


# REGRESSÃO SIMBÓLICA APLICADA À INTERROGAÇÃO DE GRADES DE PERÍODO LONGO UTILIZANDO REFLECTOMETRIA ÓPTICA NO DOMÍNIO DO TEMPO

Bruno Santana de Souza

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica, do Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, CEFET/RJ, como parte dos requisitos necessários à obtenção do título de mestre.

Orientador(a): Diego Barreto Haddad
Coorientador(a): Marco Aurélio Jucá

Rio de Janeiro

Março de 2025

REGRESSÃO SIMBÓLICA APLICADA À INTERROGAÇÃO DE GRADES DE PERÍODO LONGO UTILIZANDO REFLECTOMETRIA ÓPTICA NO DOMÍNIO DO TEMPO

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica, do Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, CEFET/RJ, como parte dos requisitos necessários à obtenção do título de mestre.

Bruno Santana de Souza

Banca Examinadora:

---

Presidente, Professor D.Sc. Diego Barreto Haddad (CEFET/RJ) (Orientador(a))

---

Professor D.Sc. Marco Aurélio Jucá da Silva Junior (CEFET/RJ) (Coorientador(a))

---

Professora D.Sc. Maria Aparecida Gonçalves Martinez (CEFET/RJ)

---

Professor D.Sc. Alexandre Bessa dos Santos (UFJF)

Rio de Janeiro

Março de 2025

Ficha catalográfica elaborada pela Biblioteca Central do CEFET/RJ

Ficha catalográfica elaborada pela Biblioteca Central do CEFET/RJ

S729 Souza, Bruno Santana de
Regressão simbólica aplicada à interrogação de grades de período longo utilizando reflectometria óptica no domínio do tempo / Bruno Santana de Souza. — 2025.
75f. + apêndice : il. color. , enc.

Dissertação (Mestrado) Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, 2025.
Bibliografia : f. 63-75
Orientador: Diego Barreto Haddad
Coorientador: Aurelio Marco Jucá

1. Inteligência artificial. 2. Detectores ópticos. 3. Sensoriamento remoto. 4. Aprendizado do computador. I. Haddad, Diego Barreto. (Orient.). II. Jucá, Marco Aurelio (Coorient.). III. Título.

CDD 006.3

Elaborada pela bibliotecária Tania Mello – CRB/7 nº 5507/04

# DEDICATÓRIA

Este trabalho é dedicado aos que passaram pela minha vida e contribuíram com conhecimentos.

# AGRADECIMENTOS

Agradeço ao Professor Dr. Diego Barreto Haddad pela excelente orientação que me foi concedida; agradeço ao professor Dr. Maurício Henrique Costa que com toda seriedade deixou um grande legado disponibilizando materiais de estudo valiosos, me conduzindo ao aprimoramento acadêmico; agradeço a professora Dra. Maria Aparecida que identificou e contribuiu para que eu fortalecesse uma característica que me agregou muito no início da jornada de pós-graduação, contribuindo para os resultados satisfatórios até então; agradeço ao professor Dr. Leandro Samyn pela paciência e sensibilidade no ato de ensinar; agradeço ao professor Dr. Thiago Americano, que valorizou a essência do conhecimento e esforço transmitido pelo aluno; agradeço ao professor Dr. João Dias, no qual admirei o grande apreço pela arte de lecionar; agradeço ao professor Dr. Marco Aurélio, que me incentivou a buscar soluções inovadoras; agradeço ao professor Dr. Gabriel de Araújo que manteve proximidade aos discentes e me conduziu transmitindo as vivências da carreira acadêmica. Agradeço aos coordenadores e todos que, de certa forma, contribuíram com gestos de generosidade durante a minha desafiadora jornada.

# RESUMO

**Regressão Simbólica Aplicada à Interrogação de Grades de Período Longo utilizando Reflectometria Óptica no Domínio do Tempo**

A regressão simbólica tem despertado o interesse por parte da comunidade científica devido à capacidade de automatização do processo de descoberta do conhecimento, contribuindo para a extração de *"insights"* e formulação de hipóteses, economizando tempo no processo de inferência, bem como recursos preciosos para os pesquisadores. Os modelos simbólicos (equações analíticas) são mais facilmente interpretáveis do que os modelos de aprendizado de máquina convencionais pois, ao invés de simplesmente ajustar parâmetros em uma equação predefinida, como em regressões lineares ou polinomiais, a regressão simbólica utiliza algoritmos para descobrir formatos funcionais mais flexíveis, com maior robustez na descrição do mapeamento entre variável dependente e variáveis independentes. Nesta dissertação, a regressão simbólica foi aplicada à interrogação óptica de um sensor, objetivando a obtenção de expressões analíticas capazes de prever a concentração salina de uma solução. A linguagem de programação Python será utilizada para modelagem e simulações, por meio da biblioteca PySR. Foi obtida uma fórmula bem mais concisa do que outra previamente proposta, com um erro padrão de 0,54 g/L e com um erro absoluto médio de 0,38 g/L.

Palavras-chave: Regressão simbólica; Interrogação de sensores ópticos; Aprendizagem de máquina.

# ABSTRACT


## Symbolic Regression Applied to Long-Period Grating Interrogation using Optical Time-Domain Reflectometry

Symbolic regression has aroused interest on the part of the scientific community due to its ability to automate the knowledge discovery process, contributing to the absorption of “insights” and formulation of hypotheses, while saving time and valuable resources during inference. Symbolic models (analytical equations) are generally more interpretable than conventional machine learning models because, instead of merely fitting parameters to a predefined equation as in linear or polynomial regressions. Symbolic regression uses algorithms to discover more flexible functional forms, with greater robustness in describing the mapping between dependent and independent variables. In this dissertation, symbolic regression was applied to the optical interrogation of a sensor, aiming to obtain analytical expressions capable of predicting the salinity concentration of a solution. The Python programming language was used for modeling and simulation through the PySR library. A significantly more concise formula than a previously proposed model was obtained, with a standard error of 0.54 g/L and a mean absolute error of 0.38 g/L.



Keywords: Symbolic regression, Interrogation of optical sensors, Machine learning.

# LISTA DE ILUSTRAÇÕES


Figura 1 – Diagrama de blocos da estrutura desta dissertação. 17

Figura 2 – Esquemático representativo de FBG gêmeas. 30

Figura 3 – Versatilidade da técnica de regressão simbólica. 41

Figura 4 – Exemplificação de cenários gráficos para a regressão simbólica com a presença de dados ruidosos. 42

Figura 5 – Exemplificação da migração entre “ilhas” correspondente ao laço externo da biblioteca PySR. 44

Figura 6 – Estrutura de dados em árvore para representação de fórmulas-exemplo; cruzamento (à esquerda) e mutação (à direita). 46

Figura 7 – Diagrama da configuração experimental contendo o sensor LPG e o equipamento OTDR. 51

Figura 8 – Valores de salinidade da água medidos e estimados em (g/L) no conjunto de dados de teste. 58

# LISTA DE TABELAS


Tabela 1 – Comparação de trabalhos relacionados contendo abordagem ampla, exclusiva e recente acerca de sensores LPGs com aplicabilidade diversa. 26

Tabela 2 – Comparação de trabalhos relacionados à Modelos Híbridos (FGB / LPG) pertinentes à aplicabilidade diversa. 27

Tabela 3 – Comparação do Sensoriamento de Salinidade para os Trabalhos de Pesquisa Relacionados em Instrumentação Óptica. 36

Tabela 4 – Comparação autoral e contributiva para os trabalhos de pesquisa relacionados à regressão simbólica. 40

Tabela 5 – Fórmulas de complexidade menor ou igual a 13 geradas pelo método de regressão simbólica. 55

Tabela 6 – Avaliação numérica dos resultados da regressão. 59

Tabela 7 – Comparação entre características das equações. 59

# LISTA DE ABREVIATURAS E SIGLAS

| | |
|---|---|
| AI | *Artificial Intelligence* - Inteligência Artificial |
| ASE | *Amplified Spontaneous Emission* - Emissão Espontânea Simplificada |
| AWG | *Arrayed Waveguided Grating* - Grade Guiada por Ondas em Matriz |
| BBS | *Broadband Source* - Fonte de Banda Larga |
| CFBG | *Chirped Fiber Bragg Grating* - Grade de Bragg em Fibra com Chirp ou com Período Variável |
| CPU | *Central Processing Unit* - Unidade Central de Processamento |
| FBG | *Fiber Bragg Grating* - Grade de Bragg de Fibra |
| FIGP | *Filter-Introduced GP* - Programação Genética Introduzida por Filtro |
| FTB | *Fiber Termination Box* - Caixa de Terminação de Fibra |
| GB | *Gigabyte* - Unidade de Medida de Armazenamento Digital |
| GECCO | *Genetic and Evolutionary Computation Conference* - Conferência de Computação Genética e Evolutiva |
| GP | *Genetic Programing* - Programação Genética |
| HBM | Hottinger Baldwin Messtechnik |
| IOT | *Internet of Things* - Internet das Coisas |
| LC | *Lightning Cure* - Cura Relâmpago |
| LPG | *Long Period Grating* - Grade de Período Longo |
| LTS | *Long Term Support* - Suporte de Longo Prazo |
| MA | *Massachusetts* |
| ML | *Machine Learning* - Aprendizado de Máquina |
| MZI | *Mach-Zehnder Interferometer* - Interferômetro de Mach-Zehnder |
| OSA | *Optical Spectrum Analyzer* - Analisador de Espectro Óptico |
| OTDR | *Optical Time-Domain Reflectometer* - Reflectômetro Óptico no Domínio do Tempo |
| PC | *Personal Computer* - Computador Pessoal |
| PMLB | *Penn Machine Learning Benchmarks* - Benchmarks de Aprendizado de Máquina da Penn |
| PMMA | *Polymethyl methacrylate* - Polimetilmetacrilato |

| | |
|---|---|
| QSAR | *Quantitative Structure-Activity Relationships* - Relações de Atividades de Estruturas Quantitativas |
| QSPR | *Quantitative Structure-Property Relationships* - Relações de Atividades de Estruturas Quantitativas |
| RAM | *Random Access Memory* - Memória de Acesso Aleatório |
| SBMO | Simpósio Brasileiro de Micro-ondas e Optoeletrônica |
| SCIMED | *Scientist-Machine Equation Detector* - Detector de Equação Cientista-Máquina |
| SFBG | *Sensing Fiber Bragg Grating* - Rede de Bragg de Fibra Sensorial |
| SLED | *Superluminescent Diode* - Diodo Emissor de Luz Superluminescente |
| SMF | *Single Mode Fiber* - Fibra Monomodo |
| SOI | *Silicon-on-Insulator* - Silício Sobre Isolante |
| SR | *Symbolic Regression* - Regressão Simbólica |
| SRBENCH | *Stream Reasoning Benchmark* - Referência para Raciocínio de Fluxo |
| UV | *Ultraviolet* - Ultravioleta |

# SUMÁRIO


| | | |
|---|---|---|
| **1** | **INTRODUÇÃO** | **14** |
| 1.1 | CONTEXTUALIZAÇÃO | 14 |
| 1.2 | OBJETIVO | 16 |
| 1.3 | ESTRUTURA DO TRABALHO | 17 |
| 1.4 | PUBLICAÇÃO ASSOCIADA A ESTA DISSERTAÇÃO | 18 |
| **2** | **FUNDAMENTOS TEÓRICOS DE INTERROGAÇÃO ÓPTICA** | **19** |
| 2.1 | ÍNDICE DE REFRAÇÃO E A LPG | 24 |
| 2.2 | TÉCNICAS POPULARES DE INTERROGAÇÃO ÓPTICA | 26 |
| 2.2.1 | INTERROGAÇÃO ÓPTICA BASEADA EM SISTEMAS HÍBRIDOS | 26 |
| 2.2.2 | TÉCNICAS DE INTERROGAÇÃO BASEADAS EM FILTROS ÓPTICOS | 28 |
| 2.2.3 | TÉCNICAS DE INTERROGAÇÃO BASEADAS EM INTERFERÔMETROS DE LUZ | 28 |
| 2.2.4 | FBGs GÊMEOS EM TÉCNICA DE INTERROGAÇÃO | 29 |
| 2.2.5 | TÉCNICA DE INTERROGAÇÃO BASEADA EM OSA | 32 |
| 2.3 | PROCEDIMENTOS PARA INTERROGAÇÃO DE SENSORES ALTERNATIVOS OU INOVADORES | 33 |
| 2.4 | ESCOPO DESTA DISSERTAÇÃO | 34 |
| 2.5 | CONSIDERAÇÕES FINAIS AOS MÉTODOS DE INTERROGAÇÃO POPULARES E INÉDITOS | 37 |
| **3** | **REGRESSÃO SIMBÓLICA - FUNDAMENTOS** | **38** |
| 3.1 | A REGRESSÃO SIMBÓLICA | 38 |
| 3.1.1 | REGRESSÃO SIMBÓLICA E A BIBLIOTECA PySR | 42 |
| 3.1.2 | REDES NEURAIS PARA REGRESSÃO SIMBÓLICA | 44 |
| 3.1.3 | ALGORITMOS GENÉTICOS E A INTELIGÊNCIA ARTIFICIAL | 45 |
| 3.2 | INTERPRETABILIDADE | 48 |

3.2.1 MÉTODOS CAIXA PRETA, CAIXA CINZA e CAIXA BRANCA 48
3.3 COMENTÁRIOS FINAIS 50

**4 METODOLOGIA E RESULTADOS 51**
4.1 MÉTODO DE INTERROGAÇÃO ÓPTICA ORIGINAL 51
4.2 APLICAÇÃO DA REGRESSÃO SIMBÓLICA 53
4.3 RESULTADOS 57
4.4 COMENTÁRIOS FINAIS 60

**5 CONCLUSÕES E TRABALHOS FUTUROS 61**

**REFERÊNCIAS 62**

**A APÊNDICE 76**
A.1 CÓDIGO PYTHON 76
A.2 FÓRMULAS OBTIDAS 77

# 1- INTRODUÇÃO

## 1.1- CONTEXTUALIZAÇÃO

Sistemas físicos ou sistemas complexos são rotineiramente aplicados e alvo de controle, seja na pesquisa científica, na engenharia e na matemática aplicada. A extração, por meio de modelagem computacional, de equações que regem os dados, ou mesmo a obtenção de seus respectivos padrões de comportamento, tem o potencial de aprofundar nossa compreensão acerca destes sistemas [1]. Simulações computacionais de sistemas físicos costumeiramente partem de modelos estabelecidos e codificados em equações diferenciais, algébricas ou em combinações dessas, as quais incorporam por vezes primeiros princípios (ou seja, princípios fundamentais) tais como balanço de força ou de energia. Tais simulações revelam soluções que, na maioria das vezes, não apresentam uma expressão analítica sucinta. Por óbvio, expressões analíticas mais simples facilitam a extração de *insights* ou percepções (tais como, relações de causa e efeito ou correlações inversas entre determinado atributo e a variável de interesse) acerca do sistema em foco. Tais *insights* são mais facilmente extraídos quando modelos com alta interpretabilidade encontram-se disponíveis. Esta é uma propriedade importante em ciência e tecnologia; afinal, um oráculo que prediz corretamente a função de toda proteína, ou as séries de rotas bioquímicas em um determinado animal, ou mesmo que antecipe resultados de ensaios de tração teria um grande impacto na ciência e na tecnologia. Entretanto, dificilmente, neste caso os cientistas estariam completamente satisfeitos, já que normalmente eles almejam compreender como tal oráculo efetuou as suas predições [2].

Recentemente, avanços em aprendizagem de máquina têm se valido de dados experimentais com o objetivo de efetuar uma descoberta automática de modelos concisos e interpretáveis [3]. Uma das frentes mais promissoras destes avanços consiste na regressão simbólica, a qual conjuga otimização evolucionária (vale dizer, programação genética) multiobjetivos com o propósito de oferecer uma ferramenta de descoberta de conhecimento orientado a dados [4].

Grosso modo, a regressão simbólica consiste na manipulação de expressões ma-

temáticas para a descoberta de funções capazes de descrever um determinado conjunto de dados. Poder-se-ia pensar que a regressão simbólica inaugura um novo paradigma, o qual pretende revelar uma relação matemática subjacente não trivial, somente a partir de dados observados, sem o concurso de conhecimento *a priori* ou de princípios fundamentais [5]. Cumpre entretanto ressaltar que não é possível a descoberta de novas teorias sem a imposição de determinadas hipóteses acerca do processo, de vieses indutivos ou de algumas restrições (de fato, este é um dogma fundamental da teoria de aprendizado estatístico, o qual é coerente com o chamado teorema da ausência de "*free lunch*") [6, 7].

A capacidade humana de, por meio de indução, extrair relações matemáticas a partir dos dados é bastante limitada, com os casos históricos tais como a formulação das leis de Newton ou de Kepler exigindo grande engenhosidade por parte do cientista. Em contraposição, o uso de ferramentas automatizadas com grande expressividade pode permitir a descoberta de fórmulas interpretáveis [8] a partir de uma base de dados, seja ela experimental, seja ela advinda de uma simulação computacional. Neste último caso, a obtenção de uma fórmula concisa capaz de emular uma relação entre os dados de entrada e o resultado de uma modelagem computacional pode evitar simulações desnecessárias no futuro, economizando tempo no processo de inferência e recursos escassos.

Em resumo, quando as soluções que emulam bem o mapeamento entrada-saída são interpretáveis, elas podem facilitar a extração de relações de causa e efeito, realçar o impacto de determinado atributo ou mesmo facilitar o controle futuro de determinados sistemas. Um exemplo de formulação interpretável é uma representação por meio de uma fórmula matemática concisa (o que não costuma ocorrer, por exemplo, quando redes neurais são adotadas) [2]. Portanto, é correto afirmar que a regressão simbólica pode ser usada como um mecanismo de interpretabilidade.

A aplicação e o resultado das técnicas de regressão simbólica é dependente do sistema físico alvo. Neste trabalho, pretendemos modelar de maneira eficiente um sistema de interrogação de sensores ópticos proposto por Juca em [9, 10]. Para contextualizar melhor o leitor, cumpre explicar o que seriam: *i*) a interrogação; *ii*) a modalidade de sensoriamento adotada e *iii*) a grandeza a ser estimada.

Concisamente, a interrogação é uma técnica de interpretação dos dados adquiridos por um sensor, a qual geralmente dá origem a uma estimativa de uma grandeza desejável. Tipicamente, dados egressos de um sensor são processados de forma a dar origem à estimativa de uma variável de interesse, tal como pressão ou temperatura.

Tal tipo de estimativa desperta grande atratividade por parte da indústria e da pesquisa científica.

Mais especificamente, o sensor adotado nesta dissertação é um sensor óptico, o qual se vale do efeito de grades de difração inscritas em fibras ópticas, particularmente, as grades de período longo (LPGs, do inglês, *long period gratings*) fabricadas com a técnica de arco elétrico, como descrito em [9]. As técnicas de interrogação óptica, especialmente, aquelas que envolvem fibras ópticas, desempenham um papel fundamental na indústria e na pesquisa devido à sua alta sensibilidade e precisão na medição de diversas quantidades físicas, como temperatura, pressão e concentração de substâncias químicas [11, 12]. A flexibilidade e facilidade de integração das fibras ópticas permitem seu uso em uma ampla gama de aplicações industriais e de pesquisa, tendo como exemplo à aplicabilidade em campos como medicina e biologia para monitoramento em tempo real de parâmetros biológicos e diagnóstico médico [13].

Por sua vez, a variável de interesse, cuja estimativa será o objeto de todo o processo de interrogação desta dissertação, é a salinidade de uma solução. A estimativa da salinidade de uma solução é fundamental em diversas aplicações industriais, especialmente na indústria alimentícia, onde a qualidade e o sabor dos produtos podem ser afetados pela concentração de sal [14, 15]. Na indústria de tratamento de água, a salinidade é um parâmetro crítico para garantir a eficiência dos processos de dessalinização e purificação [16]. Além disso, em setores como o de petróleo e gás, a salinidade da água produzida é um indicador-chave para a avaliação da viabilidade econômica da operação. Em resumo, a estimativa precisa da salinidade é essencial para garantir a qualidade dos produtos, a eficiência dos processos e a sustentabilidade das operações industriais [17]. Por fim, cabe ressaltar que a base de dados mensurados que alimentará as técnicas de regressão simbólica foi disponibilizada pelos autores de [9, 10].

## 1.2- OBJETIVO

O sensoriamento de salinidade pode ser efetuado por meio de interrogação óptica. O principal objetivo deste trabalho consiste na obtenção de fórmulas concisas que consigam estimar a salinidade de soluções por meio de dados adquiridos via sensoriamento

óptico. A obtenção de tais fórmulas facilitaria a disseminação de técnicas de interrogação óptica para este propósito.

