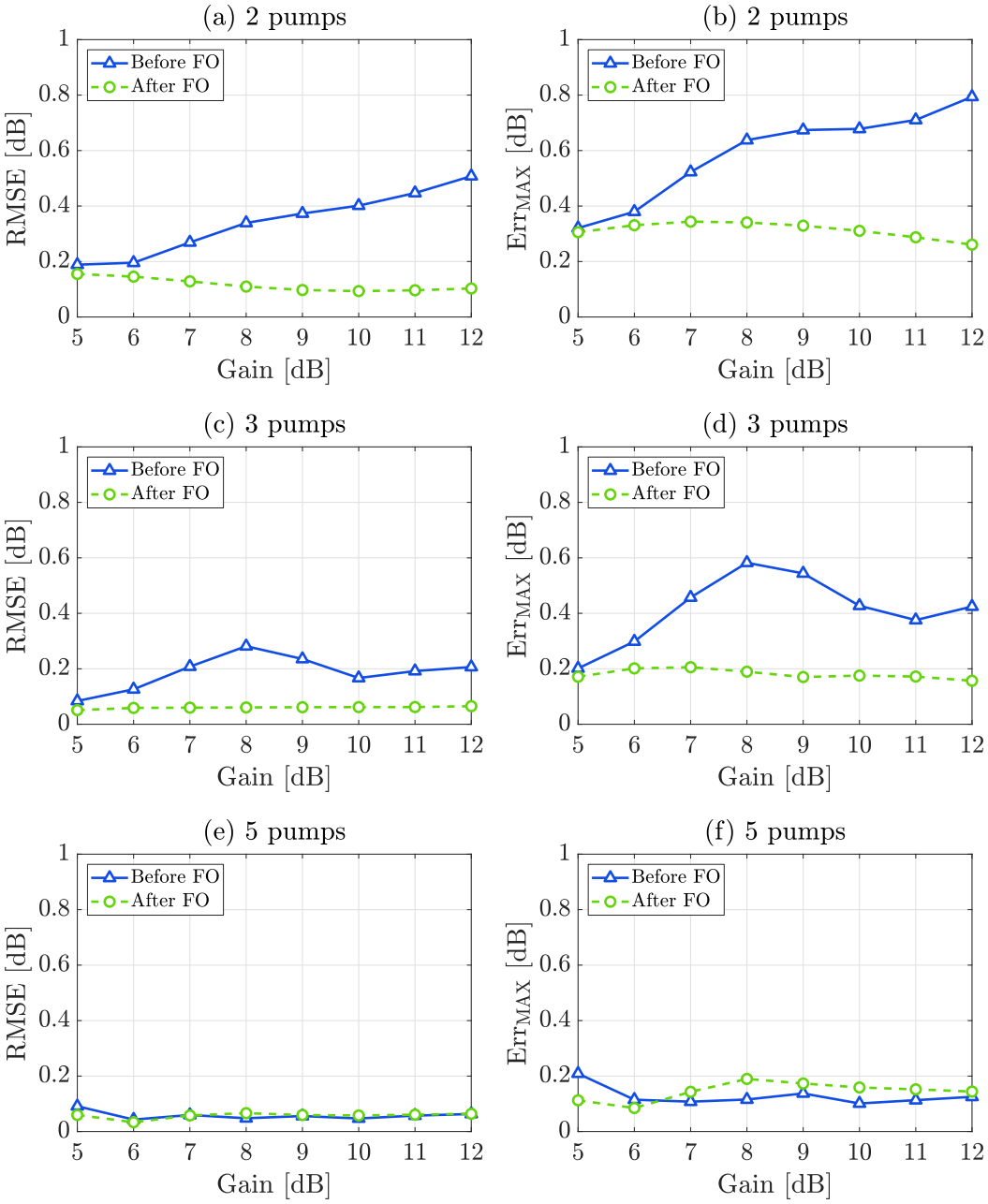


Fig. 5.8 RMSE and Err_{MAX} with respect to gain levels of *tilted* gains over the C–band before and after fine–optimization (FO) for: (a)-(b) two, (c)-(d) three and (e)-(f) five pumps.

The validated gains after fine–optimization are instead shown in Figs. 5.7(b), 5.7(d) and 5.7(f), where visually we can perceive an improvement in case of 2 and 3 pumps. The considerations made so far are confirmed if we look at the RMSE and Err_{MAX} plotted with respect to gain levels (Fig. 5.8). The increase of accuracy

following the fine–optimization is considerable in case of 2 pumps since the RMSE drops down to values below 0.2 dB and the $\mathrm{Err}_{\mathrm{MAX}}$ is always below 0.4 dB for each gain level. Similarly for 3 pumps, for which the RMSE assumes values under 0.1 dB and the $\mathrm{Err}_{\mathrm{MAX}}$ is around 0.2 dB after the fine–optimization of pumps configurations.

In case of 5 pumps, already before fine–optimization we observe a good agreement between target and validated gain profiles. Otherwise than expected, the fine–optimization does not provide any further improvement. Indeed for gains beyond 7 dB it results in higher $\mathrm{Err}_{\mathrm{MAX}}$ values (Fig. 5.8(f)). This might be explained by the fact that the ANN_{dir} used in the fine–optimization process is an approximation of the equation describing the Raman amplification effect. Therefore, the possible model mismatch between the two can propagate and affect the fine–optimization introducing some errors, which is sign of a possible overfitting.

Like for the flat gains case, also here for all the three pumping schemes (2, 3 and 5 pumps) the inverse model ANN_{inv} is able to design RAs to obtain tilted gain profiles with high accuracy. In particular, when the number of pumps is reduced, the high accuracy is guaranteed by further fine–adjusting pumps powers and wavelengths.

5.3.2 Raman Amplifier Design over C+L–band

After the analysis of the proposed machine learning framework in the C-band, in this section we investigate its robustness over the more challenging C+L–band. Here, only one pumping scheme consisting of 4 pumps is considered. Also in this case, the topology and the parameters for the inverse modeling are optimized and are reported in Table 5.2. The ANN_{inv} is trained using the random–projection method and the model consists of $N_{\mathrm{HL}} = 4$ hidden layers, $N_{\mathrm{HN}} = 900$ hidden nodes and logistic sigmoid as activation function. Model combination parameter over which performing model average is $N_{\mathrm{ANN}} = 200$. Regarding the direct model ANN_{dir} used for the fine–optimization, it is trained with the back–propagation method implementing the Levenberg–Marquardt algorithm, like in case of C–band only. The number of hidden layers is $N_{\mathrm{HL}} = 2$, the number of hidden nodes $N_{\mathrm{HN}} = 10$ and the activation function is still logistic sigmoid. Here as well, no model averaging is performed.

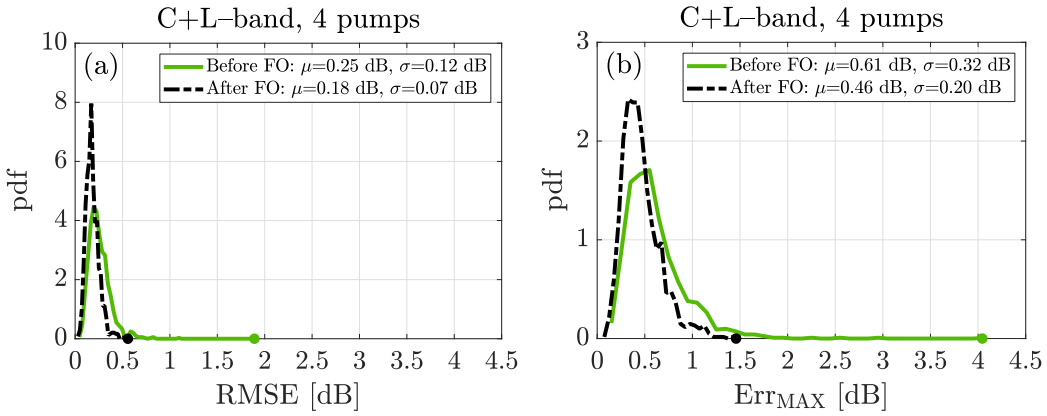


Fig. 5.9 Probability density functions (pdfs) of (a) RMSE and (b) Err_{MAX} computed between the target *arbitrary* gain profiles $\mathbf{G}^{targ}$ and the validated gain profiles $\mathbf{G}^{valid}$ over the C+L–band for four pumps.

Design of Arbitrary Gain Profiles

We start analyzing the ability of the inverse model in designing the arbitrary gains composing the testing data–set. We recall that the testing data–set has been generated considering pumps powers and wavelengths reported in Table. 5.1 and only gains within 4 dB and 12 dB have been selected. Fig. 5.9 presents the probability density functions of RMSE and Err_{MAX} computed between the target gain profiles and the gain profiles obtained validating, by means of the Raman solver, the predicted pumps configurations both before and after fine–optimization

Already without fine–optimization the ANN_{inv} is able to learn the inverse mapping with quite high accuracy. Indeed, the RMSE has average 0.25 dB and standard deviation 0.12 dB, while the Err_{MAX} has average 0.61 dB and standard deviation 0.32 dB. After fine–optimization, besides the slight improvement of average and standard deviation values, which reduce to 0.18 dB and 0.07 dB respectively for the RMSE, and to 0.49 dB and 0.20 dB respectively for the Err_{MAX}, the most relevant reduction is observed in the tails of the pdfs. In case of RMSE, the maximum value from almost 2 dB drops down to $\sim$0.5 dB, while the maximum value of Err_{MAX} from $\sim$4 dB decrease to 1.5 dB.

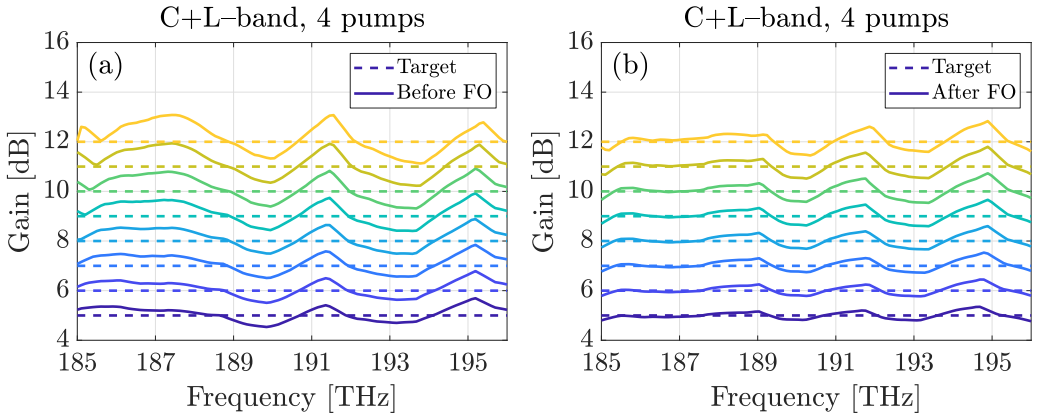


Fig. 5.10 Target and validated *flat* gain profiles over the C+L–band before and after fine–optimization (FO) in case of four pumps.

Design of Flat and Tilted Gain Profiles

Finally, we study the ability of the proposed ML framework to design RAs for desired flat and tilted gains also over the C+L–band. Like the C–band case, the gain levels are between 4 dB and 12 dB, while the slope of the tilted gains is now -0.09 dB/THz to target a 1 dB negative tilt along the entire 11 THz C+L–band.

In Figs. 5.10(a) and 5.10(b), the target and validated gain profiles given by the predicted pumps powers and wavelengths using only ANN_{inv} (before fine–optimization) and using $ANN_{inv} + ANN_{dir}$ (after fine–optimization) are illustrated, respectively. Surprisingly, already with 4 pumps it is possible to obtain quite flat gains, which are further improved in flatness when fine–optimization is performed. The corresponding RMSE and Err_{MAX} with respect to the gain levels before and after fine–optimization are instead in Fig. 5.11.

Unlike the C–band case, we can observe that the RMSE and the Err_{MAX} increase with gains also after fine–optimization, but there is still an advantage in fine–adjusting pumps configurations, since it allows to reduce the maximum RMSE value from 0.60 dB down to 0.33 dB and the maximum Err_{MAX} from 1.1 dB to 0.83 dB.

Similar considerations can be drawn in case of tilted gains, which are shown in Fig. 5.12. The performance of the inverse model is evaluated in Fig. 5.13, where the RMSE and the Err_{MAX} are plotted with respect to the gain levels before and after fine–optimization.

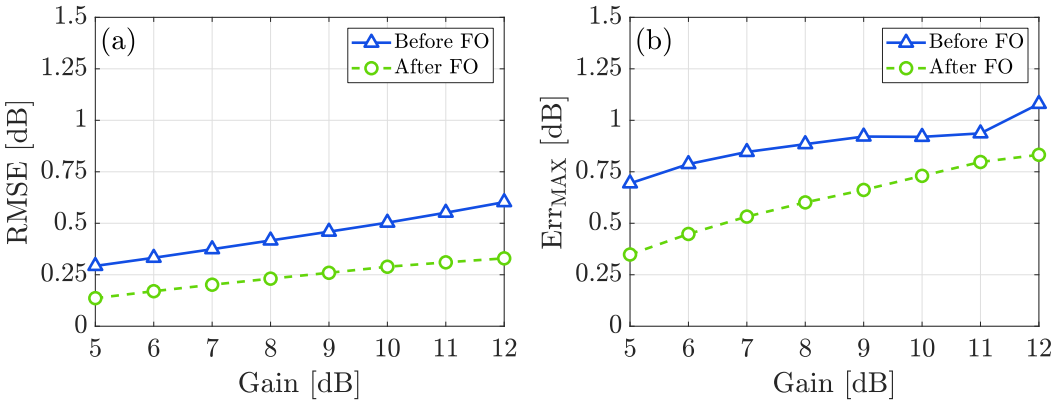


Fig. 5.11 RMSE and Err_{MAX} with respect to gain levels of *flat* gains over the C+L–band before and after fine–optimization (FO) for four pumps.

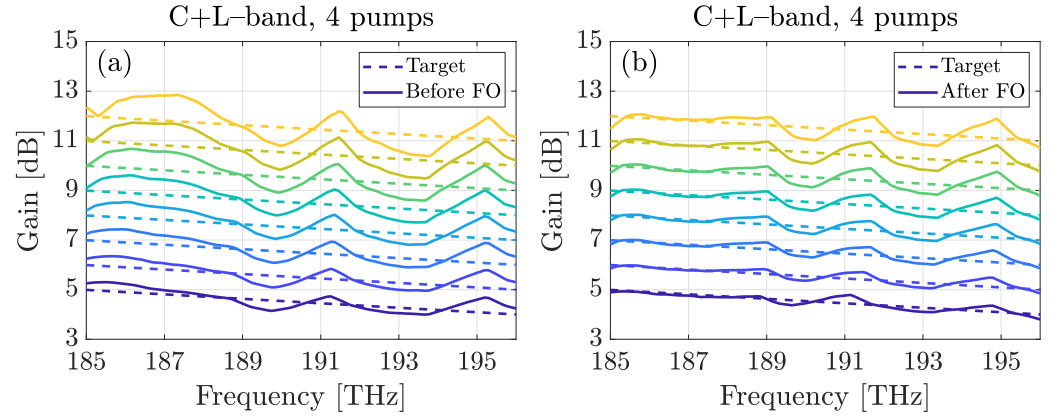


Fig. 5.12 Target and validated *tilted* gain profiles over the C–band before and after fine–optimization (FO) in case of four pumps.

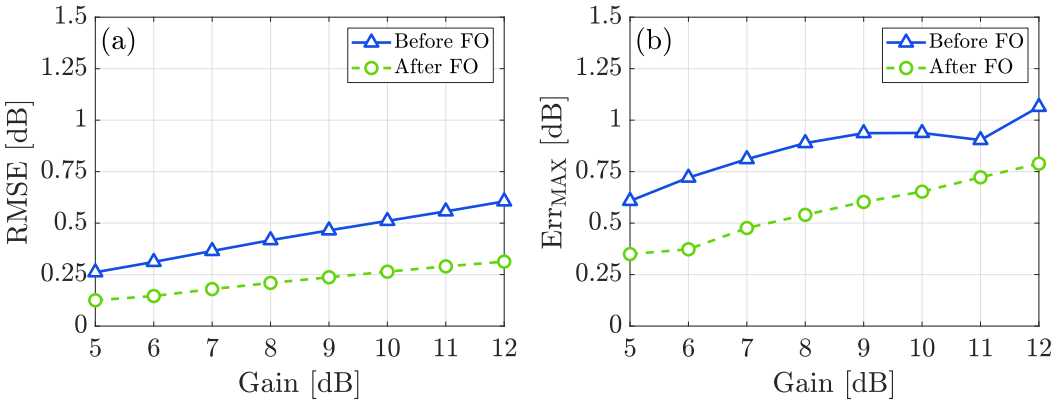


Fig. 5.13 RMSE and Err_{MAX} with respect to gain levels of *tilted* gains over the C–band before and after fine–optimization (FO) for four pumps.

Results are similar to the flat gains case: RMSE and Err_{MAX} are always below 0.61 dB and 1.1 dB respectively when the fine–optimization is not performed, while after fine–optimization these values decrease down to 0.31 dB and 0.79 dB.

Even though the results are already good, there are still some ripples in high frequencies region, especially when the gain level is higher. Thus, we believe that a larger number of pumps would cover the entire C+L–band more efficiently, improving the design of flat and titled gain profiles.

In conclusion, with performance similar to the C–band case, we have shown that the proposed ML framework for the inverse system design is able to design RAs for arbitrary, flat and titled gain profiles with high accuracy along the C+L–band as well, despite being a more challenging task due to the larger bandwidth.

5.3.3 Raman Amplifier Design over C+L–band Considering Partial Spectral Loads

In the analysis carried out so far, only full–spectral loads are considered at the input of the fiber, but in real networks traffic patterns can change over time and very rarely full–load is reached. Like in Chapter 4, where for the direct model ANN_{dir} the load–aware neural network (LA-NN) has been introduced to account for partial–loads, also for the inverse mapping we can upgrade the inverse model ANN_{inv} in order to have spectral loads included as well at the input of the ANN.

Data–sets and Machine Learning Framework

For this investigation, we assume the same simulation setup considered in section 4.3 shown in Fig. 4.9, where the study on different partial–loads has been carried out on a sub–band basis. For this reason, it is not necessary to generate further data, but we can use the training and testing *partial–load* data–sets generated in section 4.3 for the direct ANN, with the difference that, in the inverse ANN, gain profiles and pumps powers are at input and output of the ANN, respectively. Just to recall, the system operates in the 11–THz C+L–band and considers a distributed RA made of a 100-km single span of SSMF with five counter propagating laser pumps. The 220 WDM channels in the 50–GHz grid given at the input of the RA are grouped together to form 22 sub–bands, i.e. each sub–band contains 10 channels and is 500–GHz

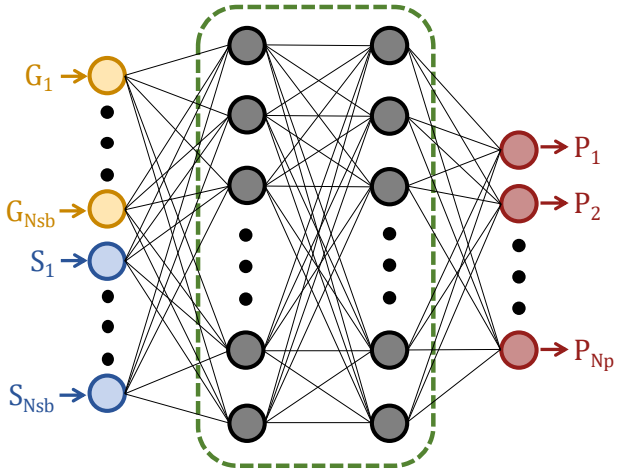


Fig. 5.14 Schematic of the inverse load-aware neural network ANN_{inv}^{LA}.

wide. Pumps frequencies and powers are those reported in Table 4.4. Each sub–band can be turned ON or OFF and when it is ON, it carries 10 mW of power. The training data–set is made of 11000 different samples, generated for different pumps power configurations and input spectral loads (based on classes and sub-classes). Similarly for the testing data–set.

The upgraded ANN model for inverse design accounting for the input spectral load is shown in Fig. 5.14, which for simplicity is referred to as ANN_{inv}^{LA}. Consequently, the inverse model not taking the spectral load at its input can be indicated as ANN_{inv}^{LU}.

According to the upgraded model, the ANN_{inv}^{LA} learns the inverse mapping $\mathbf{P} = f^{-1}(\mathbf{G},\mathbf{S})$, where $\mathbf{P} = [P_1,\ldots,P_{\mathrm{N_p}}]^{\mathrm{T}}$ is the vector of pump powers, $\mathbf{G} = [\mathrm{G}_1,\ldots,\mathrm{G}_{\mathrm{N_{sb}}}]^{\mathrm{T}}$ the vector of $\mathrm{N_{sb}}$ points taken along the gain profile and $\mathbf{S} = [\mathrm{S}_1,\ldots,\mathrm{S}_{\mathrm{N_{sb}}}]^{\mathrm{T}}$ the vector of $\mathrm{N_{sb}}$ points taken along the input signal power profile. $\mathrm{N_p}$ and $\mathrm{N_{sb}}$ are respectively the number of RA laser pumps and the number of frequency samples for input spectral load and gain (which corresponds to the number of sub–bands ON).

Like in section 4.4, where direct LA-NN has been extended to support multiple fibers, also here we use Tensorflow and Keras libraries for ML framework. The ANN model is a fully–connected network with 2 hidden layers and 300 hidden nodes per layer. The input layer is composed of 44 nodes, while the output layer has 5 nodes, corresponding to the number of RA pump lasers. The activation function

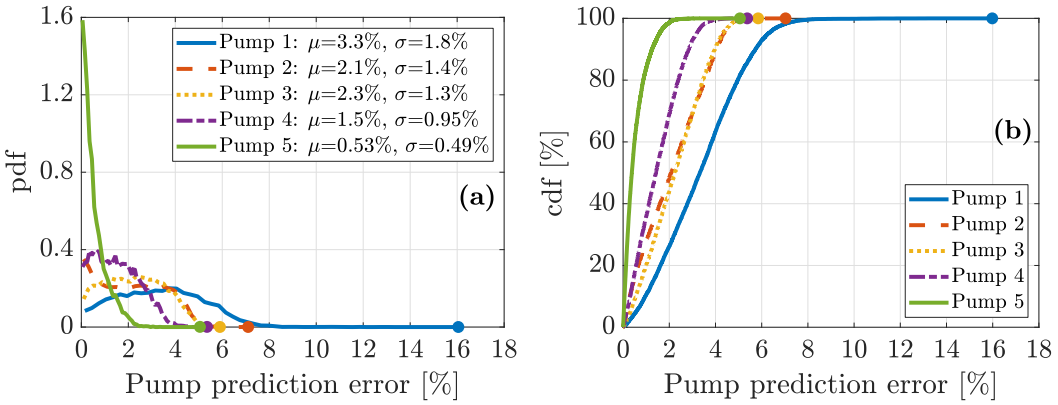


Fig. 5.15 (a) Probability density function (pdf) and (b) cumulative distribution function (cdf) of pump power prediction error for all five pump lasers.

used at input layer and hidden layers is the ReLU, while linear function is used at output layer. ANN weights are initialized based on the Keras default Glorot uniform initialization. For ANN training, we consider a validation split equals to 0.1, which means that 90% of training data–set is actually used for training and 10% for validation. To speed-up the training process, mini–batches of size $N_{batch} = 100$ are assumed. For weights optimization we consider Adam optimizer with initial learning rate $\eta = 0.001$, whereas the loss function is the MSE. After 5000 epochs, the training is stopped.

Testing and Validation Results

After training, to assess the reliability of pumps power predictions, we test the ANN on the testing data–set of arbitrary gain profiles. Fig. 5.15 shows the probability density function and the cumulative distribution function of the absolute error relative to the pump power upper bound for each pump laser. Pump lasers are ordered from higher to lower frequencies ($f_{p1,\ldots,p5}$=[210.37, 207.14, 204.01, 200.97, 198.03] THz). There, we can observe that the errors reduces with increasing pump number, i.e. with decreasing pump frequency, and the worst prediction performance is indeed experienced by pump 1, the one at highest frequency (210.37 THz).

This behavior is also observable if we plot the predicted pump powers with respect to the target true values in Fig. 5.16, where we can see that in general for pumps at higher frequency, the power is underestimated, which make us think that the ANN is not properly learning its actual contribution to the overall gain profile.

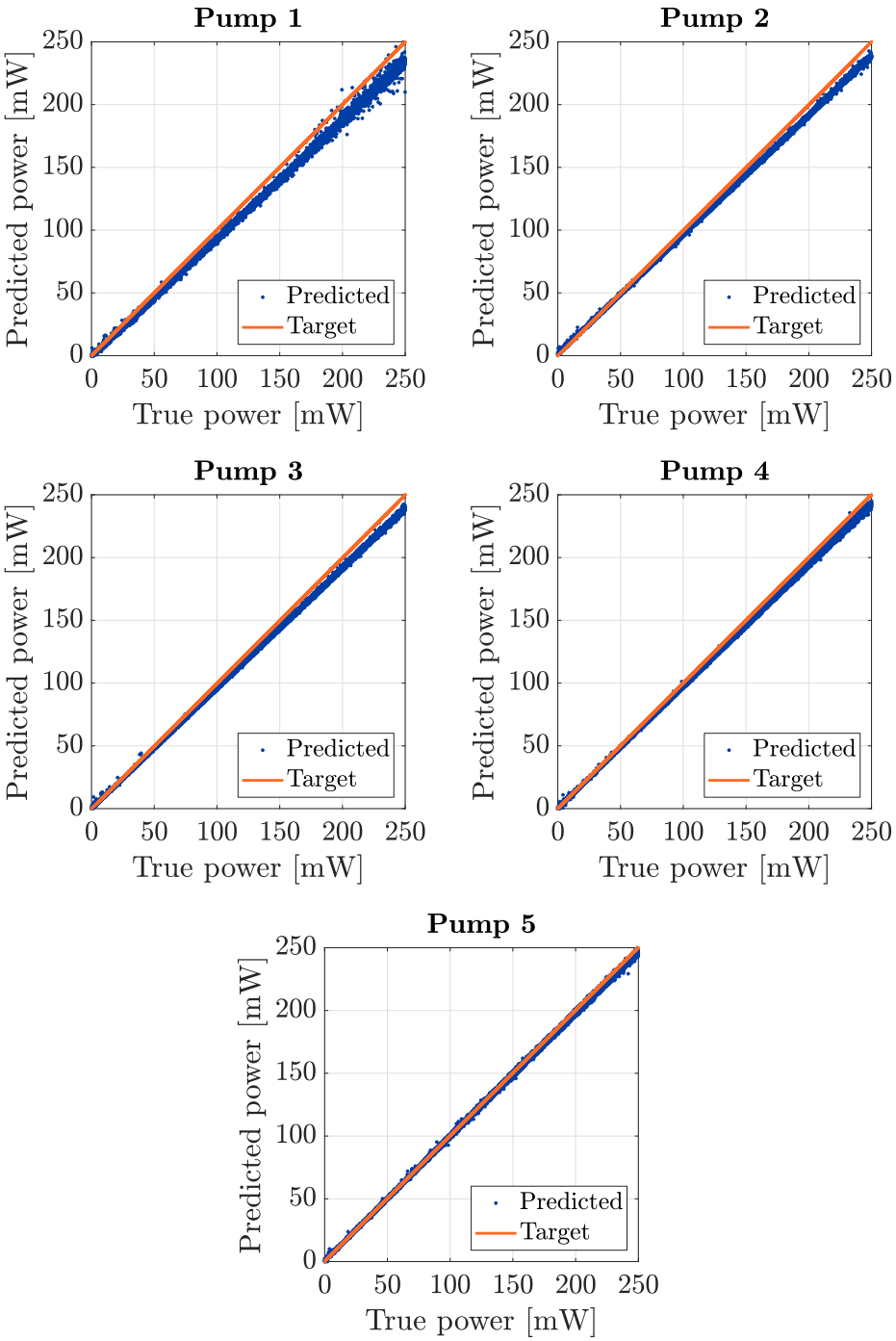


Fig. 5.16 Predicted pumps powers plotted with respect to the target true values.

