

**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA**

GUSTAVO ROSSI MARTINS

**USO DE REDES NEURAIS RECORRENTES PARA
MODELAGEM DE PROPAGAÇÃO NÃO LINEAR DE
PULSOS ÓPTICOS E GERAÇÃO DE SUPERCONTÍNUOS**

**VITÓRIA
2024**

GUSTAVO ROSSI MARTINS

**USO DE REDES NEURAIS RECORRENTES PARA
MODELAGEM DE PROPAGAÇÃO NÃO LINEAR DE
PULSOS ÓPTICOS E GERAÇÃO DE SUPERCONTÍNUOS**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica, na área de Telecomunicações e Tecnologia da Informação.

Orientador: Prof. Dr. Carlos Eduardo Schmidt Castellani

Coorientador: Prof. Dr. Helder Roberto de Oliveira Rocha

VITÓRIA
2024

Ficha catalográfica disponibilizada pelo Sistema Integrado de Bibliotecas - SIBI/UFES e elaborada pelo autor

M379u Martins, Gustavo Rossi, 1998-
Uso de redes neurais recorrentes para modelagem de propagação não linear de pulsos ópticos e geração de supercontínuos / Gustavo Rossi Martins. - 2024.
74 f. : il.

Orientador: Carlos Eduardo Schmidt Castellani.
Coorientador: Helder Roberto de Oliveira Rocha.
Dissertação (Mestrado em Engenharia Elétrica) -
Universidade Federal do Espírito Santo, Centro Tecnológico.

1. Óptica não-linear. 2. Fibras ópticas. 3. Equação de Schrodinger. 4. Solitons. 5. Inteligência artificial. 6. Redes neurais. I. Castellani, Carlos Eduardo Schmidt. II. Rocha, Helder Roberto de Oliveira. III. Universidade Federal do Espírito Santo. Centro Tecnológico. IV. Título.

CDU: 621.3

GUSTAVO ROSSI MARTINS

USO DE REDES NEURAIS RECORRENTES PARA MODELAGEM DE PROPAGAÇÃO NÃO LINEAR DE PULSOS ÓPTICOS E GERAÇÃO DE SUPERCONTÍNUOS

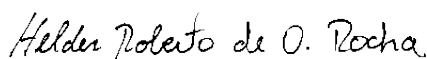
Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica, na área de Telecomunicações e Tecnologia da Informação.

Aprovado em 19 de abril de 2024.

COMISSÃO EXAMINADORA:



**Prof. Dr. Carlos Eduardo Schmidt
Castellani**
Universidade Federal do Espírito Santo
Orientador



**Prof. Dr. Helder Roberto de Oliveira
Rocha**
Universidade Federal do Espírito Santo
Coorientador



**Prof. Dr. Jorge Leonid Aching
Samatelo**
Universidade Federal do Espírito Santo
Examinador



Prof. Dr. Reginaldo Barbosa Nunes
Instituto Federal do Espírito Santo
Examinador

À minha mãe, Elisangela Rossi.

AGRADECIMENTOS

À minha família, minha avó Maria Lúcia pela sabedoria, minha irmã Iasmim pelo amor e carinho, e minha mãe Elisangela por tudo. Também aos amigos, Aiury Jureswski que encarou o desafio do mestrado comigo, Mariana Silveira pelas muitas conversas e caronas, e em especial a Eduarda Pedruzzi pela amizade e convívio (quase) diário.

Aos meus orientadores, Prof. Helder Rocha, sempre propondo uma nova rede neural ou algoritmo de otimização para testar, e Prof. Carlos Eduardo, pelas muitas conversas acompanhadas por café e teoria crítica, contribuindo na minha formação técnica e humana. Também à banca examinadora pelo tempo investido para leitura e avaliação desse trabalho.

Agradeço também aos amigos do LabTel, um espaço brasileiro-colombiano rico e alegre.

Agradeço por fim à Universidade Federal do Espírito Santo (UFES) pela minha formação, e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) pelo financiamento que me permitiu trabalhar nessa pesquisa.

*Sangrou pra ser humano
Mas secou sua lágrima
Orou pra ter engano
Mas não tinha dúvidas*

*E o homem amarelo
Que samba no sábado
Passou seu feriado
Sentindo ser último*

*Pedi pra ser segredo
Mas já era dívida
E ainda acreditava
Nesse abraço mórbido*

*E mais do que sua mágoa
Engoliu morte súbita
Deixou mulher, três filhos
E um mês pro próximo*

*Fez crescer seu valor
Em todo canto sórdido
Vendeu toda sua cópia
Como fosse o único*

*Esperança que o mundo
Não fosse tão cínico
Apagou suas memórias
Num sorriso cético*

*Vendou todo seu povo
Com mentira múltipla
Escreveu seu futuro
Em parcela de ácido*

*Condicionou seu tempo
Teorizou prática
Esculachou família
Ignorou seu naufrago*

*Queimou busão no centro
E engarrafou o tráfego
Queimaram o vacilão
Que renovou o tráfico*

*Embriagou-se a noite
Encarou dias lúcido
E prometeu aos céus
Que ainda seria próspero*

*Desconstrução
Rap dos Novos Bandidos
Sant*

RESUMO

As fibras óticas formam o *backbone* do sistema de telecomunicações moderno, interligando países e continentes via extensos cabos submarinos. Mesmo com uma atenuação de apenas 0,2 dB/km, esse meio de comunicação não é tão transparente ou passivo, e seus efeitos de dispersão e auto-modulação de fase podem ser destrutivos para os sistemas de comunicação. Tais efeitos são demasiado complexos e exigem simulações computacionais sofisticadas baseadas na equação não linear de Schrödinger (NLSE). O algoritmo estado da arte *split-step* Fourier é uma solução simples para avaliar a propagação de pulsos óticos, contudo, esse método pode ser um gargalo para experimentos em tempo real, casos muito não lineares, ou com grande número de simulações numéricas. O presente trabalho faz uso de redes neurais recorrentes para prever a evolução temporal e espectral no caso de propagação não linear de pulsos óticos, dispensando o uso de soluções numéricas no modelo.

Devido ao caráter temporal do problema de propagação, faz sentido o uso de redes recorrentes como a *long short-term memory* (LSTM). Na primeira etapa do trabalho foram testadas duas redes, a LSTM e a *convolutional* LSTM (CLSTM). Uma gama de hiperparâmetros foi testado empiricamente em busca da configuração que melhor captura as transformações óticas. Ambas redes apresentaram boa performance, com $R^2 > 96\%$, no entanto, uma análise mais profunda, estendendo as condições iniciais, e a forma de onda do pulso, evidenciou a dificuldade da CLSTM em se adaptar, destacando a LSTM, com erro máximo de $RMSE = 5,328 \times 10^{-3}$.

Na segunda etapa, buscamos generalizar a aplicação das redes também para o domínio espectral. A rede bidirecional LSTM (BiLSTM) é usada, considerando sua estrutura bidirecional, permitindo aprender dependências temporais de longo prazo e capturar padrões não lineares complexos. O domínio temporal e espectral são tratados simultaneamente nesse caso. A BiLSTM apresentou excelente desempenho, com $R^2 > 97\%$ para ambos domínios, com $RMSE = 4,39 \times 10^{-3}$ no domínio temporal e com $RMSE = 1,21 \times 10^{-2}$ no domínio espectral, superior ao LSTM para as mesmas condições. Ainda, as mesmas redes foram aplicadas para um caso de geração de supercontínuo, mais complexo, e ambas redes apresentaram bons resultados. A BiLSTM, por exemplo, apresentou um $RMSE = 1,75 \times 10^{-2}$ no domínio temporal e $RMSE = 1,32 \times 10^{-2}$ no domínio espectral.

Este trabalho contribui para as aplicações de *machine learning* em fotônica não linear, auxiliando na construção de modelos neurais para lidar com dinâmicas presentes na propagação de pulsos óticos e geração de supercontínuos.

Palavras-chave: Propagação ótica. NLSE. *Split-step* Fourier. LSTM. CLSTM. BiLSTM.

ABSTRACT

Optical fibers have become the backbone of modern telecommunication systems, linking countries, and continents through long submarine cables. Even with an attenuation of only 0.2 dB/km, this communication channel is not so transparent or passive, since its self-phase modulation and dispersive effects can be destructive for the communication system. Such effects are highly complex and require computationally demanding simulations based on the nonlinear Schrödinger equation (NLSE). The split-step Fourier state-of-the-art algorithm is a simple solution to evaluate the propagation of optical pulses, however, this method creates a severe bottleneck for real-time experiments, highly nonlinear cases, or with a great number of numerical simulations. This project uses recurrent neural networks to predict temporal and spectral evolution in the case of nonlinear pulse propagation, bypassing the need for numerical solutions.

Due to the temporal characteristics of the propagation problem, it makes sense the use of recurrent networks such as the long short-term memory (LSTM). The first part of this work tested two networks, the LSTM and the convolutional LSTM (CLSTM). A range of hyperparameters was tested empirically in search of the best configuration that captures the optical transformations. Both networks presented good performance, with $R^2 > 96\%$, however, in a deep analysis, we extended the initial conditions, and the pulse waveform, showing the CLSTM difficulty to adapt itself, highlighting the LSTM, with a maximum error of $RMSE = 5.328 \times 10^{-3}$.

In the second part, we tried to generalize the network applications also to the spectral domain. The bidirectional LSTM (BiLSTM) network is used, considering its bidirectional structure, allowing to learn long-term temporal dependencies and capture complex non-linear patterns. The temporal and spectral domains are treated simultaneously here. The BiLSTM presented excellent performance, with $R^2 > 97\%$ to both domains, the $RMSE = 4.39 \times 10^{-3}$ for the temporal domain and $RMSE = 1.21 \times 10^{-2}$ for the spectral domain, superior to the LSTM in the same conditions. Still, the same networks were applied to a supercontinuum generation case, more complex, and both networks showed good results. The BiLSTM, for example, returned an $RMSE = 1.75 \times 10^{-2}$ for the temporal domain, and an $RMSE = 1.32 \times 10^{-2}$ for the spectral domain.

This work contributes to the applications of machine learning in nonlinear photonics, helping to build a neural model to get along with dynamics in optical pulse propagation, and supercontinuum generation.

Keywords: Optical propagation. NLSE. *Split-step* Fourier. LSTM. CLSTM. BiLSTM.

LISTA DE FIGURAS

Figura 1 – Fibra óptica	18
Figura 2 – Formatos dos pulsos	20
Figura 3 – Efeito do alargamento de pulso	22
Figura 4 – Efeito do alargamento espectral	24
Figura 5 – Efeito de auto-inclinação	25
Figura 6 – Efeito Raman de atraso temporal e deslocamento de frequência	26
Figura 7 – Esquemático do método split-step Fourier	27
Figura 8 – Neurônio <i>feed-forward</i>	30
Figura 9 – Exemplo de redes <i>feed-forward</i>	31
Figura 10 – Neurônio recorrente	31
Figura 11 – Módulo interno LSTM	32
Figura 12 – Bidirecional LSTM	34
Figura 13 – Tensor 3D dos dados	40
Figura 14 – Esquemático da rede LSTM	41
Figura 15 – Esquemático da rede CLSTM	43
Figura 16 – Esquemático da rede BiLSTM	44
Figura 17 – Evolução temporal do sóliton pela LSTM	47
Figura 18 – Intensidade temporal do sóliton pela LSTM em distâncias selecionadas	47
Figura 19 – Evolução temporal gaussiano pela LSTM	49
Figura 20 – Intensidade temporal gaussiano pela LSTM em distâncias selecionadas	49
Figura 21 – Propagação do erro das redes LSTM e CLSTM na fibra	50
Figura 22 – Evolução temporal do sóliton pela BiLSTM	51
Figura 23 – Intensidade temporal do sóliton pela BiLSTM em distâncias selecionadas	52
Figura 24 – Evolução espectral do sóliton pela BiLSTM	52
Figura 25 – Intensidade espectral do sóliton pela BiLSTM em distâncias selecionadas	52
Figura 26 – Propagação do erro das redes LSTM e BiLSTM na fibra	53
Figura 27 – Evolução temporal do supercontínuo pela BiLSTM	55
Figura 28 – Intensidade temporal do supercontínuo pela BiLSTM em distâncias selecionadas	55
Figura 29 – Evolução espectral do supercontínuo pela BiLSTM	56
Figura 30 – Intensidade espectral do supercontínuo pela BiLSTM em distâncias selecionadas	56
Figura 31 – Propagação do erro das redes LSTM e BiLSTM na PCF	57

LISTA DE TABELAS

Tabela 1 – Dados gerados	37
Tabela 2 – Dados normalizados	38
Tabela 3 – Visualizando problema de regressão nos dados	39
Tabela 4 – Hiperparâmetros da LSTM	41
Tabela 5 – Hiperparâmetros da CLSTM	42
Tabela 6 – Performance da LSTM <i>vs</i> CLSTM	47
Tabela 7 – Performance estendida da LSTM <i>vs</i> CLSTM para o sóliton	48
Tabela 8 – Performance da LSTM <i>vs</i> CLSTM para o pulso gaussiano	48
Tabela 9 – Performance da LSTM <i>vs</i> BiLSTM	51
Tabela 10 – Performance da LSTM <i>vs</i> BiLSTM para a geração de supercontínuo	54
Tabela 11 – Comparação de valores em dB.	73

LISTA DE ABREVIATURAS E SIGLAS

CNN	<i>Convolutional Neural Networks</i>
CLSTM	<i>Convolutional LSTM</i>
DL	<i>Deep Learning</i>
FFT	<i>Fast Fourier Transform</i>
FWHM	<i>Full Width at Half Maximum</i>
FWM	<i>Four-Wave Mixing</i>
GNLSE	<i>Generalized Nonlinear Schrödinger Equation</i>
GPU	<i>Graphics Processing Unit</i>
GVD	<i>Group-Velocity Dispersion</i>
LSTM	<i>Long Short-Term Memory</i>
ML	<i>Machine Learning</i>
MAE	<i>Mean Absolute Error</i>
MSE	<i>Mean Squared Error</i>
NLSE	<i>Nonlinear Schrödinger Equation</i>
R^2	Coefficiente de Determinação
RMSE	<i>Root Mean Squared Error</i>
RNA	Redes Neurais Artificiais
SC	Supercontínuo
SPM	<i>Self-Phase Modulation</i>
SSF	<i>Split-Step Fourier</i>
TOD	<i>Third-Order Dispersion</i>
UFES	Universidade Federal do Espírito Santo
WDM	<i>Wavelength-Division Multiplexing</i>
ZDW	<i>Zero-Dispersion Wavelength</i>

LISTA DE SÍMBOLOS

a	Raio do núcleo
b	Raio da casca
n_1	Índice de refração do núcleo
n_2	Índice de refração da casca
Φ	Ângulo de incidência
Φ_c	Ângulo crítico
$A(z, T)$	Amplitude do pulso (W)
α	Atenuação total da fibra ótica
β_2	Parâmetro de dispersão GVD (s^2m^{-1})
β_3	Parâmetro de dispersão TOD (s^3m^{-1})
γ	Parâmetro não linear ($W^{-1}m^{-1}$)
$U(z, T)$	Amplitude normalizada do pulso (W)
α_{dB}	Atenuação total da fibra ótica (dB/km)
P_0	Potência de pico do pulso incidente (W)
T_0	Largura do pulso incidente (s)
T_{FWHM}	Largura à meia altura do pulso (s)
N	Ordem do soliton
\mathcal{F}	Transformada direta de Fourier
\mathcal{F}^{-1}	Transformada inversa de Fourier
\odot	Produto Hadamard
\star	Operador de convolução
L_D	Comprimento da dispersão
L'_D	Comprimento da dispersão associado ao TOD
L_{NL}	Comprimento não linear

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Objetivos	17
1.2	Contribuições	17
2	REFERENCIAL TEÓRICO	18
2.1	Fibra óptica	18
2.2	Equação de Propagação de Pulso	19
2.2.1	Atenuação	20
2.2.2	Dispersão cromática	21
2.2.3	Auto-modulação de fase	23
2.2.4	Efeitos de alta ordem	24
2.2.5	Auto-inclinação	24
2.2.6	Espalhamento Raman	25
2.3	Método <i>Split-Step</i> Fourier	27
2.4	Redes Neurais Artificiais	29
2.4.1	Redes Neurais <i>Feed-Forward</i>	30
2.4.2	Redes Neurais Recorrentes	31
2.4.3	Redes Neurais Recorrentes LSTM	32
2.4.4	Redes Neurais Convolucionais LSTM	33
2.4.5	Redes Neurais Bidirecionais LSTM	34
3	METODOLOGIA E ETAPAS DE DESENVOLVIMENTO	35
3.1	Modelagem da Fibra	35
3.2	Descrição dos Dados	36
3.2.1	Normalização	38
3.2.2	Série Temporal	38
3.2.3	Dimensionamento dos dados	39
3.3	Construção das Redes Neurais	40
3.3.1	LSTM vs CLSTM	40
3.3.2	LSTM vs BiLSTM	43
3.3.3	Métricas de Desempenho	44
4	RESULTADOS	46
4.1	LSTM vs CLSTM	46
4.1.1	Propagação Temporal de Pulso Solitônico	46
4.1.2	Propagação Temporal de Pulso Gaussiano	48

4.1.3	Propagação do Erro	48
4.2	LSTM vs BiLSTM	50
4.2.1	Propagação de Pulso Solitônico	50
4.2.2	Propagação do Erro	51
4.2.3	Geração de Supercontínuo	53
4.2.4	Propagação do Erro	56
4.3	Tempo Computacional	57
5	CONCLUSÕES E PROJETOS FUTUROS	58
5.1	Projetos futuros	59
	REFERÊNCIAS	61
	APÊNDICES	68
	APÊNDICE A – EQUAÇÕES DE MAXWELL	69
	APÊNDICE B – EQUAÇÃO DE PROPAGAÇÃO NÃO LINEAR	71
	APÊNDICE C – DECIBEL	73

1 INTRODUÇÃO

As fibras ópticas formam o *backbone* do sistema de telecomunicações moderno, interligando países e continentes com maior qualidade e taxa de dados via extensos cabos submarinos (COFFEN-SMOUT; HERBERT, 2000). O primeiro cabo telefônico transatlântico, o TAT-1, surgiu em 1956 levando 36 circuitos de voz por meio de cabos coaxiais arcaicos. Em 1976, o TAT-6 dispunha de 5.200 circuitos, e o TAT-7 adicionava mais 4.000 novos circuitos em 1983. O primeiro da série a ser projetado para operar com fibras ópticas foi o TAT-8 em 1988, com apenas dois pares monomodos e capacidade de 280 Mbps (HECHT, 1999). O uso de amplificadores ópticos no TAT-12 elevou a taxa de *bits* para 5,3 Gbps em 1995, e incorporando a técnica de multiplexação por divisão de comprimento de onda (WDM, de *wavelength-division multiplexing*) resultou em cabos submarinos com capacidade > 1 Tbps a partir de 2001, a exemplo do TAT-14, FLAG e Appolo (AGRAWAL, 2021). Como imaginado por Hecht (1999), as fibras ópticas tornaram-se a principal tecnologia das telecomunicações.

Sistemas de comunicação quilométricos exigem baixas perdas na transmissão, e em 1965 a melhor fibra tinha uma atenuação de 1.000 decibéis por quilômetro. A marca de 20 dB/km foi alcançada no início da década de 1970, um avanço importante para tornar possível sua aplicação em grandes distâncias, papel dividido com redes satélites e guias de ondas metálicos, mas insuficiente para predominar sobre as duas tecnologias (HECHT, 1999). A vantagem veio logo em 1976, com novas técnicas de fabricação como deposição de vapor químico e aplicação de germânio como dopante do núcleo, alcançando uma atenuação total de apenas 0,2 dB/km para comprimentos de onda de $1,55 \mu\text{m}$ (CASTELLANI, 2013).

Mesmo com uma atenuação de 0,2 dB/km, esse meio de comunicação não é tão transparente ou passivo. O *design* do sistema leva em conta os efeitos dispersivos e não lineares cumulativos da fibra, controlados devido as longas distâncias envolvidas. A potência dos *lasers* na transmissão e o perfil de dispersão são otimizados para combater esses efeitos potencialmente destrutivos para a informação (AGRAWAL, 2021).

Os efeitos dispersivos e não lineares são demasiado complexos e exigem simulações computacionais baseadas na equação de Schrödinger não linear (NLSE, de *nonlinear Schrödinger equation*). O método mais comum para tratar a equação de propagação (equação de Schrödinger) é o algoritmo *split-step* Fourier (SSF) que, através de transformações rápidas de Fourier, direta e inversa, calcula, em um número de passos, a evolução do pulso óptico (ALESHAMS; ZARIFKAR; SHEIKHI, 2005; OLIARI et al., 2020). Contudo, esse método

pode ser um gargalo severo na otimização de experimentos e situações em tempo real devido ao número elevado de iterações, por exemplo na transmissão de sinais por centenas de quilômetros, ou na simulação de lasers, que naturalmente exige milhares de voltas em uma cavidade (FANG et al., 2023). O presente trabalho faz uso de redes neurais recorrentes para o caso de propagação não linear de pulsos ópticos e geração de supercontínuos (SC), dispensando o uso de modelagem numérica, buscando dessa forma uma solução em tempo hábil para o problema.

O campo de *machine learning* tem evoluído rapidamente e suas aplicações em controle de veículos autônomos, processamento de linguagem, *health-care*, visão computacional, entre outros, tem mostrado sua capacidade e adaptabilidade para uma variedade de problemas complexos (JORDAN; MITCHELL, 2015). Trabalhos fazendo o uso dessa poderosa ferramenta em diversas aplicações tem se destacado em fotônica (KOBTSEV, 2021; GENTY et al., 2021). Por exemplo, redes neurais tem sido aplicadas para problemas de caracterização de pulsos (ZAHAVY et al., 2018; KLEINERT et al., 2019), modelagem de pulsos (KOKHANOVSKIY et al., 2019), análise de instabilidades (NÄRHI et al., 2018; SALMELA et al., 2020), *mode locking* de lasers (BAUMEISTER; BRUNTON; KUTZ, 2018), e propagação em fibras multimodo (BORHANI et al., 2018), ou ainda problemas mais complexos de predição de dinâmica não linear (SALMELA et al., 2021; MARTINS et al., 2022), predição de propagação em fibras multimodo (TEĞİN et al., 2021), e análise de dinâmica caótica (AMIL; SORIANO; MASOLLER, 2019; PATHAK et al., 2018).

Aqui, o uso do algoritmo SSF tem como finalidade gerar o banco de dados usado na construção do modelo neural. Esses dados são obtidos simulando (com o SSF) uma fibra com características físicas mantidas inalteradas (sua atenuação e outras constantes explicadas no corpo do texto), e variando o perfil do pulso limitado pela transformada de Fourier (*transform-limited*) de entrada na fibra óptica – sua potência de pico, P_0 , e sua largura de pulso, T_0 ou T_{FWHM} . O interesse do problema não se resume à evolução temporal apenas, mas também à sua contraparte no espectro.

Devido ao caráter temporal do problema de propagação, faz sentido o uso de redes recorrentes como a *long short-term memory* (LSTM) criada por Hochreiter e Schmidhuber (1997). A busca de um modelo adequado ao problema de evolução temporal/espectral óptica não se limita a abordagem com a LSTM, e podemos estendê-la para redes recorrentes mais complexas como a *convolutional LSTM* (CLSTM), ou a *bidirectional LSTM* (BiLSTM).

Para avaliar a performance do modelo, fazemos uso de métricas como o erro quadrático médio (MSE, de *mean squared error*), sua raiz quadrada (RMSE, de *root mean squared error*), entre outros, além do tempo computacional dos modelos, como forma de comparação

dos dados obtidos do SSF com os valores preditos pelo modelo neural para quantificar sua aproximação com o estado da arte.

1.1 Objetivos

O objetivo desse trabalho é elaborar um modelo neural para o problema de propagação não linear de pulsos ópticos, com desempenho similar ao algoritmo estado da arte SSF empregado, e menor custo computacional.

- Obter modelos neurais recorrentes que capturem os efeitos da transmissão na fibra;
- Comparar os modelos recorrentes LSTM, CLSTM, e BiLSTM;
- Avaliar a performance dos modelos neurais em contraposição ao modelo numérico;
- Avaliar o custo computacional dos modelos neurais em contraposição ao modelo numérico.

1.2 Contribuições

As publicações produzidas ao longo desta dissertação são listadas abaixo:

MARTINS, G. R.; SILVA, L. C. B.; SEGATTO, M. E. V.; ROCHA, H. R. O.; CASTELLANI, C. E. S. Design and analysis of recurrent neural networks for ultrafast optical pulse nonlinear propagation. **Optics Letters**, v. 47, n. 21, p. 5489–5492, 2022. Disponível em: <<https://doi.org/10.1364/OL.472267>>.

PIZZAIA, J. P. L.; MARTINS, G. R.; ROCHA, H. R. O.; CASTELLANI, C. E. S.; GOMES, A. L. Modelling, experimental validation and analysis of a multiple stage PC-PMF Sagnac loop interferometer with spectral shape versatility using an optimized fitting algorithm. **Optical Fiber Technology**, v. 80, p. 103388, 2023. Disponível em: <<https://doi.org/10.1016/j.yofte.2023.103388>>.

MARTINS, G. R.; SILVA, L. C. B.; SEGATTO, M. E. V.; ROCHA, H. R. O.; CASTELLANI, C. E. S. Simultaneous temporal and spectral evolution of ultrafast optical pulse propagation using a single bidirectional LSTM network. **Optical Fiber Technology**, v. 82, p. 103636, 2024. Disponível em: <<https://doi.org/10.1016/j.yofte.2023.103636>>.