## 1.3- ESTRUTURA DO TRABALHO

Os próximos capítulos encontram-se estruturados do seguinte modo:

A estrutura desta dissertação encontra-se na Figura 1. O Capítulo 1 contém a introdução desta dissertação. O Capítulo 2 contém os fundamentos teóricos acerca dos sensores nos quais são inscritas as grades. Por sua vez, o Capítulo 3 expõe informações relevantes sobre o método de regressão simbólica e suas aplicações contemporâneas diversas, motivando a proposta desta dissertação. Já o Capítulo 4 apresenta os resultados experimentais, evidenciando a robustez do método de regressão simbólica seguida dos resultados desta pesquisa. Por fim, o Capítulo 5 estabelece as conclusões, bem como as linhas de investigação consideradas mais promissoras desta dissertação.

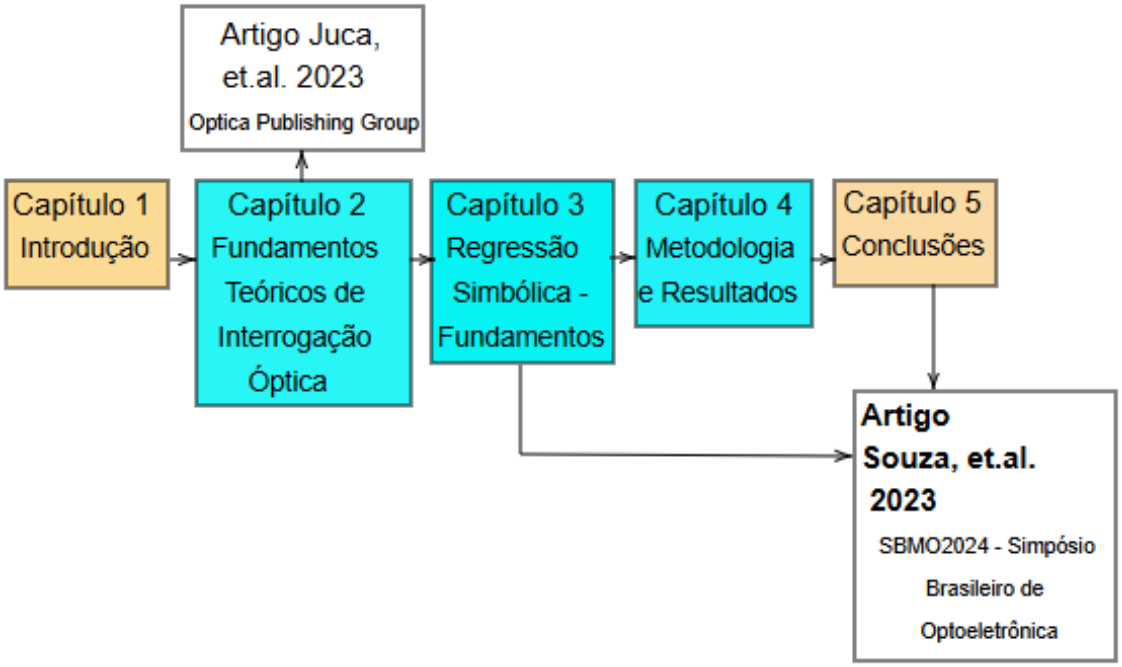



Figura 1 – Diagrama de blocos da estrutura desta dissertação.

## 1.4- PUBLICAÇÃO ASSOCIADA A ESTA DISSERTAÇÃO

**Souza, Bruno Santana de** *et al*, “Optical Interrogation by Symbolic Regression for Water Salinity Prediction”, XXI Simpósio Brasileiro de Micro-Ondas e Optoeletrônica (SBMO 2024), Bento Gonçalves - RS, p. 244-348, 2024.

Cabe ressaltar que o artigo recebeu uma premiação no SBMO 2024, tendo sido considerado o segundo melhor artigo na área de optoeletrônica.

# 2- FUNDAMENTOS TEÓRICOS DE INTERROGAÇÃO ÓPTICA

Na introdução, vimos que as técnicas de interrogação óptica, especialmente aquelas empregando fibras ópticas, têm um papel crucial tanto na indústria quanto na pesquisa devido à sua capacidade de medir com alta sensibilidade e precisão várias propriedades físicas, como temperatura, pressão e a concentração de substâncias químicas [11].

Desde a década de 1960, percebeu-se que as fibras ópticas poderiam ser uma alternativa vantajosa para o sensoriamento. Tal vantagem pode ser exemplificada por um sensor fotônico desenvolvido em 1967 que continua disponível no mercado, oferecendo "desempenho incomparável na medida de vibrações sem contato" [18]. Este sensor ilustra o fato de que, em diversas aplicações, sensores eletrônicos tradicionais não superam os sensores de fibra óptica devido à suas vantagens, dentre elas: imunidade a interferências eletromagnéticas e de radiofrequência, inércia química[1] [13, 19], resistência a choque e a abrasão [9], alta sensibilidade, leveza, confiabilidade a longo prazo[2], operação segura em ambientes onde existam materiais perigosos, resistência à corrosão química, alta largura de banda, alta robustez, possibilidade de realização de sensoriamento remoto e possibilidade de multiplexação de sensores, tendo em vista que codificam a informação mensurada em termos de comprimento de onda [9, 20].

Posteriormente, com o desenvolvimento de fibras ópticas monomodo, começou-se a pensar em sua inserção em interferômetros, o que prometia grandes vantagens na área de engenharia em relação aos seus antecessores de espaço livre montados em mesas ópticas [18].

Sensores a fibra óptica recaem em duas categorias amplas. Na primeira delas, conhecida como sensores extrínsecos, a fibra simplesmente guia a luz para uma região de sensoriamento, onde o sinal óptico incide em outro meio no qual é modulado [3]. A

[1] Consiste na resistência de uma substância reagir com outras substâncias; tal característica permite a monitorização de reações químicas sem interferir nas condições do processo, garantindo dados confiáveis, o que gera estabilidade em processos industriais, outro viés é a capacidade da inércia química evitar a degradação do sensor, mantendo a precisão das medições, favorecendo o desempenho em altas temperaturas.

[2] A confiabilidade a longo prazo justifica um investimento inicial mais alto devido aos benefícios econômicos e operacionais ao longo do tempo, caracterizando boa relação custo-benefício [12].

[3] Os parâmetros que podem ser modulados de um sinal óptico são: amplitude, frequência, fase e polarização.

luz é então coletada pela mesma ou por outra fibra após ser modulada pelo parâmetro de interesse e retornada a uma localização remota para efeitos de processamento. Em contraste, “sensores intrínsecos” mantêm a luz dentro da fibra, de modo que o parâmetro externo modula a onda eletromagnética à medida em que ela se propaga na fibra [18].

Três equipamentos importantes para a interrogação óptica são o analisador de espectro óptico (OSA, do inglês *Optical Spectrum Analyzer*) [21, 22], Braggmeter e o reflectômetro óptico no domínio do tempo (OTDR, do inglês *Optical Time Domain Reflectometer*). Cabe ressaltar que o uso do OTDR com a finalidade de interrogação simboliza uma metodologia inovadora, portanto, o mesmo não é tradicionalmente utilizado para essa finalidade.

Por sua vez, o OSA é um instrumento utilizado para analisar e caracterizar o espectro óptico de um sinal de luz. Ele é capaz de medir a intensidade de luz em diferentes comprimentos de onda ao longo do espectro óptico, permitindo uma análise detalhada da distribuição espectral da luz. Os OSAs são frequentemente adotados em diversas aplicações, incluindo telecomunicações, pesquisa em fibras ópticas, análise de *lasers*, entre outras, onde a compreensão precisa do espectro óptico é essencial para o projeto, operação e manutenção de sistemas ópticos complexos.

Braggmeters medem o comprimento de onda de Bragg de sensores FBGs (do inglês, *Fiber Bragg Gratings*) [23, 24]. Em 2017, o modelo FS22 da marca HBM foi projetado para interrogar sensores FBGs em ambientes industriais adversos. Este equipamento oferece medições estáticas e dinâmicas com taxas de amostragem de até 1000 S/s, dependendo do modelo. Em termos de capacidade de medição, braggmeters são comumente utilizados para medir deformação [25], pressão, torque, vibração, temperatura etc.

O OTDR é um equipamento óptico crucial para inspecionar fibras ópticas em redes de comunicação. Funcionando como um radar óptico, ele emite pulsos de luz na fibra e mede o tempo e a intensidade dos sinais refletidos, permitindo a identificação de eventos de reflexão, como conectores, emendas e falhas, ao longo da fibra. Essa análise detalhada da fibra ajuda os técnicos a localizar e solucionar problemas, garantindo um desempenho ideal da rede óptica. O objetivo de um OTDR é detectar, localizar e medir perdas em qualquer ponto de um enlace de fibra óptica. O OTDR precisa acessar apenas uma extremidade do enlace e atua como um sistema de radar unidimensional. Ao disponibilizar um detalhamento do traço óptico das fibras em teste, é possível obter

uma representação gráfica de todo o enlace de fibra óptica. Além da operação normal do equipamento, identificou-se a possibilidade de realização de interrogação óptica a partir dele [26, 27].

A abordagem do uso do dispositivo OTDR com a finalidade de interrogação foi proposta por Yuan *et al*. em 2013 [28], no referido artigo um sistema de medição por fibra óptica baseado em reflexão de Fresnel para monitoramento remoto foi proposto para medição do índice de refração usando a técnica de reflectometria óptica no domínio do tempo como método de interrogação. O índice de refração circundante a uma longa distância pôde ser medido facilmente usando esse sistema de sensor, que operava com base no teste de intensidade de reflexão Fresnel da interface fibra amostra. Esse sistema constituiu uma configuração simples e fácil de manusear. Resultados experimentais mostraram que o alcance desta medição pôde atingir cerca de 100,8 km, e o das sensibilidades do índice de refração foram de 38,71 dB/RIU a 304,89 dB/RIU (*Refractive Index Unit* ou Unidade de Índice de Refração). Já em 2011, Bravo *et al*. [29] propuseram e demonstraram experimentalmente um sistema no qual o equipamento comercial OTDR modelo FTB-7423B-B interrogou um sensor localizado a 253 km de distância conectado por uma fibra monomodo (Single Mode Fiber - SMF28). Em 2005, Falate [30] apresentou uma nova técnica para analisar um sensor LPG, onde a sensibilidade da grade foi determinada pelas variações na amplitude do pico de ressonância do comprimento de onda. Para medir essas variações de amplitude, utilizou-se um reflectômetro de domínio óptico. Realizou-se medições de curvatura para aplicar esse método e avaliar a sensibilidade da grade em relação a esse parâmetro físico. A técnica ofereceu vantagens como medições reflexivas que não são afetadas pela temperatura e a possibilidade de usar múltiplos sensores LPGs, simultaneamente. Em 2003, Valente *et al* [31]. propuseram e demonstraram o uso de um dispositivo comercial de reflectometria óptica no domínio do tempo com multiplexação[4] simultânea no tempo e no comprimento de onda de sensores de grade de Bragg de fibra óptica. A realização de tais sistemas proporcionou uma maneira barata de interrogar várias centenas de sensores em uma fibra única.

Uma dimensão importante para avaliação de uma proposta de interrogação óptica consiste no custo da solução. É sabido que um equipamento OTDR, frequentemente presente em laboratórios de óptica, caracteriza-se por ser menos dispendioso em relação a um OSA [21, 22]. Assim, um sistema de interrogação óptica que utilize um OTDR é

[4] É a técnica de combinar múltiplos sinais em um único meio de transmissão, otimizando o uso de recursos.

preferível a um que exija um OSA [9]. Embora existam OSAs portáteis que apresentam boa relação custo-benefício no curto prazo; contudo, possuem comumente a desvantagem econômica ao longo prazo comparativamente ao OTDR, o qual oferece maior durabilidade e menor necessidade de manutenção.

Nesta dissertação, nos restringiremos a sensores baseados em grades de difração, sendo elas as FBGs e as LPGs, inclusive sensores híbridos, que consistem na combinação de ambos.

No caso da atuação como um sensor de deformação, um sensor FBG possui características únicas. Por exemplo, quando a fibra é esticada ou comprimida, o sensor FBG irá medir deformação. Isso acontece porque a deformação de uma fibra óptica ocasionará uma variação da constante fotoelástica da fibra e uma mudança no período da microestrutura, o que resulta numa variação do comprimento de onda de Bragg.

Em termos conceituais, o comprimento de onda de Bragg é um parâmetro caracterizador do sensor FBG. Este parâmetro precisa ser detectado com precisão a fim de determinar parâmetros físicos, tais como temperatura ou deformação.

A sensibilidade a mudanças de grandezas físicas como temperatura, deslocamentos ou até mesmo à presença de elementos químicos junto às fibras ópticas [32], vinculam-se a categorização de sensores de modulação de frequência. Alterações nessas grandezas físicas modificam a frequência da luz retroespalhada ou refletida (característica comumente presente nos sensores de grade de Bragg), tal variação pode então ser detectada e traduzida.

O principal contribuinte para uma mudança no comprimento de onda de Bragg é a variação do índice de refração da sílica, induzido pelo efeito termo-óptico. Também há uma contribuição da expansão térmica, a qual altera o período da microestrutura [9]. Este efeito é, no entanto, marginal devido ao baixo coeficiente de expansão térmica da sílica [33].

Além disso, a tecnologia do sensor FBG demonstra uma grande capacidade de multiplexação e uma habilidade em promover medições absolutas [34]. Isto o torna uma alternativa às tecnologias de sensores elétricos convencionais [35, 36, 34].

Em termos descritivos, o sensor FBG é uma microestrutura periódica que atua como um espelho seletivo de comprimento de onda [37]. Isso significa que, se a luz de uma fonte de banda larga for injetada na fibra óptica, apenas a luz dentro de uma região espectral muito estreita, centrada no comprimento de onda de Bragg, será refletida. A

luz remanescente irá continuar seu caminho pela fibra óptica sem perda considerável, mantendo assim, sua capacidade de interrogação [38]. O sensor FBG é uma estrutura recíproca e, como tal, sempre irá refletir o mesmo comprimento de onda de Bragg, independente do sentido da luz.

O comprimento de onda de Bragg $\lambda_B$ é essencialmente definido pelo período da microestrutura ou periodicidade espacial da modulação do índice de refração, cuja notação é $\Lambda$. Além disso, está relacionado com o índice de refração efetivo do modo de núcleo correspondente a: $n_{\text{eff}}$. Essa relação pode ser observada através da equação [39]:

$$\lambda_B = 2n_{\text{eff}}\Lambda. \tag{1}$$

Devido aos efeitos fotoelástico e termo-óptico, tanto esforços mecânicos longitudinais quanto variações na temperatura provocam o deslocamento do comprimento de onda de Bragg. A variação, (isto é, deslocamento) do comprimento de onda de Bragg em função de variações no comprimento da fibra ($\Delta\ell$) e de variações de temperatura ($\Delta T$) é regida pela equação [9]:

$$\Delta\lambda_{\text{B}} = 2\left(\Lambda\frac{\partial n_{\text{eff}}}{\partial l} + n_{\text{eff}}\frac{\partial\Lambda}{\partial l}\right)\Delta l + 2\left(\Lambda\frac{\partial n_{\text{eff}}}{\partial T} + n_{\text{eff}}\frac{\partial\Lambda}{\partial T}\right)\Delta T. \tag{2}$$

Por outro lado, a resposta de um sensor LPG a mudanças de temperatura pode ser modelada matematicamente através da equação:

$$\Delta\lambda = \frac{\partial\lambda}{\partial T}\Delta T \tag{3}$$

Correspondentemente a uma equação comumente usada para descrever a relação entre a mudança de temperatura e o deslocamento no comprimento de onda ressonante $\Delta\lambda$.

Nesse contexto de utilização da LPG como um sensor de temperatura é de particular importância o fato de esse “dispositivo óptico” acoplar a luz do modo de núcleo fundamental com os modos de casca em co-propagação, cuja atenuação é consideravelmente maior quando comparada à de sensores FBGs, devido a grandes perdas de espalhamento na interface com a casca.

Os sensores LPGs têm uma ampla gama de aplicações bem documentadas devido

às suas altas sensibilidades e baixa retrorreflexão [5]. Para permitir o acoplamento de luz, os índices de refração efetivos dos modos devem satisfazer a condição de casamento de fase [40], dada por:

$$\beta_1 - \beta_2 = \Delta\beta = \frac{2\pi}{\Lambda}, \tag{4}$$

onde $\Lambda$ é o período espacial do LPG, $\beta_1$ e $\beta_2$ são as constantes de propagação dos modos sendo acoplados [41], e $\Delta\beta$ é a constante de propagação diferencial.

Como a principal resposta do sensor LPG a potenciais medidas está no domínio espectral, a interrogação do sensor de fibra óptica é tradicionalmente realizada por meio da análise do espectro de transmissão do dispositivo (também chamada de análise espectral ou análise harmônica) especialmente usando um analisador de espectro óptico ou dispositivos de interrogação específicos [21].

## 2.1- ÍNDICE DE REFRAÇÃO E A LPG

LPGs são muito sensíveis ao índice de refração do meio externo circundante, tornando-os muito úteis como elemento central para a construção de sensores químicos e biomédicos [42]. A mudança no índice do ambiente altera o índice efetivo do modo de casca e isso, por sua vez, leva a deslocamentos do comprimento de onda nas ressonâncias do espectro de transmissão dos sensores LPGs.

Para a produção de sensores, essa característica de sensibilidade ao índice de refração vem sendo explorada pela indústria, visando, por exemplo, realizar a medição do nível de um fluído dentro de um recipiente ou monitorar a concentração da solução de uma substância num processo químico. Adicionalmente, LPGs podem ser usadas em sensores que realizam o monitoramento em tempo real da concentração de material em solução aquosa e de materiais que estejam em locais potencialmente explosivos, para o controle de qualidade da produção industrial [43].

Via de regra, sensores químicos e biológicos requerem alguma forma de camada sensível para melhorar a capacidade de detecção [44] caracterizando um “artifício” que consiste na aplicação de uma camada adicional na casca da fibra, promovendo alteração

[5] Fenômeno que consiste num tipo específico de reflexão da luz que ocorre quando os raios luminosos incidem numa superfície e regressam à fonte de luz.

no índice de refração do sensor e, consequentemente, na resposta espectral [45].

A resposta espectral relaciona-se com parâmetros externos (tais como: temperatura, índice de refração, etc) de forma precisa, convergindo ao conceito de “modelo simbólico”, tendo em vista que a saída do algoritmo é uma expressão simbólica que pode ser usada para prever a resposta espectral do sensor LPG, dado um conjunto de condições externas.

O sensor LPG admite a aplicação de técnica que seja um modelo matemático simples e robusto, que considera todos os parâmetros espectrais do LPG, como largura de banda, comprimento de onda de ressonância e deslocamento espectral [46].

O autor Esposito em 2021, vide Tabela 1, propõe um trabalho, o qual objetiva fornecer uma análise abrangente sobre as soluções de última geração sobre sensores químicos baseados em LPG e resumir as limitações atuais e as possibilidades futuras de pesquisa. Adicionalmente, o autor discorre sobre “Técnicas de aumento da sensibilidade” e “materiais sensíveis”. O autor é eficaz na abordagem sobre o princípio de funcionamento de sensores LPGs, inclusive desenvolve uma representação esquemática de sensor químico de fibra óptica composto por um LPG revestido com um material sensível, tais como materiais polieletrolíticos (ou seja, são polímeros), para a captura do analito alvo (o qual pode ser uma substância ou um componente químico) [47].

Em termos de aprimoramentos para melhoria de sensibilidade, Zhao em 2018 destaca que essa obtenção pode ser realizada por LPGs trabalhando em proximidade de um ponto de viragem de correspondência de fase[6], corroborando para a amplificação da resposta espectral. Além disso, a diminuição de espessura da casca acaba por ser outro fator preponderante. Portanto, a sensibilidade do sensor LPG geralmente é melhorada corroendo o diâmetro da casca ou selecionando o período ideal. Além disso, a combinação dos dois métodos para melhorar a sensibilidade do sensor também é estudada e as características do sensor são melhoradas alterando a estrutura do LPG ou o material revestido. Quando esta última abordagem considera a espessura do filme, a translação do modo pode ser induzida para aumentar ainda mais e melhorar a sensibilidade. Por fim, os sensores LPGs são amplamente utilizados na detecção biológica e química. Esses dispositivos são sensíveis a alterações no índice de refração [48]. Os respectivos artigos contendo detalhes contributivos e evidenciando, inclusive, aplicações relevantes estão elucidados na tabela 1.

[6]Refere-se a uma condição específica na qual a ressonância de acoplamento entre o modo núcleo e os modos da casca sofre uma inversão de tendência.