Indeed, the introduction of partial spectral loads makes the RA inverse design even more challenging and because of the nonlinear interactions occurring due to SRS (signal-to-signal, pump-to-pump and signal-to-pump interactions), it is not easy

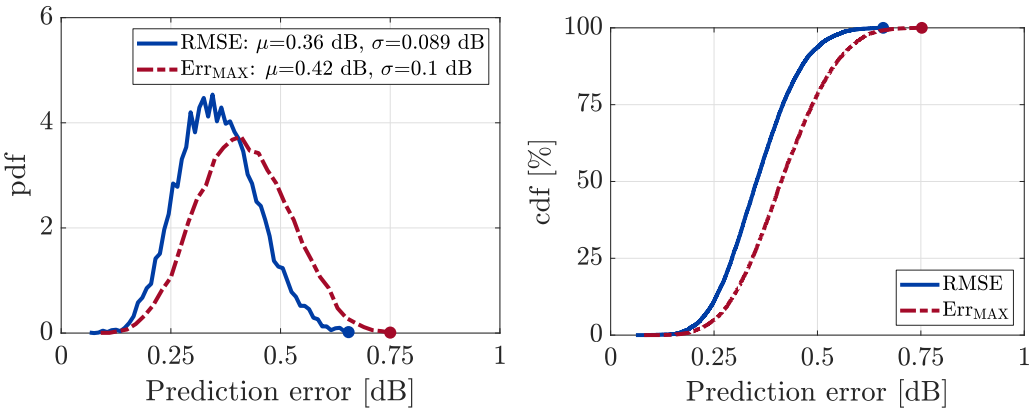


Fig. 5.17 (a) Probability density function (pdf) and (b) cumulative distribution function (cdf) of validated gain profiles given by the predicted pumps powers configurations.

to find a rigorous explanation to this phenomenon. We can try to explain it by recalling that in composite RAs, pumps at higher frequency not only are subjected to pump-signal interactions, but also to pump-pump interactions, providing a power contribution to pumps at lower frequency as well.

A possible reason might be related to the frequency position of the pumps. To explain it we can look at Fig. 2.4, showing an example case of flat gain profile assuming exactly the pumps and their frequency allocation considered here, also ordered for decreasing frequency. From there we can see that the gain contribution provided by pump 1 (210.37 THz) is limited to the upper limit of the C–band, with the gain peak at ∼197.36 THz. Moreover, pump 1 and pump 5 are separated by almost one Stokes shift (∼13 THz), therefore pump 5 is actually pumped by pump 1 due to pump-pump interactions, with part of its power transferred to pump 5. On top of that, signal-to-signal interactions occurring between high frequencies sub–bands and low frequencies sub–bands, with a power transfer from higher to lower frequencies, further contribute to the gain in the lower frequencies range. As the gains also depends on the input spectral load, this explains why the power of pump 1 is underestimated, since a power contribution is expected following the ISRS occurring among sub–bands. Also powers of pumps 2 and 3 are a little bit underestimated, for the same reason as for pump 1, with the difference that for them the phenomenon is reduced.

The predicted pumps powers configurations are further validated using the numerical Raman solver [45] to determine the corresponding gain profiles. Then, we

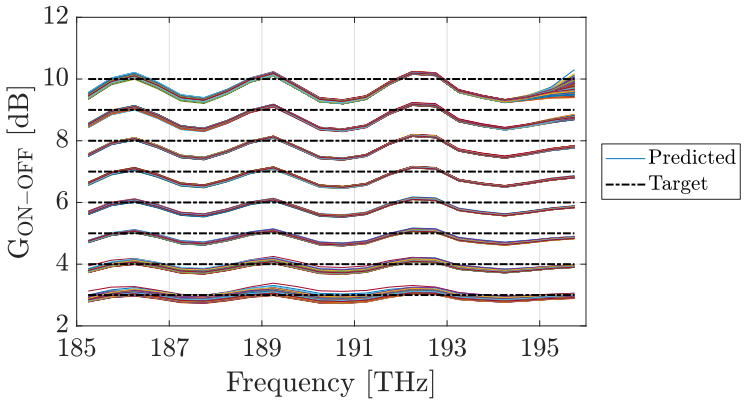


Fig. 5.18 Validated flat on–off gain profiles for predicted pumps power configurations considering different input spectral loads.

compare the validated gain profiles with the original ones of the testing data–set by computing the RMSE and Err_{MAX}. The results are reported in Fig. 5.17, where the respective pdfs and cdfs are shown. With the mean RMSE and Err_{MAX} being 0.36 dB and 0.42 dB, respectively, and with all errors always below 0.75 dB, the upgraded inverse LA-NN ANN_{inv}^{LA} is able to provide accurate predictions also in presence of different partial spectral loads. Moreover, more than 75% of cases have $\text{Err}_{\text{MAX}} < 0.5$ dB.

Afterwards, we use the trained ANN to design RAs providing flat gains from 3 dB to 10 dB. For simplicity and to have the same load distribution as the training data–set, the different partial spectral loads are obtained by randomly extracting 1000 profiles from the testing data–set. Different extractions are performed for different gain levels. As for arbitrary gain profiles, the predicted pumps powers configurations are given to the numerical Raman solver for validation. The resulting gain profiles are illustrated in Fig. 5.18. We can notice the presence of gain ripples due to pumps frequency allocation, whose amplitude increases with increasing gain level. These ripples might be reduced by increasing the number of pump lasers such that no gaps are left between Raman gain peaks.

To assess the ANN ability in designing RAs for flat gain profiles, we evaluate the RMSE and Err_{MAX} between the validated gain profiles and the corresponding target profiles. Since for each target gain level we have 1000 cases, we plot the probability density functions of RMSE and Err_{MAX}.

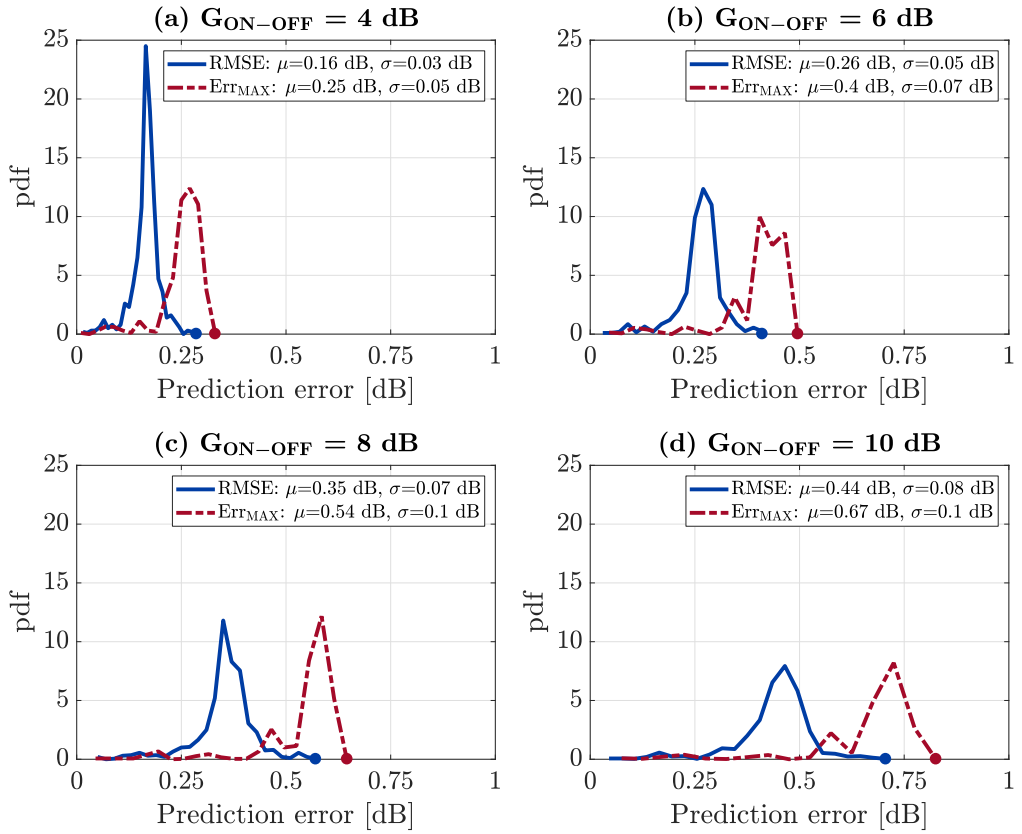


Fig. 5.19 Probability density functions (pdfs) of RMSE and Err_{MAX} of validated flat on–off gain profiles for different partial input spectral loads with respect to target flat profiles with gain levels: (a) 4 dB, (b) 6 dB, (c) 8 dB and (d) 10 dB.

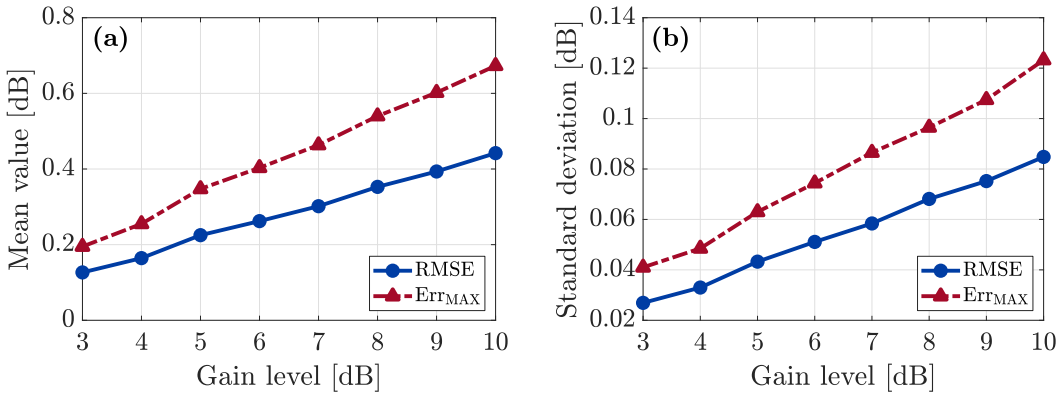


Fig. 5.20 (a) Mean value and (b) standard deviation of RMSE and Err_{MAX} of different validated gain profiles with respect to gain levels.

In Fig. 5.19, we graphically report the pdfs just for four gain levels (4, 6, 8 and 10 dB) for ease in visualization, where we can see that the mean and the standard deviation of RMSE and Err_{MAX} increase with the gain level. The results for all gain levels are instead shown in Fig. 5.20, where the mean and standard deviation of RMSE and Err_{MAX} are plotted with respect to gain levels. Although we observe this increasing trend in the RMSE and Err_{MAX} curves, errors are always below 1 dB and in Fig. 5.20(a) the mean values of RMSE and Err_{MAX} never exceed 0.44 dB and 0.67 dB, respectively.

5.4 Machine Learning for Raman Amplifier Design over Experimental Data

To analyze a more realistic scenario, the machine learning framework for the inverse design is applied to experimental data. In this section, performance over three different bands are investigated: C–band, C+L–band and S+C+L–band. The analysis over the C–band (192–196 THz) have been presented in [28, 29, 31, 117], while results over the C+L–band and the S+C+L–band in [30].

5.4.1 Raman Amplifier Design over C–band

For a first proof–of–principle, we consider the experimental set–up illustrated in Fig. 5.21. The Raman amplifier consists of a single span of SSMF of $\text{L}_{\text{span}} = 100$ km span length considering four counter–propagating pumps. Fiber parameters are the same as in simulation setup, i.e. $\alpha_s = 0.2$ dB/km attenuation for the signal, $\alpha_p = 0.25$ db/km attenuation for the pumps, $\text{A}_{\text{eff}} = 80\ \mu\text{m}^2$ effective area, D = 16.7 ps/nm/km chromatic dispersion, $\gamma = 1.26\ \text{W}^{-1}\text{km}^{-1}$ nonlinear coefficient and $\text{C}_\text{r} = 0.4125\ \text{W}^{-1}\text{km}^{-1}$ Raman gain coefficient. Unlike the simulation data, here the pumps frequencies are fixed and the only degree of freedom is the pumps power. In Table 5.3 we report the values of pump frequencies and pumps powers ranges. Specifically, pumps powers are controlled by voltages.

The signal at the input of the fiber is an arbitrary power profile generated by means of an amplified spontaneous emission (ASE) source with an input total power of 2.6 dBm and discretized over $\text{N}_{\text{ch}} = 40$ channels. The spectra at the output of

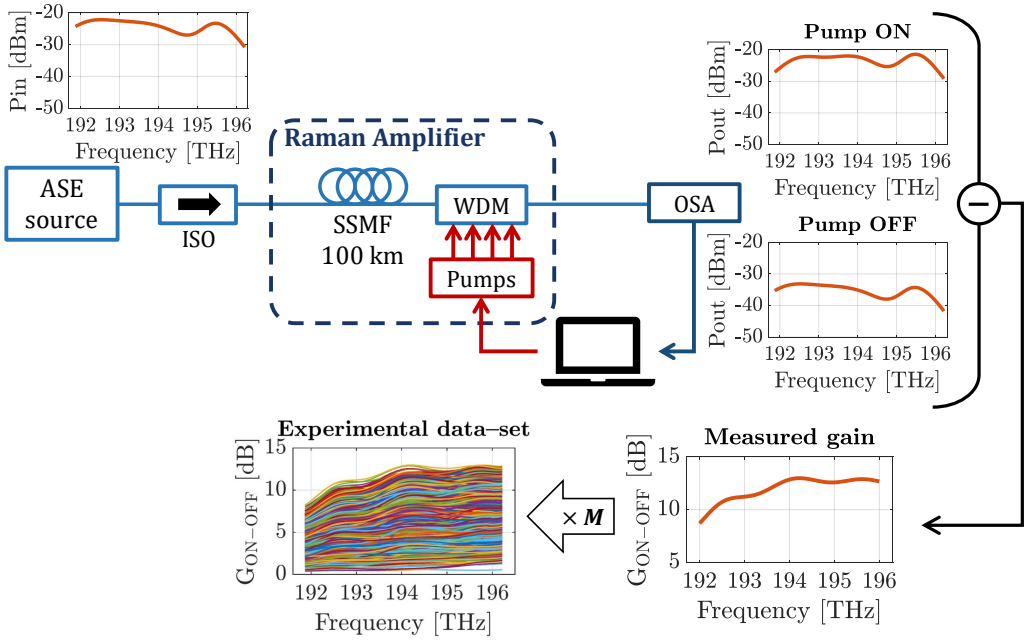


Fig. 5.21 Experimental setup used to collect data–sets and to validate the predicted pumps configuration.

	Pump 1	Pump 2	Pump 3	Pump 4
Frequency [THz]	206.1	207.5	209.0	210.6
Power [mW]	[0,145.0]	[0,158.5]	[0,180.0]	[0,152.5]

Table 5.3 Pump lasers frequencies and powers.

the fiber are captured by an optical spectrum analyzer (OSA) working at 0.1 nm resolution. Specifically, optical power spectra are recorded when pumps are both ON and OFF to compute the Raman on–off gain, which is given by the difference of the two, as show in Fig. 5.21. Also the Raman gain profile is discretized over 40 channels. We can notice that, because of the frequency position of the pump lasers, which cannot be changed, higher gains are observed at higher frequencies.

Using this experimental setup, we collect M=3000 samples corresponding to different configurations of pumps powers drawn from uniform distributions over ranges in Table 5.3. The resulting gain profiles, depicted in Fig. 5.21, are up to 13 dB. Then, these data are split into training and testing data–sets, with 1500 samples each, which results to be large enough to learn the inverse mapping. Compared to the simulation data, in this case the size of the training data–set is smaller because the dimensionality of the space to explore is reduced as pumps are fixed in frequency.

ANN model	ANN_{inv}	ANN_{dir}
Training algorithm	RP	LM
Number of hidden layers N_{HL}	4	2
Number of hidden nodes N_{HN}	700	10
Activation function	tanh	tanh
Model combination parameter N_{ANN}	200	1

Table 5.4 Hyper–parameters setting for ANN_{inv} and ANN_{dir} operating over the C–band.

The hyper–parameters considered to train the ANN_{inv} and the ANN_{dir} are shown in Table 5.4. Once trained, the ANN inverse model is used to design RAs to obtain arbitrary, flat and tilted gain profiles.

Design of Arbitrary Gain Profiles

Given the target arbitrary gains of the testing data–set, the corresponding pumps powers are predicted by the inverse model offline and then validated experimentally using the same setup used for the data–sets generation. Also for experimental data, fine–adjustment of pumps powers is performed to improve the prediction. The deviations between target and validated gain profiles are computed in terms of RMSE and $\mathrm{Err}_{\mathrm{MAX}}$, with resulting pdfs shown in Fig. 5.22 both before and after fine–optimization.

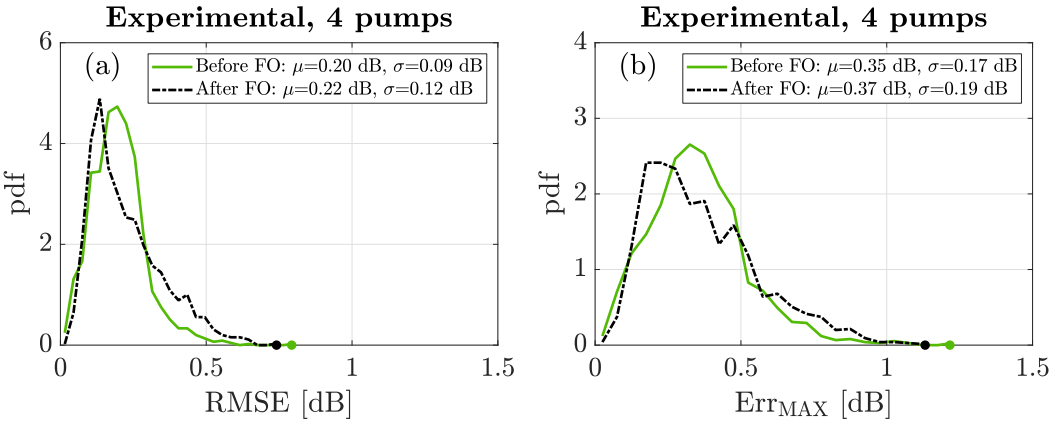


Fig. 5.22 Probability density functions (pdfs) of (a) RMSE and (b) $\mathrm{Err}_{\mathrm{MAX}}$ computed between the target *arbitrary* gain profiles $\mathbf{G}^{targ}$ and the validated gain profiles $\mathbf{G}^{valid}$ over the C–band for four pumps.

Using only the ANN_{inv}, the mean and the standard deviation of the RMSE are 0.20 dB and 0.09 dB respectively, while for the Err_{MAX} they are 0.35 dB and 0.17 dB, respectively. Applying the fine–optimization, there are no improvements, but even slightly higher values are observed, like already occurred in Fig. 5.8(f). Also in this case the ANN_{inv} might already provide the optimal configuration for error minimization. Indeed, in the simulation results section we already discussed the fact that, since the ANN_{dir} used in the fine–optimization process is an approximation of the equation describing the Raman amplification effect, there could be a mismatch between the two that can propagate in the fine–optimization process, resulting in larger errors.

Design of Flat and Tilted Gain Profiles

Finally, the inverse model is tested in the design of Raman amplifiers providing flat and tilted gains. It is important to mention that, since the pumps frequencies are fixed and the pumps powers are limited to the ranges in Table 5.3, it is not sure that completely flat gain profiles can be achieved along the entire C–band.

Fig. 5.23 shows the target and validated flat gain profiles before and after fine–optimization for gain levels from 1 dB to 6 dB. From there we can see that with the increasing gain level, also the deviation of the validated gain profile from the corresponding target one increases, showing poor design performance of the single ANN_{inv} model for gains higher than 4 dB. Better results are obtained employing $ANN_{inv} + ANN_{dir}$ with the gradient–descent routine (Fig. 5.23(b)).

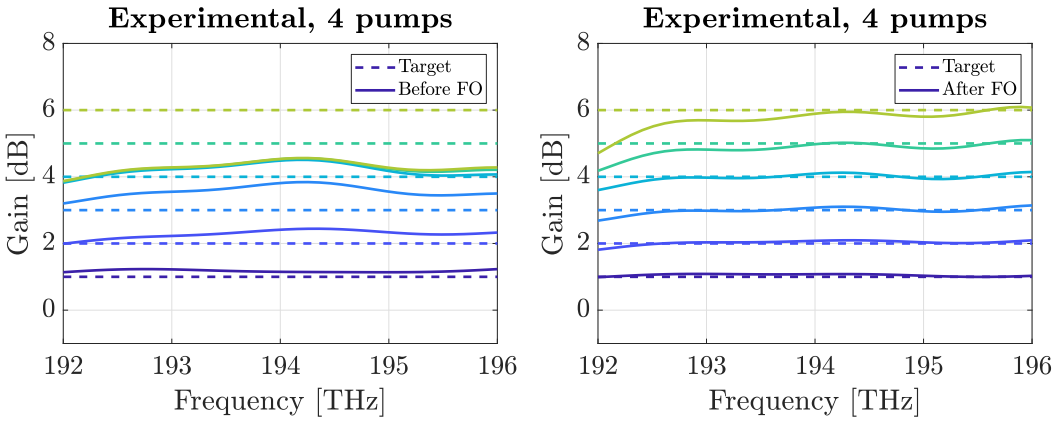


Fig. 5.23 Target and validated *flat* gain profiles over the C–band before and after fine–optimization (FO) in case of four pumps.

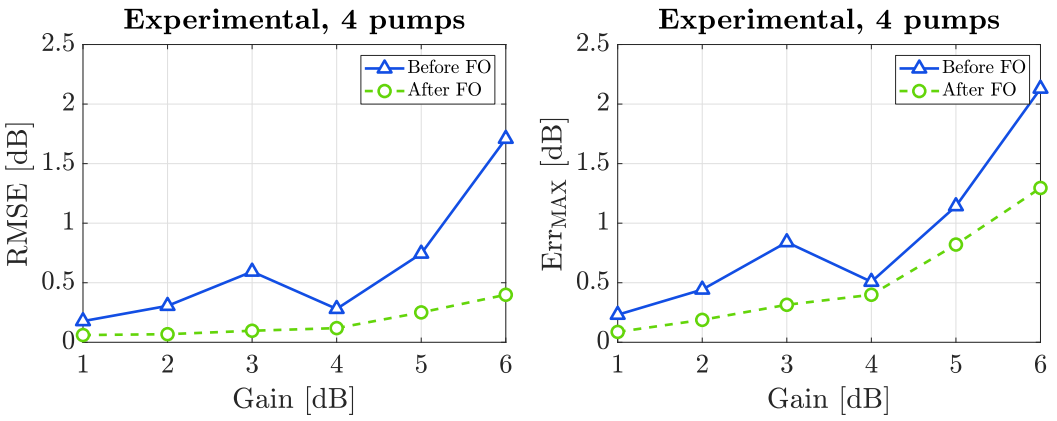


Fig. 5.24 RMSE and Err_{MAX} with respect to gain levels of *flat* gains over the C–band before and after fine–optimization (FO) for four pumps.

Although still some large deviations are present for 5 dB and 6 dB at lower frequencies, maybe due to the Raman module providing lower gains at lower frequencies, in all other cases high flatness is achieved. This is confirmed in Fig. 5.24, where the pdfs of RMSE and Err_{MAX} are plotted with respect to the gain levels, with the largest values experienced for 6 dB. Applying fine–optimization the maximum RMSE values from 1.71 dB reduces to less than 0.5 dB, while for Err_{MAX} the maximum value from 2.13 dB drop down to 1.30 dB.

Similar considerations can be drawn for the tilted gain profiles, whose results are presented in Figs. 5.25 and 5.26. In this case the errors are lower than the flat gains case because the considered tilt is positive and the target gain profiles have the same tendency as the profiles used to train the ANN models (Fig. 5.21).

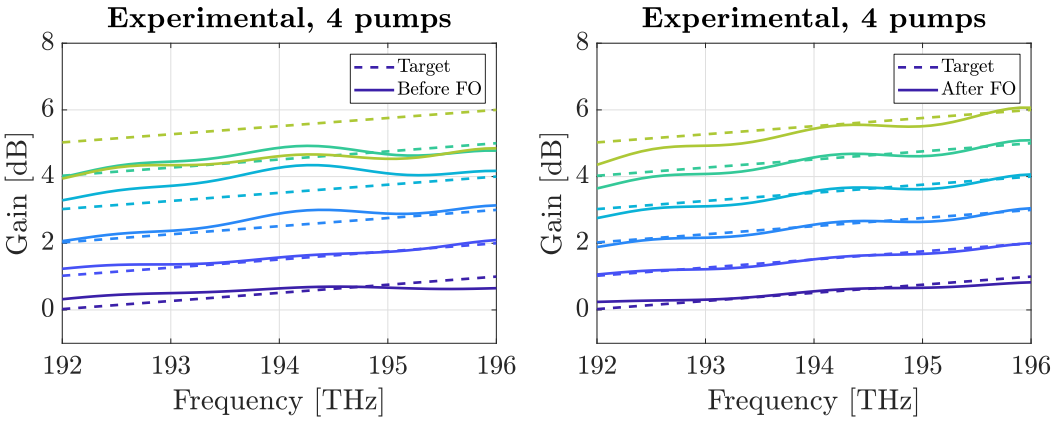


Fig. 5.25 Target and validated *tilted* gain profiles over the C–band before and after fine–optimization (FO) in case of four pumps.

