POLITECNICO DI TORINO

POLITECNICO DI TORINO

Master Degree course in Communications Engineering

Master Degree Thesis

Gray-Box Modeling for Predictive Gain and Loss in Optical Line Systems

Supervisors
Prof. Vittorio Curri
Rocco D'Ingillo
Renato Ambrosone

Candidate
Diego Aponte

Academic Year 2024-2025

Acknowledgements

Since I was a child, a deep interest in geography began to awaken in me. I knew all the flags, the capital cities, and could identify the shapes of most countries in the world. My main goal was to become a millionaire so I could travel to all those places. However, as I grew older, I started to see that dream as something utopian, and little by little, I let it go.

While studying Electronic Engineering in Colombia, I learned about the opportunity to pursue this master's degree in Italy. Despite the challenges and obstacles it entailed, I once again felt the same excitement I had as a child. It was an opportunity not only to deepen my knowledge in the field I loved most, but also to rekindle the dreams of little Diego, who longed to cross borders and discover new cultures.

My first thanks go to God, who accompanied me throughout this entire process. Above all, I thank Him for helping me overcome my fears and giving me the courage to step out of my comfort zone, allowing me to reconnect with the dreams of my childhood. Without a doubt, that little Diego would feel very proud of himself for everything he has achieved to reach this point.

I also want to thank all my professors, specially those who introduced me to the fascinating world of fiber optics. When I arrived in Turin, my initial plan was to focus on wireless communications, but thanks to the dedication and passion of my professors in this area, a new interest was sparked within me. I am specially grateful to Professor Vittorio Curri, as well as to Renato Ambrosone and Rocco D'Ingillo, who provided exceptional guidance and great support during the thesis stage. In addition to the technical knowledge they shared, I greatly appreciate their human warmth, patience, and kindness. I will always remember them fondly.

I am also grateful to all the people in Turin who taught me, supported me, and helped me when I needed it most. Specially to those who became friends or consolidated their friendship with me. Everything was brighter and easier with their presence in my life. On many difficult days, they appeared like angels to lift my spirits. I will never forget the trips around Italy and Europe, the parties -specially in December-, the movie nights, the football games, or simply the conversations we shared. You can be sure that those beautiful moments will forever remain in my heart and memory.

I also want to thank my friends around the world who, at some point, thought of me, called me, left me a message, or even came to visit. All of that felt like a warm hug to the heart. I am specially grateful to my girlfriend, who undoubtedly was the person who lived this process closest to me. Thank you for listening to me, for drying my tears, for celebrating my achievements, and above all, for helping me become a better version of

myself. The beautiful love you've given me made me feel at home, despite the distance.

Finally, I want to thank my family, who always supported me throughout this journey. My dad, my mom, and my sister, who were with me in every call, recharged my energy, and gave me advice. I always say that having them is the greatest blessing. They were there for me on many days when I felt incapable, sad, alone, or simply when I missed the warmth of home. A call, a message, a video of my niece, or a picture of our dog was enough to lift my spirits. It felt as if that long-distance connection was somehow magical, or as if the love between us was so strong that it crosses oceans and transcends all borders. You will always be a role model to me. Just as you should feel proud of this achievement, I feel deeply proud to have you in my life and to belong to a family full of love.

Acknowledgements

Desde niño, comenzó a despertarse en mí un profundo interés por la geografía. Me sabía todas las banderas, las capitales, e identificaba las formas de la mayoría de los países del mundo. Mi meta principal era ser millonario para poder recorrer todos esos lugares. Sin embargo, a medida que fui creciendo, empecé a ver este sueño como algo utópico y lo fui dejando de lado poco a poco.

Estudiando Ingeniería Electrónica en Colombia, supe de la oportunidad de realizar esta maestría en Italia. A pesar de los obstáculos y retos que implicaba, volví a sentir el entusiasmo que tenía cuando era pequeño. Era una oportunidad no solo para profundizar en el área que más me había gustado, sino también para devolverle la ilusión al pequeño Diego, que anhelaba traspasar fronteras y conocer nuevas culturas.

Mi primer agradecimiento es para Dios, quien me acompañó durante todo este proceso. Sobre todo, le agradezco por ayudarme a dejar atrás los miedos y darme la valentía de salir de mi zona de confort para reencontrarme con los sueños de mi infancia. Sin duda alguna, ese pequeño Diego se sentiría muy orgulloso de sí mismo al ver todo lo que ha logrado para llegar hasta aquí.

Quiero agradecer también a todos mis profesores, especialmente a quienes me introdujeron al fascinante mundo de la fibra óptica. Cuando llegué a Turín, mi intención era enfocarme en las comunicaciones inalámbricas, pero gracias a la vocación y dedicación de mis profesores en esta área, nació en mí un nuevo interés. Un agradecimiento muy especial para el profesor Vittorio Curri, así como para Renato Ambrosone y Rocco D'Ingillo, quienes me brindaron una guía excepcional y un gran apoyo durante la etapa de la tesis. Además del conocimiento técnico que me aportaron, destaco su calidad humana, paciencia y amabilidad. Siempre los recordaré con mucho cariño.

Agradezco también a todas las personas que, en Turín, me enseñaron, apoyaron y ayudaron cuando más lo necesité. Especialmente a quienes se convirtieron en amigos o con quienes pude consolidar la amistad. Todo fue más bonito y llevadero con su presencia en mi vida. En muchos días difíciles, aparecían como ángeles para alegrar mi jornada. Nunca olvidaré los paseos por Italia o Europa, las fiestas -especialmente en diciembre-, las idas al cine o a ver fútbol, o simplemente nuestras charlas. Pueden estar seguros de que esos lindos momentos quedarán para siempre en mi corazón y en mi memoria.

También agradezco a mis amigos repartidos por el mundo, quienes en algún momento pensaron en mí, me hicieron videollamadas, me dejaron un mensaje o incluso vinieron de visita. Todo eso se sintió como un abrazo al corazón. Agradezco especialmente a mi novia, quien sin duda fue la persona que más de cerca vivió este proceso conmigo. Gracias por escucharme, por secar mis lágrimas, por celebrar mis logros y, sobre todo,

por ayudarme a ser una mejor versión de mí. El amor tan bonito que me has brindado me hizo sentir en casa, a pesar de la distancia.

Finalmente, agradezco a mi familia, que siempre estuvo apoyándome durante todo el camino. A mi papá, mi mamá y mi hermana, quienes con cada llamada me acompañaban, me recargaban de energía y me aconsejaban. Siempre diré que tenerlos es la bendición más grande. Estuvieron para mí en muchos días de tristeza, soledad o cuando simplemente extrañaba el calor de mi hogar. Una llamada, un mensaje, un video de mi sobrina o una foto del perrito bastaban para hacerme sentir mejor. Se sentía como si esa conexión a la distancia fuera mágica, o como si el amor que nos une fuera tan fuerte que recorre los océanos y traspasa todas las fronteras. Siempre serán un ejemplo para mí. Así como ustedes deben sentirse orgullosos de este logro, yo me siento profundamente orgulloso de tenerlos en mi vida y de pertenecer a una familia llena de amor.

Abstract

Optical fiber networks have become the backbone of modern telecommunications due to their high capacity, low latency, and cost-efficiency. Ensuring reliable Quality of Transmission (QoT) in these networks requires accurate estimation of performance metrics prior to lightpath deployment. Traditional approaches often rely on worst-case assumptions to account for uncertainties that might arise, for instance, from amplified spontaneous emission (ASE) noise introduced during optical amplification, leading to conservative design margins that reduce spectral efficiency and increase operational costs. This thesis investigates the use of gray-box modeling techniques, which integrate analytical models with Machine Learning (ML), to enhance the prediction accuracy of gain and loss profiles across an Optical Line System (OLS). The work evaluates multiple model architectures to estimate amplifier gain profiles and fiber span losses, using a finite set of measured data to train and validate the models. Comparative analysis against benchmark models demonstrates that gray-box approaches can significantly reduce prediction error, enabling more accurate QoT estimation and more efficient resource allocation.

Contents

List of Tables

List of Figures

Acronyms

Adam Adaptive Moment Estimation.

AI Artificial Intelligence.

API Application Programming Interface.

ASE Amplified Spontaneous Emission.

B-GB Both features Gray-box model.

B-GB-G Both features Gray-box Gain model.

BER Bit Error Rate.

BST Booster Amplifier.

D-GB Default Gray-box model.

D-GB-G Default Gray-box Gain model.

DNN Deep Neural Networks.

DSP Digital Signal Processing.

EDFA Erbium Doped Fiber Amplifiers.

GNPy Gaussian Noise in Python.

GSNR Generalized Signal-to-Noise Ratio.

ILA In-Line Amplifier.

ISI Inter-Symbol Interference.

M-DNN Multiple Deep Neural Network model.

MAE Mean Absolute Error.

MaxAE Maximum Absolute Error.

ML Machine Learning.

MSE Mean Squared Error.

NLI Non-Linear Impairments.

OLS Optical Line System.

OMS Optical Multiplex Sections.

OSA Optical Spectrum Analyzer.

OSNR Optical Signal-to-Noise Ratio.

P-GB Power as feature Gray-box model.

P-GB-G Power as feature Gray-box Gain model.

PRE Pre Amplifier.

QoT Quality of Transmission.

ROADM Reconfigurable Optical Add-Drop Multiplexer.

S-GB State as feature Gray-box model.

S-GB-G State as feature Gray-box Gain model.

SDN Software Defined Networking.

SNR Signal-to-Noise Ratio.

SOA Semiconductor Optical Amplifiers.

SPM Self-Phase Modulation.

SRS Stimulated Raman Scattering.

VOA Variable Optical Attenuator.

WB White-box model.

WDM Wavelength Division Multiplexing.

WSS Wavelength Selective Switches.

XPM Cross-Phase Modulation.

XPolM Cross-Polarization Modulation.

Chapter 1

Introduction

Communication is one of the main aspects in the day-to-day functioning of society, and humanity has understood this since the appearance of early civilizations. Starting with smoke signals, drums, flags and even homing pigeons, communities started to transmit their messages [1] [2] [3]. As obstacles appeared, solutions were created; communication networks took shape when point-to-point communication became unfeasible and different transmission media started to be used with the advent of new technologies in order to make telecommunication faster, reliable and global. Optical fiber has been since a couple of years ago the most common transmission medium due to the huge amount of data it can carry in small time, in addition to its reach, ease of installation and costs. In an optical fiber, information is transmitted in a lightwave; each lightwave is associated with a wavelength or frequency. The route through which each lightwave roams in an optical network is called a lightpath.

Today, not only do large industries or organizations that are in charge of relevant sectors, such as transportation and banking, have a strong dependency on telecommunications, but also the average human being in his daily life is starting to create this dependency [4] [5] [6]. Thus, interruptions in communication systems can cause difficulties in different industries around the world that can even lead to economic losses for companies; therefore, Quality of Transmission (QoT) and lightpath routing design in an optical communication network are an important aspect. Estimating the QoT of the different possible lightpaths before their deployment can make a great difference in favor of an effective and flexible network design [7]. Unfortunately, estimating QoT metrics is not straightforward, as there are many sources of uncertainty. Considering assumptions or simplifications for these uncertainties will result in inaccurate metrics. As a consequence, assigning more resources to lightpaths in order to compensate for the deteriorations that the uncertainties might cause is a common practice. These extra resources are called design margins. The use of margins is detrimental to the efficiency and cost of optical networks [8]. Thus, one of the main objectives in the state-of-the-art optical fiber transmission is reducing these margins by accurately predicting the uncertainties that might appear.

An example of the uncertainties mentioned above is Amplified Spontaneous Emission (ASE) noise, a type of noise generated during the amplification process in optical networks

due to the phenomenon of spontaneous emission present in fibers. ASE noise contributions depend on the spectral load that interacts with the amplifier. Given the potentially large number of possible spectral load configurations, developing a comprehensive empirical model is impractical, as it would require an overwhelming number of measurements for estimation [9]. As a consequence, Machine Learning (ML) techniques emerge as solutions that fit properly in this context. By taking a finite set of measurements, ML models can be trained to predict the actual gain profile and, in that way, reduce the design margins [7].

A considerable number of ML models have been already created to reach this goal. Some focus just on the prediction of the components that form the gain profile (signal power and ASE noise power) using Deep Neural Networks (DNN) [10] [11]. Other investigation paths try to realize a more general analysis by complementing the DNN prediction with tools that can simulate other phenomena [7]. Moreover, other articles use algorithms and different neural network structures that can enhance the estimation [12].

The goal of this thesis work is to test the performance of gray-box models in this context, in order to accurately characterize an entire Optical Line System (OLS) of an optical network in terms of QoT. Gray-box models combine analytical models with ML and have been applied in past investigations [13] [14]. However, in this thesis, the models are not only applied to predict the gain profile, but also the loss profile.

This thesis work is structured as follows:

- Chapter 2 : **Theoretical Background** presents the concepts that are relevant to the context of the problem and to the approach used to address it.
- Chapter 3 : **Project Overview** contains a contextualization on the problem studied, the state of the art of the problem, a description of the setup and dataset used for this work, and the performance of an already existing analytical model with the described setup and dataset.
- Chapter 4 : **EDFA Analysis: Gain Profile Prediction** describes the methodology used to predict the gain profile. Different models were proposed and tested; these models are based on a benchmark model taken from the state of the art. Comparisons are made between the benchmark models and the best one among the proposed models.
- Chapter 5: **Fiber Span Analysis: Loss Profile Prediction** describes the methodology used to predict the loss profile. This chapter emphasizes on how this prediction should be done considering the existing tools and describes the models used to perform this prediction. Overall results are presented for the modeling of the whole OLS system.
- Chapter 6: **Conclusion** summarizes the work done in the thesis highlighting the main results obtained.

Chapter 2

Theoretical Background

For the purpose of a better understanding of the content of this master's thesis, a contextualization on elements and concepts in optical networks as Software Defined Networking (SDN), optical networks, fibers and signals, amplifiers, design margins, Quality of Transmission (QoT) metrics, and machine learning are presented.

2.1 Software Defined Networking

Traditionally, network elements have been closed platforms that integrated all of their features in a single device, regardless of the type of functionality. Consequently, the devices were able to perform tasks from the data, control, and management plane. For example, devices could forward packages and, at the same time, compute routes, discover topologies, and perform measurements and monitoring. This led to the design of complex devices that were proprietary depending on the vendor and could only work well with other devices because of several protocols.

Software Defined Networking (SDN) appears in the search for a more flexible and open network and in a way to simplify devices and enhance innovations in networking. The main principle of SDN is to separate the data plane from the control plane. Thus, SDN suggests having a centralized architecture in which there is a controller with knowledge of the entire network that sends commands to simple general-purpose forwarding devices. The interaction between the controller and the network elements is carried out with a common Application Programming Interface (API).

SDN can lead to open disaggregated optical networks. In these networks, an SDN controller can access each network element with common protocols and APIs. Open disaggregated networks can allow multivendor networks, a more effective network management, and the inclusion of Artificial Intelligence (AI) tools. An image illustrating an open disaggregated optical network is found in 2.1 [15].

2.2 Optical Networks

Communication networks are the infrastructure needed to perform long distance transmissions between devices. They are made up of a collection of switching nodes and

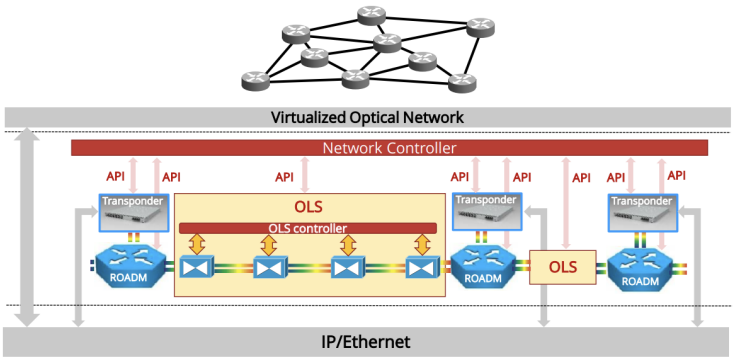

Figure 2.1. Open disaggregated optical network

connections. The optical network nodes are usually arranged in a meshed, ring, or tree topology and their connections are called Optical Line System (OLS) or Optical Multiplex Sections (OMS).

Optical networks can be divided into translucent and transparent networks. Translucent optical networks have electrical switches; therefore, all networking operations are performed electrically. Hence, in each node an electrical-optical conversion should be performed. In transparent optical networks, switches are optical and signals can be selectively directed to any output. Furthermore, networking operations are performed optically. Transparent optical networks require a coordinated operation between the physical network layer and the higher network layers. Transparent network layers have the great advantage of low power consumption and high integration, translucent networks require more hardware and might have power consumption issues [16].

The typical elements of an optical network are described in Sections 2.2.1 and 2.2.2.

2.2.1 Optical Network Nodes

In optical network nodes, we can find devices with different functions. The basic functionalities found in optical network nodes are: serving as interface between the physical layer and upper layers and as a switching functionality. Some devices perform just one of these functions, others do both, some of them perform additional functions.

The core elements of optical network nodes are the following:

- Optical transceivers
- Transponders
- Reconfigurable Optical Add-Drop Multiplexer (ROADM)

Optical transceivers are devices whose main function is to transmit and receive signals within the same module. Optical transceivers should be placed for every signal in each node in a translucent optical network.

Transponders can perform optical to electrical and electrical to optical conversions; they present wavelength converters and multiplexers. Additionally, they can receive, regenerate, and reshape signals. Transceivers can be plugged into transponders if necessary.

ROADMs are formed by an optical switch section and a local add-drop section. The optical switch section is made up of an array of Wavelength Selective Switches (WSS), another device capable of distributing signals in each wavelength through its ports. Thus, the signals are routed in different directions. One of these directions can enter or can come from the local add-drop section in the ROADM. In this section a local transceiver can either receive a signal "dropped" from the traffic in the WSSs or can transmit a signal "added" to the traffic in the WSSs. A Figure illustrating the ROADM structure is found in 2.2 [17].

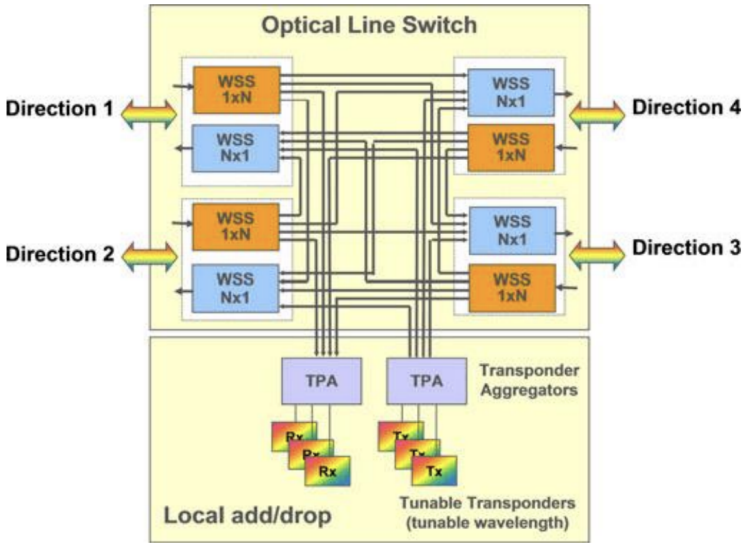

Figure 2.2. ROADM structure

2.2.2 Optical Line Systems

An Optical Line System (OLS) is a point-to-point bidirectional link that connects two nodes. The main elements of the OLS and the impairments that they introduce are better analyzed in the following sections. It is formed by:

- Spans of fiber pairs, one for each direction
- Booster Amplifier (BST), an Erbium Doped Fiber Amplifiers (EDFA) amplifier linked to the the output of the transmitting node, usually installed inside this node. It is the first amplifier through which the signal goes in the OLS.
- Pre Amplifier (PRE), an EDFA amplifier connected to the input of the receiving

node, usually installed inside this node. It is the last amplifier through which the signal goes in the OLS.

- In-Line Amplifier (ILA), set of EDFA amplifiers that are between the BST and PRE amplifiers. They come in pairs, one for each direction.

The impact of the OLS on the signal will be only in terms of adding latency and noise impairments. In other words, QoT degradation will take place mainly in this section of the network.

2.3 Optical Fibers

An optical fiber is a cylinder strand as thin as a human hair, which is generally made of very pure glass and can serve as a transmission medium for lightwaves. Its structure is composed of an inner layer called core and an outer layer called cladding. The core has a slightly higher refractive index compared to that of the cladding; thus, light is completely reflected inside the fiber and can be guided through the material [1].

The choice of glass as the fiber material is due to the low attenuation values it has at certain frequencies. Today, commercial fibers have attenuation values ranging from 0.15 dB/km to 0.22 dB/km at 1550 nm [18], far from the attenuation in copper coaxial cables (10 dB/km - 100 dB/km). A graph showing the attenuation of the fiber according to the wavelength is shown in Figure 2.3 [19]. The graph also shows the operating bands in optical fiber communication and highlights peaks in the attenuation as a result of impurities in the glass such as the water peaks.

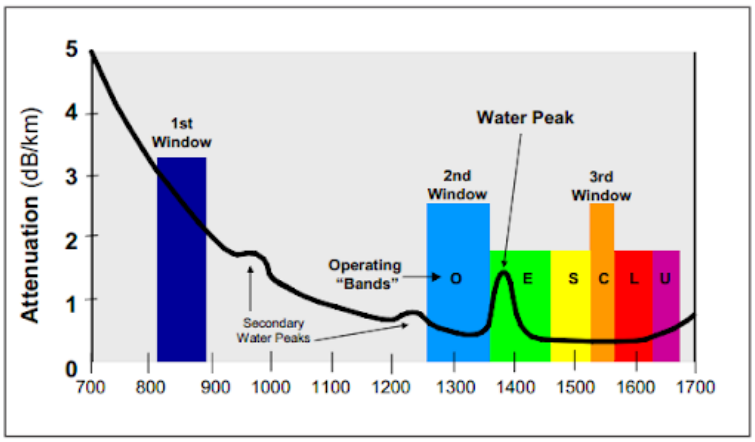

Figure 2.3. Fiber attenuation in the wavelength domain

Signal propagation loss is not the only advantage of optical fiber communication. The available bandwidth is perhaps the most interesting advantage of fibers as it is in the order of THz, different from other mediums that are in the GHz or MHz range. This large range of available bandwidth allows for a higher throughput, ergo better performances. Other

advantages are related to the fiber's weight, size, and resistance to external interferences, chemicals, and atmospheric conditions.