Tabela 1 – Comparação de trabalhos relacionados contendo abordagem ampla, exclusiva e recente acerca de sensores LPGs com aplicabilidade diversa. LPGs: Long Period Gratings.

| Referência | Contribuição | Aplicações |
|---|---|---|
| Esposito, F., 2021 [47] | O artigo promove uma revisão completa sobre o projeto, caracterização e fabricação de sensores químicos de fibra óptica baseado em tecnologia de sensor LPG. | Monitoramento de parâmetros ambientais, gases perigosos, salinidade, qualidade de alimentos e íons de metais pesados. |
| Zhao, X., 2018 [48] | Proporciona uma revisão abrangente dos princípios dos sensores LPGs, suas aplicações em sensores e telecomunicações, e as tendências futuras. Destaca a versatilidade desses sensores e a capacidade de medição de vários parâmetros. | Monitoramento biológico, químico, umidade, gases, índice de refração, detecção de ph. |

## 2.2- TÉCNICAS POPULARES DE INTERROGAÇÃO ÓPTICA

A busca por sistemas de medição que combinem precisão e robustez tem levado ao desenvolvimento de várias técnicas de interrogação óptica, cada uma com suas particularidades e vantagens.

### 2.2.1 INTERROGAÇÃO ÓPTICA BASEADA EM SISTEMAS HÍBRIDOS

No campo de sensores ópticos, a grande vantagem da utilização de FBGs, face aos LPGs, consiste na possibilidade de serem lidos em reflexão [49]. A interrogação de sensores baseados nessas tecnologias permite a utilização de microcontroladores no processamento pós-óptico, simplificando desta forma o processo de interrogação [50]. O uso da reflexão nos sensores FBGs torna sua integração com microcontroladores mais eficiente, reduzindo perdas ópticas, pois necessita apenas de uma única extremidade da fibra conectada à fonte e ao interrogador; outrossim, permite multiplexação facilitada por propiciar que múltiplos sensores sejam escritos em sequência por toda a extensão da mesma fibra ao utilizar um único interrogador. Cumpre destacar, o menor consumo de energia e maior robustez em aplicações de longo alcance (viabilizando a integração, por exemplo, do arduino ESP32 a sensores FBGs para transmitir dados em tempo real via Wi-Fi [51], Bluetooth ou LoRa).

É sabido que a combinação de potencialidades de detecção dos sensores LPGs,

Tabela 2 – Comparação de trabalhos relacionados à Modelos Híbridos (FGB / LPG) pertinentes à aplicabilidade diversa. LPGs: Long Period Gratings - Grades de Período Longo. MZI: Mach-Zehnder Interferometer - Interferômetro Mach-Zehnder.

| Referência | Tipo de Grade (Técnica) | Monitoramento |
|---|---|---|
| Possetti, G.R.C., *et al*., 2009 [54] | Emprega-se 2 (dois) sensores LPGs induzidos por um arco com período de 520 µm formando um Interferômetro de Mach-Zehnder, (do inglês, Mach–Zehnder interferometer ou MZI), além de 2 (dois) sais inorgânicos e 1 (um) orgânico (NaCl, KCl, NaCOOH), foram considerados concentrações na faixa de 0-150 g/L. | Monitora a medição de salinidade em uma solução aquosa. (Compara-se o uso da técnica de refractometria de fibras com medições obtidas usando um refractômetro de Abbe, bem como via condutividade térmica.) |
| Marin, E., *et al*., 2012 [55] | Propõe-se um sensor feito de uma grade dupla, ou seja, 1 (um) FBG padrão e 1 (um) LPG fotoescritos na mesma faixa da fibra. | Monitoramento de um processo de fabricação tendo aplicação na indústria de equipamento. |
| Gouveia, A.J.G., 2008 [49] | Aborda um sistema híbrido de grade em fibra. Sendo 1 (um) LPG e 2 (dois) FBGs são utilizados como esquema de medição simultânea. | Supervisiona parâmetros físicos ou biológicos e químicos. |
| Enriquez, D.A.C., *et al*., 2012 [56] | Desenvolve-se a configuração de uma grade de Bragg híbrida juntamente com um sensor de grade de período longo. | Inspeção e gestão da qualidade da água com aplicabilidade na indústria de alimentos e bebidas. |
| Berrettoni, C., *et al*., 2015 [53] | Foi proposto e demonstrado experimentalmente um novo sistema de detecção de fibra óptica utilizando uma configuração híbrida, ou seja, (FBG / LPG). | Aplicação em monitoramento químico e biológico. |

com as vantagens de interrogação advindas dos FBGs, são desejáveis por possibilitar a medição simultânea de dois parâmetros (como por exemplo: deformação e temperatura), conforme salientado em [52, 49]; outro exemplo, consiste na medição simultânea de temperatura e índice de refração, devidamente destacado em [53], inclusive, existem sistemas com essa proposição, vide Tabela 2.

Dentre o conjunto de vantagens de FBGs duplos, destaca-se a sensibilidade aprimorada, a qual ocorre porque o sinal diferencial entre os FBGs aumenta a resposta às variações de parâmetros externos, tais como temperatura, deformação, pressão e alteração nos índices de refração, considerando o FBG com inclinação (do inglês, *tilt*) [57, 58]; compensação de temperatura, visto que um FBG pode atuar como referência, ajudando a separar o efeito da temperatura do da deformação medida [59].

### 2.2.2 TÉCNICAS DE INTERROGAÇÃO BASEADAS EM FILTROS ÓPTICOS

Técnicas de interrogação de filtros ópticos empregam um filtro de inclinação para traduzir alterações no comprimento de onda em variações de intensidade. Esses métodos são dependentes das intensidades conhecidas de transmissão e rejeição dos filtros ópticos, as quais variam conforme o comprimento de onda. Ao mensurar tais variações de intensidade, torna-se possível realizar medições quantitativas de diversas grandezas, tais como temperatura, deformação e pressão [60].

Especificamente, para sensores de temperatura baseados em FBG, os filtros ópticos utilizados incluem: *i) Twin* FBG, *ii)* Grade de fibra de longo período, também conhecida como LPG e *iii)* Grade de fibra de Bragg com *chirp* CFBG.

### 2.2.3 TÉCNICAS DE INTERROGAÇÃO BASEADAS EM INTERFERÔMETROS DE LUZ

Técnicas de interrogação baseadas em interferômetros de luz oferecem várias vantagens, incluindo a capacidade de alcançar uma sensibilidade ajustável e maior do que sistemas de medição de índice de refração convencionais [61]; alta resolução e ampla largura de banda [60, 62]. No entanto, estas técnicas também vêm com algumas desvantagens. Dentre elas, o intervalo limitado de deslocamentos (quando comparado com o OSA, o qual pode operar sobre uma ampla faixa espectral [7], permitindo medições em uma ampla gama de comprimentos de onda com alta precisão). Outra desvantagem consiste nas mudanças de índice de refração ou variações de fase (que podem ser medidos antes que a resposta do sistema se torne ambígua ou não linear). Isso ocorre porque o princípio de funcionamento dos interferômetros está baseado na interferência coerente da luz, que pode resultar em padrões altamente oscilatórios e periodicidade limitada, o que significa possibilidade de inadequação à medição de certos parâmetros além de um certo intervalo. Além disso, as técnicas de interferometria são altamente sus-

[7] Um OSA como o Yokogawa AQ6370D cobre uma faixa espectral típica de 600 nm a 1700 nm. Comparativamente superior em relação ao interferômetro de Fabry-Pérot, o qual pode operar em faixas de 50 nm a 100 nm, enquanto um interferômetro de Michelson pode ser ainda mais restrito.

cetíveis às flutuações ambientais, comumente referidas como efeitos de desvanecimento do sinal, pois, conforme a luz avança pelo núcleo da fibra óptica, uma porção diminuta de energia penetra na casca em forma de uma onda evanescente, cuja intensidade decresce exponencialmente à medida que se distancia da interface entre a casca e o ambiente circundante. Em geral, a profundidade de penetração dessa onda no meio externo é da ordem de centenas de nanômetros [63]. Fatores externos, como vibrações, correntes de ar e variações de temperatura podem fortemente impactar a precisão e confiabilidade das medições obtidas por meio dessas técnicas [60].

Portanto, os autores de [64] são assertivos ao afirmarem que a aplicação dos sensores interferométricos é frequentemente restrita devido à dependência de temperatura e pela demodulação de sinal complexa.

### 2.2.4 FBGs GÊMEOS EM TÉCNICA DE INTERROGAÇÃO

A primeira grade de Bragg gêmea (ou *Twin Bragg Grating*) foi desenvolvida em 1980 sendo, na época, uma abordagem relativamente inovadora no campo da engenharia elétrica e optoeletrônica. São empregadas para aumentar a precisão e a estabilidade das medições, compensando as variações térmicas e outros tipos de interferências.

Esta técnica, envolve o uso de dois sensores FBGs com comprimentos de onda de Bragg análogos, conforme a Figura 2. Tal semelhança equipara-se ao espectro de reflexão de ambas no tocante ao processo de fabricação [12, 65].

Um único FBG do conjunto funciona como um filtro, enquanto o outro atua como um sensor [65]. A potência de saída depende do desacoplamento dos espectros de Bragg do filtro e do FBG sensoriador [12].

Trata-se de uma técnica de interrogação em que um espectro de banda larga de luz é projetado pela fonte de banda larga (BBS - *Broadband Source*) em uma fibra de modo único em direção a um FBG sensorial (SFBG - *Sensing Fiber Bragg Grating*).

O espectro, em seguida, chega ao filtro FBG, que reflete um comprimento de onda específico de acordo com seu comprimento de onda de Bragg. Já que o filtro e o sensor FBG são idênticos e têm o mesmo comprimento de onda de Bragg, a potência que atinge o SFBG é minimizada em condições ambientais normais. Assim, a potência

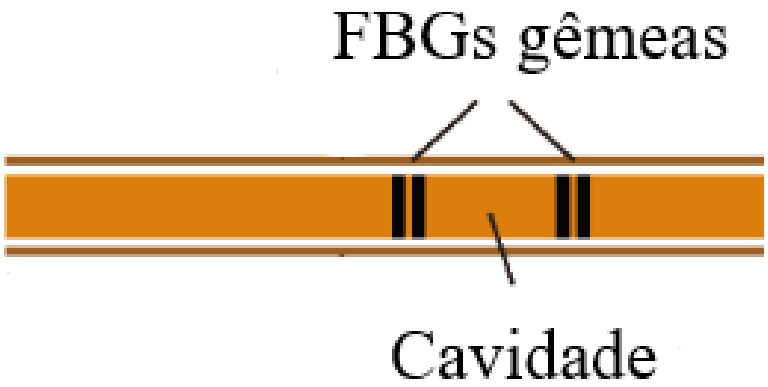



Figura 2 – Configuração de FBGs Gêmeas.

refletida pelo SFBG também é reduzida, resultando em uma saída de tensão mínima no fotodiodo. Contudo, quando há uma variação na medição (como temperatura ou tensão), o comprimento de onda de Bragg do SFBG se desloca, causando o desacoplamento dos espectros de Bragg do SFBG e do filtro FBG. A potência refletida e a tensão de saída tornam-se então dependentes do grau de desacoplamento dos espectros [12].

Segundo Allil *et al.* 2021 [66], a técnica FBG gêmea sendo aplicada para interrogar FBGs consegue a vantagem de melhorar a resolução e permite o uso de algoritmo de correlação cruzada, visando medir o atraso temporal. No mesmo sentido, Faisal [60] enfatiza que o método oferece uma resposta rápida e alta resolução.

Dentre o conjunto de vantagens, destacam-se:

- A alta precisão, haja vista que duas FBGs comumente reduzem erros, ao variar muito a separação do espectro da referência e do outro FBG sensor (exposto a uma variável externa), teoricamente, alinhados, o produtório entre eles começa a ter uma redução significativa, e consequentemente, corrobora para a precisão do parâmetro medido;

- Estabilidade ao longo prazo, pois a configuração do FBG gêmeos constitui duas grades sujeitas a mesma condição ambiental, considerando que entram em operação na mesma época, fazendo com que o sistema de medição seja mais robusto e estável;

- Mínima interferência cruzada, tendo em vista que dois FBGs estreitamente combinados minimizam a interferência entre diferentes elementos sensores do sistema;

- Monitoramento em tempo real, pois o uso da técnica permite o monitoramento simultâneo de vários parâmetros, com tempos de resposta rápidos e saída contínua de dados, é ideal para aplicações que exigem medições dinâmicas, como monitoramento da integridade estrutural ou controle de processos industriais.

Por outro lado, no conjunto de desvantagens, destacam-se:

- A complexidade do sistema de interrogação em comparação com a configuração de FBG individual, podendo exigir um processamento de sinal mais avançado para interpretar os sinais diferenciais com precisão;

- Faixa de medição limitada, essa limitação se dá pela correspondência das duas grades. Se a faixa de condições ambientais (por exemplo, temperatura ou deformação) exceder o que os FBGs podem compensar, o sistema poderá perder precisão, não obstante, a faixa de medição pode ser aumentada usando dois FBGs com características espectrais mais amplas;

- Dependência das condições ambientais, pois, embora os FBGs gêmeos possam compensar a temperatura, eles ainda dependem de fatores ambientais, como vibrações mecânicas, umidade e pressão. Às vezes, isso pode afetar o desempenho geral, especialmente se as condições ambientais variarem muito;

- Maior complexidade na fabricação das grades, isso ocorre, pois diferentemente da grade única, fabricar duas grades idênticas com características precisas é um processo mais complexo e alcançar características espectrais para as duas grades acaba sendo desafiador, aumentando o tempo e custo de fabricação. Aspectos esses dependentes da técnica utilizada para fazer as grades. Na técnica ponto a ponto, gera uma incerteza do instrumento de medida, não havendo uma garantia que a grade seja idêntica. É uma técnica mais cara e o processo de fabricação é mais longo quando comparado com outras técnicas. Outra modalidade de fabricação das grades consiste na chamada máscara de fase, onde aplica-se uma variação de laser muito forte na fibra, posteriormente a grade passa por uma máscara de fase, a qual modula o feixe UV e ao fazer isso a precisão aumenta em relação a técnica ponto a ponto. Essa por sua vez, é uma técnica cujo tempo de fabricação é menor [67];

- Demandas de processamento de sinais, pois ocorre da técnica requerer algoritmos

de processamento de sinais mais avançados, visando extrair dados significativos da configuração do FBG gêmeos e isso pode exigir mais poder de processamento e equipamentos especializados para análises precisas em tempo real.

### 2.2.5 TÉCNICA DE INTERROGAÇÃO BASEADA EM OSA

A técnica de interrogação baseada em OSA é um método comumente usado para sensores FBGs e LGPs, particularmente, em ambientes de pesquisa em laboratório.

A integração de um OSA, que opera com sinais transmitidos, e um sensor FBG, que utiliza sinais refletidos, requer a inclusão de componentes ópticos adequados. Um dos principais elementos para essa combinação é o acoplador óptico, que divide a potência óptica entre diferentes caminhos. Um acoplador 3 dB, por exemplo, pode direcionar parte do sinal para o OSA enquanto permite que a outra parte seja refletida e analisada pelo FBG. Dessa forma, é possível monitorar simultaneamente as características espectrais da luz transmitida e refletida.

Esta técnica envolve medir diretamente o comprimento de onda de Bragg do sensor FBG. Ele utiliza uma fonte de luz de banda larga, como uma fonte de emissão espontânea amplificada (ASE - *Amplified Spontaneous Emission*) ou um diodo superluminescente (SLED - *Superluminescent Light Emitting Diode*) e um OSA para análise [60], o qual tem a propriedade de medir diretamente o espectro de reflexão do sensor FBG caracterizando a metodologia mais simples, proporcionando ao profissional visualizar os espectros refletidos do FBG e possibilitando a detecção e análise das mudanças no comprimento de onda de Bragg.

Nesta técnica, a luz da fonte de banda larga é lançada em uma fibra de modo único (SMF - *Single Mode Fiber*) para um FBG de detecção (SFBG), o qual reflete um comprimento de onda específico conhecido como comprimento de onda de Bragg ou comprimento de onda central [12]. Conforme a temperatura do ambiente ao redor da SFBG muda, o comprimento de onda de Bragg também se altera de maneira correspondente [68].

O OSA pode ser integrado a um computador pessoal (PC - *Personal Computer*) para a programação da leitura do espectro e para a visualização da temperatura ao redor

do sensor FBG [12, 68].

O tempo de resposta de um OSA está diretamente relacionado à sua resolução espectral, sendo que resoluções mais finas, como 5 pm, exigem tempos de varredura (resposta) mais longos em comparação com resoluções mais amplas, como 20 pm. Em aplicações que demandam monitoramento em tempo real, a utilização do OSA portátil com resolução de 5 pm pode ser uma desvantagem em ambientes industriais que estimam a salinidade, onde variações rápidas na concentração salina exigem medições ágeis. A demora na aquisição de dados pode comprometer ajustes operacionais imediatos, afetando a eficiência do controle de salinidade e a resposta à possíveis variações indesejadas.

## 2.3- PROCEDIMENTOS PARA INTERROGAÇÃO DE SENSORES ALTERNATIVOS OU INOVADORES

Consoante [69], o sensoriamento óptico baseado em interrogação de fase com holograma de interferência de Young usando um dispositivo de microespelho digital constitui um método originalmente caracterizado pela interrogação de fase para aplicações em sensores ópticos. Trata-se de um método interferométrico com característica distintiva na geração de uma sequência de hologramas de entrada e a obtenção da sequência de hologramas perturbada na saída pelo transdutor óptico.

O método de medição de fase inclui o projeto de um sensor de holograma para transdutores ópticos, operando tanto em transmissão quanto em reflexão. Demonstrou-se excelente estabilidade na medição, com uma redução significativa na deriva e sem limitações pelo ruído do laser ou do estágio de detecção. Além disso, alcançou níveis de ruído e resolução de fase de ponta, posicionando-o no estado da arte.

No âmbito das FBGs, o estudo proposto por Li et al. 2024 emprega o AWG (*Arrayed Waveguided Grating* - Grade de guia de onda em matriz) como elemento central do sistema de interrogação, beneficiando-se da ausência de componentes mecânicos móveis para garantir uma taxa de interrogação mais rápida. Além disso, explora-se, proficientemente, o efeito termo-óptico dos materiais SOI *(silicon-on-insulator)* para ajustar o comprimento de onda do AWG, o que amplia a sobreposição espectral do dispositivo.

O sistema demonstra uma versatilidade marcante ao possibilitar a efetiva interrogação de sinais como pressão, tensão e vibração [70].

Em [71], propõe-se o desenvolvimento de um sistema de interrogação, que explora a resposta de temperatura de um diodo laser multimodo, a fim de interrogar LPGs. Sistema esse capaz de interrogar ao longo de uma janela de 10 nm, com erros de medição atingindo valores mínimos tão baixos quanto 0,4 nm, para o comprimento de onda central da grade, e 0,4 dB para sua atenuação. Dado seu baixo custo e dimensões reduzidas, o sistema desenvolvido apresenta potencial para aplicações de campo de mudanças lentas, tais como o monitoramento de reações químicas prolongadas, por exemplo, corrosão, monitoramento ambiental de longo prazo, como o monitoramento da intrusão salina em aquíferos costeiros, ou processos de resfriamento lento ou aquecimento lento, como a cura de materiais.

Ainda no âmbito dos sensores LPGs, os autores de [72] demonstraram uma técnica de interrogação óptica usando as medições das amplitudes do primeiro e segundo harmônicos do sinal no domínio elétrico. Tratou-se de uma metodologia auto-referenciada por ser independente da fonte de potência óptica; em outras palavras, apresentou insensibilidade a flutuações de potência óptica. Nesse contexto, o procedimento de interrogação apresentado nesse trabalho utilizou o modelo matemático desenvolvido e demonstrado pelo autor de [73]. O sistema proposto interrogou o sensor resolvendo uma equação analítica simples, que pôde ser realizada com um microprocessador de baixo a médio alcance. Assim, calculou o comprimento de onda de ressonância do LPG associado a um parâmetro externo e mostrou-se eficiente em medições em tempo real. Para evidenciar a abordagem de interrogação, foram realizadas as medições do índice de refração de várias soluções. Por conseguinte, o processo experimental satisfez o monitoramento em tempo real e representou um avanço tecnológico de baixo custo em relação às medições tradicionais.

## 2.4- ESCOPO DESTA DISSERTAÇÃO

Nesta dissertação, aplicaremos técnicas de regressão simbólica à estimativa de salinidade de soluções, na linha das referências que constam na Tabela 3. A estimativa da

salinidade de soluções diversas é de extrema importância para a indústria tendo em vista que a salinidade afeta diretamente a qualidade e a eficiência de processos industriais em setores como a produção de alimentos, a fabricação de produtos químicos e a dessalinização da água. Além disso, a precisão na determinação da salinidade é crucial para garantir a conformidade com padrões regulatórios, evitar danos a equipamentos e garantir a segurança e qualidade dos produtos finais. A capacidade de estimar a salinidade com precisão permite um controle mais eficaz dos processos industriais, reduzindo custos operacionais e maximizando a produtividade.