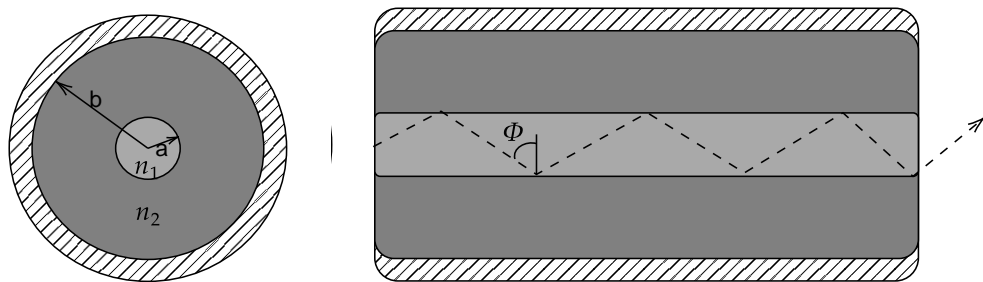
2 REFERENCIAL TEÓRICO

2.1 Fibra óptica

A fibra óptica é formada por um núcleo (ou *core*) cilíndrico vítreo de raio a e índice de refração n_1 envolto em uma casca (ou *cladding*) tubular com raio b e índice n_2 , como pode ser visto na Figura 1. Ambos núcleo e casca são constituídos de dióxido de silício (SiO_2), ou sílica, e são adicionados dopantes na sua estrutura para alterar os índices de refração de modo que $n_1 > n_2$. Dopantes como TiO_2 , GeO_2 , P_2O_5 , e Al_2O_3 aumentam o índice da sílica enquanto outros como o B_2O_3 e F o diminui (GAMBLING, 2000). O núcleo possui raio a de ordem micrométrica, tipicamente $25 \mu\text{m}$ para fibras multimodo e raio $a < 5 \mu\text{m}$ para fibras monomodo (AGRAWAL, 2019).

Apresentam características importantes como taxa de transmissão elevada, baixas perdas, imunidade a interferência eletromagnética, isolamento elétrico da sílica (dielétrico), robustez mecânica e, devido ao núcleo e casca micrométricos, são compactas (JAUREGUI; LIMPET; TÜNNERMANN, 2013).

Figura 1 – Fibra óptica



(a) Seção transversal (b) Seção longitudinal e reflexão interna total

Fonte: Próprio autor.

A relação entre os índices é fundamental para que ocorra a propagação de luz através da reflexão interna total mostrada na Figura 1. Para isso, o ângulo de incidência Φ deve ser maior que o ângulo crítico Φ_c descrito na Equação (2.1) da lei de Snell,

$$\Phi_c = \arcsin\left(\frac{n_2}{n_1}\right). \quad (2.1)$$

A reflexão interna total é útil na compreensão da fibra como um *pipeline* para a luz que se propaga por ela, mas o entendimento completo exige uma análise das equações de Maxwell em guia de ondas (Apêndice A).

2.2 Equação de Propagação de Pulso

A equação que governa a propagação de pulsos ópticos em fibra é conhecida como equação de Schrödinger não linear (NLSE) e é dada por (AGRAWAL, 2019)

$$i \frac{\partial A}{\partial z} = \frac{\beta_2}{2} \frac{\partial^2 A}{\partial T^2} + i \frac{\beta_3}{6} \frac{\partial^3 A}{\partial T^3} - i \frac{\alpha}{2} A - \gamma |A|^2 A, \quad (2.2)$$

onde A é o formato ou envelope do pulso óptico, β_2 e β_3 governam os efeitos de dispersão, α engloba as perdas na transmissão, e γ é o parâmetro das não linearidades da interação fibra-pulso. Ainda na equação, i é a unidade imaginária, z é a variável de comprimento do eixo longitudinal da fibra, e T é a medida de tempo em relação ao *frame* de referência do envelope óptico (Apêndice B).

Todas as constantes da equação acima estão relacionados a fibra óptica, sua fabricação e dopantes usados no núcleo ou casca (AGRAWAL, 2019). O termo A é referente ao envelope do pulso óptico, e a Equação (2.2) governa pulsos ultra curtos ($T_0 < 1$ ps). Por simplificação, consideramos apenas duas soluções recorrentes na literatura: pulso em formato gaussiano, ou solitônico. Além do seu formato, duas informações são importantes para descrever um pulso temporal: T_0 é sua largura inicial, e P_0 a potência de pico incidente. Ainda, as equações a seguir são normalizadas, onde $U(0, T)$ é o pulso normalizado, obtido por $A(0, T) = \sqrt{P_0} U(0, T)$.

A Equação (2.3) a seguir descreve um pulso gaussiano,

$$U(0, T) = \exp\left(-\frac{T^2}{2T_0^2}\right), \quad (2.3)$$

é comum caracterizar o pulso por sua largura à meia altura, T_{FWHM} (FWHM, de *full width at half maximum*), que se relaciona com T_0 pela equação abaixo no caso do pulso gaussiano,

$$T_{\text{FWHM}} = 2(\ln 2)^{1/2} T_0 \approx 1,665 T_0. \quad (2.4)$$

Outra equação de grande interesse é a Equação (2.5) da secante hiperbólica,

$$U(0, T) = N \operatorname{sech}\left(\frac{T}{T_0}\right), \quad (2.5)$$

onde N no contexto de sólitons é sua ordem e é dada por

$$N^2 = \frac{\gamma P_0 T_0^2}{|\beta_2|}, \quad (2.6)$$

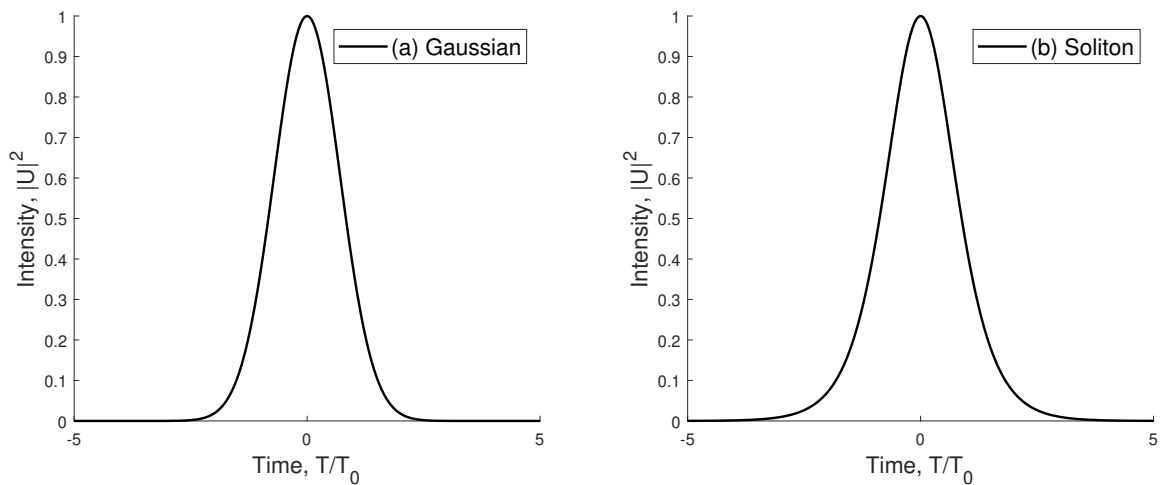
e T_{FWHM} se relaciona com T_0 através da Equação (2.7) abaixo,

$$T_{\text{FWHM}} = 2 \ln(1 + \sqrt{2}) T_0 \approx 1,763 T_0. \quad (2.7)$$

O interesse pelo pulso da forma de uma secante hiperbólica é por ele surgir naturalmente no contexto de sólitons ópticos, isto é, uma onda solitária que se propaga, em condições ideais, com formato constante. O sóliton de ordem $N = 1$ é chamado de fundamental, seu formato natural é o de uma *sech*. Idealmente, assumindo γ , P_0 e β_2 constantes, o sóliton fundamental propaga sem sofrer alterações. Para valores maiores de N , o sóliton de maior ordem apresenta mudanças periódicas temporal e espacialmente (CASTELLANI, 2013).

As equações descritas acima podem ser visualizadas na Figura 2 a seguir, os pulsos são normalizados e tem largura $T_0 = 1$.

Figura 2 – Formatos dos pulsos



Fonte: Próprio autor.

As seções 2.2.2 e 2.2.3 seguintes tratam dos efeitos de dispersão e auto-modulação de fase separadamente de forma a entender a natureza dos fenômenos que ocorrem na propagação de pulsos em fibra óptica, a ação conjunta desses efeitos lineares e não lineares não é tratado analiticamente neste capítulo, visto que o algoritmo SSF é adotado nesse caso.

2.2.1 Atenuação

A sílica usada na fabricação das fibras é um material transparente com grau de pureza elevado (GAMBLING, 2000). Ainda na década de 70, medições feitas por Kapron, Keck e Maurer (1970) mencionam fibras com 20 decibéis por quilômetro. Avanços nas técnicas empregadas permitiram reduzir ainda mais essas perdas, a maior parte devido a absorção ultravioleta em comprimentos de onda mais curtos e por íons de OH em comprimentos de onda maiores (KECK; MAURER; SCHULTZ, 1973). Nas região de 0,5 a 2 μm outro fator que contribui para a atenuação é o espalhamento Rayleigh, que devido a flutuações

aleatórias da densidade da sílica, e do seu índice de refração, é intrínseca ao processo de fabricação, da temperatura e tensões exercidas no processo (AGRAWAL, 2019).

A perda provocada por espalhamento Rayleigh pode ser estimada (em dB/km) pela equação seguinte (AGRAWAL, 2019)

$$\alpha_R = \frac{C_R}{\lambda^4}, \quad (2.8)$$

onde λ é o comprimento de onda, e C_R depende dos componentes da fibra e varia nos valores entre 0,7 e 0,9 dB/(km- μm^4). Pela Equação (2.8), fica evidente que o espalhamento Rayleigh torna inviável o uso de comprimentos de onda abaixo de 0,5 μm .

Atualmente, em fibras comerciais a atenuação para o comprimento de 1,55 μm é cerca de 0,2 dB/km, o que significa dizer que um sinal óptico teria 1% da sua potência inicial após propagar-se por 100 km (CASTELLANI, 2013). Para a melhor fibra dos anos 70, seria uma distância de apenas 1 km. A esse avanço tecnológico, foi inevitável a substituição de cabos coaxiais e sistemas satélites para telecomunicações, que além das perdas, esse último introduz atrasos na comunicação (GAMBLING, 2000).

Se P_0 é a potência na entrada da fibra de comprimento L , a potência na saída P_T é dada por (AGRAWAL, 2019)

$$P_T = P_0 \exp(-\alpha L), \quad (2.9)$$

onde α é a constante de atenuação total. Normalmente α é dado em dB/km pela relação

$$\alpha = -\frac{1}{L} \ln \left(\frac{P_T}{P_0} \right), \quad (2.10)$$

$$\alpha_{\text{dB}} = -\frac{10}{L} \log \left(\frac{P_T}{P_0} \right), \quad (2.11)$$

e a razão de α e α_{dB} , tomando $R = P_T/P_0$, é

$$\frac{\alpha_{\text{dB}}}{\alpha} = 10 \frac{\log R}{\ln R} = 10 \log e = 4,343. \quad (2.12)$$

Nas informações comerciais das fibras, é comumente usado α_{dB} , mas nos cálculos matemáticos é necessário fazer a conversão $\alpha_{\text{dB}} = 4,343\alpha$ (Apêndice C).

2.2.2 Dispersão cromática

A dispersão cromática, ou somente dispersão, é um fenômeno linear que surge da interação da onda eletromagnética com a fibra, e se manifesta através da dependência em frequência

do índice de refração $n(\omega)$. O índice de refração determina a velocidade da onda no canal de comunicação como $v = c/n$. Em razão da dispersão, um sinal óptico tem seus diferentes componentes espectrais viajando cada um em uma velocidade (AGRAWAL, 2019).

Matematicamente, os efeitos da dispersão são dados pelos componentes da constante de propagação β escrita em uma série de Taylor (AGRAWAL, 2019):

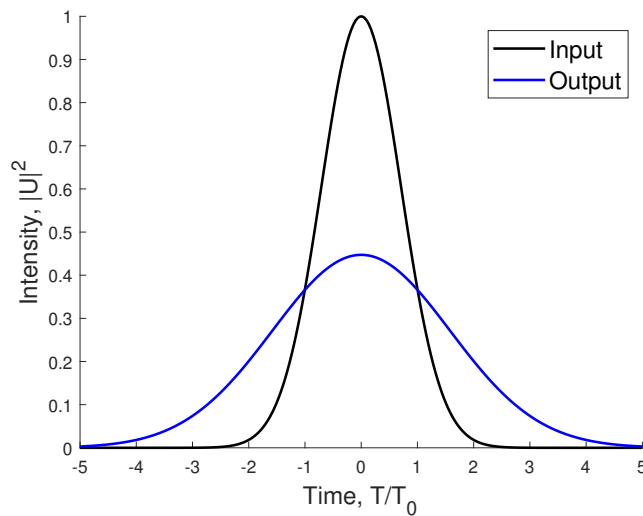
$$\beta(\omega) = n(\omega)\frac{\omega}{c} = \beta_0 + \beta_1(\omega - \omega_0) + \frac{1}{2}\beta_2(\omega - \omega_0)^2 + \dots, \quad (2.13)$$

onde

$$\beta_m = \left(\frac{d^m \beta}{d\omega^m} \right)_{\omega=\omega_0}. \quad (2.14)$$

Na equação de propagação (2.2) levamos em conta a participação de dois componentes da série de Taylor, β_2 e β_3 . O parâmetro β_2 é responsável pelo alargamento dos pulsos e é conhecido como parâmetro de dispersão de velocidade de grupo (GVD, de *group-velocity dispersion*). E o parâmetro β_3 , incluído para propagação de pulsos ultra curtos ($T_0 < 1$ ps), é chamado de parâmetro de dispersão de terceira ordem (TOD, de *third-order dispersion*) (AGRAWAL, 2019). Componentes de maior ordem são necessários em análises mais complexas como na geração de supercontínuos, métodos de caracterização baseados em misturas de quatro ondas (FWM, de *four-wave mixing*), ou quando ω se aproxima do comprimento de onda de dispersão zero, ω_{ZDW} (ZDW, de *zero-dispersion wavelength*) (DUBIETIS; COUAIRON; GENTY, 2019; GIL-MOLINA et al., 2020).

Figura 3 – Efeito do alargamento de pulso



Fonte: Próprio autor.

O efeito da dispersão cromática nos canais de comunicação é o alargamento do pulso mostrado na Figura 3. O alargamento ocorre devido a diferença de velocidade entre as

componentes de frequência. A taxa de variação temporal da fase $\phi(z, T)$ dá-se o nome de *chirp*. Quando o parâmetro GVD é positivo ($\beta_2 > 0$) o regime de propagação é normal e as componentes “vermelhas” do espectro viajam mais rápido que as componentes “azuis”. Quando o parâmetro GVD é negativo ($\beta_2 < 0$) o regime de propagação é anômalo e as componentes “azuis” viajam mais rápido que as “vermelhas” (AGRAWAL, 2019).

2.2.3 Auto-modulação de fase

Esta seção trata da auto-modulação de fase (SPM, de *self-phase modulation*) que se manifesta na propagação de pulsos de potência elevada e produz uma alteração na sua fase, e consequente alargamento espectral desses pulsos (AGRAWAL, 2019). No caso da SPM, o pulso altera seu espectro devido a sua própria intensidade; outro efeito não linear importante é a modulação cruzada de fase (XPM, de *cross-phase modulation*) que ocorre devido a intensidade de um segundo pulso (SCHNEIDER, 2004).

Simplificando a equação de propagação de modo que tenhamos apenas o efeito do termo não linear γ , ficamos com

$$\frac{\partial A}{\partial z} = i\gamma|A|^2A. \quad (2.15)$$

A Equação (2.15) pode ser resolvida pela solução geral $A = V \exp(i\phi_{\text{NL}})$,

$$e^{i\phi_{\text{NL}}} \frac{\partial V}{\partial z} + iV e^{i\phi_{\text{NL}}} \frac{\partial \phi_{\text{NL}}}{\partial z} = i\gamma V^3 e^{i\phi_{\text{NL}}}, \quad (2.16)$$

e tomando então

$$\frac{\partial V}{\partial z} = 0; \quad \frac{\partial \phi_{\text{NL}}}{\partial z} = \gamma V^2. \quad (2.17)$$

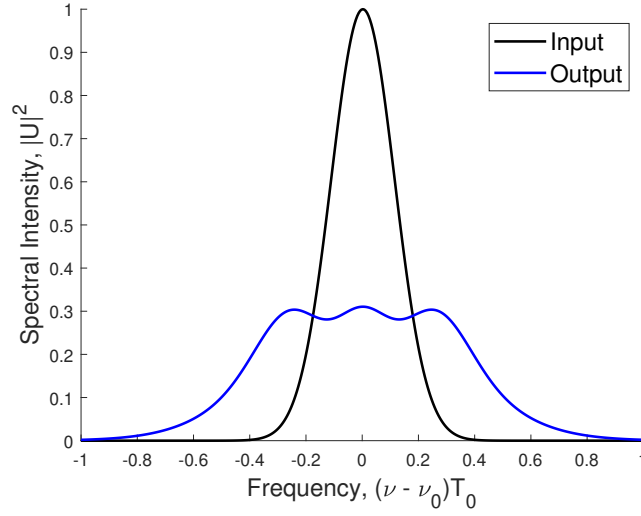
Como a derivada parcial de V em relação a z é nula, podemos afirmar que V é constante ao longo do comprimento da fibra, nesse caso a amplitude do pulso não altera durante a propagação,

$$V(T) = A(0, T); \quad \phi_{\text{NL}}(z, T) = \gamma|A(0, T)|^2 z. \quad (2.18)$$

E por isso $A(z, T) = A(0, T) \exp[i\phi_{\text{NL}}(z, T)]$.

Com as equações (2.18) demonstra-se que o SPM tem como consequência uma fase $\phi_{\text{NL}}(z, T)$ dependente da intensidade; efeito direto dessa fase é a variação no espectro do pulso. Na prática, não existe meio em que o SPM ocorra isoladamente, e sua interação com a dispersão pode dar origem a fenômenos solitônicos, geração de supercontínuos, e interferir no alargamento espectral a depender do regime dispersivo (ZHELTIKOV, 2018). A Figura 4 apresenta um exemplo do espectro de um pulso gaussiano após se propagar na fibra. Nesse caso, além de alargar a banda do pulso, o *chirp de frequência* também altera o formato do espectro.

Figura 4 – Efeito do alargamento espectral



Fonte: Próprio autor.

2.2.4 Efeitos de alta ordem

A equação de Schrödinger não linear deve ser alterada para os casos onde a propagação é mais complexa, como na geração de supercontínuos ou *optical rogue waves* (SOLLI et al., 2007), e então a equação deve ser generalizada como a seguir (AGRAWAL, 2019),

$$\frac{\partial A}{\partial z} + \frac{\alpha}{2}A + i\frac{\beta_2}{2}\frac{\partial^2 A}{\partial T^2} - \frac{\beta_3}{6}\frac{\partial^3 A}{\partial T^3} = i\gamma \left(|A|^2 A + \frac{i}{\omega_0} \frac{\partial}{\partial T} (|A|^2 A) - T_R \frac{\partial |A|^2}{\partial T} A \right), \quad (2.19)$$

a equação de Schrödinger não linear generalizada (GNLSE, de *generalized nonlinear Schrödinger equation*) é comumente escrita também da forma (DEITERDING et al., 2013)

$$\frac{\partial A}{\partial z} + \frac{\alpha}{2}A + \left(\sum_{k \geq 2} \beta_k \frac{i^{k-1}}{k!} \frac{\partial^k}{\partial T^k} \right) = i\gamma \left(1 + iS \frac{\partial}{\partial T} \right) \times \left[A \int_{-\infty}^{\infty} R(T') |A(T - T')|^2 dT' \right], \quad (2.20)$$

onde S é o efeito de auto-inclinação, R é o espalhamento Raman intra-pulso e T_R é seu primeiro momento. Nas seções seguintes abordaremos brevemente os efeitos produzidos por esses dois termos.

2.2.5 Auto-inclinação

De modo similar a análise da auto-modulação realizada previamente, a auto-inclinação pode ser estudada simplificando a Equação (2.20), e tomando apenas o termo da auto-inclinação (AGRAWAL, 2019)

$$\frac{\partial A}{\partial z} + \gamma S \frac{\partial}{\partial T} (|A|^2 A) = i\gamma |A|^2 A, \quad (2.21)$$

A Equação (2.21) pode ser resolvida tomando a solução geral $A = V_S \exp(i\phi_S)$,

$$e^{i\phi_S} \frac{\partial V_S}{\partial z} + iV_S e^{i\phi_S} \frac{\partial \phi_S}{\partial z} + 3\gamma S e^{i\phi_S} V_S^2 \frac{\partial V_S}{\partial T} + i\gamma S V_S^3 e^{i\phi_S} \frac{\partial \phi_S}{\partial T} = i\gamma V_S^3 e^{i\phi_S}, \quad (2.22)$$

e sendo $V_S^2 = I$, temos as equações acopladas

$$\frac{\partial I}{\partial z} + 3\gamma S I \frac{\partial I}{\partial T} = 0, \quad (2.23)$$

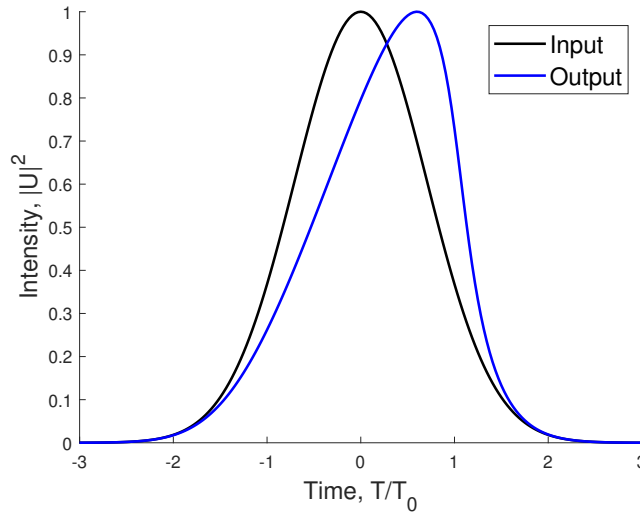
$$\frac{\partial \phi_S}{\partial z} + \gamma S I \frac{\partial \phi_S}{\partial T} = \gamma I. \quad (2.24)$$

A solução geral para a equação de intensidade (2.23) é dada por

$$I(z, T) = f(T - 3\gamma S I z), \quad (2.25)$$

de modo que cada ponto em T varia linearmente com inclinação dependente da intensidade. O efeito de auto-inclinação provocado pode ser observado na Figura 5. A partir de uma dada distância ocorre um choque óptico provocado pela inclinação (AGRAWAL, 2019). A análise pode ainda ser estendida considerando meios dispersivos (OLIVEIRA et al., 1992).

Figura 5 – Efeito de auto-inclinação



Fonte: Próprio autor.

2.2.6 Espalhamento Raman

O fenômeno do espalhamento Raman incluído na GNLSE possui duas características notáveis. Por um lado, o pulso óptico propagado sofre um atraso temporal devido a natureza não linear da resposta das moléculas a um campo óptico. Por outro, ocorre uma transferência de energia dos comprimentos de onda mais curtos para os comprimentos de onda mais longos, provocando um deslocamento para o vermelho característico do espalhamento Raman (AGRAWAL, 2019).

A resposta não linear $R(t)$ inclui as contribuições Kerr e Raman e é dada como

$$R(t) = (1 - f_R)\delta(t) + f_R h_R(t), \quad (2.26)$$

onde $f_R = 0,18$ representa a contribuição fracionária da resposta Raman atrasada para a polarização não linear, e $\delta(t)$ é a função delta de Dirac. A função de resposta Raman h_R é dada pelas vibrações moleculares da sílica e pode ser aproximado por (BLOW; WOOD, 1989),

$$h_R(t) = (\tau_1^{-2} + \tau_2^{-2})\tau_1 \exp(-t/\tau_2) \sin(t/\tau_1), \quad (2.27)$$

onde $\tau_1 = 12,2$ fs e $\tau_2 = 32$ fs. Aproximações modificadas da função de resposta Raman também foram propostas por Lin e Agrawal (2006) e Hollenbeck e Cantrell (2002).

Ainda, o termo T_R da Equação 2.19 se relaciona com a resposta Raman $R(t)$ através de

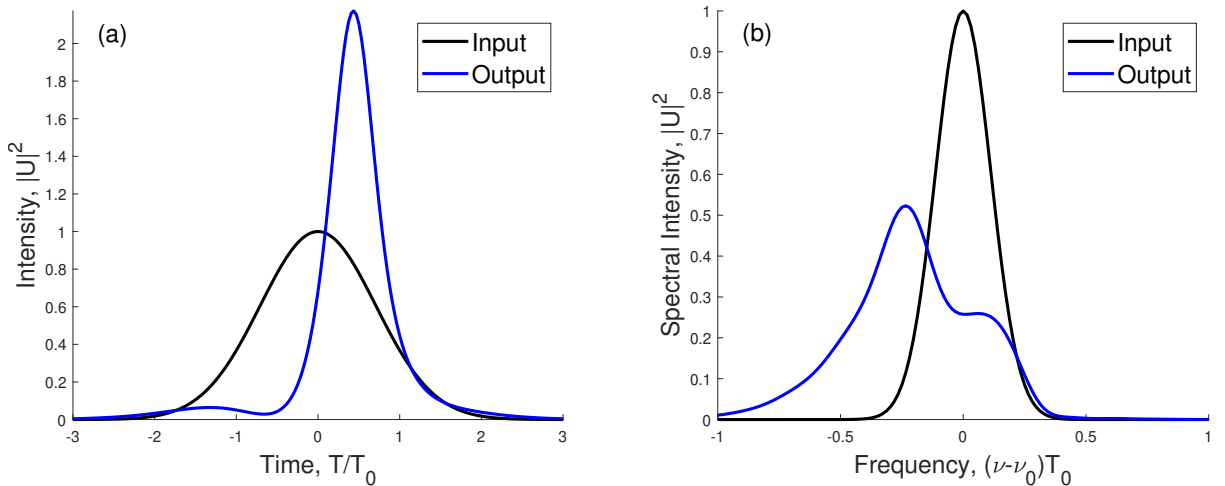
$$T_R \equiv \int_{-\infty}^{\infty} tR(t)dt \approx f_R \int_0^{\infty} t h_R(t)dt, \quad (2.28)$$

ou seja, T_R é definido como o primeiro momento da resposta Raman. E usando a Equação (2.27), temos (AGRAWAL, 2019)

$$T_R = f_R \tau_1 \tau_2^2 / (\tau_1^2 + \tau_2^2), \quad (2.29)$$

que para fibras de sílica possui valor próximo de 2 fs. O espalhamento Raman pode ser estudado na geração de supercontínuos, onde seus efeitos temporais e espectrais são bem aparentes (DUDLEY; TAYLOR, 2010). A Figura 6 mostra o atraso temporal (6a) e deslocamento para o vermelho (6b) causados pelo efeito Raman em um meio com regime de dispersão anômala.