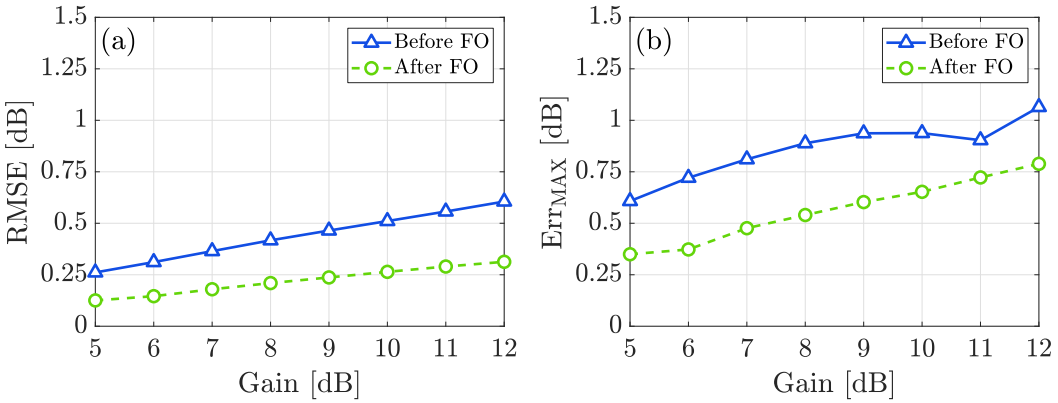


Fig. 5.26 RMSE and Err$_{MAX}$ with respect to gain levels of *tilted* gains over the C–band before and after fine–optimization (FO) for four pumps.

After fine optimization, the maximum values of RMSE and Err$_{MAX}$ become 0.28 dB and 0.67 dB, respectively.

5.4.2 Raman Amplifier Design over C–band: Extension to Different Input Signal Power Profiles and Fiber Types

To extend the ML framework discussed so far to a more generic scenario, we can study the ability of the inverse model to provide the desired gain profile also in case of different input power profiles and considering other types of fiber.

Experimental Setup

In Fig. 5.27 we present the experimental setup considered for the Raman amplifier. As in the previous experiment, the input signal is generated by means of an ASE source to cover the entire C–band between 192 THz and 196 THz, but here a wavelength selective switch (WSS) is inserted to shape the power spectral density (PSD) profile of the input signal. Examples of possible input signal PSD are shown in inset A for different tilt (flat, positively tilted and negatively tilted).

The Raman amplifier consists of an optical fiber and four counter-propagating pumps, provided by a commercial Raman pump module, which are combined with the input signal by a wavelength division multiplexer (WDM). Laser pumps frequencies are fixed while powers can vary. The relative values are the same as

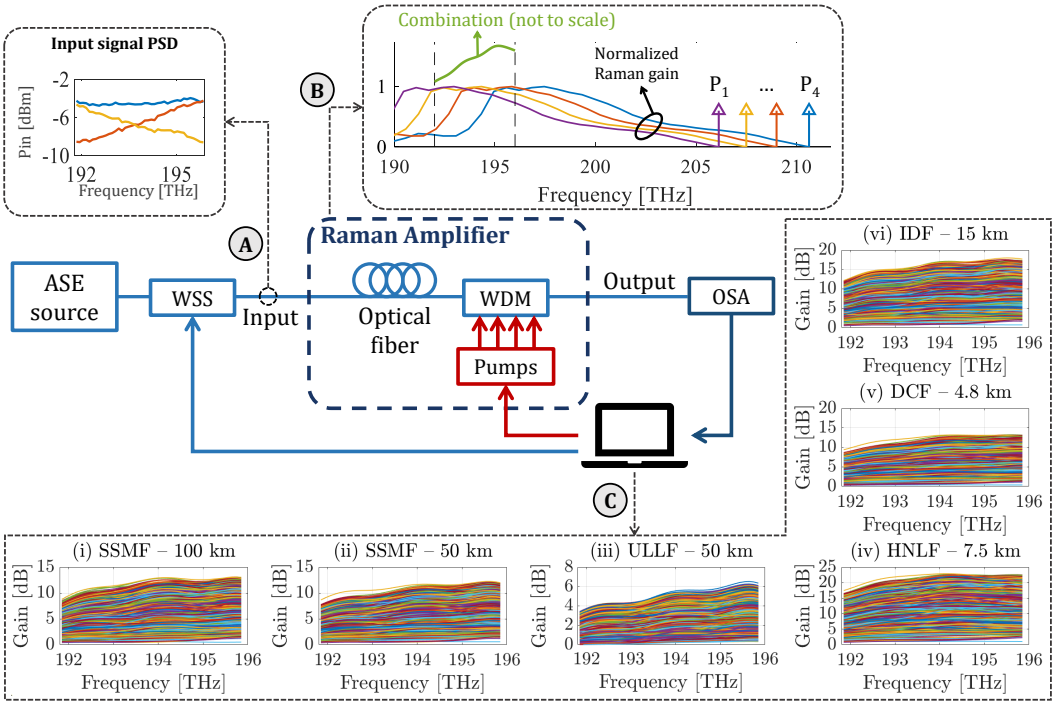


Fig. 5.27 Experimental setup used to collect data–sets for artificial neural network training, validation and testing. It is also used to further validate the predicted pumps configurations.

	Pump 1	Pump 2	Pump 3	Pump 4
Frequency [THz]	206.1	207.5	209.0	210.6
Power [mW]	[0,145.0]	[0,158.5]	[0,180.0]	[0,152.5]

Table 5.5 Pump lasers frequencies and powers.

in the experimental validation presented in section 5.4.1, but for reference they are recalled in Table 5.5.

Then, an OSA measures the resulting signal PSD at the output of the Raman amplifier at 0.1 nm resolution. Here as well, we compute the Raman on–off gain between the output signal PSD when the pumps are ON and the output signal PSD when the pumps are OFF. Signal PSDs and gain profiles are discretized over $N_{ch} = 40$ channels, where 1 corresponds to the lowest frequency and N_{ch} corresponds to the highest. In inset B, a qualitative representation of the normalized Raman gain efficiency profiles are shown for each pump together with the profile resulting from their combination. As discussed in [81, 94], according to the fiber type, the Raman gain efficiency profile should be properly scaled.

Fiber type	SSMF	ULLF	HNLF	DCF	IDF
$\alpha_{\mathbf{1550}}$ (dB/km)	0.2	0.16	1.0	0.5	0.23
$\alpha_{\mathbf{1450}}$ (dB/km)	0.25	0.2	1.2	0.8	0.31
$\mathbf{A_{eff}}$ (μm^2)	80	153	10	15	31
$\mathbf{C_r}$ (W^{-1}km^{-1})	0.8	0.52	6.3	3	1.3
$\mathbf{L_{span}}$ (km)	[50,100]	50	7.5	4.8	15

Table 5.6 Optical fibers parameters, where α: attenuation, A_{eff}: effective area, C_r: Raman gain coefficient, L_{span}: span length.

In our analysis, five types of fibers are considered: standard single mode fiber (SSMF), ultra–low loss fiber (ULLF), highly non–linear fiber (HNLF), dispersion compensating fiber (DCF) and inverse dispersion fiber (IDF). Their parameters are reported in Table 5.6, where for the SSMF two span lengths are considered: 50 km and 100 km. SSMF and ULLF are transmission fibers used for distributed Raman amplifiers, where the SSMF is used in terrestrial systems and the ULLF in submarine and unrepeatered links thanks to its lower attenuation and wider effective area. In our experiment, the total signal power at the input of these fibers is always below 9 dBm.

On the contrary, HNLF, DCF and IDF are highly non–linear fibers generally used to build discrete Raman amplifiers [57], where DCF and IDF also compensate for chromatic dispersion. Moreover, for these three types of fiber, the total input power never exceeds 3 dBm. Still in Fig. 5.27, in insets C,(i)–(vi) we show the Raman on–off gains for each fiber type and for different pumps power configurations computed from the output signal PSD collected as explained above.

As already observed in section 5.4.1, also in this case gains are greater at higher frequencies, probably because of the different contributions of each pump that sum up together. Less likely it is due to the SRS causing power transfer from higher frequency channels to lower frequency ones. Indeed, the system is operating over the C–band, where channels are close enough to prevent from these further substantial second–order effects.

Data–set Generation

For data–sets generation, we extract M different pumps configurations with powers uniformly distributed over ranges in Table 5.5. To each of them corresponds an input

signal PSD profile and a gain profile. Only linear input PSD profiles are considered. Therefore we can introduce the scalar ΔP_{in}, defined as the difference between the powers at highest and lowest frequencies, as in Eq. (5.6):

$$\Delta P_{in} = P_{in,Nch} - P_{in,1}. \tag{5.6}$$

According to the type of power PSD profile at the input of the Raman amplifier, we can identify two types of data–sets:

- $\mathscr{D}_{\Delta P_{in}}$–rand, when the PSD profile associated to each pumps configuration is randomly selected, such that $\Delta P_{in} \sim U[-5,+5]$ dB. This type of data–set is generated only for the distributed Raman amplifier considering the SSMF at 100 km, for which, depending on the value assumed by ΔP_{in}, the total input power can be between 7 dB and 9 dB;
- $\mathscr{D}_{\Delta P_{in}}$–fixed, when to each pumps configuration is associated the same flat input signal PSD profile. Instead, this kind of data–set is generated for all the other types of fibers. Moreover, since the input signal PSD does not change, the total input power is constant, being around 9 dB in case of distributed Raman amplifier using ULLF and SSMF at 50 km, and 3 dB for the discrete Raman amplifier using HNLF, DCF and IDF.

For each kind of fiber we generate $M = 3000$ samples for the $\mathscr{D}_{\Delta P_{in}}$-fixed data–set type, with the exception of SSMF at 100 km for which we generate $M = 6000$ samples for the $\mathscr{D}_{\Delta P_{in}}$-fixed data–set and $M = 10000$ samples for the $\mathscr{D}_{\Delta P_{in}}$-rand. Each data–set is split in two, one for training, referred to as $\mathscr{D}_1$, and the other one, indicated as $\mathscr{D}_2$, for testing and evaluation of the overall ANN performance.

ML–Framework and ANN Models

Depending on the information considered at the input of the inverse mapping, three types of ANN models, shown in Fig. 5.28, are analyzed:

- $\mathbf{P_{in}}$–unaware model (5.28(a)), which learns the inverse mapping between gains and pumps powers, without considering the signal PSD profile at the input of the ANN model;
- $\mathbf{P_{in}}$–aware model (5.28(b)), at whose input there are the gain and the input signal PSD profiles;

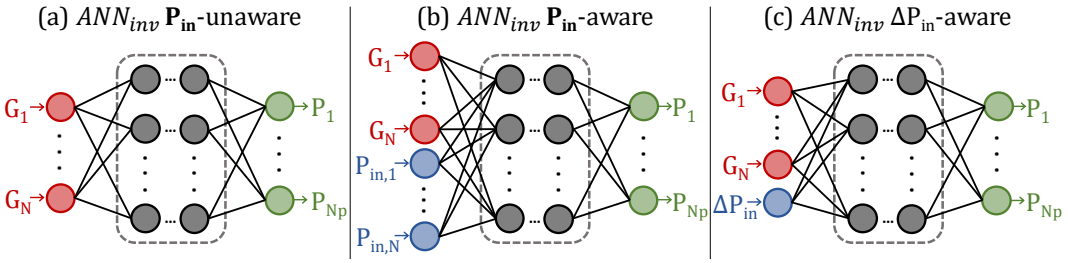


Fig. 5.28 Different *inverse* artificial neural network models ANN_{inv} depending on the information considered at the input of the network: (a) $\mathbf{P_{in}}$-unaware, (b) $\mathbf{P_{in}}$-aware and (c) ΔP_{in}-aware.

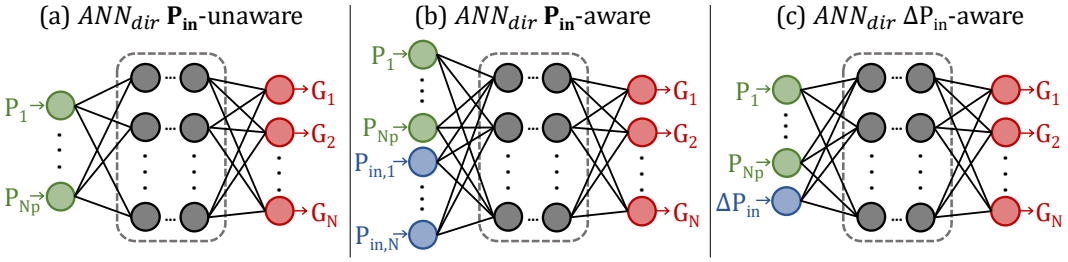


Fig. 5.29 Different *direct* artificial neural network models ANN_{dir} depending on the information considered at the input of the network: (a) $\mathbf{P_{in}}$-unaware, (b) $\mathbf{P_{in}}$-aware and (c) ΔP_{in}-aware.

- ΔP_{in}–aware model (5.28(c)), at whose input there are the gain and the tilt ΔP_{in} of the linear signal PSD profile.

The $\mathbf{P_{in}}$–unaware model is trained over $\mathscr{D}_{\Delta P_{in}}$–fixed data–sets and it is used for the generalization of the ML–framework to different types of fiber. Instead, the $\mathbf{P_{in}}$–aware and the ΔP_{in}–aware models are trained jointly with the $\mathscr{D}_{\Delta P_{in}}$–rand and $\mathscr{D}_{\Delta P_{in}}$–fixed data–sets. Once trained, all the models are tested over both $\mathscr{D}_{\Delta P_{in}}$–rand and $\mathscr{D}_{\Delta P_{in}}$–fixed. Likewise, for the direct model ANN_{dir} used in the fine–optimization routine to adjust pumps powers, a similar classification is defined, as shown in Fig. 5.29.

The optimal ANN models are selected by optimizing the hyper–parameters through grid search approach. The optimized hyper–parameters are the number of hidden layers N_{HL}, the number of hidden nodes N_{HN}, the activation function f_{act}, the regularization parameter λ_{ANN} and the standard deviation $\sigma_A NN$ of the normal distribution on which the initialization of the ANN weights is based. Also here,

Models	Input signal PSD awareness	Fiber type	Training algorithm	f_{act}	N_{HL}	N_{HN}	N_{ANN}	σ_{ANN}	λ_{ANN}
ANN_{inv}	$\mathbf{P_{in}}$–unaware	All fibers*	RP	tanh	2	600	10	5e-2	1e6
ANN_{inv}	$\mathbf{P_{in}}$–unaware	SSMF 100-km	RP	logsig	1	1300	20	5e-2	1e-8
ANN_{inv}	$\mathbf{P_{in}}$–aware	SSMF 100-km	RP	sine	1	1000	20	7e-3	1e-8
ANN_{inv}	ΔP_{in}–aware	SSMF 100-km	RP	sine	1	1000	20	2.5e-2	1e-8
ANN_{dir}	All models	All fibers	LM	tanh	2	10	1	**	***

Table 5.7 Neural network parameters for the different models.
(*) Except SSMF 100-km; (**) Nguyen-Widrow initialization algorithm [106]; (***) Dynamically modified during training according to [49].

model averaging is applied considering N_{ANN} parallel and independent ANNs to reduce the impact of the weights random initialization.

The ANN_{inv} is trained with the random–projection method and for a better model generalization, K-fold cross–validation is performed with K = 10, with 90% of the training data–set $\mathscr{D}_1$ reserved for training and 10% for model validation to prevent from overfitting.

Instead, to train the ANN_{dir} we use the back–propagation method based on the Levenberg–Marquardt algorithm and the training data–set $\mathscr{D}_1$ is divided in three parts: 70% for training, 15% for validation and 15% for testing [105]. All the optimized hyper–parameters are reported in Table 5.7 for both inverse and direct ANN models. Specifically, the hyper–parameters in the last row, which is related to the direct mapping, are the same for all the three ANN_{dir} models of Fig. 5.29.

Testing Results

The ability of the inverse ANN model to provide high prediction accuracy for different types of fiber is analyzed considering the $\mathbf{P_{in}}$–unaware models tested only over $\mathscr{D}_{\Delta \mathrm{P_{in}}}$–fixed. For each considered fiber type, the prediction accuracy is independently evaluated by computing the maximum error between the true and the predicted pumps power values. The maximum power prediction errors experienced by each pump for the different fibers are shown in Table 5.8. In general, prediction accuracy is quite high, with errors below 50 mW which corresponds to variations of gain up to 1 dB, with the exception of pumps 3 and 4 for HNLF and IDF, probably because their gains range is greater. In these cases, the errors are between 50 mW and 100 mW, which may produce gain variations from 1 dB to 2 dB.

Fiber type	SSMF 50 km	ULLF	HNLF	DCF	IDF
P_1 [mW]	19.3	50.3	28.6	17.2	31.2
P_2 [mW]	38.4	47.7	45.8	19.9	47.7
P_3 [mW]	22.1	39.4	96.5	19.8	60.7
P_4 [mW]	25.5	9.6	98.5	29.4	93.3

Table 5.8 Maximum prediction error between true and predicted pumps powers using the ANN_{inv} $\mathbf{P_{in}}$–unaware model for different types of fibers considering their relative $\mathscr{D}_{\Delta P_{in}}$-fixed data–sets.

ANN model	$\mathbf{P_{in}}$–unaware		$\mathbf{P_{in}}$–aware		ΔP_{in}–aware	
Data–set type $\mathscr{D}_{\Delta P_{in}}$	–fixed	–rand	–fixed	–rand	–fixed	–rand
P_1 [mW]	15.1	41.5	11.8	15.4	13.5	26.5
P_2 [mW]	22.6	90.2	17.6	27.9	21.3	39.0
P_3 [mW]	33.6	55.9	27.1	20.8	22.0	23.7
P_4 [mW]	41.1	60.8	38.2	27.9	31.7	41.6

Table 5.9 Maximum prediction error between true and predicted pumps powers for SSMF 100 km over $\mathscr{D}_{\Delta P_{in}}$–fixed and $\mathscr{D}_{\Delta P_{in}}$–rand data–sets for different input signal PSD awareness.

As already mentioned above, the performance of the different inverse models of Fig. 5.28 based on whether the input signal information is taken into account or not, is analyzed only for the 100 km SSMF.

In Table 5.9, we report the corresponding maximum errors between the true and the predicted pumps powers for both $\mathscr{D}_{\Delta P_{in}}$–fixed and $\mathscr{D}_{\Delta P_{in}}$–rand. From there, we can see that the ANN_{inv} $\mathbf{P_{in}}$–unaware model is able to provide good prediction accuracy only over the $\mathscr{D}_{\Delta P_{in}}$–fixed, for which the input signal PSD profiles can be neglected, as they are the same for all cases. Instead, over $\mathscr{D}_{\Delta P_{in}}$–rand the predictions are worse, with errors up to 90 mW for pump 2. Although we would expect lower errors for the $\mathscr{D}_{\Delta P_{in}}$–rand as well, since over the C–band the signal–signal and pump–pump interactions due to SRS are not that strong to produce a significant effect on the gain profile, from these results we can infer that even small SRS effects can affect the inverse model.

On the contrary, both ANN_{inv} $\mathbf{P_{in}}$–aware and ΔP_{in}–aware models are able to provide good predictions over both types of testing data–sets. Moreover, in case of $\mathscr{D}_{\Delta P_{in}}$–rand data–set, the ANN_{inv} $\mathbf{P_{in}}$–aware model presents even lower errors than the ANN_{inv} ΔP_{in}–aware model. Most likely, this is due to the fact that for the first model the entire signal PSD profile is given at the input of the ANN, while for the

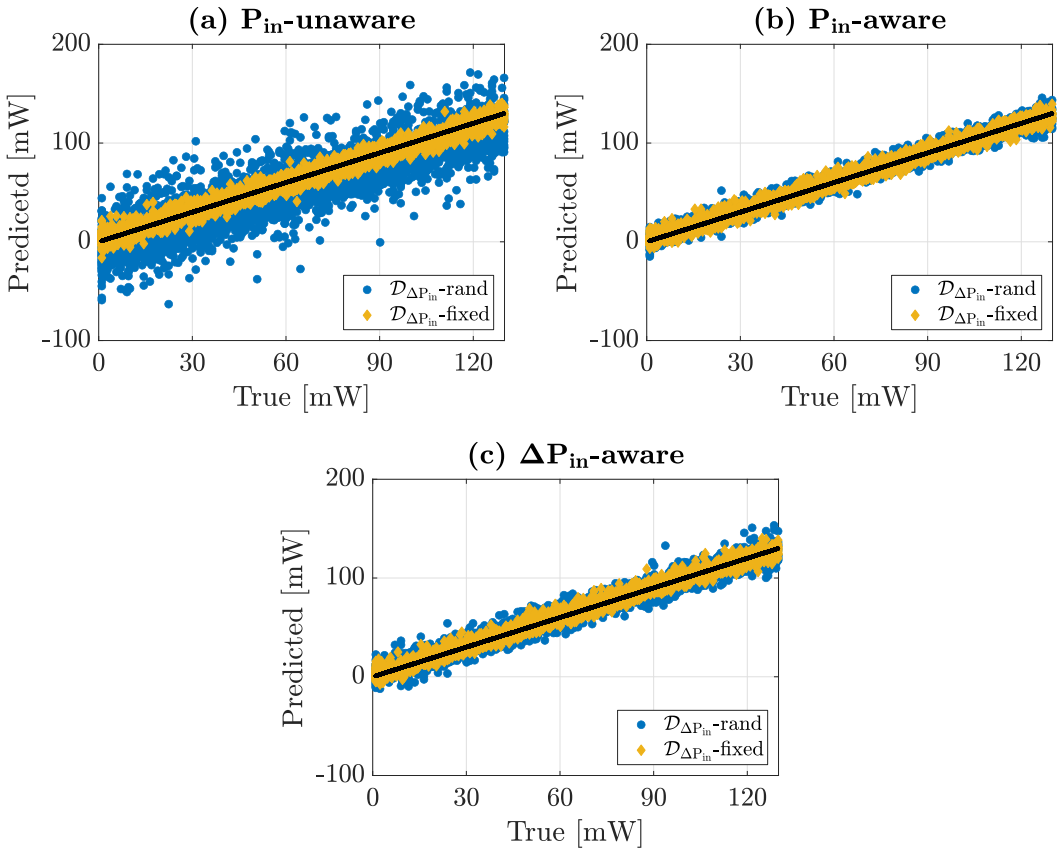


Fig. 5.30 True power versus predicted power for *pump 2* over $\mathscr{D}_{\Delta P_{in}}$-rand and $\mathscr{D}_{\Delta P_{in}}$–fixed using (a) $\mathbf{P_{in}}$–unaware, (b) $\mathbf{P_{in}}$–aware and (c) ΔP_{in}–aware inverse ANN models.

second model only the tilt is given, which might be a less meaningful and important information. Nevertheless, the ANN_{inv} ΔP_{in}–aware model returns enough accurate pumps powers predictions over the $\mathscr{D}_{\Delta P_{in}}$–rand. This is further confirmed if we look at the performance of the three models in predicting powers for pump 2 (Fig. 5.30), the one experiencing the highest maximum error when the $\mathbf{P_{in}}$–unaware model is used over the $\mathscr{D}_{\Delta P_{in}}$–rand data–set.

Indeed, compared to Fig. 5.30(a), a substantial reduction of deviations from the true values can be observed in Figs. 5.30(b) and 5.30(c). Results are similar for ANN_{inv} $\mathbf{P_{in}}$–aware and ΔP_{in}–aware model because the input signal PSD profile is linear.

In case of arbitrary non–linear input profiles, we would expect worse performance with the ΔP_{in}–aware model, since a single scalar ΔP_{in} would no longer be enough to describe the behavior of the signal at the input of the fiber.

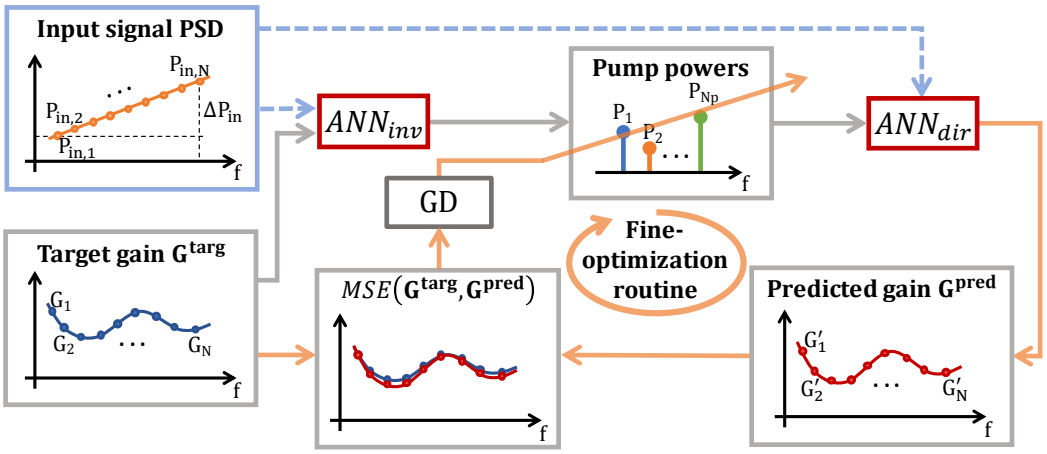


Fig. 5.31 Machine learning framework including fine–optimization procedure. According to the type of ANN models, input signal PSD information can be considered or not.

Also here, the direct model is used in combination with the gradient descent algorithm to fine–adjust pumps powers configurations, as shown in Fig. 5.31. Once trained, we test the different ANN_{dir} models over $\mathscr{D}_2$ data-sets and we evaluate the prediction accuracy in terms of absolute maximum error Err_{MAX} computed between target and predicted gain profiles. Fig. 5.32(a) shows the pdfs of Err_{MAX} for different types of fiber obtained considering the ANN_{dir} $\mathbf{P_{in}}$–unaware model over $\mathscr{D}_{\Delta P_{\text{in}}}$–fixed data–sets. From there we can see that highly accurate predictions are achieved for all fibers. Particularly, larger errors are given by fibers used in discrete RAs (HNLF, DCF and IDF), with the worst case identified in the IDF. However, in general the average (μ) and the standard deviation (σ) values of the Err_{MAX} never exceed 0.32 dB and 0.17 dB, respectively, and the Err_{MAX} is always below 1.25 dB. Instead, with mean and standard deviation of Err_{MAX} equal to 0.11 dB and 0.05 dB, respectively, and Err_{MAX} values always lower than 0.5 dB, best results are given in case of ULLF.