No que concerne a primeira referência da Tabela 3, a produção de bobinas de microfibra exige técnicas de fabricação avançadas, muitas vezes envolvendo controle rigoroso de parâmetros como tensão, alinhamento e consistência do diâmetro. Isso pode aumentar os custos de produção e os tempos de fabricação. A demora em fazer o sensor proposto decorre da inserção da bobina de microfibra, caracterizando pouca resistência, pois ficam suscetíveis a quebras ou deformações quando submetidas a forças externas. Além disso, o alto custo de manutenção e vulnerabilidade a pequenos defeitos. Experimentalmente, existe um elemento ressonador de microfibra, o qual está em volta de um cilindro PMMA e no entorno, coberta com um material epóxi de baixo índice de refração curável por UV (Luvantix, PC373) e totalmente curada usando uma lâmpada UV *(Lightning cure*, LC8) para reduzir qualquer perda desnecessária causada pela diferença no índice de refração entre o PMMA e a microfibra. A folga entre as bobinas adjacentes é de 7,36 µm e o diâmetro da haste de suporte após ser coberto com o epóxi de baixo índice de refração é de 1,3 mm [74].

A segunda referência Zhang et al. 2019, evidenciada na Tabela 3, explicita que o efeito do índice de refração com a salinidade da água vai ser semelhante. A técnica apresenta o uso clássico do interferômetro de Michelson, onde dois feixes laser percorrem diferentes trajetos ópticos e, ao se recombinarem, geram um padrão de interferência que depende da diferença de caminho entre os dois feixes.

Isso torna o sistema altamente eficiente para a demodulação de sinais, uma vez que a diferença de caminho óptico entre os dois braços do interferômetro permite uma maior sensibilidade e precisão na detecção de alterações externas. Por se tratar de um *setup* que utiliza um interferômetro de Michelson, em outras palavras, um jogo de espelhos e laser, que atravessa um reservatório que contém um líquido, trata-se inclusive de um sensor óptico e não tem fibra, sendo esse sensor feito diretamente no laser, possuindo

Tabela 3 – Comparação do Sensoriamento de Salinidade para os Trabalhos de Pesquisa Relacionados em Instrumentação Óptica.

| **Referência** | **Esquema/Técnica** |
|---|---|
| Yin et al., 2018 [74] | Requer um processo de fabricação longo e delicado. |
| Zhang et al., 2019 [75] | Baseado na diferença de caminho óptico de feixes laser usando uma configuração de interferômetro de Michelson. |
| Sun et al., 2019 [76] | Baseado em FBG revestida com poli-umida lamelar. |

um *setup* muito grande devido a montagem, também não é portátil e necessita de controle para o ar (dielétrico), visto que o laser está passando num ambiente interno e não permite medição remota.

Na terceira referência, conforme a Tabela 3, sensores ópticos baseados em FBGs revestidos com poli-imida lamelar são particularmente eficazes na medição de salinidade devido à sensibilidade aumentada proporcionada pela casca. A poli-imida lamelar protege o sensor FBG contra agentes corrosivos, como a água salgada (consolidando proteção contra danos físicos e químicos ao sensor), ao mesmo tempo em que amplifica a resposta do sensor FBG às mudanças no índice de refração causadas pela variação da salinidade; favorecendo, portanto, à aplicação em processos industriais.

A análise dos gráficos apresentados no estudo demonstra, primeiramente, que o sensor lamelar apresenta um tempo de resposta relativamente longo para estabilização completa em medições de salinidade. Esse comportamento é atribuído ao processo de formação da camada sensível em contato com a solução salina, o que pode levar aproximadamente 6 horas. Tal característica, embora possa ser uma limitação em aplicações que demandam respostas rápidas, é compensada pela alta precisão e estabilidade das medições em condições controladas. Em segundo lugar, o artigo evidencia que o comprimento de onda apresenta uma variação inicial acentuada nos primeiros instantes, seguida por uma estabilização gradativa à medida que o equilíbrio térmico e óptico é alcançado no novo meio (ou seja, quando o sensor muda de ar para água). Por fim, os dados apresentados graficamente indicam que a acurácia do sensor sofre uma deterioração significativa em condições de salinidade elevada. Ademais, enquanto o sensor mantém uma resposta precisa e consistente em níveis de salinidade baixos a moderados, observa-se uma crescente discrepância entre os valores medidos e os valores de referência, conforme a salinidade aumenta.

## 2.5- CONSIDERAÇÕES FINAIS AOS MÉTODOS DE INTERROGAÇÃO POPULARES E INÉDITOS

A principal vantagem de usar métodos alternativos de interrogação é a redução de custos, juntamente, com a portabilidade e a automação mais fácil. Os métodos de interrogação que utilizam filtragem óptica requerem algum processamento para extrair adequadamente as informações de alto nível de interesse [77]. Segundo [77], o procedimento de interrogação de sensor fazendo uso de filtros ópticos concomitante com métodos computacionais caracterizam-se pela simplicidade, consolidando eficácia metodológica. O sistema proposto em [77] foi capaz de identificar a temperatura a partir da saída do sensor, caracterizando um sistema de interrogação capaz de fornecer o valor preciso da temperatura sem medir diretamente o deslocamento do comprimento de onda de ressonância ou realizar qualquer cálculo de Fourier. Nesta dissertação, o objetivo é prever a salinidade de uma solução, por meio de medidas obtidas por interrogação óptica. Pretende-se obter a estimativa da variável de interesse por meio de uma fórmula analítica, a ser obtida via regressão simbólica. A regressão simbólica é o tema do próximo capítulo [78].

# 3- REGRESSÃO SIMBÓLICA - FUNDAMENTOS

## 3.1- A REGRESSÃO SIMBÓLICA

A regressão simbólica começou seus primeiros desenvolvimentos como uma ferramenta para a ciência nas décadas de 1970 e 1980 [79]. A ideia de uma máquina capaz de oferecer *insights* para os seres humanos foi discutida no trabalho pioneiro de Donald Michie em 1988 [80, 81]. Algoritmos de aprendizagem de máquina (ML, do inglês *machine learning*) foram classificados por Michie em três classes: ML fraco, ML forte e ML ultraforte. Algoritmos de ML classificados como “fracos” identificam o caso em que o aprendizado de máquina produz um desempenho preditivo aprimorado com quantidades crescentes de dados [2, 80]. A regressão simbólica está inserida na modalidade ML fraco já que é incapaz de extrapolar a predição para regiões de dados de entrada diferentes das do conjunto de treinamento. Em termos de popularidade, desde os anos 2010 [82], a regressão simbólica tem sido consideravelmente implementada como uma ferramenta de extração automática de fórmulas que regem determinados sistemas ou fenômenos.

Grosso modo, algoritmos de regressão simbólica almejam sintetizar fórmulas analíticas, automaticamente, podendo por conseguinte ser interpretados como técnicas específicas de mineração de dados. Tais fórmulas, que implementam um determinado conjunto discreto de primitivas (como funções exponenciais, cossenos ou tangentes hiperbólicas), devem emular a transformação entre os dados de entrada e de saída, de modo a aderir aos dados tanto quanto possível [83].

Mais especificamente, as técnicas de regressão simbólica adotadas ao longo desta dissertação representam expressões matemáticas por meio de estruturas de dados específicas, conhecidas como “árvores unárias-binárias”. Nessas árvores, os nós internos representam operadores (como a adição numérica), enquanto que os seus terminais são operandos (como variáveis ou constantes) [84, 5]. A partir de uma inicialização aleatória de possíveis fórmulas candidatas (representadas por árvores), técnicas de regressão simbólica refinam progressivamente essas fórmulas, aplicando às árvores (que as representam) algoritmos genéticos. Os algoritmos genéticos consistem num

exemplo de técnicas “meta-heurísticas bioinspiradas”, as quais inspiram-se em soluções já existentes no mundo dos seres vivos (como a evolução natural das espécies) como ferramentas de otimização matemática para problemas desafiadores encontrados na ciência e na engenharia.

Em 2021, o SRBench foi proposto como uma grande referência para a regressão simbólica. O projeto (ainda existente) consiste em promover incentivos a novas metodologias e fomentar o uso de novos conjuntos de dados tangíveis[1] e sintéticos[2]. Em sua criação, o SRBench apresentou 14 métodos de regressão simbólica, 252 conjuntos de dados do PMLB[3] (Penn Machine Learning Benchmarks) e 7 outros métodos de ML.

Em 2022, a SRBench anunciou a competição *Interpretable Symbolic Regression for Data Science*, que foi realizada na conferência GECCO em Boston (MA), durante a qual houve a comparação de diferentes métodos de acordo com propriedades, tais como: seleção de características; redescoberta de expressões exatas; insensibilidade a mínimos locais; extrapolação e sensibilidade ao ruído. A biblioteca de regressão simbólica em Python (PyRS) ficou em segundo lugar nessa competição. Devido à sua grande competitividade, facilidade para replicar resultados e potência, essa biblioteca será a utilizada nesta dissertação.

A utilização da regressão simbólica como ferramenta de modelagem baseada em dados abre caminho para o desenvolvimento de sensores fundamentais para a geração de modelos de processo em linha, os quais são inteligíveis, acurados e refletem fielmente as características do processo investigado [89]. Alguns trabalhos relevantes que exploram a regressão simbólica encontram-se descritos na Tabela 4.

A regressão simbólica funciona melhor em conjuntos de dados de baixa dimensão, mas, pode-se também estender essa abordagem para espaços de dimensões superiores utilizando a “Destilação Simbólica” de Redes Neurais, conforme explicado em [87], que enfoca problemas de $N$ corpos. É possível usar regressão simbólica, essencialmente, para converter uma rede neural a uma equação analítica. Assim, essas ferramentas apresentam simultaneamente uma maneira explícita e poderosa de interpretar redes neurais profundas [87, 90]. Essa interpretabilidade não diz respeito à precisão das

[1] Dados tangíveis também são chamados de “Dados do Mundo Real”, ou seja, dados observacionais coletados de processos físicos [85].

[2] Dados sintéticos são dados gerados computacionalmente a partir de funções estáticas ou simulações [85].

[3] É um repositório de *benchmarks* de aprendizado de máquina mantido pela Universidade da Pensilvânia. O PMLB fornece uma coleção diversificada de conjuntos de dados para a avaliação e comparação de algoritmos de aprendizado de máquina, incluindo regressão simbólica.

Tabela 4 – Comparação autoral e contributiva para os trabalhos de pesquisa relacionados à regressão simbólica. QSAR: *Quantitative Structure-Activity Relationships* - Relações de Atividades de Estruturas Quantitativas. QSPR: *Quantitative Structure-Property Relationships* - Relações de Propriedades de Estruturas Quantitativas. FIGP: *Filter-Introduced GP - Programação Genética Introduzida por Filtro. GP: Genetic Programing - Programação Genética. SR: Symbolic Regression - Regressão Simbólica. PySR: Symbolic Regression in Python and Julia. SciMED: Scientist-Machine Equation Detector - Detector de Equação Cientista-Máquina.*

| **Referência** | **Contribuição** |
|---|---|
| Kataki, K., Miyao. T., 2022 [86] | Este artigo contribuiu com a descrição do Método QSAR/QSPR interpretável baseado no uso de três filtros em GP para SR. O FIGP proposto gerou modelos QSPR mais simples do que dois métodos SR existentes (AI Feynman e GP). |
| Cornelio, C., *et al*., 2023 [8] | O referido artigo, através da equação experimental de Langmuir estabelece soluções de regressão simbólica para dois conjuntos de dados de adsorção. Refere-se à adsorção de metano em mica a uma temperatura de 90 K, enquanto a outra se refere a adsorção de isobutano em silicalita a uma temperatura de 277 K. |
| Cranmer, Miles, 2023 [79] | Destaca o algoritmo evolutivo multipopulacional contido na biblioteca de código aberto para a SR, chamada PySR. É uma biblioteca que acelera a descoberta de interpretações de modelos simbólicos. |
| Krenn, M., *et al*., 2022 [2] | O artigo contribui consolidando a compreensão científica como um dos principais objetivos da ciência e enfatiza a transformação da AI (do inglês, artificial intelligence) em agentes de compreensão caracterizando-os como esforços de investigação. A principal questão de busca do artigo é desvendar como os sistemas computacionais, e especificamente a inteligência artificial, contribuem para novas compreensões científicas, além de averiguar esta obtenção, autonomamente. |
| Keren, Simon L. *et al*., 2023 [87] | A contribuição deixada pelo artigo consiste na inovação de uma estrutura computacional de código aberto chamada Detector de Equação Científica-Máquina (SciMED), a qual integra métodos de regressão simbólica à disciplina científica; o sistema computacional SciMED apresenta a combinação de um método de seleção *"wrapper"* que é baseado em algoritmo genético. Na etapa experimental ficou evidente a robustez do sistema na descoberta de notações simbólicas, expressões dos dados e a demonstração de que a integração do conhecimento do domínio melhora o desempenho. Além disso, este novo sistema computacional agrega uma nova vertente ao alertar o usuário sobre possíveis recursos ausentes, contrariamente, ao que a maioria dos sistemas de regressão simbólica são capazes de fazer. |
| Murati, A., *et al*., 2020 [88] | A contribuição ressalta a aplicação da regressão simbólica e programação genética, concernente a obtenção de resultados na forma fisicamente significativa e interpretável provou-se inestimável. O artigo exibiu várias abordagens distintas de aplicabilidade em vários ramos da física, tais como: sensoriamento remoto, física atmosférica, não limitando-se apenas no campo dos plasmas termonucleares. Outro legado importante consiste na previsão de estudo futuro através da combinação de SR via GP também com redes neurais de topologia complexa, na tentativa de explorar a grande capacidade do aprendizado profundo. |
| Makke, N. e Chawla, S., 2024 [84] | A revisão ressalta a regressão simbólica como uma ferramenta com perspectivas promissoras, oferecendo um equilíbrio entre interpretabilidade e precisão. No entanto, ela aponta limitações na abordagem linear devido à estrutura do modelo já definida. Destaca-se que os métodos baseados em árvores, especialmente os transformadores, são os mais eficientes em dados sintéticos. Ademais, é enfatizada a influência das operações permitidas pela biblioteca na precisão das previsões do modelo, enquanto os métodos baseados em redes neurais enfrentam desafios numéricos. |

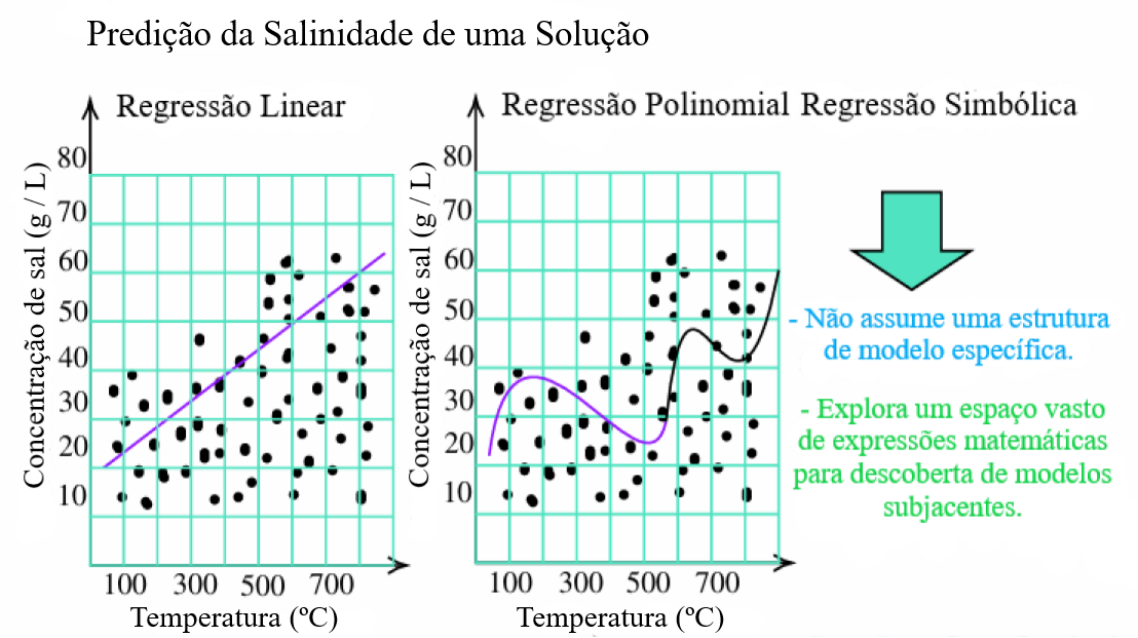



Figura 3 – Versatilidade da técnica de regressão simbólica.

estimativas nos dados de teste. Outro incentivo para o gerenciamento da complexidade são as restrições computacionais. Por exemplo, dispositivos de Internet das Coisas (IoT - *Internet of Things*) restringem o tempo de avaliação de um modelo aceitável, mesmo que isso comprometa a precisão do modelo. Em resumo, as razões para controlar a complexidade do modelo variam, assim como a noção de complexidade [91].

As abordagens de regressão simbólica baseadas em computação evolutiva buscam encontrar funções analiticamente tratáveis, diferenciando-se dos algoritmos tradicionais de regressão determinística, que apenas ajustam parâmetros [92]. Os algoritmos evolutivos de regressão simbólica procuram não apenas ajustar parâmetros, mas também determinar a forma funcional mais apropriada do modelo a partir de dados de entrada e resposta (saída), utilizando um conjunto de operadores matemáticos e operandos predefinidos, enquanto minimizam as métricas de erro.

A Figura 3 elucida que a regressão simbólica não assume uma estrutura de modelo específica. Ademais, visa à estrutura subjacente dos dados, explorando um espaço combinacional e vasto de expressões matemáticas, mas não é infinito matematicamente, pois ao se limitar a profundidade das árvores e operadores, existe uma quantidade finita, embora muito elevada, de combinações.

Dados ruidosos são aqueles que contêm informações pouco relevantes, erros ou variações aleatórias que não representam o padrão real do fenômeno estudado.

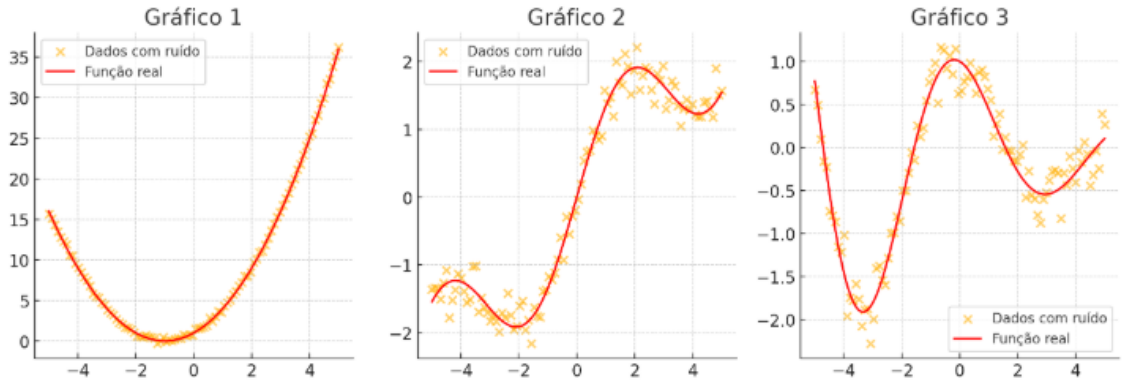



Figura 4 – Exemplificação de cenários gráficos para a regressão simbólica com a presença de dados ruidosos.

Em alguns casos, o ruído pode esconder padrões úteis, dificultando a identificação de relações reais nos dados. No entanto, em outros casos, métodos apropriados podem filtrar o ruído e extrair informações valiosas, confirmando a positiva ressalva de que em certos modelos, uma pequena quantidade de ruído pode até ajudar na generalização.

Na Figura 4 é possível observar representações gráficas para a regressão simbólica com a presença de dados ruidosos. A função real representa a relação verdadeira entre as variáveis de entrada e saída em um problema de aprendizado de máquina. No cenário em que o dado ruidoso apresenta proximidade em relação a função real (ou seja, maior aderência aos dados), caso o nível de ruído seja baixo ou se os dados forem pré-processados corretamente, o modelo pode identificar o padrão subjacente e fazer boas previsões. Contrariamente, caso haja distanciamento da função real, se houver um nível de ruído muito alto ou se o modelo não tiver técnicas adequadas [4] para lidar com o mesmo, os padrões verdadeiros podem ser ofuscados, resultando em um modelo impreciso e menos eficiente.

### 3.1.1 REGRESSÃO SIMBÓLICA E A BIBLIOTECA PySR

A biblioteca PySR (do inglês, *PySymbolicRegression*) foi desenvolvida com o objetivo de democratizar e popularizar a regressão simbólica nas ciências. Trata-se de uma biblioteca de código aberto para regressão simbólica prática. Ao aproveitar um

[4] Normalização e padronização são técnicas que podem ser utilizadas no pré-processamento dos dados. Geralmente a escolha das diversas técnicas existentes dependem do tipo de ruído específico.

poderoso algoritmo evolutivo multipopulacional, uma solução única de *loop* evolução-simplificação-otimização, um *back-end* distribuído de alto desempenho e integração com pacotes de aprendizagem profunda, o PySR é capaz de acelerar a descoberta de modelos simbólicos interpretáveis a partir de dados [79].