Figura 6 – Efeito Raman de atraso temporal e deslocamento de frequência



Fonte: Próprio autor.

2.3 Método *Split-Step* Fourier

Equações diferenciais parciais não lineares como a NLSE geralmente não possuem soluções analíticas triviais e portanto exigem métodos numéricos sofisticados (SALMELA et al., 2021). Exemplos de sistemas físicos em que a NLSE é aplicável são sistemas hidrodinâmicos, plasmas, condensados de Bose–Einstein, e superfluidos (ROGEL-SALAZAR, 2013). O algoritmo estado da arte empregado neste caso é o método *split-step* Fourier (SSF) que se utiliza da transformada rápida de Fourier (FFT, de *fast Fourier transform*), conhecida por sua rapidez no cálculo das transformadas de Fourier de sinais discretos, para calcular rapidamente os efeitos da propagação óptica não linear (FRIGO; JOHNSON, 2005).

Para entender o método, reescrevemos a Equação (2.2) na forma

$$\frac{\partial A}{\partial z} = (\hat{D} + \hat{N})A, \quad (2.30)$$

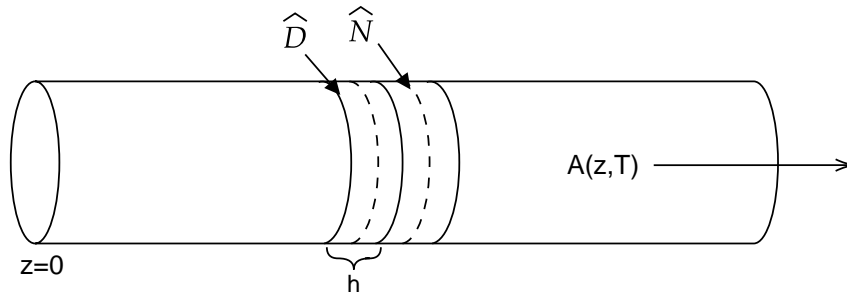
onde \hat{D} é o operador das dispersões e perdas lineares que acontecem na fibra e \hat{N} é o operador não linear da propagação. Esses operadores são dados então por

$$\hat{D} = -\frac{i\beta_2}{2} \frac{\partial^2}{\partial T^2} + \frac{\beta_3}{6} \frac{\partial^3}{\partial T^3} - \frac{\alpha}{2}, \quad (2.31)$$

$$\hat{N} = i\gamma|A|^2. \quad (2.32)$$

Para o método, assumimos que dispersão e não linearidade atuam de forma independente e uma de cada vez. Na Figura 7 apresentamos um esquemático do método. A linha sólida representa a atuação independente do efeito da dispersão (mais perdas), e a linha tracejada representa as não linearidades atuando sozinhas na fibra. Com isso, a propagação de z para $z + h$ acontece em dois passos, primeiro $\hat{N} = 0$, onde se avalia apenas os efeitos lineares, e depois $\hat{D} = 0$, onde se avalia apenas os efeitos não lineares (AGRAWAL, 2019).

Figura 7 – Esquemático do método split-step Fourier



Fonte: Próprio autor.

Então, a propagação $z + h$ é dada formalmente por (AGRAWAL, 2019)

$$A(z + h, T) = \exp[h(\hat{D} + \hat{N})]A(z, T), \quad (2.33)$$

que podemos aproximar na expressão abaixo usando a fórmula de Baker–Campbell–Hausdorff para dois operadores que não comutam ([ACHILLES; BONFIGLIOLI, 2012](#))

$$A(z + h, T) \approx \exp(h\hat{D}) \exp(h\hat{N})A(z, T). \quad (2.34)$$

O operador exponencial $\exp(h\hat{D})$ pode ser tratado no domínio de Fourier através de

$$A(z + h, T) \approx \mathcal{F}^{-1} \left\{ \exp[h\hat{D}(-i\omega)] \times \mathcal{F} \left[\exp(h\hat{N})A(z, T) \right] \right\}, \quad (2.35)$$

onde \mathcal{F} e \mathcal{F}^{-1} são as transformadas direta e inversa de Fourier, respectivamente, e $\hat{D}(-i\omega)$ é obtido substituindo o operador $\partial/\partial T$ por $-i\omega$ na Equação (2.31). Transformações similares são feitas no cálculo de $\exp(h\hat{N})A(z, T)$ para incluir efeitos não lineares de maior ordem ([DEITERDING et al., 2013](#)).

O método completo, aplicando as transformações rápidas de Fourier a cada passo resulta em ([AGRAWAL, 2019](#))

$$A(L, T) \approx \left(\prod_{m=1}^M e^{h\hat{D}} e^{h\hat{N}} \right) A(0, T), \quad (2.36)$$

onde L é o comprimento da fibra, M o número total de passos calculados, e $A(0, T)$ e $A(L, T)$ o pulso óptico na entrada e saída da fibra, respectivamente.

O algoritmo SSF pode ser reescrito de forma a aumentar a eficiência computacional – tornando-o mais complexo ([AGRAWAL, 2019](#)). No Algoritmo 1 tem-se o recorte de um código-exemplo com a parte essencial do método, nele a expressão (2.36) é alterada para

$$A(L, T) \approx e^{-h\hat{N}/2} \left(\prod_{m=1}^M e^{h\hat{N}} e^{h\hat{D}} \right) e^{h\hat{N}/2} A(0, T), \quad (2.37)$$

onde os operadores são computados na ordem $(h/2)\hat{N} \rightarrow h\hat{D} \rightarrow (h/2)\hat{N}$, com os efeitos não lineares ocorrendo na metade do passo.

Modelos numéricos para a solução da (G)NLSE foram amplamente implementados na literatura, incluindo os efeitos não lineares através de métodos eficientes de integração como o Runge–Kutta ([HULT, 2007; DUDLEY; TAYLOR, 2010; YCAS; MASER; HICKSTEIN, 2015; REDMAN et al., 2021; VOUMARD et al., 2023; BRAHMS; TRAVERS, 2023](#)).

Data:

- 1 $A = \text{pulso}(T/T0);$ ▷ Pulso óptico
- 2 $D = i(1/2)\beta_2\omega^2;$ ▷ Operador de dispersão
- 3 $N = i\gamma \text{abs}(A)^2;$ ▷ Operador não linear

Start:

- 4 $U = A \exp(N h/2);$
- 5 **for** $m = 1:M$ **do**
- 6 $U = \mathcal{F}^{-1}\{\exp(Dh) \times \mathcal{F}(U)\};$
- 7 $N = i\gamma \text{abs}(U)^2;$
- 8 $U = U \exp(Nh);$
- 9 **end**
- 10 $N = i\gamma \text{abs}(U)^2;$
- 11 $A = U \exp(-N h/2);$

Algoritmo 1: Algoritmo SSF.

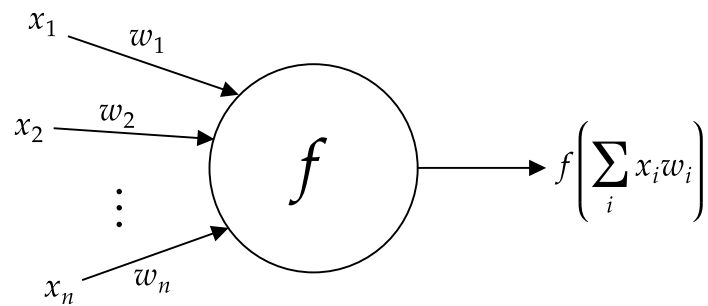
2.4 Redes Neurais Artificiais

As redes neurais artificiais (RNA), ou somente redes neurais, são técnicas de reconhecimento de padrões baseadas no processamento do cérebro humano. Os neurônios são capazes de transmitir, armazenar, e processar informação de maneira complexa e não totalmente compreendida. E sua capacidade de *aprender* atrai diversos pesquisadores (KROGH, 2008).

Primeiro, vale diferenciar dois conceitos comuns na literatura: *machine* e *deep learning*. *Machine learning* (ML) é um modelo de aprendizagem por treinamento capaz de associar padrões baseado em dados; *deep learning* (DL) é um conceito de aprendizagem de máquinas baseado em redes neurais artificiais (JANIESCH; ZSCHECH; HEINRICH, 2021).

O modelo mais simples de neurônio artificial, o perceptron de Rosenblatt (1958), demonstrou o potencial de aprendizagem do modelo. Nas RNAs, cada entrada x_i tem um peso w_i associado, e computa-se uma única saída a partir dessas entradas ponderadas ($\sum_i w_i x_i$) a partir das funções de ativação f contidas nos neurônios, a Figura 8 exemplifica isto. A saída segue para os neurônios seguintes, se houver, de acordo com a topologia da rede.

Durante o treinamento da rede, os pesos w_i são atualizados automaticamente, e o conhecimento da rede é armazenado nesses pesos. O aprendizado pode ser dividido entre: supervisionado ou não supervisionado (JANIESCH; ZSCHECH; HEINRICH, 2021).

Figura 8 – Neurônio *feed-forward*

Fonte: Próprio autor.

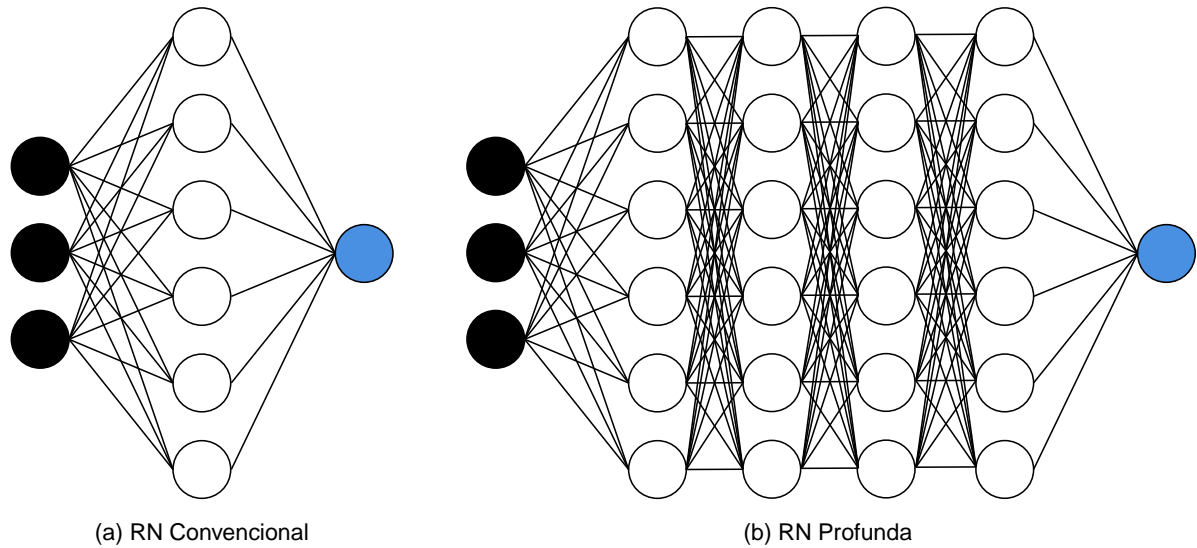
- Aprendizado supervisionado: baseado em conhecimento a priori, a rede neural recebe as entradas e as saídas esperadas (*targets*). O objetivo é ajustar os pesos e parâmetros da rede e adequar a saída prevista pela rede ao *target* (BISHOP, 2006).

A cada iteração, as saídas são comparadas e sua diferença é avaliada por uma função de perda. Os parâmetros do modelo são os pesos das conexões entre os neurônios, que são alterados gradativamente, e por isso, esse processo de aprendizagem é lento e exige um grande número de dados de treino. O algoritmo de aprendizado supervisionado mais famoso é o *backpropagation* (WERBOS, 1990; LILLICRAP et al., 2020);

- Aprendizado não supervisionado (*model-free*): Não há conhecimento a priori das respostas desejadas. A cada iteração, a rede busca identificar e agrupar os padrões dos dados de entrada. O mapa de Kohonen é um exemplo típico (KOHONEN, 2001).

2.4.1 Redes Neurais *Feed-Forward*

Modelo mais simples de redes neurais, as redes *feed-forward*, ou perceptron multicamada, são usadas principalmente para problemas de classificação e regressão (KROGH, 2008). Nela, as entradas são mapeadas nas saídas a partir das camadas intermediárias, que processam os dados passados diretamente (Figura 9). Destacam-se duas estruturas principais: a *shallow* e *deep network*. Com uma única camada intermediária, a primeira é facilmente interpretável e apresenta resultados superiores para problemas com poucos dados de treinamento e baixa dimensionalidade (JANIESCH; ZSCHECH; HEINRICH, 2021). Enquanto a rede neural profunda, composta de inúmeras camadas intermediárias, possui poder de processamento e preserva a simplicidade de construção da primeira – cada camada se conectando apenas na camada seguinte, sem *feedbacks*, por isso chamada de *feed-forward*.

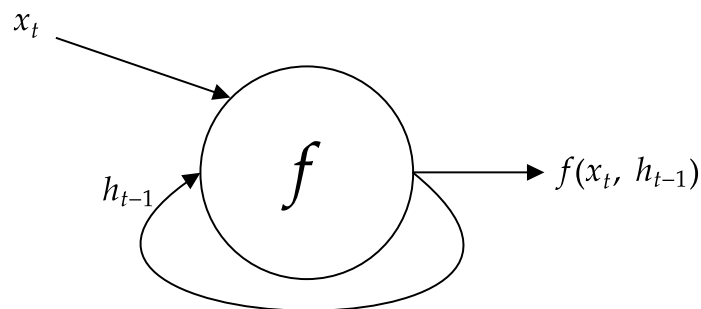
Figura 9 – Exemplo de redes *feed-forward*

Fonte: Próprio autor.

2.4.2 Redes Neurais Recorrentes

Diferentemente das redes *feed-forward*, onde cada camada se conecta apenas com a próxima, as redes neurais recorrentes (RNN, de *recurrent neural networks*) apresentam *loops* de realimentação conforme visto na Figura 10. Essa estrutura de realimentação permite armazenar informações passadas e a torna adequada para sequências temporais, e processamento de linguagem natural (JANIESCH; ZSCHECH; HEINRICH, 2021).

Figura 10 – Neurônio recorrente



Fonte: Próprio autor.

Nas RNNs, o estado oculto anterior \mathbf{h}_{t-1} e a entrada \mathbf{x}_t são usados no cálculo do estado oculto atual \mathbf{h}_t . Nesse caso, \mathbf{h}_t também é a saída do módulo, $\mathbf{h}_t = \tanh(\mathbf{x}_t, \mathbf{h}_{t-1})$. O módulo RNN padrão tem simplesmente uma função de ativação tangente hiperbólico, descrita pela Equação (2.38) seguinte (AGGARWAL, 2018)

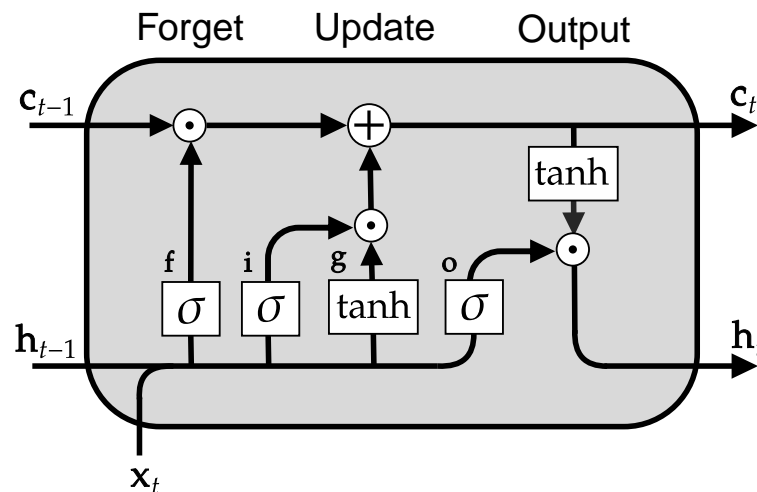
$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.38)$$

A estrutura simples da RNN dificulta a criação de memória de longo prazo, além do problema da fuga de gradiente (LECUN; BENGIO; HINTON, 2015).

2.4.3 Redes Neurais Recorrentes LSTM

A rede neural recorrente *long short-term memory* (LSTM) foi desenvolvida por Hochreiter e Schmidhuber (1997) com o intuito de mitigar o efeito de desaparecimento de gradiente, comum em problemas dinâmicos, com dependência temporal (LECUN; BENGIO; HINTON, 2015). A LSTM resolve isso com o uso de *portões* que atualizam os estados da rede.

Figura 11 – Módulo interno LSTM



Fonte: Próprio autor.

Na Figura 11, em ordem, são apresentados o portão de esquecimento (f), portão de entrada (i), célula candidata (g), e portão de saída (o). As entradas são dadas pelos vetores estado anterior da célula c_{t-1} , estado oculto anterior h_{t-1} , e entrada x_t . E os vetores saída são o estado atual da célula c_t , e o estado oculto atual h_t .

De maneira sucinta, a função dos blocos e o cálculo dos elementos são descritos a seguir:

O **portão de entrada** atualiza o estado da célula. O estado oculto anterior h_{t-1} e a entrada atual x_t são passados para os blocos g e i , seu resultado é dado pela tangente hiperbólica, $g_t = \tanh(h_{t-1}, x_t)$, e sigmoide, $i_t = \sigma(h_{t-1}, x_t)$. Para o novo estado, modifica-se o estado da célula por meio de $c_t = f_t \odot c_{t-1} + i_t \odot g_t$ (HOUDT; MOSQUERA; NÁPOLES, 2020);

O **portão de esquecimento** elimina informações. O estado oculto anterior h_{t-1} e a entrada atual x_t são passadas no portão f , seu resultado é dado pela sigmoide $f_t = \sigma(h_{t-1}, x_t)$, que

posteriormente afeta o estado da célula \mathbf{c}_t (HOUDT; MOSQUERA; NÁPOLES, 2020). A habilidade de resetar o estado da célula parece ser crítico no modelo (GREFF et al., 2017);

O **portão de saída** atualiza o estado oculto seguinte. O estado oculto anterior \mathbf{h}_{t-1} e a entrada atual \mathbf{x}_t são passadas no portão o , seu resultado é dado pela sigmoide $\mathbf{o}_t = \sigma(\mathbf{h}_{t-1}, \mathbf{x}_t)$. O estado oculto \mathbf{h}_t é obtido a partir do novo estado da célula e de \mathbf{o}_t , $\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t)$ (HOUDT; MOSQUERA; NÁPOLES, 2020).

Foram suprimidos os pesos e *biases* das funções acima. A função $\sigma(\mathbf{h}_{t-1}, \mathbf{x}_t)$ por exemplo, é calculado como $\sigma(\mathbf{W}_{x*}\mathbf{x}_t + \mathbf{R}_{h*}\mathbf{h}_{t-1} + \mathbf{b}_*)$, onde $\mathbf{W}_* \in \mathbb{R}^{N \times M}$, $\mathbf{R}_* \in \mathbb{R}^{N \times N}$, $\mathbf{b}_* \in \mathbb{R}^N$, para N blocos LSTM, e M entradas, e $*$ = (i, f, g, o). Similar para os demais cálculos. E \odot é o produto Hadamard (multiplicação elemento-a-elemento) dos vetores (GREFF et al., 2017). A função de ativação tangente hiperbólica é a mesma da Equação (2.38) e a função de ativação sigmoide é escrita na Equação (2.39) abaixo,

$$\sigma(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{1 + e^x}. \quad (2.39)$$

O resumo das equações do módulo LSTM encontra-se a seguir (GREFF et al., 2017):

$$\begin{aligned} \mathbf{i}_t &= \sigma(\mathbf{W}_{xi}\mathbf{x}_t + \mathbf{R}_{hi}\mathbf{h}_{t-1} + \mathbf{b}_i); & \mathbf{f}_t &= \sigma(\mathbf{W}_{xf}\mathbf{x}_t + \mathbf{R}_{hf}\mathbf{h}_{t-1} + \mathbf{b}_f); \\ \mathbf{g}_t &= \tanh(\mathbf{W}_{xg}\mathbf{x}_t + \mathbf{R}_{hg}\mathbf{h}_{t-1} + \mathbf{b}_g); & \mathbf{o}_t &= \sigma(\mathbf{W}_{xo}\mathbf{x}_t + \mathbf{R}_{ho}\mathbf{h}_{t-1} + \mathbf{b}_o); \\ \mathbf{c}_t &= \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \mathbf{g}_t; & \mathbf{h}_t &= \mathbf{o}_t \odot \tanh(\mathbf{c}_t). \end{aligned}$$

O estado da célula \mathbf{c}_t é alterado pelos portões f e i , e o estado oculto \mathbf{h}_t também por o .

2.4.4 Redes Neurais Convolucionais LSTM

Baseado nos avanços das redes recorrentes, em especial da LSTM, Shi et al. (2015) propõem uma nova rede convolucional LSTM (CLSTM). Outras redes associando camadas convolucionais com LSTM já haviam sido usadas (SAINATH et al., 2015; DONAHUE et al., 2017), a diferença aqui está na estrutura interna da CLSTM que integra as características de ambas camadas (convolução e recorrência), tornando mais simples a construção da rede e permitindo capturar as correlações espaço-temporais dos dados (SHI et al., 2015).

As equações que modelam a rede CLSTM são similares aquelas da LSTM, agora a entrada \mathbf{X}_t , os estados \mathbf{C}_{t-1} , \mathbf{C}_t , \mathbf{H}_{t-1} e \mathbf{H}_t , e portões \mathbf{I}_t , \mathbf{F}_t , \mathbf{G}_t , e \mathbf{O}_t , são tensores 3D, e a multiplicação matricial foi substituída pelo operador da convolução \star (SHI et al., 2015).

O resumo das equações do módulo CLSTM¹ encontra-se a seguir (SHI et al., 2015):

$$\begin{aligned}
 \mathbf{I}_t &= \sigma(\mathbf{W}_{xi} \star \mathbf{X}_t + \mathbf{R}_{hi} \star \mathbf{H}_{t-1} + \mathbf{b}_i) & \mathbf{F}_t &= \sigma(\mathbf{W}_{xf} \star \mathbf{X}_t + \mathbf{R}_{hf} \star \mathbf{H}_{t-1} + \mathbf{b}_f) \\
 \mathbf{G}_t &= \tanh(\mathbf{W}_{xg} \star \mathbf{X}_t + \mathbf{R}_{hg} \star \mathbf{H}_{t-1} + \mathbf{b}_g) & \mathbf{O}_t &= \sigma(\mathbf{W}_{xo} \star \mathbf{X}_t + \mathbf{R}_{ho} \star \mathbf{H}_{t-1} + \mathbf{b}_o) \\
 \mathbf{C}_t &= \mathbf{F}_t \odot \mathbf{C}_{t-1} + \mathbf{I}_t \odot \mathbf{G}_t & \mathbf{H}_t &= \mathbf{O}_t \odot \tanh(\mathbf{C}_t)
 \end{aligned}$$

A inclusão dessa rede como uma alternativa a LSTM permite uma comparação mais completa do problema de propagação, incluindo o cálculo da convolução.

2.4.5 Redes Neurais Bidirecionais LSTM

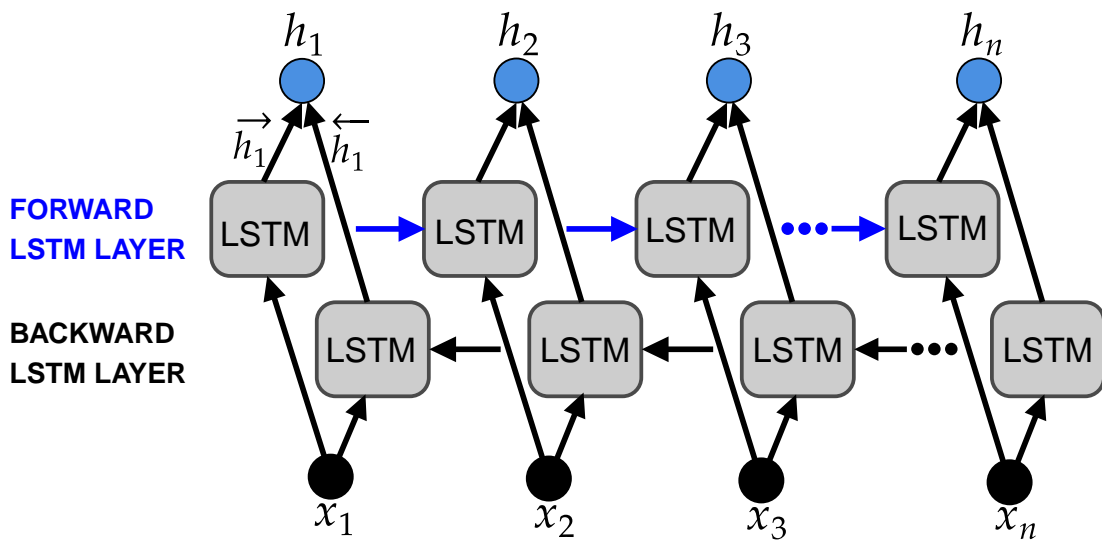
As redes recorrentes analisadas até aqui tratam os dados de modo unidirecional. Uma proposta é analisar os dados em duas direções (SCHUSTER; PALIWAL, 1997). A rede bidirecional LSTM (BiLSTM) possui duas camadas LSTM independentes, *forward* e *backward*, que processam os dados em ambas direções, do passado para o futuro, e do futuro para o passado (Figura 12) (GRAVES; SCHMIDHUBER, 2005; BIN et al., 2019).