Figs. 5.32(b)–(d) show the prediction accuracy of the three ANN_{dir} models with different input signal PSD awareness over both $\mathscr{D}_{\Delta P_{\text{in}}}$–fixed and $\mathscr{D}_{\Delta P_{\text{in}}}$–rand just for the 100 km SSMF. Interestingly, all the three models provide the same performance over the $\mathscr{D}_{\Delta P_{\text{in}}}$–fixed data–sets, while over $\mathscr{D}_{\Delta P_{\text{in}}}$–rand data–sets, prediction accuracy is lower in case of ANN_{dir} $\mathbf{P_{in}}$–unaware model, like for the inverse design case. Nevertheless, the worsening of performance from $\mathscr{D}_{\Delta P_{\text{in}}}$–fixed to $\mathscr{D}_{\Delta P_{\text{in}}}$–rand is reduced compared to the corresponding inverse ANN model, sign that the direct model is more robust to the effects caused by small SRS interactions.

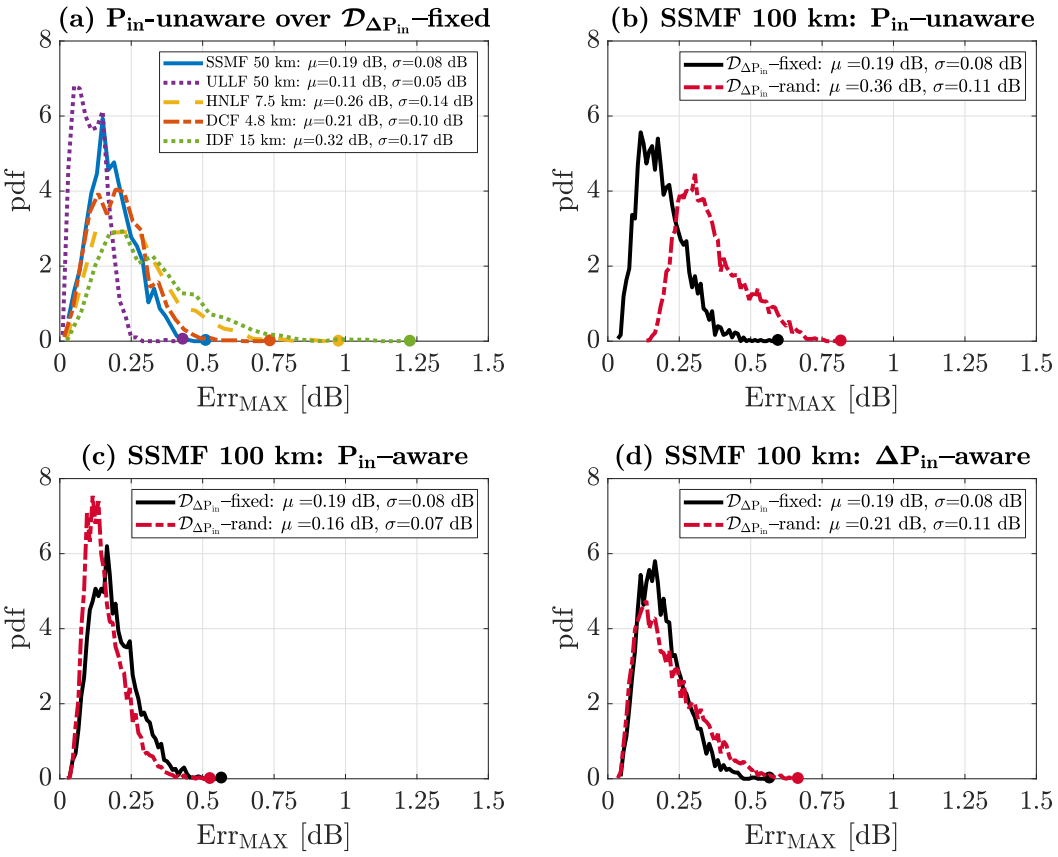


Fig. 5.32 Probability density functions of $\mathrm{Err_{MAX}}$ computed between target and predicted gain profiles for different ANN direct models ANN_{dir} over different testing data–sets: (a) $\mathbf{P_{in}}$–unaware model over $\mathscr{D}_{\Delta P_{in}}$–fixed for various optical fibers, (b) $\mathbf{P_{in}}$–unaware model, (c) $\mathbf{P_{in}}$–aware model, and (d) ΔP_{in}–aware model for 100 km SSMF over both $\mathscr{D}_{\Delta P_{in}}$–fixed and $\mathscr{D}_{\Delta P_{in}}$–rand.

After testing the inverse models over arbitrary gain profiles, we will test their ability in designing flat and tilted profiles, with gain levels between 1 and 6 dB, like done in the previous sections. It is important to remark that, due some possible limitations of the experimental setup, the design of flat and tilted gain profiles is not guaranteed.

Experimental Validation of the Predicted Pumps Configurations

Once the pumps powers have been predicted for arbitrary, flat or tilted gain profiles, with or without fine–optimization, we come back to the experimental setup of Fig. 5.27 to determine the corresponding profiles to compare with the target ones.

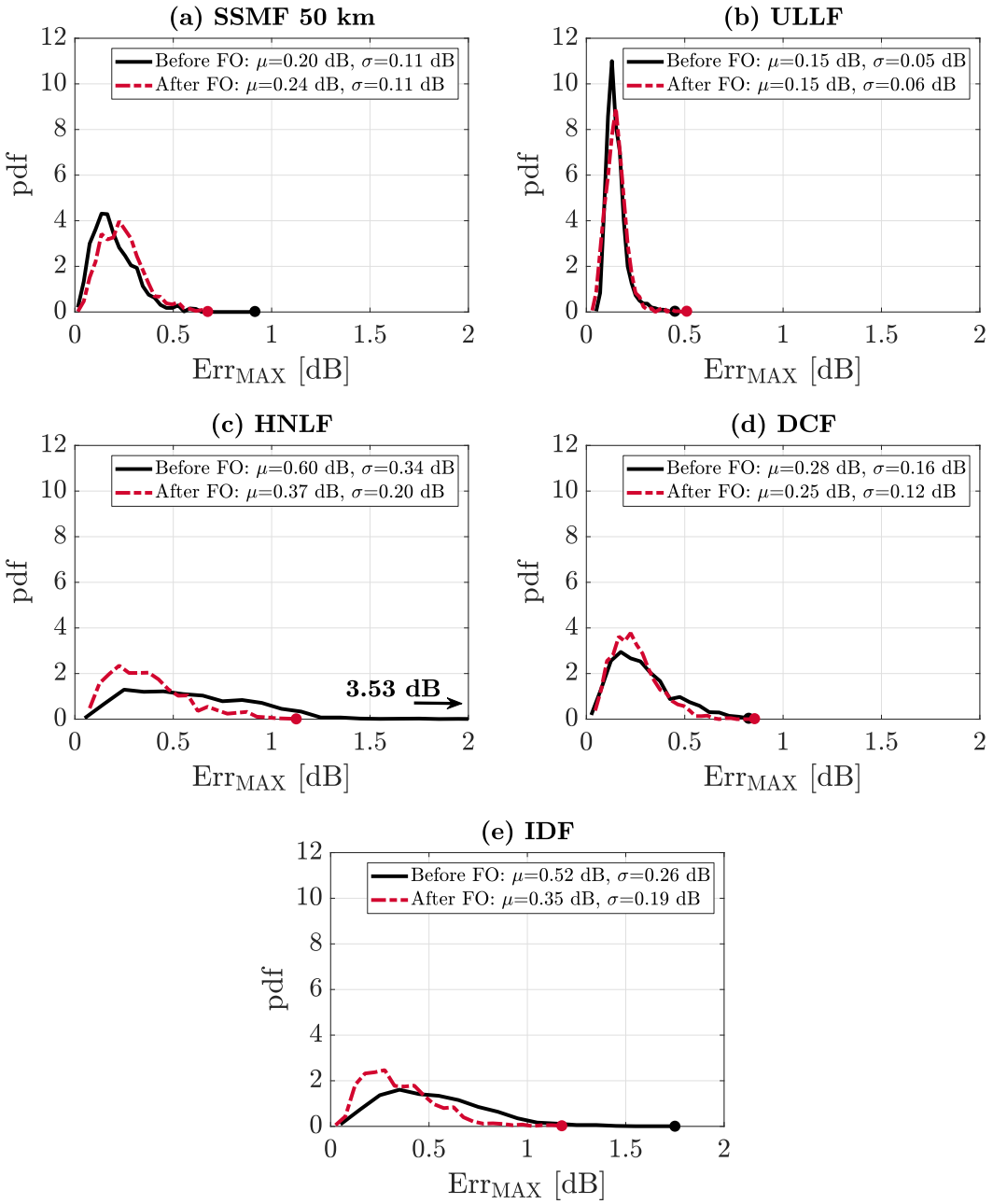


Fig. 5.33 Probability density functions of $\mathrm{Err}_{\mathrm{MAX}}$ computed between target and validated *arbitrary* gain profiles for different types of fiber before and after fine–optimization (FO). $\mathbf{P_{in}}$–unaware ANN models are considered over $\mathscr{D}_{\Delta \mathrm{P_{in}}}$–fixed data–sets.

In Fig. 5.33 we report the pdfs of $\mathrm{Err}_{\mathrm{MAX}}$ computed between target and validated arbitrary gain profiles before and after fine–optimization for different types of fiber. Here, $\mathbf{P_{in}}$–unaware models are considered for predictions over $\mathscr{D}_{\Delta \mathrm{P_{in}}}$–fixed data–

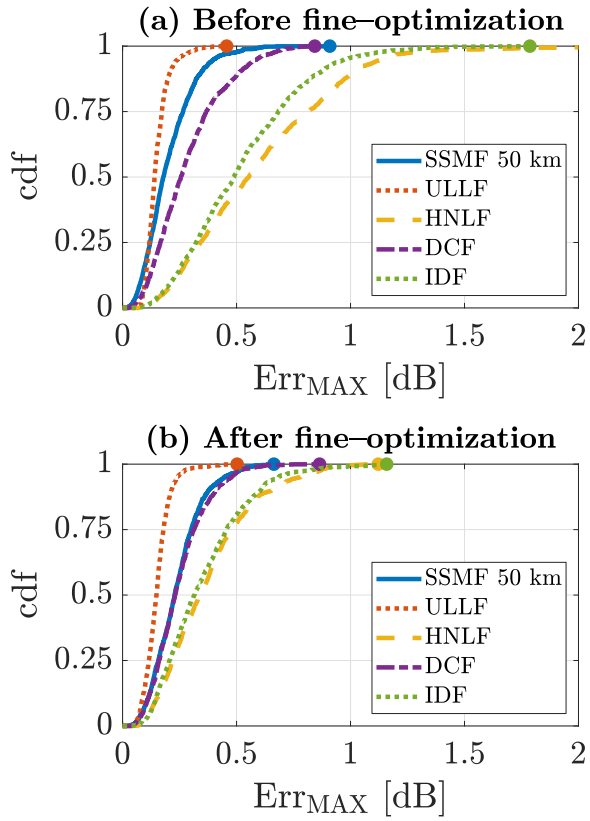


Fig. 5.34 Cumulative distribution functions of $\mathrm{Err}_{\mathrm{MAX}}$ computed between target and validated *arbitrary* gain profiles for different types of fiber before and after fine–optimization (FO). $\mathbf{P_{in}}$–unaware ANN models are considered over $\mathscr{D}_{\Delta \mathrm{P_{in}}}$–fixed data–sets.

sets, each composed of 1500 samples. The corresponding cumulative distribution functions are shown in Figs. 5.34(a) and 5.34(b) before and after fine–optimization, respectively.

In Fig. 5.34(a), we can see that the worst performance is experienced by HNLF and IDF, for which only 44% and 52% of cases have $\mathrm{Err}_{\mathrm{MAX}} < 0.5$ dB, respectively, before fine–optimization. In fact, for both types of fiber, higher errors are already experienced in the inverse design. Moreover, although for HNLF the $\mathrm{Err}_{\mathrm{MAX}}$ reaches values up to 3.53 dB, actually very few cases are above 2 dB, corresponding to just the 0.47% of the total cases. On the contrary, despite being a highly non–linear fiber as well, the DCF shows a way better performance, with 89% of cases having $\mathrm{Err}_{\mathrm{MAX}} < 0.5$ dB. Even better inverse design ability is observed for the 50 km SSMF and the ULLF. Indeed, for the SSMF 98% of case have $\mathrm{Err}_{\mathrm{MAX}} < 0.5$ dB, while for the ULLF all the $\mathrm{Err}_{\mathrm{MAX}}$ values are below 0.5 dB.

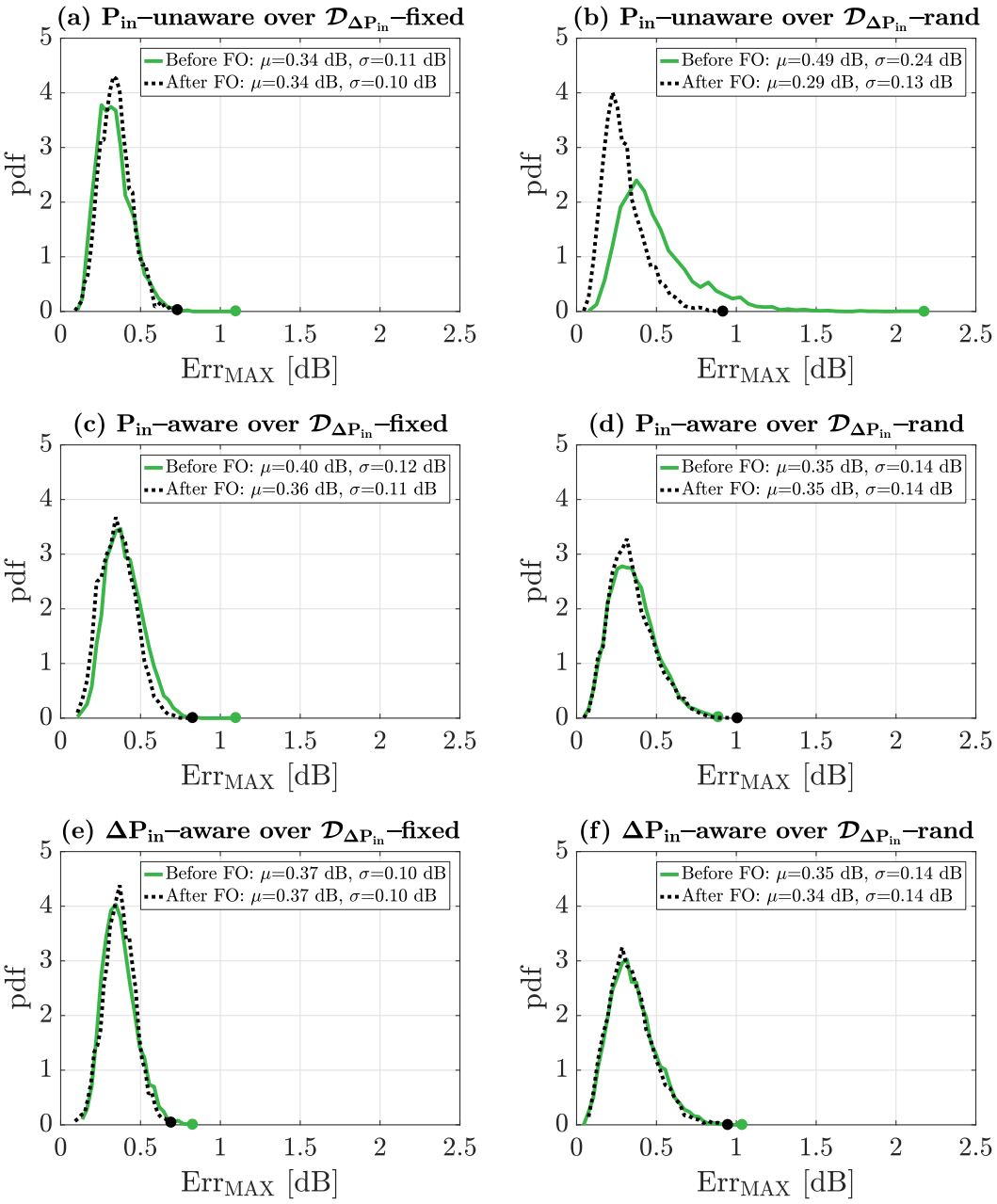


Fig. 5.35 Probability density functions (pdfs) of $\mathrm{Err_{MAX}}$ computed between target and validated *arbitrary* gain profiles for 100 km SSMF considering $\mathbf{P_{in}}$–unaware, $\mathbf{P_{in}}$–aware and ΔP_{in}–aware over both $\mathscr{D}_{\Delta P_{in}}$–fixed and $\mathscr{D}_{\Delta P_{in}}$–rand before and after fine–optimization (FO).

When we apply fine–optimization (Fig. 5.34(b)), performance considerably improves for HNLF and IDF, since the percentage of cases with $\mathrm{Err_{MAX}} < 0.5$ dB is around 80%. Instead, for the other fibers (SSMF, ULLF and IDF), very small or even

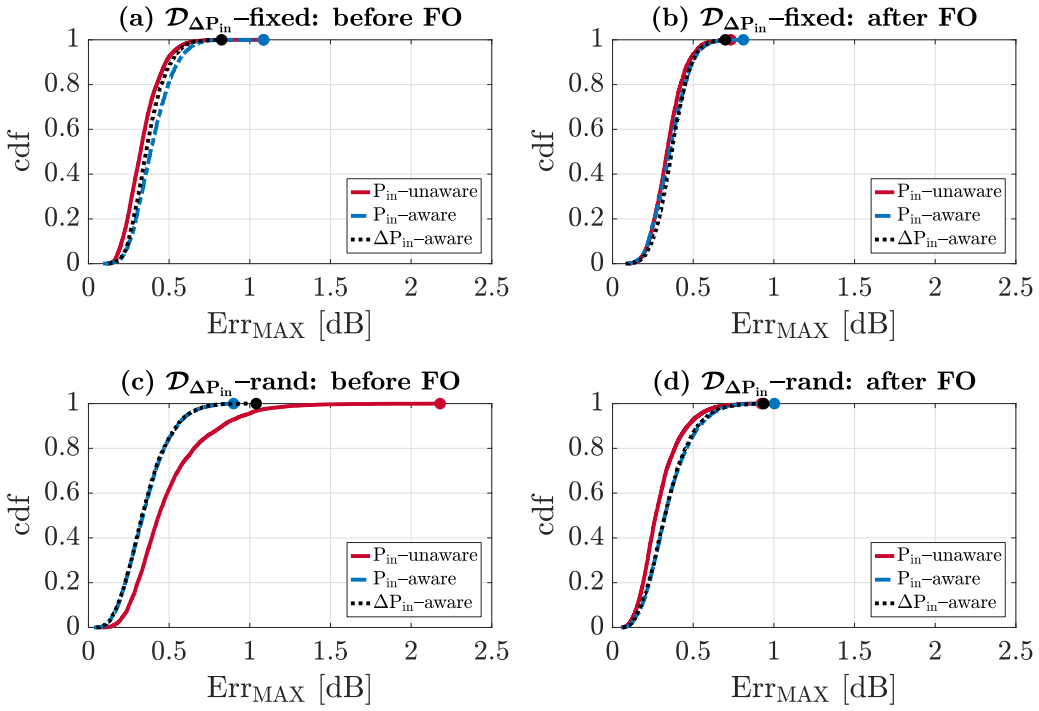


Fig. 5.36 Cumulative distribution functions (cdfs) of Err_{MAX} computed between target and validated *arbitrary* gain profiles for 100 km SSMF considering $\mathbf{P_{in}}$–unaware, $\mathbf{P_{in}}$–aware and ΔP_{in}–aware over both $\mathscr{D}_{\Delta P_{in}}$–fixed and $\mathscr{D}_{\Delta P_{in}}$–rand before and after fine–optimization (FO).

no improvement is observed after pumps powers fine–adjustment, as the solutions given by just the ANN_{inv} alone are already optimal and on a local minimum.

The capability of the inverse model in handling different input signal PSD profiles is analyzed for the 100 km SSMF in Figs. 5.35 and 5.36, where the pdfs and the cdfs of Err_{MAX} computed between target and validated arbitrary gains profiles both before and after fine–optimization are reported. In Fig. 5.35, the different input signal PSD awareness models are organized over rows, while on columns we have the two considered types of data–sets: $\mathscr{D}_{\Delta P_{in}}$–fixed and $\mathscr{D}_{\Delta P_{in}}$–rand, containing 3000 and 5000 samples, respectively.

As expected, if we do not perform fine–optimization, high accurate design is given by all models, as maximum values of $Err_{MAX}$are around or less than 1 dB, with the exception of the $\mathbf{P_{in}}$–unaware tested over $\mathscr{D}_{\Delta P_{in}}$–rand data–set (Fig. 5.35(c)). Indeed, in this case Err_{MAX} reaches values up to 2.2 dB and just 62% of cases have $Err_{MAX} < 0.5$ dB, while for the other models the percentage is always above 80%. The performance of the $\mathbf{P_{in}}$–unaware tested over $\mathscr{D}_{\Delta P_{in}}$–rand data–set improves

significantly after fine–optimization, in fact the maximum value of $\mathrm{Err}_{\mathrm{MAX}}$ reduces to less than 1 dB, like for the other models.

Instead, there are no considerable improvements for the other models, as the pumps configurations might be already optimal and the fine–optimization would just introduce random disturbance. From Figs. 5.36(d), i.e. after fine–optimization, we can observe that the $\mathbf{P_{in}}$–unaware model has even better performance than the $\mathbf{P_{in}}$–aware and the $\Delta\mathrm{P_{in}}$–aware models. This might be explained by the smaller input dimension of the $\mathbf{P_{in}}$–unaware model, therefore fewer parameters that risk to get trapped in local minima [68].

In conclusion, in case of arbitrary gain profiles design, it is always possible to achieve high accurate design performance with 80% of cases having $\mathrm{Err}_{\mathrm{MAX}}$ below 0.5 dB, also in case of different input signal PSD conditions. These results can be obtained with the $\mathbf{P_{in}}$–aware and $\Delta\mathrm{P_{in}}$–aware models without the need of pumps powers fine–optimization, otherwise with the $\mathbf{P_{in}}$–unaware jointly with fine–optimization. Interestingly, the $\Delta\mathrm{P_{in}}$–aware provides performance similar to the $\mathbf{P_{in}}$–aware model, making it attractive for its lower complexity, as only one additional input is introduced to handle the input signal PSD information.

Finally, we study the performance of the inverse model in designing flat and tilted gain profiles with gain levels between 1 and 6 dB for different types of fiber. For tilted gains, 0.25 dB/THz of slope is considered to have a total of 1 dB over the entire 4 THz C–band. For this analysis, only constant input signal PSD profiles are taken into account, corresponding to the $\mathscr{D}_{\Delta\mathrm{P_{in}}}$-fixed data–set, therefore $\mathbf{P_{in}}$–unaware models are considered. In Fig. 5.37, we show the $\mathrm{Err}_{\mathrm{MAX}}$computed between target and validated gain profiles, plotted with respect to gain levels already after fine–optimization, as we demonstrated that better performance are delivered. If we look at the flat gains case in Fig. 5.37(a), we can see that for all fibers $\mathrm{Err}_{\mathrm{MAX}}$ is below 1 dB, except for the ULLF that shows larger $\mathrm{Err}_{\mathrm{MAX}}$ values for gains higher than 4 dB. This can be easily explained by looking at inset C,(iii) of Fig. 5.27, where gain profiles for ULLF are depicted, showing that in the low frequency region, gains higher than 4 dB are not physically achievable with our experimental setup.

As already observed in the section dedicated to the experimental set–up description, because of the frequency separation between pump lasers and the limitations in their range of power, gains are greater at higher frequencies for all fiber types,

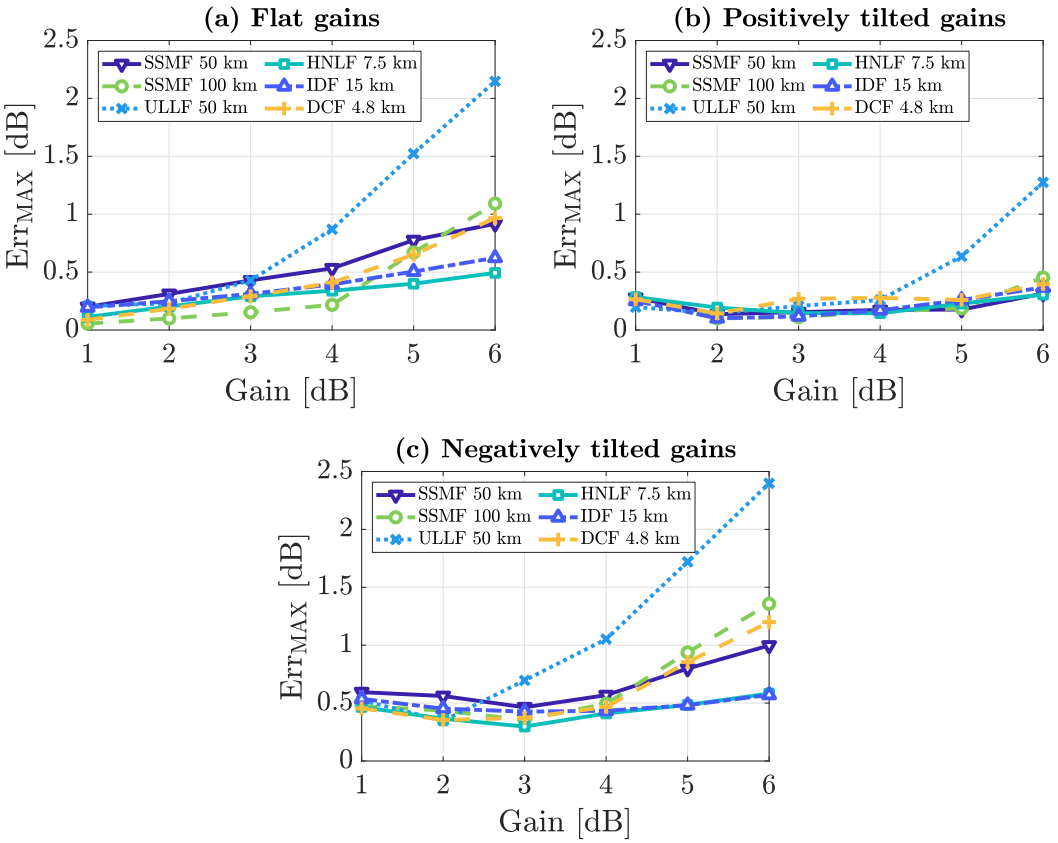


Fig. 5.37 Maximum absolute errors $\mathrm{Err}_{\mathrm{MAX}}$ with respect to gain levels for (a) flat, (b) positively tilted and (c) negatively tilted gain profiles for different types of fibers after fine–optimization. $\mathbf{P}_{\mathbf{in}}$–unaware models are considered.

resulting in positively tilted profiles. Consequently, this intrinsic tilt is learned by the inverse model, benefiting the design of positively tilted gain profiles.