O modelo multipopulacional, também chamado de “modelo de ilhas” pode ser adotado para gerenciar uma solução candidata, dentro do espaço de busca da otimização simbólica, para refinamento adicional em iterações subsequentes do processo de pesquisa; com isso, evitam-se ótimos locais [93].

Até que o armazenamento final em *buffer* seja concretizado, ocorre uma avaliação baseada em dados, com a equação $s$ e sua aptidão correspondente $f$ sendo denotadas por $(f, s)$. O par ordenado $(f, s)$ refere-se a um par de informações concernente às soluções exploradas. Essas ilhas evoluem de forma independente, permitindo a exploração paralela de diferentes regiões no espaço de equações [93]. Por conseguinte, no final de um número especificado de rodadas de evolução, a migração entre as ilhas é realizada [79].

O objetivo é permitir a especialização através da evolução separada de populações únicas, permitindo ocasionalmente que alguns membros migrem entre eles, para compartilhar essa especialização. Na Figura 5, é possível visualizar a representação de “ilhas” independentes de expressões, cada uma em evolução, bem como a migração *a posteriori* entre ilhas. Esta figura corresponde ao laço externo na lógica de programação do pacote PySR.

A simplificação vincula-se a complexidade, a qual desempenha um papel significativo na regressão simbólica. Embora o ideal seja sempre apresentar os modelos na forma mais simples, isso pode ser computacionalmente inviável. Por isso, uma comparação baseada no comportamento das funções, considerando restrições simbólicas, pode ser mais eficaz, pois essas restrições permanecem consistentes independentemente das diferentes formas de representação das expressões [94].

Quanto a aplicabilidade da referida biblioteca, destaca-se, de acordo com a literatura atualizada [79]:

1. Usa-se PySR para descobrir uma nova parametrização simbólica para formação de cobertura de nuvens [95];

2. Usa-se PySR para descobrir modelos populacionais interpretáveis de fontes de

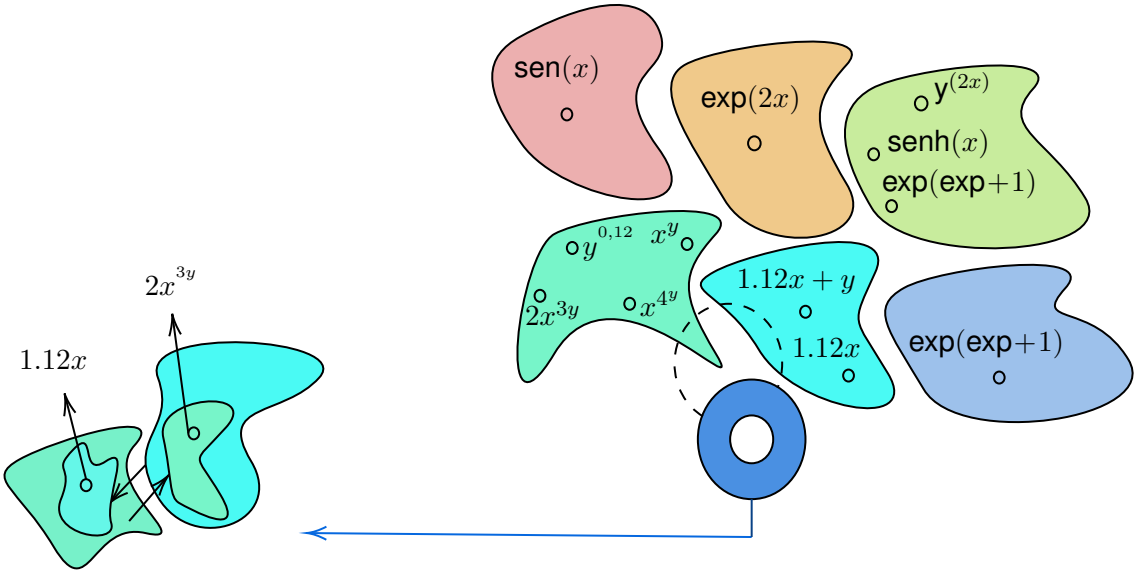



Figura 5 – Exemplificação da migração entre “ilhas” correspondente ao laço externo da biblioteca PySR.

ondas gravitacionais [96].

3. Usa-se PySR junto com outras técnicas de ML para descobrir especificações de transferência de elétrons em vários materiais [97];

4. Combina-se PySR com análise dimensional para descobrir novas relações astrofísicas [98];

5. Demonstra-se o uso da PySR para extrair equações dinâmicas aprendidas por uma equação diferencial neural [99];

6. Utiliza-se a PySR na economia, para encontrar regras eficazes que governem o comércio internacional [100].

### 3.1.2 REDES NEURAIS PARA REGRESSÃO SIMBÓLICA

A referência [101] propôs uma linha de pesquisa que investigou o potencial de redes neurais profundas em resolver o problema combinatório da regressão simbólica. A sugestão foi de uma rede neural totalmente conectada e simples, substituindo as funções

de ativação padrão por blocos de construção simbólicos (por exemplo: sen$(x)$, cos$(x)$, $+$, Identidade$(x)$).

Após o treinamento do modelo, foi possível extrair, automaticamente, uma fórmula simbólica a partir da arquitetura e dos pesos da rede. Esse método herdou a capacidade das redes neurais de lidar com dados de alta dimensão e se ajustou bem ao número de pares de entrada e saída. No entanto, demandou extensões específicas consoante [102] para lidar com funções que envolveram operações de divisão entre blocos de construção elementares e a inclusão do resultado de ativações exponenciais e logarítmicas em gradientes “explosivos” e questões numéricas.

Conforme [103], uma vez que uma expressão é amostrada, ela é simplificada usando as especificações construídas na biblioteca de manipulação simbólica SymPy. Esse método de amostragem nos permite determinar com precisão ao restringir o espaço de busca controlando a profundidade das árvores e o conjunto de operadores admissíveis, juntamente com suas probabilidades anteriores de ocorrência na expressão gerada [104].

Um exemplo recente e concreto em astronomia é a redescoberta da lei de Newton da gravitação a partir de observações de dados do mundo real, dados de planetas e luas em nosso Sistema Solar dos últimos 30 anos. Vários estudos demonstraram a extração de modelos simbólicos de dados experimentais de sistemas mecânicos, em astronomia e em sistemas quânticos. Os modelos simbólicos podem ser melhorados para propor medidas físicas razoáveis em modelos de sistemas desconhecidos que avançam compreensão científica [2].

### 3.1.3 ALGORITMOS GENÉTICOS E A INTELIGÊNCIA ARTIFICIAL

Algoritmos genéticos consistem numa técnica de busca muito utilizada para encontrar soluções aproximadas em problemas que envolvem otimização e busca. Eles consistem numa classe particular de algoritmos evolutivos que utilizam técnicas provenientes da biologia evolutiva, tais como: hereditariedade, mutação, seleção natural e cruzamento.

Na virada do século XXI, o poder computacional se tornou mais acessível, viabilizando o treinamento de algoritmos de inteligência artificial com maior quantidade de

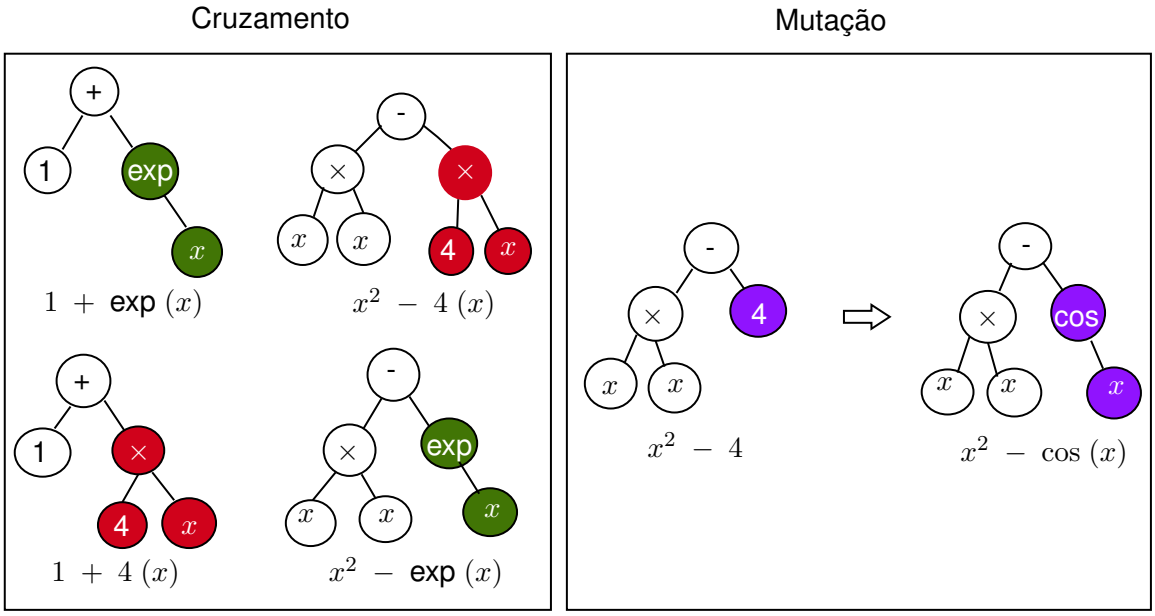



Figura 6 – Estrutura de dados em árvore para representação de fórmulas-exemplo; cruzamento (à esquerda) e mutação (à direita).

dados em uma velocidade de processamento maior. Novas pesquisas no campo recuperaram força e financiamento, e o aprendizado de máquina começou a ser utilizado em aplicações práticas por empresas e pesquisadores [105]. Com o progresso dos métodos e algoritmos de AI, a simulação e os dados experimentais assumirão um papel ainda mais significativo em projetos práticos consonante ao presságio relatado em Drikakis (2023), abarcando aplicações exigentes na ciência e tecnologia [106].

Dentre as ferramentas de aprendizado de máquina, a escolha do tipo do algoritmo que melhor se adeque aos objetivos é importante, visto que o algoritmo dita como o modelo se comporta, aprende, e como é estruturado após o treinamento [105]. No conjunto dos algoritmos mais populares estão o de árvores de decisão, florestas aleatórias, XGBoost, redes neurais, além de outros.

O conjunto de indivíduos denominado “população”, com representatividade em árvores são gerados, aleatoriamente, conforme a Figura 6. A evolução populacional tem início em T(0) ou (tempo = 0), em virtude da Programação Genética (PG - *Genetic Programming*), a qual baseia-se nos princípios da evolução biológica para resolver problemas de otimização e busca. Ou seja, através de operações evolucionárias, utilizando um conjunto de “regras de transição”. A representação matemática abaixo:

$$\{r_i : f \to f | i \in \mathbb{N}\} \tag{5}$$

é definida sobre o espaço da árvore [84]. A equação indica um conjunto de operadores $r_i$ que atuam sobre uma função $f$, onde $i$ é índice que pertence a um subconjunto dos números naturais $N$.

Conforme Calabrese *et al*. 2022 [107], os operadores genéticos para a evolução dos indivíduos são reprodução, cruzamento e mutação. O operador de reprodução consiste em copiar um indivíduo escolhido por sua pontuação de aptidão na nova população sem qualquer transformação.

Na operação de mutação, uma subárvore selecionada é substituída por outra subárvore gerada, aleatoriamente, com uma certa probabilidade [86], conforme a Figura 6, à direita. Outrossim, Heider *et al*. 2023 afirmam que operações de mutação podem modificar tanto um único nó quanto uma subárvore completa [108]. Usando-se, especificamente, o método de mutação pontual (ou seja, um nó na árvore de expressão S é escolhido, aleatoriamente, e substituído por outro valor viável [87]), no contexto da regressão simbólica, os parâmetros são substituídos por parâmetros, e funções são substituídas por funções com o mesmo número de argumentos.

Um operador de cruzamento envolve troca de conteúdo entre dois indivíduos, por exemplo, trocando uma subárvore aleatória de um indivíduo com outra subárvore aleatória de outro indivíduo [109], conforme ilustrado na Figura 6, à esquerda.

O operador de cruzamento introduz variação na população criando descendentes que incluem algumas partes de seus pais, escolhidos por método de seleção.

Em suma, no contexto da etapa de seleção, os indivíduos são rastreados de acordo com sua aptidão, haja vista que é mais fácil para um indivíduo que possui maior aptidão aparecer na próxima geração. Indivíduos com maior aptidão evoluem por meio de cruzamento e mutação [110].

O operador de mutação é uma alteração estocástica de um ou mais genes. Introduz a exploração de novos espaços na superfície de *fitness* para evitar a convergência prematura do programa. Cada operador é realizado com uma certa probabilidade [111]. Em particular, a probabilidade do operador de mutação $P_m$ não pode ser maior do que 0,1 [112], enquanto a do operador de cruzamento $P_c$ normalmente varia entre 0,6 e 0,9 [113]. A estimativa do operador de reprodução $P_r$ geralmente é complementar à soma dos percentuais probabilísticos de cruzamento e mutação. Entre as probabilidades dos

operadores mantém-se a seguinte relação [107]:

$$P_r + P_c + P_m = 1 \tag{6}$$

Em sua maioria, os desafios enfrentados no campo do aprendizado de máquina podem ser simplificados como problemas de otimização. Grande parte dos métodos populares para essa tarefa pode ser interpretada como uma busca pela minimização, resolvida por meio da modelagem de uma estrutura capaz de mapear os dados de entrada com a saída, conforme [114]. Por exemplo, árvores de decisão de regressão usam essa configuração e procuram aproximar relações matemáticas complexas, as quais podem ser representadas por estruturas de dados que minimizem o erro entre suas decisões e as respostas esperadas.

Em síntese, o algoritmo de regressão simbólica retorna um modelo “padrão” após um treinamento de aprendizado de máquina. Cabe ressaltar que cada método de aprendizado de máquina busca realizar a tarefa por meios diferentes, partindo de um espaço de busca contendo um conjunto de modelos possíveis, e o modelo final é obtido minimizando uma função de custo sobre os dados disponíveis.

## 3.2- INTERPRETABILIDADE

A interpretabilidade é um atributo de determinadas soluções de aprendizagem de máquina altamente valorizada pela comunidade científica. Ela permite, por exemplo, que a tomada de decisões em determinados sistemas de suporte sejam baseadas em regras explícitas e suficientemente simples de entender e interpretar.

### 3.2.1 MÉTODOS CAIXA PRETA, CAIXA CINZA e CAIXA BRANCA

Técnicas de aprendizado de máquina são métodos que produzem modelos que são funções das entradas e produzem saídas, onde costumeiramente o funcionamento

interno do modelo está oculto ou ininteligível para ser analisado [115]. Enquanto os modelos *Gray-Box* incorporam algum conhecimento da física dos fenômenos em estudo, os modelos *Black-Box* dependem apenas de abordagens baseadas em dados [116].

Nos modelos do tipo *Black-Box*, a lógica interna é representada em parâmetros complexos, como pesos em uma rede neural. Além disso, há dificuldade de compreender a tomada de decisão. Por outro lado, estes modelos possuem elevado desempenho preditivo, devido à capacidade de capturar relações complexas nos dados. Inclusive possuem capacidade de lidar com sistemas não lineares e reduzem problema de *overfitting*, quando o modelo é bem treinado. No panorama de aplicações adequadas para esse sistema, cuja arquitetura é oculta, podemos destacar a predição de padrões complexos em termos de uma variável de interesse (temperatura, salinidade, densidade etc.).

Os modelos de *White-Box*, por sua vez, apresentam alta interpretabilidade, pois as decisões podem ser diretamente compreendidas através de fórmulas matemáticas ou representações gráficas. Frequentemente, tais modelos são obtidos a partir de princípios fundamentais. Ademais, possuem processos internos que oferecem transparência e compreensibilidade. No âmbito de aplicações específicas (como em medicina), regulamentações que exijam explicabilidade tornam esses modelos mais viáveis. Porém, eles não funcionam bem com dados de alta dimensionalidade, caracterizando baixa escalabilidade; além disso, podem tomar decisões simplistas ou imprecisas em cenários complexos [117].

Os modelos de *Gray-Box* permitem algum grau de visualização ou explicação sobre como as decisões são tomadas. Tratam-se de uma abordagem intermediária entre os modelos de *Black-Box* (opacos e altamente complexos) e os de *White-Box* (transparentes e facilmente interpretáveis). Eles combinam desempenho preditivo robusto com interpretabilidade parcial, o que os torna especialmente úteis em cenários onde é necessário um equilíbrio entre precisão e explicação. A presença de explicabilidade efetivamente gera confiança para aplicações em setores críticos, tal como: medicina [118] e engenharia. Ou seja, aplica-se em cenários onde a precisão é importante, mas a explicação também é essencial.

## 3.3- COMENTÁRIOS FINAIS

Este capítulo descreve concisamente os conceitos fundamentais das técnicas de regressão simbólica. Parte significativa destas técnicas geram equações compreensíveis através de programação genética, capturando relações complexas nos dados. Podemos interpretá-las como ferramentas de aprendizado de máquina, o qual aprimora as equações simbólicas para obtenção de interpretabilidade e acurácia. Sendo a programação genética um caso particular de algoritmo genético, a mesma tem foco em evoluir uma estrutura de dados que representa as equações candidatas. Na abordagem a ser utilizada nesta dissertação, tal estrutura de dados é representada em uma árvore unária-binária. Por conseguinte, ao final do processo de aprendizado, são obtidas equações candidatas capazes de emular a relação do mapeamento entre entrada e saída.

Outra vertente do capítulo atual consiste no embasamento teórico proveniente da biblioteca PySR, a qual realiza regressão simbólica para encontrar expressões matemáticas que descrevem dados, visando a simplificação. Ao mesmo tempo, a biblioteca SymPy permite adicionalmente simplificar e manipular simbolicamente essas equações. Combinadas, essas ferramentas auxiliam na descoberta e análise de modelos matemáticos com maior explicabilidade.

Os diferentes métodos de aprendizagem de máquina são categorizados como caixas brancas, caixas cinzas e caixas pretas, de acordo com o grau de interpretabilidade que eles proporcionam. Métodos de aprendizado profundo e redes neurais são métodos populares de aprendizagem de máquina, que são costumeiramente classificados como métodos do tipo “caixa-preta”, dada a baixa interpretabilidade do processo de formação de suas saídas. No próximo capítulo, iremos descrever a proposta de um sistema completo de interrogação óptica baseado em regressão simbólica.

# 4- METODOLOGIA E RESULTADOS

## 4.1- MÉTODO DE INTERROGAÇÃO ÓPTICA ORIGINAL

A referência [10] propõe um sistema compacto de sensoreamento de salinidade capaz de obter boa acurácia, mesmo quando o sensoriamento é remoto ou quando há baixa concentração de sais. O sensor utilizado é um LPG fabricado com o método do arco elétrico ponto a ponto. O interrogador é baseado em reflectometria óptica no domínio do tempo, com extração de características motivada por extratores tipicamente utilizados em processamento de áudio. O sensor se encontra a 4 km da fonte de luz, ilustrando a capacidade intrínseca de sensoreamento remoto.

O procedimento experimental utiliza um OTDR Anritsu MT9083A2-073. Uma pequena seção de fibra óptica, integrada ao sensor LPG, foi inserida em um invólucro retangular fabricado com impressora 3D. O LPG foi escrito em fibra monomodo padrão. O período de *grating* é de 530 $\mu m$, de sorte a produzir uma banda de atenuação na mesma banda espectral que a fonte de luz do OTDR, cujo comprimento de onda se encontra próximo aos 1550 nm[1]. O OTDR fornece a leitura de um espectro de transmissão, o qual espera-se conter a informação relacionada à variável física de interesse (concentração salina). A Figura 7 mostra o esquema da configuração usada no estágio experimental de [10].

[1]Como a estrutura física de interrogação não é uma contribuição deste trabalho, sugerimos fortemente a leitura de [10], para maiores detalhes acerca do procedimento experimental.

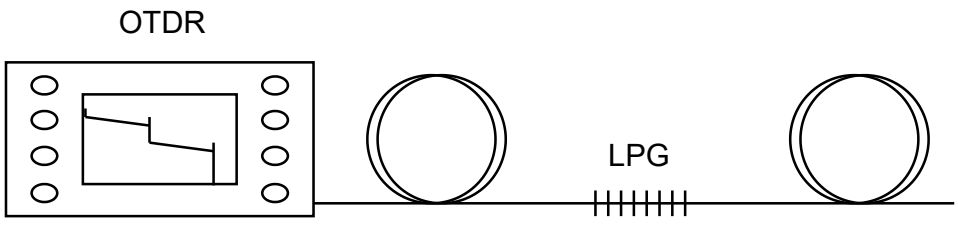



Figura 7 – Diagrama da configuração experimental.