Matematicamente, a estrutura da BiLSTM é dada simplesmente por (GRAVES; JAITLEY; MOHAMED, 2013)

$$\mathbf{y} = \mathbf{W}_{\vec{h}y} \vec{\mathbf{h}} + \mathbf{W}_{\overleftarrow{h}y} \overleftarrow{\mathbf{h}} + \mathbf{b}_y, \tag{2.40}$$

onde $\vec{\mathbf{h}}$ é a saída da camada *forward*, $\overleftarrow{\mathbf{h}}$ é a saída da *backward*, e \mathbf{y} é a saída da BiLSTM.

Figura 12 – Bidirecional LSTM



Fonte: Próprio autor.

¹ Equações seguem a formulação da *vanilla* LSTM.

3 METODOLOGIA E ETAPAS DE DESENVOLVIMENTO

3.1 Modelagem da Fibra

O algoritmo SSF foi desenvolvido no [MATLAB \(2022\)](#) em razão de sua interface simples e intuitiva, e pelo acesso na literatura de códigos numéricos já implementados para o *software* ([AGRAWAL, 2019](#)). Os parâmetros da simulação são adequados ao problema de propagação em uma fibra de 13 metros, altamente não linear. Seus coeficientes são dados por $\alpha_{dB} = 0,05 \text{ dB m}^{-1}$, $\beta_2 = -5,23 \times 10^{-27} \text{ s}^2\text{m}^{-1}$ (regime de dispersão anômalo), $\beta_3 = 4,27 \times 10^{-41} \text{ s}^3\text{m}^{-1}$ e $\gamma = 18,4 \times 10^{-3} \text{ W}^{-1}\text{m}^{-1}$, similares ao usado por [Salmela et al. \(2021\)](#) para realizar uma compressão de soliton de ordem superior. O pulso tem um perfil de secante hiperbólico (e gaussiano) com potência P_0 e largura de pulso T_0 variáveis.

Trata-se de um caso extremo e particular, com uma dinâmica não linear complexa, mas que pode ser encontrado no contexto de lasers pulsados, por exemplo ([FANG et al., 2023](#)). A escolha desse caso extremo está relacionado ao fato de serem mais difíceis de modelar e que, portanto, tornam o modelo numérico mais complexo.

Repetindo aqui a Equação (2.2) da propagação de um pulso óptico ultra curto na fibra,

$$i \frac{\partial A}{\partial z} = \frac{\beta_2}{2} \frac{\partial^2 A}{\partial T^2} + i \frac{\beta_3}{6} \frac{\partial^3 A}{\partial T^3} - i \frac{\alpha}{2} A - \gamma |A|^2 A,$$

como vamos tratar da propagação do pulso com amplitude normalizada $U(z, \tau)$, temos

$$A(z, \tau) = \sqrt{P_0} U(z, \tau), \quad (3.1)$$

então

$$i \frac{\partial U}{\partial z} = \frac{\hat{\beta}_2}{2} \frac{\partial^2 U}{\partial \tau^2} + i \frac{\hat{\beta}_3}{6} \frac{\partial^3 U}{\partial \tau^3} - i \frac{\alpha}{2} U - \hat{\gamma} |U|^2 U, \quad (3.2)$$

onde $\tau = T/T_0$,

$$\hat{\beta}_2 = \frac{\text{sgn}(\beta_2)}{L_D}, \quad \hat{\beta}_3 = \frac{\text{sgn}(\beta_3)}{L'_D}, \quad \hat{\gamma} = \frac{1}{L_{NL}}, \quad (3.3)$$

e $s = \text{sgn}(\beta_2) = \pm 1$ dependendo se a dispersão é normal ($s = 1$) ou anômala ($s = -1$), e $\hat{\beta}_2$, $\hat{\beta}_3$, e $\hat{\gamma}$ são os termos parametrizados pelos comprimentos abaixo

$$L_D = \frac{T_0^2}{|\beta_2|}, \quad L'_D = \frac{T_0^3}{|\beta_3|}, \quad L_{NL} = \frac{1}{\gamma P_0}. \quad (3.4)$$

Aqui os termos da Equação (3.4), comprimento da dispersão, L_D , comprimento da dispersão associada ao TOD, L'_D , e o comprimento não linear, L_{NL} nada mais são que escalas de

comprimento. Sua relação, representada pela ordem do sóliton $N^2 = L_D/L_{NL}$, permite estimar se o regime é dominado principalmente pela dispersão ($L_D/L_{NL} \ll 1$) ou pelas não linearidades ($L_D/L_{NL} \gg 1$) (MARTINS et al., 2022).

Por exemplo, a relação L_D/L_{NL} de um soliton com $P_0 = 26,3$ W e $T_{FWHM} = 1,1$ ps é

$$\frac{L_D}{L_{NL}} = \frac{\gamma P_0 T_0^2}{|\beta_2|} = 36,02 \gg 1,$$

ou, a relação L_D/L_{NL} de um soliton com $P_0 = 34,2$ W e $T_{FWHM} = 0,77$ ps é dada por

$$\frac{L_D}{L_{NL}} = \frac{\gamma P_0 T_0^2}{|\beta_2|} = 22,95 \gg 1.$$

Portanto, trata-se de um caso extremo e particular, com um regime predominantemente não linear. E não somente nos restringindo a propagação de pulsos, investigamos também o caso da geração de supercontínuo, modelado pela Equação (2.20), repetida aqui,

$$\frac{\partial A}{\partial z} + \frac{\alpha}{2}A + \left(\sum_{k \geq 2} \beta_k \frac{i^{k-1}}{k!} \frac{\partial^k}{\partial T^k} \right) = i\gamma \left(1 + iS \frac{\partial}{\partial T} \right) \times \left[A \int_{-\infty}^{\infty} R(T') |A(T - T')|^2 dT' \right].$$

Os parâmetros de simulação nesse caso descrevem uma fibra de cristal fotônico (PCF, de *photonic crystal fiber*) de 20 cm. A PCF é sem perda ($\alpha = 0$) e para o comprimento de onda de 810 nm está em um regime de dispersão anômala. Os coeficientes dispersivos empregados são $\beta_2 = -9,59 \times 10^{-27} \text{ s}^2\text{m}^{-1}$, $\beta_3 = 7,84 \times 10^{-41} \text{ s}^3\text{m}^{-1}$, $\beta_4 = -6,84 \times 10^{-56} \text{ s}^4\text{m}^{-1}$, $\beta_5 = -4,78 \times 10^{-70} \text{ s}^5\text{m}^{-1}$, $\beta_6 = 2,71 \times 10^{-84} \text{ s}^6\text{m}^{-1}$, $\beta_7 = -5,00 \times 10^{-99} \text{ s}^7\text{m}^{-1}$, e não linearidade $\gamma = 0,1 \text{ W}^{-1}\text{m}^{-1}$ (MARTINS et al., 2024). O pulso tem um perfil de secante hiperbólico com potência P_0 variável e largura de pulso T_0 fixa. Esse caso servirá para demonstrar a capacidade de generalização da rede neural empregada. Em se tratando de um caso óptico não linear mais complexo, será tratado a parte na Seção 4.2.3.

3.2 Descrição dos Dados

A evolução do pulso é sensível ao meio de propagação e as suas condições iniciais. A potência de pico na entrada da fibra P_0 varia no intervalo de 18,41 a 34,19 W, e a duração T_{FWHM} varia no intervalo de 0,77 a 1,43 ps (mapeando T_0 no intervalo de 0,44 a 0.81 ps), o que caracteriza um soliton de ordem N variando de 3,5 a 8,9 (SALMELA et al., 2021). Para o caso da geração de supercontínuo, a potência de pico inicial P_0 varia no intervalo de 0,5 a 2 KW e a largura de pulso é mantida fixa, $T_{FWHM} = 100$ fs (MARTINS et al., 2024).

Com um conjunto de 3.000 (40 *powers* \times 75 *widths*) condições iniciais distintas, o objetivo é gerar uma base de dados com o perfil temporal de cada passo dado dentro do algoritmo

SSF. Para isso, a fibra é dividida em 1.000 passos, e o pulso amostrado com 1.024 pontos. Para a geração de supercontínuo, caso não linear mais complexo, geramos um conjunto de 2.000 simulações (variando apenas a potência), em que a fibra é dividida em 20.000 passos, e o sinal é amostrado com 2.048 pontos. A descrição dos dados a seguir será feita para a evolução óptica solitônica. O algoritmo SSF necessita de um passo h pequeno para uma aproximação com erro pequeno. Os dados gerados pelo SSF serão usados para treinar as redes neurais, essas não tem restrição de distância propagada, e por isso, os dados serão sub-amostrados para reduzir o tempo de treinamento das redes. Por exemplo, escolhendo um número de 101 passos no interior da fibra ($h = 0,13$ m) e o pulso amostrado com 256 pontos, o resultado é uma matriz com $7,76 \times 10^7$ elementos ($101 \times 256 \times 3.000$). Um conjunto dessa magnitude favorece o aprendizado da rede neural, mas é custoso computacionalmente, e torna o processo de treinamento mais lento. Outra escolha é amostrar o pulso com 128 pontos, nesse caso o resultado é de $3,88 \times 10^7$ elementos ($101 \times 128 \times 3.000$), usaremos essa última.

Para o treinamento e teste das redes neurais, os dados são divididos em 2.900 simulações de treinamento/validação (1.900 para o SC), e cem simulações para o teste. A escolha dessa porcentagem para o conjunto de teste, 3,44% (5,26% para o SC), é condizente com a literatura, contudo, pode apresentar como fator de risco: representatividade inadequada, avaliação otimista, e sensibilidade a variações. O código SSF executa aquilo já explicado no Algoritmo 1 da Seção 2.3. O resultado é a matriz de propagação, onde cada linha representa o pulso em um passo da propagação. Os dados mostrados na Tabela 1 são de uma condição inicial apenas. O mesmo acontece para as demais condições.

Tabela 1 – Dados gerados

	0	1	2	3	...	124	125	126	127
0	2,7585	2,8886	3,0242	3,1655	...	3,3127	3,1655	3,0241	2,8885
1	2,9230	2,9532	3,0002	3,0990	...	3,2821	3,0930	2,9847	2,9488
2	2,9937	3,0184	3,0487	3,1204	...	3,2186	3,1061	3,0426	3,0115
3	3,0466	3,0661	3,0924	3,1524	...	3,2244	3,1360	3,0864	3,0594
4	3,0912	3,1079	3,1306	3,1830	...	3,2416	3,1686	3,1256	3,1000
...
96	2,3747	2,3804	2,3767	2,3763	...	2,4369	2,4093	2,3791	2,3687
97	2,3688	2,3700	2,3581	2,3657	...	2,4208	2,3874	2,3693	2,3647
98	2,3473	2,3450	2,3374	2,3417	...	2,4220	2,3968	2,3814	2,3634
99	2,3357	2,3312	2,3227	2,3254	...	2,4392	2,4165	2,3837	2,3532
100	2,3391	2,3192	2,3054	2,3139	...	2,4229	2,4185	2,4037	2,3729

101 linhas \times 128 colunas

Fonte: Próprio autor.

Com os dados gerados pelo SSF, devemos agora normalizá-los, transformar a série temporal em um problema de regressão, e por fim adequar a estrutura dos dados de acordo com a

entrada aceita pela rede neural.

3.2.1 Normalização

A normalização é um passo fundamental (e mesmo compulsório) para o treinamento de uma rede neural, pois acelera seu aprendizado (AGGARWAL, 2018). A normalização, descrita pela Equação (3.5), mapeia os dados gerados em um intervalo entre 0 (mínimo) e 1 (máximo),

$$\mathbf{X}_{\text{std}} = \frac{\mathbf{X} - X_{\min}}{X_{\max} - X_{\min}}, \quad (3.5)$$

onde \mathbf{X} é a matriz de dados, \mathbf{X}_{std} é a matriz de dados normalizados, $X_{\min} = \min(\mathbf{X})$, e $X_{\max} = \max(\mathbf{X})$. Os valores máximo e mínimo são tomados aqui do conjunto completo, antes da divisão entre dados em treinamento e testes, o que poderia introduzir informações de um conjunto em outro no processo de treinamento, superestimando assim a capacidade do modelo, processo conhecido como vazamento de dados (*data leakage*). Aplicando a normalização, obtemos os valores mapeados na Tabela 2.

Tabela 2 – Dados normalizados

	0	1	2	3	...	124	125	126	127
0	0,0071	0,0075	0,0078	0,0082	...	0,0086	0,0082	0,0078	0,0075
1	0,0076	0,0076	0,0078	0,0080	...	0,0085	0,0080	0,0077	0,0076
2	0,0078	0,0078	0,0079	0,0081	...	0,0083	0,0080	0,0079	0,0078
3	0,0079	0,0079	0,0080	0,0082	...	0,0084	0,0081	0,0080	0,0079
4	0,0080	0,0080	0,0081	0,0082	...	0,0084	0,0082	0,0081	0,0080
...
96	0,0061	0,0062	0,0061	0,0061	...	0,0063	0,0062	0,0062	0,0061
97	0,0061	0,0061	0,0061	0,0061	...	0,0063	0,0062	0,0061	0,0061
98	0,0061	0,0061	0,0060	0,0061	...	0,0063	0,0062	0,0062	0,0061
99	0,0060	0,0060	0,0060	0,0060	...	0,0063	0,0062	0,0062	0,0061
100	0,0060	0,0060	0,0060	0,0060	...	0,0063	0,0063	0,0062	0,0061

101 linhas \times 128 colunas

Fonte: Próprio autor.

3.2.2 Série Temporal

O passo seguinte é transformar o problema atual de uma série temporal, onde cada atributo (colunas) evolui no tempo (linhas) de maneira sequencial, em um problema de regressão, ou seja, dado uma sequência de observações, buscar inferir seu próximo valor.

Na Tabela 3 fica claro o que buscamos fazer, com três observações de entrada, usamos três observações (*box* vermelho) para prever uma quarta observação (*box* preto). A série temporal aumenta a quantidade de elementos da base de dados, de 101×128 para $4 \times (101 - 3) \times 128$ nesse caso, quase 4 vezes maior. A escolha do número de observações de entrada afeta diretamente no treinamento da rede.

Tabela 3 – Visualizando problema de regressão nos dados

	0	1	2	3	...	124	125	126	127
0	0,0071	0,0075	0,0078	0,0082	...	0,0086	0,0082	0,0078	0,0075
1	0,0076	0,0076	0,0078	0,0080	...	0,0085	0,0080	0,0077	0,0076
2	0,0078	0,0078	0,0079	0,0081	...	0,0083	0,0080	0,0079	0,0078
3	0,0079	0,0079	0,0080	0,0082	...	0,0084	0,0081	0,0080	0,0079
4	0,0080	0,0080	0,0081	0,0082	...	0,0084	0,0082	0,0081	0,0080
...
96	0,0061	0,0062	0,0061	0,0061	...	0,0063	0,0062	0,0062	0,0061
97	0,0061	0,0061	0,0061	0,0061	...	0,0063	0,0062	0,0061	0,0061
98	0,0061	0,0061	0,0060	0,0061	...	0,0063	0,0062	0,0062	0,0061
99	0,0060	0,0060	0,0060	0,0060	...	0,0063	0,0062	0,0062	0,0061
100	0,0060	0,0060	0,0060	0,0060	...	0,0063	0,0063	0,0062	0,0061

101 linhas \times 128 colunas

Fonte: Próprio autor.

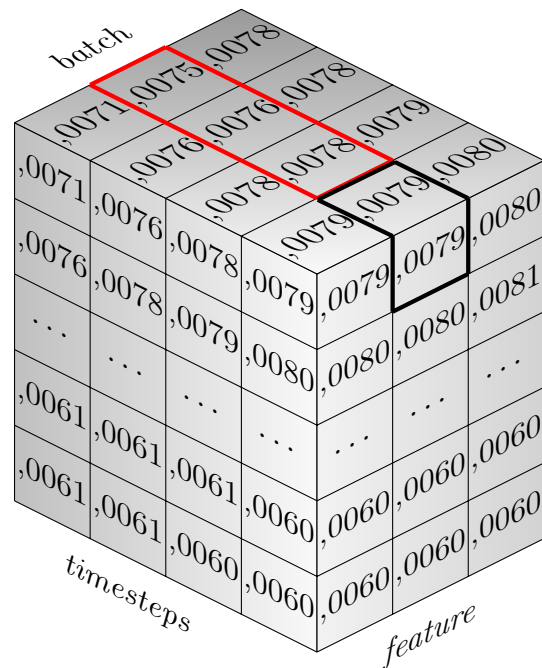
3.2.3 Dimensionamento dos dados

As redes neurais foram construídas usando o Keras (CHOLLET et al., 2015). A estrutura do LSTM no Keras espera uma entrada \mathbf{X} com três dimensões: $[batch, timesteps, feature]$. Nele, o *batch* define a quantidade de amostras de entrada, o *timestep* é o intervalo de tempo entre cada observação ou amostra na série, e *feature* são os atributos da amostra.

A rede CLSTM1D espera uma entrada \mathbf{X} com 4 dimensões: $[batch, timesteps, rows, filters]$. Enquanto a rede CLSTM2D espera 5 dimensões: $[batch, timesteps, rows, columns, channels]$. Onde *rows*, *columns*, e *filters/channels* são as linhas, colunas, e filtros/canais, semelhante ao observado no tratamento de imagens. Usaremos a estrutura CLSTM1D neste projeto.

Adequando o *dataset* da Tabela 3, obtemos o tensor da Figura 13. É possível reconhecer os dados das primeiras 3 colunas usadas para exemplificar. Cada linha da matriz define uma amostra, cada coluna é uma variação temporal, ou *timestep*, e no terceiro eixo, ortogonal aos outros dois, estão os atributos das amostras.

Figura 13 – Tensor 3D dos dados



Fonte: Próprio autor.

3.3 Construção das Redes Neurais

A construção da rede neural é empírica. Esta seção descreve o *design* e analisa as redes recorrentes propostas para o problema de propagação temporal/espectral do pulso ótico.

3.3.1 LSTM vs CLSTM

A primeira parte da análise descreve o *design* das redes LSTM e CLSTM, variando seus hiperparâmetros empiricamente, para verificar seu impacto sobre a performance do modelo. Os hiperparâmetros são parâmetros usados para construir a rede neural, de forma que ela não poderia aprender de maneira autônoma. Nesse caso, será avaliado somente a propagação temporal do pulso. A inclusão do domínio da frequência será feita posteriormente.

Os hiperparâmetros testados foram o número de células LSTM, filtros para a CLSTM, o número de camadas densas, e escolha dos *timesteps*. O número de células LSTM avaliados foram de 256, 512, 768, e 1.024. Foram adicionadas até 3 camadas densas, com 161 neurônios cada. E o horizonte de predição (*timesteps*) foi de 1, 5, e 10. A métrica escolhida para avaliar a performance da rede foi a NRMSE. Todas as redes LSTM foram treinadas com o otimizador o RMSProp (de *Root Mean Square Propagation*) em três conjuntos de

épocas, cada um com 30 épocas. Em cada conjunto, uma taxa de aprendizagem diferente foi aplicada: 1×10^{-4} , 1×10^{-5} , e 1×10^{-6} . O decaimento da taxa de aprendizagem ao longo do treinamento tem como objetivo convergir mais facilmente para a solução. A Tabela 4 sumariza os resultados obtidos.

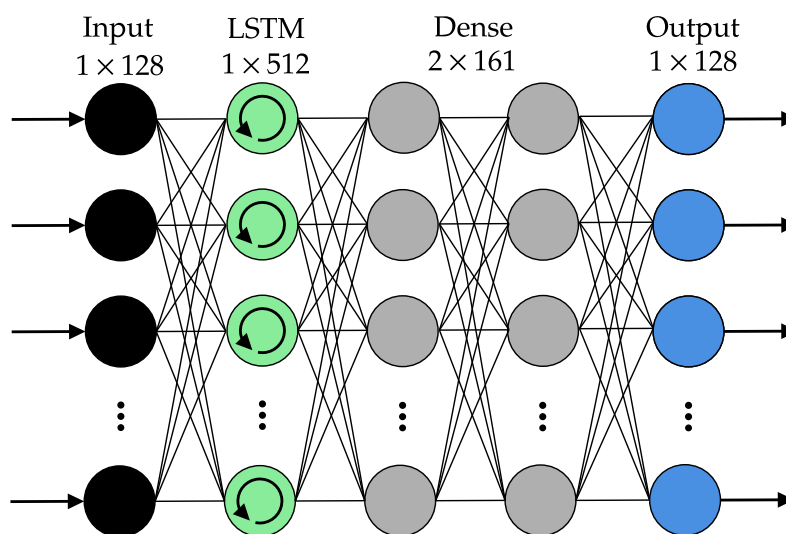
Tabela 4 – Hiperparâmetros da LSTM

LSTM	Dense	Window size	NRMSE
256	1×161	1, 5, 10	2,973355; 0,366610; 0,118528
	2×161	1, 5, 10	0,678828; 0,296355; 0,161916
	3×161	1, 5, 10	0,641981; 0,320857; 0,112304
512	1×161	1, 5, 10	2,569882; 0,819802; 0,125211
	2×161	1, 5, 10	0,816913; 0,267253; 0,090840
	3×161	1, 5, 10	0,708236; 0,147685; 0,111386
768	1×161	1, 5, 10	3,206505; 0,957628; 0,120203
	2×161	1, 5, 10	1,537784; 0,372963; 0,112285
	3×161	1, 5, 10	0,607526; 0,267638; 0,130854
1024	1×161	1, 5, 10	3,495995; 0,454984; 0,113306
	2×161	1, 5, 10	1,886774; 0,360326; 0,101754
	3×161	1, 5, 10	0,578932; 0,215696; 0,167321

Fonte: Próprio autor.

A rede que apresentou o melhor desempenho dentre aquelas testadas para o caso foi a rede com 512 células LSTM, duas camadas densas intermediárias, e com um horizonte de predição de 10 *timesteps*. As camadas recorrentes tem como função de ativação a tangente hiperbólica, a função das camadas densas é o ReLU, e a da camada de saída é a sigmoide. O esquemático da rede é apresentado na Figura 14. O melhor desempenho obtido foi $\text{NRMSE} = 0,09$.

Figura 14 – Esquemático da rede LSTM



Fonte: Próprio autor.

A mesma busca empírica foi realizada para a rede CLSTM. Similar a uma camada LSTM, mas as transformações de entrada e de recorrência são ambas convolucionais (SHI et al., 2015). O número de filtros avaliados foram de 20, 40, 80, e 100. Foram adicionadas até 3 camadas densas, com 161 neurônios cada. E o horizonte de predição foi de 5 e 10. Foi observado da Tabela 4 que o $timestep = 1$ resultou em uma performance pobre, e por isso foi dispensado. Todas as CLSTM foram treinadas com o otimizador RMSProp em dois conjuntos de épocas, cada um com 30 épocas. Em cada conjunto, uma taxa de aprendizagem diferente foi aplicada: 1×10^{-4} , e 1×10^{-5} . O decaimento da taxa de aprendizagem ao longo do treinamento acelera a convergência do modelo. A Tabela 5 sumariza os resultados obtidos.

Tabela 5 – Hiperparâmetros da CLSTM

CLSTM1D	Dense	Window size	NRMSE
20	1×161	5, 10	1,013873; 0,366006
	2×161	5, 10	0,576239; 0,299295
	3×161	5, 10	0,354939; 0,425800
40	1×161	5, 10	0,929835; 0,287829
	2×161	5, 10	0,480841; 0,163120
	3×161	5, 10	0,302953; 0,300496
80	1×161	5, 10	1,019847; 0,566825
	2×161	5, 10	0,341848; 0,303296
	3×161	5, 10	0,416118; 0,177284
160	1×161	5, 10	1,446962; 0,915385
	2×161	5, 10	0,385630; 0,351022
	3×161	5, 10	0,264444; 0,152486

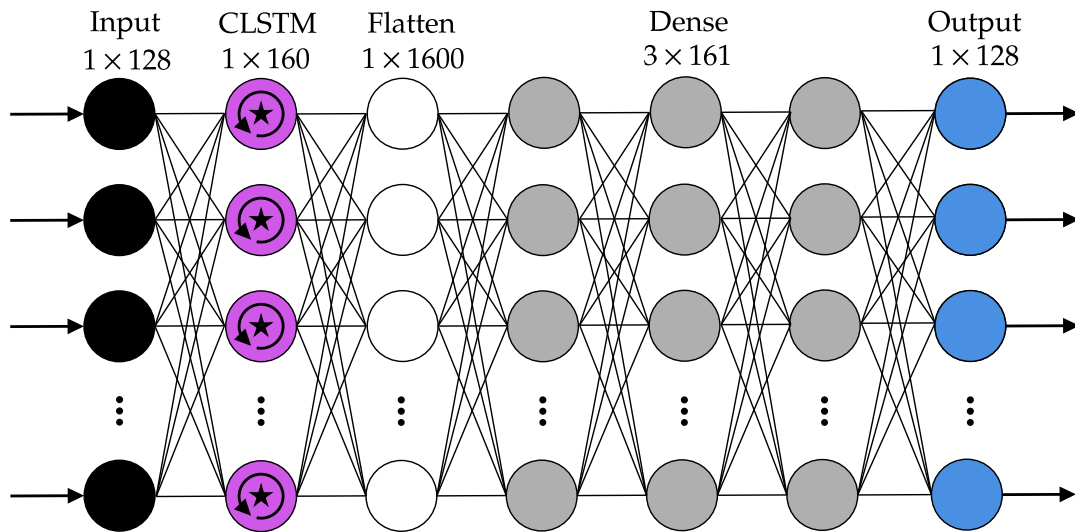
Kernel mantido constante; kernel = 5.