Apart from attenuation, fibers have more linear propagation effects such as birefringence or polarization variations due to imperfections in the fiber's shape or mechanical stresses. Chromatic dispersion is another linear propagation effect originating from the material itself and the phase evolution across the fiber. Chromatic dispersion causes the group delay to vary with frequency and can possibly lead to Inter-Symbol Interference (ISI) [20]. These linear effects can be effectively compensated with Digital Signal Processing (DSP) in the receiver [21].

Non-linear propagation effects also occur in fibers. For example, Stimulated Raman Scattering (SRS) is an effect caused by the interaction of photons with the molecular structure of fibers [22]. SRS takes place whenever energy from pump photons or a flow of propagating photons is transferred to other frequencies in the optical field, thence generating a sort of gain curve that will amplify the signals present in those frequencies. A plot displaying the generated Raman gain C_R by a pump in frequency f_p is shown in Figure 2.4 [23].

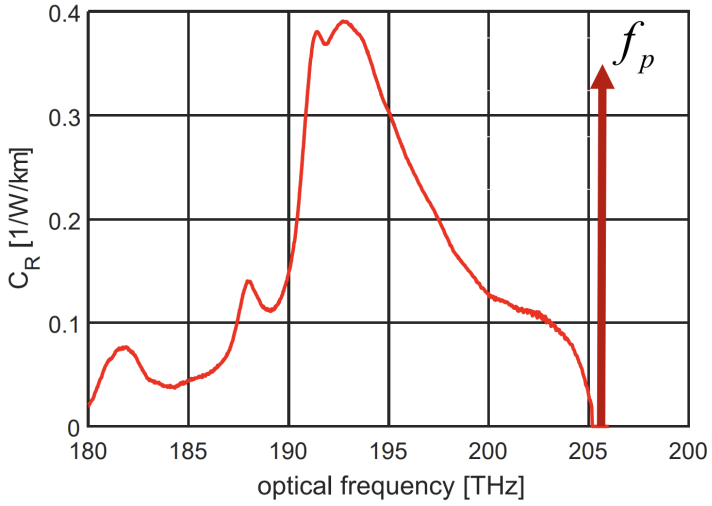

Figure 2.4. SRS effect in the spectrum

SRS can also occur among channels, as each channel can act as a pump and the generated Raman gain profiles will be summed among each other. As a result, low frequency channels experience a power gain at the expense of high frequency channels generating a tilt in the curve of the power in the frequency domain. This tilt increases with the number of spans, as seen in Figure 2.5 [23].

When considering single-channel systems, the base for modeling fiber propagation and its effects is the non-linear Schrödinger equation. It can be noticed from the equation, in Figure 2.6 [24], that apart from attenuation and dispersion effects, a non-linear effect called the Kerr effect [25] will also be present. In the absence of dispersion, the Kerr

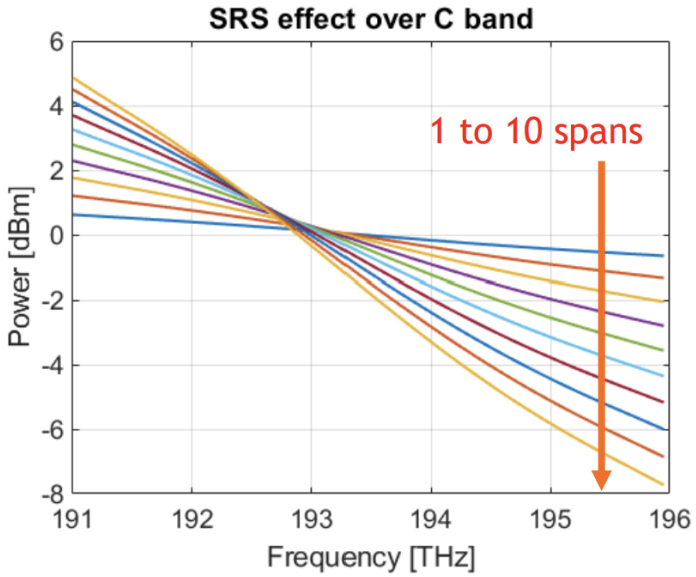

Figure 2.5. SRS effect over C band for different number of spans

effect acts by introducing a power-dependent phase modulation of the field. It can also be seen as the causative of variation in the glass refractive index due to power.

$$\frac{\partial E(z,t)}{\partial z} = \underbrace{j\frac{\beta_2}{2}\frac{\partial^2}{\partial t^2}E(z,t) + \frac{\beta_3}{6}\frac{\partial^3}{\partial t^3}E(z,t)}_{\text{DISPERSION}} \underbrace{-\alpha E(z,t)}_{\text{LOSS}} \underbrace{-j\gamma|E(z,t)|^2 E(z,t)}_{\text{KERR EFFECT}}$$

Figure 2.6. Non Linear Schrödinger Equation

If we consider a system with multiple channels, it is perceived that it is impossible to separate independent equations for each channel from the Schrödinger equation; as all equations will contain terms that connect to other channels. Therefore, non-linear contributions can come from Self-Phase Modulation (SPM), whenever phase is modulated by the power variation of the channel itself, and Cross-Phase Modulation (XPM), whenever phase is modulated by the power variation of all the other channels. This phase modulation behavior is present whenever dispersion is not considered. If we take into account dispersion, the disturbance will emulate additive Gaussian noise.

The standard model for estimating the power spectral density of Non-Linear Impairments (NLI) is the GN model [26]. GN stands for Gaussian noise. An approximated general version of the model for the n_s-th span and n_{ch}-th channel is described in 2.1.

$$G^{(1)}_{NLI,n_s,n_{ch}} = G^{(1)}_{SPM,n_s,n_{ch}} + G^{(1)}_{XPM,n_s,n_{ch}} \tag{2.1}$$

In which the number in parentheses means that the equation is computed for that number of spans. Anyway, linear accumulation with the spans is accepted. The expression can be expanded with equations 2.2 and 2.3:

$$G^{(1)}_{SPM,n_s,n_{ch}} = \frac{4}{27}\frac{\gamma^2_{n_s} P^3_{n_s,n_{ch}}}{\pi|\beta_{2,n_s}|\alpha_{n_s} R^3_{s,n_{ch}}} asinh(\frac{1}{4\alpha_{n_s}}\pi^2|\beta_{2,n_s}|R^3_{s,n_{ch}}) A^3_{pre,n_s} A_{F,n_s} A_{post,n_s} G_{n_s,n_{ch}} \tag{2.2}$$

$$\begin{aligned} G^{(1)}_{XPM,n_s,n_{ch}} = {} & \frac{4}{27}\frac{\gamma^2_{n_s} P_{n_s,n_{ch}}}{\pi|\beta_{2,n_s}|\alpha_{n_s} R_{s,n_{ch}}} \\ & \cdot \sum_{k_{ch}=1,k_{ch}\neq n_{ch}} \frac{P^2_{n_s,k_{ch}}}{R^2_{n_s,k_{ch}}} [asinh(\pi^2[2\alpha_{n_s}]^{-1}|\beta_{2,n_s}|[f_{k_{ch}} - f_{n_{ch}} + R_{s,k_{ch}}/2]R_{s,n_{ch}}) \\ & - asinh(\pi^2[2\alpha_{n_s}]^{-1}|\beta_{2,n_s}|[f_{k_{ch}} - f_{n_{ch}} - R_{s,k_{ch}}/2]R_{s,n_{ch}})] \cdot A^3_{pre,n_s} A_{F,n_s} A_{post,n_s} G_{n_s,n_{ch}} \end{aligned} \tag{2.3}$$

In which:

- γ is the non-linearity coefficient
- P is the average transmitted power
- β_2 is the dispersion coefficient
- α is the fiber loss coefficient
- R_s is the symbol rate
- f_{ch} is the central frequency of channel ch
- A_{pre} are the lumped losses before the fiber
- A_F are the losses from the fiber itself
- A_{post} are the lumped losses after the fiber
- G is the gain of the amplifier

Looking at the model, it can be concluded that:

1. NLI scales with the cube of the power
2. NLI is inversely proportional to dispersion (the factor present in the $asinh$ argument has little impact)
3. NLI is proportional to the square of the fiber's non-linear coefficient

Another effect that is also part of NLI and is present only when the dispersion is low is called four-wave mixing [27]. In four-wave mixing, a narrow-band signal is generated at a frequency f_i due to the presence of stronger narrow-band signals at frequencies f_m, f_n and f_k. Where $f_i = f_m - f_n + f_k$. Cross-Polarization Modulation (XPolM) is another NLI effect that happens if we consider a dual-polarization; SPM and XPM occur on one polarization because of the power in the other polarization. Although the approximate GN model shown in equation 2.1 only takes into account SPM and XPM for NLI, the complete actual version of the GN model also considers four-wave mixing and XPolM.

2.4 Amplifiers

Optical amplifiers are devices capable of intensifying an optical signal without performing a conversion to electronics. They are required to account for the losses generated by fiber propagation. The most common types of optical amplifiers are Erbium Doped Fiber Amplifiers (EDFA), Semiconductor Optical Amplifiers (SOA), and Raman amplifiers. In this thesis work, EDFAs are the type of amplifiers studied; then, this section will refer entirely to these amplifiers.

Amplification is possible as a consequence of an effect called stimulated emission. In stimulated emission, a new photon with an energy equal to that of another incoming photon originates from an atomic system. Not only will the new photon have the same energy as the incoming one but they will also share frequency and phase. This will only happen if some atoms are already present in an upper energy level; for that reason, the absorption process must take place. In absorption, photons are pumped in order to excite the atoms into higher energy levels. Stimulated emission is not the only way a photon can be generated inside the fiber, as there is another phenomenon called spontaneous emission. In spontaneous emission, there is no necessity to have an incoming photon, energy comes from an atom spontaneously descending to a lower energy level. Photons from spontaneous emission follow random directions, so they act as noise; furthermore, they can be amplified by stimulated emission, forming in this way what is known as Amplified Spontaneous Emission (ASE) noise. An image illustrating the phenomena of absorption, spontaneous and stimulated emission is present in Figure 2.7 [28].

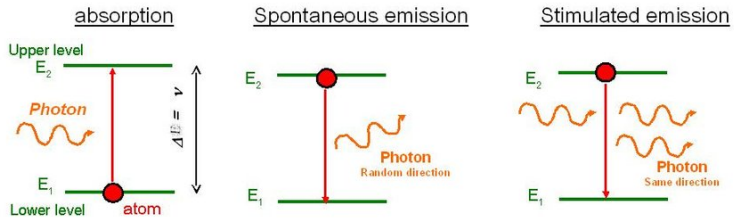

Figure 2.7. Phenomena of absorption, spontaneous emission and stimulated emission

The most important components of an EDFA are a fiber span that contains erbium

ions in its core and a pump laser. Thanks to erbium ion doping, the fiber will present energy levels suitable for amplification, while the pump laser is used to excite the atoms into higher energy levels. EDFAs can also contain Variable Optical Attenuator (VOA). As their name indicates, VOAs attenuate the signals; therefore, it is important to acknowledge this setting when configuring the EDFAs.

A simple model for amplification is present in equation 2.4

$$P_{out}(t) = G \cdot P_{in}(t) + P_{ASE}(t) \tag{2.4}$$

In which:

- P_{out} is the power coming out from the amplifier
- P_{in} is the power coming in to the amplifier
- P_{ASE} is the power coming from ASE noise
- G is the gain of the amplifier

ASE noise is white and Gaussian and can be described for a bandwidth B as in equation 2.5

$$P_{ASE} = h \cdot f \cdot B \cdot (G - 1) \cdot F \tag{2.5}$$

Where:

- h is Planck's constant
- f is the optical frequency (usually the central frequency of the band)
- G is the amplifier power gain
- F is the noise figure

The noise figure of an amplifier specifies how much the device degrades the Signal-to-Noise Ratio (SNR) [29]. Its value can also be described as twice the spontaneous emission factor n_{sp}, a noise factor that is always greater than 1. Other parameters for the EDFAs are gain and tilt, that differently from the noise figure, they are part of the input settings. The target gain specifies how much will be applied to the incoming power in the amplifier, while the tilt specifies the inclination of the gain profile in a specific band, generally in dB/THz. An example of the usage of tilt is to compensate for the SRS effect already shown in Figure 2.5. Despite being indicated as constant numbers in the datasheets or setting configurations, neither gain nor noise figure are constant in frequency, even if the tilt is 0 dB/Hz. This is shown in Figures 2.8, 2.9, 2.10, 2.11 [30].

In order to characterize the gain profile, a model is suggested in [31]. The model is present in equation 2.6.

$$G(f) = G + \frac{T}{B} \cdot (f - f_0) + g(f) \tag{2.6}$$

In which:

- G is the gain target or set gain

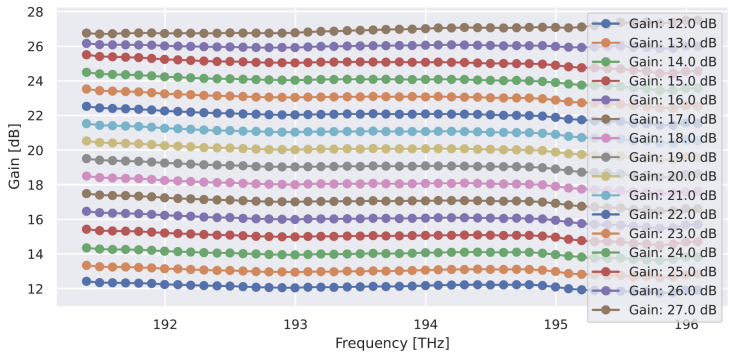

Figure 2.8. Gain profile of different gain settings and tilt set to 0 dB/Hz

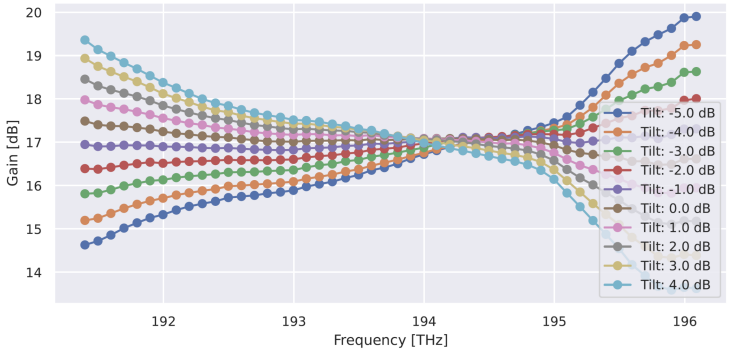

Figure 2.9. Gain profile of gain set to 17 dB and varying tilt

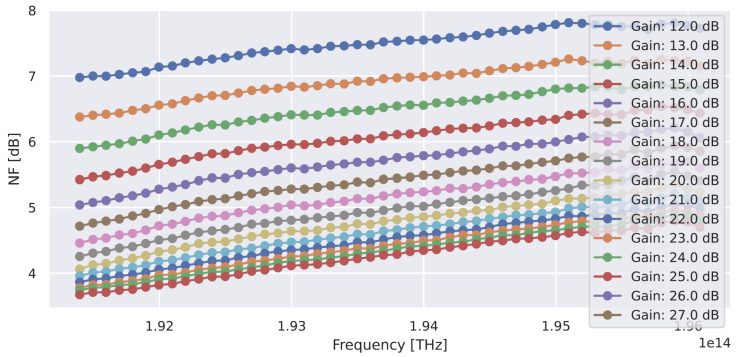

Figure 2.10. Noise figure spectrum with varying gain and tilt set to 0 dB/Hz

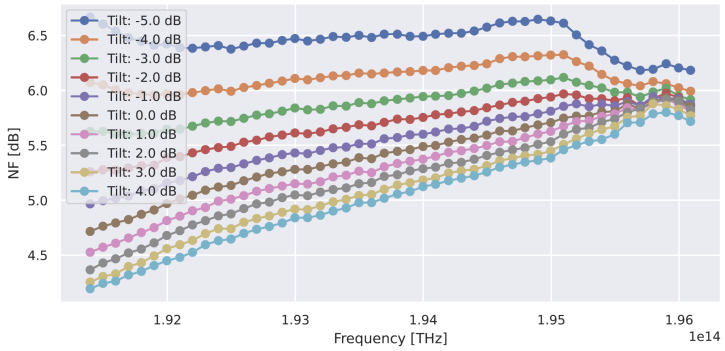

Figure 2.11. Noise figure spectrum with varying tilt and gain set to 17 dB/Hz

- T is the tilt target or set tilt
- B is the bandwidth
- f_0 is the band center
- g is the gain ripple

An illustration showing the model parameters is shown in Figure 2.12 [31].

2.4.1 Spectral load's effect on EDFAs

The distribution of a spectral load can determine how different devices will work. For example, the gain profile of an EDFA will have a different behavior when there are only four channels on and the rest off, compared to a full spectral load (all channels on, zero channels off), even if the setting of the EDFA is the same for both cases [10]. Unfortunately, in the already mentioned EDFA case, designing an analytical model for the amplifier that takes into account these variations is very difficult to handle, as the idea is to define these models at a system level; hence, there is access to parameters such as gain, tilt, total power in/out. It is not only difficult to have an accurate model due to the huge number of spectral load possibilities that can exist, but also because the spectral load is not the only factor that can affect the gain profile at the output of an EDFA. Other factors that might affect slightly can be the temperature or even the vendor and model of the device.

2.5 Optical Signals

As in every communication system, information is transmitted through electromagnetic waves associated with a carrier frequency. In optical fibers, as these waves come in the form of light, they are called lightwaves. It is also important to recall that in the optical

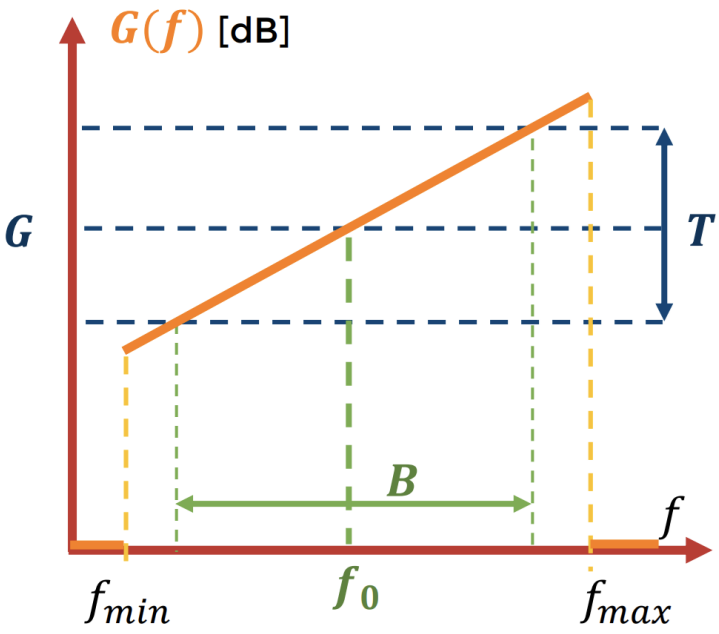

Figure 2.12. Gain profile modeling representation

domain the usage of wavelengths λ instead of frequencies f is more common. In which the relationship between both is:

$$f = \frac{c}{\lambda} \tag{2.7}$$

Where c is the constant for the speed of light.

A lightpath is defined as the route that a lightwave takes from source to destination in a transparent optical network. Lightpaths should respect a pair of conditions in order to be properly deployed. These conditions are known as wavelength continuity. Wavelength continuity states that the same wavelength should be available for the lightwave in every OLS crossed from source to destination. At the same time, wavelength continuity implies that the lightwave can be properly directed through all the crossed switches.

Wavelength continuity is important as it plays a main role in the routing and wavelength assignment. The routing and wavelength assignment must not only take into account the topology of the network and its current status regarding availability of paths and wavelengths but also needs knowledge on the QoT for each feasible path, so in that way paths are selected with a criteria associated with effectiveness and not in an arbitrary way.

Reserving an entire fiber for a single lightwave is very inefficient, and this is why the most common multiplexing technique in optical transmission is Wavelength Division Multiplexing (WDM). In WDM, a lightwave has a dedicated channel for itself. A channel in a WDM comb is associated with a wavelength or frequency. The set of channels in WDM is called WDM comb or it can also be called the spectral load. The standard for

channel spacing in a WDM comb established by the ITU-T is of multiples of 12.5 GHz.

2.5.1 Digital Twin

A digital twin is a precise high-fidelity model of a system and the impact of its external environment. Its main purpose is to emulate a real system or its "twin". In the SDN context, digital twins can serve as a great complement to SDN controllers. For example, it can help in the planning process where choices regarding the equipment used and the traffic management are relevant, or it can help in the control of the network by setting the devices in their optimal status. Route and wavelength assignment with a digital twin can be done more efficiently. Digital twins can also be exploited along with AI. AI can be tested on digital twins in order to set optimal points in the devices; in turn, digital twins can provide useful information in order to improve AI training.

A tool that can be used as a digital twin for the physical layer of the network is the Gaussian Noise in Python (GNPy) project. GNPy is originally meant to be a vendor-neutral design and planning tool.

Physical models of the network elements are needed to use a digital twin of an optical network. This is the reason why different models are proposed and studied for these elements.