Para efeitos de extração de características, a referência [10], motivada por extratores tipicamente utilizados em processamento de sinais de áudio, propõe a adoção de quatro extratores $\{\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4\}$, obtidos a partir do espectro de transmissão:

$$\hat{x}_1 = \frac{\sum_{n=1}^{N} P(\ell_n)\,\ell_n}{\sum_{n=1}^{N} P(\ell_n)}, \tag{7}$$

$$\hat{x}_2 = \frac{1}{N}\sum_{n=1}^{N} P^2(\ell_n), \tag{8}$$

$$\hat{x}_3 = \frac{\sum_{n=1}^{N} |x_1 - \ell_n| P(\ell_n)}{\sum_{n=1}^{N} P(\ell_n)}, \tag{9}$$

$$\hat{x}_4 = \sum_{n-2}^{N} |P(\ell_n) - P(\ell_{n-1})|^2, \tag{10}$$

onde $\ell_n$ é a abscissa do $n$-ésimo ponto selecionado do traço, isto é, a distância do OTDR para o $n$-ésimo ponto na região relevante do traço, $P(\ell_n)$ é a potência refletida originando no ponto $\ell_n$, e $N$ é o número de pontos na região selecionada. Os valores $\{\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4\}$ são normalizados para o intervalo $[0, 1]$ dando origem às variáveis normalizadas respectivas $\{\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4\}$.

A fórmula final de regressão proposta por [10] pode ser descrita como:

$$\begin{aligned} y = {} & 576.000 + 279.017 \cdot \tilde{x}_0 - 2073.524 \cdot \tilde{x}_1 - 1265.795 \cdot \tilde{x}_2 \\ & + 298.787 \cdot \tilde{x}_3 - 30.722 \cdot \tilde{x}_0^2 + 1591.529 \cdot \tilde{x}_1^2 - 291.119 \cdot \tilde{x}_2^2 \\ & - 99.631 \cdot \tilde{x}_1^3 + 2029.320 \cdot \tilde{x}_2^3 - 67.432 \cdot \tilde{x}_0^4 + 11.132 \cdot \tilde{x}_0^6 \\ & - 191.084 \cdot \tilde{x}_1^6 - 2631.655 \cdot \tilde{x}_2^6 + 48.744 \cdot \tilde{x}_0^7 + 1665.088 \cdot \tilde{x}_2^7 \\ & - 289.309 \cdot \tilde{x}_0\tilde{x}_2 - 17.764 \cdot \tilde{x}_0\tilde{x}_2^7 + 135.788 \cdot \tilde{x}_0^3\tilde{x}_2^7 \\ & - 39.848 \cdot \tilde{x}_0^7\tilde{x}_2^2 - 547.288 \cdot \tilde{x}_0\tilde{x}_3^2 + 2102.660 \cdot \tilde{x}_1\tilde{x}_2 \\ & + 55.362 \cdot \tilde{x}_1^7\tilde{x}_3^7. \end{aligned} \tag{11}$$

Onde $y$ é a estimativa da salinidade. O erro padrão do modelo é de 0,60 g/L, nos dados de teste.

A conjunção da metodologia experimental proposta em [10], conjugada com a Equação (11), dá origem a um método de sensoriamento baseado em fibra óptica que é compacto, portável, barato, seguro, resistente à corrosão e à oxidação, imune à interferência eletromagnética, tem um baixo tempo de resposta, permite o sensoriamento remoto de maneira direta, requer baixa manutenção e usa poucos componentes.

A Equação (11) é uma equação complexa. Seu formato funcional (soma ponderada de produtos das variáveis de entrada elevadas a expoentes inteiros) é proposto *a priori*, sendo um truncamento de uma série de Volterra. O processo de otimização almeja estimar os coeficientes de uma função cujo formato é preestabelecido. A hipótese fundamental desta dissertação poderia ser assim resumida: assume-se que a regressão simbólica seria capaz de apresentar uma equação mais concisa (e, possivelmente, mais acurada) do que a Equação (11).

## 4.2- APLICAÇÃO DA REGRESSÃO SIMBÓLICA

Para aplicação da regressão simbólica, será usada a mesma base de dados que a utilizada em [10]. Ao contrário de [10], não efetuaremos normalização dos descritores, porque a regressão simbólica é capaz de escalá-los, automaticamente, caso julgue necessário. Destinamos 80% do conjunto de dados para treinamento, reservando 20% deles para teste.

Dada a dificuldade do problema, definimos o número de iterações para 15.000, o que exigiu 353 horas para executar o experimento no seguinte sistema: uma máquina virtual com 16 CPUs e 64 GB de RAM rodando o sistema operacional Ubuntu 22.04.4 LTS com kernel 5.15.0-107-genérico x86 64. Tanto o número de populações quanto o tamanho das populações foram fixados em 100, a fim de promover uma maior exploração do espaço de busca. Vale ressaltar que os valores de referência para estes parâmetros são 15 (para o número de populações) e 33 (para o tamanho de cada população), que são insuficientes para obter soluções precisas. Para distorcer a busca por termos de baixa complexidade nas soluções, definimos a profundidade máxima para árvores candidatas para 8. O parâmetro `maxsize` limita a complexidade máxima de equações residentes no espaço de busca. Seu valor padrão é 20. Como o problema da interrogação é bastante desafiador para o estimador resultante, as equações de complexidade 20 são insuficientes para alcançar um bom desempenho. Por essa razão, aumentamos o valor desse parâmetro para 40. Permitimos os operadores binários de adição, multiplicação, subtração, divisão e exponenciação. Em relação aos operadores unários, permitiu-se que a pesquisa usasse as funções $\cos(x)$, $\exp(x)$, $\mathrm{sen}(x)$, $\log(x)$, $\mathrm{sqrt}(x)$, $\sinh(x)$, $\cosh(x)$,

$\tanh(x)$, $\operatorname{atan}(x)$, $\operatorname{asinh}(x)$, e $\operatorname{erf}(x)$, onde:

$$\operatorname{erf}(x) \triangleq \frac{2}{\sqrt{\pi}} \int_0^x \exp\left(-t^2\right) dt. \tag{12}$$

Como ferramenta para a regressão simbólica, foi utilizado o pacote PySR, um dos mais populares na comunidade acadêmica. O código Python encontra-se no Apêndice B.

Interessantemente, algoritmos de regressão simbólica não oferecem uma solução única. Eles oferecem diversas fórmulas, ordenadas por sua complexidade, de modo que o especialista (ou seja, o profissional que vai implementar o processo de inferência via regressão simbólica) possa escolher o melhor "ponto de operação" em duas dimensões: *i*) a complexidade da equação (cujo aumento reduz a interpretabilidade); *ii*) sua aderência aos dados (que é incrementada quando a complexidade das fórmulas aumenta). Adicionalmente, a complexidade é definida como o número de nós na árvore.

Fazendo a mudança de variáveis: $\hat{x}_1 \to x_0$, $\hat{x}_2 \to x_1$, $\hat{x}_3 \to x_2$, $\hat{x}_4 \to x_3$, após mais de 2 (duas) semanas, ou mais precisamente 353 horas, de execução do sistema de regressão simbólica, obtivemos algumas fórmulas, com as mais simples sendo descritas na Tabela 5.[2]

Dentre as fórmulas obtidas, a fórmula de menor complexidade que obtém um erro padrão similar à fórmula (11) é:

$$y = \theta_1(x_0, x_2) + \theta_2(x_2) + \theta_3(x_2, x_3), \tag{13}$$

onde

$$\begin{aligned}
\theta_1(x_0, x_2) &\triangleq -2.43462881636892 \\
&\quad \times[\cos(14.843932 \cdot x_2) + 1.6268982] \\
&\quad \times\cos(3.0025335077485 \cdot x_0), && (14)\\
\theta_2(x_2) &\triangleq \sinh(\exp(\cosh(\cos(0.4025822 \cdot x_2)))), && (15)\\
\theta_3(x_2, x_3) &\triangleq \operatorname{asinh}\left[\frac{1176037.4 \cdot \cos(1.059711 \cdot x_2)}{x_3}\right]. && (16)
\end{aligned}$$

Assim, temos que a equação (11) (proposta em [10]) e a equação (13) (obtida, *de maneira inédita*, por regressão simbólica, nesta dissertação), apresentam um erro padrão similar.

Podemos efetuar uma breve discussão acerca da equação proposta nesta dissertação

[2]Todas as fórmulas obtidas, mesmo as mais complexas, encontram-se no Apêndice A.

Tabela 5 – Fórmulas de complexidade menor ou igual a 13 geradas pelo método de regressão simbólica. Fórmulas mais complexas (que não cabem nesta tabela) são apresentadas no Apêndice, no formato SymPy.

| Complexidade | Fórmula |
|---|---|
| 1 | $49.9998$ |
| 3 | $x_1^{(-0,4818793)}$ |
| 5 | $80,722305 - 89312.336 \cdot x_3$ |
| 6 | $70,81623 - (e^{x_1})^{5528,247}$ |
| 7 | $74,79686 - e^{153,17398 \cdot \sqrt{x_3}}$ |
| 8 | $74,82488 - 15,607775 \cdot \mathsf{senh}(2872.4497 \cdot x_3)$ |
| 9 | $75,298935 - x_2 \cdot \mathsf{senh}(\mathsf{senh}(1636.5731 \cdot x_3))$ |
| 10 | $50892,58 - 11,864567 \cdot (3270,7036 \cdot \mathsf{senh}(x_3) + x_0)$ |
| 11 | $(\mathsf{asenh}(8,6985576e7 \cdot \mathsf{senh}(55,199825 \cdot \mathsf{sen}(\mathsf{log}_{10}(x_1)))) + 19,314484$ |
| 13 | $\mathsf{asenh}(9,967798e7 \cdot \mathsf{senh}(55,07948 \cdot \mathsf{sin}(\mathsf{log}_{10}(x_1)))) + 4302,44 - x_0$ |
| 14 | $1,33 \times 10^8 \cdot \mathsf{asenh}(52.427 \cdot \mathsf{senh}(\mathsf{sen}(\mathsf{log}_{10}(x_1))))$ $+ \left(20,67919 - \frac{1}{\mathsf{cos}((0,7153872 \cdot x_0 - x_2)}\right)$ |
| 30 | $y = \theta_1(x_0, x_2) + \theta_2(x_2) + \theta_3(x_2, x_3)$, onde: $\theta_1(x_0, x_2) \triangleq -2.43462881636892 \times [\mathsf{cos}(14.843932 \cdot x_2) + 1.6268982] \times \mathsf{cos}(3.0025335077485 \cdot x_0)$, $\theta_2(x_2) \triangleq \mathsf{sinh}(\mathsf{exp}(\mathsf{cosh}(\mathsf{cos}(0.4025822 \cdot x_2))))$, $\theta_3(x_2, x_3) \triangleq \mathsf{asinh}\left[\frac{1176037.4 \cdot \mathsf{cos}(1.059711 \cdot x_2)}{x_3}\right]$ |

(isto é, a Equação (13)), ressaltando os seguintes pontos:

- ⋆ Embora ambas as equações apresentem desempenho equivalente, haja vista que o erro quadrático médio de ambas equações não destoam entre si, a equação proposta (13) é significativamente mais simples, facilitando a compreensão do fenômeno de interrogação e enfatizando o papel de cada descritor na estimativa. Observe que a Equação (13) depende tão somente da soma de três termos, com o primeiro deles sendo dependente de $x_2$ e $x_0$, o segundo sendo dependente apenas de $x_2$, o terceiro sendo dependentes dos atributos $x_2$ e $x_3$;

- ⋆ A equação proposta (13) requer menos descritores do que a equação defendida por [10] (já que $x_1$ não é necessário na equação proposta). Assim, poder-se-ia *conjecturar* que a fórmula proposta é menos propensa a *overfitting*, já que é mais simples e utiliza menos descritores, com um desempenho similar.

- ⋆ Cumpre ressaltar a habilidade da regressão simbólica em efetuar automaticamente um procedimento de *seleção de características*, já que ela, automaticamente, descartou o descritor $x_1$, enquanto que este descritor foi utilizado (como $\tilde{x}_2$) na formulação de [10]. Na prática, sabemos que descartar determinados atributos é uma tarefa bastante desafiadora, que a regressão simbólica, ao penalizar soluções demasiado complexas, implicitamente, executa [119].

- ⋆ Cabe enfatizar a capacidade de a regressão simbólica encontrar uma expressão analítica sucinta que adere aos dados que é inviável para o especialista humano, que dificilmente imaginaria um funcional similar ao da equação proposta (13).

Importa notar adicionalmente que em nossos experimentos não observamos grandes discrepâncias entre o erro quadrático médio nos conjuntos de treinamento e de teste. Essa similaridade é um forte indício de que a técnica de regressão simbólica é razoavelmente robusta ao problema de *overfitting* (isto é, um excesso de aderência aos dados de treinamento, devido ao ajuste de seus parâmetros, que prejudica a capacidade de generalização). Entendemos que tal robustez possa ser atribuída à otimização, intrínseca às técnicas de regressão simbólica, que almeja encontrar fórmulas concisas, as quais não conteriam um número suficiente de parâmetros capazes de gerar *overfitting*.

## 4.3- RESULTADOS

A regressão simbólica leva em consideração a Frente de Pareto Ótima de soluções eficientes considerando a complexidade do modelo (ou seja, o número de nós na árvore unária-binária) versus ajuste aos dados [4]. Assim, a técnica retorna qualquer expressão muito simples (com baixa aderência aos dados) ou expressões mais complexas (que, embora tenham maior precisão, são extremamente complexas para permitir a extração de *insights* sobre o fenômeno de interesse). Uma avaliação numérica de resultados da regressão é mostrada na Tabela 6. Esta, por sua vez, revela que uma equação cuja complexidade é 30 está bem posicionada para tal troca. Observe que embora a equação de complexidade 35 apresente um erro absoluto médio menor, obtém um ligeiro aumento no erro padrão em comparação com a equação cuja complexidade é 30. Isso se deve ao fato de a Tabela 6 relatar resultados no conjunto de teste (ou seja, em dados que não foram usados em treinamento), onde uma variação estocástica das métricas de avaliação é esperada.

Baseando-se nessas considerações, a Equação (13) de complexidade 30 se destaca como candidata que combina aderência aos dados com menor complexidade.

A Equação (13) proposta apresenta diversas vantagens em comparação com a Equação original (11). Em primeiro lugar, apresenta um erro absoluto médio mais baixo e um erro padrão mais baixo (ver Tabela 6) no conjunto de dados de teste. A Figura 8 mostra a situação real e valores estimados de salinidade no conjunto de dados de teste, revelando uma boa adesão ao *ground truth*.

Também é mais concisa, exigindo 7 constantes, enquanto a Equação (11) exige 23 constantes (ver Tabela 7 para mais detalhes). Tal consideração é importante, visto que, na prática, este estimador seria implementado em um sistema embarcado com capacidades de processamento limitadas. Embora a função $\theta_2(x_2)$ possa parecer exigir mais recursos computacionais, devido ao cálculo recursivo de várias funções incomuns, é uma função par de uma única variável, o que permite uma descrição sucinta em termos de uma série de Taylor unidimensional, sem termos exponenciais ímpares. Assim, uma aproximação em torno da origem pode ser escrita como:

$$\hat{\theta}_2(x_2) \approx 53.8256 - 148.0121 \cdot x_2^2 + 307.8418 \cdot x_2^4 - 532.4656 \cdot x_2^6, \tag{17}$$

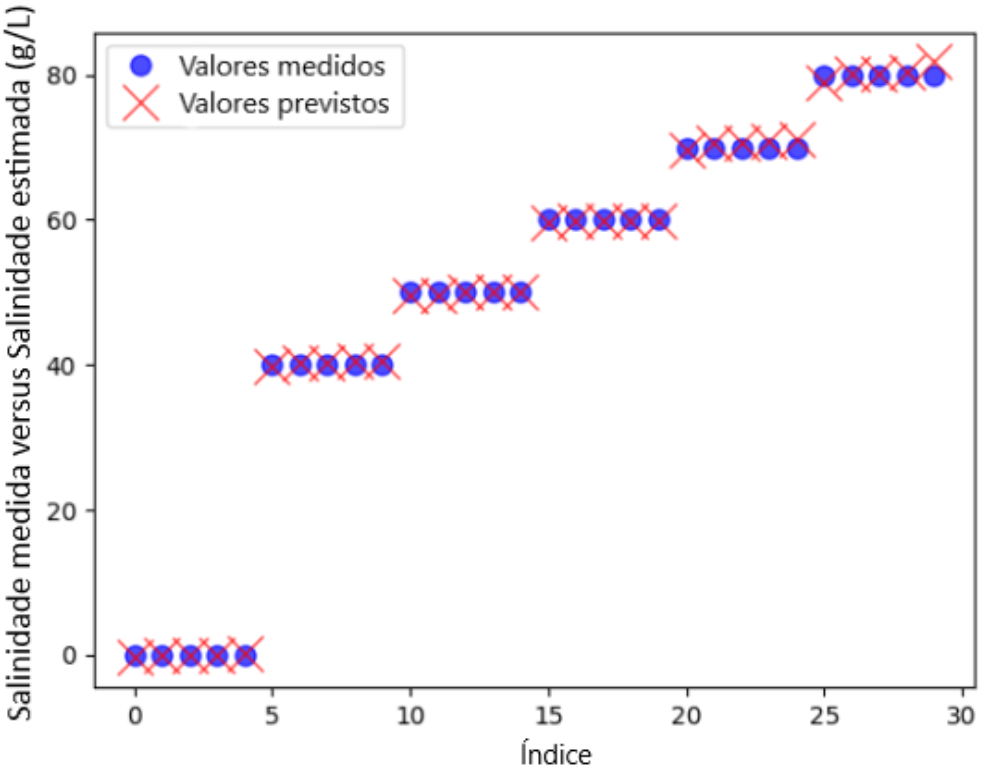



Figura 8 – Valores de salinidade da água medidos e estimados em (g/L) no conjunto de dados de teste. Índice refere-se ao índice dos dados de entrada e saída no conjunto de teste.

Tabela 6 – Avaliação numérica dos resultados da regressão. #i denota a complexidade da equação “número de nós” obtida por meio de regressão simbólica.

| **Equação / Complexidade** | **Erro Médio Absoluto (g/L)** | **Erro Padrão (g/L)** |
|---|---|---|
| [10] | 0.49 | 0.60 |
| #25 | 0.64 | 1.18 |
| **#30** | **0.38** | **0.54** |
| #35 | 0.34 | 0.58 |

Tabela 7 – Comparação entre características das equações. 1 - número de operadores unários; 2 - número de operadores binários; 3 - dimensão de entrada; 4 - número de constantes necessárias.

| **Equação / Complexidade** | **#UO**[1] | **#BO**[2] | **#ID**[3] | **#NC**[4] |
|---|---|---|---|---|
| [10] | 18 | 51 | 4 | 23 |
| #25 | 11 | 6 | 2 | 5 |
| #30 | 9 | 9 | 3 | 7 |
| #35 | 12 | 10 | 3 | 7 |

que requer 4 constantes, 2 subtrações, 1 adição e 6 multiplicações (em uma implementação otimizada). A respeito de plataformas embarcadas comerciais, através de emuladores percebemos que tanto a fórmula (11) quanto a fórmula proposta (13) não são adequados para implementação em sistemas mais simples (como o Arduino Uno de 8 bits ou Arduino Mega), mas são viáveis para implementação em sistemas embarcados de 32 bits, como o ESP32 ou Arduino Due.

A Equação (13) também depende de menos descritores, uma vez que não depende de $x_1$ (ou seja, a potência média). Esse fato ilustra uma vantagem adicional das técnicas de regressão simbólica: sua capacidade de realizar a seleção de recursos, o que mitiga o risco de *overfitting* e reduz o problema de multicolinearidade entre as variáveis. Além disso, remover recursos menos relevantes torna o modelo resultante mais robusto ao ruído ou informações irrelevantes nos dados, levando a previsões mais confiáveis ao facilitar a análise exploratória de dados. Para o especialista, a eliminação de $x_1$ traz um *insight* importante: isso significa que a potência média provavelmente não é um parâmetro significativo para prever a salinidade, quando o centroide espectral, a largura de banda e o fluxo espectral estão disponíveis.

Outra diferença importante é que a Equação (13) não requer normalização dos dados de entrada. Essa é uma vantagem interessante, já que normalizar dados de entrada em técnicas de aprendizado de máquina pode apresentar vários desafios quando o algoritmo é testado em um conjunto de testes ainda não vistos. Esses desafios incluem

o potencial de vazamento de dados, em que a normalização vaza inadvertidamente informações sobre o conjunto de testes, levando a estimativas de desempenho otimistas. Além disso, a normalização pode influenciar a importância relativa de cada recurso para a inferência geral. Tal problema é particularmente verificado se a técnica de normalização (ou padronização) não se ajusta a distribuição de probabilidade subjacente aos dados, que geralmente é desconhecida [120]. Consequentemente, a ausência de normalização torna a regressão proposta mais robusta para prever novos dados observacionais [121].