Fonte: Próprio autor.

A rede com o melhor desempenho foi a rede com 160 filtros e $kernel = 5$, três camadas densas intermediárias, e com um horizonte de predição de 10 $timesteps$. A rede também inclui uma camada de achatamento (*flatten*) responsável por transformar a saída da camada CLSTM de um tensor de tamanho ($1 \times 10 \times 160$) para um vetor de tamanho (1×1600). As camadas recorrentes tem como função de ativação a tangente hiperbólica, a função das camadas densas é o ReLU, e a da camada de saída é a sigmoide. O esquemático da mesma é apresentado na Figura 15. Nesse caso, percebe-se que a rede construída é mais pesada que a anterior, com maior número de parâmetros, e camadas. O melhor desempenho obtido foi $NRMSE = 0,15$.

As redes descritas foram utilizadas para prever a evolução óptica temporal apenas, a inclusão do domínio espectral tem resultados ruins, e por isso foi desconsiderado até aqui.

Figura 15 – Esquemático da rede CLSTM



Fonte: Próprio autor.

3.3.2 LSTM vs BiLSTM

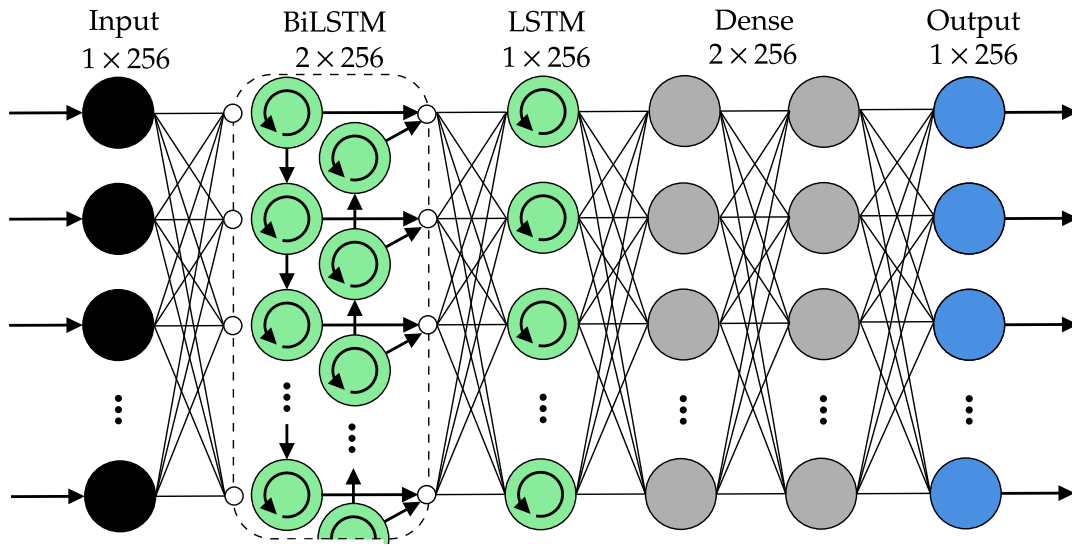
Para generalizar a análise ao domínio da frequência, vamos utilizar a BiLSTM. Por se tratar de uma rede com estrutura mais complexa, tratando os dados de modo bidirecional, a rede BiLSTM foi escolhida para ampliar a análise. Não realizamos uma busca comparativa extensa nesse caso.

A estrutura da rede BiLSTM é apresentada na Figura 16. A rede possui duas camadas recorrentes, sendo uma BiLSTM e outra LSTM unidirecional, 2 camadas densas, e 10 *timesteps*. Todas as camadas possuem 256 células LSTM. As camadas recorrentes tem como ativação a função tanh, a função das camadas densas é o ReLU, e a da saída é a sigmoide. O esquemático da rede é apresentado na Figura 16. O número de células foi duplicado em razão da entrada agora ser um tensor agregando as evoluções temporais e espectrais simultaneamente ($128 + 128$). Durante o treinamento foi utilizado o otimizador Adam em três conjuntos de épocas, os dois primeiros com 30 épocas e o terceiro com dez épocas. Em cada conjunto, uma taxa de aprendizagem diferente foi utilizada: 1×10^{-4} , 1×10^{-5} e 1×10^{-6} . Mais uma vez, O decaimento da taxa de aprendizagem é feito para que o modelo atinja a convergência mais rapidamente.

Também usaremos uma rede LSTM para comparar o desempenho obtido, nesse caso a LSTM terá a mesma estrutura vista na Figura 16, substituindo a camada bidirecional por 2 camadas LSTM, de modo a realizar uma comparação justa.

Ainda, para o caso da geração de supercontínuo, o número de neurônios foi alterado para

Figura 16 – Esquemático da rede BiLSTM



Fonte: Próprio autor.

descrever melhor a dinâmica não linear e efeitos de maior ordem, nesse caso a entrada tem dimensão $(190 + 248)$. Durante o treinamento, foram realizadas 60 épocas usando o otimizador Adam com taxa de aprendizagem adaptativa (1×10^{-4}) , e paciência de 20 épocas uma vez que alcança um platô).

3.3.3 Métricas de Desempenho

Com a rede treinada, os dados de teste (isolados do treinamento) são passados para a rede, que deve fazer a predição das respostas. A seguir temos um conjunto de métricas usadas.

As duas equações mais recorrentes na literatura são o erro quadrático médio (MSE) e sua raiz (RMSE), descrito pela Equação (3.6) e Equação (3.7) abaixo

$$\text{MSE} = \frac{1}{N} \sum_{m=1}^N (x_m - \hat{x}_m)^2 \quad (3.6)$$

$$\text{RMSE} = \sqrt{\text{MSE}} \quad (3.7)$$

onde x_m é o valor obtido pela SSF e \hat{x}_m o valor obtido pela rede para cada previsão m . O MSE acelera o treinamento, no entanto é sensível a *outliers* (JADON; PATIL; JADON, 2022). O RMSE é facilmente diferenciável, mas penaliza os erros de treinamento menos que o MSE (JADON; PATIL; JADON, 2022).

Uma outra métrica, similar ao erro quadrático, é o erro absoluto médio (MAE). O MAE é mais simples e menos sensível a *outliers*, tratando os erros igualmente, mas não é

diferenciável em zero (JADON; PATIL; JADON, 2022). Descrito pela Equação (3.8), temos

$$\text{MAE} = \frac{1}{N} \sum_{m=1}^N |x_m - \hat{x}_m| \quad (3.8)$$

Para os índices acima, quanto menor, melhor o ajuste. De outra forma, para o coeficiente de determinação R^2 descrito na Equação (3.9) a seguir, que mede a redução da variabilidade total da saída, $0 \leq R^2 \leq 1$, queremos que R^2 se aproxime de 1,

$$R^2 = 1 - \frac{\sum_{m=1}^N (x_m - \hat{x}_m)^2}{\sum_{m=1}^N (x_m - \bar{x})^2} \quad (3.9)$$

sendo que \bar{x} denota a média dos valores de x_m . A métrica R^2 fornece uma compreensão intuitiva, percentual, para o leitor. Contudo, assume relacionamento linear entre as variáveis, podendo não ser aplicável em alguns casos (SC).

4 RESULTADOS

Neste capítulo são discutidos os resultados obtidos com a rede LSTM, CLSTM, e BiLSTM, sua acurácia ao longo da evolução óptica, e como o erro de predição se propaga. Serão tratados os problemas de propagação temporal e espectral de um pulso óptico solitônico e pulso óptico gaussiano, além da geração de supercontínuo. Por fim, o tempo computacional do modelo neural será comparado ao do método de simulação numérico.

4.1 LSTM vs CLSTM

Tomando como ponto de partida apenas a propagação temporal, será demonstrado o resultado da performance obtida com as redes LSTM e CLSTM, cuja busca empírica de seus hiperparâmetros foi realizada na Seção 3.3.1. Descreveremos o *dataset* a ser aplicado a um pulso solitônico, em seguida esse *dataset* será estendido para demonstrar a generalização da rede neural proposta. Também o mesmo *dataset* será aplicado a um pulso gaussiano, desta vez sem estendê-lo. Finalmente, será realizada uma análise do erro ao longo da fibra.

4.1.1 Propagação Temporal de Pulso Solitônico

Com as arquiteturas da LSTM e CLSTM, obtemos os resultados observados na Tabela 6. Nesse caso, o *dataset* é composto por $P_0 \in [18,41; 34,19]$ W, e $T_{\text{FWHM}} \in [0,77; 1,43]$ ps. As mesmas arquiteturas foram retreinadas para uma base de dados com maior variação de potência e largura de pulso, $P_0 \in [35,0; 40,0]$ W, $T_{\text{FWHM}} \in [0,5; 5,0]$ ps para o LSTM e $T_{\text{FWHM}} \in [0,5; 3,0]$ ps para o CLSTM. Os resultados estão apresentados na Tabela 7.

No primeiro caso, as redes capturam bem a dinâmica de propagação temporal, tendo ambas R^2 acima de 96%, e $\text{RMSE} = 3,494 \times 10^{-3}$ para o LSTM, e $\text{RMSE} = 6,032 \times 10^{-3}$ para o CLSTM. A qualidade de predição é evidenciada nas Figuras 17 e 18, que mostram a evolução temporal do sóliton ao longo da fibra, e sua forma de onda em distâncias selecionadas. O sóliton tem como condições iniciais $P_0 = 27,31$ W e $T_{\text{FWHM}} = 1,07$ ps. Devido a compressão do pulso, o pico atinge um máximo de 250 W durante a propagação. A complexa simulação numérica SSFM é acompanhada de perto pela predição da rede LSTM, e a diferença fica em torno de apenas 15 W. A predição é feita praticamente sem erro na maior parte da evolução óptica, como evidenciado na Figura 17(c). A Figura

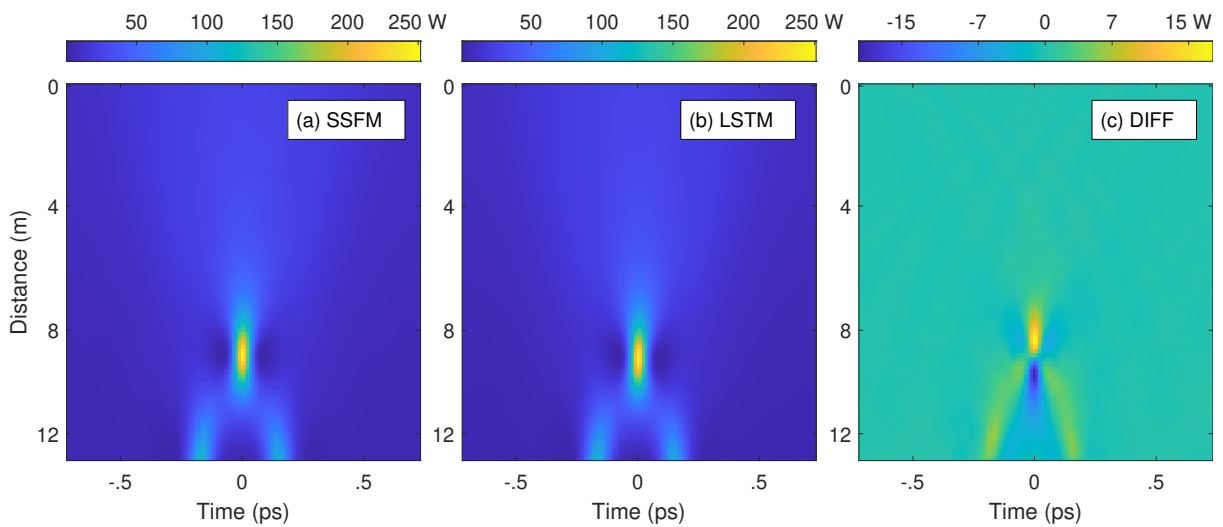
18 corrobora com isso, mostrando a forma do pulso predito alinhado com a simulação numérica para as três distâncias numéricas propagadas.

Tabela 6 – Performance da LSTM vs CLSTM

Rede Neural	RMSE	MSE	MAE	R ²
LSTM	$3,494 \times 10^{-3}$	$1,22 \times 10^{-5}$	$1,94 \times 10^{-3}$	98,68%
CLSTM	$6,032 \times 10^{-3}$	$3,64 \times 10^{-5}$	$3,07 \times 10^{-3}$	96,11%

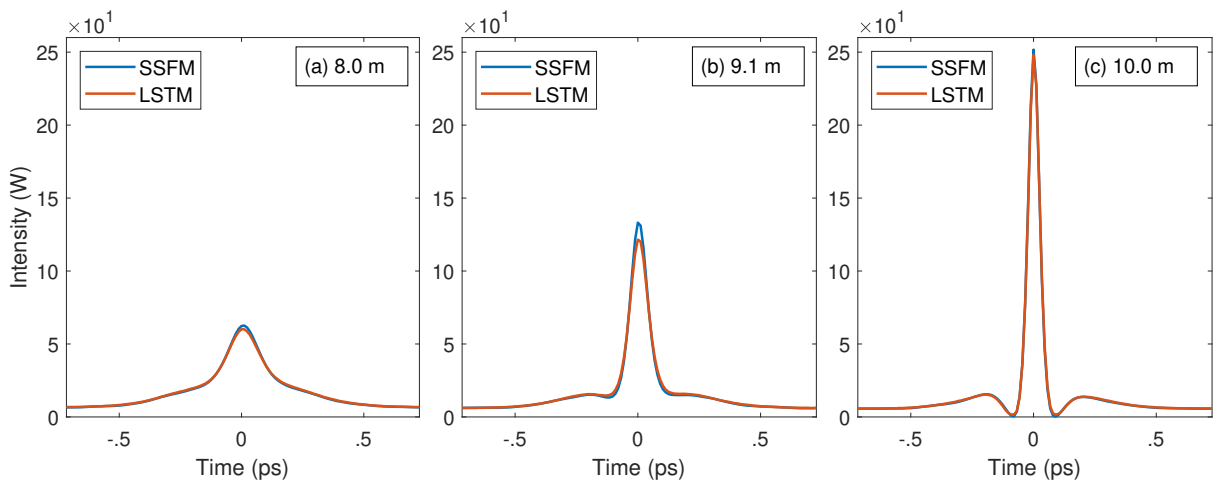
Fonte: Próprio autor.

Figura 17 – Evolução temporal do sóliton pela LSTM



Fonte: Próprio autor.

Figura 18 – Intensidade temporal do sóliton pela LSTM em distâncias selecionadas



Fonte: Próprio autor.

No segundo caso, submetendo as redes sob regimes não lineares mais complexos, com valores de potência de pico de entrada mais elevados, realizamos novo treinamento das redes LSTM e CLSTM, os resultados são mostrados na Tabela 7. A rede LSTM ainda é adequada para esse novo regime de propagação, apresentando $RMSE = 5,328 \times 10^{-3}$ e

$R^2 = 95,17\%$, em contraposição, a performance da rede CLSTM é fortemente afetada pelo novo regime, com $RMSE = 1,835 \times 10^{-2}$ e $R^2 = 66,81\%$.

Tabela 7 – Performance estendida da LSTM vs CLSTM para o sóliton

Rede Neural	RMSE	MSE	MAE	R^2
LSTM	$5,328 \times 10^{-3}$	$2,84 \times 10^{-5}$	$1,92 \times 10^{-3}$	95,17%
CLSTM	$1,538 \times 10^{-2}$	$2,37 \times 10^{-4}$	$6,57 \times 10^{-3}$	66,81%

Fonte: Próprio autor.

4.1.2 Propagação Temporal de Pulso Gaussiano

Além de estender o regime de propagação dos sólitons, utilizamos o *dataset* original e alteramos a forma de onda para um pulso gaussiano, treinando novamente as redes. A mesma estrutura das redes foi usada, apenas alteramos seus pesos. O resultado é mostrado na Tabela 8. Em comparação com o pulso solitônico, as redes LSTM e CLSTM tem maior dificuldade em prever a evolução gaussiana, mas ainda se obtém um bom desempenho. A rede LSTM é adequada para o pulso gaussiano, apresentando $RMSE = 3,310 \times 10^{-3}$ e $R^2 = 97,69\%$, além disso, a CLSTM possui $RMSE = 5,828 \times 10^{-3}$ e $R^2 = 92,83\%$. Um exemplo de evolução óptica é visto nas Figuras 19 e 20. O pulso gaussiano tem como condições iniciais $P_0 = 34,19$ W e $T_{FWHM} = 1,08$ ps. Devido a compressão do pulso, o pico ultrapassa a potência de 350 W. O erro de predição fica abaixo de 15 W em toda evolução. A simulação numérica SSFM é acompanhada de perto pela predição da LSTM, e o ajuste das curvas na Figura 20 evidencia isso para as distâncias demarcadas.

Tabela 8 – Performance da LSTM vs CLSTM para o pulso gaussiano

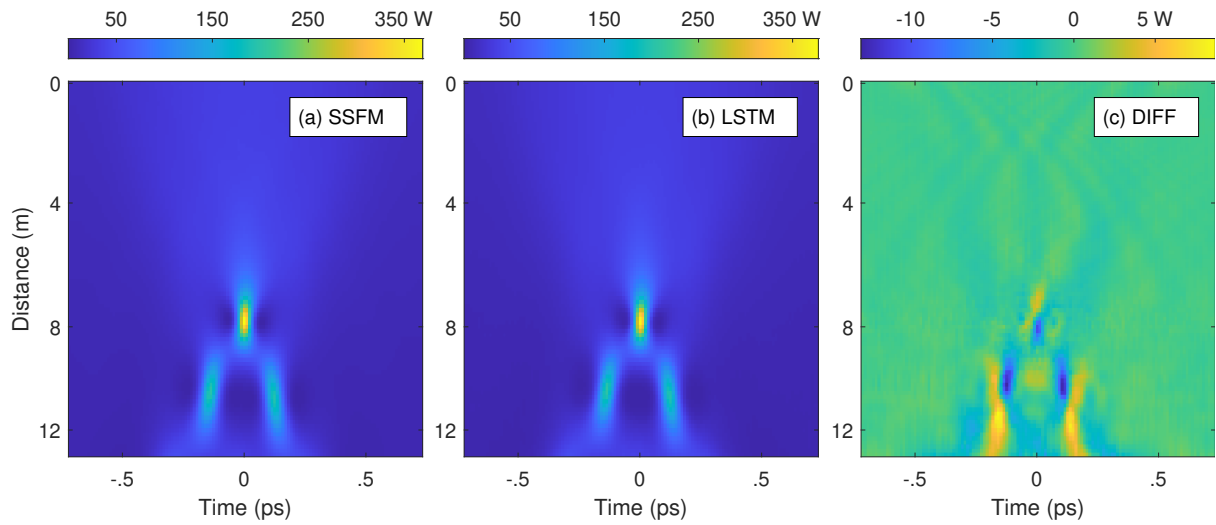
Rede Neural	RMSE	MSE	MAE	R^2
LSTM	$3,310 \times 10^{-3}$	$4,20 \times 10^{-5}$	$1,70 \times 10^{-3}$	97,69%
CLSTM	$5,828 \times 10^{-3}$	$1,05 \times 10^{-4}$	$2,70 \times 10^{-3}$	92,83%

Fonte: Próprio autor.

4.1.3 Propagação do Erro

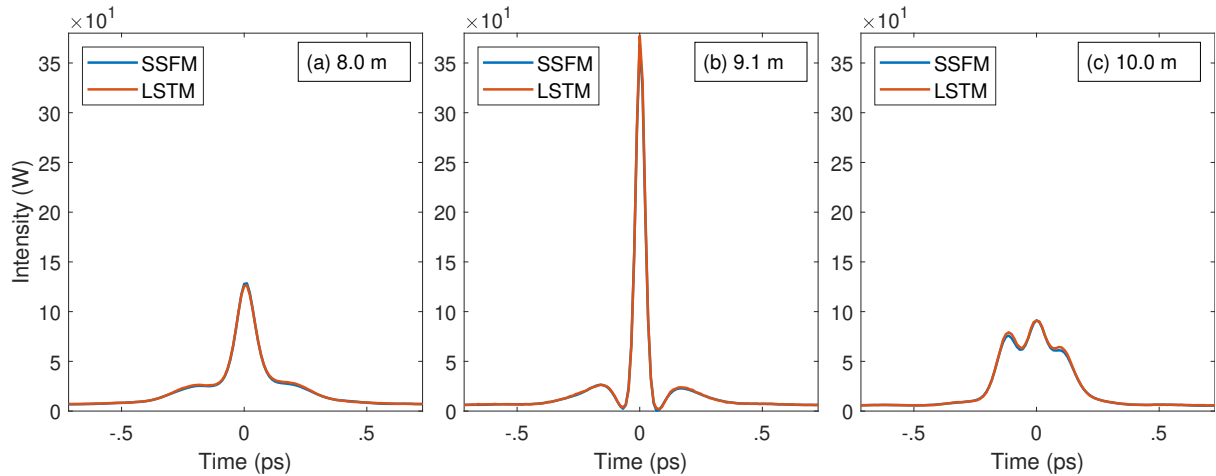
Por fim, fazemos uma análise da propagação do erro ao longo da fibra. Para isso, calculamos o erro RMSE a cada passo propagado, $RMSE_{STEP}$, e sua média, $RMSE_{AVG}$. A Figura 21 mostra o crescimento desse erro na fibra para as duas redes, LSTM e CLSTM. Como já vimos, a LSTM é mais adequada para descrever a evolução óptica, fato este percebido aqui onde o erro $RMSE_{STEP}$ não ultrapassa $6,5 \times 10^{-3}$, e conseqüentemente, seu valor médio é pequeno e não atinge $3,0 \times 10^{-3}$. O comportamento do erro $RMSE_{AVG}$ é monotonicamente

Figura 19 – Evolução temporal gaussiano pela LSTM



Fonte: Próprio autor.

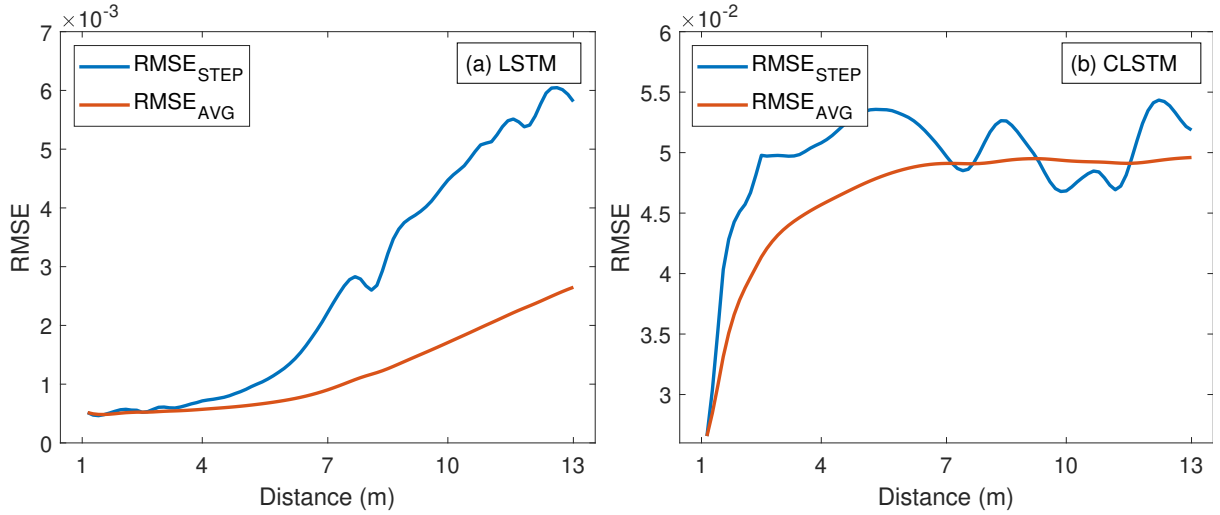
Figura 20 – Intensidade temporal gaussiano pela LSTM em distâncias selecionadas



Fonte: Próprio autor.

crecente, ou seja, a rede LSTM não é adequada para comprimentos acima de 13 m. A Figura 21(b) mostra um crescimento do erro mais acentuado e precoce para a CLSTM, alcançando um erro $RMSE_{STEP} = 5,0 \times 10^{-2}$ no início da fibra, e variando em torno desse valor ao longo da propagação, atingindo um máximo de $RMSE_{STEP} = 5,5 \times 10^{-2}$ e também $RMSE_{AVG} = 5,0 \times 10^{-2}$. O erro mais elevado torna a CLSTM menos recomendada para o problema de propagação óptica.

Figura 21 – Propagação do erro das redes LSTM e CLSTM na fibra



Fonte: Próprio autor.

4.2 LSTM vs BiLSTM

Incluiremos agora o domínio espectral, e demonstraremos o resultado da performance obtida com as redes neurais LSTM e BiLSTM, cujos hiperparâmetros foram tratados na Seção 3.3.2. Apenas o pulso no formato solitônico foi considerado nas análises seguintes. Além da dinâmica de propagação, estudaremos ainda um caso mais complexo de geração de supercontínuo.