2.6 Design Margins and Quality of Transmission Metrics

2.6.1 Design Margins

Planning and designing a network is a complicated task. Just for one of the considerations that must be taken for this task as deploying lightpaths with an acceptable QoT, it is necessary to consider all the impairments that the signals will suffer through the network, mainly the ones present in the OLS like ASE noise and NLI. In addition, network designers should consider how these impairments evolve when new signals are added or dropped in the OLS, making the analysis of the spectral load's effect on devices more relevant. Other important factors that should be considered while planning and designing are aging of equipment and maintenance operations such as fiber repairs.

In order to account for all of these impairments, worst-case scenarios are taken into account to guarantee lightpaths with admissible QoT. Design margins are established to overprovision lightpaths with the help of simplified models or estimations. Unfortunately, as models and estimations might deal with uncertainties, the computation of these margins is not accurate enough, and usually networks operate far from these conservative design margins. As a result, networks are designed with extra equipment or expensive equipment with unnecessary complexity, which means additional costs. The goal of modern networks is then seeking to reduce these design margins to "actual (just enough)" margins [8].

2.6.2 Quality of Transmission Metrics

Optical Signal-to-Noise Ratio

Optical Signal-to-Noise Ratio (OSNR) is a QoT metric defined as:

$$OSNR = \frac{P_S}{P_{ASE}} \tag{2.8}$$

In which:

- P_S is the average power of the signal
- P_{ASE} is the power of the ASE noise in an equivalent reference bandwidth

Usually, the reference bandwidth used in OSNR is 12.5 GHz due to legacy reasons.

OSNR is generally used when the QoT analysis does not involve non-linear effects or if non-linear effects are analyzed separately.

Generalized Signal-to-Noise Ratio

A most complete QoT metric for lightpaths in optical networks is Generalized Signal-to-Noise Ratio (GSNR). GSNR takes into account non-linear impairments.

As mentioned in Chapter 2.3, dispersion is a propagation effect that accumulates through the fiber spans. Dispersion does not affect the spectrum, but it strongly affects the time-domain signal. When a transmission is performed without compensating the dispersion, it can be noticed that the signal will resemble Gaussian noise. This means that non-linear interferences generated by the crosstalk among channels and by the channel itself will also act as zero-mean additive Gaussian noise. As NLI is statistically independent from ASE and on top of that it acts as Gaussian noise, similarly to ASE, it can be summed in the denominator of the OSNR. Therefore, the definition of GSNR is:

$$GSNR = \frac{P_S}{P_{ASE} + P_{NLI}} \tag{2.9}$$

Where P_{NLI} corresponds to the power of NLI "noise" in the overall signal.

Other metric called Bit Error Rate (BER) can use either OSNR or GSNR to be computed apart from other data concerning the modulation of the signal.

2.7 Machine Learning

Machine Learning (ML) is a branch of AI in which algorithms can effectively recognize patterns or "learn" from a collection of data with the help of statistics. In that way, these algorithms can be generalized and applied to data that have not been studied previously by themselves. ML algorithms must have a performance metric that should improve with their experience performing a task. Training a ML model involves using a set of data and iteratively finding the proper parameters for the model to predict correctly. A ML model represents a mathematical relationship between the data and the expected results. The

model captures important and quantifiable features from the input and processes them to make predictions.

Data is crucial for deep learning; therefore, it should be handled thoroughly. An strict separation between the data that will be used for training, validation (used to check performance and to take design choices in the model), and testing (used for the final test of the model) should be done. In addition, normalization and adjustments in the distribution of the data must be made if necessary.

ML models can be tuned with design choices called hyperparameters. Hyperparameters are usually tuned in the validation set and have a large impact on network performance. The most traditional way to tune hyperparameters is with a method called grid search. In a grid search, different combinations for hyperparameters are tested in the validation set to check which set of hyperparameters work better. This technique usually takes a long time, as it has to train the model for each possible combination. Clearly, the time for the grid search will increase with the number of hyperparameters that will be tested.

When referring to models for a system, ML models are referred as black-box models because the machine takes a decision based on the data and the hyperparameters set as input, but the developer knows little about the actual operations done inside the network or the actual behavior of the model. Opposite to an analytical model, in which the construction of the model allows the person using the model to understand in detail how a system behaves and how it can be handled. Analytical models are referred as white-box models. Gray-box models combine both the analytical and the ML approach, taking advantage of the accuracy that can be reached with black-box models and the transparency and explainability of white-box models.

The ML strategy used in this thesis is the artificial neural network. A neural network is a collection of neurons arranged in parallel and in layers. A neuron is a representation of a function that depends on a weighted sum of several inputs summed by a bias. Generally, the function represented by the neuron, called activation function, is a non-linear function. The mathematical representation of the neuron is present in equation 2.10 [32].

$$y = f(\sum_{i=1}^{N_{in}} w_i x_i + b) \tag{2.10}$$

Where:

- y is the output of the neuron
- f is the activation function
- N_{in} is the number of inputs
- w_i is the weight that corresponds to input x_i
- b is the bias

In a neural network, each neuron computes a different weighted sum and each layer receives as input the output of the previous layer. The first layer of neurons is called

input layer and the last output layer. All the layers between the input layer and the output layer are called hidden layers. The deeper the network, the more complex the function it represents. This is depicted in Figure 2.13 [33].

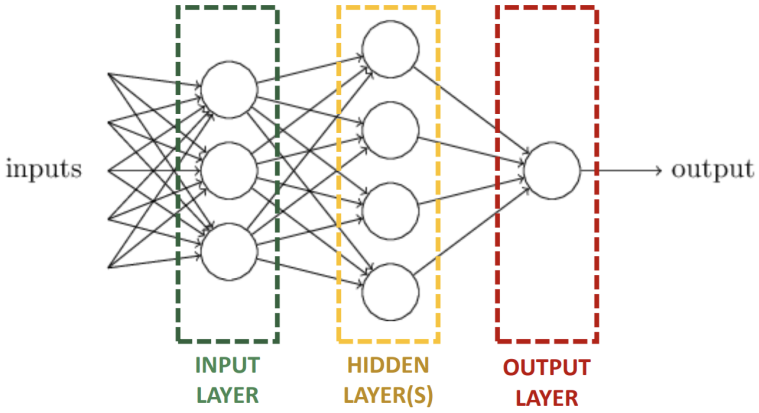

Figure 2.13. Typical neural network structure

As mentioned in the beginning of this chapter, performance should improve iteratively when training a ML model. Usually, this is controlled with a cost function, also called loss function. This function outputs a scalar value that measures the difference between the neural network's output and the target output. Thus, the main goal while training the neural network is finding the weights and biases that minimize the cost function.

Different algorithms can be used to minimize the functions. The gradient descent algorithm is an example of algorithm used to minimize these multivariable functions, it basically updates the parameters (weights and biases) by subtracting the current parameters with the gradient of the cost function with respect to the current parameters and multiplied by a factor called the learning rate. The learning rate defines how fast the parameters are updated; if it is too fast, possibly the algorithm will not converge, whereas if it is too slow, the computational complexity will be high. Gradients in neural networks are computed with another algorithm called backpropagation [34]. Backpropagation takes advantage of the chain rule of derivatives and computes the derivative of the cost with respect to a weight, from the derivative of the cost with respect to the output of the neural network. Backpropagation is already implemented in specialized Python libraries such as TensorFlow or PyTorch (the one used in this thesis).

In order to reduce computational complexity while performing the gradient descent algorithm, batches of data are used. The use of batches allows to avoid processing all the entire training set to compute the gradient, a sample of the set can be used, and the gradient will behave as a "noisy" version of the true complete set gradient despite being faster in the training. The number of iterations required to process the entire training set at once is called an epoch. Training is usually done in tens or a few hundreds of epochs. An option to enhance training is the use of a hyperparameter called patience. Patience defines the number of epochs to wait before stopping and restarting the training

from that point if there is no progress. Other hyperparameters in neural networks are the learning rate, the batch size, the choice for the activation function, the number of neurons and layers in the network, among others.

Problems can arise in the training of ML models, one of the most common ones is called overfitting. Overfitting occurs when a model is trained and it matches very closely with the training set, causing the model to fail with other data different from the training set. Underfitting is the opposite problem to overfitting; the model is trained but matches poorly with the training set; as a result, it does not learn a dominant trend from the data and it fails with every other dataset. The best way to control whether overfitting or underfitting is happening is by controlling how the cost or loss function behaves as epochs are trained. A descriptive image explaining this is found in 2.14 [35].

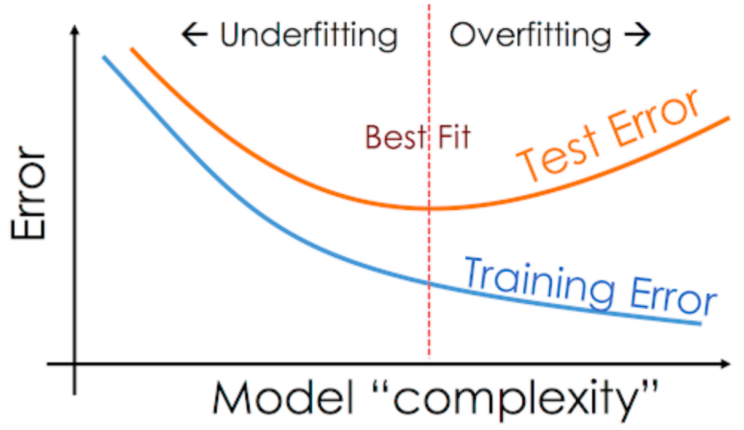

Figure 2.14. Overfitting and underfitting recognized from the learning curve

A way to overcome overfitting can be through the use of regularization techniques. One of the most common ones is called dropout. In this technique, the random nodes in the layer are set to zero in the training process, reducing the size and complexity of the neural network and preventing unnecessary learning of irrelevant information.

Chapter 3

Project Overview

Currently, optical networks must be able to carry massive amounts of data traffic taking into account new technologies and applications that require higher bandwidths. Hence, capacity and optimal exploitation of the network infrastructure appear to be relevant goals for optical fiber communications. These goals can be dealt with a disaggregated optical network. This type of networks are characterized by their flexibility, multi-vendor or open approach, and customizable nature. SDN controllers play an important role in disaggregated networks, as they are the ones in charge of setting an optimal working point for the different elements of the network. In this context, lightpath deployment can be carried out effectively, but the controller in charge should have knowledge of the potential QoT of the lightpaths before deployment. This is only possible with a QoT estimator. A more accurate estimator allows for a smaller difference between design margins and actual margins. GSNR is the common metric used for QoT. In addition to the signal power in the channel, the GSNR has ASE noise and NLI as elements. The degradation of GSNR is mainly caused by ASE noise, as its effect is twice compared to that of NLI [7] [10]. Furthermore, ASE noise is more difficult to predict compared to NLI. In fact, NLI can be approximated with the help of tools using the GN model approach, which is implemented in tools such as GNPy. ASE noise strongly depends on the working point of the EDFAs within the OLSs [7]. In turn, the working point of EDFAs depends on the spectral load. To illustrate this, the data used in this thesis work will be taken as an example; considering two different spectral loads and the whole OLS, it can be seen that in the input of the OLS (Figure 3.1), when channels are turned on, both spectrums align. Therefore, if the same configuration was applied to both spectral loads throughout the OLS, in the output there should be the same alignment between the two spectrums in this spectral portion. But in Figure 3.2, it is evident that the spectral load #18 was amplified a bit more and even a tilt modification can be seen, as the first portion of the channels that are turned on seem to have a smaller gap between the spectrums compared to the last portion.

Investigating how each unique amplifier behaves with the different possible spectral loads before performing a lightpath deployment is evidently not practical at all. Fortunately, this is an ideal scenario for ML and can be considered to characterize the EDFA

and predict its output accurately. In that way, GSNR can also be estimated more accurately, lightpaths can be effectively deployed, and design margins can be set closer to actual margins.

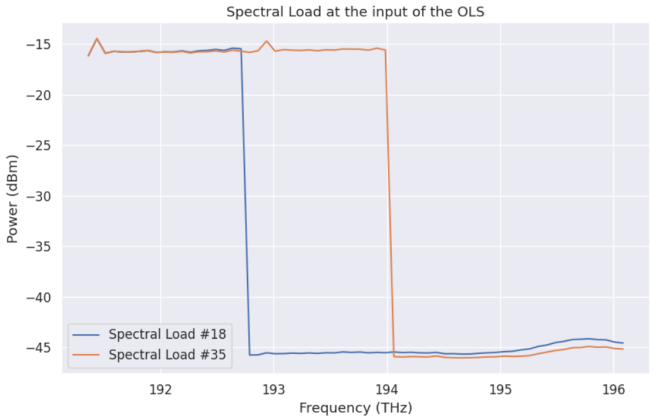

Figure 3.1. Comparison between two different spectral loads at the input of the OLS

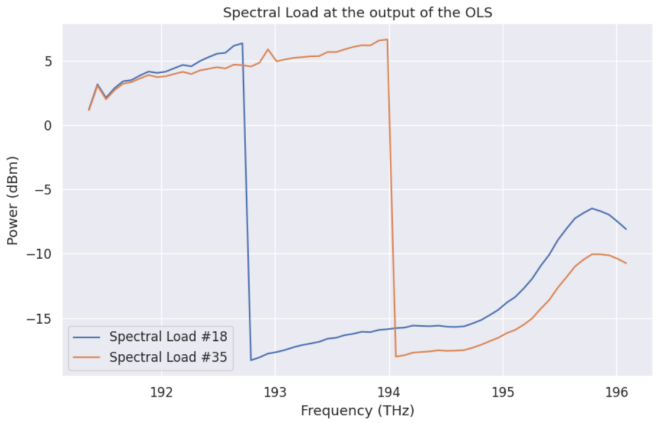

Figure 3.2. Comparison between two different spectral loads at the output of the OLS

3.1 Related Work

Various strategies have been considered to achieve the goals mentioned above. In [10], DNNs are used to predict the OSNR. In this study, the authors decided to focus just on ASE's impact and neglected NLI by using VOAs instead of fiber spans. The features

used for their DNN included the variances of the received signal power and ASE noise. Meanwhile, in [7], NLI was analyzed with the help of GNPy and data from the OLSs, and this analysis was combined with the DNN prediction of the output power coming from EDFAs, in order to perform a more general analysis on GSNR. The channel status was used as a DNN feature. In [11], the output power, including ASE noise and crosstalk from other channels, is also predicted; thus, an analysis of OSNR can be done. Channel status was used as DNN feature as well, but power incoming to the EDFA and EDFA configuration for gain and tilt were also considered as features. In [36], different constructions for DNN models are made, specifically an end-to-end model (complete OLS seen as black-box), a model in which there is a neural network for the BST and the PRE amplifiers and one neural network for all the spans in-between these amplifiers (in a cascaded form), a model that uses a neural network for each amplifier and measurements to account for the loss profile in fiber spans, and a model that uses neural networks for each amplifier and each fiber span also in a cascaded way, in that way having the possibility to model both gain and loss profile for each span. In all the cases mentioned above, EDFAs or OLSs were considered as black-boxes.

In [12], authors emphasize in the losses introduced by the connectors and in the gain profile as the main inaccuracies that lead to large design margins. Taking this into account, they propose an algorithm that estimates these quantities, and they also combine it with a neural network approach. Apart from this, the paper also introduces a novel idea in which the neural network is not just the conventional fully connected sequential neural network, but they also test a structure similar to ResNets called RatioNets. In ResNets, the input of a layer is added to the output of a posterior layer, forcing the neural network to learn the difference between that input and output. RatioNets follow the same idea, but the neural network learns the ratio between input and output instead of the difference.

In [13], [14] and [37], a gray-box approach is tested. Specifically in [13], the idea is to predict with neural networks the difference between the output power when using a gain profile analytical formula and the real measured output power. Then, this value can be summed with the analytical result to obtain the prediction of the actual output power. Two approaches were carried out; the first one is a global approach (end-to- end) in which only one neural network is used to predict all the OLS behavior, while the other is a local approach (span-by-span) that predicts the error per amplifier and, after computing the actual output power and accounting for fiber losses, will be used as input for the next neural network model for the next amplifier. Additionally, in this study, it is shown that as the spectral load gets full, the accuracy of the estimation increases. In [14], the authors are more focused on achieving good results with small datasets, but the idea is similar to that presented in [13]. In [37], the analytical formula used for gain is a very simple one based on the center of mass of a single channel and a fully channel loaded ripple function. This article also presents an innovative way for the neural network architecture as it considers a neural network per channel, instead of a unique neural network for all channels.

As indicated in Section 2.4, the noise figure of an EDFA is not a constant parameter over the wavelenghts. Consequently, to reduce more uncertainties for the design margins,

models on the noise figure have been studied in the master's thesis [38] and in the publication [39]. During the development of this thesis, a constant value of 4 will be considered for the noise figure.

3.2 Experimental Setup and Dataset

The experimental setup used for the data extraction of this project is illustrated in Figure 3.3, it is the same setup used in [13]. It comprises a Booster Amplifier (BST) at the input and a Pre Amplifier (PRE) at the output of the OLS used as setup. In between, there are five In-Line Amplifier (ILA). All amplifiers are connected by spans of standard single-mode fiber, the lengths of the fiber spans are specified in Table 3.1. A C-band WDM comb generated with the help of an ASE noise source and a WSS are present at the input of the BST. This comb has 64 channels and has a 75 GHz channel spacing. Four channels inside this comb, specifically channels #2, #22, #42, #62, are actually being generated by a transponder.

A Python-based OLS controller manages the EDFAs and collects telemetry data, including gain and tilt settings via vendor-specific interfaces [40].

Channel power levels are measured at the OLS terminations and at each EDFA's input and output for all the spectral loads, using an optical switch and an Optical Spectrum Analyzer (OSA).

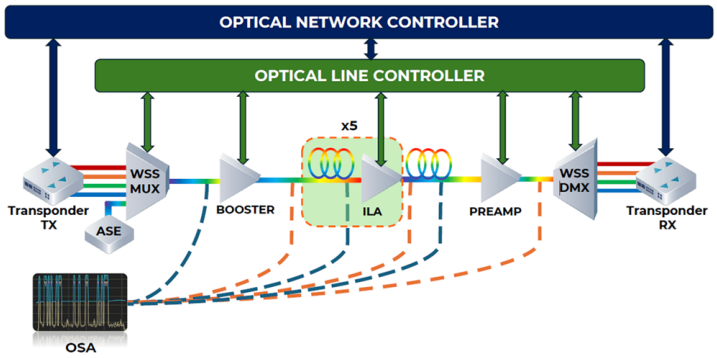

Figure 3.3. Experimental Setup

Regarding the dataset used for this project, 500 spectral loads were taken as samples. For each amplifier in the OLS, each spectral load has:

- ID - created as a way to uniquely identify each spectral load. The ID is actually formed by all the values present in the "Channel Status" column concatenated.
- Channel Status - the state of each channel in terms of presence of signal, ASE noise and crosstalk. In 3.2, a table indicates the meaning of each value in this column.

Span	Length (km)
BST → ILA1	72.22
ILA1 → ILA2	68.97
ILA2 → ILA3	79.02
ILA3 → ILA4	75.70
ILA4 → ILA5	67.44
ILA5 → PRE	71.05

Table 3.1. Fiber span lengths

Value	Meaning
0	Only ASE noise present
$0.3\bar{3}$	ASE noise and crosstalk from one neighbor channel present
$0.6\bar{6}$	ASE noise and crosstalk from two neighbor channels present
1	ASE noise and signal present
$1.3\bar{3}$	ASE noise, signal and crosstalk from one neighbor channel present
$1.6\bar{6}$	ASE noise, signal and crosstalk from two neighbor channels present

Table 3.2. Meaning of values in "Channel Status"

- Ch# Input Signal - the total input power (ASE noise + signal) value for each channel
- Ch# Output Signal - the total output power (ASE noise + signal) value for each channel
- Ch# ON or OFF - indicates if whether each channel contains a signal (value equal to 1) or not (value equal to 0)
- Gain - gain set in the amplifier
- Tilt - tilt set in the amplifier

From the previously mentioned features and other parameters, such as frequency and tilt ripple, more useful data were included in the dataset as:

- Ch# Expected Output - the expected output power per channel estimated using the analytical formula present in equation 2.6.
- Ch# Difference - the difference per channel between the measurement of the actual output power and the one calculated with the analytical formula in 2.6 .

Other data found in the set were:

- Frequency of each of the 4 channels under test (mean, standard deviation, maximum and minimum of different measurements in a lapse of time)

- Chromatic dispersion of each of the 4 channels under test (mean, standard deviation, maximum and minimum of different measurements in a lapse of time)
- Transmitting power of each of the 4 channels under test (mean, standard deviation, maximum and minimum of different measurements in a lapse of time)
- Total power input (mean, standard deviation, maximum and minimum of different measurements in a lapse of time)
- Per channel ASE noise power input and signal power input
- BER of each of the 4 channels under test (mean, standard deviation, maximum and minimum of different measurements in a lapse of time)
- GSNR of each of the 4 channels under test (mean, standard deviation, maximum and minimum of different measurements in a lapse of time)
- Receiving power of each of the 4 channels under test (mean, standard deviation, maximum and minimum of different measurements in a lapse of time)
- Total power output (mean, standard deviation, maximum and minimum of different measurements in a lapse of time)
- Per channel ASE noise power output and signal power output

The data extraction or telemetry process was already done prior to the development of the thesis, so this work will not focus on this procedure.

3.3 White-box Analysis

As mentioned in Section 2.4, an analytical model for the behavior of the amplifier can be used in order to compute the output power profile. The equation of the model is present in 2.6. This model will be referred to as White-box model (WB) throughout this thesis. This model is characterized by being flexible, in the sense that all the parameters can be specific to the EDFA that will be modeled, except for the tilt ripple, which is usually related to a family of EDFAs. In addition to flexibility, the model is also descriptive, as it is easy to understand the behavior just by looking at the equation, and it is low cost as only two measurements per EDFA are required to compute the model. By using the dataset given, an analysis on the errors of this WB approach is done in this section. This analysis will be useful for later comparison with the gray-box approach.