Finalmente, a Equação (13) decompõe o estimador em três componentes aditivos (enquanto a Equação (11) requer 23 termos), todos os três dependentes de $x_2$ (ou seja, a largura de banda). Este fato enfatiza a largura de banda como um recurso mais relevante para a estimativa prevista. Além disso, a variável $x_0$ influencia apenas um desses termos, assim como o descritor $x_3$. Isso significa que é possível divorciar os efeitos das variáveis $x_0$ e $x_3$. Além disso, a sensibilidade do estimador a estas variáveis pode ser mais facilmente calculada, uma vez que as derivadas parciais em relação a elas são simplificadas.

## 4.4- COMENTÁRIOS FINAIS

Neste capítulo, utilizamos a regressão simbólica para aprimorar um processo de interrogação óptica. Conseguimos obter, como resultado, uma equação mais simples, que utiliza menos descritores e que apresenta um desempenho análogo ao da fórmula proposta em [10], o qual se justifica em virtude da aproximação do erro quadrático médio, constatado experimentalmente. Entendemos que tal resultado sugere, fortemente, a adequabilidade da metodologia proposta (isto é, adoção de técnicas de regressão simbólica conjugada a processos de interrogação óptica).

# 5- CONCLUSÕES E TRABALHOS FUTUROS

A estimativa de salinidade em soluções apresenta grande potencial de aplicação na área industrial. A interrogação óptica surge como uma alternativa para este problema, com algumas vantagens relevantes, tais como: capacidade de sensoriamento remoto, segurança, baixo risco de contaminação à solução e robustez a interferências eletromagnéticas.

Nesta dissertação, foram utilizadas técnicas de regressão simbólica empregadas para aprimorar uma sistema de interrogação óptico baseado em reflectometria no domínio do tempo para prever a salinidade da água. Descobriu-se uma fórmula encontrada via regressão simbólica que prevê a salinidade com mais precisão do que a proposta na literatura pregressa. Sendo mais concisa, a fórmula requer menos constantes, não necessita normalização de dados e fornece alguns *insights* importantes sobre a importância relativa dos descritores. Particularmente, é observado que a potência média é um descritor redundante (considerando os outros descritores presentes no conjunto de dados de entrada) e largura de banda desempenha um papel crucial no processo de interrogação.

Embora a regressão simbólica seja um método que exige recursos computacionais significativos, os autores acreditam que múltiplas execuções do método de regressão simbólica usando o método de validação cruzada K-fold permitiria uma avaliação estatística mais rigorosa quanto à precisão das fórmulas, a importância de cada variável e a presença ou ausência de sobreajuste. Ao analisar a validação cruzada para a regressão simbólica não é tão direta quanto para regressores neurais (já que em cada execução do método as fórmulas resultantes podem ser notavelmente distintas), os autores entendem que tal conjunção permitiria a extração de *insights* importantes para uma abordagem robusta de soluções baseadas em regressão simbólica.

Finalmente, os autores acreditam que as técnicas de regressão simbólica são particularmente promissoras para o aprimoramento dos diferentes processos de interrogação óptica, entendendo que a aplicação descrita neste artigo é um estudo de caso que fortalece essa hipótese. Isto é especialmente notável para técnicas de interrogação alternativas ou novas, considerando a importância que a redução dos custos de interrogação pode ter promovendo uma implantação generalizada de sistemas baseados em sensores

de fibra óptica.

# Referências Bibliográficas

[1] X. Zhang, J. Guan, Y. Liu, and G. Wang, “MORL4PDEs: Data-driven discovery of PDEs based on multi-objective optimization and reinforcement learning,” *Chaos, Solitons & Fractals*, vol. 180, p. 114536, 2024.

[2] M. Krenn, R. Pollice, S. Y. Guo, M. Aldeghi, A. Cervera-Lierta, P. Friederich, G. d. P. Gomes, F. Häse, A. Jinich, A. Nigam, Z. Yao, and A. Aspuru-Guzik, “On scientific understanding with artificial intelligence,” *Nature Reviews Physics*, vol. 4, pp. 761–769, 12 2022.

[3] S. Manti and A. Lucantonio, “Discovering interpretable physical models using symbolic regression and discrete exterior calculus,” *Machine Learning: Science and Technology*, vol. 5, p. 015005, jan 2024.

[4] D. B. Laucelli, L. Enríquez, J. Saldarriaga, and O. Giustolisi, “Using symbolic machine learning to assess and model substance transport and decay in water distribution networks,” *Scientific Reports*, vol. 14, Feb. 2024.

[5] T. Tohme, D. Liu, and K. Youcef-Toumi, “Gsr: A generalized symbolic regression approach,” *Transactions on Machine Learning Research*, Jan. 2023.

[6] D. J. Bartlett, H. Desmond, and P. G. Ferreira, “Exhaustive symbolic regression,” *IEEE Transactions on Evolutionary Computation*, vol. 28, p. 950–964, Aug. 2024.

[7] P. Lemos, N. Jeffrey, M. Cranmer, S. Ho, and P. Battaglia, “Rediscovering orbital mechanics with machine learning,” *Machine Learning: Science and Technology*, vol. 4, p. 045002, oct 2023.

[8] C. Cornelio, S. Dash, V. Austel, T. Josephson, J. Goncalves, K. L. Clarkson, N. Megiddo, B. El Khadir, and L. Horesh, “Combining data and theory for derivable scientific discovery with ai-descartes,” *Nature Communications*, vol. 14, 4 2023.

[9] M. Jucá, *Aplicações de Técnicas Computacionais à Interrogação de Sensores Ópticos Baseados em Grades de Período Longo*. PhD thesis, Universidade Federal de Juiz de Fora/UFJF, 4 2023.

[10] M. A. Jucá, I. V. C. Pereira, P. C. G. Spelta, and A. B. dos Santos, “Water salinity measurement using a long-period grating and optical time-domain reflectometry,” *Journal of the Optical Society of America B*, vol. 40, p. C8, feb 2023.

[11] S. N. Khonina, N. L. Kazanskiy, and M. A. Butt, “Optical fibre-based sensors - an assessment of current innovations,” *Biosensors*, vol. 13, p. 835, Aug. 2023.

[12] M. Elsherif, A. E. Salih, M. G. Muñoz, F. Alam, B. AlQattan, D. S. Antonysamy, M. F. Zaki, A. K. Yetisen, S. Park, T. D. Wilkinson, and H. Butt, “Optical fiber sensors: Working principle, applications, and limitations,” *Advanced Photonics Research*, vol. 3, July 2022.

[13] X. Zhang, C. Wang, T. Zheng, H. Wu, Q. Wu, and Y. Wang, “Wearable optical fiber sensors in medical monitoring applications: A review,” *Sensors*, vol. 23, p. 6671, July 2023.

[14] W. J. Peveler, “Food for thought: Optical sensor arrays and machine learning for the food and beverage industry,” *ACS Sensors*, vol. 9, p. 1656–1665, Apr. 2024.

[15] P. N. S. S. Ja’afar, N. M. Razali, S. Ambran, and F. Ahmad, “Sodium chloride concentration measurement via optical fiber tip sensor,” *IOP Conference Series: Materials Science and Engineering*, vol. 1051, p. 012030, Feb. 2021.

[16] S. Herrera-León, C. Cruz, M. Negrete, J. Chacana, L. A. Cisternas, and A. Kraslawski, “Impact of seawater desalination and wastewater treatment on water stress levels and greenhouse gas emissions: The case of chile,” *Science of The Total Environment*, vol. 818, p. 151853, Apr. 2022.

[17] M. Grossi, “Measurement of water salinity using a capacitively coupled contactless conductivity sensor,” *P-ESEM*, p. 657, 2021.

[18] B. Culshaw and A. Kersey, “Fiber-optic sensing: A historical perspective,” *Journal of Lightwave Technology*, vol. 26, no. 9, pp. 1064–1078, 2008.

[19] C. C. C. Lam, R. Mandamparambil, T. Sun, K. T. V. Grattan, S. V. Nanukuttan, S. E. Taylor, and P. A. M. Basheer, “Optical fiber refractive index sensor for chloride ion monitoring,” *IEEE Sensors Journal*, vol. 9, no. 5, pp. 525–532, 2009.

[20] J. Guo, Y. Ding, X. Xiao, L. Kong, and C. Yang, “Multiplexed static fbg strain sensors by dual-comb spectroscopy with a free running fiber laser,” *Optics Express*, vol. 26, p. 16147, June 2018.

[21] M. A. Jucá, I. V. C. Pereira, D. B. Haddad, and A. B. dos Santos, “Weight classifier using optical time domain reflectometry and a long period grating sensor,” *Optical Fiber Technology*, vol. 67, p. 102694, December 2021.

[22] P. dos Santos, P. Jorge, J. de Almeida, and L. Coelho, “Low-cost interrogation system for long-period fiber gratings applied to remote sensing,” *Sensors*, vol. 19, p. 1500, mar 2019.

[23] R. L. F. Filho, J. R. Almeida, F. O. Barino, D. Campos, A. B. d. Santos, A. Marcato, L. Honorio, C. Duarte, R. Capelini, V. Pamplona, and C. C. d. Santos, “Segurança em usinas hidreletricas: Monitoramento em tempo real utilizando sensores de fibra optica,” *Simpósio Brasileiro de Micro-Ondas e Optoeletrônica*, 2024.

[24] M. Rahimi, M. Malekmohammad, M. Taghavi, M. Noori, and G.-M. Mohammad Parsanasab, “Design and fabrication of a differential moems accelerometer based on fabry–pérot micro-cavities,” *IEEE Sensors Journal*, vol. 22, p. 14779–14785, Aug. 2022.

[25] J. H. Osório, R. Oliveira, S. Aristilde, G. Chesini, M. A. Franco, R. N. Nogueira, and C. M. Cordeiro, “Bragg gratings in surface-core fibers: Refractive index and directional curvature sensing,” *Optical Fiber Technology*, vol. 34, p. 86–90, Mar. 2017.

[26] K. Dey, S. Roy, P. Kishore, M. S. Shankar, R. K. Buddu, and R. Ranjan, “Analysis and performance of edge filtering interrogation scheme for fbg sensor using sms fiber and otdr,” *Results in Optics*, vol. 2, p. 100039, Jan. 2021.

[27] M. T. M. R. Giraldi, C. S. Fernandes, M. S. Ferreira, M. J. de Sousa, P. A. S. Jorge, J. C. W. A. Costa, J. L. C. O. Santos, and O. Frazao, “Fiber loop mirror sensors interrogated and multiplexed by otdr,” *Journal of Lightwave Technology*, vol. 33, p. 2580–2584, June 2015.

[28] J. Yuan, C. Zhao, M. Ye, J. Kang, Z. Zhang, and S. Jin, “A fresnel reflection-based

optical fiber sensor system for remote refractive index measurement using an OTDR," *Photonic Sensors*, vol. 4, p. 48–52, Nov. 2013.

[29] M. Bravo, J. M. Baptista, J. L. Santos, M. Lopez-Amo, and O. Frazão, "Ultralong 250 km remote sensor system based on a fiber loop mirror interrogated by an optical time-domain reflectometer," *Optics Letters*, vol. 36, p. 4059, Oct. 2011.

[30] R. Falate, "Optical bend sensor based on a long-period fiber grating monitored by an optical time-domain reflectometer," *Optical Engineering*, vol. 44, p. 110502, Nov. 2005.

[31] L. Valente, A. Braga, A. Ribeiro, R. Regazzi, W. Ecke, C. Chojetzki, and R. Willsch, "Combined time and wavelength multiplexing technique of optical fiber grating sensor arrays using commercial OTDR equipment," *IEEE Sensors Journal*, vol. 3, no. 1, pp. 31–35, 2003.

[32] R. Yen-Ni Wong, D. Hu Juan Juan, M. Ibsen, and P. Ping Shum, *Applications of Optical Fibers for Sensing*, ch. Optical Fibre Long-Period Grating Sensors Operating at and around the Phase Matching Turning Point. IntechOpen, Apr. 2019.

[33] L. Jin, R. Li, Y. Wang, Y. Tan, and Z. Zhou, "Theoretical error model of fbg-based surface temperature measurement and its accuracy enhancement," *IEICE Electronics Express*, vol. 21, p. 20240098–20240098, July 2024.

[34] T. Komljenovic, M. Bosiljevac, and Z. Sipus, "Cost effective fbg based optical sensor," in *2007 19th International Conference on Applied Electromagnetics and Communications*, IEEE, 2007.

[35] P. S. Macheso and M. Zekriti, "Modelling and analysis of fiber bragg grating temperature sensor for internet of things applications (fbg-4-iot)," *International Journal of Intelligent Networks*, vol. 5, p. 224–230, 2024.

[36] P. J. G. Orr, *Magnetophotonic sensor systems for fibre-optic magnetic diagnostics in tokamak fusion reactors*. PhD thesis, University of Strathclyde, 2011.

[37] J. Orfanidis, *Electromagnetic Waves and Antennas*. Rutgers University, USA, 2016.

[38] A. Cascardo Carneiro and A. P. López Barbero, "Princípio de funcionamento dos sensores ópticos baseados em grades em fibras ópticas e sua importância nas diferentes áreas tecnológicas," *Engevista*, vol. 16, p. 389, Apr. 2014.

[39] G. M. Rego, “Temperature dependence of the thermo-optic coefficient of GeO2-doped silica glass fiber,” *Sensors*, July 2024.

[40] R. Falate, *Produção e Caracterização de Redes de Período Longo para Aplicações em Sensoriamento*. PhD thesis, Universidade Federal de Educação Tecnológica/UFPR, 8 2002. Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial.

[41] Y. A and P. Yeh, *Photonics: Optical Electronic in Modern Communications*. New York: Oxford University Press, USA, 2006.

[42] J. V. d. C. Nascimento, R. L. F. Filho, D. E. D. Campos, F. O. Barino, J. P. E. Ferreira, A. C. d. P. Fernandes, A. B. d. Santos, and W. G. Quirino, “Sensor de umidade baseado em revestimento de Óxido de grafeno rico em carbonila em redes de período longo,” *Simpósio Brasileiro de Micro-Ondas e Optoeletrônica*, 2024.

[43] Y. Chen, L. Yue, X. Yang, C. L. Jones, and Z. Wang, “Superlens-assisted laser nanostructuring of long period optical fiber gratings (lpgs) for enhanced refractive index sensing,” *Optics & Laser Technology*, vol. 176, p. 111001, Sept. 2024.

[44] M. S. Sulaiman, P. Thirunavakkarasu, J.-L. Auguste, G. Humbert, F. S. Roslan, and N. Saidin, “Long period fiber grating for refractive index sensing,” *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 30, p. 154–162, Apr. 2023.

[45] M. S. M. Nor, A. A. Khan, S. Mohamad, and P. Thirunavakkarasu, “Development of optical fiber sensor for water salinity detection,” *International Journal of Technology*, vol. 14, p. 1247, Oct. 2023.

[46] V. Oliveira, A. Cascardo, A. López-Barbero, V. Silva, and R. Ribeiro., “Técnica de interrogação de lpgs auto-referenciável por análise espectral,” in *Anais de XXXV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, SBrT2017, Sociedade Brasileira de Telecomunicações, 2017.

[47] F. Esposito, “Chemical sensors based on long period fiber gratings: A review,” *Results in Optics*, vol. 5, p. 100196, Dec. 2021.

[48] X.-W. Zhao and Q. Wang, “Mini review: Recent advances in long period fiber grating biological and chemical sensors,” *Instrumentation Science & Technology*, vol. 47, p. 140–169, Sept. 2018.

[49] C. A. D. J. Gouveia, *Sensores em fibra Óptica para Monitorização Ambiental*. PhD thesis, Universidade da Madeira, 2008.

[50] H. Patrick, G. Williams, A. Kersey, J. Pedrazzani, and A. Vengsarkar, “Hybrid fiber Bragg grating/long period fiber grating sensor for strain/temperature discrimination,” *IEEE Photonics Technology Letters*, vol. 8, p. 1223–1225, Sept. 1996.

[51] I. A. Avila Castro, Y. H. Chen, B. R. Hayes-Gill, S. P. Morgan, S. Korposh, D. Gomez, and R. Correia, “Ambulatory fibre bragg grating interrogator with applications in blood pressure monitoring,” in *Biophotonics in Exercise Science, Sports Medicine, Health Monitoring Technologies, and Wearables V* (B. Shadgan and A. H. Gandjbakhche, eds.), p. 23, SPIE, Mar. 2024.

[52] H.-Y. Chang, C.-H. Yeh, H. Chuan-Ying, M.-Y. Fu, C.-W. Chow, and W.-F. Liu, “In-fiber long-period grating and fiber bragg grating-based sensor for simultaneously monitoring remote temperature and stress,” *Sensors and Materials*, p. 23, 2018.

[53] C. Berrettoni, C. Trono, V. Vignoli, and F. Baldini, “Fibre tip sensor with embedded FBG-LPG for temperature and refractive index determination by means of the simple measurement of the fbg characteristics,” *Journal of Sensors*, vol. 2015, p. 1–8, 2015.

[54] G. R. C. Possetti, R. C. Kamikawachi, C. L. Prevedello, M. Muller, and J. L. Fabris, “Salinity measurement in water environment with a long period grating based interferometer,” *Measurement Science and Technology*, vol. 20, p. 034003, Feb. 2009.

[55] E. Marin, L. Robert, S. Triollet, and Y. Ouerdane, “Liquid resin infusion process monitoring with superimposed fibre bragg grating sensor,” *Polymer Testing*, vol. 31, p. 1045–1052, Dec. 2012.

[56] D. A. Chamorro Enríquez, A. R. da Cruz, and M. T. M. Rocco Giraldi, “Hybrid FBG-LPG sensor for surrounding refractive index and temperature simultaneous discrimination,” *Optics & Laser Technology*, vol. 44, p. 981–986, June 2012.

[57] N. Ramalingam, H. Yang, K.-S. Lim, and H. Ahmad, “Cantilever-based biaxial FBG inclinometer with low cross-axis sensitivity,” *Optica Applicata*, vol. 54, no. 2, 2024.

[58] D. A. Resen and M. F. Altemimi, “Dual FBGs, one solution: THz radiation and multi-sensing innovation,” *Journal of Optics*, vol. 53, p. 2656–2660, Feb. 2024.

[59] L. Lu, L. jianxin, g. chen, L. Zhu, C. Weiqiang, A. Chunhua, and c. Daping, “Full-range displacement sensor with dual FBGs combined cantilever beam based on magnetic grating,” *Optical Fiber Technology*, June 2024.

[60] M. Faisal, “Fiber bragg grating temperature sensor and its interrogation techniques,” *July 2023*, vol. 4, p. 1–11, July 2023.

[61] D. Aydin, J. A. Barnes, and H.-P. Loock, “In-fiber interferometry sensors for refractive index,” *Applied Physics Reviews*, vol. 10, Feb. 2023.

[62] D. Srivastava, U. Tiwari, and B. Das, “Interferometric interrogation of $\pi$-phase shifted fiber bragg grating sensors,” *Optics Communications*, vol. 410, p. 88–93, Mar. 2018.

[63] S. Korposh, S. James, R. Tatam, and S.-W. Lee, *Current Developments in Optical Fiber Technology*, ch. Fibre-Optic Chemical Sensor Approaches Based on Nanoassembled Thin Films: A Challenge to Future Sensor Technology. InTech, June 2013.

[64] K. Yang, Y. Liu, Z. Wang, Y. Li, Y. Han, and H. Zhang, “Twist sensor based on long period grating and tilted bragg grating written in few-mode fibers,” *IEEE Photonics Journal*, p. 1–1, 2018.

[65] G. M. Mattos, “Sistema de interrogação com compensação de temperatura para sensores baseados em FBG utilizando um atuador termoelétrico,” Master’s thesis, Universidade Federal do Rio de Janeiro, 2018.

[66] A. S. Allil, A. Dante, C. C. Carvalho, R. C. d. S. B. Allil, and M. M. Werneck, “FBG-based sensor applied to volumetric flow rate measurements on flare systems / sensor baseado em FBG aplicado a medições de vazão volumétrica em sistemas de queima,” *Brazilian Applied Science Review*, vol. 5, no. 2, p. 955–964, 2021.

[67] H.-K. Choi, Y.-J. Jung, B.-A. Yu, J.-H. Sung, I.-B. Sohn, J.-Y. Kim, and M. S. Ahsan, “Femtosecond-laser-assisted fabrication of radiation-resistant fiber bragg grating sensors,” *Applied Sciences*, vol. 12, p. 886, Jan. 2022.

[68] M. M. Werneck, *Current Trends in Short- and Long-period Fiber Gratings*, ch. A Guide to Fiber Bragg Grating Sensors. InTech, May 2013.

[69] T. Girerd, F. Mandorlo, C. Jamois, T. Benyattou, L. Ferrier, and L. Berguiga, “Optical sensing based on phase interrogation with a young’s interference hologram using a digital micromirror device,” *Optics Express*, vol. 32, p. 3647, jan 2024.

[70] S. Li, P. Yuan, K. Li, T. Li, and L. Zhu, “Awg-based large dynamic range fiber bragg grating interrogation system,” *Optoelectronics Letters*, vol. 20, p. 129–134, Jan. 2024.