4.2.1 Propagação de Pulso Solitônico

Com as arquiteturas da LSTM e BiLSTM, obtemos os resultados observados na Tabela 9. Nesse caso, o *dataset* é composto por $P_0 \in [18,41; 34,19]$ W, e $T_{FWHM} \in [0,77; 1,43]$ ps. O tratamento nos domínios do tempo e frequência ocorrem simultaneamente nesse caso, e para o cálculo das métricas separamos antes os domínios. A inclusão do domínio espectral permite maior generalização da rede.

Para o domínio temporal, a rede LSTM tem um pior desempenho que a BiLSTM, tendo $RMSE = 1,093 \times 10^{-2}$ e $R^2 = 82,23\%$, enquanto a rede neural BiLSTM apresenta um erro $RMSE = 4,390 \times 10^{-3}$ e $R^2 = 97,26\%$. Para o domínio espectral, a situação é semelhante, a LSTM possui um $RMSE = 3,788 \times 10^{-2}$ e $R^2 = 91,16\%$, enquanto a rede BiLSTM apresenta um erro $RMSE = 1,211 \times 10^{-2}$ e $R^2 = 98,84\%$. A inclusão do domínio espectral tem como consequência direta uma piora na performance da LSTM simples, devido a maior complexidade da base de dados, nesse caso a BiLSTM deve ser aplicada.

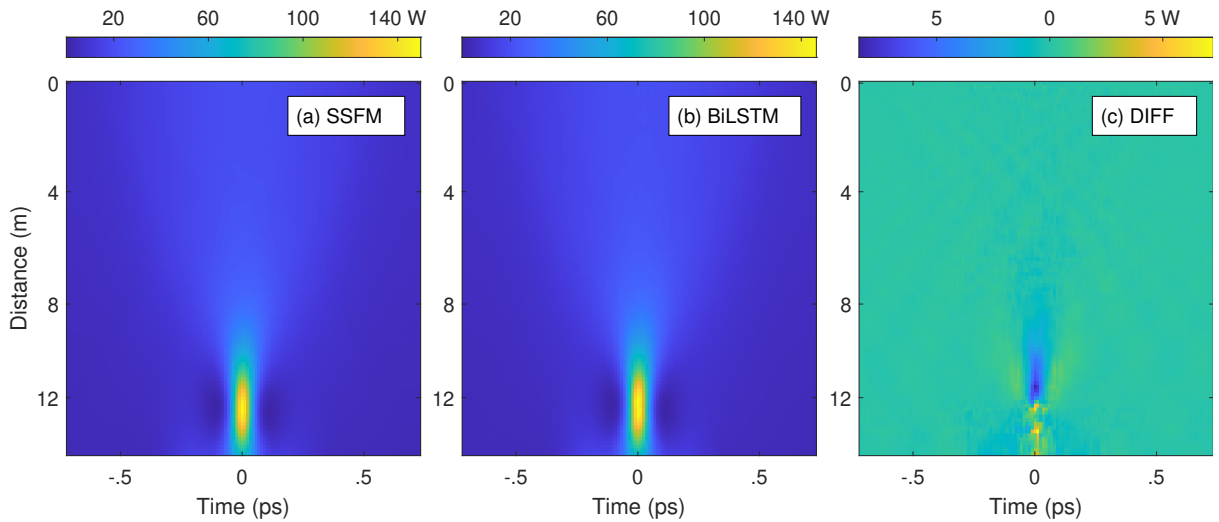
Tabela 9 – Performance da LSTM *vs* BiLSTM

Rede Neural	Domínio	RMSE	MSE	MAE	R ²
LSTM	Temporal	$1,093 \times 10^{-2}$	$2,52 \times 10^{-4}$	$5,16 \times 10^{-3}$	82,23%
	Espectral	$3,788 \times 10^{-2}$	$1,52 \times 10^{-3}$	$2,24 \times 10^{-2}$	91,16%
BiLSTM	Temporal	$4,390 \times 10^{-3}$	$3,90 \times 10^{-5}$	$1,77 \times 10^{-3}$	97,26%
	Espectral	$1,211 \times 10^{-2}$	$1,59 \times 10^{-4}$	$6,11 \times 10^{-3}$	98,84%

Fonte: Próprio autor.

A predição é mostrada nas Figuras 22 e 23 que retratam o domínio temporal. O sóliton mostrado tem como condições iniciais $P_0 = 20,03$ W e $T_{FWHM} = 0,98$ ps. O pico atinge um máximo de 145 W durante a propagação. A complexa simulação numérica SSFM é acompanhada de perto pela predição da rede BiLSTM, e a diferença fica em torno de apenas 5 W. O mesmo pode ser visto no espectro, Figuras 24 e 25. O espectro possui limite inferior de -55 dB, que facilita o aprendizado das redes e acelera a convergência do modelo. Ao longo da propagação na fibra, ocorre um alargamento espectral devido ao SPM. A rede BiLSTM é capaz, portanto, de realizar a predição correta e simultânea da propagação óptica em ambos domínios temporal e espectral.

Figura 22 – Evolução temporal do sóliton pela BiLSTM

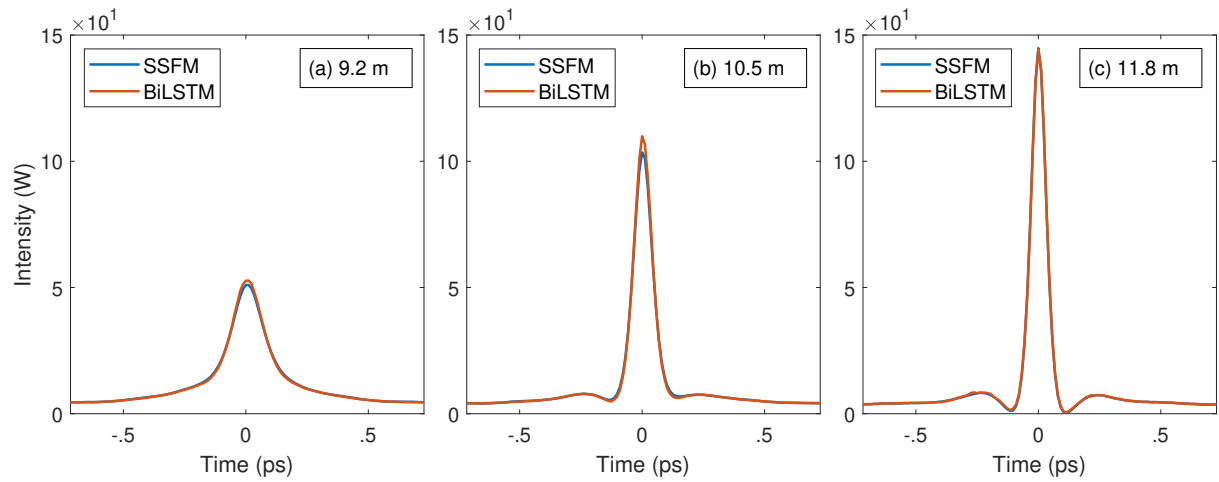


Fonte: Próprio autor.

4.2.2 Propagação do Erro

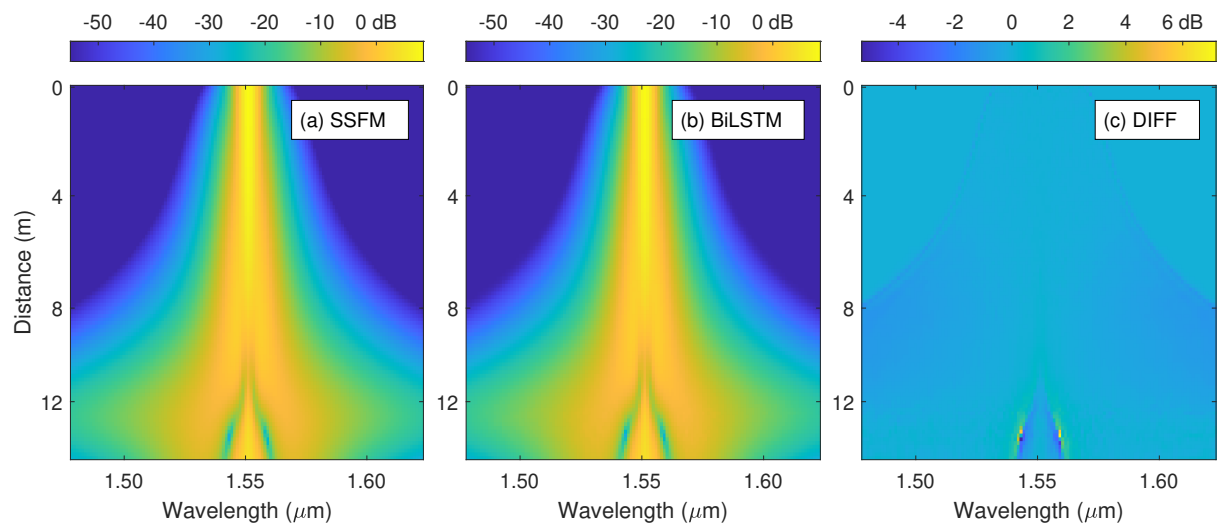
Finalmente, faremos outra vez uma análise da propagação do erro ao longo da fibra. Calculamos abaixo apenas os erros $RMSE_{STEP}$. A Figura 26 mostra o crescimento desse erro na fibra para as duas redes, LSTM e BiLSTM. A LSTM é menos adequada para descrever a evolução (simultânea) temporal e espectral, e a curva de propagação do erro é pelo menos o dobro do valor em comparação com a BiLSTM. Temporalmente, o

Figura 23 – Intensidade temporal do sóliton pela BiLSTM em distâncias selecionadas



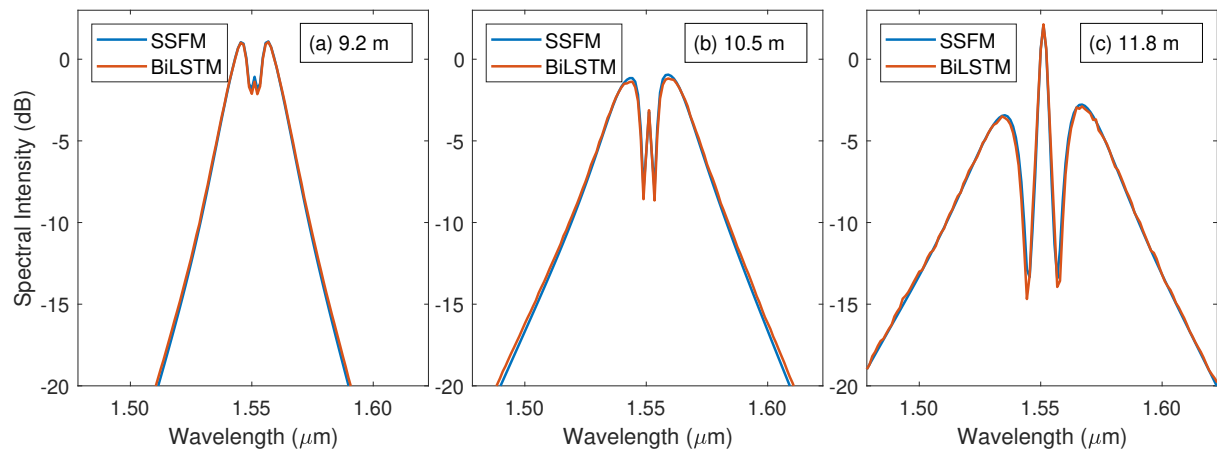
Fonte: Próprio autor.

Figura 24 – Evolução espectral do sóliton pela BiLSTM



Fonte: Próprio autor.

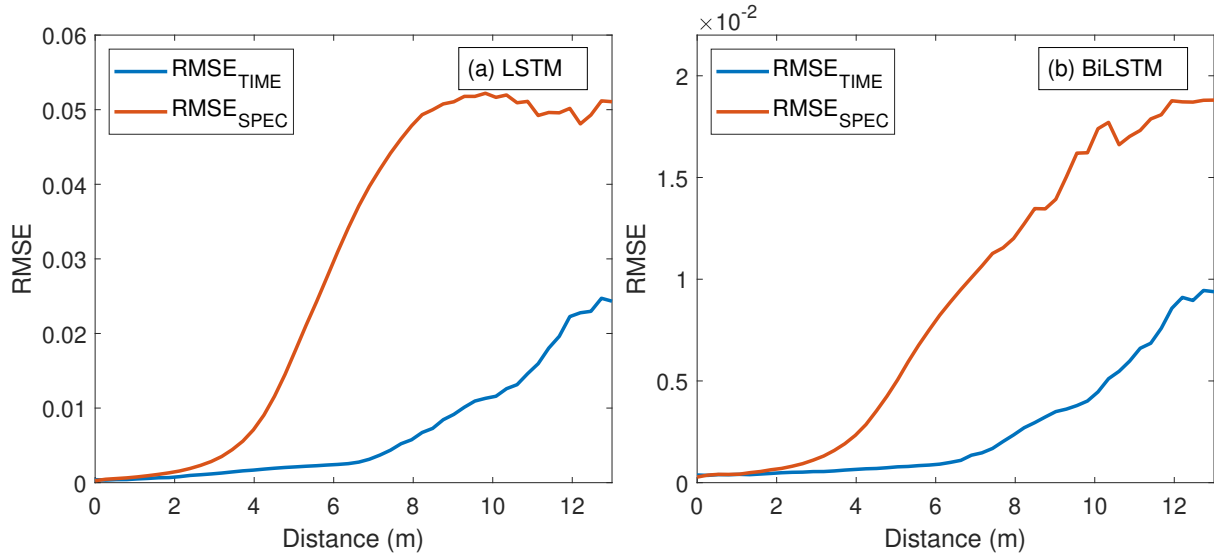
Figura 25 – Intensidade espectral do sóliton pela BiLSTM em distâncias selecionadas



Fonte: Próprio autor.

erro $\text{RMSE}_{\text{TIME}}$ permanece baixo ao longo de toda a fibra, e alcança 1×10^{-2} ao final da propagação. A evolução espectral da BiLSTM tem um rápido crescimento do erro $\text{RMSE}_{\text{SPEC}}$ já no início da fibra, com apenas 4 metros, e atinge um máximo de $1,9 \times 10^{-2}$.

Figura 26 – Propagação do erro das redes LSTM e BiLSTM na fibra



Fonte: Próprio autor.

4.2.3 Geração de Supercontínuo

Entre os processos de alargamento espectral e geração de novas frequências, entre eles a auto-modulação de fase ou os processos de espalhamento Raman e Brillouin, por exemplo, a geração de supercontínuos se destaca em razão de sua banda contínua e que se entende por diferentes comprimentos de onda (DUDLEY; GENTY; COEN, 2006). Podem ser aplicados em espectroscopia (ZORIN et al., 2022; KREBBERS et al., 2022), compressão de pulsos (HUA-QIANG et al., 2016), e para multiplexação por divisão de comprimento de onda (NGUYEN-THE; MATSUURA; KISHI, 2014; SAGHAEI, 2017). Para a geração de supercontínuos, se faz necessário uma alta não linearidade do material aliado a uma área efetiva pequena, e curva de dispersão adequada. No regime de dispersão anômala, o supercontínuo é comumente gerado em razão do processo de fissão que ocorre com sólitons de alta ordem (KOROBKO et al., 2023). O mesmo processo de fissão e consequente geração de supercontínuo não ocorre no caso de dispersão normal, já que esse regime impede que ocorra instabilidade de modulação e formação de sólitons. O mecanismo de geração, nesse caso, é devido a ação da auto-modulação de fase e mistura de quatro ondas, e quebra de onda óptica (SUKHOIVANOV et al., 2014; HEIDT et al., 2017). Aqui examinamos os efeitos provocados na evolução de um pulso curto incidente em um meio com dispersão anômala que experimenta o alargamento espectral não linear da geração de supercontínuo.

Obtemos os resultados observados na Tabela 10. A métrica R^2 foi removida devido a não expressar bem o ajuste da curva não linear, retornando valores sem sentido. O coeficiente R^2 assume implicitamente que o relacionamento entre as variáveis é linear. Nesse caso, o *dataset* é composto por pulsos solitônicos com $P_0 \in [0,5; 2]$ kW e $T_{FWHM} = 0,1$ ps como condições iniciais. O tratamento nos domínios do tempo e frequência ocorrem simultaneamente nesse caso, e para o cálculo das métricas os tensores dos resultados são divididos. Ainda, a amostragem do sinal temporal é reduzido para 190 pontos no intervalo $t = [-0,1; 1,2]$ ps, enquanto a amostragem do sinal espectral é reduzido para 248 pontos em $\lambda = [550, 1000]$ nm, de forma a capturar melhor a evolução não linear.

Para o domínio temporal, a rede LSTM tem um desempenho semelhante ao da rede BiLSTM, tendo $RMSE = 1,27 \times 10^{-2}$ e $MAE = 5,65 \times 10^{-3}$, enquanto a rede BiLSTM apresenta um erro $RMSE = 1,75 \times 10^{-2}$ e $MAE = 6,91 \times 10^{-3}$. Para o domínio espectral, a situação é semelhante, com a rede LSTM possuindo uma pequena superioridade com $RMSE = 9 \times 10^{-3}$ e $MAE = 5,64 \times 10^{-3}$, enquanto a rede BiLSTM apresenta um erro $RMSE = 1,32 \times 10^{-2}$ e $MAE = 7,18 \times 10^{-3}$. Vemos que a LSTM apresenta uma boa performance, no entanto, considerando o caso da propagação solitônica anterior, vemos que a BiLSTM compreende melhor ambas dinâmicas não lineares, ou seja, apresenta maior generalização.

Tabela 10 – Performance da LSTM vs BiLSTM para a geração de supercontínuo

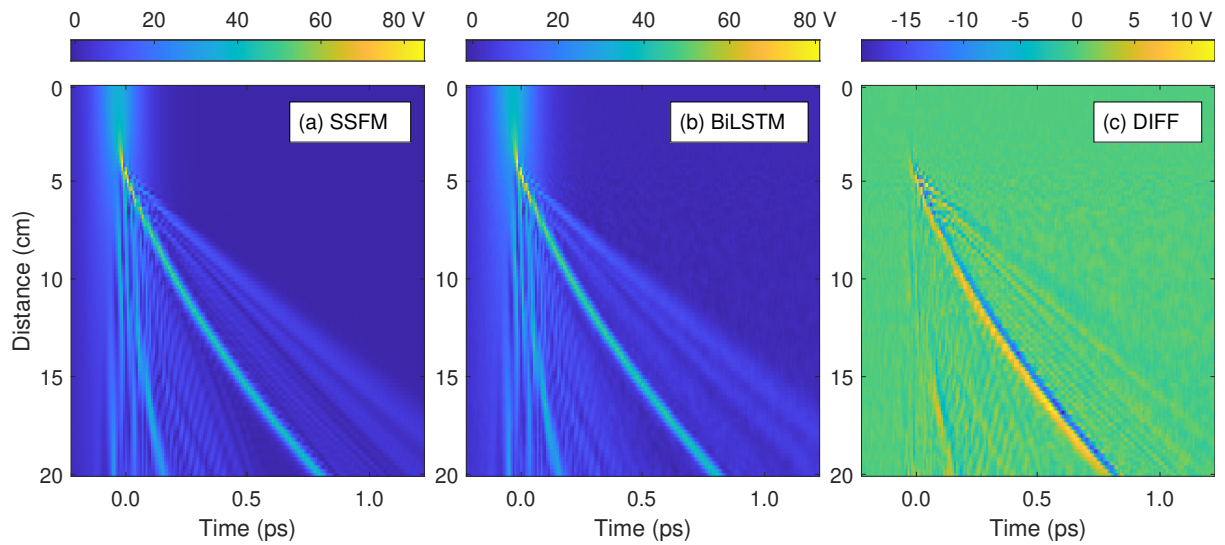
Rede Neural	Domínio	RMSE	MSE	MAE
LSTM	Temporal	$1,273 \times 10^{-2}$	$2,05 \times 10^{-4}$	$5,65 \times 10^{-3}$
	Espectral	$9,005 \times 10^{-3}$	$9,50 \times 10^{-5}$	$5,64 \times 10^{-3}$
BiLSTM	Temporal	$1,751 \times 10^{-2}$	$3,51 \times 10^{-4}$	$6,91 \times 10^{-3}$
	Espectral	$1,320 \times 10^{-2}$	$2,05 \times 10^{-4}$	$7,18 \times 10^{-3}$

Fonte: Próprio autor.

A dinâmica da geração de supercontínuo pode ser observada a seguir. Nas Figuras 27 e 28 temos a evolução temporal, enquanto nas Figuras 29 e 30 temos a evolução espectral. Fica explícito nas Figuras 27 e 28 o efeito de atraso temporal produzido pelo espalhamento Raman, onde o pulso inicial com $P_0 = 1,3$ kW e largura de pulso de 100 fs tem um desvio de quase 1 ps ao final da PCF. Constatamos como a rede neural consegue capturar bem esta dinâmica altamente complexa, com erro presente apenas nos picos mais elevados. Na Figura 28 vemos que mesmo os múltiplos picos são bem ajustados ao modelo numérico.

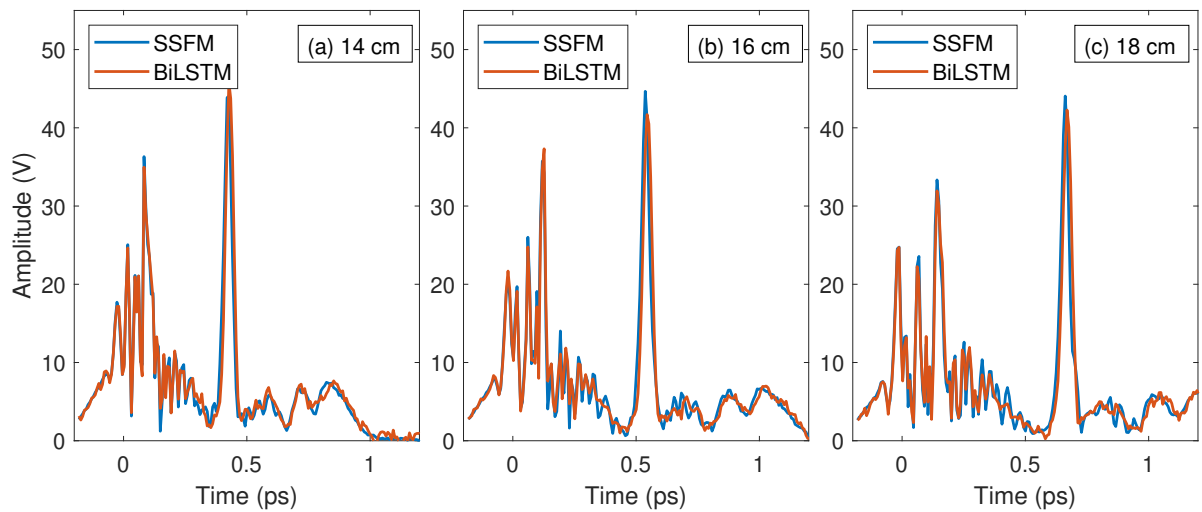
Nas Figuras 29 e 30 o deslocamento para o vermelho produzido pelo efeito Raman pode ser observado. Ainda, o supercontínuo se caracteriza pelo alargamento da banda do sinal, que chega a 400 nm para uma potência inicial de $P_0 = 1,3$ kW. Constatamos como a rede neural consegue capturar bem esta dinâmica espectral altamente complexa, e a comparação das Figuras 29(a) e 29(b) vemos que a BiLSTM parece suavizar a figura de propagação

Figura 27 – Evolução temporal do supercontínuo pela BiLSTM



Fonte: Próprio autor.

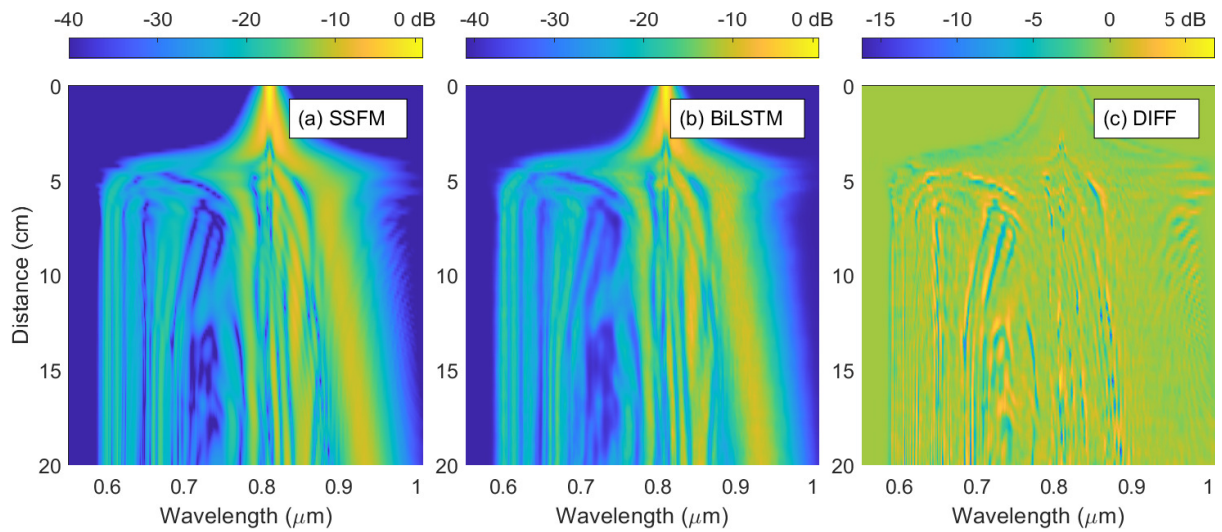
Figura 28 – Intensidade temporal do supercontínuo pela BiLSTM em distâncias selecionadas



Fonte: Próprio autor.