Table 3.3 contains the Mean Absolute Error (MAE) and Maximum Absolute Error (MaxAE) of this white-box approach. Combining the values per amplifier present in this table will result in a considerable error for the whole OLS, either if we consider MAE or MaxAE. This makes the introduction of ML in models more attractive, as it can also improve performance.

In order to have a more visual way to analyze the WB approach, a plot showing the predicted and the real spectral load with the maximum error, corresponding to one of the spectral loads in the BST, is present in Figure 3.4.

Amplifier	MAE	MaxAE
BST	0.80 dB	3.29 dB
ILA1	0.48 dB	4.67 dB
ILA2	0.33 dB	2.47 dB
ILA3	0.39 dB	2.79 dB
ILA4	0.26 dB	1.49 dB
ILA5	0.25 dB	2.07 dB
PRE	0.32 dB	3.06 dB

Table 3.3. Mean Absolute Error and Maximum Absolute Error per amplifier with the white-box approach

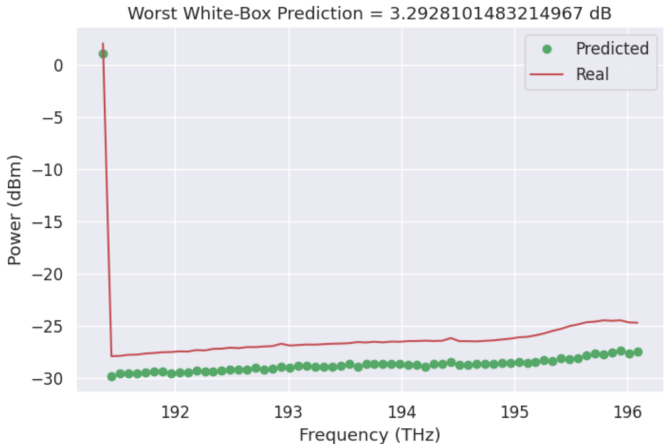

Figure 3.4. Worst configuration for white-box approach

It can be noticed in Figure 3.5 that if the MAE is analyzed taking into account how many channels are on, that as the spectral occupation gets full (all channels are turned on), the better the performance of the model. Actually, it can be seen that the model performs really poorly whenever the number of channels turned on is very small. This can be further evident when comparing Figures 3.6 and 3.4. In Figure 3.6 the predicted and the real full spectral load of the same BST analyzed in Figure 3.4 are shown.

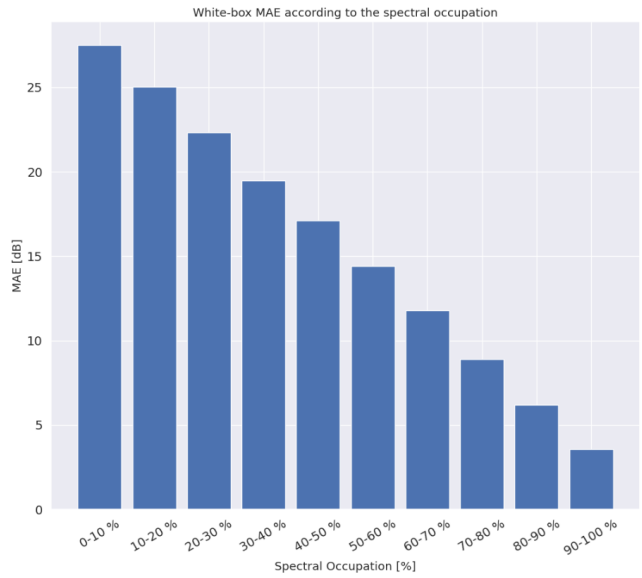

Figure 3.5. MAE according to spectral occupation with the white-box approach

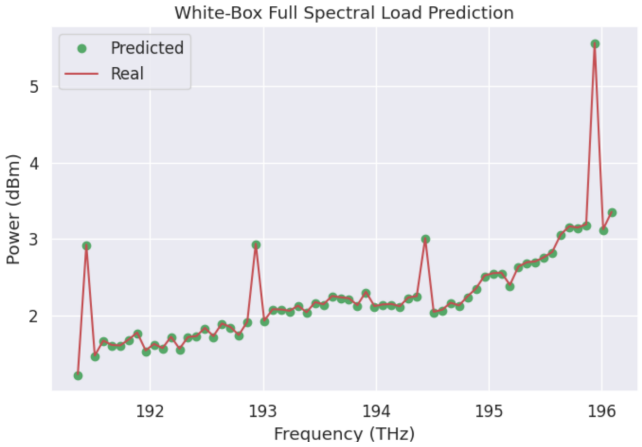

Figure 3.6. Full spectral load configuration for the white-box approach

Chapter 4

EDFA Analysis: Gain Profile Prediction

In this thesis work, a local perspective is taken into account, so fiber spans and EDFAs will be handled separately, meaning that each amplifier and fiber span will have its own separate neural network and model. In this chapter, the methodology for building the neural network and overall the model for EDFAs is described.

In order to start with the analysis, the work done in the paper "Enhancing Optical Multiplex Section QoT Estimation Using Scalable Gray-box DNN" ([13]) was used as a reference. The gray-box model proposed in this paper was taken as an initial guide; therefore, in this thesis work it will be called Default Gray-box model (D-GB).

The procedure carried out can be summarized as follows:

1. **Default model structure and results -** Test the default model and then store its results. Write down how this model is structured and its parameters.

2. **Feature and label possibilities -** Annotate all the possible features and labels that can be taken into account

3. **Hyperparameters -** Write down all the possible hyperparameters to take into account.

4. **Gray-box models -** Think in variations or new ideas for the structure of the DNN model inside the gray-box model.

5. **Hyperparameter testing -** Taking into account the different models, test them with a brief analysis of their optimal hyperparameters. Test a specialized Python library to find optimal hyperparameters like Optuna and compare it to the results of a traditional grid search in terms of accuracy and training time.

6. **Best model results -** With the best hyperparameters, verify if there is no overfitting and considering the test set, take the final results for the best proposal in terms of performance.

4.1 Default model structure and results

4.1.1 Model structure

As already mentioned in the introduction of this chapter, this model follows a gray-box approach in which the analytical model of the gain profile, described in section 2.4 and specifically in equation 2.6, and a neural network are employed to obtain an estimate for the output power per channel. The main strategy is making the neural network predict the difference between the output of the analytical formula and the actual measured output power, so then this result can be summed to the output power that the analytical formula suggests, and thus the measured output power can be obtained. In fact, this idea will be the basis for the models presented throughout this thesis. The diagram in Figure 4.1 describes the default model structure for just one of the amplifiers. In the next chapter, the complete model considering fiber spans will be shown.

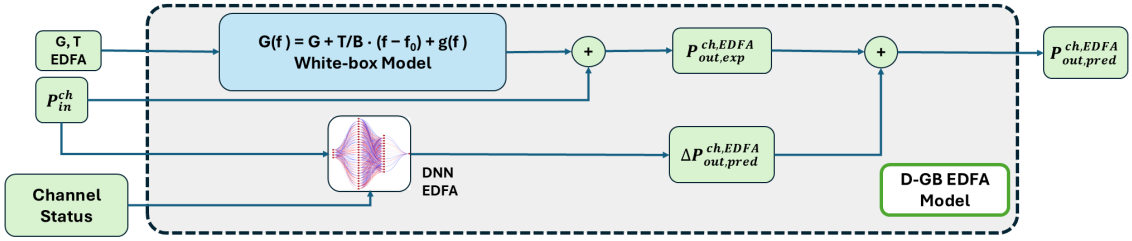

Figure 4.1. D-GB model structure

The main characteristics of the neural network present in the D-GB are :

- **Inputs (Features):** 128 features divided into 64 channel status and 64 channel input powers.
- **Output (Labels):** 64 labels that correspond to the error per channel between the prediction of the output power of the analytical formula and the measured output power.
- **Training, testing and validation set:** The whole dataset is divided following a proportion of 70/15/15. 70% of the dataset is for the training set, and the validation and test set get 15% of the dataset each.
- **# of epochs:** 100 epochs
- **Batch size:** 16
- **# of hidden layers:** 2 layers
- **# of neurons per hidden layer:** 1024 neurons
- **Learning rate:** 0.0001

- **Patience:** 10 epochs
- **Activation Function:** ReLu function defined as: $f(x) = max(0, x)$

4.1.2 Results

The performance of this gray-box model in terms of MAE and MaxAE for each amplifier is displayed in Table 4.1.

Amplifier	MAE	MaxAE
BST	0.11 dB	0.74 dB
ILA1	0.07 dB	0.69 dB
ILA2	0.05 dB	0.36 dB
ILA3	0.06 dB	0.47 dB
ILA4	0.05 dB	0.45 dB
ILA5	0.05 dB	0.39 dB
PRE	0.07 dB	0.52 dB

Table 4.1. Mean Absolute Error and Maximum Absolute Error per amplifier with the default gray-box approach

As a way to analyze in a more visual form the performance of the model, the predicted and measured power output of the configuration containing the worst MaxAE in the BST, the amplifier with the worst performance, was plotted in Figure 4.2. For the predictions, we include both the WB and D-GB model. A comparison between the MAE and MaxAE in the configuration of Figure 4.2 is presented in table 4.2. In addition, the MAE according to the spectral occupation in the BST was plotted in Figure 4.3. The predicted and measured power outputs for the full spectral load case in the BST were also plotted in Figure 4.4. Again, both WB and D-GB predictions were considered and the table indicating the metrics for this configuration is present in 4.3. The total time to train all the neural networks was 57.17 seconds.

Model	MAE	MaxAE
WB	0.77 dB	1.44 dB
D-GB	0.18 dB	0.86 dB

Table 4.2. Mean Absolute Error and Maximum Absolute Error of the configuration shown in Fig. 4.2 for the D-GB and WB model

A big difference regarding accuracy can be seen in this gray-box approach when comparing it with respect to the WB approach, specially in configurations with low spectral occupation. In the WB approach, the MaxAE of all amplifiers exceeded 1 dB, while in the gray-box approach none of them surpasses this value. Therefore, it can be stated that most of the error of the WB model comes from noise prediction. Taking into account the MAE values for each amplifier, in the gray-box approach, the combination

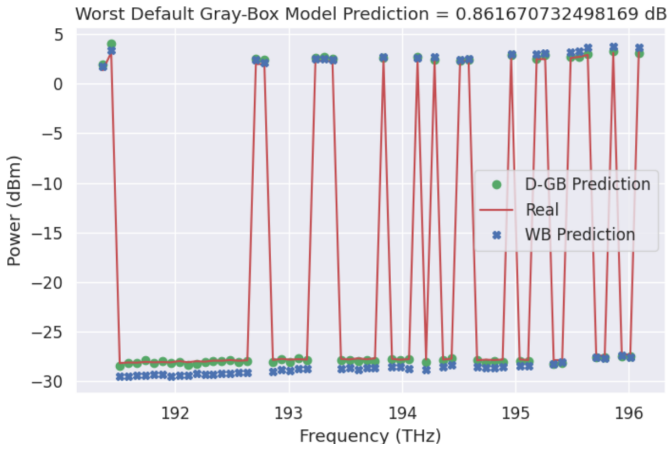

Figure 4.2. Worst configuration for default gray-box model

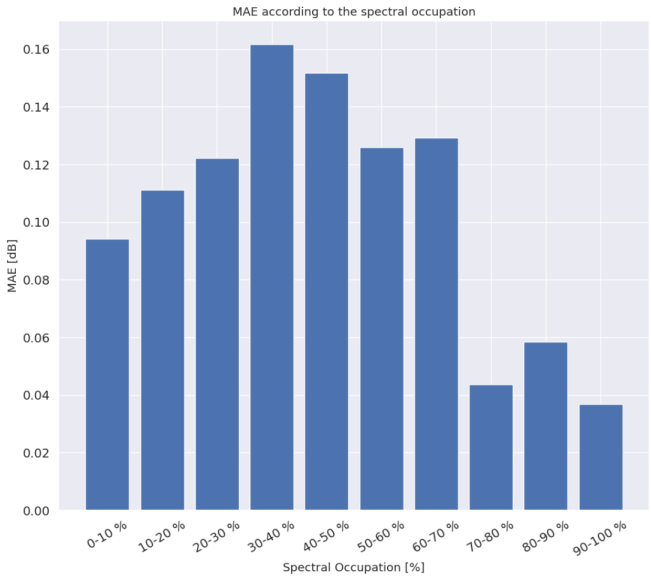

Figure 4.3. MAE according to spectral occupation with the default gray-box model

of errors would not result in a large total value as the one that can result from the WB approach. Visually, the difference in accuracy between both strategies is also observed in Figure 4.2. This is further confirmed with the information presented in Table 4.2.

In both approaches, the best performances are obtained when dealing with a full spectral load. In fact, it seems that as the analytical formula works so well by itself

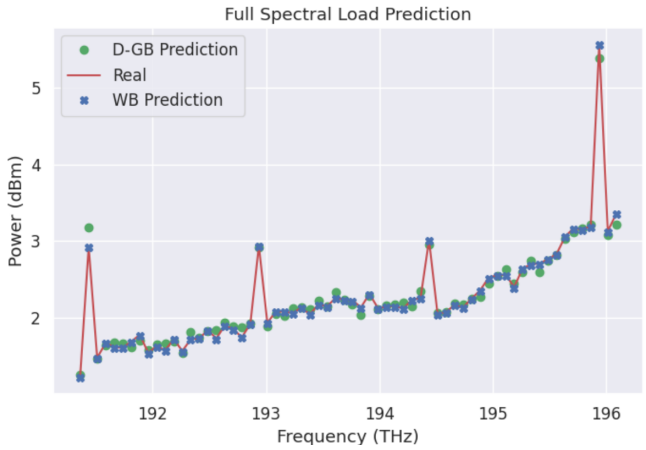

Figure 4.4. Full spectral load configuration for the default gray-box model

Model	MAE	MaxAE
WB	0.01 dB	0.01 dB
D-GB	0.05 dB	0.25 dB

Table 4.3. Mean Absolute Error and Maximum Absolute Error of the configuration shown in Fig. 4.4 for the D-GB and WB model

in this case, it causes an impact in the gray-box approach too. Actually, by comparing Figures 3.6 and 4.4, it is noticed that performance is better in the WB approach for this case. This is numerically corroborated as the MAE in the full spectral load configuration for the WB approach is 0.01 dB, while for the gray-box case the value is 0.05 dB, as stated in Table 4.3. Despite this, it can be confirmed with the tables and plots presented for both strategies that overall the best performances are obtained with the gray-box approach.

Regardless of having good results with the D-GB, performance can possibly be further enhanced, and this is why different gray-box model variations are tested in the next sections.

4.2 Feature and label possibilities

In order to better organize the ideas for creating new models, the data extracted described in Section 3.2 were divided into possible features or labels for the neural networks. The division resulted in the following:

4.2.1 Features

- Signal, noise and crosstalk state

- Channel under test frequency
- Channel under test chromatic dispersion
- Channel under test transmitting power
- Amplifier's tilt
- Amplifier's total power input
- Amplifier's per channel power input (signal, ASE noise, signal + ASE noise)
- Analytical model power output
- Channel on/off state (without considering crosstalk)

4.2.2 Labels

- Channel under test BER
- Channel under test GSNR
- Channel under test receiving power
- Amplifier's gain
- Amplifier's total power output
- Amplifier's per channel power output (signal, ASE noise, signal + ASE noise)
- Difference between analytical and measured power output

4.3 Hyperparameters

Regarding the possible hyperparameters that can be tuned to get better performances, there are:

- Activation function
- Batch size
- Number of hidden layers
- Number of neurons per hidden layer (constant number of neurons through the layers will be considered for this work)
- Learning rate
- Number of epochs
- Patience

- Choice for the optimizer algorithm, this is the algorithm used to update weights inside the neural network
- Coefficients of regularization techniques, if they are present, as dropouts to avoid overfitting

4.4 Gray-box models

Apart from the already analyzed D-GB, other DNN structures based on the ideas proposed in the related work for this topic were tested. In the next subsections, a short description and a diagram for each of the DNNs considered will be presented.

4.4.1 Variation of the DNN output or label

The idea of this neural network is to follow the same logic as the D-GB one but instead of selecting as DNN label the difference between the analytical output power and the measurement of the ouput power, the goal for the DNN will be to predict the difference between the analytically estimated gain and the actual measured gain. Then this difference will be summed to the analytical gain to obtain a final estimated gain for this model.

Taking into account this, a new model can be derived that will be referred throughout this thesis as:

- Default Gray-box Gain model (D-GB-G) - The variation of the D-GB model with the difference of gain as label. Shown in Figure 4.5.

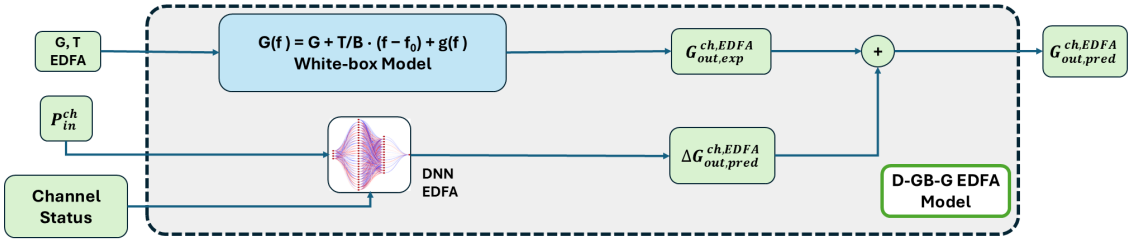

Figure 4.5. D-GB-G model structure

4.4.2 Analytical information included in the DNN features

The idea for the neural networks of the models of this variation is to incorporate analytical estimates as features for the DNN. This can be done either for the label of the difference of measured and analytical output power or gain. Other features that can be included, apart from the analytical information, are both the input power and the channel status, or just one of these.

Considering this, new models can be derived that will be referred throughout this thesis as:

- Both features Gray-box model (B-GB) - Includes as features the analytical information and both the input power and the channel status and as label the difference of powers. Shown in Figure 4.6.

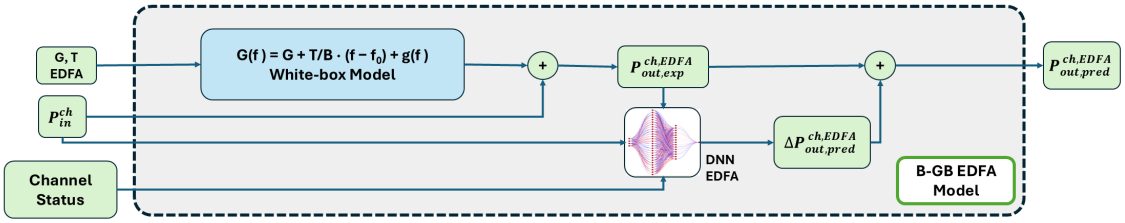

Figure 4.6. B-GB model structure

- Both features Gray-box Gain model (B-GB-G) - Includes as features the analytical information and both the input power and the channel status and as label the difference of gains.Shown in Figure 4.7.

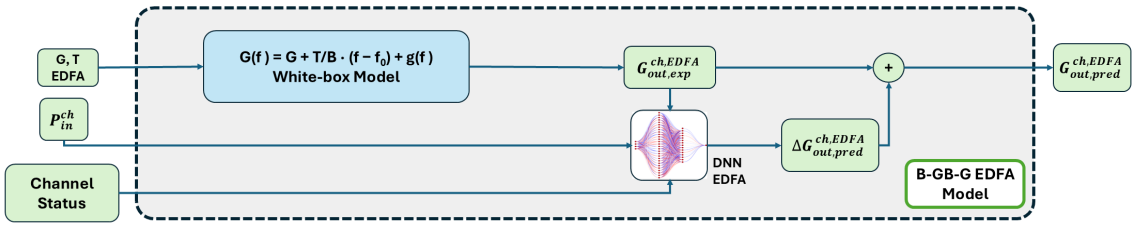

Figure 4.7. B-GB-G model structure

- Power as feature Gray-box model (P-GB) - Includes as features the analytical information and the input power and as label the difference of powers. Shown in Figure 4.8.
- Power as feature Gray-box Gain model (P-GB-G) - Includes as features the analytical information and the input power and as label the difference of gains. Shown in Figure 4.9.
- State as feature Gray-box model (S-GB) - Includes as features the analytical information and the channel status and as label the difference of powers. Shown in Figure 4.10.

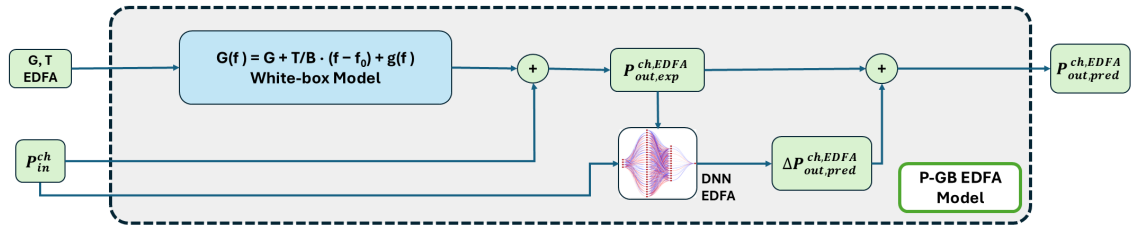

Figure 4.8. P-GB model structure

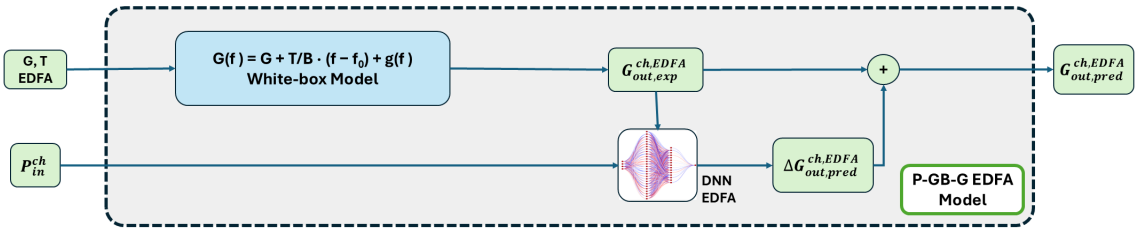

Figure 4.9. P-GB-G model structure

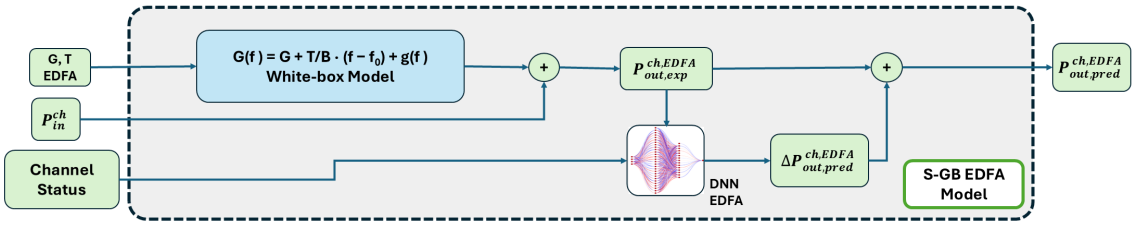

Figure 4.10. S-GB model structure

- State as feature Gray-box Gain model (S-GB-G) - Includes as features the analytical information and the channel status and as label the difference of gains. Shown in Figure 4.11.