This is confirmed by the lower $\mathrm{Err}_{\mathrm{MAX}}$ values shown in Fig. 5.37(b), which are always below 0.5 dB except for ULLF, for which same considerations as above hold. On the contrary, larger errors are given in case of design of negatively tilted gain profiles (Fig. 5.37(c)), for which, excluding the ULLF, $\mathrm{Err}_{\mathrm{MAX}}$ values up to almost 1.5 dB are reached.

With this extensive analysis considering different types of fiber and different input power spectral distributions, we have shown the robustness of the proposed ML–framework to different conditions. Moreover, two possible approaches have been proposed to improve RA design. The first one consists in performing a further fine–adjustment of pumps powers configuration based on gradient–descent optimization in

case of ANN models not considering the input signal profile information. Instead, the second approach avoids the iterative optimizations required in the first approach and relies on input signal PSD aware models, which include the signal PSD information at the input of both inverse and direct mapping.

5.4.3 Raman Amplifier Design over C–band with Joint Noise Figure Prediction

Generally, when a transmitted optical signal is amplified, ASE noise is generated that adds up to the original signal. The presence of noise significantly affects transmission performance, therefore it is fundamental to evaluate it in order to monitor the SNR. For this purpose, in Chapter 4, dedicated to RA modelling, we have introduced a dedicated artificial neural network to predict ASE noise profiles generated for different pumps configurations in an ultra–fast and accurate way.

In designing RAs for desired gain profiles, adjustments of pump lasers powers imply variations in noise profiles and consequently in the resulting SNR. Thus, here as well, it is crucial to estimate the corresponding noise profile. To do this, we can take advantage of the ANN_{dir} for the ASE noise profiles prediction introduced and extensively studied in Chapter 4. Alternatively, a more sophisticated inverse ANN can be implemented, having at its output the noise information together with pumps configurations.

In this section, we provide a proof of principle for this second approach considering a single case of study: 100 km single span SSMF distributed RA with flat input optical signal. Nevertheless, the proposed inverse ANN model can be easily upgraded to handle different input optical signal and different types of fiber. Noise contribution is assessed in terms of noise figure (NF), for which more details will be given in the following.

Experimental Setup and Data–set Generation

The experimental setup considered for this analysis is shown in Fig. 5.38. It is similar to the experimental setup used in the previous section (Fig. 5.27), with the distributed RA operating in counter–propagation scheme implemented in the same way: same single span of SSMF with $L_{span} = 100$ km and same commercial RA

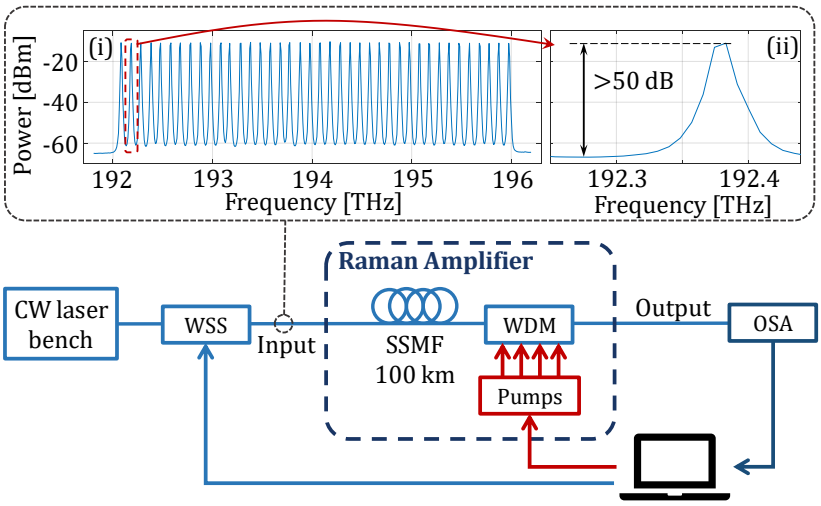


Fig. 5.38 Experimental setup.

pump module with $N_p = 4$ pump lasers (pumps frequency allocation and ranges of power in Table 5.5). $N = 40$ continuous–wave (CW) lasers are used to generate the input signal in the 100–GHz ITU-T grid between 192 THz and 196 THz, which is then flattened by a WSS. The corresponding spectrum at the input of the fiber is shown in Fig. 5.38(i), with the ratio between signal and noise power highlighted in Fig. 5.38(ii), from which we can observe that a high optical SNR is experienced by input signals, thanks to an excursion larger than 50 dB. For this reason, the input signals can be considered nearly shot–noise–limited and the NF can be evaluated using the measured spectral density of the ASE and the signal gain [9]. The noise figure is defined as:

$$\mathrm{NF} = 10\log_{10}(F_n) \tag{5.7}$$

where F_n is the noise factor, generally depending on the optical frequency ν and the baseband frequency f, which is given by:

$$F_n(\nu, f) = \frac{\mathrm{SNR_{in}}}{\mathrm{SNR_{out}}(\nu, f)}. \tag{5.8}$$

$\mathrm{SNR_{in}}$ and $\mathrm{SNR_{out}}(\nu, f)$ are the SNR at the input and at the output of the amplifier, respectively. In particular, we can notice that $\mathrm{SNR_{in}}$ is fixed.

To emphasize the noise advantage of distributed Raman amplifier (which increases with increasing on–off gain) we consider the effective noise factor, as defined

in [51]:

$$F_{\text{eff}} = F_0 \, e^{-2\alpha L} \tag{5.9}$$

where F_0 is the noise factor at the beginning of the fiber span, α the fiber attenuation coefficient at the signal frequency (wavelength) and L the fiber length. Expressing it in decibel, we have the effective noise figure [51]:

$$\text{NF}_{\text{eff}} = \text{NF} - \mathscr{L} \tag{5.10}$$

where $\mathscr{L}$ is the span loss. Interestingly, from the definition $\text{NF}_{\text{eff}} < 1$ dB, which is the property that makes distributed amplification attractive for long–haul WDM systems.

The experimental data–set is generated using the described setup to which we provide a set of $M = 7000$ pumps configurations with powers drawn from uniform distribution in ranges reported in Table 5.5. At the output of the fiber, the corresponding output power spectra are collected through an OSA with 0.1 nm resolution. Then, we compute the relative NF and Raman on–off gain profiles. Like done so far, the latter are computed as the difference between output power spectra in decibel units with pump lasers turned on and off. In our data–set, gain profiles are between 0 dB and 13.5 dB. The resulting data–set is split into two sets of same size (3500 samples each), one used for ANN training and validation and the other for ANN testing.

ML–Framework and ANN Models

Fig. 5.39 illustrates the inverse and direct ANN models, which, compared to models in Figs. 5.28(a) and 5.28(b) of section 5.4.2, introduce the NF information at their output.

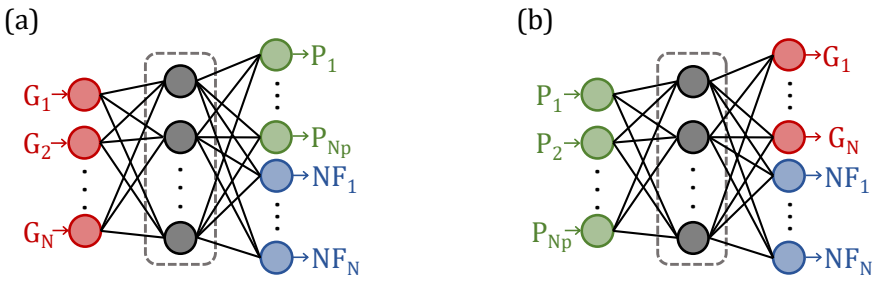


Fig. 5.39 (a) Inverse ANN model ANN_{inv} and (b) direct ANN model ANN_{dir}.

ANN model	f_{act}	N_{HL}	N_{HN}	N_{ANN}	σ_{ANN}	λ_{ANN}
ANN_{inv}	sine	1	1000	20	0.1	10^{-8}
ANN_{dir}	sine	1	1000	1	0.1	10^{-2}

Table 5.10 Hyper-parameters for inverse and direct ANN models.

Indeed, we can see that this modification is present also in the ANN_{dir} used in combination with the GD algorithm in the fine–optimization procedure to improve RA design.

Both models are fully–connected single–hidden layer ANNs and are trained using the random projection method applying K–fold cross–validation with K = 10. This means that 90% of the training data–set is used for training and the remaining 10% for validation. Also for this analysis, ANN hyper–parameters are optimized and the resulting values are reported in Table 5.10.

Experimental Validation of the Predicted Pumps Configurations

Once trained, the ANN_{inv} is tested over the testing data–set of arbitrary gain profiles to assess its prediction performance and is used to design RAs for flat and tilted gain profiles. Afterwards, the predicted pumps powers are given to the experimental setup to collect new measurements.

The resulting *validated* gain profiles are compared to the target ones of the testing data–set, while the effective noise figure profiles are compared with the $\mathrm{NF_{eff}}$ profiles predicted by the ANN_{inv}. Prediction performance is evaluated spectrally in terms of absolute maximum error $\mathrm{Err_{MAX}}$. Different is the estimation of the $\mathrm{NF_{eff}}$ profile accuracy when fine–optimization is performed, which instead is compared to the updated prediction given by the ANN_{dir}. In Figs. 5.40 and 5.41, we report the probability density functions and the cumulative distribution functions, respectively, of $\mathrm{Err_{MAX}}$ for validated gain profiles and $\mathrm{NF_{eff}}$ profiles predictions, both with and without fine–optimization.

If fine–optimization is not performed, similar prediction accuracy is observed between gain and effective noise figure profiles, with similar values of $\mathrm{Err_{MAX}}$ mean and standard deviation. Such quite accurate predictions are also confirmed by the

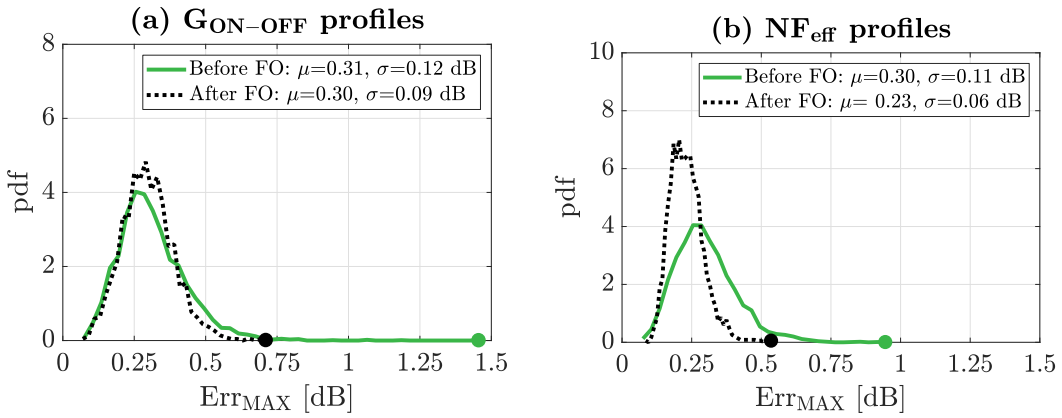


Fig. 5.40 Probability density functions (pdfs) of Err_{MAX} before and after fine–optimization (FO) for (a) Raman on–off gain profiles and (b) effective noise figure NF_{eff} profiles.

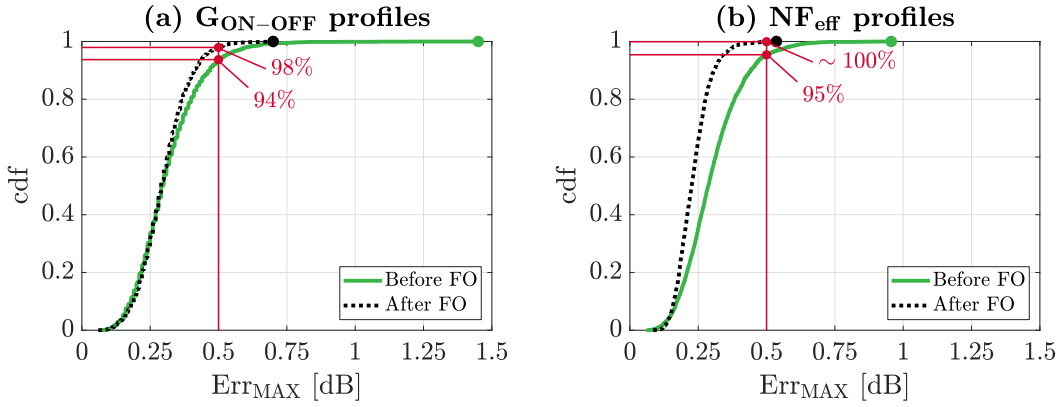


Fig. 5.41 Cumulative distribution functions (cdfs) of Err_{MAX} before and after fine–optimization (FO) for (a) Raman on–off gain profiles and (b) effective noise figure NF_{eff} profiles.

cdfs trends, with 94% and 95% of cases having $Err_{MAX} \leq 0.5$ dB in case of gain and effective noise figure profiles, respectively.

If we apply fine–optimization of pumps powers, in case of gain design, no further substantial improvements are observed overall the testing data–set population. In fact, we can just notice a reduction of the maximum value of Err_{MAX} from 1.45 dB to 0.50 dB, and a growth to 98% of the percentage of cases with Err_{MAX} below 0.5 dB (Fig. 5.41(a)). Instead, in case of effective noise figure, mean and standard deviation of Err_{MAX} significantly reduce from 0.30 dB to 0.23 dB and from 0.11 dB to 0.06 dB, respectively. Moreover, for almost 100% of cases $Err_{MAX} \leq 0.5$ dB.

Particularly, we look at how the worst case (i.e. when the largest Err_{MAX} is experienced) before fine–optimization improves after fine–optimization.

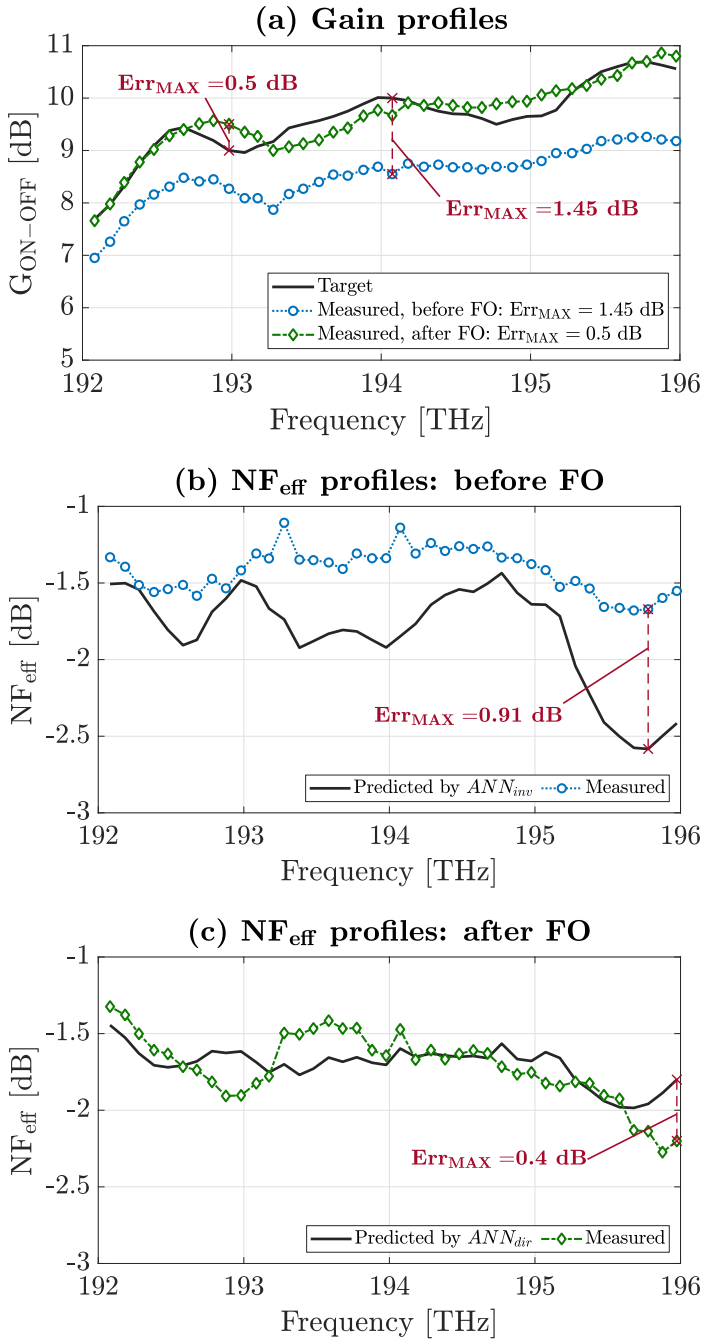


Fig. 5.42 Worst case inverse design (using ANN_{inv}): (a) on–off gain profiles before and after fine–optimization (FO), effective noise figure NF_{eff} (b) before FO and (c) after FO.

In Fig. 5.42(a), we can see that the measured gain profile obtained validating the pumps powers predicted by the single ANN_{inv} model is all shifted towards lower gain values with respect to the target one. While, a pretty good match is obtained after fine–optimization with almost completely overlapped profiles.

For the noise figure, two different figures are shown, one comparing the noise figure profile predicted by the single ANN_{inv} with the corresponding measured profile (Fig. 5.42(b)), and the other one comparing the noise figure profile predicted by the ANN_{dir} after pumps powers fine–adjustment and the corresponding measured profile (Fig. 5.42(c)). As expected, also for the noise figure, better performance is obtained after fine–optimization, with a better matching between predicted and measured noise figure profiles and a reduced Err_{MAX}value (from 0.91 dB before fine–optimization down to 0.4 dB after fine–optimization).

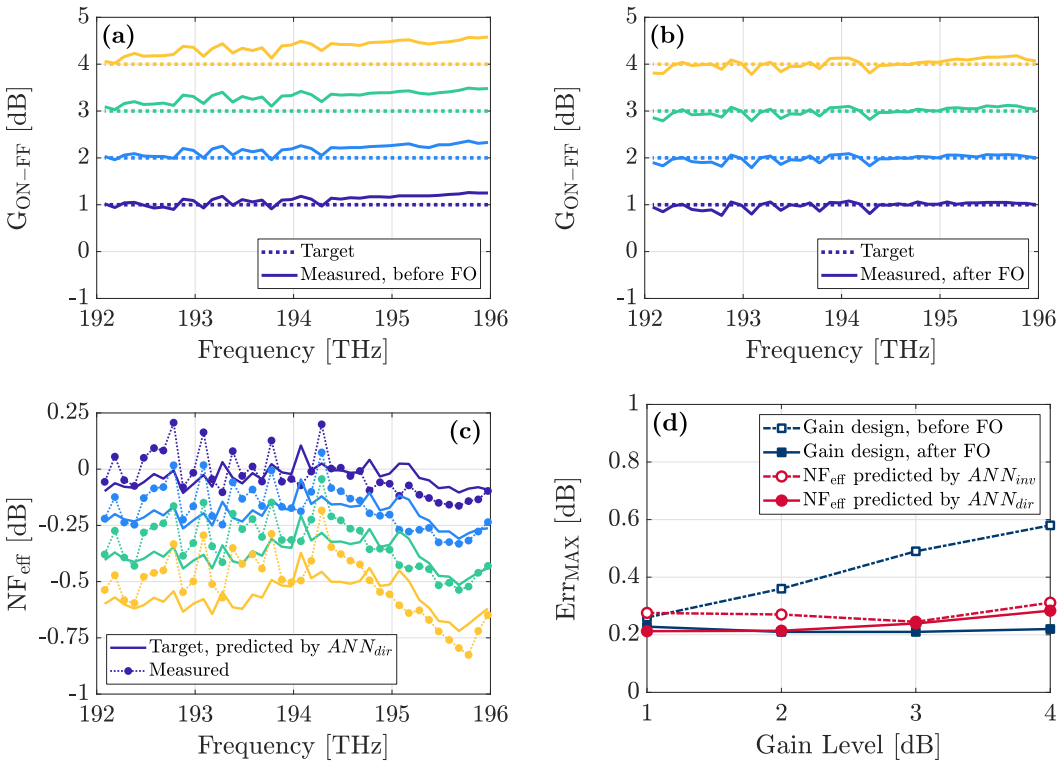


Fig. 5.43 (a) Target and corresponding measured gain profiles before fine–optimization (FO), (b) target and corresponding measured gain profiles after FO, (c) target (predicted by ANN_{dir}) and corresponding NF_{eff} profiles, and (d) Err_{MAX} with respect to flat gain levels for gain design (before and after FO) and NF_{eff} profiles predictions both by ANN_{inv} and ANN_{dir}.

Validation results of pumps powers predicted for flat gain profiles with gain levels from 1 dB to 4 dB are reported in Fig. 5.43. Also in this case, it is not possible to achieve flat gains higher than 4 dB for hardware limitations. If we do not fine–optimize pumps powers (Fig. 5.43(a)), the measured gain profiles show a deviation from the corresponding target, which is larger for higher frequencies. Moreover, we can observe larger deviations for increasing gain levels, which is clearly visible in Fig. 5.43(d), where the Err_{MAX} is plotted with respect to gain levels. Indeed, the Err_{MAX} increases from 0.26 dB, for 1 dB gain level, up to 0.58 dB, for 4 dB gain level.

Although already satisfactory results are obtained in terms of flat gain design without applying the fine–optimization routine, since Err_{MAX} never exceeds 0.6 dB, as expected, higher flatness is obtained after fine–optimization. The measured gain profiles illustrated in Fig. 5.43(b) show a better match with respect to the relative target gain profiles, with Err_{MAX} values $\sim$0.22 dB, as illustrated in Fig. 5.43(d). There, we can also see that Err_{MAX} is flat with gain levels.

Afterwards, we analyze the effective noise figure profiles generated by the designed RAs. In Fig. 5.43(c), we show the NF_{eff} profiles predicted by the ANN_{dir} after the fine–optimization routine (target) and the corresponding measured profiles, which actually are the effective noise figure profiles of designed RAs providing gains in Fig. 5.43(b). The same color in Figs. 5.43(b) and 5.43(c) refers to the same gain level. Thus, at higher gain levels correspond lower NF_{eff} values, showing the dependency of the noise figure on the gain and confirming the advantage of distributed Raman amplification. In general, we can observe that smoother curves are obtained in the higher frequency region, where already gain profiles are smoother, with a good match with target profiles thanks to pumps powers fine–optimization. Likewise, gain fluctuations in the lower frequency region affect the measured NF_{eff}. Nevertheless, Err_{MAX} values are kept low and are almost constant with gain levels, specifically they are always below 0.3 dB, as reported in Fig. 5.43(d).

If fine–optimization is not applied, NF_{eff} profiles do not significantly differ from the target ones. This is proven by the corresponding Err_{MAX} values in Fig. 5.43(d), which are similar to those obtained after fine–optimization.

5.4.4 Raman Amplifier Design over C+L–band

After the comprehensive analysis over the C–band considering experimental data, in this section the study of the performance of the inverse model in designing Raman amplifiers in the C+L–band is carried out.

Experimental Setup and Data–sets Generation

In Fig. 5.44(a) the experimental setup is shown, where the RA can operate as a distributed or a discrete amplifier. For the distributed RA, a 75 km single span of SSMF is considered, while for the discrete RA, a 7.5 km single span of IDF is used. Five non–tunable semiconductor pump laser diodes operating in counter–propagating configuration are employed to achieve gains over the desired frequencies range.

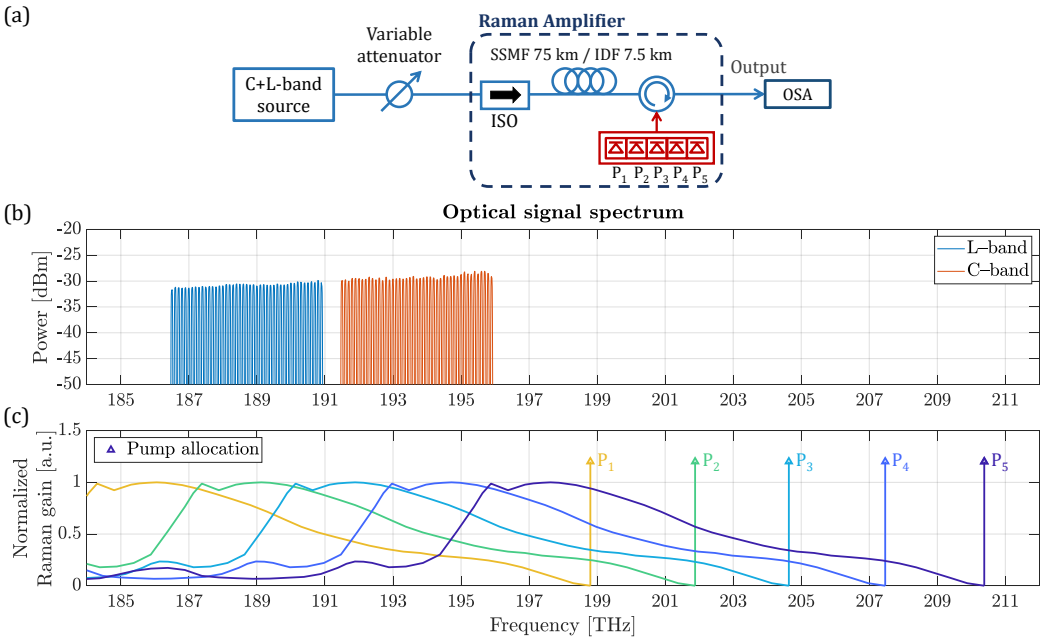


Fig. 5.44 (a) Experimental setup for C+L–band Raman amplifier, (b) input optical signal spectrum and (c) qualitative representation of the normalized pump lasers spectrum and their expected contribution to the overall Raman gain.

Pump	Wavelength [nm]	Frequency [THz]	Currents [A]	
			Distributed RA	Discrete RA
P_1	1508	198.8	[0.2,1.0]	[0.2,0.9]
P_2	1485	201.9	[0.2,1.0]	[0.2,0.8]
P_3	1465	204.6	[0.2,1.2]	[0.2,1.2]
P_4	1445	207.5	[0.2,1.5]	[0.2,1.4]
P_5	1425	210.4	[0.2,1.5]	[0.2,1.5]

Table 5.11 Pump lasers wavelengths, frequencies and current ranges for both distributed and discrete RAs.