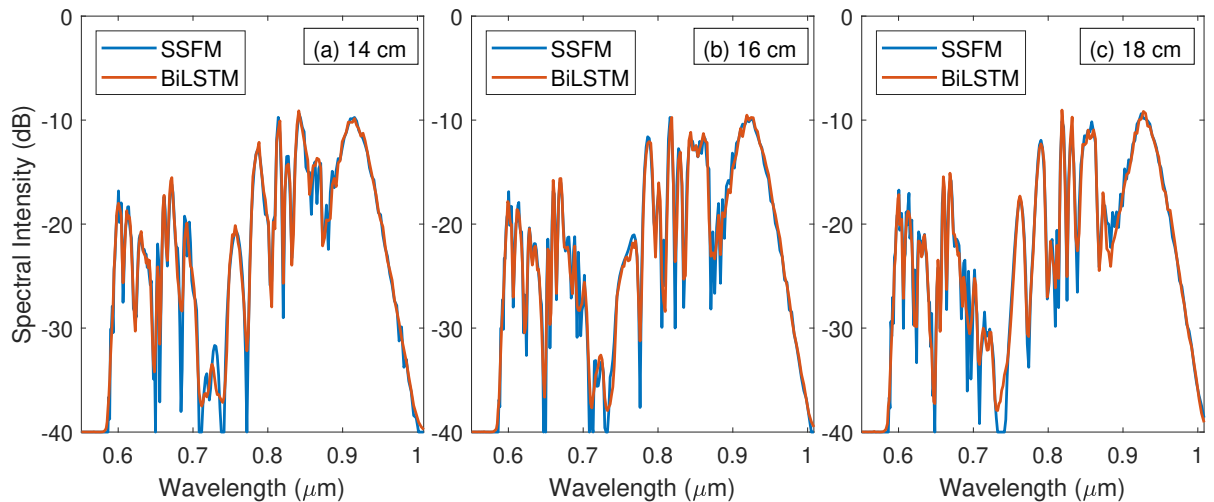
do SSFM, isso ocorre devido ao modelo neural não conseguir reproduzir os picos, o que já era esperado pelo que foi constatado em casos anteriores. A necessidade de funções de ativação suaves e diferenciáveis nas redes neurais, e.g. sigmoide e ReLU, pode limitar sua capacidade de interpolar curvas com picos pronunciados ou comportamentos não lineares extremos. A Figura 29(c) exibe a diferença entre os dois modelos, e fica claro que o erro é praticamente nulo em toda a figura. Na Figura 30 vemos que o espectro é bem ajustado ao longo de diferentes distâncias selecionadas da PCF, e que a BiLSTM não possui o mesmo *fundo de escala* que o SSFM, também por isso os dados foram limitados a uma intensidade de -40 dB.

Figura 29 – Evolução espectral do supercontínuo pela BiLSTM



Fonte: Próprio autor.

Figura 30 – Intensidade espectral do supercontínuo pela BiLSTM em distâncias selecionadas



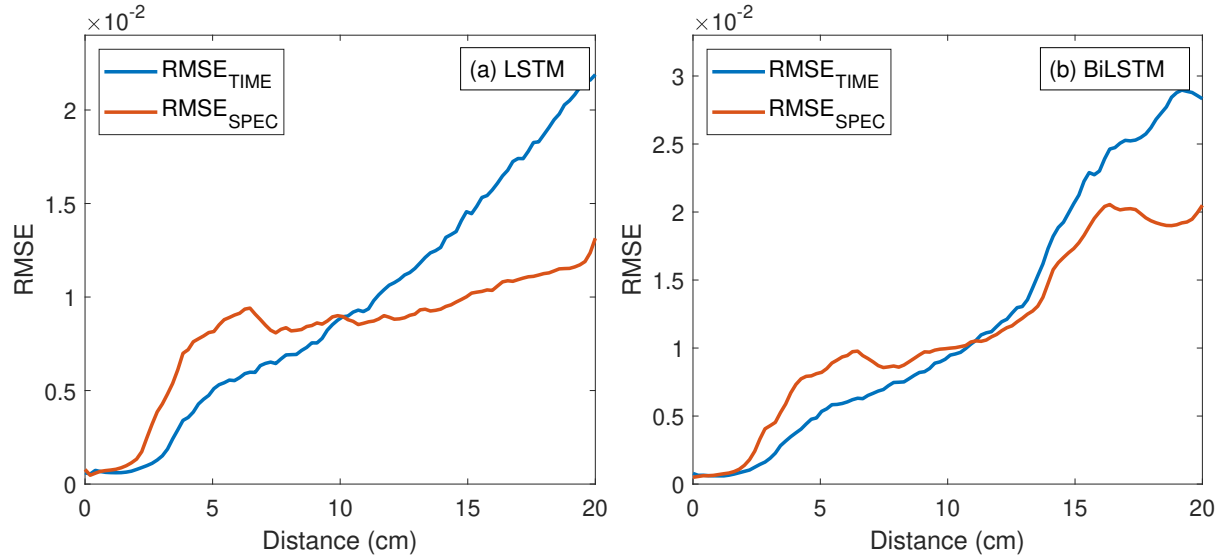
Fonte: Próprio autor.

4.2.4 Propagação do Erro

A análise da propagação do erro ao longo da PCF também foi realizada aqui. Calculamos abaixo apenas o erro $\text{RMSE}_{\text{STEP}}$. A Figura 31 mostra o crescimento desse erro para as duas redes, LSTM e BiLSTM. Ambas descrevem a evolução (simultânea) temporal e espectral de modo bastante similar. O erro observado no domínio temporal $\text{RMSE}_{\text{TIME}}$ cresce ao longo da propagação, e diferente dos casos vistos anteriormente, a não linearidade afeta a evolução já no início da PCF. O erro ao fim da PCF é então $\text{RMSE}_{\text{TIME}} = 2,2 \times 10^{-2}$ e $2,8 \times 10^{-2}$ para a LSTM e BiLSTM, respectivamente. No domínio da frequência, pelo contrário, o erro parece não crescer dramaticamente ao longo do comprimento da fibra,

isso fica mais claro observando a predição da LSTM na Figura 31 (em vermelho), em que o erro não aumenta muito a partir de 5 cm. O erro ao fim da PCF é $\text{RMSE}_{\text{SPEC}} = 1,3 \times 10^{-2}$ e $\text{RMSE}_{\text{SPEC}} = 2 \times 10^{-2}$ para a LSTM e BiLSTM, respectivamente.

Figura 31 – Propagação do erro das redes LSTM e BiLSTM na PCF



Fonte: Próprio autor.

4.3 Tempo Computacional

Finalmente, faremos brevemente uma comparação do tempo computacional exigido de cada modelo, SSFM e LSTM. Aqui será usado o modelo bidirecional LSTM, visto que possui um maior número de parâmetros em comparação com o modelo unidirecional. As simulações foram realizadas em uma mesma máquina, um *hardware* com processador Intel Core i9 e unidade de processamento gráfico (GPU, de *graphics processing unit*) RTX2080Ti. Para a dinâmica solitônica, sem termos dispersivos ou efeitos não lineares de alta ordem, o SSFM toma 1,30 s em média para simular uma única propagação, considerando 10.000 passos. No caso do supercontínuo, onde efeitos não lineares e lineares de alta ordem são considerados, o SSFM demora 5,44 s em média para simular uma única propagação, com 20.000 passos. O modelo BiLSTM leva 5,35 s e 5,98 s em média para simular 100 casos de propagação solitônica e geração de supercontínuo, respectivamente, já que apenas 100 passos são suficientes no caso da rede neural, ou seja, um ganho na velocidade computacional de 24 vezes no caso da propagação solitônica e 91 vezes no caso da geração de supercontínuo.

5 CONCLUSÕES E PROJETOS FUTUROS

O objetivo desse trabalho foi propor um modelo neural para o problema de propagação não linear de pulsos óticos, com desempenho similar ao algoritmo estado da arte *split-step* Fourier, a fim de obter um modelo de menor custo computacional e mais flexível.

Na primeira etapa, foi avaliado um conjunto de diferentes redes neurais, variando parâmetros como o número de neurônios, de camadas densas, o número de passos usados na predição (*timesteps*) e também o tipo de camada (LSTM ou CLSTM). A métrica principal para comparar essas redes foi a RMSE, mas também foram calculadas a MSE, MAE e R^2 . A CLSTM é menos adequada para tratar o problema de propagação óptica. Considerando um conjunto de treinamento avaliado para potências óticas $P_0 < 35$ W, ambas as redes capturam bem a dinâmica temporal, com R^2 acima de 96%, e $RMSE = 3,494 \times 10^{-3}$ para o LSTM, e $RMSE = 6,032 \times 10^{-3}$ para o CLSTM. Estendendo o problema para um regime não linear mais complexo, $35 \leq P_0 \leq 40$ W, e retreinando as redes, a rede LSTM se adapta para esse novo regime, apresentando $RMSE = 5,328 \times 10^{-3}$ e $R^2 = 95,17\%$, em contraposição, a CLSTM é fortemente afetada, com $RMSE = 1,835 \times 10^{-2}$ e $R^2 = 66,81\%$, uma queda expressiva da performance. Além disso, um novo treinamento é realizado para prever a evolução de pulsos gaussianos, o desempenho das redes diminui, mas ainda se obtém um bom desempenho. A rede neural LSTM apresenta um $RMSE = 3,31 \times 10^{-3}$ e $R^2 = 97,7\%$, e a CLSTM apresenta um $RMSE = 5,828 \times 10^{-3}$ e $R^2 = 92,8\%$.

Em seguida, a generalização para o domínio da frequência é proposta com o uso de uma estrutura mais complexa, bidirecional, a rede BiLSTM. Nesse caso, os domínios temporal e espectral são tratados simultaneamente. Para o domínio temporal, a rede LSTM tem um pior desempenho que a BiLSTM, tendo $RMSE = 1,093 \times 10^{-2}$ e $R^2 = 82,23\%$, enquanto a rede BiLSTM apresenta um erro $RMSE = 4,39 \times 10^{-3}$ e $R^2 = 97,26\%$. Para o domínio espectral a situação é semelhante, a LSTM possui um $RMSE = 3,788 \times 10^{-2}$ e $R^2 = 91,2\%$, enquanto a rede BiLSTM apresenta um erro $RMSE = 1,211 \times 10^{-2}$ e $R^2 = 98,84\%$. A inclusão do domínio espectral tem como consequência direta uma piora na performance da LSTM simples, devido a maior complexidade da base de dados. Também tratamos do caso da geração de supercontínuo, caracterizado por uma ampla largura de banda, ao incluir efeitos não lineares de alta ordem. Nesse caso, as redes LSTM e BiLSTM apresentaram performances semelhantes. A BiLSTM, por exemplo, obteve no domínio temporal um $RMSE = 1,75 \times 10^{-2}$ e $MAE = 6,91 \times 10^{-3}$, e no domínio espectral, consistentemente, obteve um $RMSE = 1,32 \times 10^{-2}$ e $MAE = 7,18 \times 10^{-3}$.

Seguindo, uma análise da propagação do erro ao longo da fibra é realizada. Na comparação entre a LSTM e CLSTM, a diferença é de uma ordem de magnitude em favor da LSTM, sendo esta mais apropriada para descrever a evolução óptica temporal. O crescimento do erro $\text{RMSE}_{\text{STEP}}$ não ultrapassa $6,5 \times 10^{-3}$, e conseqüentemente, seu valor médio RMSE_{AVG} se mantém pequeno, ficando abaixo de $3,0 \times 10^{-3}$. O mesmo não ocorre no caso da CLSTM, que apresenta um crescimento do erro mais acentuado e precoce, atingindo logo um máximo de $\text{RMSE}_{\text{STEP}} = 5,5 \times 10^{-2}$ e $\text{RMSE}_{\text{AVG}} = 5,0 \times 10^{-2}$. Na comparação entre a LSTM e BiLSTM utilizamos apenas o $\text{RMSE}_{\text{STEP}}$ na análise temporal/espectral. A BiLSTM apresentou uma melhor performance nesse caso, com erro temporal $\text{RMSE}_{\text{TIME}}$ abaixo de 1×10^{-2} ao longo de toda a fibra, enquanto o erro espectral $\text{RMSE}_{\text{SPEC}}$ se manteve abaixo de 2×10^{-2} ao longo de toda a propagação. E por fim, para o SC vemos que ambas redes LSTM e BiLSTM tem performance similar. O erro temporal da PCF ao final da propagação é $\text{RMSE}_{\text{TIME}} = 2,2 \times 10^{-2}$ e $2,8 \times 10^{-2}$ para a LSTM e BiLSTM, respectivamente. Para o espectro, o erro ao fim da PCF é $\text{RMSE}_{\text{SPEC}} = 1,3 \times 10^{-2}$ e 2×10^{-2} para a LSTM e BiLSTM, respectivamente.

Por fim, uma comparação do tempo computacional exigido por cada modelo, simulação numérica e rede neural. Escolhemos a rede BiLSTM nessa comparação. Para a dinâmica solitônica, o SSFM toma 1,30 s em média para simular uma única propagação, considerando 10.000 passos. No caso do supercontínuo, o SSFM demora 5,44 s em média para simular uma única propagação, com 20.000 passos. O modelo BiLSTM leva 5,35 s e 5,98 s em média para simular 100 casos de propagação solitônica e geração de supercontínuo, respectivamente. A grande vantagem de utilizar as redes neurais, como vimos, é evitar a necessidade de um passo de propagação pequeno para mitigar erros de cálculo, reduzindo com isso o tempo computacional do modelo.

Com este trabalho, realizamos uma extensa análise da aplicabilidade de redes neurais recorrentes no contexto de óptica não linear, descrevendo estruturas capazes de lidar com regimes não lineares complexos, destacando pelo menos duas, a LSTM e a BiLSTM, que, para os problemas tratados, podem ser usadas no lugar do SSFM.

5.1 Projetos futuros

Deste trabalho, podem ser derivados os seguintes problemas de pesquisa:

1. **Redes *feed-forward* multicamada:** O problema foi tratado aqui diretamente com redes recorrentes, vale utilizar uma rede perceptron multicamada, *feed-forward*, para

observar sua capacidade de prever os valores de propagação. Geralmente essas redes, mais simples, são ainda capazes de uma aproximação.

2. **Normalização Z-Score:** Utilizar uma normalização min-max no conjunto completo da base de dados pode acarretar o já mencionado *data leakage*, isto é, no vazamento de informação para o conjunto de testes. Em vista disso, a aplicação da normalização Z-Score visa a independência total dos conjuntos, tendo em conta que nesse caso a função sigmoide não poderá ser utilizada como função de ativação na saída da rede.
3. **Telecomunicações:** O problema tratado no corpo do texto foi o de uma fibra de somente 13 metros com propagação predominantemente não linear. Podemos adaptar a rede construída para o problema de telecomunicações, onde ocorre a transmissão de pulsos ópticos por milhares de quilômetros, e a propagação é dispersiva, ou linear.
4. **Lasers:** O caso apresentado neste trabalho foi de uma propagação de um pulso óptico, mas não abordamos a geração desse pulso, o que é geralmente feita empregando lasers. Na construção de um laser, ocorre a amplificação e absorção do feixe de luz repetidas vezes em um anel formado por fibra óptica. Nele, os efeitos não lineares e lineares se combinam de maneira complexa. Por isso, a aplicação de redes neurais pode contribuir nos estudos dessa área (FANG et al., 2023).
5. **Fibras multimodos:** A partir da rede desenvolvida por Salmela et al. (2021) para pulsos em fibra monomodo, Teġin et al. (2021) propõem o reuso dessa estrutura para predições de dinâmicas não lineares espaço-temporais em fibras multimodos.

REFERÊNCIAS

- ACHILLES, R.; BONFIGLIOLI, A. The early proofs of the theorem of Campbell, Baker, Hausdorff, and Dynkin. **Archive for History of Exact Sciences**, p. 295–358, 2012. Disponível em: <<https://doi.org/10.1007/s00407-012-0095-8>>. Citado na página 28.
- AGGARWAL, C. C. **Neural Networks and Deep Learning: A Textbook**. 1. ed. [S.l.]: Springer Cham, 2018. Citado 2 vezes nas páginas 31 e 38.
- AGRAWAL, G. P. **Nonlinear Fiber Optics**. [S.l.]: Academic Press, 2019. Citado 11 vezes nas páginas 18, 19, 21, 22, 23, 24, 25, 26, 27, 28 e 35.
- AGRAWAL, G. P. **Fiber-Optic Communication Systems**. [S.l.]: John Wiley & Sons, 2021. Citado na página 15.
- ALESHAMS, M.; ZARIFKAR, A.; SHEIKHI, M. Split-step Fourier transform method in modeling of pulse propagation in dispersive nonlinear optical fibers. In: **Proceedings of CAOL 2005. Second International Conference on Advanced Optoelectronics and Lasers, 2005**. [s.n.], 2005. v. 2, p. 124–126 vol. 2. Disponível em: <<https://doi.org/10.1109/CAOL.2005.1553936>>. Citado na página 15.
- AMIL, P.; SORIANO, M. C.; MASOLLER, C. Machine learning algorithms for predicting the amplitude of chaotic laser pulses. **Chaos**, v. 29, p. 113111, 2019. Disponível em: <<https://doi.org/10.1063/1.5120755>>. Citado na página 16.
- BAUMEISTER, T.; BRUNTON, S. L.; KUTZ, J. N. Deep learning and model predictive control for self-tuning mode-locked lasers. **Journal of the Optical Society of America B**, v. 35, n. 3, p. 617–626, 2018. Disponível em: <<https://doi.org/10.1364/JOSAB.35.000617>>. Citado na página 16.
- BIN, Y.; YANG, Y.; SHEN, F.; XIE, N.; SHEN, H. T.; LI, X. Describing video with attention-based bidirectional LSTM. **IEEE Transactions on Cybernetics**, v. 49, n. 7, p. 2631–2641, 2019. Disponível em: <<https://doi.org/10.1109/TCYB.2018.2831447>>. Citado na página 34.
- BISHOP, C. M. **Pattern Recognition and Machine Learning**. 1. ed. [S.l.]: Springer New York, NY, 2006. Citado na página 30.
- BLOW, K.; WOOD, D. Theoretical description of transient stimulated Raman scattering in optical fibers. **IEEE Journal of Quantum Electronics**, v. 25, n. 12, p. 2665–2673, 1989. Disponível em: <<https://doi.org/10.1109/3.40655>>. Citado na página 26.
- BORHANI, N.; KAKKAVA, E.; MOSER, C.; PSALTIS, D. Learning to see through multimode fibers. **Optica**, v. 5, n. 8, p. 960–966, 2018. Disponível em: <<https://doi.org/10.1364/OPTICA.5.000960>>. Citado na página 16.
- BRAHMS, C.; TRAVERS, J. C. **Luna.jl**. Zenodo, 2023. Disponível em: <<https://doi.org/10.5281/zenodo.10323704>>. Citado na página 28.

- CASTELLANI, C. E. S. **All-fibre wavelength versatile short pulsed laser sources**. Tese (Doutorado) — Imperial College London, U.K., 2013. Disponível em: <<https://doi.org/10.25560/24135>>. Citado 3 vezes nas páginas 15, 20 e 21.
- CHOLLET, F. et al. **Keras**. 2015. Disponível em: <<https://keras.io>>. Citado na página 39.
- COFFEN-SMOUT, S.; HERBERT, G. J. Submarine cables: a challenge for ocean management. **Marine Policy**, v. 24, n. 6, p. 441–448, 2000. Disponível em: <[https://doi.org/10.1016/S0308-597X\(00\)00027-0](https://doi.org/10.1016/S0308-597X(00)00027-0)>. Citado na página 15.
- DEITERDING, R.; GLOWINSKI, R.; OLIVER, H.; POOLE, S. A reliable split-step Fourier method for the propagation equation of ultra-fast pulses in single-mode optical fibers. **Journal of Lightwave Technology**, v. 31, n. 12, p. 2008–2017, 2013. Disponível em: <<https://doi.org/10.1109/JLT.2013.2262654>>. Citado 2 vezes nas páginas 24 e 28.
- DONAHUE, J.; HENDRICKS, L. A.; ROHRBACH, M.; VENUGOPALAN, S.; GUADARRAMA, S.; SAENKO, K.; DARRELL, T. Long-term recurrent convolutional networks for visual recognition and description. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 39, n. 4, p. 677–691, 2017. Disponível em: <<https://doi.org/10.1109/TPAMI.2016.2599174>>. Citado na página 33.
- DUBIETIS, A.; COUAIRON, A.; GENTY, G. Supercontinuum generation: introduction. **Journal of the Optical Society of America B**, v. 36, n. 2, p. SG1–SG3, 2019. Disponível em: <<https://doi.org/10.1364/JOSAB.36.000SG1>>. Citado na página 22.
- DUDLEY, J.; TAYLOR, R. **Supercontinuum Generation in Optical Fibers**. [S.l.]: Cambridge University Press, 2010. Citado 2 vezes nas páginas 26 e 28.
- DUDLEY, J. M.; GENTY, G.; COEN, S. Supercontinuum generation in photonic crystal fiber. **Reviews of Modern Physics**, v. 78, p. 1135–1184, 2006. Disponível em: <<https://doi.org/10.1103/RevModPhys.78.1135>>. Citado na página 53.
- FANG, Y.; HAN, H.-B.; BO, W.-B.; LIU, W.; WANG, B.-H.; WANG, Y.-Y.; DAI, C.-Q. Deep neural network for modeling soliton dynamics in the mode-locked laser. **Optics Letters**, v. 48, n. 3, p. 779–782, 2023. Disponível em: <<https://doi.org/10.1364/OL.482946>>. Citado 3 vezes nas páginas 16, 35 e 60.
- FRIGO, M.; JOHNSON, S. The design and implementation of FFTW3. **Proceedings of the IEEE**, v. 93, n. 2, p. 216–231, 2005. Disponível em: <<https://doi.org/10.1109/JPROC.2004.840301>>. Citado na página 27.
- GAMBLING, W. The rise and rise of optical fibers. **IEEE Journal of Selected Topics in Quantum Electronics**, v. 6, n. 6, p. 1084–1093, 2000. Disponível em: <<https://doi.org/10.1109/2944.902157>>. Citado 3 vezes nas páginas 18, 20 e 21.
- GENTY, G.; SALMELA, L.; DUDLEY, J. M.; BRUNNER, D.; KOKHANOVSKIY, A.; KOBTSEV, S.; TURITSYN, S. K. Machine learning and applications in ultrafast photonics. **Nature Photonics**, v. 15, p. 91–101, 2021. Disponível em: <<https://doi.org/10.1038/s41566-020-00716-4>>. Citado na página 16.

GIL-MOLINA, A.; CASTAÑEDA, J. A.; LONDONO-GIRALDO, D. F.; GABRIELLI, L. H.; CÁRDENAS, A. M.; FRAGNITO, H. L. High-order dispersion mapping of an optical fiber. **Optics Express**, v. 28, n. 3, p. 4258–4273, 2020. Disponível em: <<https://doi.org/10.1364/OE.379512>>. Citado na página 22.

GRAVES, A.; JAITLEY, N.; MOHAMED, A.-r. Hybrid speech recognition with deep bidirectional LSTM. In: **2013 IEEE Workshop on Automatic Speech Recognition and Understanding**. [s.n.], 2013. p. 273–278. Disponível em: <<https://doi.org/10.1109/ASRU.2013.6707742>>. Citado na página 34.

GRAVES, A.; SCHMIDHUBER, J. Framewise phoneme classification with bidirectional LSTM networks. In: **Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005**. [s.n.], 2005. v. 4, p. 2047–2052 vol. 4. Disponível em: <<https://doi.org/10.1109/IJCNN.2005.1556215>>. Citado na página 34.

GREFF, K.; SRIVASTAVA, R. K.; KOUTNÍK, J.; STEUNEBRINK, B. R.; SCHMIDHUBER, J. LSTM: A search space odyssey. **IEEE Transactions on Neural Networks and Learning Systems**, v. 28, n. 10, p. 2222–2232, 2017. Disponível em: <<https://doi.org/10.1109/TNNLS.2016.2582924>>. Citado na página 33.

HECHT, J. **City of Light: The Story of Fiber Optics**. [S.l.]: Oxford University Press, 1999. (Sloan Technology). Citado na página 15.

HEIDT, A. M.; FEEHAN, J. S.; PRICE, J. H. V.; FEURER, T. Limits of coherent supercontinuum generation in normal dispersion fibers. **Journal of the Optical Society of America B**, v. 34, n. 4, p. 764–775, 2017. Disponível em: <<https://doi.org/10.1364/JOSAB.34.000764>>. Citado na página 53.

HOCHREITER, S.; SCHMIDHUBER, J. Long Short-Term Memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, 1997. Disponível em: <<https://doi.org/10.1162/neco.1997.9.8.1735>>. Citado 2 vezes nas páginas 16 e 32.

HOLLENBECK, D.; CANTRELL, C. D. Multiple-vibrational-mode model for fiber-optic Raman gain spectrum and response function. **Journal of the Optical Society of America B**, v. 19, n. 12, p. 2886–2892, 2002. Disponível em: <<https://doi.org/10.1364/JOSAB.19.002886>>. Citado na página 26.

HOUDT, G. V.; MOSQUERA, C.; NÁPOLES, G. A review on the long short-term memory model. **Artificial Intelligence Review**, v. 53, n. 10, p. 5929–5955, 2020. Disponível em: <<https://doi.org/10.1007/s10462-020-09838-1>>. Citado 2 vezes nas páginas 32 e 33.

HUA-QIANG, Z.; PENG, W.; WEN-JUN, L.; YI-LEI, Y.; ZHI-JING, X.; JIAN, L. 19-fs pulse generated by supercontinuum compression. **Chinese Physics B**, v. 25, n. 2, p. 024209, 2016. Disponível em: <<https://doi.org/10.1088/1674-1056/25/2/024209>>. Citado na página 53.

HULT, J. A fourth-order Runge–Kutta in the interaction picture method for simulating supercontinuum generation in optical fibers. **Journal of Lightwave Technology**, v. 25, n. 12, p. 3770–3775, 2007. Disponível em: <<https://doi.org/10.1109/JLT.2007.909373>>. Citado na página 28.