4.4.3 Multiple DNN

The idea of this model is to create a neural network predicting the difference between analytical and measured output power or gain for each channel. This means that since there are 64 channels studied in this thesis work, there will be 64 neural networks. These

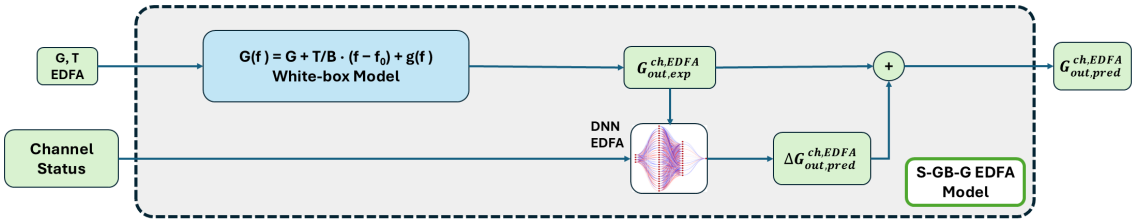

Figure 4.11. S-GB-G model structure

neural networks can have the structure of the neural networks described in the previous subsections. The multiple DNN model will be referred to as Multiple Deep Neural Network model (M-DNN). A figure describing the structure is given in 4.12. This figure follows the S-GB structure for each neural network.

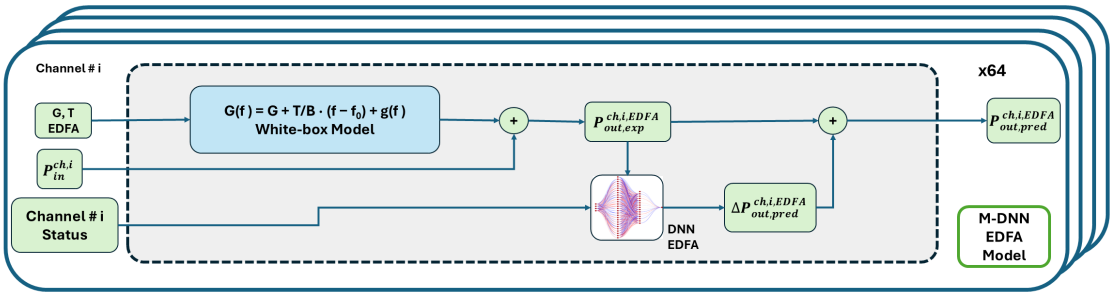

Figure 4.12. M-DNN model structure

4.5 Hyperparameter testing

Two methodologies are going to be tested to see which is the best set of hyperparameters for each of the proposed models. One is the traditional grid search, in which each possible combination of hyperparameter values is tested one by one. This strategy is clearly neither fast nor efficient, as the number of combinations can be large and the values for each numerical hyperparameter come from a discrete set that the developer of the code defines. The other strategy is by using the Python library Optuna, a specialized library for hyperparameter search. In Optuna, the range or number of options for the hyperparameter values used for the test could be larger because the number of trials within the test is fixed and set by the code developer. The library ignores combinations that will result in poor performances and includes in its analysis the combinations that

can potentially result in better performances. The biggest advantage of the grid search compared with Optuna is its exhaustiveness as it tests every possible combination defined by the code developer, leaving no doubt on what would be the best option among those possible combinations; in Optuna results may vary slightly if two different tests are done with the same ranges and options for the values of the hyperparameters (except if all the hyperparameters are of categorical nature and not numerical), therefore, generally there is not a unique best result.

In Section 4.3, a list of possible hyperparameters was registered. Considering all these hyperparameters in the grid search will cause the process to extend considerably in time. While, if we consider each one of them in Optuna, it might cause the process to become more complex; therefore, a greater number of trials should be configured; otherwise, the best set of hyperparameters indicated might not be the most optimal.

Both the number of epochs and the patience are hyperparameters that can be controlled depending on the evolution of the learning process of the model. For example, if we notice that the performance is not improving after a long number of epochs or overfitting is starting to be evident, then it should be considered lowering the patience value or the number of epochs or both; otherwise if the learning process seems to be interrupted at an early stage, then increasing the number of epochs or the patience or both is a good option. Overfitting can also be controlled with dropouts or other regularizers, but this strategy might worsen performance. Taking into account this, these hyperparameters will only be tuned if considered necessary, but they will not be included in the studies done with the grid search nor with Optuna.

One of the most widely used algorithms for the optimizer in deep learning is called Adaptive Moment Estimation (Adam). Adam modifies the learning rate of each parameter in each time step. As Adam adapts well for this context, it will be used and will not take part of the hyperparameter tuning.

Then, the hyperparameters that will be tuned with the grid search and Optuna are:

- Activation function
- Batch size
- Number of hidden layers
- Number of neurons per hidden layer (constant number of neurons through the layers will be considered for this work)
- Learning rate

The range and type of values considered for the optimization process in Optuna are indicated in Table 4.4.

In the learning rate, the value is sampled following a uniform probability in that range in the logarithmic domain, so there is no step as shown in Table 4.4. Regarding the activation functions considered in Optuna, in Table 4.5, the name and formula with the values set for other variables for the contemplated options are shown.

In Table 4.5, in RReLU, the parameter a is randomly sampled from $v(a, b)$, which means a uniform distribution from a to b.

Hyperparameter	Type	Minimum	Maximum	Step
Learning rate	float (real number)	10^{-6}	10^{-1}	None
Number of layers	integer	1	5	1
Number of neurons	integer	64	2048	32
Batch size	integer	16	128	16

Table 4.4. Ranges and types for hyperparameter tuning in Optuna

Activation function	Formula	Other variables
ReLU	$ReLU(x) = \max(0, x)$	
Leaky ReLU	$LeakyReLU(x) = \max(0.1x, x)$	
ELU	$ELU(x) = \begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$	$\alpha = 1$
ReLU6	$ReLU6(x) = \begin{cases} x & 0 \leq x \leq 6 \\ 6 & x > 6 \\ 0 & x < 0 \end{cases}$	
Hard Swish	$Hardswish(x) = x\frac{ReLU6(x+3)}{6}$	
PReLU	$PReLU(x) = \max(0, x) + a * min(0, x)$	$a = 0.25$
RReLU	$RReLU(x) = \max(0, x) + a * min(0, x)$	a sampled from $\upsilon(\frac{1}{8}, \frac{1}{3})$
SELU	$SELU(x) = scale * (max(0, x) + min(0, \alpha * (e^x - 1)))$	$\alpha \approx 1.67$, $scale \approx 1.05$
GELU	$GELU(x) = x * \frac{1}{2}[1 + erf(x/\sqrt{2})]$	

Table 4.5. Activation functions considered in the Optuna hyperparameter search

It can be noticed that the number of possible combinations of the set of hyperparameters in Optuna is huge. This can not be replicated in grid search as it will take several days to run completely the algorithm. Consequently, in the grid search, three values for the number of layers (1,2,3), three for the number of neurons (128,1024,2048), three for learning rate (10^{-6},10^{-4},10^{-2}), two for batch size (16,32) and three activation functions (ReLU, Leaky ReLU, and ELU) were tested. This means 162 combinations, for all of the possible DNNs that can be derived from the descriptions in Section 4.4, except M-DNN. There still seem to be a lot of combinations, and indeed the time taken to run the grid search algorithm for a single model is less than a couple of hours. Anyway, compared to the more than 117,000 possible combinations in Optuna, the number of combinations for the grid search is small.

As mentioned above, the idea is to perform the hyperparameter search with these two methods for each possible model that can be derived from the ideas proposed in Section 4.4. The exception will be the models derived from M-DNN, as a single hyperparameter search for a M-DNN takes a considerable time to run. Therefore, only one neural network from the M-DNN will be tested in the hyperparameter search with the methodology that gave the best results. In addition, to save time, only one of the possible M-DNNs that can be constructed will be tested. The structure of this neural network will be the one

that gave the best results in the other DNN models.

The results of the best hyperparameters obtained for each model and methodology are summarized in Tables 4.6, 4.7, 4.8, 4.9, 4.10, 4.11, 4.12, 4.13.

	Grid Search	Optuna
Average MAE among amplifiers [dB]	0.065	0.060
Time [s]	27.23	21.02
Learning Rate	0.0001	0.0002
Activation Function	LeakyReLU	RReLU
Batch Size	32	80
Number of hidden layers	3	4
Number of neurons per hidden layer	2048	1696

Table 4.6. D-GB results

	Grid Search	Optuna
Average MAE among amplifiers [dB]	0.196	0.191
Time [s]	59.16	30.34
Learning Rate	0.0001	0.0003
Activation Function	ELU	SELU
Batch Size	16	48
Number of hidden layers	3	4
Number of neurons per hidden layer	2048	1568

Table 4.7. D-GB-G results

	Grid Search	Optuna
Average MAE among amplifiers [dB]	0.068	0.061
Time [s]	53.36	21.89
Learning Rate	0.0001	0.0006
Activation Function	Leaky ReLU	ELU
Batch Size	16	80
Number of hidden layers	3	5
Number of neurons per hidden layer	2048	1632

Table 4.8. B-GB results

It is important to recall that the row called "Time" in these tables corresponds to the time taken to train the neural network just for the best set of hyperparameters, not the total time of the hyperparameter search.

Until now, it has been evident that the performances are better when the label is the difference between the analytical and measured output power compared to the one of the gain. Also, it is important to indicate that in most of the tests Optuna did a better job

	Grid Search	Optuna
Average MAE among amplifiers [dB]	0.199	0.181
Time [s]	48.88	30.94
Learning Rate	0.0001	0.0003
Activation Function	ELU	SELU
Batch Size	16	48
Number of hidden layers	3	3
Number of neurons per hidden layer	2048	1600

Table 4.9. B-GB-G results

	Grid Search	Optuna
Average MAE among amplifiers [dB]	0.087	0.079
Time [s]	52.97	25.08
Learning Rate	0.0001	0.0013
Activation Function	ReLU	GELU
Batch Size	16	64
Number of hidden layers	3	3
Number of neurons per hidden layer	2048	1696

Table 4.10. P-GB results

	Grid Search	Optuna
Average MAE among amplifiers [dB]	0.207	0.210
Time [s]	58.52	30.01
Learning Rate	0.0001	0.0004
Activation Function	ELU	PReLU
Batch Size	16	48
Number of hidden layers	3	2
Number of neurons per hidden layer	2048	2016

Table 4.11. P-GB-G results

looking for the best hyperparameter set compared to the grid search. For this reason, the grid search analysis will be avoided for the M-DNN test.

From these tables, it can be seen that the best performance is obtained for the S-GB model. Therefore, this structure will be the one used for neural networks in the M-DNN test. The neural network selected for the hyperparameter search will be the one corresponding to channel # 62, as this is one of the actual channels under test, that is not being emulated only by the WSS and the ASE noise source. The results of the hyperparameter search in the M-DNN are present in Table 4.14.

	Grid Search	Optuna
Average MAE among amplifiers [dB]	0.066	0.059
Time [s]	23.17	17.89
Learning Rate	0.0001	0.0003
Activation Function	ReLU	PReLU
Batch Size	32	128
Number of hidden layer	3	4
Number of neurons per hidden layer	2048	1824

Table 4.12. S-GB results

	Grid Search	Optuna
Average MAE among amplifiers [dB]	0.274	0.241
Time [s]	14.97	20.34
Learning Rate	0.01	0.0045
Activation Function	ELU	SELU
Batch Size	32	48
Number of hidden layers	1	1
Number of neurons per hidden layer	2048	1408

Table 4.13. S-GB-G results

	Optuna
Average MAE among amplifiers [dB]	0.091
Learning Rate	0.0052
Activation Function	Hardswish
Batch Size	128
Number of hidden layers	3
Number of neurons per hidden layer	416

Table 4.14. Best hyperparameters for a single neural network in the M-DNN

4.6 Best model results

Finally, it can be verified that the best performance was achieved in the Optuna test for the S-GB model. All previous tests to find the best hyperparameter set were performed with the validation set. Now, the performance of the best model obtained will be retrieved with the test set, but a check should be done on whether the model is overfitting or not.

First, it is important to mention that since the values for the performance of the M-DNN just corresponded to the performance of one portion of the structure, this type of model was also checked with the test set with the hyperparameters from Table 4.14. In the end, the S-GB model remained the best.

In order to verify the presence of overfitting, the value for the loss across the epochs

for the training and validation set will be plotted. The curves formed are called learning curves. As there are seven amplifiers, there are seven neural networks to train and seven plots for the learning curves. In Figure 4.13, the neural network learning curve can be seen for ILA5. This was the neural network that reached the smallest value of the loss among the other neural networks. Anyway, all the other learning curves had a similar behavior as the one shown in 4.13. The loss function selected was Mean Squared Error (MSE). The MSE is defined as in equation 4.1.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{4.1}$$

Where:

- n is the number of elements
- Y_i is the observed value
- $\hat{Y}_i$ is the predicted value

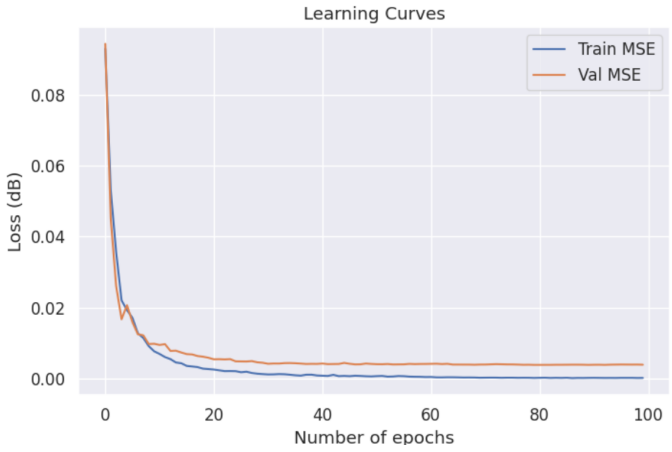

Figure 4.13. Learning curves for the training and validation set in the neural network of ILA5

By comparing the results obtained in Figure 4.13 and the reference image in Figure 2.14, it can be stated that the model is neither overfitting nor underfitting. Both curves follow a similar behavior across the epochs, and at the final epochs there are no signs that the validation MSE increases its value. In addition, the difference between both MSE is small.

Proceeding with the test of the model and hyperparameter set found as the most optimal, the results for the MAE and MaxAE for each amplifier are presented in Table 4.15. Following what has been done throughout this thesis, the plots for the predicted and measured output powers corresponding to the configuration containing the worst MaxAE and the full spectral load case are presented, in Figures 4.14 and 4.15 respectively; as

well as the performance according to the spectral occupation shown in Figure 4.16. All of these graphs are from the amplifier with the maximum MaxAE, in this case the ILA1.

Amplifier	MAE	MaxAE
BST	0.07 dB	0.55 dB
ILA1	0.06 dB	0.65 dB
ILA2	0.04 dB	0.24 dB
ILA3	0.04 dB	0.37 dB
ILA4	0.04 dB	0.31 dB
ILA5	0.04 dB	0.34 dB
PRE	0.06 dB	0.52 dB

Table 4.15. Mean Absolute Error and Maximum Absolute Error per amplifier with the most optimal gray-box model (S-GB model)

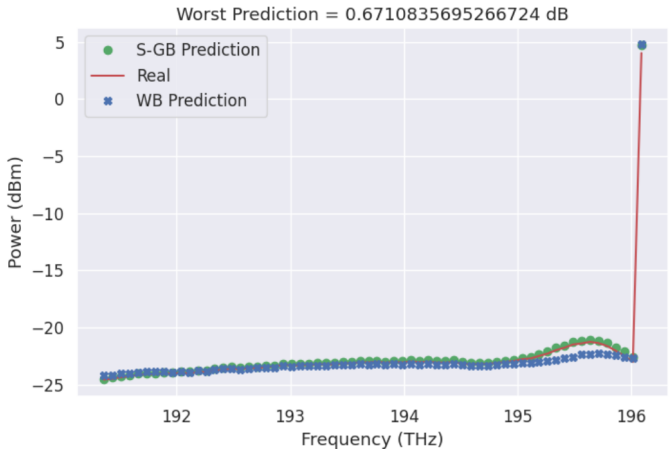

Figure 4.14. Worst configuration for the optimal gray-box amplifier model (S-GB)

Model	MAE	MaxAE
WB	0.27 dB	1.08 dB
S-GB	0.15 dB	0.67 dB

Table 4.16. Mean Absolute Error and Maximum Absolute Error of the configuration shown in Fig. 4.14 for the S-GB and WB model

The time taken to train all the neural networks of each amplifier is 61.30 seconds. It can be noticed that this time is way longer than the ones seen in the tables of the hyperparameter search. The reason behind this is that the neural network models in the final test are being saved in a local server. In the hyperparameter search process, model

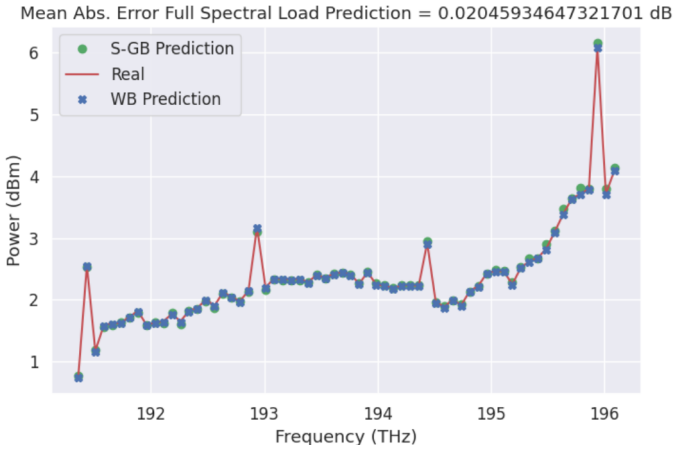

Figure 4.15. Full spectral load configuration for the optimal gray-box amplifier model(S-GB)

Model	MAE	MaxAE
WB	0.01 dB	0.01 dB
S-GB	0.02 dB	0.10 dB

Table 4.17. Mean Absolute Error and Maximum Absolute Error of the configuration shown in Fig. 4.15 for the S-GB and WB model

saving was avoided as eventually it would fill the memory of the server. The time of this optimal configuration (S-GB) is just 4.14 seconds more than the D-GB.

If a comparison is made among the values for the MAE and MaxAE in the D-GB and optimal gray-box (S-GB) model (Tables 4.1 and 4.15 respectively), it is evident that better results are obtained with the new structure and the optimized hyperparameters. In Figure 4.14, plots for the prediction of the gain profile by the WB and optimal (S-GB) model and the real spectral load are depicted. This time, the plot for D-GB was avoided as visually the difference is minimum with respect to the one of S-GB. Anyways as mentioned before, in general S-GB performs better. It can be seen that the configuration with the worst MaxAE is the one with just one channel turned on. This value for the MaxAE is still smaller than the one from the D-GB. The table containing the metrics for this configuration is present in Table 4.16. Regarding the full spectral load configuration, visually it seems that a better fit between both predicted and real output powers is found in the optimal gray-box amplifier model (S-GB). This is also confirmed numerically as the MAE for this configuration in this model is around 0.03 dB less than the D-GB and 0.01 dB more than the WB model. The MAE and MaxAE values for these specific configuration are present in Table 4.17. In Figure 4.16, it can still be seen that the best results are obtained when more than 70% of the channels are on. As well as in the D-GB, this might be the effect of the analytical part in the gray-box model, as we saw before in Figure 3.5 that the analytical formula works really well as more channels are turned on

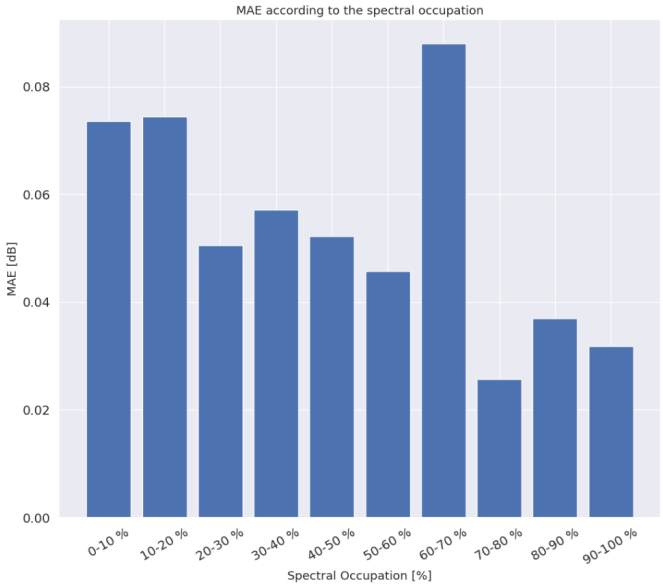

Figure 4.16. MAE according to spectral occupation with the optimal gray-box amplifier model (S-GB)

in the WB model.