A qualitative representation of their spectral allocation and normalized Raman gain efficiency profile contributing to the overall gain is illustrated in Fig. 5.44(c). Pump lasers frequencies (wavelengths) and ranges of driving currents controlling the output power are reported in Table 5.11. The resulting output powers assume values between $\sim$ 16 dBm and $\sim$ 27 dBm.

Two ASE sources for C and L bands are used to generate the optical signal at the input of the Raman amplifier, covering a total of 9.4 THz (77 nm) bandwidth. Then, by means of a WSS, this optical signal is shaped into $N_{ch} = 90$ channels in the 100 GHz ITU-T grid. Its spectrum, measured at $\Delta\lambda = 0.1$ nm resolution, is depicted in Fig. 5.44(b) with two different colors for the different bands. A variable optical attenuator (VOA) is used to adjust the power of the input signal to 10 dBm and 0 dBm for distributed and discrete Raman amplifiers, respectively. Moreover, in case of discrete RA, the IDF is preceded by an isolator (ISO) that avoids the laser pumps powers to enter the C+L–band signal source and that minimizes the multi–path interference induced by the double Rayleigh back–scattering [57]. At the output of the RA, an OSA collects the corresponding output optical spectrum both with pumps turned on and off. Afterwards these optical spectra are used to compute the on–off gain profile, defined as the difference of the two.

For the inverse design ML–framework, we follow the same procedure presented and applied in previous sections. To generate the experimental data–sets, one for each RA type, $M = 5600$ different pumps currents configurations are drawn from uniform distributions with ranges indicated in Table 5.11 and the corresponding gain profiles are measured, discretized over the number of channels. Among the 5600 samples, a set of them are reserved for the ANN training process (training

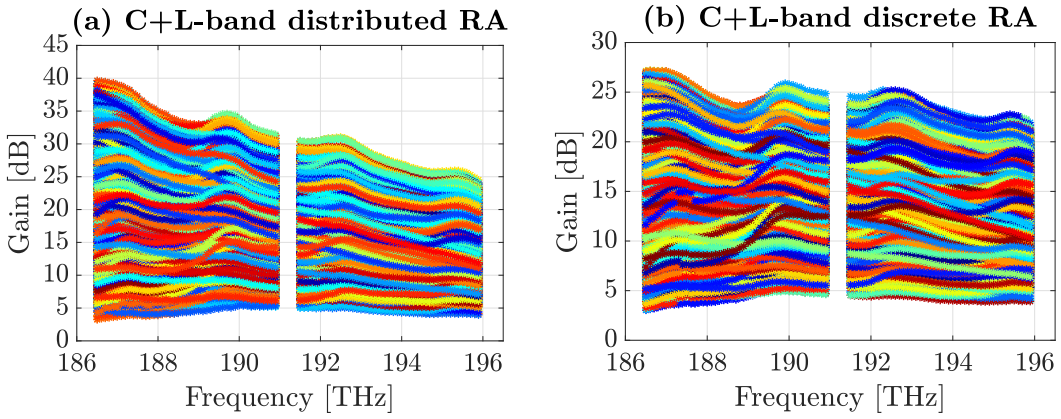


Fig. 5.45 Measured on–off gain profiles for different pump lasers configurations for (a) distributed and (b) discrete Raman amplifiers over the C+L–band.

data–set), more precisely 3464 for distributed RA and 3000 for discrete RA. The remaining data (testing data–set), 2100 and 2600 for distributed RA and discrete RA, respectively, are used for ANN testing with further experimental validation, as already done in previous sections.

Figs. 5.45(a) and 5.45(b) show all the measured on–off gains for distributed and discrete Raman amplifiers, respectively. From there, we can see that lower gain levels are achieved by the discrete RA compared to the distributed RA, because a limitation in the output pumps powers is required to guarantee stable gains. Moreover, unlike the distributed RA case, for the discrete RA higher gain levels are not observed in the low frequencies region due to the decreased driving currents for pumps P_1 and P_2, which leads to a lower depletion of pumps at higher frequency position.

ML–Framework and ANN Models

Like in section 5.4, two ANN models are used to obtain highly accurate Raman amplifiers design: the inverse model ANN_{inv}, which learns the mapping between gain profiles and pump lasers currents, and the direct model ANN_{dir} for fine–design optimization, that learns the mapping between pumps currents and gain profiles. The ANN_{inv} model is trained with the RP method applying the 10–fold cross–validation, meaning that 90% of the training data–set is used for training and 10% for model validation. To obtain better results, hyper–parameters optimization is performed and the model providing the lowest prediction error, evaluated in terms of RMSE, is selected. Instead, the ANN_{dir} model is trained with the most widely used back–

Amplification scheme	**Distributed RA**		**Discrete RA**	
ANN model	ANN_{inv}	ANN_{dir}	ANN_{inv}	ANN_{dir}
Training algorithm	RP	LM	RP	LM
f_{act}	logsig	tanh	sine	tanh
N_{HL}	1	2	1	2
N_{HN}	760	10	500	10
N_{ANN}	20	1	20	1
σ_{ANN}	$6.0 \cdot 10^{-3}$	*	$2.6 \cdot 10^{-3}$	*
λ_{ANN}	$1.0 \cdot 10^{9}$	**	$1.0 \cdot 10^{3}$	**

Table 5.12 Artificial neural network parameters for both distributed and discrete RAs. f_{act}: activation function, N_{HL}: number of hidden layers, N_{HN}: number of hidden nodes, N_{ANN}: model combination parameter, σ_{ANN}: standard deviation of normal distribution considered for ANN weights initialization, λ_{ANN}: regularization parameter. (*) Nguyen-Widrow initialization algorithm [106]. (**) Dynamically modified during training according to [49].

propagation method based on Levenberg–Marquardt (LM) algorithm. Table 5.12 summarizes the parameters for both ANN models.

Testing Results

After training, both models are tested over the respective arbitrary gain profiles testing data–sets and their prediction performance is evaluated. For the ANN_{inv} model, prediction accuracy is estimated for each pump laser as the absolute error computed between predicted and target current values relative to the maximum current excursion. The pdf and cdf over all cases of the testing data–sets are illustrated in Figs. 5.46(a) and 5.46(b) for distributed and discrete RAs, respectively. From there, we can see that high prediction accuracy is achieved for both amplification schemes, with average and standard deviation of pump prediction errors assuming low values. Moreover, for 99% of cases pump prediction errors are below 2.51% and 2.79% for distributed RA and discrete RA, respectively.

Prediction ability of the direct model ANN_{dir} is instead evaluated computing the RMSE and the maximum absolute error $\mathrm{Err}_{\mathrm{MAX}}$ between predicted and target gain profiles. The respective pdfs are shown in Figs. 5.47(a) and 5.47(b), where most of the $\mathrm{Err}_{\mathrm{MAX}}$ values are below 0.5 dB with slightly better performance for the discrete RA. Indeed, the mean of $\mathrm{Err}_{\mathrm{MAX}}$ is 0.29 dB for the distributed RA, while 0.14 dB for the discrete RA.

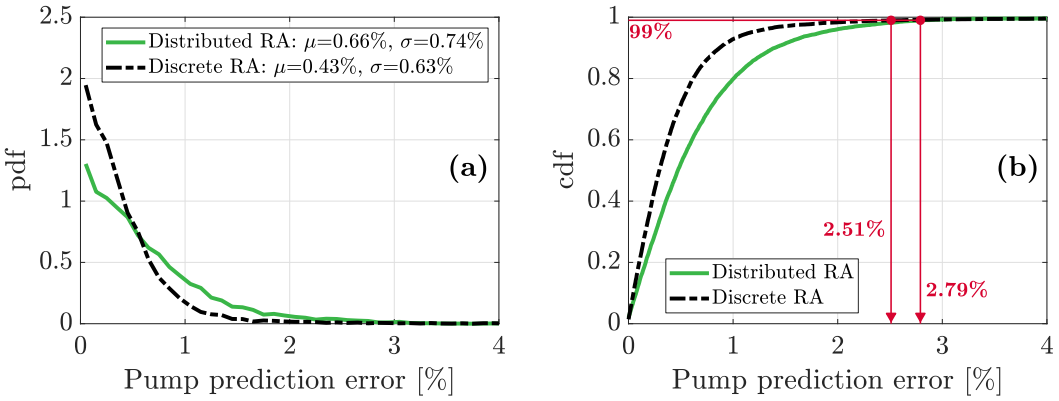


Fig. 5.46 (a) Probability density functions (pdfs) and (c) cumulative distribution functions (cdfs) of the relative absolute errors of pumps currents predicted with the ANN_{inv} for both distributed and discrete RAs.

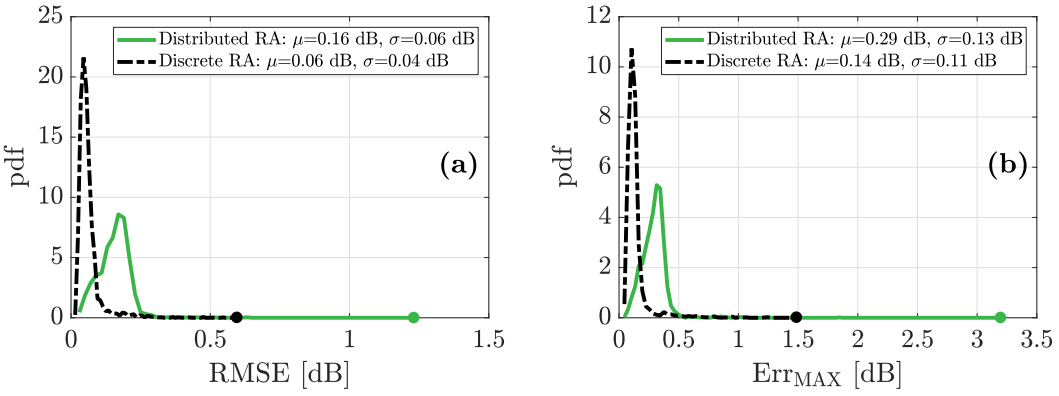


Fig. 5.47 Probability density functions (pdfs) of (a) RMSE and (b) Err_{MAX} for the arbitrary gain profiles predicted by the ANN_{dir} for both distributed and discrete RAs.

The ANN_{inv}, together with the ANN_{dir}, is then used to design RAs providing flat and tilted gain profiles. For the distributed RA, gain levels from 6 dB to 16 dB are considered in steps of 1 dB, instead for the discrete RA, gain levels range from 7 dB to 15 dB, still in step of 1 dB. For the tilted gain profiles, a negative tilt of $\sim$1 dB is assumed over each band, corresponding approximately to a slope of -0.24 dB/THz. This results to be a challenging task, since there is no certainty that the desired flat and tilted gain profiles can be physically achieved by our experimental set–up, for which pump lasers frequency tuning is not possible.

This is confirmed by looking at the number of cases in the training data–sets with gain profiles "close to flatness", such that:

$$(\max([G_1,\ldots,G_{N_{ch}}]) - \min([G_1,\ldots,G_{N_{ch}}])) \leq 1.2 \text{ dB}, \tag{5.11}$$

which are 13 out of 3464 for the distributed RA and 61 out of 3000 for the discrete RA. With such low number of examples, to predict pumps currents for flat gains, the ANN_{inv} has to perform an extrapolation, which explains the poor prediction accuracy, as generally on the contrary multi–layer neural networks are good at interpolating. For this reason, fine–optimization routine is fundamental to improve currents predictions. Similarly in case of tilted gain profiles.

Experimental Validation of the Predicted Pumps Configurations

The predicted pumps configurations are validated experimentally and the resulting gain profiles are compared to the targeted gain profiles. Probability density functions and cumulative distributed functions of root-mean-squared-error per bandwidth (RMSE/BW) and maximum absolute error per bandwidth (Err_{MAX}/BW) for arbitrary gain profiles are shown in Figs. 5.48 and 5.49, respectively, for both distributed and discrete RAs before and after fine–optimization.

For the distributed RA, fine–optimization significantly improves prediction performance since the average of RMSE/BW (Fig. 5.48(a)) reduces from 0.048 dB/THz down to 0.026 dB/THz and the average of Err_{MAX}/BW per bandwidth (Fig. 5.49(a)) decreases from 0.074 dB/THz to 0.044 dB/THz. Further confirmations are given if we look at the cdfs. In fact, from Fig. 5.48(b), whether before fine–optimization only 11.9% of cases have RMSE/BW below 0.04 dB/THz, after fine–optimization the percentage of cases increases up to 98.3%. Similarly for the Err_{MAX}/BW in Fig. 5.49(b), where the percentage of cases for which $Err_{MAX}/BW \leq 0.06$ dB/THz grows from 13.5% to 94.7%.

On the other hand, for the discrete RA, highly accurate predictions of pumps currents are already obtained using the ANN_{inv} only, with mean RMSE/BW equals to 0.023 dB/THz (Fig. 5.48(c)) and mean Err_{MAX}/BW corresponding to 0.035 dB/BW (Fig. 5.49(c)). Indeed, fine–adjustment of pumps currents does not give improvements but slightly worsen performance. We already provided a motivation to this behavior: the inverse model ANN alone already achieves the optimal pumps currents

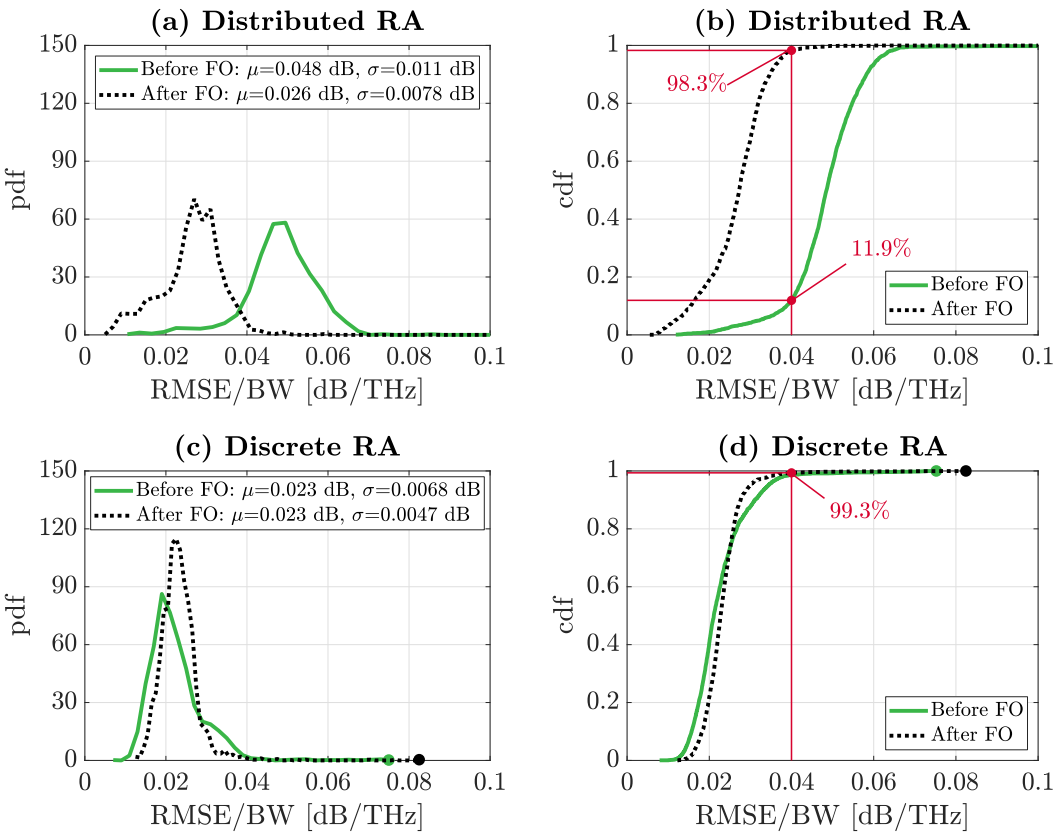


Fig. 5.48 Probability density functions and cumulative density functions of RMSE for *arbitrary* gain profiles before and after fine–optimization for (a),(b) distributed RA and (c),(d) discrete RA.

configuration, thus fine–optimization only adds small perturbations to the predictions. Overall, better performance are obtained in case of discrete RA instead of distributed RA, as confirmed by the 99.3% of cases having RMSE/BW $\leq$0.04 dB/THz (Fig. 5.48(d)) and the 97.2% of cases with Err_{MAX}/BW below 0.06 dB/THz (Fig. 5.49(d)).

Whereupon, we analyze the validation results for flat and tilted gain profiles. Fig. 5.50 shows predicted and target profiles in case of flat and tilted gains for distributed RA (Fig. 5.50(a),(b)) and discrete RA (Fig. 5.50(c),(d)) directly after fine–optimization, for reasons given above. For ease in visualization, only a subset of gains are illustrated in steps of 2 dB, with corresponding pumps configurations reported in Tables 5.13 and 5.14 for distributed and discrete RAs, respectively.

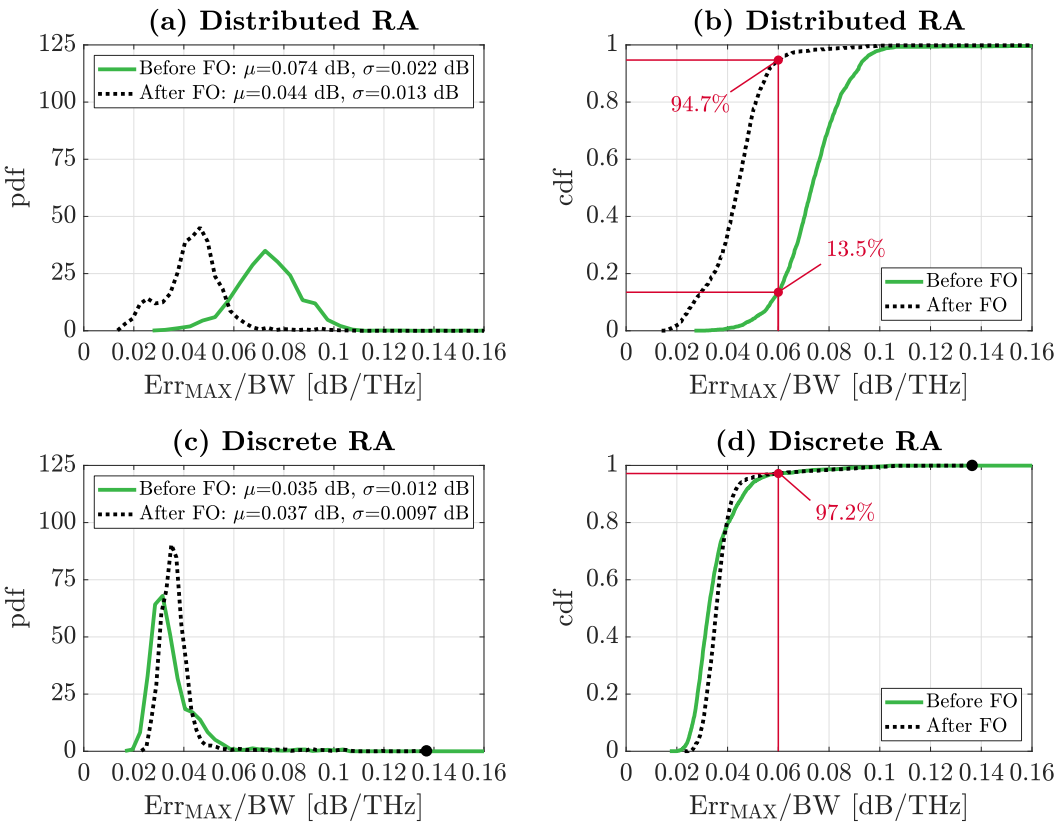


Fig. 5.49 Probability density functions and cumulative density functions of $\mathrm{Err_{MAX}}$ for *arbitrary* gain profiles before and after fine–optimization for (a),(b) distributed RA and (c),(d) discrete RA.

Gain Level	Flat gains					Tilted gains				
	I_1	I_2	I_3	I_4	I_5	I_1	I_2	I_3	I_4	I_5
[dB]	[A]	[A]	[A]	[A]	[A]	[A]	[A]	[A]	[A]	[A]
6	0.43	0.20	0.23	0.42	0.22	0.49	0.20	0.21	0.22	0.20
8	0.49	0.20	0.29	0.54	0.30	0.60	0.20	0.27	0.39	0.21
10	0.52	0.21	0.35	0.65	0.39	0.66	0.20	0.34	0.53	0.26
12	0.50	0.22	0.40	0.74	0.53	0.67	0.21	0.41	0.66	0.35
14	0.45	0.24	0.43	0.82	0.70	0.63	0.23	0.46	0.77	0.47
16	0.39	0.28	0.44	0.87	0.92	0.57	0.26	0.49	0.87	0.63

Table 5.13 Pumps currents configurations for flat and tilted gain profiles after fine–optimization for distributed RA.

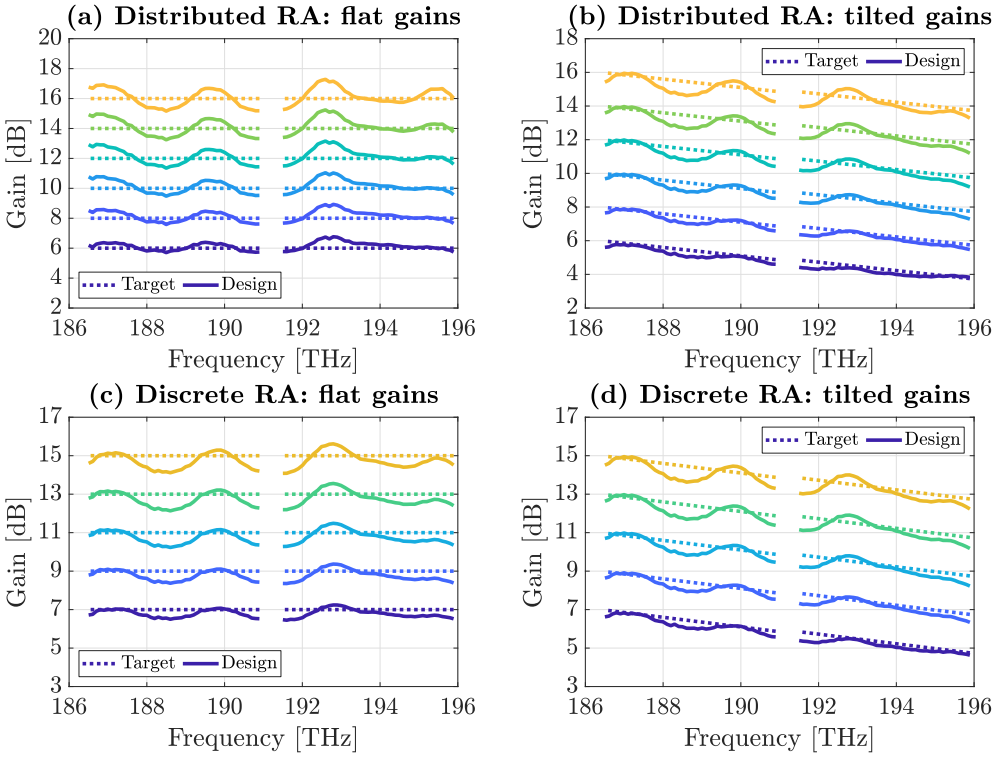


Fig. 5.50 Validated and target profiles for flat and tilted gains in case of (a),(b) distributed RA and (c),(d) discrete RA.

Gain Level	**Flat gains**					**Tilted gains**				
	I_1	I_2	I_3	I_4	I_5	I_1	I_2	I_3	I_4	I_5
[dB]	[A]	[A]	[A]	[A]	[A]	[A]	[A]	[A]	[A]	[A]
7	0.51	0.26	0.27	0.45	0.29	0.65	0.24	0.24	0.29	0.20
9	0.55	0.29	0.34	0.59	0.38	0.71	0.30	0.32	0.45	0.23
11	0.56	0.31	0.40	0.71	0.51	0.72	0.34	0.39	0.60	0.32
13	0.52	0.35	0.45	0.82	0.68	0.69	0.39	0.46	0.73	0.45
15	0.46	0.41	0.48	0.92	0.89	0.63	0.44	0.50	0.85	0.63

Table 5.14 Pumps currents configurations for flat and tilted gain profiles after fine–optimization for discrete RA.

In Fig. 5.50, ripples are observed around the target gain profile, which show growing amplitude with increasing gain level. Afterwards, the goodness of pumps currents predictions are analyzed computing the RMSE/BW and the Err_{MAX}/BW for each level of gain. Fig. 5.51(a) and 5.51(b) show the RMSE/BW with respect

to gain levels for both flat and tilted gains in case of distributed and discrete RAs, respectively. We can see that RMSE/BW values are kept below ∼0.06 dB/THz, with slightly better performance in case of tilted gain profiles.

Moving to the Err_{MAX}/BW curves in Figs. 5.51(c) and 5.51(d), although distributed and discrete RAs have a similar trend in case of tilted gains, with values always below 0.1 dB/THz, this is no longer true in case of flat gain profiles, where the distributed RA presents higher values compared to discrete RA, up to almost 0.14 dB/THz for larger gain levels. To explain this, we have to recall that, as already seen previously, the design of RAs providing flat gain profiles can be challenging because of limitations in our experimental set–up. Therefore, an increase of degrees of freedom, such as the possibility of changing pumps frequencies allocation or a larger number of pump lasers, might result in a more suitable design to achieve the targeted flatness.

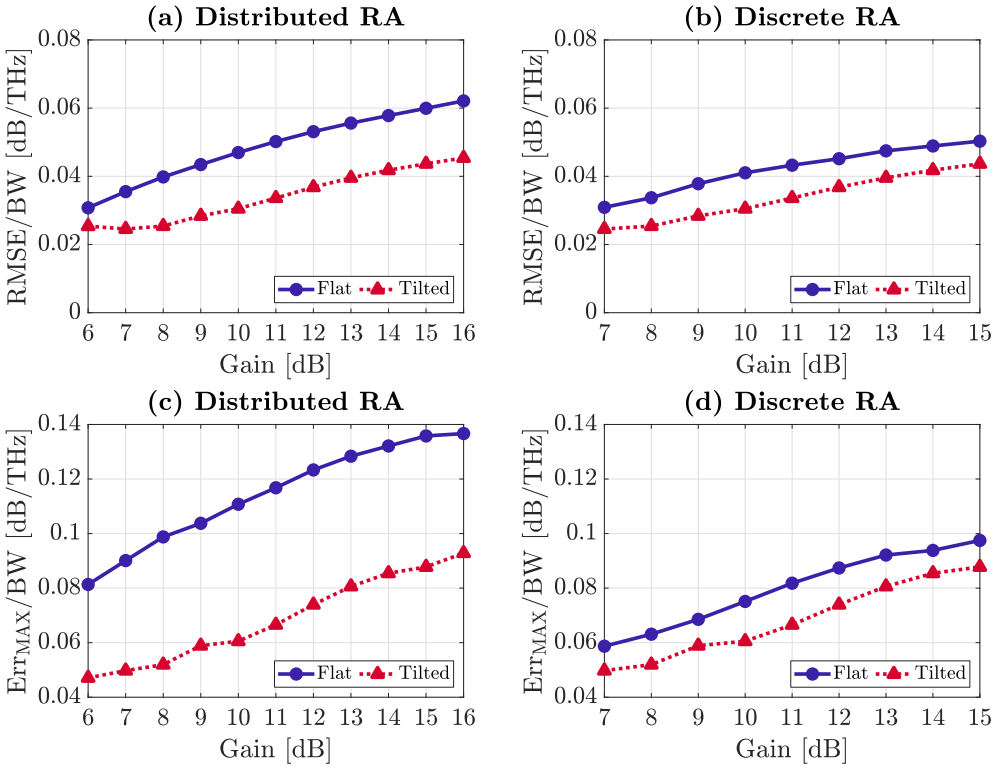


Fig. 5.51 (a),(b) RMSE/BW and (c),(d) Err_{MAX}/BW with respect to gain levels for flat and tilted gain profiles for both distributed and discrete RAs.