- JADON, A.; PATIL, A.; JADON, S. A comprehensive survey of regression based loss functions for time series forecasting. **arXiv preprint arXiv:2211.02989**, 2022. Disponível em: <<https://doi.org/10.48550/arXiv.2211.02989>>. Citado 2 vezes nas páginas 44 e 45.
- JANIESCH, C.; ZSCHECH, P.; HEINRICH, K. Machine learning and deep learning. **Electronic Markets**, v. 31, p. 685–695, 2021. Disponível em: <<https://doi.org/10.1007/s12525-021-00475-2>>. Citado 3 vezes nas páginas 29, 30 e 31.
- JAUREGUI, C.; LIMPET, J.; TÜNNERMANN, A. High-power fibre lasers. **Nature Photonics**, v. 7, p. 861–867, 2013. Disponível em: <<https://doi.org/10.1038/nphoton.2013.273>>. Citado na página 18.
- JORDAN, M. I.; MITCHELL, T. M. Machine learning: Trends, perspectives, and prospects. **Science**, v. 349, n. 6245, p. 255–260, 2015. Disponível em: <<https://doi.org/10.1126/science.aaa8415>>. Citado na página 16.
- KAPRON, F. P.; KECK, D. B.; MAURER, R. D. Radiation losses in glass optical waveguides. **Applied Physics Letters**, v. 17, n. 10, p. 423–425, 1970. Disponível em: <<https://doi.org/10.1063/1.1653255>>. Citado na página 20.
- KECK, D.; MAURER, R.; SCHULTZ, P. On the ultimate lower limit of attenuation in glass optical waveguides. **Applied Physics Letters**, v. 22, n. 7, p. 307–309, 1973. Disponível em: <<https://doi.org/10.1063/1.1654649>>. Citado na página 20.
- KLEINERT, S.; TAJALLI, A.; NAGY, T.; MORGNER, U. Rapid phase retrieval of ultrashort pulses from dispersion scan traces using deep neural networks. **Optics Letters**, v. 44, n. 4, p. 979–982, 2019. Disponível em: <<https://doi.org/10.1364/OL.44.000979>>. Citado na página 16.
- KOBTSEV, S. M. Perspective paper: Can machine learning become a universal method of laser photonics? **Optical Fiber Technology**, v. 65, p. 102626, 2021. ISSN 1068-5200. Disponível em: <<https://doi.org/10.1016/j.yofte.2021.102626>>. Citado na página 16.
- KOHONEN, T. **Self-Organizing Maps**. 3. ed. Springer Berlin, Heidelberg, 2001. Disponível em: <<https://doi.org/10.1007/978-3-642-56927-2>>. Citado na página 30.
- KOKHANOVSKIY, A.; BEDNYAKOVA, A.; KUPRIKOV, E.; IVANENKO, A.; DYATLOV, M.; LOTKOV, D.; KOBTSEV, S.; TURITSYN, S. Machine learning-based pulse characterization in figure-eight mode-locked lasers. **Optics Letters**, v. 44, n. 13, p. 3410–3413, 2019. Disponível em: <<https://doi.org/10.1364/OL.44.003410>>. Citado na página 16.
- KOROBKO, D.; KAMYNNIN, V.; SALGANSKI, M.; SYSOLIATIN, A.; ZHLUKTOVA, I.; ZVEREV, A.; ZOLOTOVSKII, I.; TSVETKOV, V. (invited) control of supercontinuum generation due to soliton propagation in fibers with varying dispersion. **Optik**, v. 287, p. 171032, 2023. Disponível em: <<https://doi.org/10.1016/j.ijleo.2023.171032>>. Citado na página 53.
- KREBBERS, R.; LIU, N.; JAHROMI, K. E.; NEMATOLLAHI, M.; BANG, O.; WOYESSA, G.; PETERSEN, C. R.; ROOIJ, G. van; HARREN, F. J. M.; KHODABAKHSH, A.; CRISTESCU, S. M. Mid-infrared supercontinuum-based Fourier transform spectroscopy for plasma analysis. **Scientific Reports**, v. 12, n. 9642, 2022. Disponível em: <<https://doi.org/10.1038/s41598-022-13787-w>>. Citado na página 53.

KROGH, A. What are artificial neural networks? **Nature Biotechnology**, v. 26, p. 195–197, 2008. Disponível em: <<https://doi.org/10.1038/nbt1386>>. Citado 2 vezes nas páginas 29 e 30.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, p. 436–444, 2015. Disponível em: <<https://doi.org/10.1038/nature14539>>. Citado na página 32.

LILICRAP, T. P.; MARRIS, L.; SANTORO, A.; MARRIS, L.; AKERMAN, C. J.; HINTON, G. Backpropagation and the brain. **Nature Reviews Neuroscience**, v. 21, p. 335–346, 2020. Disponível em: <<https://doi.org/10.1038/s41583-020-0277-3>>. Citado na página 30.

LIN, Q.; AGRAWAL, G. P. Raman response function for silica fibers. **Optics Letters**, v. 31, n. 21, p. 3086–3088, 2006. Disponível em: <<https://doi.org/10.1364/OL.31.003086>>. Citado na página 26.

MARTINS, G. R.; SILVA, L. C. B.; SEGATTO, M. E. V.; ROCHA, H. R. O.; CASTELLANI, C. E. S. Design and analysis of recurrent neural networks for ultrafast optical pulse nonlinear propagation. **Optics Letters**, v. 47, n. 21, p. 5489–5492, 2022. Disponível em: <<https://doi.org/10.1364/OL.472267>>. Citado 2 vezes nas páginas 16 e 36.

MARTINS, G. R.; SILVA, L. C. B.; SEGATTO, M. E. V.; ROCHA, H. R. O.; CASTELLANI, C. E. S. Simultaneous temporal and spectral evolution of ultrafast optical pulse propagation using a single bidirectional LSTM network. **Optical Fiber Technology**, v. 82, p. 103636, 2024. Disponível em: <<https://doi.org/10.1016/j.yofte.2023.103636>>. Citado na página 36.

MATLAB. **Version 9.13.0.2166757 (R2022b)**. Natick, Massachusetts: The MathWorks Inc., 2022. Citado na página 35.

NÄRHI, M.; SALMELA, L.; TOIVONEN, J.; BILLET, C.; DUDLEY, J. M.; GENTY, G. Machine learning analysis of extreme events in optical fibre modulation instability. **Nature Communications**, v. 9, n. 1, p. 4923, 2018. Disponível em: <<https://doi.org/10.1038/s41467-018-07355-y>>. Citado na página 16.

NGUYEN-THE, Q.; MATSUURA, M.; KISHI, N. WDM-to-OTDM conversion using supercontinuum generation in a highly nonlinear fiber. **IEEE Photonics Technology Letters**, v. 26, n. 18, p. 1882–1885, 2014. Disponível em: <<https://doi.org/10.1109/LPT.2014.2339932>>. Citado na página 53.

OLIARI, V.; GOOSSENS, S.; HÄGER, C.; LIGA, G.; BÜTLER, R. M.; HOUT, M. van den; HEIDE, S. van der; PFISTER, H. D.; OKONKWO, C.; ALVARADO, A. Revisiting efficient multi-step nonlinearity compensation with machine learning: An experimental demonstration. **Journal of Lightwave Technology**, v. 38, n. 12, p. 3114–3124, 2020. Disponível em: <<https://doi.org/10.1109/JLT.2020.2994220>>. Citado na página 15.

OLIVEIRA, J. R.; MOURA, M. A.; HICKMANN, J. M.; GOMES, A. S. L. Self-steepening of optical pulses in dispersive media. **Journal of the Optical Society of America B**, v. 9, n. 11, p. 2025–2027, 1992. Disponível em: <<https://doi.org/10.1364/JOSAB.9.002025>>. Citado na página 25.

PATHAK, J.; HUNT, B.; GIRVAN, M.; LU, Z.; OTT, E. Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. **Physical Review Letters**, v. 120, p. 024102, 2018. Disponível em: <<https://doi.org/10.1103/PhysRevLett.120.024102>>. Citado na página 16.

REDMAN, P.; ZATORSKA, M.; PAWLOWSKI, A.; SZULC, D.; MAJCHROWSKA, S.; TARNOWSKI, K. **gnlse-python: Open Source Software to Simulate Nonlinear Light Propagation In Optical Fibers**. 2021. Citado na página 28.

ROGEL-SALAZAR, J. The Gross–Pitaevskii equation and Bose–Einstein condensates. **European Journal of Physics**, v. 34, n. 2, p. 247, 2013. Disponível em: <<https://dx.doi.org/10.1088/0143-0807/34/2/247>>. Citado na página 27.

ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. **Psychological Review**, v. 65(6), p. 386–408, 1958. Disponível em: <<https://doi.org/10.1037/h0042519>>. Citado na página 29.

SAGHAEI, H. Supercontinuum source for dense wavelength division multiplexing in square photonic crystal fiber via fluidic infiltration approach. **Radioengineering**, v. 26, n. 1, p. 16–22, 2017. Disponível em: <<https://doi.org/10.13164/re.2017.0016>>. Citado na página 53.

SAINATH, T. N.; VINYALS, O.; SENIOR, A.; SAK, H. Convolutional, long short-term memory, fully connected deep neural networks. In: **2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. [s.n.], 2015. p. 4580–4584. Disponível em: <<https://doi.org/10.1109/ICASSP.2015.7178838>>. Citado na página 33.

SALMELA, L.; LAPRE, C.; DUDLEY, J. M.; GENTY, G. Machine learning analysis of rogue solitons in supercontinuum generation. **Scientific Reports**, v. 10, p. 9596, 2020. Disponível em: <<https://doi.org/10.1038/s41598-020-66308-y>>. Citado na página 16.

SALMELA, L.; TSIPINAKIS, N.; FOI, A.; BILLET, C.; DUDLEY, J. M.; GENTY, G. Predicting ultrafast nonlinear dynamics in fibre optics with a recurrent neural network. **Nature Machine Learning**, v. 3, p. 344–354, 2021. Disponível em: <<https://doi.org/10.1038/s42256-021-00297-z>>. Citado 5 vezes nas páginas 16, 27, 35, 36 e 60.

SCHNEIDER, T. Self- and cross-phase modulation. In: _____. **Nonlinear Optics in Telecommunications**. Berlin, Heidelberg: Springer, 2004. p. 143–165. Disponível em: <https://doi.org/10.1007/978-3-662-08996-5_6>. Citado na página 23.

SCHUSTER, M.; PALIWAL, K. Bidirectional recurrent neural networks. **IEEE Transactions on Signal Processing**, v. 45, n. 11, p. 2673–2681, 1997. Disponível em: <<https://doi.org/10.1109/78.650093>>. Citado na página 34.

SHI, X.; CHEN, Z.; WANG, H.; YEUNG, D.-Y.; WONG, W.-k.; WOO, W.-c. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In: **Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1**. MIT Press, 2015. (NIPS'15), p. 802–810. Disponível em: <<https://doi.org/10.48550/ARXIV.1506.04214>>. Citado 3 vezes nas páginas 33, 34 e 42.

- SOLLI, D. R.; ROPERS, C.; KOONATH, P.; JALALI, B. Optical rogue waves. **Nature**, v. 450, n. 7172, p. 1054–1057, 2007. Disponível em: <<https://doi.org/10.1038/nature06402>>. Citado na página 24.
- SUKHOIVANOV, I. A.; IAKUSHEV, S. O.; SHULIKA, O. V.; ANDRADE-LUCIO, J. A.; DíEZ, A.; ANDRÉS, M. Supercontinuum generation at 800 nm in all-normal dispersion photonic crystal fiber. **Optics Express**, v. 22, n. 24, p. 30234–30250, 2014. Disponível em: <<https://doi.org/10.1364/OE.22.030234>>. Citado na página 53.
- TEĞİN, U.; DİNÇ, N. U.; MOSER, C.; PSALTIS, D. Reusability report: Predicting spatiotemporal nonlinear dynamics in multimode fibre optics with a recurrent neural network. **Nature Machine Intelligence**, v. 3, p. 387–391, 2021. Disponível em: <<https://doi.org/10.1038/s42256-021-00347-6>>. Citado 2 vezes nas páginas 16 e 60.
- VOUMARD, T.; LUDWIG, M.; WILDI, T.; AYHAN, F.; BRASCH, V.; VILLANUEVA, L. G.; HERR, T. Simulating supercontinua from mixed and cascaded nonlinearities. **APL Photonics**, v. 8, n. 3, p. 036114, 2023. Disponível em: <<https://doi.org/10.1063/5.0135252>>. Citado na página 28.
- WERBOS, P. Backpropagation through time: what it does and how to do it. **Proceedings of the IEEE**, v. 78, n. 10, p. 1550–1560, 1990. Disponível em: <<https://doi.org/10.1109/5.58337>>. Citado na página 30.
- YCAS, G.; MASER, D.; HICKSTEIN, D. **Pynlo: Nonlinear optics modelling for python**. 2015. Citado na página 28.
- ZAHAVY, T.; DIKOPOLTSEV, A.; MOSS, D.; HAHAM, G. I.; COHEN, O.; MANNOR, S.; SEGEV, M. Deep learning reconstruction of ultrashort pulses. **Optica**, v. 5, n. 5, p. 666–673, 2018. Disponível em: <<https://doi.org/10.1364/OPTICA.5.000666>>. Citado na página 16.
- ZHELTIKOV, A. Analytical insights into self-phase modulation: beyond the basic theory. **Optics Express**, v. 26, n. 13, p. 17571–17577, 2018. Disponível em: <<https://doi.org/10.1364/OE.26.017571>>. Citado na página 23.
- ZORIN, I.; GATTINGER, P.; EBNER, A.; BRANDSTETTER, M. Advances in mid-infrared spectroscopy enabled by supercontinuum laser sources. **Optics Express**, v. 30, n. 4, p. 5222–5254, 2022. Disponível em: <<https://doi.org/10.1364/OE.447269>>. Citado na página 53.

Apêndices

APÊNDICE A – EQUAÇÕES DE MAXWELL

Para todos os fenômenos eletromagnéticos, inclusive fenômenos ópticos, nos apoiamos nas equações de Maxwell para o desenvolvimento matemático necessário. São elas:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad \nabla \cdot \mathbf{D} = \rho_f, \quad (\text{A.1})$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}, \quad \nabla \cdot \mathbf{B} = 0, \quad (\text{A.2})$$

onde \mathbf{E} é o vetor campo elétrico, e \mathbf{H} o vetor campo magnético. E também \mathbf{D} e \mathbf{B} são os vetores densidade de fluxo elétrico e magnético correspondentes. Ainda, \mathbf{J} é o vetor densidade de corrente, e ρ_f a densidade de carga elétrica. Sendo a fibra ótica um meio dielétrico, temos portanto um meio sem cargas livres. Nesse caso $\mathbf{J} = \mathbf{0}$, e $\rho_f = 0$.

Os vetores densidade de fluxo se relacionam com seus respectivos campos através das Equações A.3 e A.4,

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}, \quad (\text{A.3})$$

$$\mathbf{B} = \mu_0 \mathbf{H} + \mathbf{M}, \quad (\text{A.4})$$

onde ε_0 é a permissividade elétrica do vácuo, μ_0 é a permeabilidade magnética do vácuo, e \mathbf{P} e \mathbf{M} são as polarizações elétrica e magnética. Na fibra ótica, $\mathbf{M} = \mathbf{0}$, então $\mathbf{B} = \mu_0 \mathbf{H}$.

A partir daí, buscamos obter a equação de onda que descreve a propagação de luz na fibra,

$$\begin{aligned} \nabla \times (\nabla \times \mathbf{E}) &= \nabla \times \left(-\frac{\partial \mathbf{B}}{\partial t} \right), \\ &= -\mu_0 \frac{\partial (\nabla \times \mathbf{H})}{\partial t}, \\ &= -\mu_0 \frac{\partial^2 \mathbf{D}}{\partial t^2}, \\ &= -\frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} - \mu_0 \frac{\partial^2 \mathbf{P}}{\partial t^2}, \end{aligned}$$

e foi usado a relação $\mu_0 \varepsilon_0 = 1/c^2$. No caso de interesse nos efeitos não lineares, a polarização \mathbf{P} inclui uma parcela linear \mathbf{P}_L , e outra não linear \mathbf{P}_{NL} ,

$$\mathbf{P}(\mathbf{r}, t) = \mathbf{P}_L(\mathbf{r}, t) + \mathbf{P}_{NL}(\mathbf{r}, t). \quad (\text{A.5})$$

E aplicando a identidade vetorial,

$$\nabla \times \nabla \times \mathbf{E} \equiv \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E}, \quad (\text{A.6})$$

obtemos a seguinte equação de onda,

$$\nabla^2 \mathbf{E} - \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} = \mu_0 \frac{\partial^2 \mathbf{P}_L}{\partial t^2} + \mu_0 \frac{\partial^2 \mathbf{P}_{NL}}{\partial t^2}. \quad (\text{A.7})$$

Por fim, simplificamos a Equação (A.7) realizando a transformação de Fourier de $\mathbf{E}(\mathbf{r}, t)$, e relacionando os demais termos com esse vetor. Isso resume a equação de Helmholtz abaixo,

$$\nabla^2 \tilde{\mathbf{E}} + \varepsilon(\omega) k_0^2 \tilde{\mathbf{E}} = 0, \quad (\text{A.8})$$

onde $k_0 = \omega/c$, $\tilde{\mathbf{E}}(\mathbf{r}, \omega)$ é a transformada de Fourier $\mathbf{E}(\mathbf{r}, t)$, e a constante dielétrica é dada por $\varepsilon(\omega) = 1 + \chi^{(1)} + \varepsilon_{NL}$, com as contribuições lineares e não lineares da constante dielétrica, relacionado as respectivas polarizações.

Nas equações acima, assume-se muitas simplificações, cobrindo de maneira introdutória o assunto, recomenda-se a leitura de outras fontes para se aprofundar no tema.

APÊNDICE B – EQUAÇÃO DE PROPAGAÇÃO NÃO LINEAR

Partindo da equação de Helmholtz, vamos desenvolver matematicamente a Equação (2.2),

$$\nabla^2 \tilde{\mathbf{E}} + \varepsilon(\omega) k_0^2 \tilde{\mathbf{E}} = 0. \quad (\text{B.1})$$

Assumindo uma solução do tipo,

$$\tilde{\mathbf{E}}(\mathbf{r}, \omega - \omega_0) = F(x, y) \tilde{A}(z, \omega - \omega_0) \exp(i\beta_0 z), \quad (\text{B.2})$$

onde $\tilde{A}(z, \omega)$ é uma função que varia lentamente em z , $F(x, y)$ é a distribuição modal, e β_0 é o número de onda. Substituindo (B.2) em (B.1), e conhecendo o operador laplaciano ∇^2 ,

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}, \quad (\text{B.3})$$

por meio do método de separação de variáveis, temos

$$\frac{\partial^2 [F \tilde{A} \exp(i\beta_0 z)]}{\partial x^2} + \frac{\partial^2 [F \tilde{A} \exp(i\beta_0 z)]}{\partial y^2} + \frac{\partial^2 [F \tilde{A} \exp(i\beta_0 z)]}{\partial z^2} + \varepsilon(\omega) k_0^2 F \tilde{A} \exp(i\beta_0 z) = 0,$$

$$\tilde{A} \exp(i\beta_0 z) \left(\frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2} + \varepsilon(\omega) k_0^2 F \right) + F \frac{\partial^2 [\tilde{A} \exp(i\beta_0 z)]}{\partial z^2} = 0,$$

$$\tilde{A} \exp(i\beta_0 z) \left(\frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2} + \varepsilon(\omega) k_0^2 F \right) + F \exp(i\beta_0 z) \left(\frac{\partial^2 \tilde{A}}{\partial z^2} + 2i\beta_0 \frac{\partial \tilde{A}}{\partial z} - \beta_0^2 \tilde{A} \right) = 0,$$

$$\tilde{A} \left(\frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2} + \varepsilon(\omega) k_0^2 F \right) + F \left(2i\beta_0 \frac{\partial \tilde{A}}{\partial z} - \beta_0^2 \tilde{A} \right) = 0,$$

$$\tilde{A} \left(\frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2} + \varepsilon(\omega) k_0^2 F \right) - \tilde{A} (F \tilde{\beta}^2) + F \left(2i\beta_0 \frac{\partial \tilde{A}}{\partial z} - \beta_0^2 \tilde{A} \right) + F (\tilde{A} \tilde{\beta}^2) = 0,$$

$$\tilde{A} \left(\frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2} + [\varepsilon(\omega) k_0^2 - \tilde{\beta}^2] F \right) + F \left(2i\beta_0 \frac{\partial \tilde{A}}{\partial z} + (\tilde{\beta}^2 - \beta_0^2) \tilde{A} \right) = 0,$$

onde a segunda derivada $\partial^2 \tilde{A} / \partial z^2$ foi eliminada visto que assumimos $\tilde{A}(z, \omega)$ variando lentamente em função de z . Portanto,

$$\frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2} + [\varepsilon(\omega) k_0^2 - \tilde{\beta}^2] F = 0, \quad (\text{B.4})$$

$$2i\beta_0 \frac{\partial \tilde{A}}{\partial z} + (\tilde{\beta}^2 - \beta_0^2) \tilde{A} = 0, \quad (\text{B.5})$$

em que $F(x, y)$ pode ser aproximada por $F \approx \exp[-(x^2 + y^2)/w^2]$, e $A(z, \omega)$ é obtida da Equação (B.5).

A seguir, fazemos algumas considerações,

$$\begin{aligned}\tilde{\beta}(\omega) &\approx \beta(\omega) + \Delta\beta_0, \\ \tilde{\beta}^2 - \beta_0^2 &\approx 2\beta_0(\tilde{\beta} - \beta_0),\end{aligned}$$

que, substituindo na Equação (B.5), obtemos

$$\frac{\partial \tilde{A}}{\partial z} = i[\beta(\omega) + \Delta\beta_0 - \beta_0]\tilde{A}, \quad (\text{B.6})$$

e $\beta(\omega)$ é normalmente dado em uma série de Taylor,

$$\beta(\omega) = \beta_0 + (\omega - \omega_0)\beta_1 + \frac{1}{2}(\omega - \omega_0)^2\beta_2 + \frac{1}{6}(\omega - \omega_0)^3\beta_3 + \dots, \quad (\text{B.7})$$

onde

$$\beta_m = \left(\frac{d^m \beta}{d\omega^m} \right)_{\omega=\omega_0} \quad (m = 1, 2, \dots). \quad (\text{B.8})$$

É suficiente para o caso de propagação tratado aqui valores de $m \leq 3$, termos de maior ordem podem ser ignorados, sendo assim

$$\frac{\partial \tilde{A}}{\partial z} = i[(\omega - \omega_0)\beta_1 + \frac{1}{2}(\omega - \omega_0)^2\beta_2 + \frac{1}{6}(\omega - \omega_0)^3\beta_3 + \Delta\beta_0]\tilde{A}, \quad (\text{B.9})$$

e portanto, finalmente fazendo a transformação inversa de Fourier na Equação (B.9), usando $(\omega - \omega_0) \leftrightarrow i\partial/\partial t$,

$$\frac{\partial A}{\partial z} + \beta_1 \frac{\partial A}{\partial t} + i\frac{\beta_2}{2} \frac{\partial^2 A}{\partial t^2} - \frac{\beta_3}{6} \frac{\partial^3 A}{\partial t^3} = i\Delta\beta_0 A, \quad (\text{B.10})$$

com o termo $\Delta\beta_0$ incluindo os efeitos de atenuação e não linearidade, sendo substituído por α e γ , respectivamente,

$$\frac{\partial A}{\partial z} + \beta_1 \frac{\partial A}{\partial t} + i\frac{\beta_2}{2} \frac{\partial^2 A}{\partial t^2} - \frac{\beta_3}{6} \frac{\partial^3 A}{\partial t^3} + \frac{\alpha}{2} A = i\gamma(\omega_0)|A|^2 A. \quad (\text{B.11})$$

Por fim, fazendo a substituição pelo frame de referência $T = t - z/v_g \equiv t - \beta_1 z$. Pela regra da cadeia, temos

$$\begin{aligned}\frac{\partial A}{\partial t} &= \frac{\partial A}{\partial T} \frac{\partial T}{\partial t} + \frac{\partial A}{\partial z} \frac{\partial z}{\partial t} = \frac{\partial A}{\partial T}, \\ \frac{\partial A}{\partial z} &= \frac{\partial A}{\partial z} \frac{\partial z}{\partial z} + \frac{\partial A}{\partial T} \frac{\partial T}{\partial z} = \frac{\partial A}{\partial z} - \beta_1 \frac{\partial A}{\partial T},\end{aligned}$$

e finalmente, obtemos a equação de propagação não linear

$$\frac{\partial A}{\partial z} + i\frac{\beta_2}{2} \frac{\partial^2 A}{\partial T^2} - \frac{\beta_3}{6} \frac{\partial^3 A}{\partial T^3} + \frac{\alpha}{2} A = i\gamma|A|^2 A. \quad (\text{B.12})$$

APÊNDICE C – DECIBEL

O uso da unidade de medida relativa adimensional decibel (dB) na engenharia facilita as comparações de valores com grande *range*, como potência e frequência. Outra aplicação comum são as perdas de transmissão em telecomunicações. Uma razão $R > 0$ pode ser escrita em decibéis através da Equação C.1,

$$R_{\text{dB}} = 10 \log_{10} R. \quad (\text{C.1})$$

Nesse caso, com o logaritmo de base dez, conseguimos comparar diferentes ordens de grandezas de maneira simples, conforme exibido na Tabela 11.

Tabela 11 – Comparação de valores em dB.

R_{dB}	R
30	= 1.000
20	= 100
10	= 10
6	≈ 4
3	≈ 2
0	= 1
-3	$\approx 1/2$
-6	$\approx 1/4$
-10	= 1/10
-20	= 1/100
-30	= 1/1.000

Fonte: Próprio autor.