It is important to recall that these results correspond only to the analysis for the amplifiers; therefore, in order to perform the complete analysis of the OLS, the fiber spans must also be analyzed. This will be done in the next chapter.

Chapter 5

Fiber Span Analysis : Loss Profile Prediction

Now that there is a gray-box model that predicts accurately the gain profile, a loss profile model should be considered to model the whole OLS. In this chapter, the methodology for building the loss profile model is described.

The procedure carried out can be summarized as follows:

1. **Benchmark: Performance using measurements -** Model the loss profile from the data present in the dataset. These results will later be used as a reference.
2. **GNPy -** Use the GNPy tool to model the loss profile.
3. **Loss profile gray-box models -** Incorporate the ideas for DNN presented in the previous chapter to improve performance.
4. **Hyperparameter testing -** Taking into account the different models, test them with a brief analysis of their optimal hyperparameters. Test a specialized Python library to find optimal hyperparameters like Optuna and compare it to the results of a traditional grid search in terms of accuracy and training time.
5. **Best model results -** With the best hyperparameters, verify if there is no overfitting and considering the test set, take the final results for the best proposal in terms of performance.

5.1 Benchmark: Performance using measurements

In order to have a reference on what the performance of the whole OLS model will be just by using the models of the EDFAs, a simple strategy is applied to model the loss profile. The strategy consists in considering the loss profile as the subtraction of the measurement of the amplifier's output power at the input of the fiber span and the next amplifier's input power at the output of the fiber span. This idea considering the S-GB structure for the amplifiers' model is illustrated in Figure 5.1. The S-GB model will be the one used for the amplifiers in this entire section.

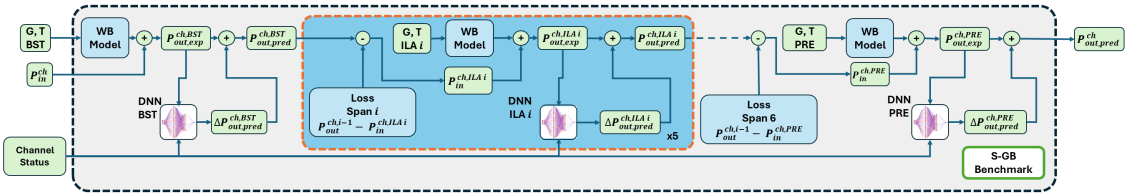

Figure 5.1. OLS model structure that will serve as benchmark

The results for this Benchmark OLS model are summarized in Table 5.1. A plot for the worst predicted configuration, in terms of MaxAE, and for the full spectral load case are included as well in Figures 5.2 and 5.3, respectively. A figure showing the error according to the spectral occupation is also shown in 5.4.

Model	MAE	MaxAE
OLS Benchmark	0.18 dB	2.77 dB

Table 5.1. Mean Absolute Error and Maximum Absolute Error of the OLS Benchmark model

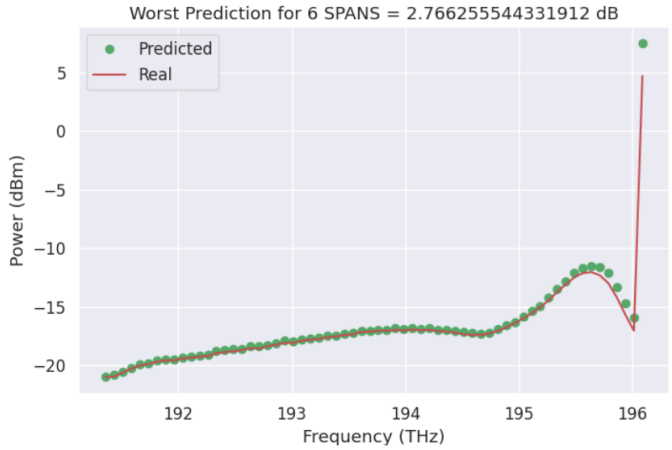

Figure 5.2. Worst predicted configuration in terms of MaxAE for the Benchmark model

Again, it can be seen from the three figures (5.2, 5.3, 5.4) that the best performances come from the spectral loads with more channels on. In fact, the one with the worst MaxAE is one of the configurations with only one channel turned on. The results obtained in Table 5.1 will be treated as the ideal scenario for the rest of the chapter.

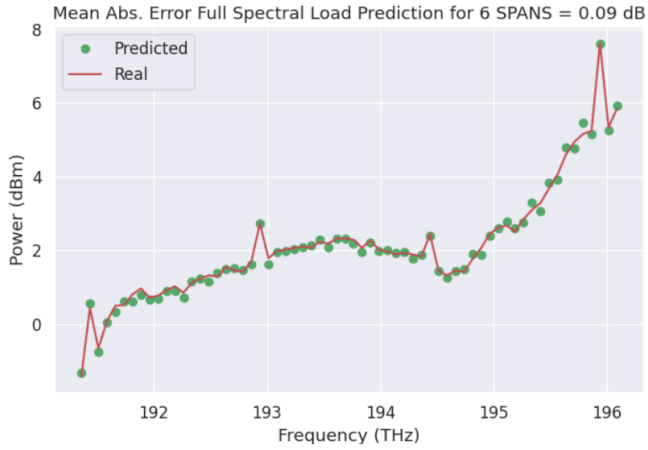

Figure 5.3. Prediction of the full spectral load configuration for the Benchmark model

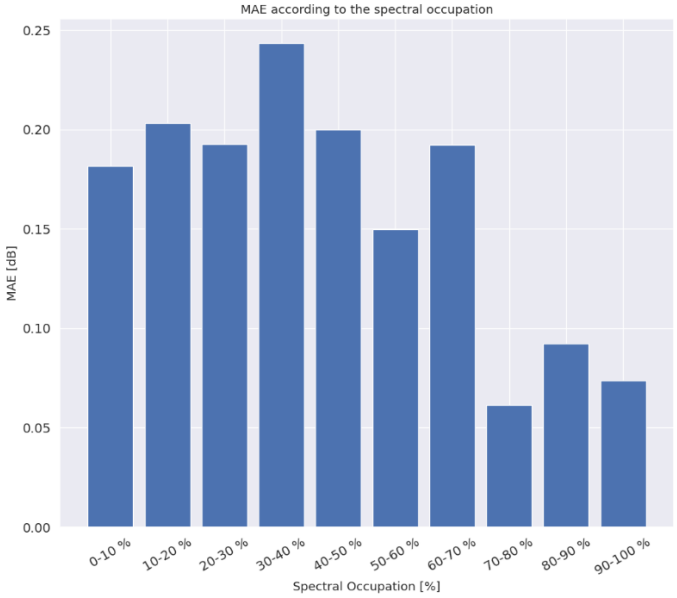

Figure 5.4. MAE according to the spectral occupation for the Benchmark model

5.2 GNPy

As mentioned in Section 2.5.1, GNPy is a tool that can serve as the digital twin for the physical layer of an optical network. Taking this into account, GNPy can be used to

emulate all the phenomena present in the fiber span, and therefore a prediction of the loss profile can be taken from this tool.

The GNPy library includes algorithms and models that simulate a great part of the effects mentioned in Sections 2.3 and 2.4. Fiber attenuation, chromatic dispersion, SRS, and other NLI effects are examples of events modeled by GNPy. This is achieved thanks to the parameters and information of the real devices present in the physical setup. Therefore, it can be ensured that the accuracy of the models in GNPy strongly depends on the accuracy of the information provided. This is also emphasized in [12], where they claim that large errors in models might come from poor characterizations of aspects like the losses from connectors; as in general, these kinds of parameters are not periodically measured due to costs. In addition to this, models inside GNPy can also have limitations in some scenarios. For example, the model used in GNPy for the EDFA is the WB model presented in this thesis. As seen in the previous chapter, this model is accurate only in the full spectral load case; otherwise it is extremely inaccurate, and other strategies such as the gray-box models proposed in this thesis should be implemented. This is the reason why a combination between the EDFA gray-box model and GNPy will be tested in this section. The general image illustrating how the entire OLS model is structured considering the S-GB model for the amplifiers and with the GNPy inclusion is present in Figure 5.5.

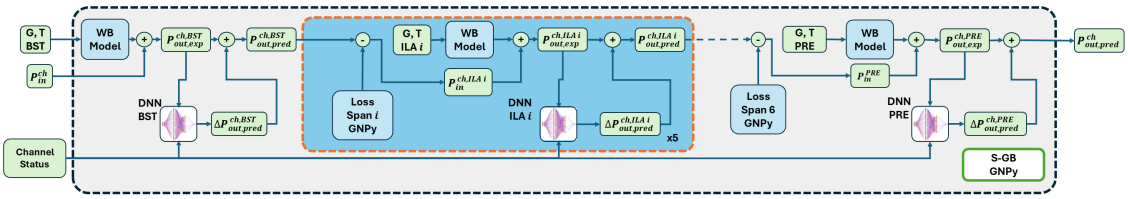

Figure 5.5. OLS model structure using GNPy for loss profile

The results for the MAE and MaxAE of the OLS model with the GNPy tool for the prediction of the loss profile are shown in Table 5.2. As usual, a plot with the configuration containing the worst MaxAE as well as the full spectral load configuration are included in Figures 5.6 and 5.7, respectively. These plots show the GNPy and the benchmark prediction along with the real measurement . Tables specifying the MAE and MaxAE of these configurations are present in 5.3 and 5.4. Finally, the MAE analysis according to the spectral occupation is also performed and shown in Figure 5.8.

Model	MAE	MaxAE
OLS GNPy	0.43 dB	4.00 dB

Table 5.2. Mean Absolute Error and Maximum Absolute Error of the OLS GNPy model

It is clear from the results obtained that the performance using just GNPy to model

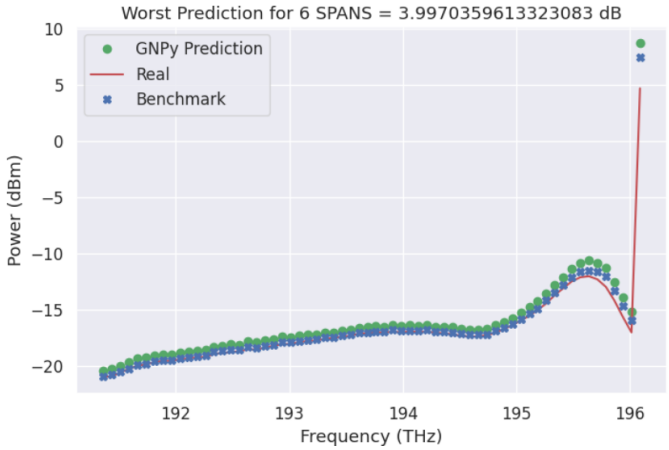

Figure 5.6. Worst predicted configuration in terms of MaxAE for the GNPy model

Model	MAE	MaxAE
Benchmark	0.24 dB	2.77 dB
GNPy	0.80 dB	4.00 dB

Table 5.3. Mean Absolute Error and Maximum Absolute Error of the configuration shown in Fig. 5.6 for the Benchmark and GNPy model

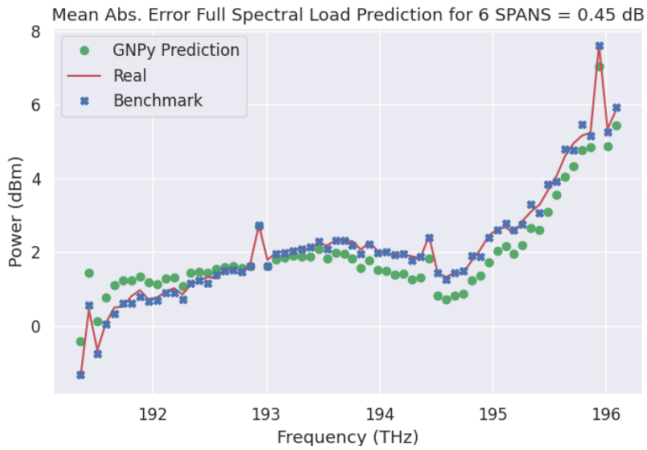

Figure 5.7. Prediction of the full spectral load configuration for the GNPy model

the loss profile is not as good as expected, being far from the benchmark results. Specifically comparing Tables 5.1 and 5.2, it can be observed that the error increased in the

Model	MAE	MaxAE
Benchmark	0.09 dB	0.31 dB
GNPy	0.45 dB	1.01 dB

Table 5.4. Mean Absolute Error and Maximum Absolute Error of the configuration shown in Fig. 5.7 for the Benchmark and GNPy model

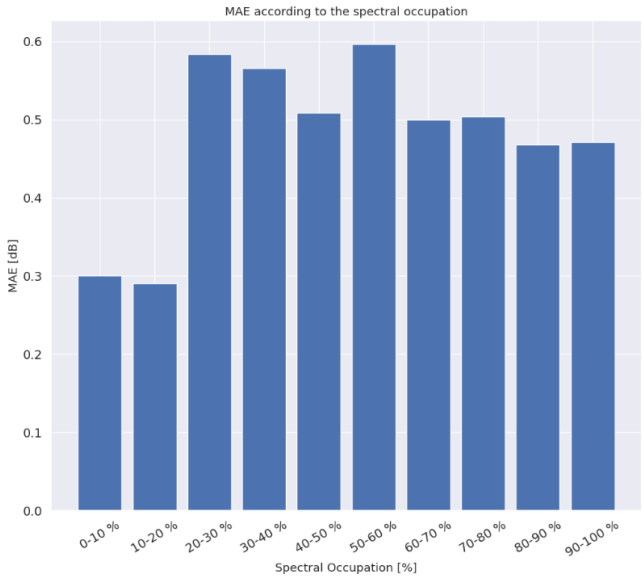

Figure 5.8. MAE according to the spectral occupation for the GNPy model

GNPy model for the MAE in 0.25 dB and for the MaxAE in 1.23 dB. This is further confirmed by looking at Figure 5.6 together with Table 5.3. At first glance in the plot, it might seem that the GNPy prediction was not that bad except for the last channel, the one of the MaxAE, that might have worsen the overall average. Anyway, this also happens in the Benchmark model, and in fact, when looking at Table 5.3, there is still a big difference between the Benchmark and GNPy models.

Other results that may also seem unexpected are present in Figures 5.7 and 5.8. Until now for all the analyses performed, the full spectral load configuration was among the best in terms of performance; but with the inclusion of GNPy, this changed as the prediction is far from the Benchmark prediction and farthest from the real measurements. This is observed visually in Figure 5.7 and numerically in Table 5.4. Also different from the previous analyses, in Figure 5.8, it can be noticed that configurations with low spectral occupation are the ones with the best performances.

In order to understand this performance, a more detailed analysis of the prediction of GNPy fiber loss is given in the next subsection.

5.2.1 Loss profile prediction

To understand whether the error generated by GNPy is more related to an offset or to an inaccurate loss profile construction, the MAE and MaxAE per channel were plotted and are shown in Figures 5.9 and 5.10, respectively. It is important to emphasize that the results of this subsection correspond uniquely to the errors coming from the fiber loss prediction; here, we are not taking into account the errors coming from the gain profile prediction.

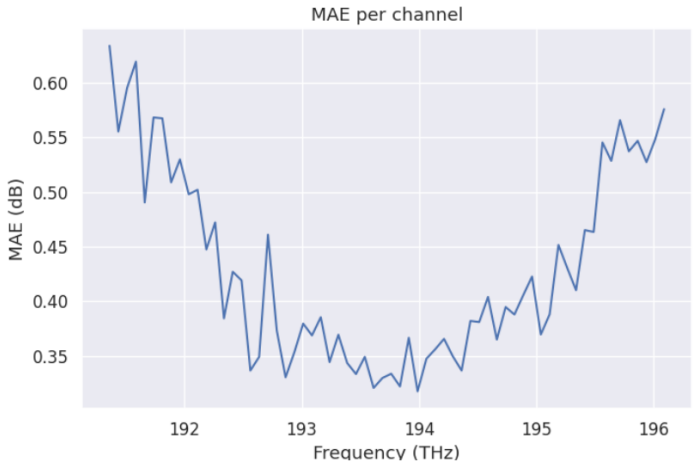

Figure 5.9. MAE per channel in the GNPy model

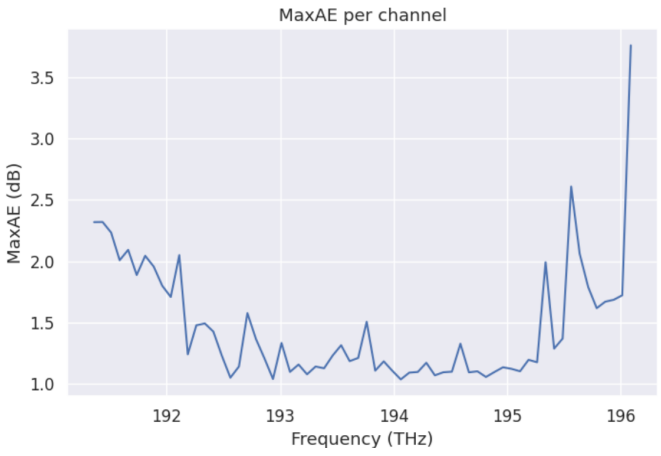

Figure 5.10. MaxAE per channel in the GNPy model

Looking at the plots in 5.9 and 5.10, it can be seen that both have a kind of concave-up

parabola shape. In addition to this, there is a considerable amount of dBs between the highest and lowest MAE and MaxAE. This indicates that the error is not a constant offset over the wavelengths but rather comes from a poor characterization of the loss profile. What these plots are suggesting is that most of the loss profiles are being predicted with a different slope with respect to the real loss profile, forming a figure resembling a cross or even an "X", making the first and last channels more prone to big errors compared to the middle channels. This is confirmed by taking a quick look at the graphs that contain the real loss profiles and the GNPy predicted loss profiles for each configuration. The plots of some of the configurations are found in Figure 5.11.

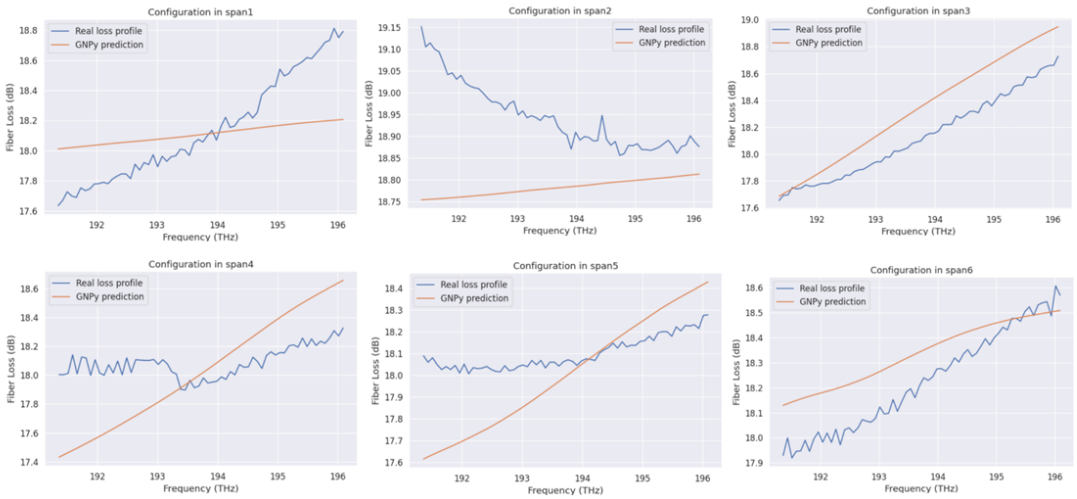

Figure 5.11. GNPy predicted and real loss profiles for some of the configurations in different spans

Apart from this analysis, an examination of the MAE according to the spectral occupation per span was also performed. This was done because it was important to understand if the unexpected results obtained in Figure 5.8 were also coming from the GNPy modeling inclusion. Then, in Figure 5.12, the plots can be found per span.

From the plots in Figure 5.12, it can be seen that despite the fact that in the first and last span the behavior of the MAE follows the expected results (better performance as the full spectral load is reached), the other spans appear to have almost a uniform error for all percentages of spectral occupation, or it appears that the behavior is slightly contrary to the expected one. Therefore, it is difficult to conclude why the results obtained in Figure 5.8 have this unexpected behavior, as there is no clear relationship. However, it can be stated that not all the spans followed the expected MAE behavior and clearly this will not derive in a general expected result in Figure 5.8.

In conclusion, it can be deduced that GNPy by itself is not enough to model the loss profile of the OLS. The analysis indicates that in general the GNPy-generated loss profiles do not follow the same slope, sometimes even generating contrary slopes, with respect to

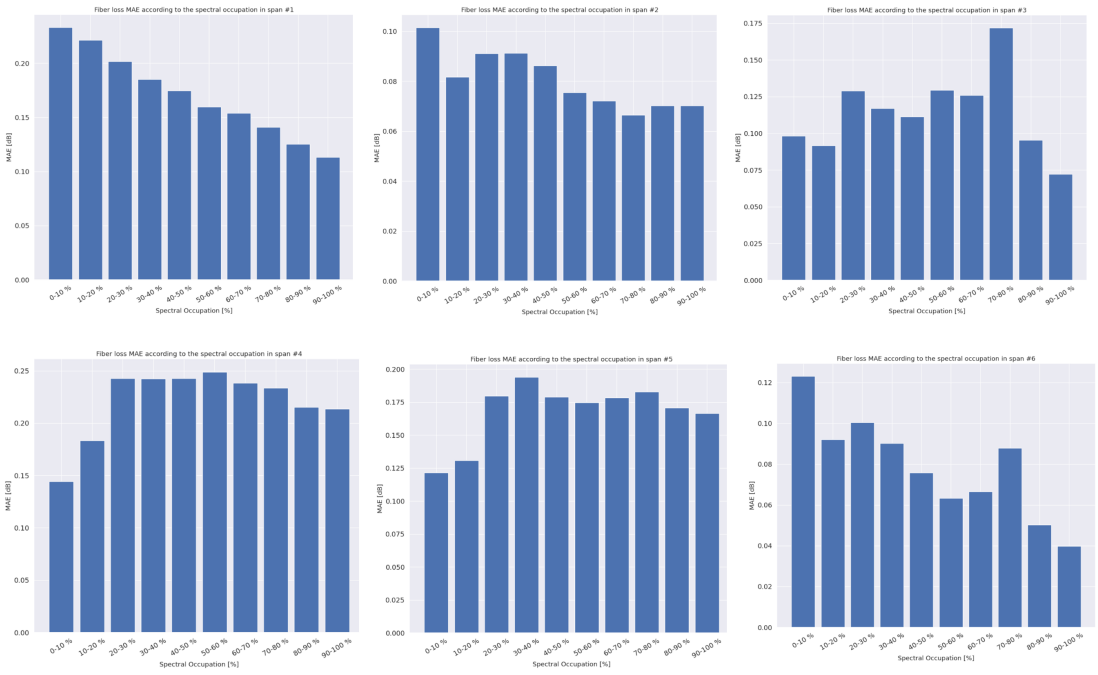

Figure 5.12. MAE according to the spectral occupation for the GNPy model per span

the real measurements. Apart from this, the GNPy inclusion worsen the results for the full spectral load configurations, something that was a strength obtained from the WB model.