5.4.5 Raman Amplifier Design over S+C+L–band

In this section, the ability of the inverse ANN model to design Raman amplifiers is investigated for an even larger bandwidth corresponding to 17.6 THz (140.7 nm) over the S+C+L–band.

Experimental Setup and Data–sets Generation

The considered experimental set–up, illustrated in Fig. 5.52(a), is an upgrade of the experimental set–up in Fig. 5.45(a) working over the C+L–band. Here, the multi–band RA consists of a dual stage sequential discrete RA employing $N_p = 8$ pump lasers operating in counter–propagating pumping scheme. In particular, since we have two sequential stages, the strong depletion of higher frequency (or shorter wavelength) pumps is reduced [67]. The first stage provides gains in the S–band and is made of a 7.5 km span of IDF and three pump lasers, which we refer to as P_6, P_7 and P_8. Instead, the second stage provides gains over the C+L–band and is the same used in section 5.4.4 made of the 7.5 km IDF span.

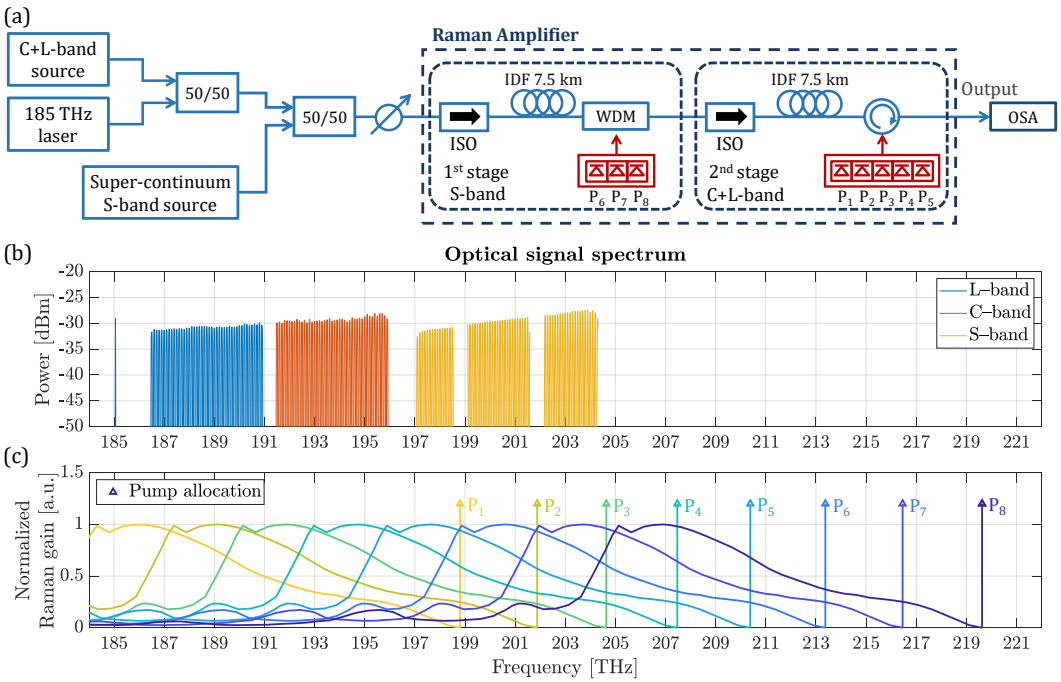


Fig. 5.52 (a) Experimental set–up for the dual stage discrete Raman amplifier over the S+C+L–band, (b) input optical signal spectrum and (c) qualitative representation of the normalized pump lasers spectrum and their expected contribution to the overall Raman gain.

	Frequency [THz]	Wavelength [nm]	Current [A]	Voltage [V]
P_1	198.8	1508	[0.20,1.00]	-
P_2	201.9	1485	[0.20,0.80]	-
P_3	204.6	1465	[0.20,1.00]	-
P_4	207.5	1445	[0.20,1.40]	-
P_5	210.4	1425	[0.20,1.40]	-
P_6	213.4	1405	[0.20,1.20]	-
P_7	216.5	1385	[0.60,1.30]	-
P_8	221.1	1365	-	[1.80,2.40]

Table 5.15 Pumps lasers parameters for the S+C+L–band Raman amplifier.

In Fig. 5.53(c), a qualitative representation of Raman pump laser spectrum and corresponding normalized Raman gain efficiency profiles are reported. Pumps lasers P_1 to P_7 are semiconductor laser diodes driven by currents providing power to the RA from $\sim$16 dBm to $\sim$27 dBm, like in the experimental set–up working in the C+L–band, while pump P_8 is a Raman–based fiber laser controlled by voltage providing power from $\sim$20 dBm to $\sim$27 dBm. Pump lasers parameters are reported in Table 5.15.

To generate the optical signal at the input of the dual–stage RA, the optical signal generated by the C+L–band source is combined together with a single frequency laser operating at 185 THz and the optical signal generated by a supercontinuum S–band source [36]. Since pumps P_1 and P_2 of the first stage fall within the S–band signal, we have to remove some channels in the S–band to prevent from the overlap with the Rayleigh back–scattered components of pumps [59], resulting in $N_{ch} = 148$ channels in the 100 GHz ITU-T grid. By means of a VOA, the input signal power is set to 7 dBm. The spectrum of the multi–band input optical signal is depicted in Fig. 5.53(b), where we can clearly see the gaps left in the S–band.

For data–set generation, we follow the same procedure as in previous sections. In particular, $M = 4025$ different pumps currents and voltages are drawn from uniform distributions over ranges in Table 5.15 and the corresponding gain profiles are measured, each with 148 points. Gain profiles of the resulting data–set $\mathscr{D}^{M\times(N_{ch}+N_{N_p})} = \{(G_1^i,\ldots,G_{Nch}{}^i, I_1^i,\ldots,I_7^i, V_8^i \mid i = 1,\ldots,M)\}$ are illustrated in Fig. 5.53. The collected data–set is then divided into training and testing data–sets, containing 3000 and 1025 samples, respectively.

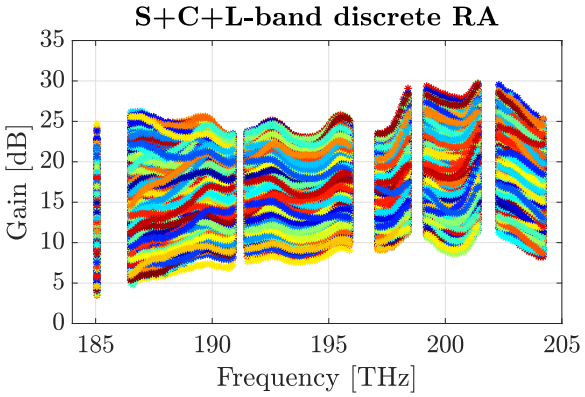


Fig. 5.53 Measured on–off gain profiles for different pump lasers configurations for dual–stage discrete Raman amplifier over the S+C+L–band.

ANN model	ANN_{inv}	ANN_{dir}
Training algorithm	RP	RP
f_{act}	sine	tanh
N_{HL}	1	1
N_{HN}	500	500
N_{ANN}	20	1
σ_{ANN}	$1.0 \cdot 10^{-2}$	$1.0 \cdot 10^{-3}$
λ_{ANN}	$1.0 \cdot 10^{4}$	$1.0 \cdot 10^{8}$

Table 5.16 Artificial neural network parameters for S+C+L–band dual–stage discrete RA. f_{act}: activation function, N_{HL}: number of hidden layers, N_{HN}: number of hidden nodes, N_{ANN}: model combination parameter, σ_{ANN}: standard deviation of normal distribution considered for ANN weights initialization, λ_{ANN}: regularization parameter.

ML–Framework and ANN Models

Similarly to the C+L–band RAs case, ANN_{inv} and ANN_{dir} models, learning respectively the inverse mapping between gain profiles and pumps configurations (currents and voltages) and the direct mapping between pumps configurations and gain profiles, are optimized to improve predictions. The optimized hyper–parameters are shown in Table 5.16, where we can see that random–projection (RP) method is used not only to train the ANN_{inv} inverse model, as in previous section, but the direct model ANN_{dir} as well. Unlike the C+L–band RAs case, we prefer to use RP instead of the Levenberg–Marquardt based BP method for convergence time reasons. Indeed, due

to larger input and output dimensions of RA operating in the S+C+L–band, more time would be required by the second approach to converge, while RP already proved to be faster.

Testing Results

Once trained, to assess prediction performance of ANN models we test them over the respective testing data–sets of arbitrary gain profiles. The prediction accuracy is evaluated as done for the C+L–band RAs. In Fig. 5.54(a), the pdf and cdf of relative absolute errors of all cases of pumps currents and voltages predicted by the ANN_{inv} are shown. With an average percentage of pump prediction error equals to 0.67% and with 99% of cases having relative pump prediction error always below 3.57%, results are similar to those obtained for RAs working on the C+L–band, confirming the robustness of the ML–based inverse design also in multi–band scenarios.

Prediction performance of the ANN_{dir}, that will be used in the fine–optimization routine, are instead analyzed in Fig. 5.54(b), depicting the pdfs of RMSE and Err_{MAX} computed between target and predicted gain profiles. Compared to the predictions of gain profiles for the RAs working in C+L–band, the trend of RMSE curve is similar to the discrete RA with very low values. However, worse performance are observed for the Err_{MAX}, as for most of the cases, the largest Err_{MAX} is experienced at 185 THz, which indeed corresponds to the single frequency line generated separately. A possible reason can be that gains at 185 THz are more subjected to fluctuations with respect to gains in other frequency channels, due to a less stable source.

Afterwards, the ANN_{inv} and the ANN_{dir} are used to design RAs to obtain flat and tilted gains with gain levels from 14 dB and 20 dB. For tilted gains, a negative slope of -0.20 dB/THz is considered to have a tilt of $\sim$1 dB over each band.

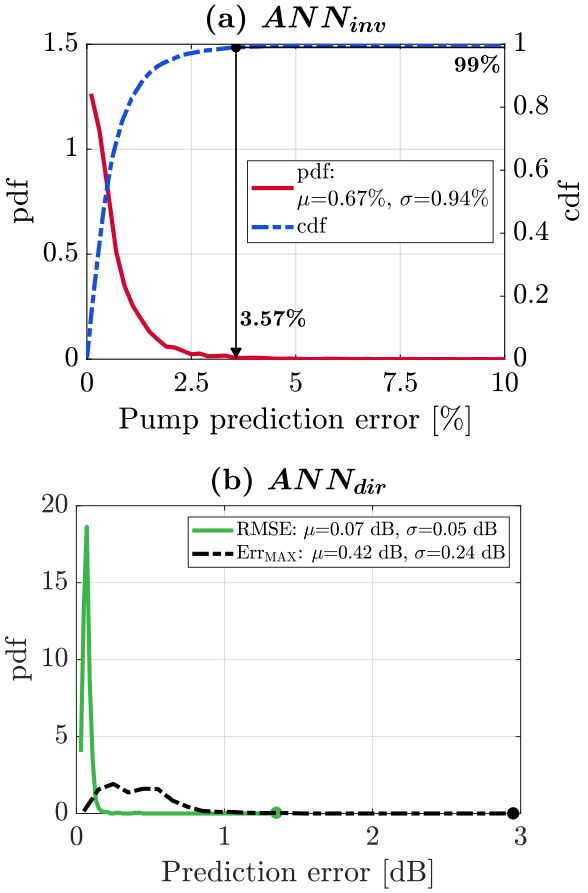


Fig. 5.54 (a) Probability density function (pdf) and cumulative distribution function (cdf) of relative absolute errors of pumps currents and voltages predicted with the ANN_{inv}, and (b) pdfs of RMSE and Err_{MAX} for the arbitrary gain profiles predicted by the ANN_{dir} for S+C+L–band dual–stage discrete RA.

Experimental Validation of the Predicted Pumps Configurations

Finally, we provide the results of experimental validation of predicted pumps currents and voltages for arbitrary, flat and tilted gain profiles. Prediction accuracy is evaluated computing the RMSE/BW and the Err_{MAX}/BW between validated and target gains. For arbitrary gain profiles, the corresponding pdfs and cdfs are illustrated in Fig. 5.55 both applying and not applying fine–optimization of pumps configurations. As already seen for the C+L–band discrete RA, with mean RMSE/BW and mean Err_{MAX}/BW equal to 0.0047 dB/THz and 0.016 dB/RHz respectively, already with the single ANN_{inv} highly accurate predictions are obtained.

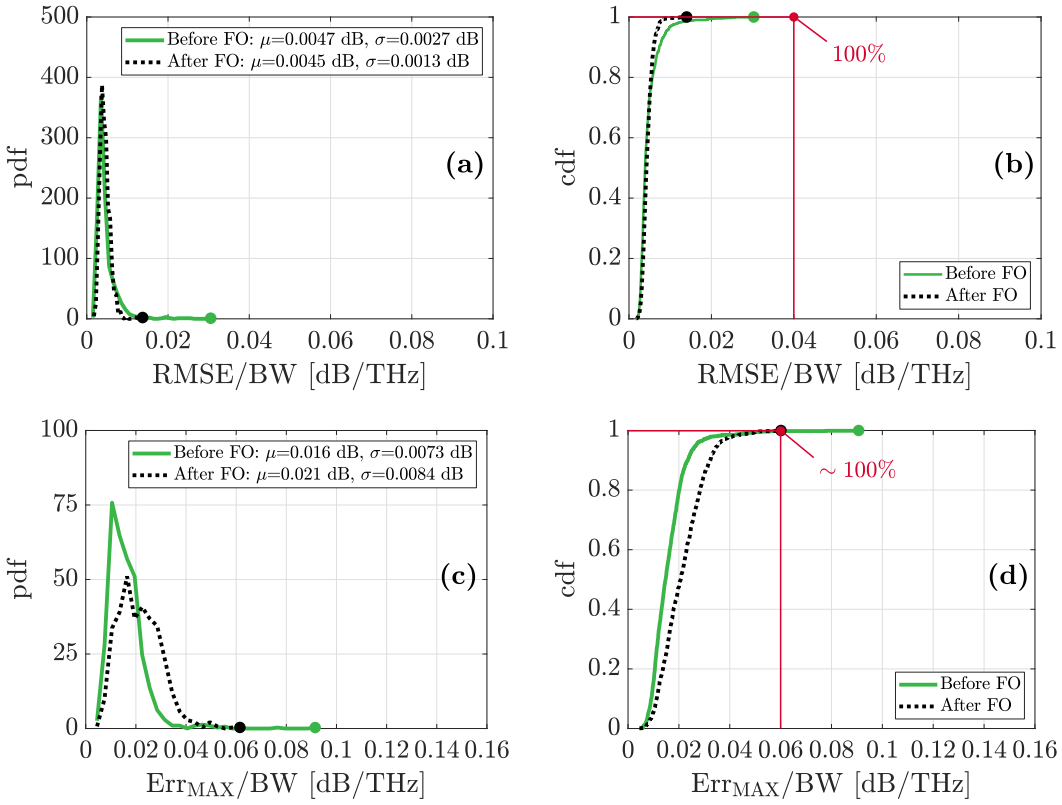


Fig. 5.55 Probability density functions and cumulative density functions of (a)-(b) RMSE and (c)-(d) $\mathrm{Err_{MAX}}$ for *arbitrary* gain profiles before and after fine–optimization for S+C+L–band dual–stage discrete RA.

This is also showed by the cdfs, for which all cases have RMSE/BW below 0.04 dB/THz and almost 100% of cases with $\mathrm{Err_{MAX}/BW} \leq 0.06$ dB/THz. Moreover, compared to the C+L–band discrete RA case, we observe approximately a 2 dB reduction for mean and standard deviation. In fact, although the ANN_{inv} models of the two amplification schemes provide similar prediction accuracy in terms of $\mathrm{Err_{MAX}}$, the bandwidth by which the errors are divided is larger in case of S+C+L–band (almost twice the C+L–band). If errors are not normalized with respect to bandwidth, prediction accuracy for C+L–band RA and S+C+L–band RA is qualitatively comparable. Here as well, fine–optimization does not bring substantial improvement.

On the contrary, results related to flat and tilted gain profiles, that will be presented in the following, are all obtained fine–adjusting pumps configurations.

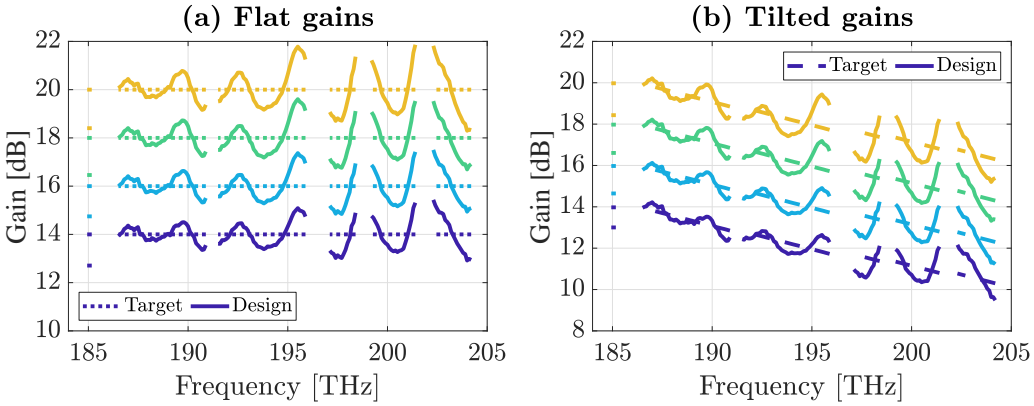


Fig. 5.56 Validated and target profiles for flat and tilted gains in case of S+C+L–band dual–stage discrete RA.

Validated flat and tilted gain profiles are plotted in Figs. 5.56(a) and 5.56(b), respectively, together with the targeted gains. Like in section 5.4.4, for a better visualization, only a subset of gains is shown. The corresponding pumps design can be found in Table 5.17 in case of flat gain profiles and in Table 5.18 in case of tilted gains.

Gain Level [dB]	I_1 [A]	I_2 [A]	I_3 [A]	I_4 [A]	I_5 [A]	I_6 [A]	I_7 [A]	V_8 [V]
14	0.72	0.47	0.49	0.72	0.61	0.37	0.60	2.01
16	0.70	0.49	0.56	0.82	0.73	0.37	0.60	2.23
18	0.67	0.50	0.62	0.92	0.89	0.38	0.64	2.40
22	0.65	0.49	0.65	1.07	1.09	0.41	0.85	2.40

Table 5.17 Pumps currents and voltage configurations for flat gain profiles after fine–optimization.

Gain Level [dB]	I_1 [A]	I_2 [A]	I_3 [A]	I_4 [A]	I_5 [A]	I_6 [A]	I_7 [A]	V_8 [V]
14	0.94	0.51	0.47	0.63	0.42	0.38	0.60	1.80
16	0.88	0.55	0.56	0.72	0.56	0.39	0.60	1.98
18	0.85	0.57	0.65	0.82	0.69	0.39	0.60	2.19
22	0.82	0.58	0.72	0.93	0.83	0.39	0.62	2.40

Table 5.18 Pumps currents and voltage configurations for tilted gain profiles after fine–optimization.

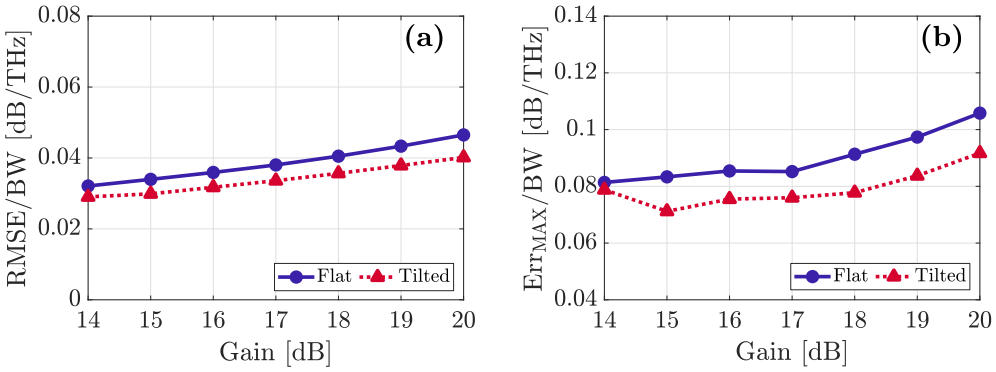


Fig. 5.57 (a) RMSE/BW and (b) Err$_{MAX}$/BW with respect to gain levels for flat and tilted gain profiles for for S+C+L–band dual–stage discrete RA.

Considerations similar to those provided in section 5.4.4 hold here as well, such as ripples around target gain profiles, with amplitude considerably increasing with increasing gain level, especially in high frequency region, up to 2 dB of maximum error with respect to the target. This phenomenon might be related to the power instabilities observed in the supercontinuum S–band source and in the Raman–based fiber laser P_8.

In Fig. 5.57, we report the RMSE/BW and Err$_{MAX}$/BW of flat and tilted gain profiles for all gain levels. As already observed for the C+L–band discrete RA, for both RMSE/BW and Err$_{MAX}$/BW, there is an increasing trend with the growing gain level, with slightly lower errors in case of tilted gains. Though we observe this increasing trend, errors are still kept low, since RMSE/BW never exceeds 0.05 dB/THz and Err$_{MAX}$/BW is always below 0.11 dB/THz.

5.5 Conclusions

This Chapter have shown the use of ML and ANN to design RAs for different target gain profiles: arbitrary, flat and tilted. The method has been extensively investigated considering both synthetic and experimental data for different bandwidth in the context of multi–band systems scenario. The analysis over synthetic data is performed considering distributed RA operating over C– and C+L–band. In case of C–band, RA configurations with two, three and five pumps have been studied, while in case of C+L–band, four pumps are used to build the RA. Moreover, still for synthetic data, the inverse ANN has been extended to partial spectral loads.

Regarding the analysis over experimental data, the approach has been investigated for systems operating in C–, C+L– and S+C+L–band. In particular, both discrete RAs (in C+L– and S+C+L–band) and distributed RAs (in C– and C+L–band) have been employed. For systems operating in C–band, a further extensive analysis considering different fiber types and input power profile has been carried out. In addition, a model simultaneously designing the RA and evaluating the resulting noise figure has been proposed.

For all these different scenarios, the method proved to be able to accurately design RAs for desired gain profiles, also thanks to the application of a further gradient–based optimization procedure to fine–tune pumps configurations. For some cases, the design of RAs for flat and tilted gain profiles was more challenging, or even poor, because some specific gain profiles are not physically feasible due to limitations in the available hardware, in particular in the pumps allocation.

Chapter 6

Conclusions

In this thesis, we applied ML and ANNs to different optical communication applications, with a particular interest in multi–band systems. Indeed, extending the transmission bandwidth beyond the standard C–band towards multiple bands (O–, E–, S–, C– and L–bands) is one of the viable and widely investigated solution for capacity increase to keep up with the continuously growing Internet data traffic. Nevertheless, the extension of the transmission bandwidth requires the adoption of other amplification technologies different from EDFAs, which in fact are characterized by a limited operational region (C–band and an extension in the L–band). Among possible solutions, in this thesis we considered Raman amplifiers, especially thanks to their ability of providing gains potentially over all optical bands, with adjustable gain profiles by just properly tuning pump lasers frequency (wavelength) and power, and for cost effective advantages, as Raman amplification exploiting SRS effect can occur in any fiber, allowing the use of already deployed systems.

At the control plane level, optical networks are constantly monitored to assess their QoT performance, requiring a real-time evaluation of linear and nonlinear impairments affecting the system, and to estimate the behavior of their components, such as RAs, when changes in condition and configuration occur for a proper network response. These operations, based on approximated models and numerical solutions of set of ODEs, can become time and computational demanding in a multi-band scenario, where the number of WDM channels and pumping RA lasers is larger, with the risk of not satisfying the speed requirement of real-time systems. For this reason,

in this thesis we have proposed the use of ML and ANNs as alternative solutions to standard approaches.

More precisely, in Chapter 3, we introduced two ANN models to determine, in an ultra-fast and highly accurate way, power profiles evolution along the optical fiber in frequency and distance when ISRS effect occurs among WDM channels. In particular, we have shown that, with a proper design of the ANN architecture, we were able to achieve a target prediction absolute error of 0.1 dB between predicted and true power profiles. The prediction accuracy and the computational time of the obtained results make the proposed method suitable to be used in support of real-time QoT assessment and network performance monitoring for the estimation of the impact of linear and nonlinear impairments.

Then, in Chapter 4, we illustrated the use of ML to analyze the behavior of Raman amplifiers, particularly to predict their gain and ASE noise profiles for a given RA configuration. The analysis has been conducted in full–load and partial–load scenarios, proving the need of an ANN model accounting for the input spectral load when partial–loads are present. A comprehensive statistical analysis of prediction errors distribution with respect to spectral occupation and classes of loads was presented for the partial–load scenario. Moreover, we introduced a generalized ANN model handling different fiber types and span length, showing high prediction abilities especially for SSMF and ULLF.

Finally, in Chapter 5 we applied ML and ANNs to the design of RAs to obtain desired arbitrary, flat and tilted gain profiles. The study was carried out for different multi–bands scenarios over both synthetic and experimental data and considering diverse RA types (distributed and discrete) and configurations. The results showed very high design ability of the inverse ANN and a significant robustness of the method when larger bandwidth have been considered, with some difficulties in case of flat and tilted gain profiles due to hardware limitations (in case of experimental data). In some cases, substantial improvements were observed when the inverse ANN was used in combination to the direct ANN to fine-tune pumps configurations by means of gradient-descent optimization algorithm.