[71] L. H. Silva, P. Santos, L. C. C. Coelho, P. Jorge, and J. M. Baptista, “Development of a long period fiber grating interrogation system using a multimode laser diode,” *Sensors*, vol. 21, p. 749, Jan. 2021.

[72] V. N. H. Silva, V. A. De Oliveira, W. De Souza Gaia Filho, A. B. Do Santos, A. P. L. Barbero, and A. C. Carneiro, “Interrogation technique using harmonic equations in the electrical domain for long period grating sensors,” in *2023 SBMO/IEEE MTT-S International Microwave and Optoelectronics Conference (IMOC)*, p. 46–48, IEEE, Nov. 2023.

[73] P. X. Neto, A. C. Carneiro, A. P. Lopez-Barbero, V. N. H. Silva, R. M. Ribeiro, and A. B. dos Santos, “A mathematical model for the interrogation of lpg fiber optical sensors based on electrical harmonic analysis,” *IEEE Sensors Journal*, vol. 20, p. 4237–4244, Apr. 2020.

[74] Y. Yin, S. Li, J. Ren, G. Farrell, E. Lewis, and P. Wang, “High-sensitivity salinity sensor based on optical microfiber coil resonator,” *Optics Express*, vol. 26, p. 34633, Dec. 2018.

[75] L. Zhang, Z. Zhu, T. Yang, W. Chen, and Y. Dong, “Water temperature and salinity measurement using frequency comb,” *IEEE Photonics Journal*, vol. 13, p. 1–5, Aug. 2019.

[76] M.-Y. Sun, H.-T. Jiang, B. Shi, G.-Y. Zhou, H. I. Inyang, and C.-X. Feng, “Development of fbg salinity sensor coated with lamellar polyimide and experimental study on salinity measurement of gravel aquifer,” *Measurement*, vol. 140, p. 526–537, jul 2019.

[77] M. A. Jucá, D. B. Haddad, and A. B. dos Santos, “Interrogation system for optical

sensor using filter bank and artificial neural network," *Microwave and Optical Technology Letters*, vol. 62, p. 4015–4020, Aug. 2020.

[78] M. Vastl, J. Kulhánek, J. Kubalík, E. Derner, and R. Babuška, "Symformer: End-to-end symbolic regression using transformer-based architecture," *IEEE Access*, vol. 12, p. 37840–37849, 2024.

[79] M. Cranmer, "Interpretable machine learning for science with pysr and symbolicregression.jl," 2023.

[80] S. H. Muggleton, U. Schmid, C. Zeller, A. Tamaddoni-Nezhad, and T. Besold, "Ultra-strong machine learning: comprehensibility of programs learned with ilp," *Machine Learning*, vol. 107, p. 1119–1140, May 2018.

[81] D. Michie, "Machine learning in the next five years," in *EWSL*, 1988.

[82] S. Kubosawa, T. Onishi, and Y. Tsuruoka, "AI and simulation for soft sensors and process control," *KAGAKU KOGAKU RONBUNSHU*, vol. 48, p. 141–151, July 2022.

[83] Z. Huang, Y. Mei, and J. Zhong, "Semantic linear genetic programming for symbolic regression," *IEEE Transactions on Cybernetics*, vol. 54, no. 2, pp. 1321–1334, 2024.

[84] N. Makke and S. Chawla, "Interpretable scientific discovery with symbolic regression: a review," *Artificial Intelligence Review*, vol. 57, Jan. 2024.

[85] W. La Cava, P. Orzechowski, B. Burlacu, F. O. de França, M. Virgolin, Y. Jin, M. Kommenda, and J. H. Moore, "Contemporary symbolic regression methods and their relative performance," 2021.

[86] K. Takaki and T. Miyao, "Symbolic regression for the interpretation of quantitative structure-property relationships," *Artificial Intelligence in the Life Sciences*, vol. 2, p. 100046, Dec. 2022.

[87] L. Keren, A. Liberzon, and T. Lazebnik, "A computational framework for physics-informed symbolic regression with straightforward integration of domain knowledge," *Scientific Reports*, 1 2023.

[88] A. Murati, E. Peluso, G. P, J. Vega, and M. Gelfusa, "Data driven theory for knowledge discovery in the exact sciences with applications to thermonuclear fusion," *Scientific Reports*, 11 2020.

[89] H. Kay, S. Kay, M. Mowbray, A. Lane, C. Mendoza, P. Martin, and D. Zhang, “Constructing a symbolic regression-based interpretable soft sensor for industrial data analytics and product quality control,” *Industrial & Engineering Chemistry Research*, vol. 63, p. 4083–4092, Feb. 2024.

[90] S. Kim, P. Lu, S. Mukherjee, M. S. m. I. Gilbert, L. Jing, V. M. I. Čeperié, and M. Soljačić, “Integration of neural network-based symbolic regression in deep learning for scientific discovery,” *IEEE Transactions on Neural Networks and Learning Systems*, pp. 4166–4177, 9 2021.

[91] A. S. Sambo, R. M. A. Azad, Y. Kovalchuk, V. P. Indramohan, and H. Shah, “Evolving simple and accurate symbolic regression models via asynchronous parallel computing,” *Applied Soft Computing*, vol. 104, p. 107198, June 2021.

[92] H. Vaddireddy, A. Rasheed, A. E. Staples, and O. San, “Feature engineering and symbolic regression methods for detecting hidden physics from sparse sensor observation data,” *Physics of Fluids*, vol. 32, Jan. 2020.

[93] P. Shojaee, K. Meidani, S. Gupta, A. B. Farimani, and C. K. Reddy, “Llm-sr: Scientific equation discovery via programming with large language models,” 2024.

[94] C. Fox, N. Tran, N. Nacion, S. Sharlin, and T. R. Josephson, “Incorporating background knowledge in symbolic regression using a computer algebra system,” 2023.

[95] A. Grundner, T. Beucler, P. Gentine, and V. Eyring, “Data-driven equation discovery of a cloud cover parameterization,” *Journal of Advances in Modeling Earth Systems*, vol. 16, Feb. 2024.

[96] K. W. K. Wong and M. Cranmer, “Automated discovery of interpretable gravitational-wave population models,” 2022.

[97] Y. Li, H. Wang, Y. Li, H. Ye, Y. Zhang, R. Yin, H. Jia, B. Hou, C. Wang, H. Ding, X. Bai, and A. Lu, “Electron transfer rules of minerals under pressure informed by machine learning,” *Nature Communications*, vol. 14, Mar. 2023.

[98] K. T. Matchev, K. Matcheva, and A. Roman, “Analytical modelling of exoplanet transit specroscopy with dimensional analysis and symbolic regression,” 2021.

[99] P. Kidger, “On neural differential equations,” 2022.

[100] S. Verstyuk and M. R. Douglas, “Machine learning the gravity equation for international trade,” *SSRN Electronic Journal*, 2022.

[101] G. Martius and C. H. Lampert, “Extrapolation and learning equations,” *ArXiv*, vol. abs/1610.02995, 2016.

[102] S. S. Sahoo, C. H. Lampert, and G. Martius, “Learning equations for extrapolation and control,” 2018.

[103] A. Meurer, C. P. Smith, M. Paprocki, O. Čertík, S. B. Kirpichev, M. Rocklin, A. Kumar, S. Ivanov, J. K. Moore, S. Singh, T. Rathnayake, S. Vig, B. E. Granger, R. P. Muller, F. Bonazzi, H. Gupta, S. Vats, F. Johansson, F. Pedregosa, M. J. Curry, A. R. Terrel, v. Roučka, A. Saboo, I. Fernando, S. Kulal, R. Cimrman, and A. Scopatz, “Sympy: symbolic computing in python,” *PeerJ Computer Science*, vol. 3, p. e103, jan 2017.

[104] L. Biggio, T. Bendinelli, A. Neitz, A. Lucchi, and G. Parascandolo, “Neural symbolic regression that scales,” in *Proceedings of the 38 th International Conference on Machine Learning, PMLR 139, 2021*, 2021.

[105] A. O. Sodré, “Aplicação de algoritmos de inteligência artificial para determinação de correlações empíricas em modelos de transferência de calor e massa,” 2022.

[106] D. Drikakis and F. Sofos, “Can artificial intelligence accelerate fluid mechanics research,” *Fluids*, vol. 8, p. 212, July 2023.

[107] F. Calabrese, A. Regattieri, R. Piscitelli, M. Bortolini, and F. G. Galizia, “Genetic programming-based feature construction for system setting recognition and component-level prognostics,” *Applied Sciences*, vol. 12, p. 4749, May 2022.

[108] C. Haider, F. de Franca, B. Burlacu, and G. Kronberger, “Shape-constrained multi-objective genetic programming for symbolic regression,” *Applied Soft Computing*, vol. 132, p. 109855, jan 2023.

[109] T. N. Mundhenk, M. Landajuela, R. Glatt, C. P. Santiago, D. M. Faissol, and B. K. Petersen, “Symbolic regression via neural-guided genetic programming population seeding,” 2021.

[110] S. Zhou, B. Yang, S. Xiao, G. Yang, and T. Zhu, “Crack growth rate model derived from domain knowledge-guided symbolic regression,” *Chinese Journal of Mechanical Engineering*, vol. 36, Mar. 2023.

[111] O. A. Carmona Cortes and J. C. d. Silva, “Unconstrained numerical optimization using real-coded genetic algorithms: a study case using benchmark functions in R from scratch,” *Revista Brasileira de Computação Aplicada*, vol. 11, p. 1–11, Sept. 2019.

[112] L. Türkler, T. Akkan, and L. O. Akkan, “Usage of evolutionary algorithms in swarm robotics and design problems,” *Sensors*, vol. 22, p. 4437, June 2022.

[113] A. Hassanat, K. Almohammadi, E. Alkafaween, E. Abunawas, A. Hammouri, and V. B. S. Prasath, “Choosing mutation and crossover ratios for genetic algorithms—a review with a new dynamic approach,” *Information*, vol. 10, p. 390, Dec. 2019.

[114] G. S. I. Aldeia, “Avaliação da interpretabilidade em regressão simbólica,” Master’s thesis, Universidade Federal do ABC, 12 2021.

[115] M. Affenzeller, B. Burlacu, V. Dorfer, S. Dorl, G. Halmerbauer, T. Königswieser, M. Kommenda, J. Vetter, and S. Winkler, *White Box vs. Black Box Modeling: On the Performance of Deep Learning, Random Forests, and Symbolic Regression in Solving Regression Problems*, p. 288–295. Springer International Publishing, 2020.

[116] M. De Florio, I. G. Kevrekidis, and G. E. Karniadakis, “AI-lorenz: A physics-data-driven framework for black-box and gray-box identification of chaotic systems with symbolic regression,” *Chaos, Solitons &; Fractals*, vol. 188, p. 115538, Nov. 2024.

[117] A. Fajar, S. Yazid, and I. Budi, “Comparative analysis of black-box and white-box machine learning model in phishing detection,” 2024.

[118] N. Ahmadi Daryakenari, M. De Florio, K. Shukla, and G. E. Karniadakis, “AI-aristotle: A physics-informed framework for systems biology gray-box identification,” *PLOS Computational Biology*, vol. 20, p. e1011916, Mar. 2024.

[119] D. Brkić, P. Praks, R. Praksová, and T. Kozubek, “Symbolic regression approaches for the direct calculation of pipe diameter,” *Axioms*, vol. 12, p. 850, Aug. 2023.

[120] K. Cabello-Solorzano, I. Ortigosa de Araujo, M. Peña, L. Correia, and A. J. Tallón-Ballesteros, “The impact of data normalization on the accuracy of machine learning algorithms: A comparative analysis,” in *18th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2023)*

(P. García Bringas, H. Pérez García, F. J. Martínez de Pisón, F. Martínez Álvarez, A. Troncoso Lora, Á. Herrero, J. L. Calvo Rolle, H. Quintián, and E. Corchado, eds.), (Cham), pp. 344–353, Springer Nature Switzerland, 2023.

[121] V. Cherkassky and F. M. Mulier, *Learning from data*. Wiley - IEEE, Chichester, England: Wiley-Blackwell, 2 ed., Aug. 2007.

# A- APÊNDICE

## A.1- CÓDIGO PYTHON

```
from pysr import *
import numpy as np

arr = np.loadtxt( "desc.csv" , delimiter = ',' , dtype = float)


X = arr[ : ,  [ 1 , 2 , 3 , 4 ] ]

y = arr[ : , 0 ]

model = PySRRegressor(
    progress=False,
    niterations=200000,
    populations=100,
    population_size=100,
    maxdepth=8,
    annealing=True,
    alpha=0.1,
    maxsize=40,
    warm_start=True,
    model_selection='score', # 'accuracy', 'best', or 'score'
    nested_constraints={"cos":{"cos": 0},"cos":{"sin": 0},"sin":{"cos": 0},
    "sin":{"sin": 0},"exp":{"exp": 0},"log10":{"log10": 0},"erf":{"erf": 0}},
    binary_operators=["+", "*","-","^"],
    unary_operators=[
        "cos",
```

```
        "exp",
        "sin",
        "inv(x) = 1/x",
        "log10",
        "sqrt",
        "sinh",
        "cosh",
         "tanh",
        "atan",
        "asinh",
        "erf",
    ],
    extra_sympy_mappings={"inv": lambda x: 1 / x},
)

model.fit(X, y)
```

## A.2- FÓRMULAS OBTIDAS

As fórmulas obtidas pelo sistema de regressão simbólica, no estilo SymPy, são descritas a seguir. Cada proposta é apresentada no formato $x, y,"z"$, com $x$ sendo a complexidade, $y$ sendo o MSE (no conjunto de treinamento) e $z$ consistindo na fórmula propriamente dita.

```
Complexity,Loss,Equation
1,666.6665,"49.9998"
3,199.08224,"(x1 ^ -0.4818793)"
5,71.383286,"(80.722305 - (x3 * 89312.336))"
6,66.567154,"(70.81623 - (exp(x1) ^ 5528.247))"
7,51.868397,"(74.79686 - exp(sqrt(x3) * 153.17398))"
8,46.006725,"(74.82488 - (sinh(x3 * 2872.4497) * 15.607775))"
9,42.029163,"(75.298935 - (sinh(sinh(x3 * 1636.5731)) * x2))"
```

```
10,34.885628,"(50892.58 - ((sinh(x3 * 3270.7036) + x0) * 11.864567))"
11,3.588265,"(asinh(sinh(sin(log10(x1)) * 55.199825) * 8.6985576e7) + 19.314484)"
13,2.9818504,"(asinh(sinh(sin(log10(x1)) * 55.07948) * 9.967798e7) + (4302.44 - x0))"
15,2.5330517,"((asinh(sin(log10(x1)) * 3.3557775e9) + ((-4328.284 + x0) * log10(x1)))
+ -122.61706)"
16,1.8107678,"(asinh(sinh(sin(log10(x1)) * 55.120033) * 1.6341915e8)
+ (16.934017 - (cos(x0) * 5.580241)))"
17,0.6518277,"(asinh(sinh(sin(log10(x1)) * 51.342773) * 1.4060733e8)
+ (20.700397 - inv(cos(x0 - x2))))"
18,0.48998857,"(asinh(sinh(sin(log10(x1)) * 51.58133) * 1.3050332e8)
+ (21.118917 - asinh(inv(-0.44342083 - sin(x2)))))"
19,0.43575343,"(asinh(sinh(sin(log10(x1)) * 52.427002) * 1.3298419e8)
+ (20.67919 - inv(cos((x0 * 0.7153872) - x2))))"
20,0.410569,"(asinh(sinh(sin(log10(x1)) * 51.71754) * 1.1981211e8)
+ (21.177692 - asinh(inv(-0.4284867 - sin(x1 ^ -0.1971107)))))"
21,0.3633313,"(asinh(sinh(sin(log10(x1)) * 52.77625) * 1.14869464e8)
+ (20.635637 - inv(cos((x0 * 0.56013745) - (x1 ^ -0.21567109)))))"
22,0.35819045,"(asinh(sinh(sin(log10(x1)) * 52.907684) * 1.1328165e8)
+ (20.506447 - inv(sinh(cos((x0 * 0.56013745) - (x1 ^ -0.21562037))))))"
23,0.2926612,"(asinh(sinh(sin(log10(x1)) * 51.677727) * 1.5618773e8)
+ (20.656149 - asinh(inv(cos(-2.0262372 - x2)) - sin(x1 ^ -0.42998666))))"
24,0.21937193,"(asinh(sinh(sin(log10(x1)) * 50.94194) * 1.8341722e8)
+ ((20.876968 - inv(cos((x0 * 0.62737405) - x2))) + sin(x1 ^ -0.429771)))"
25,0.1957585,"(asinh(sinh(sin(log10(x1)) * 51.411648) * 1.5986416e8) +
((19.609394 - inv(cos((x0 * 0.65818006) - x2))) + exp(sin(x1 ^ -0.42918566))))"
27,0.18565384,"(asinh(sinh(sin(log10(x1)) * 51.59716) * 1.5047323e8)
+ ((19.525272 - inv(cos((x0 * 0.6831185) - x2)))
+ sqrt(sinh(exp(sin(x1 ^ -0.42908016))))))"
28,0.18453838,"(asinh(sinh(sin(log10(x1)) * 51.58133) * 1.4917754e8)
+ ((19.519094 - inv(cos((x0 * 0.6831186) - x2)))
+ (sinh(exp(sin(x1 ^ -0.42908016))) ^ 0.5232409)))"
30,0.15076181,"(((asinh(sin(log10(x1)) * 1.8899657e9)
+ ((-45.43336 + cos(x0 * -2.1426246)) * log10(x1))) + -124.98491)
```

```
- (exp(cos(log10(x1) * 106.39591)) * sin(x1 ^ -0.41782326)))"
31,0.14740913,"(((asinh(sin(log10(x1)) * 2.0232978e9)
+ ((-45.437958 + cos(x0 * 2.1294265)) * log10(x1))) + -125.07586)
- (exp(erf(cos(log10(x1) * -106.37373))) * sin(x1 ^ -0.4176216)))"
32,0.14113979,"(((asinh(sin(log10(x1)) * 2.0705837e9)
+ ((-45.437958 + cos(x0 * 2.1294265)) * log10(x1))) + -125.08714)
- ((sinh(cos(log10(x1) * -106.369255)) - -1.1311253) * sin(x1 ^ -0.41759008)))"
33,0.13434596,"(((asinh(sin(log10(x1)) * 2.0511625e9)
+ ((-45.431664 + cos(x0 * -2.1426246)) * log10(x1))) + -125.08697)
- (((cos(log10(x1) * -106.37373) * 1.3084371)
- -1.0785301) * sin(x1 ^ -0.4176297)))"
34,0.13434583,"(((asinh(sin(log10(x1)) * 2.0511625e9)
+ ((-45.431664 + cos(x0 * -2.1426246))  log10(asinh(x1)))) + -125.08697)
- (((cos(log10(x1) * -106.37373) * 1.3084371) - -1.0785301) * sin(x1 ^ -0.4176297)))"
35,0.11753266,"(((asinh(sin(log10(x1)) * 2.0232978e9)
+ ((-45.426105 + cos(x0 * 2.1748984)) * log10(x1))) + -125.05383)
- ((cosh(cos(x1 * 72652.42)) + cos(log10(x1) * 106.37469)) * sin(x1 ^ -0.41759813)))"
36,0.11338293,"(((asinh(sin(log10(x1)) * 1.9488641e9)
+ ((-45.437958 + cos(x0 * -2.171961)) * log10(x1)))
+ -125.073235) - ((cosh(cos(x1 * -72691.18))
+ sinh(cos(log10(x1) * 106.38756))) * sin(x1 ^ -0.41767934)))"
37,0.09415784,"(((asinh(sin(log10(x1)) * 2.0544187e9) +
((-45.406048 + cos(x0 * -2.1382205)) * log10(x1))) + -125.07741)
- ((cosh(cos(73009.57 * x1)) * (cos(log10(x1) * 106.39591) + 0.8357919))
* sin(x1 ^ -0.41747266)))"
38,0.09415781,"(((asinh(sin(log10(x1)) * 2.0544187e9)
+ ((-45.406048 + cos(x0 * -2.1382205)) * log10(sin(x1)))) + -125.07586)
- ((cosh(cos(73009.57 * x1)) * (cos(log10(x1) * 106.39591) + 0.8357919))
* sin(x1 ^ -0.41747266)))"
39,0.08891192,"(((asinh(sin(log10(x1)) * 1.9817834e9)
+ ((-45.430363 + cos(x0 * 2.1836994)) * log10(x1)))
-125.044426)
- ((cosh(cos(73098.02 * x1)) * (cos(log10(x1) * 106.37992) - cos(log10(x1))))
```

```
* sin(x1 ^ -0.41743696)))"
40,0.088911645,"(((asinh(sin(log10(x1)) * 1.9817834e9)
+ ((-45.430363 + cos(x0 * 2.1836994)) * log10(atan(x1)))) + -125.044426)
- ((cosh(cos(73098.02 * x1)) * (cos(log10(x1) * 106.37992) - cos(log10(x1))))
* sin(x1 ^ -0.41743696)))"
```