As a consequence, another strategy will be tested. This strategy will combine ML and GNPy, so basically it will be a gray-box model for the loss profile. This idea is extended in the next section.

5.3 Loss profile gray-box models

In the previous chapter, it was seen how the different gray-box models tested were derived from the D-GB. In this model, an analytical WB model was considered to generate a reference output power; apart from the WB model, a DNN predicted the difference between this reference output power and the actual measurement of the output power. In the end, this prediction of the difference was summed to the reference output power and thus the output of the whole D-GB model was obtained.

Now the idea for the prediction of the loss profile is similar, but in this case, GNPy will take the place of the WB model and the DNN will predict the difference between the GNPy-generated loss profile and the actual measured loss profile. Then, this predicted difference will be summed to the GNPy-generated loss profile and thus the output of the

gray-box model of the loss profile will be obtained. A diagram representing the whole OLS system with this new idea is shown in Figure 5.13. It is important to mention that the S-GB model was kept for the amplifier because it was the one with the best performance.

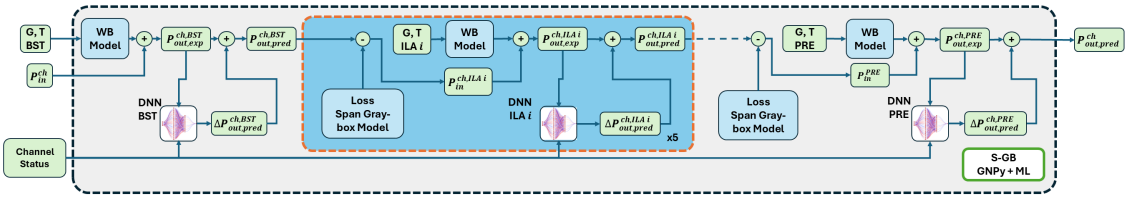

Figure 5.13. OLS model structure considering the loss profile gray-box model

Similarly to what was done in the previous chapter, different variations for the D-GB will be tested, its optimal hyperparameters will be collected and finally the best results will be presented.

The description and explanatory diagrams for each loss profile gray-box model proposed are found in the following subsections. The diagrams shown will represent only the block from loss span gray-box model inside the whole OLS system model.

5.3.1 Loss Profile: Default Gray-box model

As explained in the second paragraph of this section, the idea here is replicating the D-GB, but replacing the WB model with the GNPy tool. The features for the DNN will be the channel status and the power per channel incoming to the fiber and the label will be the difference between GNPy's prediction and the actual loss profile. At the end, this DNN output will be summed to GNPy's prediction in order to get the final output of this gray-box model. This model will be referred as "Loss Profile D-GB". The diagram is found in Figure 5.14.

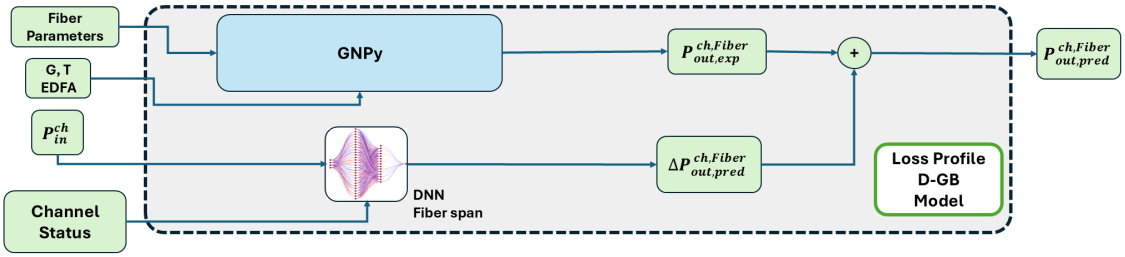

Figure 5.14. Loss profile D-GB model structure

5.3.2 Loss Profile: GNPy information included in the DNN features

The idea for the neural networks present in this type of model variation is incorporating the GNPy output as feature for the DNN. Other features that can be included, apart from the analytical information, are both the input power and the channel status or just one of these.

Considering this, new models can be derived that will be referred throughout this thesis as:

- Loss Profile B-GB - Includes as features the GNPy information and both the input power and the channel status and as label the difference of output powers. Shown in Figure 5.15.

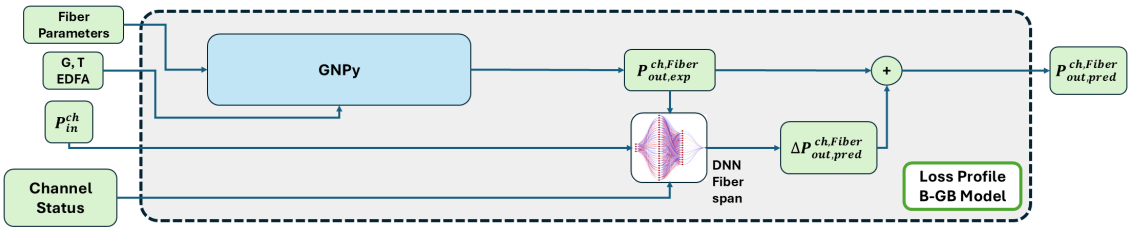

Figure 5.15. Loss Profile B-GB model structure

- Loss Profile P-GB - Includes as features the GNPy information and the input power and as label the difference of output powers. Shown in Figure 5.16.

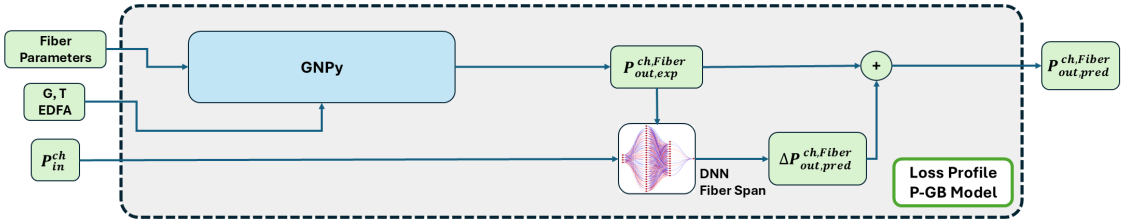

Figure 5.16. Loss Profile P-GB model structure

- Loss Profile S-GB - Includes as features the GNPy information and the channel status and as label the difference of output powers. Shown in Figure 5.17.

5.3.3 Loss Profile: Multiple DNNs

The idea for this model is to create a neural network predicting the difference between GNPy-generated and measured fiber output power for each channel. This means that

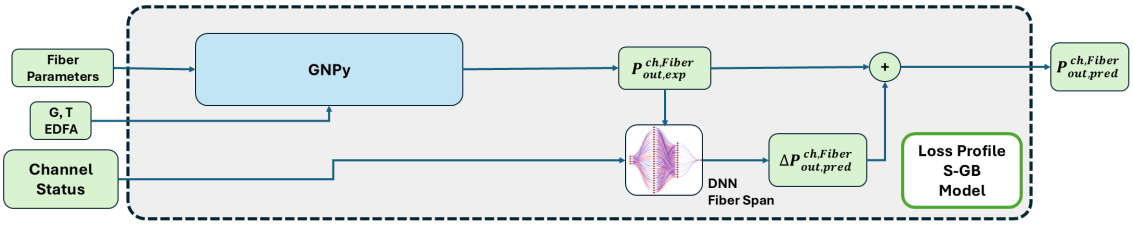

Figure 5.17. Loss Profile S-GB model structure

for this thesis work, as there are 64 channels, there will be 64 neural networks. These neural networks can have the structure of the neural networks described in the previous subsections. This multiple DNN model will be referred as "Loss Profile M-DNN". A figure describing the structure is present in 5.18. This figure follows the Loss Profile S-GB structure for each neural network.

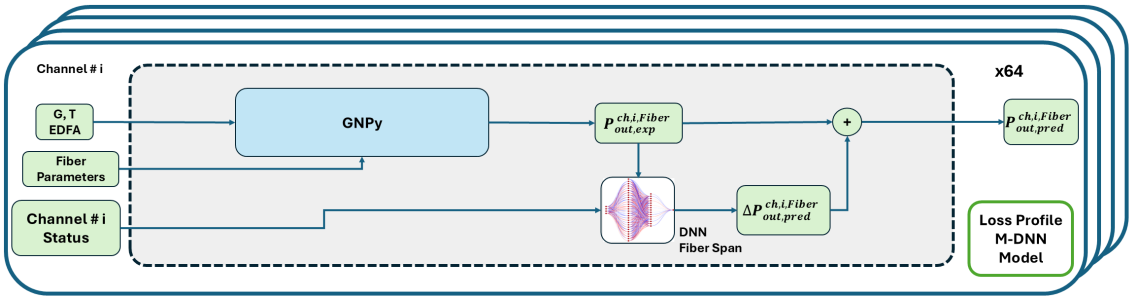

Figure 5.18. Loss Profile M-DNN model structure

5.4 Hyperparameter testing

As well as in Chapter 4, a grid search analysis and Optuna were used to look for the best hyperparameter set in each of the model structures proposed in the previous section. The hyperparameters tuned and the optimizer chosen are also the same as the ones described in Section 4.5.

The values of the hyperparameters analyzed in the grid search are also the same as the ones described in Section 4.5, and the hyperparameter values analyzed in Optuna are the same ones described in Table 4.4 and Table 4.5 (regarding activation functions).

Another thing replicated from Section 4.5 is how the M-DNN variation will be handled. The hyperparameter analysis for the Loss Profile M-DNN will be done just for one of the

neural networks in the structure and only one of the possible variations among Loss Profile M-DNN will be tested, and it will be the model that obtained the best results.

The results for the best hyperparameters obtained for each model and methodology are summarized in Tables 5.5, 5.6, 5.7, 5.8.

	Grid Search	Optuna
Average MAE among amplifiers [dB]	0.178	0.033
Time [s]	60.32	10.35
Learning Rate	0.0001	1.83×10^{-5}
Activation Function	LeakyReLU	LeakyReLU
Batch Size	16	128
Number of hidden layers	3	5
Number of neurons per hidden layer	2048	1696

Table 5.5. Loss Profile D-GB results

	Grid Search	Optuna
Average MAE among amplifiers [dB]	0.174	0.032
Time [s]	59.71	73.50
Learning Rate	0.0001	1.89×10^{-6}
Activation Function	LeakyReLU	ReLU
Batch Size	16	16
Number of hidden layers	3	5
Number of neurons per hidden layer	2048	1664

Table 5.6. Loss Profile B-GB results

	Grid Search	Optuna
Average MAE among amplifiers [dB]	0.174	0.032
Time [s]	42.81	14.45
Learning Rate	0.0001	1.55×10^{-5}
Activation Function	ReLU	ReLU
Batch Size	16	112
Number of hidden layers	3	5
Number of neurons per hidden layer	2048	1248

Table 5.7. Loss Profile P-GB results

Again, it is important to recall that the row called "Time" in these tables corresponds to the time taken to train the neural network just for the best set of hyperparameters, not the total time of the hyperparameter search.

Similar to the hyperparameter analysis performed for the EDFA modeling, Optuna

	Grid Search	Optuna
Average MAE among amplifiers [dB]	0.172	0.031
Time [s]	17.69	14.44
Learning Rate	0.0001	1.67×10^{-5}
Activation Function	ReLU	ReLU6
Batch Size	32	80
Number of hidden layers	3	5
Number of neurons per hidden layer	128	1088

Table 5.8. Loss Profile S-GB results

presented significantly better results while looking for the best hyperparameter set compared to the grid search. For this reason, the grid search analysis will be avoided for the Loss Profile M-DNN test.

From these tables, it can be noticed that the best performance is obtained for the Loss Profile S-GB model. Despite not presenting a significant improvement, the S-GB structure was the best in both gain and loss profile prediction. This structure will be the one used for the neural networks in the Loss Profile M-DNN test. The neural network selected for the hyperparameter search will be the one corresponding to channel # 62, as this is one of the actual channels under test, that is not being emulated only by the WSS and the ASE noise source. The results for the hyperparameter search in the Loss Profile M-DNN are present in Table 5.9.

	Optuna
Average MAE among amplifiers [dB]	0.036
Learning Rate	1.71×10^{-5}
Activation Function	PReLU
Batch Size	96
Number of hidden layers	5
Number of neurons per hidden layer	1248

Table 5.9. Best hyperparameters for a single neural network in the Loss Profile M-DNN

5.5 Best model results

Replicating what was done in the previous chapter, both the best structure with its optimal hyperparameters and the Loss Profile M-DNN with its best hyperparameters, found in Table 5.9, were analyzed in the test set. It is important to recall that all the results from the hyperparameter analysis were done in the validation set. The reason why the Loss Profile M-DNN is also tested in the test set is because the results from the hyperparameter analysis corresponded to just one portion of the whole structure; therefore, just to verify the whole model's structure it will also be analyzed in the test set.

Anyway, when checking both results, Loss Profile S-GB remained as the best performing model, so only the final test set results for this model will be shown.

First, the learning curves will be analyzed, verifying if there is overfitting happening. These curves are found in Figure 5.19. The curves shown correspond to the neural network part of the model of the fiber span connecting ILA1 and ILA2, as this was the neural network that reached the lowest training loss value. However, the behavior seen in Figure 5.19 is similar for all the other neural networks that are part of the models of the other fiber spans. As mentioned in the previous chapter the loss function used is the MSE defined in Equation 4.1.

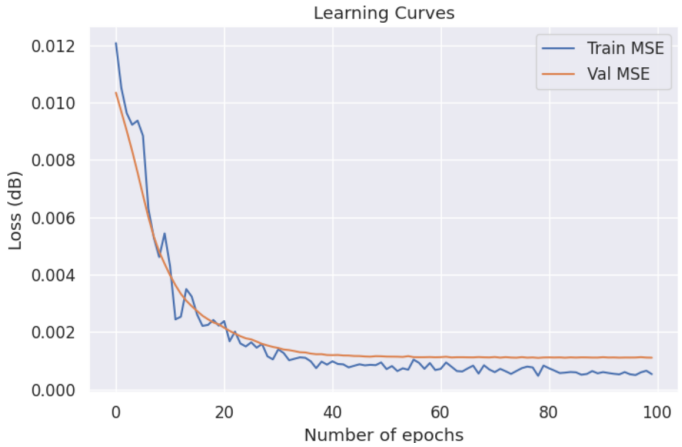

Figure 5.19. Learning curves for the training and validation set in the neural network of the span connecting ILA1 and ILA2

By comparing the results obtained in Figure 5.19 and the reference image in Figure 2.14, it can be stated that the model is neither overfitting nor underfitting. Both curves follow a similar behavior across the epochs and at the final epochs there are no signs for the validation MSE to rise its value. Also, it is noticed that the difference between both MSE is small.

Now that a normal behavior from the learning curves is verified, the results of the best model can be shown. These results will be divided in two parts, the first one showing just the loss profile prediciton results and the second one focusing on the whole OLS model performance.

5.5.1 Loss profile prediction performance

In Section 5.2.1, it was evident that GNPy by itself was not a reliable option to model the loss profile. Hence, ML was considered to adjust GNPy prediction. The best model tested was the Loss Profile S-GB, therefore in this subsection we will compare the results obtained with this model and those of Section 5.2.1.

The first thing noticed in Section 5.2.1 was not only that the values for the MAE

and MaxAE per channel were high, but also that the largest errors were present in the first and last channels forming a concave up parabola shape. Now, for the new Loss Profile S-GB model, it is evident in the MAE per channel figure (Figure 5.20), that the error decreased significantly and that the parabola shape is no longer a parabola as it is flatter. Furthermore, in the MaxAE figure (Figure 5.21), both GNPy and Loss Profile S-GB follow a same behavior; in any case the MaxAE plot did not have a prominent parabola as the MAE plot. Still, it can be seen for the MaxAE that Loss Profile S-GB model present smaller values per channel.

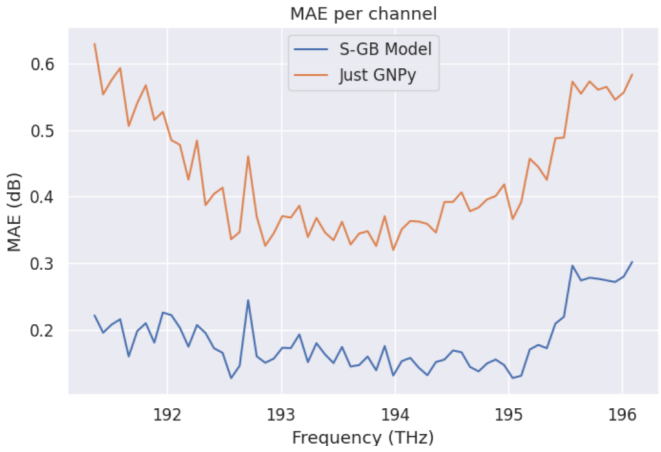

Figure 5.20. MAE per channel for the loss profile prediction of GNPy and the Loss Profile S-GB model

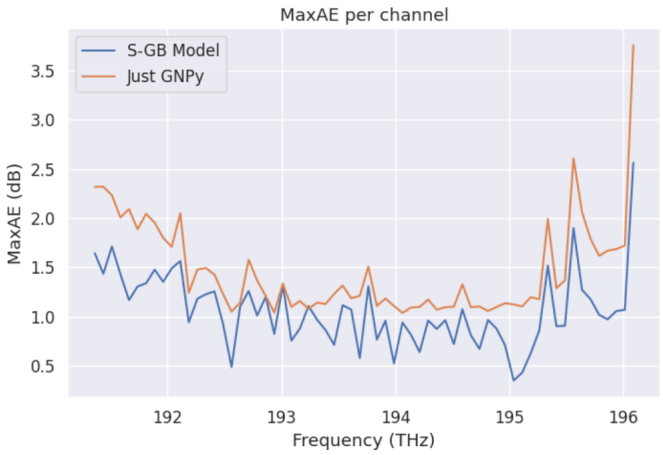

Figure 5.21. MaxAE per channel for the loss profile prediction of GNPy and the Loss Profile S-GB model

The better performance of the Loss Profile S-GB model is also evident if we plot the prediction for some of the configurations. These plots are found in Figure 5.22. It can be noticed that the Loss Profile S-GB is a better fit of the real loss profile, even following the shape and slope of the real loss curve, something that did not happen with just the GNPy tool. This explains why in Figure 5.20, the Loss Profile S-GB model does not present a parabola shape as the GNPy MAE behavior.

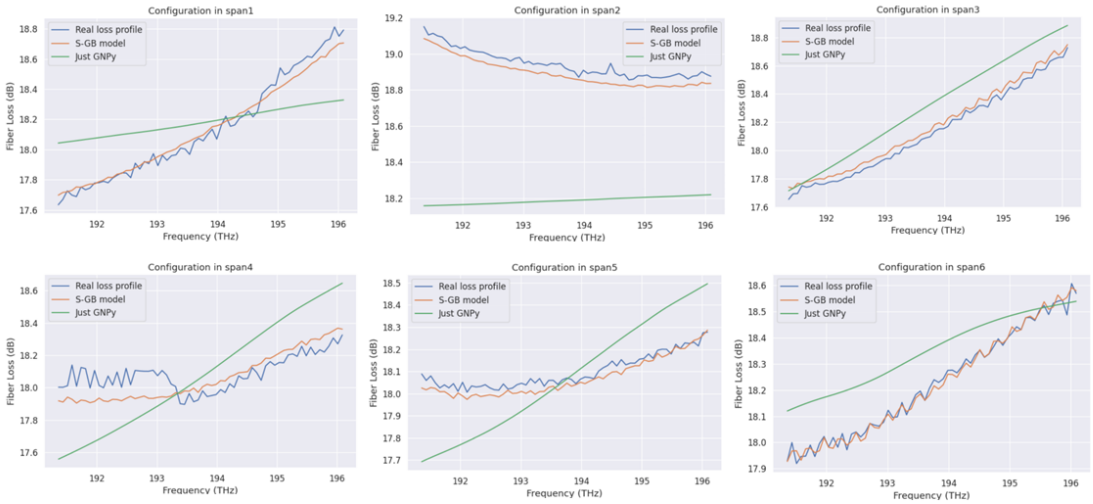

Figure 5.22. Loss Profile S-GB model, GNPy predicted and real loss profiles for some of the configurations in different spans

Following what was done in Section 5.2.1, an analysis per span of the MAE according to the spectral occupation was also performed considering the Loss Profile S-GB model. When this analysis was done for the GNPy prediction per span, it was found that some of the spans followed an expected behavior, in which the lowest error values were found as the full spectral load case was reached, but other spans don't and in general the MAE values tended to have a uniform behavior despite the percentage of occupancy. Therefore, it was complicated to find a relation between plots in Figure 5.12 and Figure 5.8. Now, for the Loss Profile S-GB model case in Figure 5.23. It can be seen that for most spans the lowest error values are found when most channels are set on, approaching the full spectral load case, but at the same time it is seen that in general all the MAE values per spectral occupation are low and their difference is not that much. This means that there is still a kind of uniform behavior across spectral occupation percentages and this can actually be considered a good thing as the error values are low and it means that the model works well for every spectral load configuration. In the next subsection, it will be verified if the same thing happens for the whole OLS model.