The obtained results are very promising and suggest that the proposed solutions based on ML and ANNs represent a possible technology to be adopted in the context of multi-band systems. Future works and further extensions of the proposed methodologies would include the experimental validation of ANN Model1 and

Model 2 introduced in Chapter 3 for the power profiles prediction in presence of ISRS, also considering different values of launch power per channel, as a stronger tilt in the power profile is observed for higher powers. Moreover, also the simulation analysis could be further extended to handle a more generalized scenario, considering multiple types of fibers and different values of launch power per channel.

The robustness of direct and inverse ANN models for RA analysis and design could be also further studied with respect to different values of power per channel and extended towards the other transmission bands not investigated in this thesis.

Furthermore, as the ANNs here presented have been used in specific scenarios, other ML learning techniques, such as *transfer learning*, could be implemented to improve the generalization capabilities in handling different scenarios and make them even more suitable and robust for deployment in on-field systems. In addition, in the context of multi-span links characterizing real optical fiber systems, a further possible step would consist in implementing an ANN optimizing the RA pumps to maximize the generalized signal-to-noise ratio (GSNR), in particular dealing with different partial spectral loads.

References

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from https://www.tensorflow.org/.

[2] G. P. Agrawal. *Nonlinear Fiber Optics, 3rd Edition*. Academic press, 2001.

[3] G. P. Agrawal. *Fiber-Optic Communication Systems*. Wiley, 2010.

[4] G. P. Agrawal. *Nonlinear Fiber Optics, 5th Edition*. Academic press, 2012.

[5] R. C. Alferness. *The Evolution of Optical Transport Networks*, pages 1–19. Springer International Publishing, Cham, 2020.

[6] J. D. Ania-Castañón, A. A. Pustovskikh, S. M. Kobtsev, and S. K. Turitsyn. Simple design method for gain-flattened three-pump Raman amplifiers. *Opt. Quant. Electron.*, 39(3):213–220, 2 2007.

[7] A. Arnould et al. 103 nm Ultra-Wideband Hybrid Raman/SOA Transmission Over 3 x 100 km SSMF. *J. Lightwave Technol.*, 38(2):504–508, 1 2020.

[8] N. W. Ashcroft and N. D. Mermin. *Solid State Physics*. Philadelphia, PA: Saunders College, 1976.

[9] D. M. Baney, P. Gallion, and R. S. Tucker. Theory and Measurement Techniques for the Noise Figure of Optical Amplifiers. *Optical Fiber Technology*, 6(2):122–154, 2000.

[10] C. M. Bishop. *Pattern recognition and machine learning*. Springer, 2006.

[11] G. Borraccini, S. Straullu, A. D'Amico, A. Nespola, S. Piciaccia, A. Tanzi, G. Galimberti, and V. Curri. Autonomous Raman amplifiers in multi-band software-defined optical transport networks. *Journal of Optical Communications and Networking*, 13(10):E53–E62, 2021.

[12] G. Bosco, V. Curri, A. Carena, P. Poggiolini, and F. Forghieri. On the Performance of Nyquist-WDM Terabit Superchannels based on PM-BPSK, PM-QPSK, PM-8QAM or PM-16QAM subcarriers. *Journal of Lightwave Technology*, 29(2):53–61, 1 2011.

[13] R. W. Boyd. *Nonlinear Optics, 3th Edition*. Academic press, 2008.

[14] J. Bromage. Raman amplification for fiber communications systems. *Journal of Lightwave Technology*, 22(1):79–93, 1 2004.

[15] M. Cantono, D. Pilori, A. Ferrari, C. Catanese, J. Thouras, J.-L. Augé, and V. Curri. On the Interplay of Nonlinear Interference Generation With Stimulated Raman Scattering for QoT Estimation. *Journal of Lightwave Technology*, 36(15):3131–3141, 2018.

[16] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri. EGN model of non-linear fiber propagation. *Opt. Express*, 22(13):16335–16362, 6 2014.

[17] J. Chen and H. Jiang. Optimal Design of Gain-Flattened Raman Fiber Amplifiers Using a Hybrid Approach Combining Randomized Neural Networks and Differential Evolution Algorithm. *IEEE Photonics Journal*, 10(2):1–15, 2018.

[18] Y. Chen, J. Du, Y. Huang, K. Xu, and Z. He. Intelligent gain flattening in wavelength and space domain for FMF Raman amplification by machine learning based inverse design. *Opt. Express*, 28(8):11911–11920, 4 2020.

[19] J. Cho and P. J. Winzer. Probabilistic Constellation Shaping for Optical Fiber Communications. *Journal of Lightwave Technology*, 37(6):1590–1607, 2019.

[20] F. Chollet et al. Keras. https://keras.io, 2015.

[21] F. Chollet et al. Keras Adam optimizer. https://keras.io/api/optimizers/adam/, 2015. Accessed: 2021-05-07.

[22] F. Chollet et al. Keras Layer weight initializers. https://keras.io/api/layers/initializers/, 2015. Accessed: 2021-05-07.

[23] D. N. Christodoulides and R. B. Jander. Evolution of stimulated Raman crosstalk in wavelength division multiplexed systems. *IEEE Photonics Technology Letters*, 8(12):1722–1724, 1996.

[24] Cisco. Cisco Annual Internet Report. https://www.cisco.com/c/en/us/solutions/executive-perspectives/annual-internet-report/index.html. Accessed: 2021-10-29.

[25] Cisco. Practical Aspects of Raman Amplifier. https://www.cisco.com/c/en/us/support/docs/optical-networking/ons-15454-m6-multiservice-transport-platform-mstp/212834-practical-aspects-of-raman-amplifier.pdf. Accessed: 2022-06-02.

[26] V. Curri and A. Carena. Merit of Raman Pumping in Uniform and Uncompensated Links Supporting NyWDM Transmission. *Journal of Lightwave Technology*, 34(2):554–565, 2016.

[27] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif. Properties of nonlinear noise in long, dispersion-uncompensated fiber links. *Opt. Express*, 21(22):25685–25699, 11 2013.

[28] U. C. de Moura, A. M. R. Brusin, A. Carena, D. Zibar, and F. D. Ros. Simultaneous gain profile design and noise figure prediction for Raman amplifiers using machine learning. *Opt. Lett.*, 46(5):1157–1160, 3 2021.

[29] U. C. de Moura, F. Da Ros, A. M. Rosa Brusin, A. Carena, and D. Zibar. Experimental Demonstration of Arbitrary Raman Gain-Profile Designs using Machine Learning. In *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3, 2020.

[30] U. C. de Moura, M. A. Iqbal, M. Kamalian, L. Krzczanowicz, F. Da Ros, A. M. Rosa Brusin, A. Carena, W. Forysiak, S. Turitsyn, and D. Zibar. Multi–Band Programmable Gain Raman Amplifier. *Journal of Lightwave Technology*, 39(2):429–438, 2021.

[31] U. C. de Moura, F. D. Ros, A. M. Rosa Brusin, A. Carena, and D. Zibar. Experimental Characterization of Raman Amplifier Optimization Through Inverse System Design. *Journal of Lightwave Technology*, 39(4):1162–1170, 2021.

[32] E. Desurvire, editor. *Erbium-Doped Fiber Amplifiers: Principles and Applications*. Wiley, 2002.

[33] G. Di Rosa, R. Emmerich, M. Ribeiro Sena, J. K. Fischer, C. Schubert, R. Freund, and A. Richter. Characterization, Monitoring, and Mitigation of the I/Q Imbalance in Standard C-Band Transceivers in Multi-band Systems. *Journal of Lightwave Technology*, pages 1–1, 2022.

[34] V. V. Dvoyrin, V. M. Mashinsky, and S. K. Turitsyn. Bismuth-Doped Fiber Amplifier Operating in the Spectrally Adjacent to EDFA Range of 1425-1500 nm. In *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3, 2020.

[35] S. Echeverri-Chacón, J. J. Mohr, J. J. V. Olmos, P. Dawe, B. V. Pedersen, T. Franck, and S. B. Christensen. Transmitter and Dispersion Eye Closure Quaternary (TDECQ) and Its Sensitivity to Impairments in PAM4 Waveforms. *J. Lightwave Technol.*, 37(3):852–860, 2 2019.

[36] A. E. El-Taher, J. D. Ania-Castañón, V. Karalekas, and P. Harper. High efficiency supercontinuum generation using ultra-long Raman fiber cavities. *Opt. Express*, 17(20):17909–17915, 9 2009.

[37] R. Emmerich, M. Sena, R. Elschner, C. Schmidt-Langhorst, I. Sackey, C. Schubert, and R. Freund. Enabling S-C-L-Band Systems With Standard C-Band Modulator and Coherent Receiver Using Coherent System Identification and Nonlinear Predistortion. *Journal of Lightwave Technology*, 40(5):1360–1368, 2022.

[38] Y. Emori, S. Kado, and S. Namiki. Simple gain control method for broadband Raman amplifiers gain-flattened by multi-wavelength pumping. In *Proc. Eur. Conf. Opt. Commun.*, volume 2, pages 158–159 vol.2, 2001.

[39] A. Feldmann, O. Gasser, F. Lichtblau, E. Pujol, I. Poese, C. Dietzel, D. Wagner, M. Wichtlhuber, J. Tapiador, N. Vallina-Rodriguez, O. Hohlfeld, and G. Smaragdakis. A Year in Lockdown: How the Waves of COVID-19 Impact Internet Traffic. *Commun. ACM*, 64(7):101–108, 6 2021.

[40] A. Ferrari, A. Napoli, J. K. Fischer, N. Costa, A. D'Amico, J. Pedro, W. Forysiak, E. Pincemin, A. Lord, A. Stavdas, J. P. F.-P. Gimenez, G. Roelkens, N. Calabretta, S. Abrate, B. Sommerkorn-Krombholz, and V. Curri. Assessment on the Achievable Throughput of Multi-Band ITU-T G.652.D Fiber Transmission Systems. *Journal of Lightwave Technology*, 38(16):4279–4291, 2020.

[41] G. C. M. Ferreira, S. P. N. Cani, M. J. Pontes, and M. E. V. Segatto. Optimization of distributed Raman amplifiers using a hybrid genetic algorithm with geometric compensation technique. *IEEE Photonics Journal*, 3(3):390–399, 2011.

[42] C. R. S. Fludger. *Linear Noise Characteristics. In: Islam M.N. (eds) Raman Amplifiers for Telecommunications 1*. Springer, 2004.

[43] K. Fukuchi et al. 10.92-Tb/s (273 x 40-Gb/s) triple-band/ultra-dense WDM optical-repeatered transmission experiment. In *Proc. Opt. Fiber Commun. Conf.*, page PD24. Optical Society of America, 2001.

[44] L. Galdino, A. Edwards, W. Yi, E. Sillekens, Y. Wakayama, T. Gerard, W. S. Pelouch, S. Barnes, T. Tsuritani, R. I. Killey, D. Lavery, and P. Bayvel. Optical Fibre Capacity Optimisation via Continuous Bandwidth Amplification and Geometric Shaping. *IEEE Photonics Technology Letters*, 32(17):1021–1024, 2020.

[45] GitHub repository of GNPy. DOI: 10.5281/zenodo.3458320, https://github.com/Telecominfraproject/oopt-gnpy.

[46] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[47] V. Gordienko et al. Demonstration of an Ultra-Flat Raman-Enhanced Fibre Optical Parametric Amplifier (FOPA) with >110nm Gain-Bandwidth. In *Proc. Eur. Conf. Opt. Commun.*, pages 1–3, 2016.

[48] A. Géron. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition*. O'Reilly Media, Inc., 2019.

[49] M. T. Hagan and M. B. Menhaj. Training feedforward networks with the Marquardt algorithm. *IEEE Trans. Neural Netw.*, 5(6):989–993, 1994.

[50] F. Hamaoka, M. Nakamura, S. Okamoto, K. Minoguchi, T. Sasai, A. Matsushita, E. Yamazaki, and Y. Kisaka. Ultra-Wideband WDM Transmission in S-, C-, and L-Bands Using Signal Power Optimization Scheme. *Journal of Lightwave Technology*, 37(8):1764–1771, 2019.

[51] C. Headley and G. Agrawal. *Raman amplification in fiber optical communication systems*. Academic press, 2005.

[52] G.-B. Huang, D. H. Wang, and Y. Lan. Extreme learning machines: a survey. *International Journal of Machine Learning and Cybernetics*, 2(2):107–122, 6 2011.

[53] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang. Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(2):513–529, 2012.

[54] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew. Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1):489–501, 2006. Neural Networks.

[55] M. Ionescu. Machine Learning for Ultrawide Bandwidth Amplifier Configuration. In *2019 21st International Conference on Transparent Optical Networks (ICTON)*, pages 1–4, 2019.

[56] M. Ionescu, A. Ghazisaeidi, and J. Renaudier. Machine Learning Assisted Hybrid EDFA-Raman Amplifier Design for C+L Bands. In *2020 European Conference on Optical Communications (ECOC)*, pages 1–3, 2020.

[57] M. A. Iqbal, M. A. Z. Al-Khateeb, L. Krzczanowicz, I. D. Phillips, P. Harper, and W. Forysiak. Linear and Nonlinear Noise Characterisation of Dual Stage Broadband Discrete Raman Amplifiers. *Journal of Lightwave Technology*, 37(14):3679–3688, 2019.

[58] M. A. Iqbal, L. Krzczanowicz, I. Phillips, P. Harper, and W. Forysiak. 150nm SCL-Band Transmission through 70km SMF using Ultra-wideband Dual-stage Discrete Raman Amplifier. In *Proc. Opt. Fiber Commun. Conf.*, page W3E.4. Optical Society of America, 2020.

[59] M. A. Iqbal, L. Krzczanowicz, I. Phillips, P. Harper, and W. Forysiak. Evaluation of Performance Penalty from Pump-Signal Overlap in S+C+L band Discrete Raman Amplifiers. In *Proc. Opt. Fiber Commun. Conf.*, page W4B.2. Optical Society of America, 2020.

[60] M. N. Islam. *Overview of Raman Amplification in Telecommunications*, pages 1–34. Springer New York, New York, NY, 2004.

[61] H.-M. Jiang, K. Xie, and Y.-F. Wang. Pump scheme for gain-flattened Raman fiber amplifiers using improved particle swarm optimization and modified shooting algorithm. *Opt. Express*, 18(11):11033–11045, 5 2010.

[62] P. Johannisson and M. Karlsson. Perturbation Analysis of Nonlinear Propagation in a Strongly Dispersive Optical Communication System. *Journal of Lightwave Technology*, 31(8):1273–1282, 2013.

[63] C. K. Kao and G. A. Hockham. Dielectric-fibre surface waveguides for optical frequencies. In *Institution of Electrical Engineers*, volume 113, page 1151–1158, 1966.

[64] F. N. Khan, Q. Fan, C. Lu, and A. P. T. Lau. An Optical Communication's Perspective on Machine Learning and Its Applications. *Journal of Lightwave Technology*, 37(2):493–516, 2019.

[65] D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. https://arxiv.org/abs/1412.6980, 2014.

[66] T. Kobayashi, S. Shimizu, M. Nakamura, T. Umeki, T. Kazama, R. Kasahara, F. Hamaoka, M. Nagatani, H. Yamazaki, H. Nosaka, and Y. Miyamoto. Wide-Band Inline-Amplified WDM Transmission Using PPLN-Based Optical Parametric Amplifier. *Journal of Lightwave Technology*, 39(3):787–794, 2021.

[67] P. M. Krummrich, R. E. Neuhauser, and C. Glingener. Bandwidth limitations of broadband distributed Raman fiber amplifiers for WDM systems. In *Proc. Opt. Fiber Commun. Conf.*, volume 1, page MI3, 2001.

[68] H. Li, Z. Xu, G. Taylor, C. Studer, and T. Goldstein. Visualizing the Loss Landscape of Neural Nets. https://arxiv.org/abs/1712.09913, 2017.

[69] J.-H. Lu, D.-P. Cai, Y.-L. Tsai, C.-C. Chen, C.-E. Lin, and T.-J. Yen. Genetic algorithms optimization of photonic crystal fibers for half diffraction angle reduction of output beam. *Opt. Express*, 22(19):22590–22597, 9 2014.

[70] G. Marcon, A. Galtarossa, L. Palmieri, and M. Santagiustina. Model-Aware Deep Learning Method for Raman Amplification in Few-Mode Fibers. *Journal of Lightwave Technology*, 39(5):1371–1380, 2021.

[71] S. Molesky, Z. Lin, A. Y. Piggott, W. Jin, J. Vucković, and A. W. Rodriguez. Inverse design in nanophotonics. *Nature Photonics*, 12(11):659–670, 11 2018.

[72] A. Mowla and N. Granpayeh. Design of a flat-gain multipumped distributed fiber Raman amplifier by particle swarm optimization. *J. Opt. Soc. Am. A*, 25(12):3059–3066, 12 2008.

[73] A. Napoli, N. Costa, J. K. Fischer, J. ao Pedro, S. Abrate, N. Calabretta, W. Forysiak, E. Pincemin, J. P.-P. Gimenez, C. Matrakidis, G. Roelkens, and V. Curri. Towards multiband optical systems. In *Advanced Photonics 2018*

(BGPP, IPR, NP, NOMA, Sensors, Networks, SPPCom, SOF), page NeTu3E.1. Optical Society of America, 2018.

[74] B. Neto, A. L. J. Teixeira, N. Wada, and P. S. André. Efficient use of hybrid Genetic Algorithms in the gain optimization of distributed Raman amplifiers. *Opt. Express*, 15(26):17520–17528, 11 2007.

[75] OFS. TeraWave® SCUBA 125 Ocean Optical Fiber. https://fiber-optic-catalog.ofsoptics.com/documents/pdf/TeraWave-SCUBA-125-Single-mode-Fiber-fap-170-web.pdf, 03 2018. Accessed: 2021-05-19.

[76] OFS. TeraWave® SCUBA 150 Ocean Optical Fiber. https://stage.fiber-optic-catalog.ofsoptics.com/documents/pdf/TeraWave-Scuba-150-Ocean-Fibers-fiber-168-web.pdf, 03 2018. Accessed: 2021-05-19.

[77] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[78] D. Pérez, I. Gasulla, L. Crudgington, D. J. Thomson, A. Z. Khokhar, K. Li, W. Cao, G. Z. Mashanovich, and J. Capmany. Multipurpose silicon photonics signal processor core. *Nature Communications*, 8(1):636, 2017.

[79] V. E. Perlin and H. G. Winful. Optimal Design of Flat-Gain Wide-Band Fiber Raman Amplifiers. *J. Lightwave Technol.*, 20(2):250, 2 2002.

[80] E. Pincemin. Towards Multi-Band WDM Systems with Hundreds of Terabit/s Transported on one Single-Mode Fiber. In *2021 IEEE Photonics Society Summer Topicals Meeting Series (SUM)*, pages 1–4, 2021.

[81] E. Pincemin, D. Grot, L. Bathany, S. Gosselin, M. Joindot, S. Bordais, Y. Jaouen, and J.-M. Delavaux. Raman gain efficiencies of modern terrestrial transmission fibers in S-, C- and L-band. In *Nonlinear Guided Waves and Their Applications*, page NLTuC2. Optical Society of America, 2002.

[82] P. Poggiolini. A generalized GN-model closed-form formula. https://arxiv.org/abs/1810.06545, 2018.

[83] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri. The GN-Model of Fiber Non-Linear Propagation and its Applications. *Journal of Lightwave Technology*, 32(4):694–721, 2014.

[84] B. J. Puttnam, R. S. Luís, G. Rademacher, M. Mendez-Astudilio, Y. Awaji, and H. Furukawa. S, C and Extended L-Band Transmission with Doped Fiber and Distributed Raman Amplification. In *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3, 2021.

[85] S. Radic. *Forward, Bidirectional, and Higher-Order Raman Amplification*, pages 253–290. Springer New York, New York, NY, 2004.

[86] D. Rafique and L. Velasco. Machine learning for network automation: overview, architecture, and applications [Invited Tutorial]. *Journal of Optical Communications and Networking*, 10(10):D126–D143, 2018.

[87] M. Ranjbar Zefreh, F. Forghieri, S. Piciaccia, and P. Poggiolini. Accurate Closed-Form Real-Time EGN Model Formula Leveraging Machine-Learning Over 8500 Thoroughly Randomized Full C-Band Systems. *Journal of Lightwave Technology*, 38(18):4987–4999, 2020.

[88] M. Ranjbar Zefreh and P. Poggiolini. A Real-Time Closed-Form Model for Nonlinearity Modeling in Ultra-Wide-Band Optical Fiber Links Accounting for Inter-channel Stimulated Raman Scattering and Co-Propagating Raman Amplification. https://arxiv.org/abs/2006.03088, 2020.

[89] J. Renaudier. 100nm Ultra-Wideband Optical Fiber Transmission Systems Using Semiconductor Optical Amplifiers. In *Proc. Eur. Conf. Opt. Commun.*, pages 1–3, 2018.

[90] A. M. Rosa Brusin, V. Curri, D. Zibar, and A. Carena. An ultra-fast method for gain and noise prediction of raman amplifiers. In *45th European Conference on Optical Communication (ECOC 2019)*, pages 1–4, 2019.

[91] A. M. Rosa Brusin, U. C. de Moura, V. Curri, D. Zibar, and A. Carena. Introducing Load Aware Neural Networks for Accurate Predictions of Raman Amplifiers. *Journal of Lightwave Technology*, 38(23):6481–6491, 2020.

[92] A. M. Rosa Brusin, U. C. de Moura, A. D'Amico, V. Curri, D. Zibar, and A. Carena. Load Aware Raman Gain Profile Prediction in Dynamic Multi-Band Optical Networks. In *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3, 2020.

[93] A. M. Rosa Brusin, M. Ranjbar Zefreh, P. Poggiolini, S. Piciaccia, F. Forghieri, and A. Carena. Machine Learning for Power Profiles Prediction in Presence of Inter-channel Stimulated Raman Scattering. In *2021 European Conference on Optical Communication (ECOC)*, pages 1–4, 2021.

[94] K. Rottwitt, J. Bromage, A. J. Stentz, L. Leng, M. E. Lines, and H. Smith. Scaling of the Raman gain coefficient: applications to germanosilicate fibers. *Journal of Lightwave Technology*, 21(7):1652–1662, 2003.

[95] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 10 1986.

[96] R. Ryf and C. Antonelli. *Space-Division Multiplexing*, pages 353–393. Springer International Publishing, Cham, 2020.

[97] M. Secondini and E. Forestieri. Analytical Fiber-Optic Channel Model in the Presence of Cross-Phase Modulation. *IEEE Photonics Technology Letters*, 24(22):2016–2019, 2012.

[98] D. Semrau, R. I. Killey, and P. Bayvel. A Closed-Form Approximation of the Gaussian Noise Model in the Presence of Inter-Channel Stimulated Raman Scattering. *Journal of Lightwave Technology*, 37(9):1924–1936, 2019.

[99] D. Semrau, E. Sillekens, R. I. Killey, and P. Bayvel. A Modulation Format Correction Formula for the Gaussian Noise Model in the Presence of Inter-Channel Stimulated Raman Scattering. *Journal of Lightwave Technology*, 37(19):5122–5131, 2019.

[100] J. M. Senior. *Optical Fiber Communications: Principles and Practice, 3rd Edition*. Pearson, 2009.

[101] P. Serena and A. Bononi. A Time-Domain Extended Gaussian Noise Model. *Journal of Lightwave Technology*, 33(7):1459–1472, 2015.

[102] S. Tariq and J. C. Palais. A computer model of non-dispersion-limited stimulated Raman scattering in optical fiber multiple-channel communications. *Journal of Lightwave Technology*, 11(12):1914–1924, 1993.

[103] The MathWorks Inc. Choose a Multilayer Neural Network Training Function. https://www.mathworks.com/help/deeplearning/ug/choose-a-multilayer-neural-network-training-function.html#bss4gz0-26, 2021. Accessed: 2021-02-12.

[104] The MathWorks Inc. Deep Learning Toolbox™ R2019b. https://www.mathworks.com/products/deep-learning.html, 2021. Accessed: 2021-12-02.

[105] The MathWorks Inc. Fit Data with a Shallow Neural Network. https://it.mathworks.com/help/deeplearning/gs/fit-data-with-a-neural-network.html, 2021. Accessed: 2021-12-02.

[106] The MathWorks Inc. Nguyen-widrow layer initialization function. https://www.mathworks.com/help/deeplearning/ref/initnw.html, 2021. Accessed: 2021-05-19.

[107] L. Velasco. Machine Learning for Network Automation : Overview , Architecture , and Applications [Invited Tutorial]. 10(10):126–143, 2018.

[108] Y. Wang, N. K. Thipparapu, D. J. Richardson, and J. K. Sahu. Broadband Bismuth-Doped Fiber Amplifier with a Record 115-nm Bandwidth in the O and E Bands. In *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3, 2020.

[109] A. Willner, editor. *Optical Fiber Telecommunications VII*. Academic Press, 2019.

[110] P. J. Winzer and D. T. Neilson. From Scaling Disparities to Integrated Parallelism: A Decathlon for a Decade. *Journal of Lightwave Technology*, 35(5):1099–1115, 2017.

[111] P. J. Winzer, D. T. Neilson, and A. R. Chraplyvy. Fiber-optic transmission and networking: the previous 20 and the next 20 years [Invited]. *Opt. Express*, 26(18):24190–24239, 9 2018.

[112] Xiang Zhou, Chao Lu, Ping Shum, and T. H. Cheng. A simplified model and optimal design of a multiwavelength backward-pumped fiber Raman amplifier. *IEEE Phot. Tech. Lett.*, 13(9):945–947, 2001.

[113] X. Ye, A. Arnould, A. Ghazisaeidi, D. Le Gac, and J. Renaudier. Experimental Prediction and Design of Ultra-Wideband Raman Amplifiers using Neural Networks. In *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3, 2020.

[114] T. Zami. Co-Optimizing Allocation of Nyquist Superchannels and Physical Impairments Aware Placement of Regenerators in Elastic WDM Networks. *Journal of Lightwave Technology*, 32(16):2830–2840, 1 2014.

[115] L. Zhuang, C. G. H. Roeloffzen, M. Hoekman, K.-J. Boller, and A. J. Lowery. Programmable photonic signal processor chip for radiofrequency applications. *Optica*, 2(10):854, 2015.

[116] D. Zibar, A. Ferrari, V. Curri, and A. Carena. Machine Learning-Based Raman Amplifier Design. In *2019 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3, 2019.

[117] D. Zibar, A. M. Rosa Brusin, U. C. de Moura, F. Da Ros, V. Curri, and A. Carena. Inverse System Design Using Machine Learning: The Raman Amplifier Case. *Journal of Lightwave Technology*, 38(4):736–753, 2020.