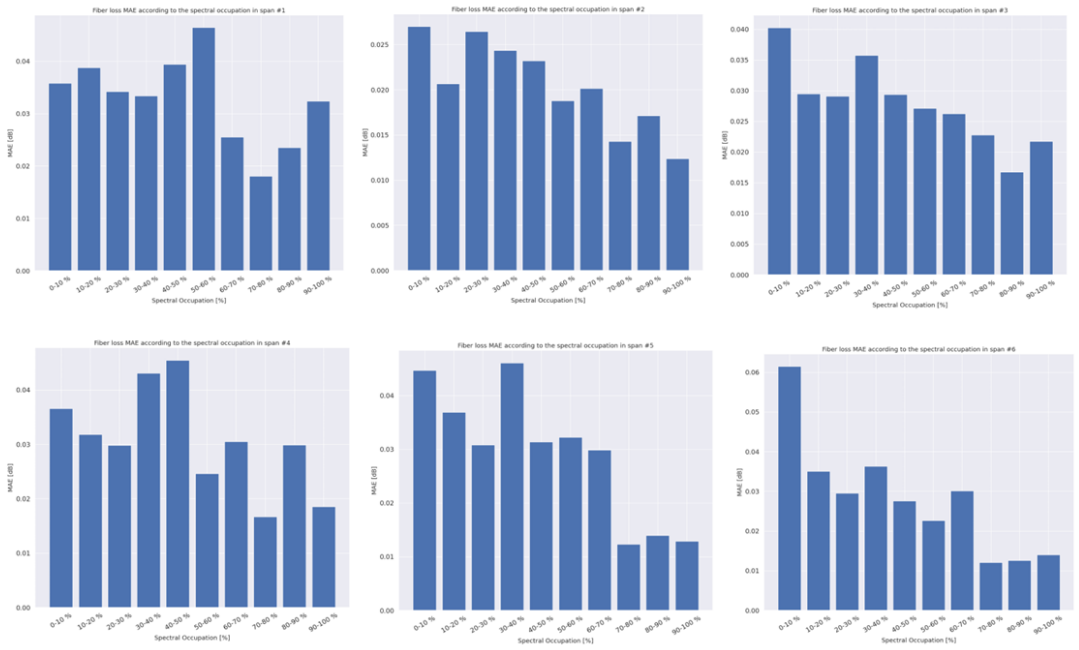

Figure 5.23. MAE according to the spectral occupation for the Loss Profile S-GB model per span

5.5.2 OLS model performance

With the observed improvement of the Loss Profile S-GB model in loss profile prediction over the use of the GNPy tool alone, the analysis of the complete OLS system model can now proceed. It is important to look back on the structure being tested; a picture depicting it is found in Figure 5.13. In this figure, the S-GB model, whose structure is found in Figure 4.10 but can still be seen in Figure 5.13, is used to model the EDFAs; while the fiber spans, modeled inside the "Loss Span Gray-box Model" box in the diagram, are modeled with the Loss Profile S-GB model, whose structure can be found in Figure 5.17.

Taking into account the best hyperparameter sets found in the previous sections and chapters, the results for the MAE and MaxAE are shown in Table 5.10. In order to have a visual reference, the plots for the OLS GNPy and OLS S-GB model predictions as well as the measured output powers corresponding to the configuration containing the worst MaxAE and the full spectral load case are presented, in Figures 5.24 and 5.25 respectively. Tables specifying the MAE and MaxAE of the models shown in these plots are found in 5.11 for the worst configuration case and 5.12 for the full spectral load case. The OLS Benchmark model was not included in the plots, as it mostly overlaps with the curves of other models, however its MAE an MaxAE values are being included in the tables, so a

comparison can be done. The performance according to the spectral occupation is shown in Figure 5.26.

Model	MAE	MaxAE
S-GB OLS	0.19 dB	2.56 dB

Table 5.10. Mean Absolute Error and Maximum Absolute Error of the S-GB OLS model

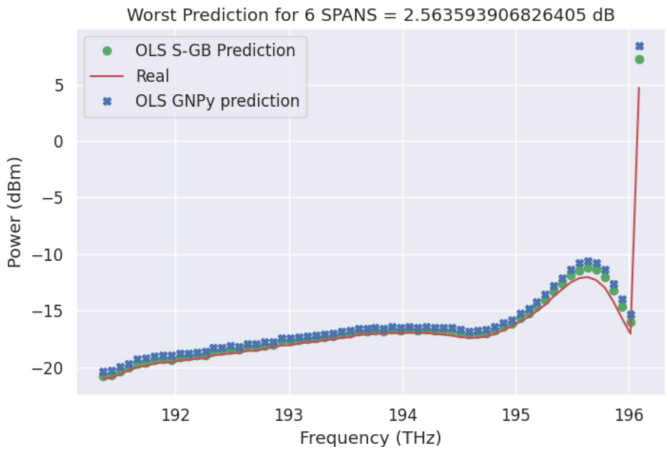

Figure 5.24. Worst predicted configuration in terms of MaxAE for the OLS S-GB model

Model	MAE	MaxAE
OLS Benchmark	0.24 dB	2.53 dB
OLS GNPy	0.79 dB	3.76 dB
OLS S-GB	0.43 dB	2.56 dB

Table 5.11. Mean Absolute Error and Maximum Absolute Error of the configuration shown in Fig. 5.24 for the OLS Benchmark, OLS GNPy model and OLS S-GB model

Model	MAE	MaxAE
OLS Benchmark	0.08 dB	0.29 dB
OLS GNPy	0.45 dB	1.03 dB
OLS S-GB	0.08 dB	0.19 dB

Table 5.12. Mean Absolute Error and Maximum Absolute Error of the configuration shown in Fig. 5.25 for the OLS Benchmark, OLS GNPy, and OLS S-GB model

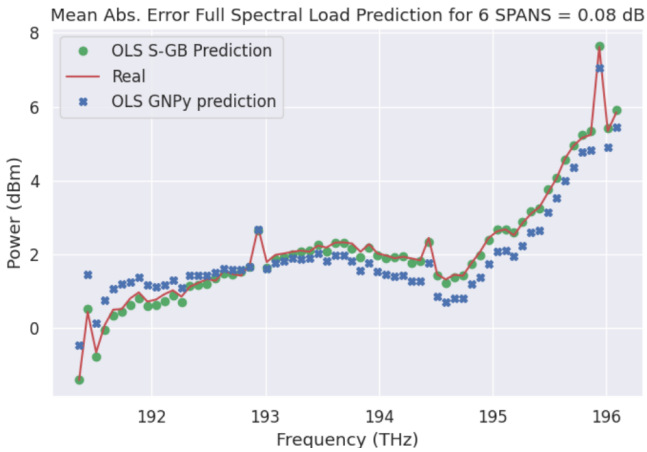

Figure 5.25. Prediction of the full spectral load configuration for the OLS S-GB model

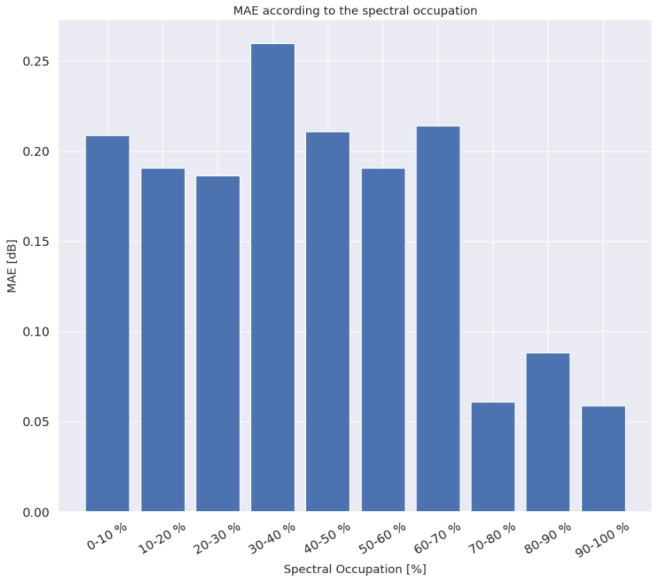

Figure 5.26. MAE according to the spectral occupation for the OLS S-GB model

From the overall results found in Table 5.10, it can be noticed that the improvement with respect to the OLS GNPy model (Table 5.2) is significant. In fact, it can be stated that the results are very close to the optimal ones as the difference with the results of the OLS Benchmark model, used as "ideal" reference and found in Table 5.1, is minimum.

This is also evident when looking at Figure 5.24 and Table 5.11. Even though visually

it seems there is not so much difference between the models' prediction; it is evident in the table that the OLS S-GB model performs better than the OLS GNPy one, as the first one has a difference with respect to the OLS Benchmark model of just 0.19 dB, while the latter has a difference of 0.36 dB in the MAE. It can still be seen that the worst configuration in terms of MaxAE is the one with just one channel turned on, making it clear again that low spectral occupation represents the biggest challenge for these models. The reason behind this is that it is difficult to model the noise profile of EDFAs. As mentioned in the introduction, uncertainties come from the ASE noise contributions. A good starting point to model this effect is needed to overcome this challenge. In Figure 5.25 and Table 5.12, the excellent performance of the OLS S-GB model is also noticeable. The model even reaches the MAE metric for the OLS Benchmark model and improves the MaxAE metric of the OLS Benchmark model. Stating this, it is also clear that the improvement with respect to the OLS GNPy model is noteworthy.

Finally, in Figure 5.26, the behavior that was normally observed throughout the thesis, except for the OLS GNPy analysis, that is low MAE value for the highly occupated spectral loads was found again. This can be explained by looking at Figure 5.23, in which for every span this behavior happened as well, despite having small MAE value differences among the spectral occupations. When doing this same analysis for the OLS GNPy model, the low MAE value for the highly occupated spectral loads was not so evident for some of the spans and this might have caused the unexpected behavior. Yet, due to the variety of behaviors found in the MAE analysis according to the spectral occupation per span in the OLS GNPy model (Figure 5.12), it was difficult to found a relation between these plots and the entire OLS MAE analysis according to the spectral occupation of the GNPy model (Figure 5.8).

Chapter 6

Conclusion

This thesis work demonstrates how local span-by-span gray-box models represent a good prospect to characterize OLS in optical networks. The results obtained are comparable with the results obtained for black-box approaches as in [11] and they show better performance compared to other gray-box approaches as in [13]. The final model that was presented seized good results by modeling each of the elements inside the OLS; thus, the characterization of the entire OLS does not lose flexibility or generalization. The accurate characterization achieved contributes in the reduction of uncertainties when estimating QoT metrics and moreover in the reduction of design margins. In this way, optical network design, planning and management will reduce its costs and use effectively its resources.

During this thesis work it was evident that despite the transparency and explainability of analytical models or other white-box approaches, they are not enough for a good characterization of the elements in the OLS. This happens due to the constraints the model might have. For example, in the particular case of the gain profile white-box model, accurate results were only obtained in configurations close to the full spectral load, as the formula was derived for this case; in any other case, this model presented a large error.

Despite the previous statement, it was proven that gray-box models, the joint use of ML along with white-box models or other tools that can provide more information on the elements that are being characterized like GNPy, are strongly beneficial in terms of model performance. Therefore, analytical models play a vital role not only because of their explainability but also because they can enhance ML black-box models.

In the context of this thesis, gray-box models performed better in configurations where there was a high spectral occupation. This might be a consequence of the gain profile white-box model derived from the full spectral load case. The only case where this did not happen was when GNPy was used to model the fiber loss profile. It was observed that the MAE values with respect to the spectral occupation per span coming from the GNPy prediction did not have a clear behavior and this could have caused the overall unexpected behavior. This shows that white-box models have a certain impact on the gray-box models in which they are being included.

Different structures were tested for the gray box model. From this it was observed

that for EDFAs, models that predict the difference between the measured power and the analytically predicted power perform better compared to those that do the same exercise with gain. Besides that, for both gain profile and loss profile prediction, the other variations tested did not present a significant difference on performance with the model that was taken as a reference. Anyway, the models that presented the best performances were part of the variations proposed. In particular, these models included as features the analytical prediction and the channel status. The inclusion of the analytical results as a feature of the ML black box validates also the second conclusion in which it was emphasized that the improvement of models come from the information provided by this white-box approaches.

This thesis highlights that local span-by-span gray-box models are a promising and effective approach for accurately characterizing OLS in optical networks, enabling reduced uncertainties, design margins, and overall network costs. Regarding further future improvements, it was mentioned during the thesis that the noise figure is not a constant parameter over the spectrum, furthermore it is also affected by gain and tilt. In this work a constant value for the noise figure was considered; therefore, integrating into the analysis a noise figure prediction can possibly enhance the results obtained. Apart from this, in this thesis work there was a focus on total power and therefore the GNPy tool was just considered to analyze the loss profiles. GNPy can be also useful to do a separate analysis on signal power, ASE noise power and NLI power and eventually analyzing directly the GSNR metric.

Bibliography

[1] G. P. Agrawal, *Fiber-Optic Communication Systems*. Hoboken, New Jersey: John Wiley and Sons, Inc., 4th ed., 2010.

[2] U. of Michigan, "Instrument encyclopedia, cultural heritage for community outreach," 2006.

[3] W. M. Levi, *The Pigeon*. Levi Publishing Co, Inc., 1977.

[4] R. E. Schuler, "Transportation and telecommunications networks: Planning urban infrastructure for the 21st century," *Urban Studies*, vol. 29, no. 2, pp. 297–310, 1992.

[5] M. Lewan, "The internet as an enabler of fintech," in *The Rise and Development of FinTech: Accounts of Disruption from Sweden and Beyond* (R. Teigland, S. Siri, A. Larsson, A. M. Puertas, and C. I. Bogusz, eds.), pp. 190–204, 2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN: Routledge, 2018.

[6] DataReportal, W. A. Social, and Meltwater, "Average daily time spent using the internet by online users worldwide from 3rd quarter 2015 to 3rd quarter 2024 (in hours.minutes) [graph]," tech. rep., Kepios, 2025.

[7] A. D'Amico, S. Straullu, G. Borraccini, E. London, S. Bottacchi, S. Piciaccia, A. Tanzi, A. Nespola, G. Galimberti, S. Swail, and V. Curri, "Enhancing lightpath qot computation with machine learning in partially disaggregated optical networks," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 564–574, 2021.

[8] P. Soumplis, K. Christodoulopoulos, M. Quagliotti, A. Pagano, and E. Varvarigos, "Network planning with actual margins," *Journal of Lightwave Technology*, vol. 35, no. 23, pp. 5105–5120, 2017.

[9] F. Musumeci, C. Rottondi, A. Nag, I. Macaluso, D. Zibar, M. Ruffini, and M. Tornatore, "An overview on application of machine learning techniques in optical networks," *IEEE Communications Surveys and Tutorials*, vol. 21, pp. 1383–1408, Secondquarter 2019.

[10] A. D'Amico, S. Straullu, A. Nespola, I. Khan, E. London, E. Virgillito, S. Piciaccia, A. Tanzi, G. Galimberti, and V. Curri, "Using machine learning in an open optical line system controller," *Journal of Optical Communications and Networking*, vol. 12, no. 6, pp. C1–C11, 2020.

[11] R. D'Ingillo, A. D'Amico, R. Ambrosone, E. Virgillito, V. Gatto, S. Straullu, F. Aquilino, and V. Curri, "Deep learning gain and tilt adaptive digital twin modeling of optical line systems for accurate osnr predictions," in *2024 International Conference on Optical Network Design and Modeling (ONDM)*, pp. 1–3, 2024.

[12] N. Morette, H. Hafermann, Y. Frignac, and Y. Pointurier, "Machine learning enhancement of a digital twin for wavelength division multiplexing network performance prediction leveraging quality of transmission parameter refinement," *Journal of Optical Communications and Networking*, vol. 15, no. 6, pp. 333–343, 2023.

[13] R. D'Ingillo, A. D'Amico, R. Ambrosone, S. Straullu, F. Aquilino, and V. Curri, "Enhancing optical multiplex section qot estimation using scalable gray-box dnn," in *IEEE Photonics Conference (IPC 2024)*, pp. 1–2, 2024.

[14] Y. Liu, X. Liu, Y. Zhang, M. Cai, M. Fu, X. Zhong, L. Yi, W. Hu, and Q. Zhuge, "Building a digital twin of an edfa for optical networks: a gray-box modeling approach," *Journal of Optical Communications and Networking*, vol. 15, no. 11, pp. 830–838, 2023.

[15] V. Curri, "Snr-based network abstraction." Extracted from slides in Open Optical Networks lessons, 2023. Politecnico di Torino.

[16] R. Sadeghi, B. Correia, A. Souza, N. Costa, J. Pedro, A. Napoli, and V. Curri, "Transparent vs translucent multi-band optical networking: Capacity and energy analyses," *Journal of Lightwave Technology*, vol. 40, no. 11, pp. 3486–3498, 2022.

[17] V. Curri, "Basic concepts on optical networking." Extracted from slides in Open Optical Networks lessons, 2023. Politecnico di Torino.

[18] Y. Yamamoto, Y. Kawaguchi, and M. Hirano, "Low-loss and low-nonlinearity pure-silica-core fiber for c- and l-band broadband transmission," *Journal of Lightwave Technology*, vol. 34, pp. 321–326, Jan 2016.

[19] "Understand fiber attenuation." Via Internet, June 2015. Extracted from http://fowiki.com/b/understand-fiber-attenuation/.

[20] U. Gliese, S. Norskov, and T. Nielsen, "Chromatic dispersion in fiber-optic microwave and millimeter-wave links," *IEEE Transactions on Microwave Theory and Techniques*, vol. 44, no. 10, pp. 1716–1724, 1996.

[21] V. Curri, P. Poggiolini, A. Carena, and F. Forghieri, "Dispersion compensation and mitigation of nonlinear effects in 111-gb/s wdm coherent pm-qpsk systems," *IEEE Photonics Technology Letters*, vol. 20, no. 17, pp. 1473–1475, 2008.

[22] K. Blow and D. Wood, "Theoretical description of transient stimulated raman scattering in optical fibers," *IEEE Journal of Quantum Electronics*, vol. 25, no. 12, pp. 2665–2673, 1989.

[23] P. Poggiolini, "Advanced topics." Extracted from slides in Optical Fiber Communications lessons, 2023. Politecnico di Torino.

[24] V. Curri, "Fiber propagation." Extracted from slides in Open Optical Networks lessons, 2023. Politecnico di Torino.

[25] S. Al-Dabagh and M. Abdalla, "Numerical solutions aspect of nonlinear schrodinger equation in monomode optical fiber," in *2006 2nd International Conference on Information and Communication Technologies*, vol. 2, pp. 2099–2103, 2006.

[26] 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*, vol. 32, no. 4, pp. 694–721, 2014.

[27] J. Bromage, "Raman amplification for fiber communications systems," *Journal of Lightwave Technology*, vol. 22, no. 1, pp. 79–93, 2004.

[28] R. De Matos, "High efficiency solar pumped laser through a ring array concentrator," July 2017.

[29] H. Haus, "The noise figure of optical amplifiers," *IEEE Photonics Technology Letters*, vol. 10, pp. 1602–1604, Nov 1998.

[30] V. Curri, "Real ols." Extracted from slides in Open Optical Networks lessons, 2023. Politecnico di Torino.

[31] G. Borraccini, V. Gatto, A. D'Amico, S. Straullu, F. Aquilino, A. Nespola, S. Piciaccia, A. Tanzi, G. Galimberti, and V. Curri, "Gain profile characterization and modeling for dual-stage edfa abstraction and control," *IEEE Photonics Technology Letters*, vol. 36, no. 2, pp. 107–110, 2024.

[32] *9 A Logical Calculus of the Ideas Immanent in Nervous Activity (1943)*, pp. 79–88. MIT Press, 2021.

[33] D. Valsesia, "Deep neural networks." Extracted from slides in Signal, image and video processing and learning lessons, 2022. Politecnico di Torino.

[34] Y. L. Cun, B. Boser, J. S. Denker, R. E. Howard, W. Habbard, L. D. Jackel, and D. Henderson, *Handwritten digit recognition with a back-propagation network*, pp. 396–404. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1990.

[35] A. Kumar, "Overfitting and underfitting in machine learning." via Internet, August 2024. Extracted from https://vitalflux.com/overfitting-underfitting-concepts-interview-questions/.

[36] Z. Wang, Y.-K. Huang, S. Han, T. Wang, D. Kilper, and T. Chen, "Multi-span optical power spectrum prediction using ml-based edfa models and cascaded learning," in *2024 Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, March 2024.

[37] S. Zhu, C. Gutterman, A. D. Montiel, J. Yu, M. Ruffini, G. Zussman, and D. Kilper, "Hybrid machine learning edfa model," in *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, March 2020.

[38] A. Castronovo, "Edfa noise figure and wss dgd modeling," Master's thesis, Politecnico di Torino, Corso Duca degli Abruzzi, 24, July 2024.

[39] G. Borraccini, Y.-K. Huang, A. D'Amico, E. Ip, T. Wang, and K. Asahi, "Optical line system physical digital model calibration using a differential algorithm," in *2025 Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, April 2025.

[40] R. Ambrosone, G. Borraccini, A. D'Amico, S. Straullu, F. Aquilino, D. Breuer, R. Schatzmayr, G. Grammel, and V. Curri, "Open line controller architecture in partially disaggregated optical networks," in *2023 International Conference on Photonics in Switching and Computing (PSC)*, pp. 1–3, Sep. 2023.